- The role of outcome expectancy and credibility beliefs in the outcome of Cognitive Behavioural Therapy –

Master’s thesis

Maria Austermann

June 2018

1st Supervisor dr. P.M. ten Klooster
2nd Supervisor dr. P.A.M. Meulenbeek

Faculty of Behavioural Sciences
Positive Psychology and Technology
Abstract

Background. A patient’s outcome expectancy and treatment credibility beliefs are considered as non-specific factors having explanatory value for the outcomes of different psychological approaches. A consensus about their impact on the outcomes of Cognitive Behavioural Therapy (CBT) is missing to date. The present study provides a systematic review on existing findings about the roles of expectancy and credibility in the outcomes of CBT.

Method. The systematic search of articles identified 14 longitudinal studies that measured the relationships between CBT outcomes and expectancy or credibility. Most studies (N = 13) referred to patients with affective disorders. A careful check of the variety of measuring tools utilized to assess expectancy (N = 5) and credibility (N = 5) revealed that two of the 14 studies had interchanged the conceptions of both constructs and that one further study had disregarded their conceptual distinction. Findings from this latter study were subsequently excluded from analysis. The remainder of analysis was conducted with 13 studies, six solely measuring credibility, five solely measuring expectancy and two measuring both constructs.

Results. Weak (e.g. β = -.17) to very strong (e.g. r = -.71) associations between CBT outcomes and expectancy as well as credibility were found. Expectancy appeared to be a robust, moderate predictor of general symptoms of anxiety and was more frequently and more strongly related to the outcomes of CBT for anxiety disorders than credibility. Credibility was more frequently and more strongly related to the outcomes of CBT for anxiety disorders than to the outcomes of CBT for depression. Credibility weakly to moderately predicted very specific outcome variables (e.g. coping skills), whereas its effects on general symptoms of anxiety and depression were only weak or insignificant. One study showed that homework compliance mediated the expectancy-outcome relationship. Another study showed that adherence mediated the credibility-outcome relationship. Single studies further revealed that alliance was moderately related to expectancy and very strongly related to credibility.

Conclusion. Non-specific factors are not as non-specific as originally believed. Expectancy and credibility appear to be related, but distinct, patient characteristics having different explanatory values for the outcomes of CBT that vary across mental disorders and types of outcomes. Both might exert their ameliorative effects through a greater involvement with CBT techniques and may be more meaningful to the outcomes of CBT than the widely accepted non-specific factor alliance. Expectancy could be a more robust predictor of changes in general symptoms than credibility. Yet, credibility seems to be relevant for specific outcome variables which may reflect the effectiveness of specific CBT techniques.

Keywords: outcome expectancy, credibility beliefs, CBT, outcome, non-specific factors
# Table of Contents

1. Introduction ........................................................................................................... 1

2. Method ...................................................................................................................... 6
   2.1 Search procedure .................................................................................................. 6
   2.2 Literature screening ............................................................................................. 7
      2.2.1 Eligibility criteria ......................................................................................... 7
   2.2.2 Phases of Screening ....................................................................................... 8
   2.2.3 Quality assessment ......................................................................................... 9
   2.2.4 Data extraction ............................................................................................... 10

3. Results ..................................................................................................................... 11
   3.1 Study characteristics ........................................................................................... 11
      3.1.2 Descriptive characteristics ......................................................................... 12
      3.1.3 Methodological characteristics .................................................................... 14
   3.2. Answer to the research questions ...................................................................... 25
      3.2.1 Answer to research question 1 and research question 2 .............................. 25
      3.2.2 Answer to research question 3 ..................................................................... 29

4. Discussion ............................................................................................................... 31

References ................................................................................................................... 37

APPENDIX .................................................................................................................. 45
1. Introduction

For psychotherapy to be as effective as possible, patients should have an inner mobilization of hope for positive change and confidence in the treatment they enter (Frank, 1961; Kirsch, 1978).

CBT has been designed under the assumption that psychopathological symptoms and related distress are a product of cognitive misappraisals and dysfunctional behaviours. Practitioners of CBT aim to change these conditions by teaching new information-processing skills and coping strategies (Meichenbaum, Carlson & Kjos, 2007; Field, Beeson & Jones, 2015). For this objective, they can draw on a variety of evidence-based techniques such as exposure, psychoeducation and cognitive restructuring (Meichenbaum et al., 2007).

CBT has proved to be effective in the treatment of most mental disorders (Steinert & Leichsenring, 2016) and has yielded superior outcomes to alternative forms of psychotherapies in numerous meta-analytic studies (e.g. Hofmann, Asnaani, Vonk, Sawyer & Fang, 2012; Hunsley & Di Giulio, 2002; Tolin, 2010). However, up to now, it is not fully clear why CBT works so well and why some patients benefit more from this treatment than others (Tolin, 2010).

Investigators present well-founded arguments to conclude that the outcome of CBT is mainly a result of its specific techniques (e.g. O'Donohue, Fisher, 2008; Tolin, 2010). However, there is also some evidence suggesting that the outcome of any treatment approach is based less on its respective techniques but more on non-specific factors, shared by all forms of psychotherapy (e.g. Ahn & Wampold, 2001; Laska, Gurman & Wampold, 2014). These non-specific factors can be classified into the broad categories of treatment structure, change processes, therapist qualities, therapeutic relationship variables and patient characteristics (Greencavage & Norcross, 1990).

Some non-specific factors are well-studied in the context of CBT. For example, it is known that therapeutic relationship variables (e.g. alliance) as well as several clinical (e.g. symptom severity), demographic (e.g. gender) and personal (e.g. religiosity, personality) patient characteristics have robust explanatory value for the outcomes of CBT (Linden, 2008; Bowen, Baetz & D'arcy, 2006; Olatunji, Davis, Powers & Smits 2013; Spek, Nykliček, Cuijpers & Pop, 2008). By contrast, less is known about how, and to which extent, the outcomes of CBT are determined by a patient’s perceptions on the credibility of CBT as well as his expectancy of the symptom improvement that will be achieved by this treatment.
Indeed, *credibility beliefs* and *outcome expectancy* have long been neglected, wrongly conceptualized or flawly measured in psychotherapy research (Weinberger & Eig, 1999) and a consensus about their impact on the outcomes of psychological treatments is missing to date. In other fields like medicine, however, it is already widely accepted that these malleable patient characteristics have a significant impact on frequently measured outcomes, such as those referring to pain, function and quality of life (Shapiro, 1971; Licciardone & Russo, 2006; Weinberger & Eig, 1999).

Meanwhile there is a growing assumption that expectancy and credibility are not only associated with adaptive treatment processes and outcomes of CBT (Constantino, Penek, Bernecker & Overtree, 2014), but also provide explanations for the superior specific effects of this treatment (Dew & Bickman, 2005; Kirsch, 1978). Furthermore, a greater number of psychotherapy outcome studies, considering measurements of both constructs, exist (Newman & Fisher, 2010). This enables to compile a holistic picture of findings about the respective roles of expectancy and credibility in the outcomes of CBT and simultaneously gives an opportunity to clarify the proper conceptions and measurements of both constructs.

Outcome expectancy reflects a patient’s expected symptom improvement by receiving a specific psychological treatment and constitutes an important influencing factor for the decision to start a therapy (Constantino, 2012; Vogel, Wester, Wei & Boysen, 2005). A patient who enters a therapy is assumed to expect positive change (Jacobson & Baucom, 1977). The magnitude of this expectancy is not only determined by the patient’s clinical characteristics (e.g. symptom severity, psychological mindedness) but is also presumed to develop, to a greater degree, upon how *credible* the treatment seems (Constantino et al., 2014; Tsai, Ogrodniczuk, Söchting & Mirmiran, 2014). Credibility beliefs reflect a patient’s perceptions about the logic, suitability and plausibility of a specific treatment and have been proved to strongly correlate with outcome expectancy (Ametrano, 2011; Safren, Heimberg & Juster, 1997). However, due to their strong correlation and the fact that both reflect perceptions about a specific treatment, some disagreement exists about the extent to which expectancy and credibility overlap (Constantino, 2012). Although theoretically separable, the factors were often used interchangeably in the literature (Sandell, Clinton, Frövenholt & Bragesjö, 2011; Jacobson & Baucom, 1977).

Jacobson and Baucom (1977) were one of the first who clarified that outcome expectancy and credibility beliefs are likely distinct. Whereas they viewed expectancy as an organismic variable which refers to the “assessment of the probability that a particular treatment will help” (p. 714) and therefore reflects an individual’s perception about *himself,*
credibility, on the other hand, was considered to be “a statement about a treatment procedure rather than a judgement as to that procedure’s probable effects on a given individual” (p. 714) that reflects a perception about the treatment itself.

In more recent literature, the distinction of expectancy and credibility is mainly explained by emphasizing that both constructs are products of different processes (e.g. Ametrano, 2011; Constantino, 2012; Devilly & Borkovec, 2000). More specifically, expectancy is understood to be a product of affective processes that are akin to hope or faith and represent “what a patient feels will happen” (Ametrano, 2011, p. 3). Credibility, by contrast, is assumed to be a product of cognitive processes, that are akin to logical reasoning and represent “what a patient thinks will happen” (Ametrano, 2011, p. 3). A further difference is that estimations about the credibility of a treatment cannot develop until a patient has experienced, at least, a sample of the treatment or gained some understanding of its rationale, whereas expectations of potential improvements can exist prior to learning about a treatment’s characteristics (Constantino, 2012). However, this early expectancy has been found to change by experiencing the practical application of a treatment rationale, which supports the assumption that expectancy partially develops upon how credible the treatment seems (Ametrano, 2011).

Empirical evidence for the distinction of credibility and expectancy is provided by the finding that the two constructs differ in their relationships with clinical outcome variables. Despite the fact that credibility has been demonstrated to have a significant impact on treatment outcomes such as quit-smoking success in nicotine-replacement therapy (Titter, Fitzgeorge, De Jesus, Harper & Prapavessis, 2014) as well as post-treatment levels of anxiety, general psychopathological symptoms and depression in psychodynamic therapies (Mooney, Gibbons, Gallop, Mack & Crits-Christoph, 2014), it is assumed to be less frequently related to the outcomes of psychotherapies than expectancy (e.g. in Ametrano, 2011; in Devilly & Borkovec, 2000). Nevertheless, this assumption is solely based on findings from clinical outcome studies measuring both constructs in homogenous samples of patients treated for anxiety disorders (Borkovec & Costello, 1993; Borkovec & Mathews, 1988, Ametrano, 2011). The extent to which the impacts of expectancy and credibility might differ between psychological disorders, is unknown and merits systematic investigation.

The latter applies also to the question if the effects of expectancy and credibility vary across treatment approaches. Researchers often merged data from patients of different treatment conditions to assess relationships between outcomes and expectancy as well as credibility (e.g. Borkovec & Costello, 1993; Borkovec & Mathews, 1988). However, non-
specific factors are assumed to differ in their importance for psychotherapies (Dew & Bickman, 2005) which is why their impacts on outcomes should be examined for each psychological treatment separately.

The notion that expectancy and credibility could have explanatory value for the (superior) outcomes of CBT inter alia bases on the finding that both patients with a mental illness (Irankunda & Heatherington, 2017; Wanigaratne & Barker, 1995) and therapy-inexperienced people from the public (Frövenholt, Bragesjö, Clinton & Sandell, 2007) ranked CBT as the most promising and credible treatment approach after receiving information on the rationales of different treatments (e.g. psychodynamic, cognitive therapy, pharmacotherapy, CBT). In contrast to therapy rationales which do not provide for a fixed treatment content and rely on theoretical constructs that are difficult to prove (e.g. in psychodynamic therapy) (Jacobson, 2013), the CBT rationale “inform[s] clients what treatment will be like, describe[s] the mechanisms of action underlying the treatment, instil[s] confidence that treatment will be beneficial, and align[s] agreement between the client and the therapist on the goals and tasks of therapy” (Newman & Fisher, 2010, p. 2), in this way likely targeting a more positive manifestation of expectancy and credibility. These manifestations, in turn, may function as mechanisms of positive change and could be directly or indirectly related to the outcomes of CBT (Newman & Fisher, 2010). Interestingly, the theories supporting the respective potential change mechanisms of expectancy and credibility partially differ from each other, again emphasizing the related but distinct characteristics of both constructs.

According to the social influence theory (Strong & Claiborn, 1982), psychotherapy is “an interpersonal influence process whereby therapists gain influence through establishing credibility with clients and subsequently use that influence to bring about desired change in client behaviour and ways of thinking” (in Söchting, Tsai & Ogrodniczuk, 2016, p. 8). In this connection, credibility would constitute a premise to fully comprehend and incorporate the techniques of CBT and might therefore be directly related to the outcomes of this treatment. As expectancy influences the extent to which a patient gets involved with therapy (Vogel et al., 2005), it could represent a further premise to exhaust the possibilities of CBT and might also have a predictive effect on the outcomes of this treatment. With reference to the perceptual processes that are assumed to underly expectancy and credibility, it could further be assumed that credibility might be more strongly related to outcome variables reflecting potential changes in cognitive interpretations such as general thoughts about the self and others (Boelen, van den Bout & van den Hout, 2003), whereas expectancy might rather be
more strongly related to affective outcome variables such as depression and anxiety.

The ameliorative effects of expectancy and credibility could also be exerted through working alliance. Alliance is a widely accepted predictor of psychotherapy outcomes (Horvath & Symonds, 1991). Credibility is used by the therapists to establish a productive alliance (Strong & Claiborn, 1982) and a positive outcome expectancy is assumed to facilitate the build-up of this working relationship (Joyce & Piper, 1998). Alliance could therefore exert a mediating effect on the relationships between CBT outcomes and expectancy as well as credibility.

Furthermore, both constructs could be related to adherence and homework compliance, variables essential for the successful application of CBT (Schmidt & Woolaway-Bickel, 2000). This hypothesis is supported by the expectancy-value theory, which assumes that an individual’s predictions about the potential outcomes of his actions as well as the value he places on these actions, strongly influence the likelihood to which these actions are executed (Wigfield, 1994, Parjares, 1996). A patient highly valuing to achieve symptom improvement, could therefore be more compliant and adherent to therapy and hence might have an increased likelihood to experience a positive change. Likewise, perceiving a rationale as credible could increase an individual’s engagement in therapy (Elkin, Yamaguchi, Arnkoff, Glass, Sotsky & Krupnick, 1999), which supports the hypothesis that adherence and homework compliance might also mediate the credibility-outcome relationship in CBT.

Since CBT became increasingly popular and is known as efficacious treatment (Gaudiano, 2008), some researchers even hypothesized that patients could simply improve due to the fact that they view CBT as highly credible and hence, expect that it will be helpful (Gaudiano, 2008; Newman & Fisher, 2010). In this case, patients with higher expectancy and credibility would improve in CBT regardless of the content and scope of treatment or the therapist employed, and this improvement would presumably occur within a disproportionately short time period.

The abovementioned findings and assumptions emphasize that expectancy and credibility are potentially directly and indirectly related to the outcomes of CBT. A comprehensive investigation of existing findings on these relationships could further increase the understanding of why CBT can exert ameliorative, or even superior, effects on an individual’s mental health.

The research questions (RQ) defined for the purposes of this systematic review were formulated as follows:
RQ 1: In which way and to which extent are outcome expectancy and credibility beliefs related to the outcome of CBT?

RQ 2: Do the effects of outcome expectancy and credibility beliefs on the outcomes of CBT vary across psychological disorders?

RQ 3: Do outcome expectancy and credibility beliefs differ in their relationships with different outcomes of CBT?

2. Method

In the present systematic review, findings from studies which had measured the impacts of outcome expectancy and credibility beliefs on the outcomes of CBT, were comprehensively gathered and examined. In accordance with evidence-based research, the protocol and methodology of this study followed the “Preferred Reporting Items for Systematic Reviews and Meta-Analysis guidelines” (PRISMA) of Moher, Liberati, Tetzlaff and Altman (2009).

2.1 Search procedure

The systematic database search was conducted in Scopus, PubMed and Web of Science in spring 2018 (March – April 2018). Considering the research questions, the key index terms used for the database searches referred to concepts of “CBT”, “outcome expectancy”, “credibility beliefs”, “outcome”, “prediction”, “moderation” and “mediation”. As an initial scan of index lists revealed that most of the key index terms could have been identified with different labels, truncation terms were used to receive a greater number of relevant records. To arrange comprehensive search queries, the Boolean operators “AND” as well as “OR” were used to combine all labels. Two search queries were compiled.

The first search query was designed to conduct an automatized search of relevant titles, abstracts and keywords in the mentioned databases and was defined as: ((CBT OR “Cognitive Beh* Therapy“) AND (outcome* OR effect*) AND (predict* OR moderat* OR mediat*) AND (credib* OR expect*)). To assure that all relevant records were included in the final data pool, a second search query was compiled to run an automatized full-text screening on all existing labels referring to outcome expectancy or credibility beliefs. This second query was defined as: ((CBT OR “Cognitive Beh* Therapy“) AND (outcome* OR effect*) AND (predict* OR moderat* OR mediat*) AND (“outcome expect*“ OR “treatment expect*“ OR “credibility beliefs“ OR “perceived credibility“ OR “treatment credibility“ OR “rationale credibility“)). Labels inside the last bracketing of this query were selected to be searched in full-texts instead of title, abstract and keywords only.
Subtle adaptations were made for the two search queries with regard to the respective operating principles of the different databases. No limits were applied to the date of publication.

2.2 Literature screening

2.2.1 Eligibility criteria
The eligibility criteria of the current literature review were partially adapted from a systematic review conducted by Lingiardi, Muzi, Tanzilli and Carone (2018) which addressed the role of therapists characteristics in psychodynamic therapy outcomes. Records were considered eligible for this systematic review if they had presented quantitative statistical analyses providing insights about the roles of outcome expectancy and/or credibility beliefs in the outcomes of CBT.

In specific, articles had to conform to the following criteria: a) the abstract had to be available; b) the full text of the records had to be available in English; c) original data had to be reported, or in case of a secondary analysis, reanalysed data had to be retrieved from a single rather than multiple studies; d) participants had to be diagnosed with - or had to show symptoms of a mental or behavioural disorders listed in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-IV); e) participants had to be aged 18 years or older; f) studies had to be either observational CBT outcome studies or had to employ at least one experimental condition that included CBT (individual or group-based) as treatment modality under investigation with samples larger than 1; g) studies had to employ a “pure” version of CBT; h) data relevant to the relationships between outcome expectancy and/or credibility beliefs and the outcomes of CBT had to be reported in terms of effect sizes; i) expectancy and credibility had to be assessed as separate constructs; j) at least some information had to be provided about the measuring tools used to assess expectancy or credibility.

Studies were excluded if they a) were not published in English language; b) pertained to a single case series, were qualitative, cross-sectional, a meta-analysis, a research proposal, a book chapter, a review article or a correction paper; c) employed populations without observable or diagnosable symptoms of disease (e.g. a random student sample) or subjects diagnosed with neurodevelopmental disorders (e.g. ADHD), neurocognitive disorders (e.g. dementia) or psychological problems due to medical conditions (e.g. depression in cancer patients); d) did not specify that CBT had been employed; e) only employed internet-based CBT, hybrid CBT versions (e.g. mindfulness based CBT) or just a single technique picked from the CBT approach (e.g. exposure); f) involved participants under 18 years; g) focused on
expectancies and beliefs other than outcome expectancy and credibility beliefs; h) did not report data that specified the respective roles of expectancy and/or credibility in the outcomes of CBT (e.g. comparative studies of CBT and other treatments that combined data from different conditions to assess relationships between expectancy or credibility and outcomes); j) did not separate expectancy and credibility when assessed together (e.g. mean scores of scales measuring both constructs were not stratified by factors); k) did not provide any information concerning the measuring tools used to assesses expectancy and credibility.

When studies comprised insufficient information to assess whether the eligibility criteria had been met, they were excluded from the review.

2.2.2 Phases of Screening

Step 1. After the database query had been conducted by using the defined search terms, all records obtained were incorporated into the software tool Endnote (Clarivate Analytics, 2018), which was designed to manage bibliographies, citations and references. The pool of records was then de-duplicated through carrying out an automatized record scan via the Endnote software.

Step 2. The abstracts of remaining records were assessed on the eligibility criteria. Records that were clearly not related to the purpose of this study were removed.

Step 3. The eligibility criteria were again utilized to screen the full-texts of the records left. Once more, records that did not comply with the eligibility criteria were removed. Records, which full-texts had not been made accessible by any database on the internet, were also excluded from further analysis.

Step 4. Reference lists of the remaining records were screened on articles that had not been found by use of the search queries but might contain data relevant for the purpose of the present study.

The flow diagram of the screening phases and resulting numbers of articles is illustrated below (Figure 1).
2.2.3 Quality assessment

A quality assessment was conducted to enable systematic judgements of the characteristics of the outcome studies included in the final data pool. Quality criteria were partially adapted from Lingardi et al (2018). Eight criteria were set. Those referred to the quality of sample sizes utilized; psychotherapists employed; drop-out analyses conducted; measuring tools and measurement points used to assess relationships between outcome or process variables and expectancy or credibility; and statistical analyses performed to measure the strength of
relationships. Scores for each criterion ranged from 0 to 1 and 0 to 2, respectively with higher scores reflecting higher quality. A maximum sum-score of $\Sigma = 10$ was possible. An overview of the quality criteria and scores obtained are provided in the Appendix (Table 1 and Table 2 in Appendix A).

2.2.4 Data extraction

After records included in the final data pool had been read in detail, two tables were set to depict the main characteristics of each study. Table 1 provides descriptive study characteristics like each study’s author(s), year of publication, country, design, patient sample and therapists. Table 2 provides methodological study characteristics, clarifying which measuring tools and measurement points were utilized to assess expectancy and credibility, which outcome variables were considered, and which statistical analyses were conducted, including the effect sizes found.

Cohen’s (1992) heuristics were used to interpret the magnitude of effect sizes (e.g. Pearson’s $r$, Cohen’s $d$, $\beta$). According to Cohen (1992), a $r \geq 0.1$ as well as a $d \geq 0.2$ represent small effect sizes, a $r \geq .3$ as well as a $d \geq 0.5$ represent medium effect sizes and a $r \geq 0.5$ as well as a $d \geq 0.8$ represent large effect sizes. Rosenthal’s (1996) suggestions to interpret a $r \geq 0.7$ as well as a $d \geq 1.3$ as very large effect sizes were also considered. This enabled a more precise interpretation and comparison of effect sizes found. Based on the recommendation of Stellefson, Hanik, Chaney and Chaney (2008), standardized $\beta$ coefficients were interpreted by utilizing Cohen’s (1992) heuristics for zero-order correlations.
3. Results

Reasons for record rejection are illustrated in Figure 1. One thousand twenty-six records were obtained through the initial database search. After 521 duplicates were removed from the data pool, remaining unique records from the initial search (N = 505) were screened on titles, abstracts and keywords to identify studies eligible for full-text retrieval. At this stage, 353 records were excluded due to not meeting the eligibility criteria. Main reasons for record rejection in this phase were that studies did not employ CBT or only employed hybrid versions of CBT (N = 103), were a review (N = 52), comprised subjects with physical illnesses (N = 44) or children and adolescents (N = 43).

The remaining records (N = 152) were retrieved for full-text screening. Of those, 139 records were excluded since full-text screens revealed they did not meet the eligibility criteria. Main reasons for record rejection at this stage were that studies did not measure expectancy or credibility (N = 80), measured other expectancies and beliefs (N = 23) or merged data from CBT patients and patients of other treatment conditions to assess relationships between outcome expectancy or credibility beliefs and treatment outcomes (N = 12). Moreover, seven further records were excluded because their full-texts were not freely available. Reference lists of remaining full-text records were screened and one additional study was found that matched with the eligibility criteria of the present investigation (Dozois & Westra, 2005). Finally, 14 records were suitable for further qualitative analysis. These records were asterisked in the reference list.

3.1 Study characteristics

3.1.1 Total scores obtained in quality assessment

No study obtained the full score of 10 points in the quality assessment (see Table 2 in Appendix A). Total scores ranged from three to eight points. The lowest total score was obtained by Andersson, Carlbring and Grimlund (2008) (score = 3), as the researchers only considered a sample of small size (N = 25), did not provide data on drop-out comparisons and utilized an unvalidated measuring tool to assess credibility. The highest total score (score = 8) was obtained by Chambless et al. (2017) but the researchers failed to perform drop-out analyses and partially employed unlicensed therapists for the implementation of CBT. Most of the remaining studies (N = 7) obtained moderate scores (score = 5 to score = 7) in the quality assessment.
3.1.2 Descriptive characteristics

Country, year of publication. Most of studies were conducted in the USA (N = 6) and in Canada (N = 5). The remaining studies were conducted in the UK (Hardy et al., 1995), in Sweden (Andersson, Carlbring & Grimlund, 2008) and in Australia (Riley, 2015). All studies were published in English language between 1995 (Hardy et al.) and 2018 (Vislă, Constantino, Newkirk, Ogrodniczuk & Söchting). Many authors were involved in two or more studies included in the final data pool.

Study design. Half of the studies (N = 7) were experimental and conducted as randomised controlled trial (RCT). The other studies were observational. Nearly all studies applied strict in- and exclusion criteria and were categorized as highly selective. An exception was the observational study of Webb, Beard, Auerbach, Menninger and Björgvinsson (2014), which was rather unselective in nature as in- and exclusion criteria had been exclusively deployed for the diagnosis of the sample.

Four studies utilized pretest, posttest, follow-up designs. Three studies adopted a multiple repeated measure design, meaning that assessments had been performed during the whole course of treatment. That was either at each day (Westra, Dozois & Marcus, 2007) or at some selected days of treatment (Webb et al., 2014; Westra, Constantino & Aviram, 2011). The remaining studies either utilized standard pretest, posttest designs (Dozois & Westra, 2005; Andersson, Carlbring & Grimlund, 2008) or incorporated additional measurement points at mid-treatment (Söchting, Tsai & Ogrodniczuk, 2016; Meyerhoff & Rohan, 2016; Vislă et al., 2018) as well as at mid-treatment and follow-up (Chambless, Tran & Glass, 1997; Chambless et al., 2017).

Sample size. Total sample sizes ranged from 38 (Westra, Dozois & Marcus, 2007) to 177 (Meyerhoff & Rohan, 2016). Sample sizes for the CBT conditions alone ranged from 25 (Andersson, Carlbring, & Grimlund, 2008) to 150 (Hundt et al., 2014). Most of studies (N = 10) obtained a moderate quality score for the size of their sample (score = 1, range: 0 – 2).

Age. The reported mean age ranged from 35 (SD = 9.7) (Chambless, Tran & Glass, 1997) to 67 (SD = 5.9) (Hundt et al., 2013). Three studies did not provide data on age but reported that samples comprised adults only (Hundt et al., 2014; Riley, 2015; Vislă et al., 2018). Two studies focused on older adults explicitly (Hundt et al., 2013; Hundt et al., 2014).

---

1 Eight of the studies included in the final data pool comprised secondary analyses of existing data. Original studies were screened on information useful for the purpose of the present review, but primary analyses did not present data providing as detailed insights in the relationships between expectancy or credibility and outcomes as secondary analyses, for example since only baseline scores of expectancy or credibility assessments were reported. Therefore, most of original studies were not further considered in the qualitative synthesis.
Gender. Except from the two studies who did not specify the gender of subjects (Riley, 2015; Vîslă et al., 2018), in nearly all study samples, females were more strongly represented than males. Only the sample of Andersson, Carlbring and Grimlund (2008) comprised slightly more males (N = 30) than females (N = 19).

Diagnosis. Most of study samples were composed of individuals with specific types of anxiety disorders (N = 8), such as generalized anxiety disorder (GAD) (Dozois & Westra, 2005; Westra, Dozois & Marcus, 2007; Westra, Constantino & Aviram, 2011), panic disorder (PD) with (Westra, Dozois & Marcus, 2007; Chambless et al., 2017) or without agoraphobia (Andersson, Carlbring & Grimlund, 2008) and social anxiety disorder (SAD) (Chambless, Tran & Glass, 1997; Westra, Dozois & Marcus, 2007). Next to anxiety disorders, a number of studies focused on patients with major depressive disorder (MDD) (N = 5). Of those, one study sample was composed of patients with the specific subtype of seasonal affective disorder (SEAD) (Meyerhoff & Rohan, 2016). The only study that did not refer to affective disorders was provided by Riley (2015), who had focused on a sample of patients with pathological gambling disorder (PGD).

Treatment. Either individual CBT (N = 7) or group CBT (N = 7) was employed. Most studies reported that CBT had been conducted in accordance with the CBT manual (N =10). A few studies employed flexible CBT (Webb et al., 2014), exposure-based (Chambless, Tran & Glass, 1997; Riley, 2015) or skill-based versions of CBT (Hundt et al., 2013).

Scope of treatment. The quantity of total treatment sessions ranged from five (Webb et al., 2014) to 24 sessions (Chambless et al., 2017). However, slightly more than half of the studies conducted CBT within the context of ten sessions (N = 8). Only Chambless, Tran and Glass (1997) did not provide data on the number of CBT sessions conducted.

Therapists. In five studies, CBT was conducted by experienced mental health professionals, clinical psychologists or psychiatrists exclusively. These studies obtained the full score in the quality assessment of therapist (score = 2, range: 0 - 2). The remaining studies either received a zero score in the quality assessment as CBT had been solely performed by unlicensed therapists (e.g. psychology students) (Chambless, Tran & Glass, 1997; Meyerhoff & Rohan, 2016; Westra, Constantino & Aviram, 2011; Hundt et al., 2013; Riley, 2015) or they received a moderate score as they had deployed mixed groups of licensed and unlicensed therapists for the implementation of CBT (Andersson, Carlbring & Grimlund, 2008; Webb et al., 2014; Vîslă et al., 2018). Only Hundt et al., (2014) did not report about the therapists who had executed the CBT treatment provided.
3.1.3 Methodological characteristics

Expectancy and credibility measuring tools. Expectancy was either assessed with the expectancy subscale of the credibility/expectancy questionnaire (CEQ) from Devilly and Borkovec (2000), with the expectancy subscale of the expectancy rating scale (ERS) from Hundt et al. (2013), with the credibility/expectancy scale (CES) from Borkovec and Nau (1972), with the anxiety change expectancy scale (ACES) from Dozois and Westra (2005) or with the outcome expectancy scale (OES) from Ogrodniczuk and Söchting (2010). Credibility was either assessed with the credibility subscale of the CEQ (Deville & Borkovec, 2000), with the credibility subscale of the ERS (Hundt et al., 2013), with the CES (Borkovec & Nau, 1972), with the credibility scale (CS) from Söchting, Tsai and Ogrodniczuk (2016), with the treatment credibility form (TCF) from Morrision and Shapiro (1987) or with the treatment-endorsement scale of the opinions about psychotherapy questionnaire (OPP) from Pistrang and Barker (1992).

Of all questionnaires utilized, the ACES and CEQ were the only ones that had been assessed on psychometric properties. Both questionnaires are reliable and validated measuring tools (Dozois & Westra, 2005; Devilly & Borkovec, 2000). Six studies had deployed one of these questionnaires and thus obtained the full score in the quality assessment of measuring tools (score = 1, range: 0 -1). The majority of these studies focussed on expectancy. A reliable and validated measurement of credibility was only provided by Riley (2015). An overview of questionnaires utilized is provided in the Appendix (Table 3 in Appendix A).

Internal consistency of measuring tools. When reported, internal consistency of the scales ranged from unacceptable ($\alpha < 0.5$) to excellent ($\alpha \geq 0.9$) (Cronbach, 1951). Internal consistencies of the CEQ-E ($\alpha = .89$), OPP ($\alpha = .87$), CS ($\alpha = .89$) and ERS-C ($\alpha = .89$) were good. Internal consistency of the OES was good ($\alpha = .89$) to excellent ($\alpha = .92$). Excellent internal consistency was also found for the ACES ($\alpha = .91$). Internal consistency of the CES ranged from unacceptable ($\alpha = .39$) to good ($\alpha = .85$) to excellent ($\alpha = .91$). However, the low alpha value for the CES was only found when credibility had been assessed at post-treatment, a measurement point conflicting with the conception of this construct. No alphas were reported for the TCF, ERS-E and CEQ-E scales.

Content of measuring tools. To try to increase the validity of the present research, it was examined if the various questionnaires deployed, considered the established definitions as well as conceptual distinctions of outcome expectancy and credibility beliefs. For this, scales used to measure expectancy and credibility were comprehensively studied. Their respective contents were compared with the contents of the validated questionnaires ACES and CEQ as
well as with the conceptions of outcome expectancy and credibility beliefs (see Appendix B).

The main findings were that the CES, which had either been utilized to assess expectancy (Chambless, Tran & Glass, 1997; Meyerhoff & Rohan, 2006) or credibility (Andersson, Carlbring & Grimlund, 2008), solely contained items matching to the conceptualization of credibility. Consequently, in the present research, results from the studies of Chambless, Tran and Glass (1997) as well as Meyerhoff and Rohan (2006) were interpreted as findings on credibility instead of findings on expectancy. Moreover, it became apparent that the OES - a three-item scale used for the assessment of outcome expectancy - incorporated only one item measuring expectancy and two items measuring credibility (see Appendix B). As the good to excellent internal consistency of the OES can be explained with the relatedness of expectancy and credibility and the scale has yet not been assessed on psychometric properties, it was decided to exclude findings from Vislă et al. (2018) due to disregard of construct conceptualisation. Moreover, findings that based on the OPP scale, which had been deployed to measure credibility (Hardy et al., 1995), were also excluded from further analyses as no single item of the OPP matched to the conceptualisation of credibility (or expectancy).

**Studies eligible to answer the research questions.** When regarding the conclusions drawn from comparative analyses, the final data pool consisted of six studies exclusively referring to credibility and five studies exclusively referring to expectancy. Originally, three further studies considered measurements of both constructs. However, to avoid multicollinearity, Hundt et al. (2013) excluded the expectancy scale from further analyses due to its high correlation with the credibility scale ($r = .67$, $p < .01$). In total, slightly more studies measured credibility ($N = 8$) than expectancy ($N = 7$).

**Measurement points.** Expectancy was assessed early in treatment ($N = 4$) or at baseline ($N = 3$). Only Riley (2015) as well as Chambless, Tran and Glass (1997) assured that in-treatment assessments of expectancy had been conducted after provision of the CBT rationale. Credibility was mainly assessed early in treatment, after provision of the CBT rationale ($N = 5$), or at pre-treatment ($N = 3$). Hardy et al. (1995) conducted multiple measures to study potential changes of credibility from baseline, to pre-randomization to the early treatment phase. Meyerhoff and Rohan (2016) studied potential changes of credibility by assessing it at baseline, mid-treatment and post-treatment.

Most of studies selected an adequate measurement point to examine expectancy and/or credibility, thereby obtaining the full score in the quality assessment of measurement points (score = 1, range: 0 – 1). Eight studies received additional quality points for considering the
performance of measurements after rationale provision (score = 1, range: 0 – 1).

**Outcome and process variables.** Outcome variables that had been assessed on their respective relationships with expectancy referred to symptoms of anxiety, depression and gambling behaviour. Process variables that had been assessed in relation to expectancy referred to alliance (Westra, Constantino & Aviram, 2011; Webb et al., 2014) and homework compliance (Westra, Dozois & Marcus, 2007). Expectancy was more frequently assessed in samples of patients with anxiety disorders (N = 7) than in samples of patients with depression (N = 1).

Outcome variables that had been assessed on their respective relationships with credibility referred to symptoms of anxiety, depression, general psychopathology and gambling behaviour as well as to quality of life and satisfaction with treatment. Process measures that had been assessed in relation to credibility referred to alliance (Söchting et al., 2016) and adherence (Hundt et al., 2013). Credibility was nearly as frequently assessed in samples of patients with anxiety disorders (N = 4) as in samples of patients with depression (N = 3). In total, more outcome variables were assessed on potential relationships with credibility than on potential relationships with expectancy (see Table 4 in Appendix A).

Nearly all outcome and process variables were measured with established, validated and reliable measuring tools. Therefore, most of studies obtained the full score in the quality assessment on measuring tools utilized for the assessments of outcome/process variables (score = 1, range: 0 – 1). Only Chambless, Tran and Glass (1997) developed novel questionnaires for the purpose of their study (all fully written-out questionnaires in Table 4 in Appendix A). Those scales had not been assessed on psychometric properties, but researchers at least proved that each scale loaded on a single factor.

**Drop-out analyses.** Five studies obtained the full score in the quality assessment of drop-out analyses (score = 1, range: 0 – 1) as they had compared drop-outs with completers regarding their respective scores on expectancy or credibility. Results from Meyerhoff and Rohan (2016) suggested that drop-outs had lower baseline credibility scores (M = 5.69, SD = 2.34) than completers (M = 7.31, SD = 1.81) but the significance of this difference was not examined. Moreover, Westra, Dozois and Marcus (2007) reported that drop-outs had lower expectancy scores than completers (t (113) = 1.89, p = .06) but the found difference is only borderline significant when a standard alpha level (α = .05) is used as benchmark. No further differences were found.

**Statistical analyses performed.** Slightly more than half of studies (N = 8) examined the predictive effects of expectancy or credibility on outcomes via multiple regression
analyses, thereby allowing for a direct comparison of effect sizes between different independent variables. These studies received the full score in the quality assessment of statistical analyses (score = 1, range: 0 – 1). Zero scores were assigned to studies that examined relationships by conducting univariate regression or correlational analyses exclusively (Webb et al. 2014; Riley, 2015). Remarkably, Webb et al. (2014) and Riley (2015) measured the relationships between expectancy or credibility and outcome solely by using post-treatment values of outcome variables as dependent variables. This was different in all other studies, which either considered baseline scores of outcome variables as further independent variable or computed (residual) gain scores of outcome variables to serve as dependent variables. Analyses conducted in these studies were therefore considered to be higher in quality than the analyses conducted by Webb et al. (2014) and Riley (2015).
<table>
<thead>
<tr>
<th>Study</th>
<th>Study Design</th>
<th>Sample</th>
<th>Diagnosis</th>
<th>Treatment</th>
<th>Therapists</th>
<th>Notes</th>
<th>Score*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chambless et al. (1997) USA</td>
<td>Observational (highly selective, one group pretest, mid-test, posttest, follow-up design)</td>
<td>N = 62&lt;br&gt;Age: M = 35, SD = 9.7&lt;br&gt;Gender: F = 35</td>
<td>SAD (DSM-III-R)</td>
<td>Exposure-based group CBT (manualized) Scope: NR</td>
<td>PhD students</td>
<td>Dropouts (N = 5) were not included in main analysis.</td>
<td>5/10</td>
</tr>
<tr>
<td>Chambless et al. (2017) USA</td>
<td>Experimental (RCT, pretest, mid-test, posttest, follow-up design, conditions: CBT, PFPP)</td>
<td>N = 161&lt;br&gt;(CBT: N=80)&lt;br&gt;Age: M = 39.40, SD = 13.25&lt;br&gt;Gender: F = 104</td>
<td>PD/A (NR)</td>
<td>Individual CBT (manualized) Scope: 24 x 45 minutes</td>
<td>Trained therapists</td>
<td>Data drawn from: Milrod et al. (2016)</td>
<td>8/10</td>
</tr>
<tr>
<td>Dozois &amp; Westra (2005) CAN</td>
<td>Observational (one-group pretest, posttest design)</td>
<td>N = 43&lt;br&gt;Age: M = 37.86, SD = 10.42&lt;br&gt;Gender: F = 34</td>
<td>GAD (DSM-IV)</td>
<td>Group CBT (manualized) Scope: 8 x 150 minutes</td>
<td>CBT experienced mental health professionals</td>
<td></td>
<td>6/10</td>
</tr>
<tr>
<td>Meyerhoff &amp; Rohan (2016) USA</td>
<td>Experimental (RCT, pretest, mid-test, posttest design, conditions: CBT, LT)</td>
<td>N = 177&lt;br&gt;(CBT: N = 88)&lt;br&gt;Age: M = 45.6, SD = NR&lt;br&gt;Gender: F = 147</td>
<td>MDD + SEAD (DSM-IV-TR)</td>
<td>Group CBT (manualized) Scope: 12 x 90 minutes</td>
<td>PhD students, psychological interns</td>
<td>Data collected as part of RCT by Rohan et al. (2013). Dropouts (N = 12) had a lower baseline outcome expectancy (M = 5.69, SD = 2.34) than completers (M = 7.31, SD = 1.81).</td>
<td>5/10</td>
</tr>
<tr>
<td>Vîslă et al. (2018) CAN</td>
<td>Observational (highly selective, one-group pretest, mid-test, posttest design)</td>
<td>N = 91&lt;br&gt;Age: ≥ 18&lt;br&gt;Gender: NR</td>
<td>MDD (DSM-V)</td>
<td>Group CBT (manualized) Scope: 10 x 120 minutes</td>
<td>Doctorate-level registered psychologist, psychiatrists</td>
<td>Data drawn from: Tsai et al. (2014). Dropouts (N = 15) did not significantly differ from completers in ratings of outcome expectancy at baseline t (78) = 0.59, p = 0.56, or session 3, t (66) = 0.05, p = 0.96.</td>
<td>7/10</td>
</tr>
<tr>
<td>Webb et al. (2014) USA</td>
<td>Observational (one group repeated measures design)</td>
<td>N = 103&lt;br&gt;Age: M = 36.02, SD = 13.71&lt;br&gt;Gender: F = 66</td>
<td>MDD (DSM-IV)</td>
<td>Group CBT (flexible) Scope: 5 x 50 minutes per weak</td>
<td>Different professions (e.g. social workers, psychologists)</td>
<td>Patients who did not complete the multiple mid-treatment assessments (N = 4) were not included in the final sample.</td>
<td>6/10</td>
</tr>
<tr>
<td>Study</td>
<td>Study Design</td>
<td>Sample</td>
<td>Diagnosis</td>
<td>Treatment</td>
<td>Therapists</td>
<td>Notes</td>
<td>Score*</td>
</tr>
<tr>
<td>---------------</td>
<td>------------------------------------------------------------------------------</td>
<td>--------------------------------------</td>
<td>-----------</td>
<td>----------------------------</td>
<td>------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>Westra et al. (2007) CAN</td>
<td>Observational (highly selective, one group repeated measures design)</td>
<td>N = 67; Age: M = 41.03, SD = NA; Gender: F = 43</td>
<td>GAD, PD/A, SAD (DSM-IV)</td>
<td>Group CBT (manualized) Scope: 10 x 120 minutes</td>
<td>CBT experienced mental health professionals</td>
<td>Dropouts (N = 18) and completers (N = 30) without full data were not included in the final sample. Dropouts had lower expectancy scores than completers, t (113) = 1.89, p = .06.</td>
<td>6/10</td>
</tr>
<tr>
<td>Westra et al. (2011) CAN</td>
<td>Experimental (secondary analyses of CBT-arm of a RCT multiple repeated measure design)</td>
<td>N = 38; Age: M = 40.89, SD = 11.73; Gender: F = 27</td>
<td>GAD (DSM-IV-TR)</td>
<td>Individual CBT Scope: 6 x 120 minutes and 2 x 60 minutes</td>
<td>Trained PhDs, graduate students</td>
<td>Data drawn from RCT by Westra, Arkowitz &amp; Dozois, (2009). Drop-outs (N = 6) were not included in the final sample.</td>
<td>4/10</td>
</tr>
<tr>
<td>Andersson et al. (2008) SWEDEN</td>
<td>Experimental (RCT pretest, posttest, follow-up design, conditions: CBT, iCBT)</td>
<td>N = 49; (CBT: N = 25); Age: M = 35.8, SD = 9.3; Gender: F = 19</td>
<td>PD (DSM-IV)</td>
<td>Individual CBT (manualized) Scope: 10 sessions (duration: NR)</td>
<td>Clinical psychologists, graduate students, psychologists in training</td>
<td>Data drawn from Carlbring et al. (2005)</td>
<td>3/10</td>
</tr>
<tr>
<td>Hardy et al. (1995) UK</td>
<td>Experimental (RCT pretest, posttest, follow-up design, conditions: CBT (long/brief), IPT (long/brief))</td>
<td>N = 117; (CBT: N = NR); Age: M = 40.5, SD = 9.5; Gender: F = 61</td>
<td>MDD (DSM-III)</td>
<td>Individual CBT brief: 8 sessions long:16 sessions; (Duration): NR</td>
<td>Trained clinical psychologists</td>
<td>Data drawn from Shapiro et al., (1990). Sample consisted of well-positioned workers only.</td>
<td>5/10</td>
</tr>
<tr>
<td>Söchting et al. (2016) CAN</td>
<td>Observational (highly selective, one group pretest, mid-test, posttest design)</td>
<td>N = 80; Age: M = 47.82, SD = 10.58; Gender: F = 58</td>
<td>MDD (DSM-IV-TR)</td>
<td>Group CBT (manualized) Scope:10 x 120 minutes</td>
<td>Clinical psychologist, psychiatrists</td>
<td>Drop-outs (N = 15) did not differ from completers in credibility ratings (t (60) = – 0.62, p = .54) and were excluded from main analysis.</td>
<td>7/10</td>
</tr>
<tr>
<td>Hundt et al. (2013) USA</td>
<td>Experimental (RCT pretest, posttest, follow-up design, conditions: CBT, EUC)</td>
<td>N = 103; (CBT: N = 60); Age: M = 67.3, SD = 5.9; Gender: F = 83</td>
<td>GAD (DSM-IV)</td>
<td>Individual CBT (flexible) Scope: 10 sessions (duration: NR)</td>
<td>Bachelor’s and master’s level therapists, predoctoral interns</td>
<td>Data drawn from Author (2009). Drop-outs and completers without full data were not included in the final sample (N = 40).</td>
<td>5/10</td>
</tr>
</tbody>
</table>
### Table 1. Descriptive statistics (part 3/3)

<table>
<thead>
<tr>
<th>Study</th>
<th>Study Design</th>
<th>Sample</th>
<th>Diagnosis</th>
<th>Treatment</th>
<th>Therapist</th>
<th>Notes</th>
<th>Score*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hundt et al. (2014) USA</td>
<td>Experimental (secondary analyses of CBT-arm of a RCT pretest, posttest design)</td>
<td>N = 150 Age: ≥ 60 years Gender: F = 75</td>
<td>GAD (DSM-IV)</td>
<td>Individual skill-based CBT (manualized) Scope: 10 sessions (duration: NR) + telephone support</td>
<td>NR</td>
<td>Data drawn from RCT by Stanley et al. (2014)</td>
<td>4/10</td>
</tr>
<tr>
<td>Haley et al. (2015) AUS</td>
<td>Observational (highly selective, one-group pretest, posttest, follow-up design)</td>
<td>N = 74 Age: adult Gender: NR</td>
<td>PGD (NR)</td>
<td>Individual ERP-based CBT (manualized) Scope: 12 x 60 minutes</td>
<td>Social workers, psychology graduates</td>
<td>Treatment drop-outs (N = 29) and treatment-engagers (N = 45) did not significantly differ in baseline assessments and were not included in the final sample.</td>
<td>6/10</td>
</tr>
</tbody>
</table>

*Note: *Total score obtained in quality assessment. A total score of 10 was possible (see Table 1 and Table 2 in Appendix A). DSM: Diagnostic and Statistical Manual of Mental Disorders; ERP: exposure and response prevention; EUC: enhanced usual care; GAD: generalized anxiety disorder; iCBT: internet-based CBT; IPT: interpersonal therapy; LT: light therapy; MDD: major depressive disorder; PDA: Panic disorder with agoraphobia; PD: Panic disorder; PGD: pathological gambling disorder; PFPP: panic focused psychodynamic psychotherapy; RCT: randomized control trial; SAD: social anxiety disorder; SEAD: seasonal affective disorder.
<table>
<thead>
<tr>
<th>Study</th>
<th>E/C variables, measurement points &amp; internal consistency</th>
<th>Outcome/Process variables &amp; measurement points</th>
<th>Statistical analyses performed to assess direct or indirect relationships between E/C and outcome of CBT.</th>
<th>Significant relationships found</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chambless et al. (1997)</td>
<td>CES* (4 items derived): post-session 1+ (α: NA)</td>
<td>anxious apprehension, speech anxiety, speech skill, (dyad) anxiety and skill: baseline, post-treatment and follow-up</td>
<td>Bivariate analyses (Pearson correlation) and multivariate analyses (multiple regression analyses) were performed to measure relationships between expectancy and residual gain scores (pre-to-post-treatment, pre-to follow-up) of each outcome variable.</td>
<td>Initial expectancy was weakly positively correlated with pre-to post-treatment changes in speech skills (r = .26*), pre-to post-treatment (r = .29*) and pre-to follow-up changes in anxious apprehension (r = .29*) and moderately positively correlated with pre-to follow-up change in dyad anxiety &amp; skill (r = .39**). When multiple predictors** were considered, initial expectancy only had a small, positive effect on pre-to-follow-up change in dyad anxiety &amp; skill (β = NA).</td>
</tr>
<tr>
<td>Chambless et al. (2017)</td>
<td>CEQ-E: session 2+ (α: NA)</td>
<td>Δ PDSS: week 1, 5, 9 and follow-up</td>
<td>Multivariate analyses (MLM) was performed to measure relationship between expectancy and rate of change in panic disorder severity. Moderation analysis (via MLM) was performed to measure if the expectancy-change relationship varied with condition (CBT, PFPP) or site.</td>
<td>When multiple predictors were considered**, initial expectancy was a strong predictor of change in panic disorder severity (d = -1.05, CI 95% [-1.50, -0.60]). This effect was moderated by an interaction with treatment condition (t (83) = -2.19*). An investigation of the interaction revealed that patients with lower initial expectancy improved more in CBT than in PFPP (d = .52, CI 95% [0.09, 0.95]).</td>
</tr>
<tr>
<td>Dozois &amp; Westra (2005)</td>
<td>ACES: baseline (α = .91)</td>
<td>BAI, PSWQ: baseline and post-treatment</td>
<td>Bivariate analyses (Pearson correlation) were performed to measure relationships between expectancy and pre-to post-treatment change scores of each outcome variable. Multivariate analyses (hierarchical regression analyses) were performed to measure expectancy-outcome relationships.</td>
<td>Baseline expectancy was moderately positively correlated with pre-to post-treatment changes in anxiety (r = .44**) and worry severity (r = .46**). When multiple predictors were considered**, baseline expectancy had a moderate negative effect on anxiety (β = -.46**) and worry (β = -.36*) at post-treatment.</td>
</tr>
<tr>
<td>Meyerhoff &amp; Rohan (2016)</td>
<td>CES (adapted version): baseline* (α = .85), mid-treatment (α = .91) and post-treatment (α = .39)</td>
<td>BDI: baseline, mid- and post-treatment</td>
<td>Bivariate analyses (Pearson correlation) were performed to measure relationships between depression and expectancy at mid-treatment and post-treatment. Multivariate analyses (structural equitation modelling) were performed to measure relationships between expectancy and depression at each measurement point.</td>
<td>Post-treatment expectancy was moderately negatively correlated with depression at post-treatment (r = -.32*). When multiple predictors were considered**, expectancy at mid-treatment had a small, negative effect on depression at post-treatment (β = -.17*).</td>
</tr>
<tr>
<td>Study</td>
<td>E/C variables, measurement points &amp; internal consistency</td>
<td>Outcome/Process variables &amp; measurement points</td>
<td>Statistical analyses performed to assess direct or indirect relationships between E/C and outcome of CBT.</td>
<td>Significant relationships found</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Visla et al.</td>
<td>OES&lt;sup&gt;)&lt;/sup&gt;; baseline (α = .92) and post-session 3&lt;sup&gt;+&lt;/sup&gt; (α = .89)</td>
<td>BDI-II, BAI, IPP-28; baseline and post-treatment, WAI; session 1 and 5</td>
<td>Mediation analysis (via bootstrapping) were performed to measure if alliance mediated the relationship between baseline expectancy and outcome. Serial multiple mediation analysis was performed to measure if alliance (1st mediator) and post-session 3 expectancy (2nd mediator) mediated the relationships between baseline expectancy and outcome (dummy codes were used for this analyses).</td>
<td>Higher initial expectancy related to better mid-treatment alliance, which in turn related to lower anxiety at post-treatment (CI 95% [-5.086, -2.66]). Better initial alliance related to higher subsequent expectancy, which in turn related to fewer interpersonal problems at post-treatment (CI 95% [-.221, -.042]). Higher baseline expectancy related to better early to mid-treatment alliance, which in turn related to lower depression at post-treatment (CI 95% [-2.933, -.159]). Higher baseline expectancy related to better initial alliance, which in turn related to fewer interpersonal problems at post-treatment (indirect effect = -.049, CI 95% [-.2136, -.0004]). Effect sizes of indirect effects were weak to moderate.</td>
</tr>
<tr>
<td>Webb et al.</td>
<td>CEQ-E: first day of treatment (α = .89)</td>
<td>CES-D-10: day 1, 2, 4, 7 and post-treatment, WAI: day 2, 4, 7 and post-treatment</td>
<td>Bivariate analyses (Pearson correlation) were performed to measure relationships between expectancy and depression as well as between expectancy and alliance at each measurement point.</td>
<td>Initial expectancy was weakly negatively correlated with depression at post-treatment (r = -.27**). Correlations between initial expectancy and mid-treatment assessments of depression as well as alliance were also significant.</td>
</tr>
<tr>
<td>Westra et al.</td>
<td>ACES: baseline (α = NA)</td>
<td>Δ FNEB, Δ ASI, Δ PSWQ; each session</td>
<td>Multivariate analyses (partial correlations) were performed to measure relationships between expectancy, initial change (session at which first drop in symptoms occurred) and speed of response as well as between expectancy and fear of negative evaluation at post-treatment. Mediation analyses (via multiple regression analyses) was performed to measure if homework compliance mediated the relationships between expectancy and symptoms at first point of change. Covariates: baseline score of each outcome variable</td>
<td>Baseline expectancy was moderately negatively correlated with speed of response (r = -.39**) (with negative associations indicating a faster response), and very strongly negatively correlated with fear of negative evaluations at post-treatment (r = -.71**). The weak to moderate negative predictive effects of baseline expectancy on anxiety (β = -.28*) and worry at first point of change (β = -.30*), were no longer significant when homework compliance was considered. Higher baseline expectancy related to higher homework compliance, which in turn related to greater initial changes in anxiety and worry.</td>
</tr>
</tbody>
</table>
Table 2. Methodological characteristics and results (part 3/4)

<table>
<thead>
<tr>
<th>Study</th>
<th>E/C variables, measurement points &amp; internal consistency</th>
<th>Outcome/Process variables &amp; measurement points</th>
<th>Statistical analyses performed to assess direct or indirect relationships between E/C and outcome of CBT.</th>
<th>Significant relationships found</th>
</tr>
</thead>
<tbody>
<tr>
<td>Westra et al. (2011)</td>
<td>CEQ-E (single item derived); post-session 1,3,5 and 7 ((\alpha = NA))</td>
<td>PSWQ: baseline and post-treatment, CALPAS: session 1, 3, 5 and 7</td>
<td>Mediation analyses (via MCMAM) was performed to measure if the relationships between alliance rupture and outcome as well as between initial expectancy and outcome are mediated by subsequent expectancy (i.e. expectancy after an alliance rupture was observed). Covariates: baseline score of each outcome variable</td>
<td>Alliance rupture related to lower subsequent expectancy, which in turn related to higher worry severity at post-treatment (indirect effect = .17, CI 95% [ .39, 12.45]). Higher initial expectancy related to higher subsequent expectancy, which in turn related to lower worry severity at post-treatment (indirect effect = -.27, CI 95% [-.35, .01]). There was no significant direct relationship between initial expectancy and worry severity at post-treatment.</td>
</tr>
<tr>
<td>Andersson et al. (2008)</td>
<td>CES: baseline ((\alpha = NA))</td>
<td>ACQ, BSQ: baseline, post-treatment and follow-up</td>
<td>Multivariate analyses (semi-partial correlations) were performed to measure relationships between baseline credibility and residual gain scores (pre-to-post-treatment, pre-to-follow-up) of each outcome variable.</td>
<td>When multiple predictors were considered**, baseline credibility was moderately positively correlated with pre-to post-treatment change in body sensations ((r = .34^*)).</td>
</tr>
<tr>
<td>Hardy et al. (1995)</td>
<td>OPP-TR(^a): baseline ((\alpha = .87)) TCF: post-randomization and post-session (^r) ((\alpha = NA))</td>
<td>BDI, SCL-90, SE, IPP; baseline, post-treatment and follow-up</td>
<td>Multivariate analyses (hierarchical regression analyses, partial correlations) were performed to measure relationships between credibility and outcome. Moderation analyses were performed to measure if the relationship between credibility and outcome varied with condition (IP, CBT).</td>
<td>Prediction of outcome by credibility at baseline was moderated by treatment condition and only significant for the IP condition. In the CBT condition, no direct relationships between credibility and outcome were found. No further statistics were available.</td>
</tr>
<tr>
<td>Söchting et al. (2016)</td>
<td>CS: post-session (^1) ((\alpha = .89))</td>
<td>BDI, BAI, QOLI, IPP-28: baseline and post-treatment, WAI: session 1 and 5</td>
<td>Multivariate analyses (hierarchical regression analyses) was performed to measure relationships between credibility and pre-to post-treatment change scores of respective outcome variables. Mediation analyses (via multiple regression analyses) was performed to measure if alliance mediated these relationships. Covariates: baseline scores of respective outcome variables.</td>
<td>When multiple predictors were considered**, initial credibility had a moderate, positive effect on pre-to post-treatment change in interpersonal problems ((\beta = .34^<em>)) and very strong, positive effects on early alliance quality ((\beta = .73^{</em><strong>})) and alliance quality at mid-treatment ((\beta = .74^{</strong>*})). Mediation analysis was ceased as alliance was not significantly associated with change in interpersonal problems.</td>
</tr>
</tbody>
</table>
**Table 2. Methodological characteristics and results (part 4/4)**

<table>
<thead>
<tr>
<th>Study</th>
<th>E/C variables, measurement points &amp; internal consistency</th>
<th>Outcome/Process variables &amp; measurement points</th>
<th>Statistical analyses performed to assess direct or indirect relationships between E/C and outcome of CBT.</th>
<th>Significant relationships found</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hundt et al. (2013)</td>
<td>ERS: post-session 1+ (credibility subscale: α = .89, expectancy subscale: α = NA)</td>
<td>PSWQ: baseline and post-treatment, AWTS, CSQ: post-treatment and follow-up</td>
<td>Multivariate analyses (multiple regression analyses) was performed to measure relationship between credibility and treatment satisfaction. Mediation analysis (via bootstrapping) were performed to measure if adherence mediated the relationship between credibility and satisfaction at post-treatment as well as the relationship between credibility and pre-to post-treatment change score of worry severity.</td>
<td>When multiple predictors were considered**, initial credibility had a moderate, positive effect on satisfaction with treatment (β = .39***). Higher initial credibility related to greater adherence which in turn related to higher satisfaction at post-treatment (CI 95% [.25, 1.33]).</td>
</tr>
<tr>
<td>Hundt et al. (2014)</td>
<td>ERS: baseline (credibility subscale: α = .89, expectancy subscale: α = NA)</td>
<td>PSWQ-A, STAI-T: baseline and post-treatment</td>
<td>Bivariate analyses (simple linear regression analyses) and multivariate analyses (multiple regression analyses) were performed to measure relationships between expectancy and outcome as well as between credibility and outcome.</td>
<td>Baseline credibility had a negative effect on anxiety at post-treatment (b = -.38*, SE(b) = .18). This effect remained significant in completers when multiple predictors were considered**.</td>
</tr>
<tr>
<td>Riley (2015)</td>
<td>CEQ: post-session 1+ (α = NA)</td>
<td>VGS-HS: baseline, post-treatment and follow-up</td>
<td>Bivariate analyses (Pearson correlation) were performed to measure relationships between expectancy and outcome as well as between credibility and outcome.</td>
<td></td>
</tr>
</tbody>
</table>

Note: *p < .05, **p < .01, and ***p < .001; a qualitative analyses of measuring tools utilized to assess credibility and expectancy, indicated that the CES rather measures credibility than expectancy; b qualitative analyses indicated that the OES disregards the distinction of credibility and expectancy which is why results from this study were excluded from further analyses; c scale does not meet the conceptualisation of credibility; d researchers excluded the expectancy subscale from analyses due to its strong correlation with the credibility subscale (r = .67***), e researchers assured this measure was taken after provision of CBT rationale. ** further predictors/covariates considered in multivariate analyses will be discussed in text.

ACES: Anxiety Change Expectancy Scale; ACQ: Agoraphobic Cognitions Questionnaire; ASI: Anxiety Sensitivity Index; AWTS: Adherence with Treatment Scale; BAI: Beck Anxiety Inventory; BDI-(II): Beck Depression Inventory; BSQ: Body Sensations Questionnaire; CALPAS: California Psychotherapy Alliance Scale; CES: Credibility/Expectancy Scale; CES-D-10: Centre for the Epidemiological Studies of Depression-10; CEQ-E: Credibility Expectancy Questionnaire, expectancy subscale; CS: Credibility Scale; CSQ: Client Satisfaction Questionnaire; ERS: Expectancy Rating scale (ERS-E Expectancy subscale; ERS-C credibility subscale); FNEB: Fear of Negative Evaluation Scale, brief version; IPP-(28): inventory of interpersonal problems; MCAMAM: Monte Carlo Method for Mediation; MLM: multilevel modelling; OES: Outcome Expectancies Scale; OPP-TESS: Opinions about Psychological Problems Questionnaire, Treatment Endorsement subscale; PDSS: panic disorder severity scale; PSWQ: Penn State Worry Questionnaire (PSWQ-A: abbreviated version); QOLI: Quality of Life Inventory; SE: measure of self-esteem; SCL-90: symptom checklist of psychological problems; STAI-T: Spielberger State-Trait Anxiety Inventory, trait subscale; TCF: Treatment credibility form; VGS-HS: Victorian Gambling Screen Harm to Self subscale; WAI: Working Alliance Inventory
3.2. Answer to the research questions

First, it was examined in which way and to which extent expectancy and credibility were related to the outcomes of CBT (RQ 1) and if their respective effects had been indicated to vary across psychological disorders (RQ 2). Main findings from these investigations were then compared to determine if expectancy and credibility differ in their relationships with different CBT outcomes (RQ 3).

3.2.1 Answer to research question 1 and research question 2

**Outcome expectancy.** The strength of significant relationships between expectancy and CBT outcomes ranged from weak (Webb et al., 2014) to very strong (Westra, Dozois & Marcus, 2007). However, most of relationships proved to be significant, reflected moderate associations between expectancy and changes in symptoms of anxiety disorders. Only the findings from Hundt et al. (2014) did not support these associations. Nevertheless, contrary to all other studies on expectancy, Hundt et al. (2014) did not deploy a validated measuring tool to assess expectancy, challenging the meaningfulness of their findings.

Higher expectancy was moderately related to greater pre-to-post-treatment changes of anxiety ($r = .44^{**}$) and worry ($r = .46^*$) in patients with GAD. The strength of these relationships remained moderate when they were controlled for baseline symptoms and hopelessness (Dozois & Westra, 2005). Although the unique effects of baseline anxiety ($\beta = .51^{**}$) and worry ($\beta = .46^{**}$) were stronger in this latter model, the findings showed that expectancy predicted changes in post-treatment anxiety ($\beta = -.46^{**}$) and worry ($\beta = -.36^*$) over and above baseline symptomatology. Furthermore, findings on a sample of patients with GAD, PD and SAD, showed that expectancy does not only influence to which degree worry ($\beta = -.30^*$) and anxiety symptoms ($\beta = -.28^{**}$) positively change, but also provides an indication about how much time it might take for these changes to occur ($r = -.39^{**}$) with patients having higher expectancies, improving more strongly and more quickly (Westra, Dozois & Marcus, 2007). However, this does not mean that individuals, being less optimistic about their potential improvements, were unresponsive to CBT. As shown by Chambless et al. (2017), patients with PD perceiving lower expectancy to improve by treatment, experienced moderately greater changes in panic severity when they were treated by CBT than when they were treated by psychodynamic therapy ($d = 0.52$, CI 95% [0.09, 0.95]).

The strongest relationships between expectancy and anxiety symptoms was revealed by Westra, Dozois and Marcus (2007), who examined the effect of expectancy on post-treatment fear of negative evaluation in patients with SAD. The very strong association found

*Note: $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$*
(r = - .71**), however, was revealed by exploratory analyses and was solely significant for patients who had SAD as secondary but not as principal diagnosis. This group of patients might have been less hampered to shed their social fears, or their fears might have been less pronounced originally, which is why this finding should be interpreted with caution.

Next to being directly related to treatment outcomes, expectancy was also indirectly related to outcomes. Westra, Dozois and Marcus (2007) showed that the weak to moderate predictive effects of expectancy on anxiety (β = -.28*) and worry (β = -.30*) at first point of change were no longer significant when homework compliance was considered. Higher baseline expectancy related to higher homework compliance which in turn related to lower anxiety and worry. Moreover, a secondary analysis on a subset of patients with GAD that had experienced an alliance rupture (Westra, Constantino & Aviram, 2011) revealed that alliance ruptures related to lower subsequent expectancy, which in turn related to higher worry severity at post-treatment (CI 95% [.39, 12.45]). However, patients with higher initial expectancy also had higher expectancies at post-rupture as well as a lower worry severity at post-treatment than patients with lower initial expectancy. That is, the effect of initial expectancy on post-treatment worry was mediated by a patient’s expectancy at post-rupture (CI 95% [-.35, .01]).

Only one study assessed the effects of expectancy on the outcomes of CBT for depression. Higher expectancy was only weakly related to lower depression at post-treatment (r = -.27**) (Webb et al., 2014). In this study, relationships between expectancy and mid-treatment assessments of depression were stronger (r = -.31**; r = -.41**). The mean length of the treatment provided by Webb et al. (2014) turned out short (M = 11.80 days, SD = 3.81) indicating that the predictive power of initial expectancy on depression diminished quickly over the course of CBT. Nevertheless, higher expectancy was moderately related to higher early (r = .31**) and mid-treatment alliance (r = .30**) which in turn both were moderately related to lower depression at post-treatment (r = -.34**; r = -.41**), suggesting that alliance might has been a potential mediator in the relationship between expectancy and post-treatment depression.

With regards to RQ 1, it was concluded that expectancy is both directly and indirectly related to the outcomes of CBT. The direct effects of expectancy were mainly of medium strength and referred to post-treatment levels of anxiety and worry in patients with GAD and PD. These effects were also shown to be mediated by homework compliance. Findings from Webb et al. (2014) as well as Westra, Constantino and Aviram (2011) further suggested that alliance could play a role in the relationships between expectancy and post-treatment worry as
well as depression.

With regards to RQ 2, it was carefully concluded that expectancy could play a more important role in changes in symptoms of anxiety than in changes in depression. However, more findings than those of Webb et al. (2014) are needed to ascertain if expectancy is generally weakly related to post-treatment depression. Especially as worry, which can be regarded as a symptom of depression, was shown to be moderately related to expectancy.

The findings further suggested that expectancy was unrelated to the outcomes of CBT for PGD. In patients with PGD, expectancy had no explanatory value for gambling behaviour at post-treatment (Riley, 2015). However, as with depression, more research is needed to ascertain the representativeness of this finding.

**Credibility beliefs.** As, in nearly all studies, credibility was assessed with unvalidated measuring tools, most of the following findings should be interpreted with caution. Only Riley (2015) deployed an adequate scale to measure the relationship between credibility and outcomes of CBT for PGD, which, however, turned out to be statistically insignificant. The remaining studies referred to affective disorders. Relationships proved to be significant in these samples, reflected weak (Chambless, Tran & Glass, 1997) to moderate associations (e.g. Andersson, Carlbring & Grimalnd, 2008) between credibility and very specific outcome variables. However, the findings from Chambless, Tran and Glass (1997) should be interpreted with particular caution since the researchers performed unvalidated measurements of both, credibility and outcome variables.

Chambless, Tran and Glass (1997) revealed that, in patients with SAD, credibility was weakly related to a greater pre-to post-treatment change in speech skills (r = .26*), pre-to post-treatment (r = .29*) and pre-to follow-up changes in anxious apprehension (r = .29*) and moderately related to a greater pre-to follow-up change in anxiety coping skills (r = .39**). Most of these relationships failed to remain significant when depressive symptoms were considered. Then, credibility was only weakly related to anxiety coping skills (β = NA). In line with these findings, Hundt et al. (2014) revealed that credibility was rather weakly related to post-treatment anxiety in patients with GAD (b = -.38*, SE(b) = .18) at least when comparing the univariate effects of credibility with the effects other univariate predictors, like adherence (b = 7.77*, SE(b) = 3.85), homework compliance (b = -1.77*, SE(b) = .66) and baseline symptoms (b = .98*, SE(b) = .07) had on this outcome variable. However, Hundt et al. (2014) reported that the small effect of credibility on anxiety remained significant when all predictors were considered simultaneously (b = NA, p < .05).
Two medium-sized associations between credibility and outcomes of CBT for anxiety disorders were found. Credibility moderately predicted greater treatment satisfaction ($\beta = .39^{***}$) over and above adherence ($\beta = .42^{***}$), social support ($\beta = .12$, $p = \text{n.s}$) and changes in worry ($\beta = .07$, $p = \text{n.s}$) in patients with GAD (Hundt et al., 2013). Adherence, in turn, mediated the relationship between credibility and treatment satisfaction (CI 95% [.25, 1.33]). Credibility related to greater adherence which in turn related to higher satisfaction. Moreover, Andersson, Carlbring and Grimlund (2008) showed that, in patients with PD, higher credibility was moderately related to a greater change in anxious body sensations (partial $r = .34^{*}$), when agoraphobic symptoms (partial $r = .55^{**}$) verbal fluency (partial $r = -.24$, n.s) and personality disorder (partial $r = -.24$, n.s) were considered. However, credibility was unrelated to agoraphobic cognitions in this sample (Andersson, Carlbring & Grimlund, 2008).

Next to being related to outcomes of CBT for anxiety disorders, Söchting, Tsai and Ogrodniczuk (2016) reported that higher credibility moderately predicted changes in interpersonal problems ($\beta = .35^{*}$) in patients with MDD and that this predictive effect remained unchanged ($\beta = .35^{*}$) when baseline symptoms were controlled statistically ($\beta = \text{NA}$, $p = \text{NA}$). Moreover, credibility exerted a strong, positive effect on early ($\beta = .73^{*}$) and mid-treatment ratings of alliance ($\beta = .74^{*}$) in the depressive sample. In contrast to credibility, alliance had no explanatory value for interpersonal problems (Söchting, Tsai & Ogrodniczuk, 2016).

Nearly all other associations assessed between credibility and variables reflecting depressive symptoms failed to be significant. Initial credibility was neither related to post-treatment levels of self-esteem, psychopathological symptoms (Hardy et al., 1995), anxiety, quality of life (Söchting, Tsai & Ogrodniczuk, 2016) and general depression (Söchting, Tsai & Ogrodniczuk, 2016; Hardy et al., 1995; Meyerhoff & Rohan 2016) in patients with MDD and SEAD, nor was it related to post-treatment levels of worry in patients with GAD (Hundt et al., 2013; Hundt et al., 2014). Meyerhoff and Rohan (2016) could only show that post-treatment depression was moderately related ($r = -.32^{*}$) to post-treatment credibility and weakly related to mid-treatment credibility ($\beta = -.17^{*}$) but unrelated to initial credibility. This finding suggests that initial credibility could have lost its predictive power quickly in the course of CBT for depression.

With regards to RQ 1, it can be concluded that credibility is both directly and indirectly related to the outcomes of CBT. Credibility weakly affected changes in anxiety coping skills in patients with SAD as well as changes in anxiety in patients with GAD. Moreover, credibility moderately affected changes in anxious body sensations in patients with
PD, changes in interpersonal problems in depressive patients and treatment satisfaction in patients with GAD. All these effects were over and above baseline symptoms. However, credibility had no impact on worry in patients with GAD, on agoraphobic cognitions in patients with PD, on several outcome variables reflecting specific and general symptoms of depression in patients with MDD and it also failed to add explanatory value to the effects of baseline symptoms on anxious apprehension, speech anxiety and speech skills in patients with SAD.

With regards to RQ 2 it was concluded that credibility was more frequently and more strongly related to the outcomes of CBT for anxiety disorders than to the outcomes of CBT for depression. The results further suggested that credibility was unrelated to the outcomes of CBT for PGD. Overall, the findings indicated that credibility might rather be related to very specific outcome variables (e.g. anxious body sensations) than to outcome variables reflecting general symptoms of mental disorders (e.g. anxiety, worry).

3.2.2 Answer to research question 3

Expectancy and credibility were both directly related to the outcomes of CBT for MDD, GAD, PD and SAD or their effects on outcomes were mediated by resembling process variables such as adherence and homework compliance. Moreover, both constructs were unrelated to the outcomes of CBT for PGD. A further commonality between expectancy and credibility was that both were associated with alliance. Nevertheless, the two constructs differed in their relationships with different types of outcomes of CBT.

Expectancy was mainly moderately associated with changes in anxiety (Westra et al., 2007; Dozois & Westra, 2005) whereas the effects of credibility on post-treatment anxiety were only small (Hundt et al., 2014). Expectancy was moderately related to worry (Westra et al., 2007; Dozois & Westra, 2005; Chambless et al., 2017), whereas no significant relationships between credibility and worry were found (Hundt et al., 2013; Hundt et al., 2014). Moreover, although it was already emphasized that findings from Chambless, Tran and Glass (1997) as well as Webb et al., (2014) should be interpreted with caution, the researchers showed that expectancy was very strongly related to changes in fears in patients with SAD, whereas credibility was either weakly related to the outcomes of CBT for SAD or its effects on these outcomes were cancelled out when baseline symptoms were considered.

When regarding the results emphasizing the superior effects of expectancy, it must be noted that most of findings on expectancy were revealed by univariate analyses. Only Dozois and Westra (2005) considered baseline symptoms and hopelessness as potential covariates
when assessing the impact of expectancy on anxiety and worry, but the predictive effects of expectancy yet remained medium sized and were above and beyond baseline symptoms and hopelessness. In general, expectancy appears to be more frequently and more strongly related to the outcomes of CBT for anxiety disorders than credibility.

Whether this pattern also matches to the outcomes of CBT for depression, was more difficult to examine since expectancy was only measured in one study on depression. The findings, however, indicated that the effects of expectancy and credibility on depression both diminished quickly over the course of CBT (Meyerhoff & Rohan, 2016; Webb et al., 2014). Nevertheless, whereas initial expectancy was weakly related post-treatment depression (Webb et al., 2014), initial credibility was unrelated to this outcome measure (Meyerhoff & Rohan, 2016). More findings than those of Webb et al. (2014) are needed to examine if expectancy indeed has more explanatory value for post-treatment depression than credibility.

Nevertheless, credibility seems to play an especially important role for the therapeutic relationship in CBT for MDD. Credibility was namely very strongly related to alliance in depressive patients (Söchting, Tsai & Ogrodniczuk, 2016) whereas expectancy was only moderately related to this process measure (Webb et al., 2014).

Overall, credibility was less frequently and less strongly related to CBT outcomes than expectancy even though much more associations between credibility and outcomes had been examined. With regards to RQ 3, it was concluded that, whereas expectancy appears to be a robust predictor of changes in general symptoms of anxiety and perhaps also in general symptoms of depression, credibility might rather be related to very specific outcome variables than to outcome variables reflecting general symptoms of affective disorders. Nevertheless, since expectancy was not assessed on its association with these specific outcome variables, more research is needed to examine how expectancy and credibility differ in their impact on such specific outcome variables.
4. Discussion

This systematic review presents valuable insights about the roles of expectancy and credibility in the outcomes of CBT that may increase the understanding of why CBT can exert ameliorative, or even superior, effects on an individual’s mental health. The consideration and adequate measurement of expectancy and credibility might provide an important indication about a patient’s success in CBT. Both constructs are malleable and could be positively influenced and utilized by the therapist to achieve a desirable treatment outcome in CBT. In this connection it should, however, be considered that expectancy and credibility appear to have different explanatory values for the outcomes of CBT which vary across mental disorders and types of outcomes.

The gathered findings extent earlier results from Borkovec and Costello (1993) as well as Borkovec and Mathews (1988) by showing that expectancy was not only more frequently related to treatment outcomes in patients with anxiety disorders but was, in general, also more strongly related to these outcomes than credibility. Moreover, the present research could reveal that expectancy and credibility can be associated with the outcomes of CBT for depression even though both constructs may have more explanatory value for the outcomes of CBT for anxiety disorders. Nevertheless, expectancy might be more frequently and more strongly related to changes in general symptoms of depression than credibility. Yet the ameliorative effects of credibility may have been underestimated, as credibility was shown to predict changes in very specific outcomes of CBT for affective disorders over and above baseline symptoms (e.g. Söchting, Tsai & Ogrodniczuk, 2016).

Furthermore, findings from comparative outcome studies included in this review, supported the assumption that the predictive effects of expectancy and credibility vary across psychological treatments. In patients with panic disorder, those who had lower initial expectancy experienced greater changes in CBT than in psychodynamic therapy (Chambless et al., 2017). It can only be speculated why lower expectancy predicted worse outcome for psychodynamic therapy than for CBT. Chambless et al., (2017) assumed that a CBT therapist may have used cognitive restructuring to increase a patient’s expectancy. It could therefore be possible that the patient’s initial expectancy had positively changed. Earlier studies support this assumption by showing that both expectancy as well as credibility increased in the course of CBT for anxiety disorders (Newman & Fisher, 2010). To increase the explanatory power of expectancy and credibility, it might therefore be useful to assess both constructs after the patient experienced a sample of CBT. That is, after the CBT rationale was explained and after the patient had experienced the application of some CBT techniques.
Hardy et al.’s (1995) study on depressive patients showed that credibility was only related to the outcomes of interpersonal therapy but not to the outcomes of CBT. Why credibility was unrelated to CBT outcomes in this study is not clear. The sample of Hardy et al. (1995) exclusively consisted of well-positioned workers. It is possible that the credibility of a treatment rationale was very important for these highly educated patients. The interpersonal therapist may have recognised this and might have made a greater use of credibility to achieve a desirable treatment outcome. Another explanation could be that credibility usually has only little explanatory value for general symptoms of depression. This was further supported by the findings of several studies included in the present review (e.g. Söchting, Tsai & Ogrodniczuk, 2016). Only Meyerhoff and Rohan (2016) found a significant relationship between credibility and post-treatment depression. However, this relationship was weak and only significant for mid-treatment assessments of credibility but not for initial assessments of credibility.

The findings of Chambless et al. (2017) as well as Hardy et al. (1995) support the hypothesis that non-specific factors may not be that non-specific as researchers think they are (Dew & Bickman, 2005). The unique characteristics of a treatment may determine in which way and to which extent the ameliorative abilities of expectancy and credibility become unfolded and are utilized to achieve desirable outcomes. It is therefore important to emphasize which outcome and process variables are significantly associated with expectancy and credibility. Moreover, it is relevant to deliberate why such associations turned out to be significant in order to provide practical implications for a successful implementation of CBT.

Expectancy was a moderate and robust predictor of changes in anxiety and worry symptoms in patients with anxiety disorders (Dozois & Westra, 2005). Patients expecting greater improvements in anxiety symptoms were not only less worried and anxious when finishing CBT (Westra, Dozois & Marcus, 2007), they also improved more quickly than their counterparts (Dozois & Westra, 2005). As the effects of expectancy on anxiety and worry were over and above baseline symptoms, patients with anxiety disorders could benefit from targeting expectancy early in treatment, regardless of the severity of their symptoms.

A secondary analysis on a portion of patients that experienced an alliance rupture (i.e. a significant drop in alliance ratings) in the course of CBT for GAD, further revealed that individuals with higher initial expectancy achieved more positive outcomes after an alliance rupture than patients with lower initial expectancies (Westra, Constantino & Aviram 2011). The finding indicates that patients with higher expectancy levels might be less dejected, and hence more robust to potential setbacks in CBT. This is further supported by the expectancy-
value theory, emphasizing that individuals placing high values on a specific outcome, are more resilient to failures (Scholz, Blumer & Brand, 2012).

The findings of Westra, Constantino and Aviram (2011) also raised the question if expectancy could exert an influence on the likelihood for alliance-ruptures to occur, which should be examined in future research. Alliance is often considered as foundation for therapy that is strongly related to improvement (Castonguay, Constantino & Holtforth, 2006; Shirk, Gudmundsen, Kaplinski & McMakin, 2008). The present review found that both, expectancy as well as credibility exerted an influence on alliance in patients with depressive disorders. Credibility was a very strong predictor of alliance (Söchting, Tsai & Ogrodniczuk, 2016) which suggests that CBT therapists could make use of the credible CBT rationale to establish a material working relationship (Strong & Clairborn, 1982). Moreover, the moderate association that was found between early expectancy and mid-treatment alliance (Webb et al., 2014) supported the assumption that a positive expectancy could facilitate the build-up of this working relationship (Joyce & Piper, 1998).

Nevertheless, the role of alliance in CBT might be overestimated. Earlier research showed that alliance was only marginally related to post-treatment levels of anxiety and depression (e.g. Linden, 2008; Shirk et al., 2008; Shafran et al., 2009). In contrast, the findings of the present review indicated that expectancy was a moderate to strong predictor of changes in anxiety (e.g. Dozois & Westra, 2005; Chambless et al., 2017) and that credibility moderately predicted specific outcome measures of anxiety (e.g. anxious body sensations) (Andersson, Carlbring & Grimlund, 2008) as well as interpersonal problems in patients with depression, whereas alliance was unrelated to this latter outcome measure (Söchting, Tsai & Ogrodniczuk, 2016). Further research should examine if expectancy and credibility indeed have a greater impact on the outcomes of CBT for affective disorders than the widely accepted measure of therapeutic alliance. This could also provide an explanation for the fact that the outcomes of internet-based CBT are independent of the frequency of therapist support (Klein et al., 2009) and generally equal to the outcomes of face-to-face CBT (Andersson, Cuijpers, Carlbring, Riper & Hedman, 2014). It could therefore also be interesting to examine how expectancy and credibility are related to the outcomes of internet-based CBT.

The evidence-based techniques of CBT could overshadow the relevance of the therapeutic relationship (Shirk et al., 2008) and targeting expectancy and credibility early in treatment could increase a patient’s usage of these techniques. This is further supported by the finding that homework compliance mediated the relationship between expectancy and symptoms of anxiety (Westra, Dozois & Marcus, 2007) as well as the finding that adherence
mediated the relationship between credibility and treatment satisfaction in patients with anxiety disorders (Hundt et al., 2013). Both process variables are assumed to reflect the involvement with CBT techniques (Detweiler & Whisman, 1999). This implies that higher expectancy and credibility could enhance the likelihood to exhaust the positive effects of the evidence-based CBT techniques.

It can only be speculated why expectancy was, in general, more frequently and more strongly related to CBT outcomes for anxiety disorders than credibility. CBT demands a relatively high effort in therapy as this treatment entails a great degree of treatment sessions and homework assignments in a relatively fixed time period (Kazantzis, Whittington & Dattilio, 2010). Earlier research revealed that the effort involved in therapy as well as the willingness to undergo that effort will lead to changes through the reduction of cognitive dissonance (Axsom & Cooper, 1985). A patient’s idea about his outcome in a psychological treatment might have a greater impact on the patient’s effort in therapy than his perception of a treatment rationale (Axsom & Cooper, 1985; Weinberger & Eig, 1999).

Expectancy was mainly moderately associated with changes in anxiety (Westra et al., 2007; Dozois & Westra, 2005) whereas the effects of credibility on post-treatment anxiety were only small (Hundt et al., 2014). Expectancy was moderately related to worry (Westra et al., 2007; Dozois & Westra, 2005; Chambless et al., 2017), whereas no significant relationships between credibility and worry were found (Hundt et al., 2013; Hundt et al., 2014). Moreover, even though the effects of expectancy and credibility on depression both diminished quickly over the course of CBT (Meyerhoff & Rohan, 2016; Webb et al., 2014) the findings showed that initial expectancy was weakly related post-treatment depression (Webb et al., 2014) whereas initial credibility was unrelated to this outcome measure (Meyerhoff & Rohan, 2016). Meyerhoff and Rohan (2016) could only show that post-treatment depression was weakly related to mid-treatment credibility. Nearly all other associations assessed between credibility and variables reflecting depressive symptoms failed to be significant (Söchting, Tsai & Ogrodniczuk, 2016; Hardy et al., 1995; Meyerhoff & Rohan 2016). Nevertheless, credibility weakly affected changes in anxiety coping skills (Chambless, Tran & Glass, 1997) and moderately affected satisfaction (Hundt et al., 2013) and changes in anxious body sensations (Anderrson, Carlbring & Grimlund, 2008) in patients with anxiety disorders. Moreover, credibility moderately affected changes in interpersonal problems (Söchting, Tsai & Ogrodniczuk, 2016) in patients with depression. All these impacts were above and beyond baseline symptoms indicating that targeting credibility could affect changes in these outcome variables regardless of a patient’s symptom severity.
Credibility might rather be related to outcome variables that reflect the effectiveness of specific CBT techniques than to variables reflecting general symptoms of anxiety and depression. For example, social skills training directly addresses interpersonal problems by teaching techniques to prevent depressive patients from future issues in everyday interactions. Placing a higher credibility in such techniques could ensure that a patient makes a greater use of them in his personal life, thereby primarily reducing his interpersonal problems. In the long run, a reduction of interpersonal problems might cause a decline in depressive symptoms. This could be an explanation for the finding that credibility is more strongly related to specific variables such as interpersonal problems than to general symptoms such as depression.

Why expectancy and credibility are both more frequently and more strongly related to the outcomes of CBT for anxiety disorders than to the outcomes of CBT for depression, is not clear. However, patients with anxiety disorders are assumed to have good chances to recover from symptoms (Wlazlo, 1995) whereas the recidivism rate of major depression is comparatively high (Mueller et al., 1996). As the CBT rationale provides a general prognosis of an individual’s success in therapy (Newman & Fisher, 2010), patients with anxiety disorders might be more motivated to engage in CBT and make more use of the specific techniques than depressive patients. As expectancy and credibility both predict outcomes over and above baseline symptoms (e.g. Dozois & Westra, 2005; Söchting, Tsai and Ogrodniczuk (2016) patients with depressive disorders could benefit from a more positive but credible approach of psychotherapy that increases their outcome expectancy and hence, their chance of a more successful treatment outcome. Future research should examine if the effects of expectancy and credibility differ between “pure” versions of CBT and more strength-based, positive CBT approaches like Acceptance and Commitment Therapy (ACT). ACT namely turned out to be viable and promising form of CBT which patients even perceive as more satisfying than original versions of CBT (A-tjak et al., 2015; Wetherell et al., 2011; Craigie, Rees, Marsh & Nathan, 2008).

However, too few studies assessed expectancy in samples of patients with depressive disorders which is why this conclusion should be interpreted with caution. Moreover, when it comes to the superior specific effects of expectancy when compared to credibility, it must be noted that credibility was mainly assessed with unvalidated measuring tools. The CEQ, which measures both credibility and expectancy separately (Devilly & Borkovec, 2000) is, so far, the only valid and reliable assessment tool to measure credibility. However, the CEQ was rarely utilized to assess credibility. The CS scale from Söchting, Tsai and Ogrodniczuk (2016)
could also be a promising questionnaire to assess credibility. The CS not only contains questions about the credibility of a treatment approach but also includes questions about the credibility of the therapist in charge. Such measurements could be extremely valuable when considering that credibility is strongly related to working alliance (Söchting, Tsai & Ogrodniczuk, 2016). Next to the expectancy subscale of the CEQ, the ACES questionnaire from Dozois and Westra (2005) can be utilized to assess expectancy in a valid and reliable manner. All other questionnaires that were used have not been assessed on psychometric properties. Unfortunately, only five of the 14 studies included in this analysis employed valid and reliable assessment tools to measure the constructs of interest.

The inadequate assessment of both expectancy and credibility provides a great threat to the meaningfulness of findings included in the present review. Researchers still interchange the conceptions of expectancy and credibility or they assess both as a combined construct. Moreover, researchers still tend to merge data from patients of different conditions to assess relationships between outcomes and credibility as well as expectancy even though this study could emphasize that both constructs differ in their relevance for the outcomes of different treatments. A strength of the present study was that its validity was tried to increase by a careful check of the assessment tools utilized and the subsequent exclusion of studies that disregarded the conception and distinction of expectancy and credibility (Visla et al., 2018).

A limitation of this systematic review was the potential presence of a publication bias. That is, relevant findings on the relationships between CBT outcomes and expectancy as well as credibility were not published by authors based on the nature and direction of the results (Dwan et al., 2008). Therefore, it is not fully clear how much insignificant relationships between expectancy as well as credibility and CBT outcomes had been concealed. Moreover, it is so far unknown how both constructs relate to mental disorders other than affective disorders. Only one study considered both constructs in a sample of patients with PGD (Riley, 2015). Here, all assessed associations turned out to be insignificant. More data is needed to examine in which mental disorders expectancy and credibility exert their ameliorative influence.

So far, it was concluded that expectancy could be a robust indicator of changes in general symptoms of affective disorders whereas credibility might have more explanatory value for variables reflecting the effectiveness of specific CBT techniques. It should be examined in future research, if expectancy also has explanatory value for the specific variables credibility was significantly associated with. By now, it can be carefully assumed that expectancy is the more robust predictor for the outcomes of CBT. Measurements of
expectancy could therefore be more useful to make predictions about a patient’s outcome in CBT than measurements of credibility.

However, it should not be forgotten that expectancy develops upon how credible a treatment seems (Ametrano, 2011). That is, targeting credibility likely helps to increase, unfold and utilize the ameliorative effects of outcome expectancy. Moreover, credibility has been found to strongly influence the quality of working alliance (Söchting, Tsai & Ogrodniczuk, 2016). This finding is extremely valuable for therapist in training who are often unsure about how to build an adequate working relationship with patients. Emphasizing the credibility of a treatment rationale could also help therapists to create a material working alliance with patients that have greater difficulties to build a relationship with the therapist as for example borderline patients (Yeomans, 1994). However, when expectancy and credibility indeed have greater explanatory value for the outcomes of CBT than alliance, less attention should be paid to the dynamics of the therapeutic relationship and even more focus should be set on the build-up of positive credibility beliefs and outcome expectancy to facilitate a greater engagement with the CBT techniques. The armamentarium of CBT readily provides techniques to increase a patient’s expectancy and credibility like cognitive restructuring and psychoeducation (Chambless et al., 2017; Gonzalez-Pinto et al., 2004) thereby enabling the patient to tap the full potential of CBT.

Much research existed about why patients fail in CBT instead of why they succeed. Research on expectancy and credibility provides a strength-based approach that might entail highly realizable practical indications to increase a patient’s chance of achieving a successful outcome in CBT (Demetriou & Schmitz-Sciborski, 2011) possibly regardless of his clinical characteristics and the quality of the alliance with the therapist in charge.

References


Frank, JD (1961): Persuasion and Healing, Baltimore: Johns Hopkins Press


APPENDIX

Appendix A.

Table 1.
Quality criteria applied on records in the final data pool.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1 Sample size</td>
<td>N &lt; 30</td>
</tr>
<tr>
<td></td>
<td>30 ≤ N &lt; 100</td>
</tr>
<tr>
<td></td>
<td>N ≥ 100</td>
</tr>
<tr>
<td></td>
<td>not reported</td>
</tr>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>NR</td>
</tr>
<tr>
<td>#2 CBT was conducted by clinical psychotherapists, psychiatrists or</td>
<td>not the case</td>
</tr>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>professionals.</td>
<td>partially applies</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>applies</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>not reported/clarified</td>
</tr>
<tr>
<td></td>
<td>NR</td>
</tr>
<tr>
<td>#3 Completers and drop-outs were compared on initial assessments of</td>
<td>no comparison</td>
</tr>
<tr>
<td>outcome expectancy and/or credibility beliefs.</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>comparison</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>data of full sample available</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>not reported/no drop-out info</td>
</tr>
<tr>
<td></td>
<td>NR</td>
</tr>
<tr>
<td>#4 Validated and reliable measuring tools were used to assess outcome</td>
<td>not used</td>
</tr>
<tr>
<td>expectancy and credibility beliefs.</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>used</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>#5 Outcome expectancy and credibility beliefs were assessed after</td>
<td>not the case</td>
</tr>
<tr>
<td>patients had entered the treatment.</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>applies</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>#6 If #5 applies: Clarification if early-in-treatment assessments of</td>
<td>not clarified</td>
</tr>
<tr>
<td>outcome expectancy or credibility beliefs had been assessed after</td>
<td>0</td>
</tr>
<tr>
<td>rationale was provided by therapist in charge.</td>
<td>clarified</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>#7 Validated and reliable measuring tools had been used to measure</td>
<td>not used</td>
</tr>
<tr>
<td>process and outcome variables that were assessed on their respective</td>
<td>0</td>
</tr>
<tr>
<td>relationships with outcome expectancy and credibility beliefs.</td>
<td>used</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>#8 Quality/appropriateness of statistical analyses conducted to assess</td>
<td>lower quality/appropriateness</td>
</tr>
<tr>
<td>direct and indirect relationships.</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>higher quality/appropriateness</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Sum-score</td>
<td>∑ =10</td>
</tr>
</tbody>
</table>
Table 2. Results from quality assessment.

<table>
<thead>
<tr>
<th></th>
<th>#1 Sample size (0 – 2)</th>
<th>#2 Therapist (0-2)</th>
<th>#3 Dropout-analysis (0 - 1)</th>
<th>#4 Validity and reliability of assessment tool used to measure expectancy or credibility (0 - 1)</th>
<th>#5 Measurement point (0 - 1)</th>
<th>#6 Rationale provision considered (0 – 1)</th>
<th>#7 Validity and reliability of assessment tool used to measure outcome/process variables (0 – 1)</th>
<th>#8 Quality/appropriateness of statistical analyses conducted (0 – 1)</th>
<th>Total (0 – 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Chambless et al. (1997)</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>2.</td>
<td>Chambless et al. (2017)</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>3.</td>
<td>Dozois &amp; Westra (2005)</td>
<td>1</td>
<td>2</td>
<td>NR</td>
<td>1</td>
<td>0</td>
<td>-</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>4.</td>
<td>Meyerhoff &amp; Rohan (2016)</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>5.</td>
<td>Vîslă et al. (2018)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>6.</td>
<td>Webb et al. (2014)</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>7.</td>
<td>Westra et al. (2007)</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>-</td>
<td>1</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>8.</td>
<td>Westra et al. (2011)</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>9.</td>
<td>Andersson et al. (2008)</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>
Table 2. (continued)
Results from quality assessment.

<table>
<thead>
<tr>
<th>Sample size (0 – 2)</th>
<th>Therapist (0-2)</th>
<th>Dropout-analysis (0 - 1)</th>
<th>Validity and reliability of assessment tool used to measure expectancy or credibility (0 - 1)</th>
<th>Measurement point (0 - 1)</th>
<th>Rationale provision considered (0 – 1)</th>
<th>Validity and reliability of assessment tool used to measure outcome/process variables (0 – 1)</th>
<th>Quality/appropriateness of statistical analyses conducted (0 – 1)</th>
<th>Total (0 – 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardy et al., (1995)</td>
<td>1</td>
<td>2</td>
<td>NR</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Söchting et al. (2016)</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Hundt et al. (2013)</td>
<td>1</td>
<td>0</td>
<td>NR</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Hundt et al. (2014)</td>
<td>2</td>
<td>NR</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Riley (2015)</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>16/28</strong></td>
<td><strong>12/28</strong></td>
<td><strong>6/14</strong></td>
<td><strong>5/14</strong></td>
<td><strong>9/14</strong></td>
<td><strong>8/14</strong></td>
<td><strong>13/14</strong></td>
<td><strong>8/14</strong></td>
</tr>
</tbody>
</table>
Table 3.  
*Overview of questionnaires utilized to assess expectancy and credibility.*

<table>
<thead>
<tr>
<th>Outcome expectancy</th>
<th>times utilized</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEQ-E⁺</td>
<td>3 x</td>
</tr>
<tr>
<td>ACES⁺</td>
<td>2 x</td>
</tr>
<tr>
<td>OES</td>
<td>1 x</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Credibility beliefs</th>
<th>times utilized</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS</td>
<td>1 x</td>
</tr>
<tr>
<td>TCF</td>
<td>1 x</td>
</tr>
<tr>
<td>OPP</td>
<td>1 x</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Expectancy &amp; credibility</th>
<th>times utilized</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERS</td>
<td>2 x</td>
</tr>
<tr>
<td>CES</td>
<td>3 x</td>
</tr>
</tbody>
</table>

ACES: Anxiety Change Expectancy Scale; CES: Credibility/Expectancy Scale; CEQ-E: Credibility Expectancy Questionnaire, expectancy subscale; CS: Credibility Scale; OES: Outcome Expectancies Scale; OPP-TES: Opinions about Psychological Problems Questionnaire, Treatment Endorsement subscale; TCF: Treatment credibility form
Table 4. List of outcome and process variables examined on respective relationships with expectancy or credibility.

<table>
<thead>
<tr>
<th>Outcome variables assessed</th>
<th>Outcome expectancy</th>
<th></th>
<th></th>
<th>Credibility Beliefs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>times assessed</td>
<td>significantly related to expectancy</td>
<td>times assessed</td>
<td>significantly related to credibility</td>
</tr>
<tr>
<td>ACQ</td>
<td>-</td>
<td>-</td>
<td>1 x</td>
<td>-</td>
</tr>
<tr>
<td>Anxiety &amp; skill (observer rated)</td>
<td>-</td>
<td>-</td>
<td>1 x</td>
<td>-</td>
</tr>
<tr>
<td>Anxious apprehension</td>
<td>-</td>
<td>-</td>
<td>1 x</td>
<td>1 x</td>
</tr>
<tr>
<td>ASI</td>
<td>1 x</td>
<td>1 x</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BAI</td>
<td>1 x</td>
<td>1 x</td>
<td>1 x</td>
<td>-</td>
</tr>
<tr>
<td>BDI</td>
<td>-</td>
<td>-</td>
<td>3 x</td>
<td>1 x</td>
</tr>
<tr>
<td>BSQ</td>
<td>-</td>
<td>-</td>
<td>1 x</td>
<td>1 x</td>
</tr>
<tr>
<td>CESD-10</td>
<td>1 x</td>
<td>1 x</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CSQ</td>
<td>-</td>
<td>-</td>
<td>1 x</td>
<td>1 x</td>
</tr>
<tr>
<td>Dyad anxiety &amp; skill</td>
<td>-</td>
<td>-</td>
<td>1 x</td>
<td>1 x</td>
</tr>
<tr>
<td>FNE-B</td>
<td>1 x</td>
<td>1 x</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>IPP-28</td>
<td>-</td>
<td>-</td>
<td>1 x</td>
<td>1 x</td>
</tr>
<tr>
<td>PDSS</td>
<td>1 x</td>
<td>1 x</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PSWQ</td>
<td>4 x</td>
<td>3 x</td>
<td>1 x</td>
<td>-</td>
</tr>
<tr>
<td>PSWQ-A</td>
<td>-</td>
<td>-</td>
<td>1 x</td>
<td>-</td>
</tr>
<tr>
<td>QOLI</td>
<td>-</td>
<td>-</td>
<td>1 x</td>
<td>-</td>
</tr>
<tr>
<td>SCL-90</td>
<td>-</td>
<td>-</td>
<td>1 x</td>
<td>-</td>
</tr>
<tr>
<td>SE</td>
<td>-</td>
<td>-</td>
<td>1 x</td>
<td>-</td>
</tr>
<tr>
<td>speech anxiety</td>
<td>-</td>
<td>-</td>
<td>1 x</td>
<td>-</td>
</tr>
<tr>
<td>speech skill</td>
<td>-</td>
<td>-</td>
<td>1 x</td>
<td>1 x</td>
</tr>
<tr>
<td>STAI-T</td>
<td>1 x</td>
<td>-</td>
<td>1 x</td>
<td>1 x</td>
</tr>
<tr>
<td>VGS-HS</td>
<td>1 x</td>
<td>-</td>
<td>1 x</td>
<td>-</td>
</tr>
<tr>
<td>Process variables assessed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AWTS</td>
<td>-</td>
<td>-</td>
<td>1 x</td>
<td>1 x</td>
</tr>
<tr>
<td>CALPAS</td>
<td>1 x</td>
<td>1 x</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Homework compliance</td>
<td>1 x</td>
<td>1 x</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>WAI</td>
<td>1 x</td>
<td>1 x</td>
<td>1 x</td>
<td>1 x</td>
</tr>
</tbody>
</table>

ACQ: Agoraphobic Cognitions Questionnaire; ASI: Anxiety Sensitivity Index; AWTS: Adherence with Treatment Scale; BAI: Beck Anxiety Inventory; BDI-(II): Beck Depression Inventory; BSQ: Body Sensations Questionnaire; CALPAS: California Psychotherapy Alliance Scale; CES-D-10: Centre for the Epidemiological Studies of Depression-10; CSQ: Client Satisfaction Questionnaire; FNE: Fear of Negative Evaluation Scale (FNEB: brief version); IPP-(28): inventory of interpersonal problems; PDSS: panic disorder severity scale; PSWQ: Penn State Worry Questionnaire (PSWQ-A: abbreviated version); QOLI: Quality of Life Inventory; SE: measure of self-esteem; SCL-90: symptom checklist of psychological problems; STAI-T: Spielberger State-Trait Anxiety Inventory, trait subscale; VGS-HS: Victorian Gambling Screen Harm to Self subscale; WAI: Working Alliance Inventory
Appendix B.

Content of questionnaires

Note: A comparative analyses of questionnaires was conducted to assess their respective appropriateness to measure expectancy or credibility. The definitions of outcome expectancy (i.e. a patients’ expected improvement through a specific treatment) and credibility beliefs (i.e. estimation about the logic, plausibility, suitability of a specific treatment) as well as the validated and reliable scales of the ACES and CEQ were used as benchmarks for this comparative analysis. Items matching to the conceptualization of credibility were underlined and a squiggly line was put under items matching to the conceptualization of expectancy.

Credibility Expectancy Questionnaire: CEQ (Devilly & Borkovec, 2000, partially adapted from Borkovec & Nau, 1972)

Expectancy:
1. How much improvement in your symptoms do you think will occur.
2. How much do you really feel that therapy will help you to reduce your symptoms.
3. How much improvement in your symptoms do you really feel will occur.

Credibility:
1. How logical does the therapy offered to you seem?
2. How successfully do you think this treatment will be in reducing your symptoms?
3. How confident would you be in recommending this treatment to a friend?

Anxiety change expectancy scale: ACES (Dozois & Westra, 2005 partially adapted from Borkovec & Nau, 1972)

Expectancy:
1. I feel pessimistic that my anxiety problems could ever change for the better.
2. Even though I try, nothing seems to help with my anxiety.
3. It would be extremely difficult or impossible to solve my problems with anxiety.
4. I have had some positive experiences with being able to control my anxiety through talking positively to myself.
5. My problems with anxiety are too severe to benefit from treatment.
6. Self-help methods may help others control their anxiety, but they won’t work for me.
7. I don’t believe I will ever feel truly relaxed and not worried.
8. Facing my fears has never helped me to reduce my anxiety.
9. When I force myself to do something that scares me, often it’s not as bad as I thought.
10. I have had some success in reducing my anxiety.
11. There is very little anyone could do to help me solve my anxiety problems.
12. Even when I try to talk positively to myself, it doesn’t help my anxiety.
13. Positive thinking is helpful to me in managing my anxiety.
14. There is no solution to my anxiety problems.
15. I am optimistic that my anxiety can change for the better.
16. I have found that I can reduce my anxiety by telling myself to relax or by using relaxation exercises.
17. I’ll never be able to control my anxiety and worry.

**Credibility Expectancy Scale:** CES (Borkovec & Nau, 1972)

**Credibility/Expectancy:**
1. How logical does this type of treatment seem to you?
2. How confident would you be that this treatment would be successful in eliminating fear of speaking before a group?
3. How confident would you be in recommending this treatment to a friend who was extremely anxious about making speeches?
4. If you were extremely anxious in speech situations, would you be willing to undergo such treatment?
5. How successful do you feel this treatment would be in decreasing a different fear; for example, strong anxiety about taking test?

**Expectancy rating scale:** ERS (Hundt et al. 2013 partially adapted from Borkovec & Nau, 1972)

**Expectancy:**
1. How much improvement do you expect to result from treatment?

**Credibility:**
2. 3 items measuring beliefs about treatment rationale

**Outcome expectancy scale:** OES (Ogrodniczuk & Söchting, 2010 partially adapted from Borkovec & Nau, 1972)

**Expectancy:**
1. How much do you expect to recover from your problems after therapy?
2. How successful do you think this therapy will be in helping with your problems?
3. How confident are you that this therapy will help you?

**Credibility Scale:** CS (Söchting, 2009)

**Credibility:**
1. How **credible** do you **think** this therapy is?
2. How **logical** does this therapy seem to you?
3. How much do you **trust** this therapy to help you?
4. How **credible** does your therapist seem?
5. How much do you **trust your therapist** to be able to help you?
6. How **knowledgeable** does your therapist appear?

**Treatment credibility form:** TCF (Morrision & Shapiro, 1987 partially adopted from Borkovec & Nau, 1972)

**Credibility:**
1. How **logical** does this treatment seem to you?
2. How **useful** does this treatment seem to you?
3. How **confident** are you that this treatment will be **successful**?
4. How **confident** would you be in **recommending** this treatment to a friend with similar difficulties?

**Opinions about psychotherapy questionnaire – treatment endorsement subscale:** OPP-TES (Pistrang & Barker, 1992)

**Credibility:**

A good way to help me with my problems would be…

1. learning to replace negative thoughts with positive ones
2. tackling the problem in a step by step way
3. having an expert to teach one specific ways of changing one’s behaviour
4. having an expert to point out that how one thinks about oneself can sometimes be wrong