(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

J. Behav. Ther. & Exp. Psychiat. 43 (2011) 540–547

Contents lists available at ScienceDirect

Journal of Behavior Therapy and Experimental Psychiatry

journal homepage: www.elsevier.com/locate/jbtep

Examining the effects of thought records and behavioral experiments in instigating belief change

F. McManus [a,b,*], K. Van Doorn [c,1], J. Yiend [d,2]

[a] Oxford Cognitive Therapy Centre, Warneford Hospital, Oxford OX3 7JX, UK
[b] Oxford University, Department of Psychiatry, Warneford Hospital, Oxford OX3 7JX, UK
[c] Clinical Psychology, Oxfordshire & Buckinghamshire Mental Health NHS Foundation Trust, Warneford Hospital, Oxford OX3 7JX, UK
[d] Division of Psychological Medicine and Psychiatry, Institute of Psychiatry, PO 63, De Crespigny Park, London SE5 8AF, UK

ARTICLE INFO

Article history:
Received 1 April 2011
Received in revised form
6 June 2011
Accepted 5 July 2011

Keywords:
Cognitive behavior therapy
Component analysis
Thought record
Behavioral experiment

ABSTRACT

Objective: While the efficacy and effectiveness of CBT protocols are well established, much less is known about the comparative contribution of the various techniques within CBT. The present study examined the relative efficacy, in comparison to a control condition, of two central techniques in CBT: thought records (TRs) and behavioral experiments (BEs).
Method: A mixed within and between participants design was used to compare the efficacy of a single session TR and a single session BE intervention with a control intervention, in a non-clinical sample. Ninety one participants were randomly allocated to one of the three conditions.
Results: The overall pattern of results suggests that both TR and BE had a beneficial therapeutic impact in comparison to the control condition on beliefs, anxiety, behavior and a standardized measure of symptoms. There was evidence of a small advantage of the BE over the TR intervention in that the target belief changed earlier and change generalized to beliefs about others as well as the self.
Conclusions: The findings confirm the utility of both TR and BE interventions and point to BEs as more useful in effecting belief change in that the change in the BE condition occurred sooner and generalized further.

1. Introduction

Although Cognitive behavior therapy (CBT) has demonstrated efficacy as a treatment package (Roth & Fonagy, 2004), however little is known about the efficacy of individual components of CBT. Most CBT efficacy research focuses on the effect of multi-component protocols (e.g., Westbrook & Kirk, 2005) and while this approach reflects clinical practice, it cannot identify the critical ingredients responsible for the efficacy or compare the efficacy of the different components. There are a limited number of studies that attempt to identify the active ingredients of CBT (Longmore & Worrell, 2007), particularly in areas other than the treatment of depression (Dobson & Khatri, 2000). It is important to evaluate the efficacy of individual treatment components independently of the overall treatment program for three reasons. First, while CBT has proven to be effective, there remains room for improvement – all trials of CBT have found that a significant proportion of treated patients remain symptomatic and/or fail to reach an optimal level of functioning. Hence, evaluating individual components may help to show how to improve efficacy. Second, recent research suggests that therapists may be particularly poor at implementing components of CBT that require exposure to feared situations and thus produce a temporary increase in patients' distress (Becker, Zayfert, & Anderson, 2004; Schulte & Eifert, 2002; Waller, 2009). This suggests that the choice of technique or intervention may be determined by therapists' preferences rather than the intervention's demonstrated efficacy. The third reason for evaluating efficacy of individual treatment components is to enhance the cost-effectiveness of CBT protocols by eliminating any less effective or unnecessary procedures. New cost effective procedures are being developed reducing the need for face to face intervention (e.g., Andersson & Cuijpers, 2008; Lester, Mathews, Davison, Burgess, & Yiend, 2011; Mathews, Ridgeway, Cook, & Yiend, 2007; Warmerdam, Smit, van Straten, Riper, & Cuijpers, 2010). In this context it is important that intensive face to face techniques

* Corresponding author. Oxford University, Department of Psychiatry, Warneford Hospital, Oxford OX3 7JX, UK. Tel.: +44 1865 613 145; fax: +44 1865 226 331.
E-mail addresses: freda.mcmanus@psych.ox.ac.uk (F. McManus), karlijn.van-doorn@hmc.ox.ac.uk (K. Van Doorn), jenny.yiend@kcl.ac.uk (J. Yiend).
[1] Tel.: +44 1865 226 333.
[2] Tel.: +44 2078 485 119; fax: +44 (0) 2078 481 000.

0005-7916/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jbtep.2011.07.003

continue to develop their efficacy and demonstrate their mechanisms of action. Evaluations of individual treatment components will help to do this by identifying the most effective elements of existing procedures.

The aim of the present study was to evaluate the relative efficacy of two commonly used CBT components, namely completing thought records (TRs) and carrying out behavioral experiments (BEs). The completion of a TR involves exploring the identified belief, questioning the patient's evidence in support of this belief, using Socratic questioning to identify evidence that does not support the belief and finally, specifying an alternative more balanced belief that incorporates both the evidence that does and does not support the original belief. The information generated during the TR is detailed on a written record (see Greenberger & Padesky, 1995 for a fuller description and examples of a seven-column thought record). In contrast, BEs are planned, experiential activities designed to test the validity of an identified belief. To structure the BE, the therapist and patient collaboratively complete a BE record sheet identifying the patient's specific prediction and designing an experiment to test that prediction (see Bennett-Levy et al., 2004 for further information and examples). After carrying out the experiment, they record the outcome of the experiment and review the implications for the belief being tested. BEs differ from exposure in that the theoretical model on which exposure interventions are based is a behavioral one that suggests that their therapeutic efficacy is achieved via the habituation of anxiety, whereas BEs are established within a cognitive rationale and attempt to achieve symptom change via cognitive change (belief disconfirmation) arising as a result of the BE (see McMillan & Lee, 2010 for a review of the comparative efficacy of BEs vs exposure).

TRs can be said to be the core technique in CBT as originally conceived by Beck (1976) and Beck, Rush, Shaw, and Emery (1979) and have been a central component of the CBT protocols used in the randomized controlled trials demonstrating the efficacy of CBT (Elkin et al., 1989). However, although exposure is a well established CBT treatment procedure, it is only in recent years that BEs have been incorporated into CBT protocols. The efficacy of BEs as a component of CBT protocols has been demonstrated (McManus et al., 2009; Salkovskis, Hackmann, Wells, Gelder, & Clark, 2007; Wells et al., 1995) and it has been suggested that they may be a particularly powerful means of achieving belief change and symptom relief: "*Behavioral experimentation is widely regarded as the single most powerful way of changing cognitions*" (Waller, 2009, p. 602); "*behavioral strategies offer the most powerful means to cognitive change in cognitive therapy*" (Wells, 1997, p. 78). Furthermore the merit of verbal strategies alone has been questioned: "*Beliefs rarely change as a result of intellectual challenge, but only through engaging emotions and behaving in new ways that produce evidence that confirms new beliefs*" (Chadwick, Birchwood, & Trower, 1996, p. 37).

While both TRs and BEs are commonly used techniques in CBT, there have been few direct comparisons of their relative efficacy and there is little empirical evidence to help CBT therapists in choosing which technique to employ in order to best achieve the desired belief change within the limited time available. Two early studies found no difference in the efficacy of interventions that correspond to TRs and BEs (Jarrett & Nelson, 1987; O'Donohue & Szymanski, 1993). In contrast, more recent studies have suggested an advantage of BEs over TRs (Bennett-Levy, 2003; Tang & Harvey, 2006). There are a number of possible reasons for these inconsistent findings. First, small sample sizes in the earlier studies mean that these may have been underpowered to detect a difference between two active treatments (sample sizes ranged between 27 and 48). Second, the lack of control groups means the efficacy of the interventions over the passage of time/non-specific factors was not established. Third, both Jarrett and Nelson (1987) and O'Donohue and Szymanski (1993) used group interventions which may have smaller effect sizes than individual CBT (e.g., Mortberg, Clark, Sundin, & Wistedt, 2007; Stangier, Heidenreich, Peitz, Lauterbach, & Clark, 2003). Furthermore, the results may not generalize to the individual format in which CBT interventions are typically carried out. Fourth, BEs have only been a focus of interest in CBT in recent years (indeed the first book on BEs was published in 2004 by Bennett-Levy and colleagues). Thus the technique of BEs has undergone refinement in recent years involving increased specification, dissemination and use. It is likely that this more widespread use will have led to refinement of the technique that has enhanced its efficacy (Salkovskis, 2002).

1.1. The present study: design and aims

The present study used a larger sample size with random allocation to one of the two active interventions (BE or TR) or a no-treatment control group. A mixed within and between participants design with individual treatment was used, with assessment of outcome immediately after the intervention and at one week follow-up. Each intervention was manualized and limited to a single session, to provide a 'pure' comparison that avoids the inherent variation that can occur when treatment is delivered over a large number of sessions (Shapiro & Startup, 1992).

The primary aim of the study was to compare the efficacy of TRs and BEs to a control intervention in effecting belief and analog symptom change from a single session intervention. It was hypothesized that both the TR and BE interventions would produce greater reductions in participants' belief, anxiety and behavior ratings, and scores on standardized symptom measures, than the control intervention. Further, it was hypothesized that the BE intervention may produce greater reductions on these measures than the TR intervention.

2. Method

2.1. Design

In order to recruit a large sample, and most importantly to keep the content of the interventions constant across participants and within each condition, it was decided to focus on a specified belief that is commonly held in the general population but that could be considered at least somewhat irrational e.g., superstitious beliefs (Newport & Strausberg, 2001) or subclinical OCD related beliefs (Gibbs, 1996). An email survey of the authors' associates ($n = 128$) was conducted to determine the prevalence of such beliefs in the local population. To ensure that belief change was possible in the interventions a belief rating of >60% was used as a cut off and the belief most frequently endorsed at this level was 'Not washing your hands after going to the toilet will make you ill', which was given a belief rating of 60% or above by 32% of the 128 pilot study participants. Whilst there is some validity to this belief, research shows that a significant proportion of the population do not wash their hands after going to the toilet (Drankiewicz & Dundes, 2003) and that 'normal' hand-washing is unlikely to be sufficient to remove pathogens (Moe, Christmas, Echols, & Miller, 2001). This belief is also ethically and practically amenable to intervention via either TR or BE in a single session (in contrast to, for example, the belief that seeing one magpie brings sorrow but two brings joy, which would be practically difficult to arrange) and emulates the clinical situation of testing out a belief that is likely to generate some anxiety about possible negative consequences.

The study used a mixed within/between participants design with random allocation[3] to one of the three conditions: (i) TR intervention (ii) BE intervention and (iii) Control group (no active intervention). To assess the impact of the three conditions participants completed belief, anxiety and behavior ratings, and self-report symptom measures pre-intervention, post-intervention and at one week follow-up.

2.2. Participants and recruitment

Participants were recruited via advertisements in the local press and Universities. Via online screening, potential participants rated, from 0 = 'not at all' to 100 = 'completely', how much they endorsed the target belief. Participants were excluded if they were not fluent in English or reported a current or past history of psychiatric disorder. Ninety-eight of the 248 (41.88%) responders gave a target belief rating of 60% or more and were invited to participate. Participants were reimbursed £7.00 and, if relevant, received course credits for completing the study.

2.3. Measures

2.3.1. Belief, anxiety and behavior ratings

Outcome measures were chosen to reflect those commonly used in clinical practice (e.g., Greenberger & Padesky, 1995; Westbrook, Kennerly, & Kirk, 2007). Participants rated their degree of conviction in a statement that reflected the belief identified as the target of the intervention (belief ratings). They also rated how much emotion they would experience should a relevant situation occur (here, anxiety ratings). Hence, the primary outcome measure was (i) the degree of endorsement of the target belief "not washing your hands after going to the toilet will make you ill". Additional outcomes were (ii) the degree of endorsement of the related belief "not washing your hands after going to the toilet will make *someone else* ill" and (iii) the amount of anxiety that the participant would feel about having to go to the toilet without washing their hands. In addition, to assess the likely impact of the interventions on actual behavior, participants were asked to rate how likely they would be to engage in the target situation (going to the toilet without washing their hands) in the next week. Belief ratings were made on a 9-point Likert scale from 0 = 'do not believe this at all' to 9 = 'strongly believe this to be true'. Anxiety ratings were made on a 9-point Likert scale from 0 = 'not at all anxious' to 9 = 'extremely anxious'. Behavior ratings were made on a 9-point Likert scale from 0 = 'not at all likely' to 9 = 'extremely likely'. Belief, anxiety and behavior ratings were made pre- and post-intervention and at one week follow-up.

2.3.2. Standardized symptom measures

To assess the wider impact of the interventions, standardized measures of obsessional and irrational beliefs, the Obsessional Beliefs Questionnaire-44 (OBQ-44; Obsessive Compulsive Cognitions Working Group, 2005) and Irrational Belief Inventory (IBI; Koopmans, Sanderman, Timmerman, & Emmelkamp, 1994) were completed pre-intervention, post-intervention and at follow-up. The OBQ-44 consists of 44 obsessional beliefs which are rated on a 7-point Likert scale, and has been shown to have adequate psychometric properties (Tolin, Brady, & Hannan, 2008). The IBI is a 50-item measure of general irrational beliefs that uses a 5-point Likert scale and has been shown to have adequate psychometric properties (Bridges & Sanderman, 2002).

2.3.3. Therapeutic alliance and credibility of interventions

In order to determine whether any differences between the TR and BE conditions were due to differences in the alliance or credibility of the interventions, appropriate measures were administered immediately following the interventions. The Scale to Assess Therapeutic Relationship patient version (STAR; McGuire-Snieckus, McCabe, Catty, Hansson & Priebe, 2007) was used to assess the therapeutic alliance. The STAR is a 12-item scale measuring the relationship between therapist and patient on a 4-point Likert scale. The STAR purports to overcome shortcomings of previous measures of therapeutic alliance and has been shown to have adequate psychometric properties (McGuire-Snieckus et al., 2007). The Credibility Expectancy Questionnaire (CEQ; Borkovec & Nau, 1972) was used to assess participants' perceived credibility of the interventions. The CEQ is a 5-item scale that assesses the credibility and expectancy of interventions on a 9-point Likert scale, and it has been shown to have adequate psychometric properties (Devilly & Borkovec, 2000).

2.4. Procedure

Participants were consented, randomly allocated[3] to one of the three experimental conditions and asked to complete the pre-intervention measures (belief ratings, anxiety ratings, behavior ratings and standardized symptom measures). Interventions were carried out by one of two graduate students, fully trained and supervised in the procedures by the first author, a Director of a CBT training program (University of Oxford) who has substantial experience of carrying out manualized CBT treatments in randomized controlled trials and of training therapists in CBT interventions. Both of the graduate students delivered both of the interventions so that any effects could not be attributed to the therapist.

2.5. Experimental conditions[4]

All experimental conditions were of 30 min duration.

2.5.1. Thought record (TR) intervention

The TR intervention involved the experimenter guiding the participant through the completion of a thought record (in the manner described by Greenberger and Padesky (1995)). Participants were asked to rate how much they believed the target belief ('not washing your hands after going to the toilet will make you ill') and to specify the details and timescale of any illness they might get from not washing their hands. Then the experimenter asked them to identify any evidence that supported their belief (e.g., parents' beliefs, information in the media, personal experiences) and any that did not support their belief (e.g., observations of the frequency of omissions or ineffectiveness of hand-washing, personal experience of instances where people have not washed their hands but have not become ill). Participants were prompted to identify further evidence and reflect on their own experiences of not washing their hands after going to the toilet. After reviewing the evidence for and against the belief in detail participants formulated a 'balanced alternative belief' summarizing both the evidence for

[3] To ensure that the three conditions were balanced with regard to baseline belief rating of the target belief and gender, Pocock's minimization method (1983, p. 84–85) was used to allocate participants to the condition that 'minimized' any existing imbalance between the three conditions.

[4] To ensure adherence and treatment fidelity 30% of TR and BE interventions were recorded and categorized by an independent rater, blind to participants' allocation. The content of the two active intervention groups was clearly distinguishable and adhered to the manual (100% correctly classified as TR or BE).

and against the target belief e.g., "Although I would feel dirty if I did not wash my hands after going to the toilet, I most likely would not get ill from it".

2.5.2. Behavioral experiment (BE) intervention

The BE intervention involved the experimenter guiding the participant through the completion of a BE record (in the manner described by Bennett-Levy et al. (2004)). The BE intervention was identical to the TR intervention until the discussion of evidence for and against the target belief. At this point BE participants were asked to devise an experiment to test the validity of the target belief (e.g., to pass urine without washing their hands afterward to find out if they did become ill). As part of completing the behavioral experiment record sheet participants specified exactly what they would do during the experiment and how they would judge the outcome in relation to the target belief (e.g., how they would know if they became ill or not). Participants were then asked to carry out the experiment during the session. They then reviewed the implications of the experiment for their target belief. In line with the principles of BEs (Bennett-Levy et al., 2004) the experimenter encouraged participants to test their belief as fully as possible (e.g., if they believed that they were more likely to become ill from not washing their hands after going to the toilet if they then touched their face, they were encouraged to test this out).

2.5.3. Control group (no active intervention)

The control participants did not receive an active intervention but spent 30 min reading a neutral text, answering questions relating to it and completing mathematical calculations.

Immediately after all of the interventions, participants completed the post-intervention measures (belief ratings, anxiety ratings, behavior ratings and standardized symptom measures) and made an appointment to return in one week's time to complete the same measures a final time at follow-up.

2.6. Data analytic strategy

A MANOVA was conducted with Group (TR, BE, control) as a between participants factor and Time (pre-intervention, post-intervention, follow-up) as a within participants factor. This was followed up with corresponding univariate repeated measures Group × Time ANOVAS for each dependent measure separately. Significant interactions were followed up with one way ANOVAs comparing all three groups at each time point separately. Finally Bonferroni corrected comparisons were used to identify significant between group differences. This approach controls the family wise Type I error rate in the manner recommended by Howell (2010).

3. Results

3.1. Participant characteristics

Of the 98 eligible participants who responded to the adverts, 91 (92.9%) accepted the invitation to participate in the study. Thirty one (6 male, 25 female) were allocated to the BE group, 30 (6 male, 24 female) to the TR group and 30 (9 male, 21 female) to the control group. In the BE group follow-up data is missing for 3 participants, who failed to attend the follow-up appointment. Table 1 shows participants' characteristics prior to intervention.

As expected, there were no significant differences between groups on: gender, previous psychological treatment, ethnic group, education level, or age (all p's $> .4$). Nor were there any differences between the groups on the OBQ ($F(2,88) = 0.001$, $p = .99$) or the IBI ($F(2,88) = 0.12$, $p = .89$). Most importantly there was no difference between the groups on the primary outcome measure, the endorsement of the target belief ($p > .6$).

3.2. Intervention checks

Measures of therapeutic alliance (STAR) and credibility of the intervention (CEQ) were compared for the two active interventions (TR and BE). TR was perceived to be marginally more credible than BE (mean rating of 32.77 (SD = 6.98) vs 29.82 (SD = 6.42)) but this difference failed to reach significance ($t(57) = 1.71$, $p = .09$). There was no difference between the TR and BE groups in STAR ratings (35.64 (SD = 8.61) vs 34.61 (SD = 5.92), $t(57) = 0.54$, ns).

Table 1
Participant characteristics and scores at baseline.

Measure (scale)	Group			$F(2,88)$	p Value
	TR mean (SD)	BE mean (SD)	Control mean (SD)		
Characteristics					
Sex (frequency)				1.21 (χ^2)	ns
Women	24	25	21		
Men	6	6	9		
Age (years)	24.22 (8.71)	22.73 (7.96)	23.14 (9.79)	0.265	ns
Ethnic group (frequency)				0.09 (χ^2)	ns
Caucasian	22	22	21		
Other	8	9	9		
Highest education (frequency)				1.51 (χ^2)	ns
BSc or more	11	14	9		
A-levels or less	19	17	21		
Previous therapy (frequency)				1.68 (χ^2)	ns
Yes	5	8	4		
No	25	23	26		
Belief rating at screening (0–100)	77.63 (13.69)	78.97 (12.30)	79.33 (11.43)	0.154	ns
Belief, anxiety & behavior ratings					
Not washing hands after going to the toilet will make you ill (1–9)	7.00 (1.08)	6.90 (1.09)	6.70 (1.34)	0.50	ns
Not washing hands after going to the toilet will make others ill (1–9)	6.03 (1.92)	6.03 (1.81)	6.13 (1.63)	0.03	ns
Anxiety if I had to go to toilet without washing hands afterward (1–9)	5.33 (2.29)	5.39 (1.96)	4.93 (2.26)	0.39	ns
Likelihood of using the toilet without washing hands next week[a] (1–9)	7.77 (1.57)	8.19 (0.91)	8.10 (1.40)	0.88	ns
Standardized measures					
OBQ-44	157.97 (38.84)	158.00 (40.79)	158.40 (35.32)	0.00	ns
IBI	147.43 (18.93)	146.45 (19.61)	145.10 (18.10)	0.12	ns

[a] Scores reversed.

Table 2
Means and standard deviations for the three groups (TR, BE and Control) at pre-intervention, post-intervention and follow-up.

Measure	Group	Pre-intervention mean (SD)	Post-intervention mean (SD)	Follow-up mean (SD)
Belief ratings				
Not washing your hands after going to the toilet will make you ill (1–9)	TR	7.03 (1.09)	5.21 (1.78)	5.14 (1.60)
	BE	6.96 (1.07)	4.82 (1.72)	4.79 (1.79)
	C	6.70 (1.34)	6.17 (1.88)	6.40 (1.71)
Not washing your hands after going to the toilet will make others ill (1–9)	TR	6.03 (1.95)	4.86 (1.98)	4.52 (1.88)
	BE	6.11 (1.77)	4.50 (2.03)	4.14 (1.88)
	C	6.13 (1.63)	5.60 (2.33)	5.57 (2.22)
Anxiety rating (1–9)	TR	5.24 (2.28)	4.97 (2.49)	4.76 (2.57)
	BE	5.38 (1.80)	4.79 (2.14)	4.28 (2.15)
	C	4.93 (2.26)	4.90 (2.51)	5.13 (2.58)
Behavior rating (1–9) (reversed)	TR	7.76 (1.60)	7.24 (1.96)	7.41 (2.11)
	BE	8.17 (0.93)	7.21 (2.16)	7.14 (2.12)
	C	8.10 (1.40)	8.13 (1.43)	8.13 (1.07)
Symptom measures				
OBQ	TR	155.31 (36.65)	148.76 (36.05)	147.34 (32.37)
	BE	160.55 (40.95)	147.66 (42.18)	143.86 (40.18)
	C	158.40 (35.32)	153.07 (32.13)	154.60 (35.27)
IBI	TR	146.79 (18.93)	145.31 (18.29)	144.03 (17.00)
	BE	146.70 (20.02)	143.34 (21.86)	145.86 (21.24)
	C	145.10 (18.10)	144.80 (19.22)	144.07 (17.90)

TR = thought record intervention; BE = behavioral experiment intervention; C = control group.

3.3. Main analyses

A MANOVA on the dependent measures (shown in Table 2) revealed a significant effect of Time $F(12,73) = 16.87$, $p < .001$, partial $\eta^2 = 0.74$, showing that participants' scores across all measures decreased significantly over time. There was also a significant Time × Group interaction, $F(24,148) = 2.43$, $p < .005$, partial $\eta^2 = 0.28$.

3.3.1. Belief ratings

A univariate repeated measures ANOVA on target belief ratings revealed a significant main effect of Time, $F(2,84) = 76.87$, $p < .001$, partial $\eta^2 = 0.48$ and Group $F(2,84) = 3.18$, $p < .05$, partial $\eta^2 = 0.07$, and a significant Group × Time interaction, $F(4,168) = 10.55$, $p < .001$, partial $\eta^2 = 0.20$. As shown in Fig. 1, immediately following the intervention the BE group endorsed this belief significantly less than did controls $t(59) = -2.81$, $p < .01$. In contrast the TR group's ratings of this belief did not differ significantly from the control group's $t(59) = -2.06$, ns. At one week follow-up both the BE and TR groups had significantly lower belief ratings than the control group ($t(57) = -3.36$, $p < .005$ and $t(57) = -2.92$, $p < .01$ respectively).

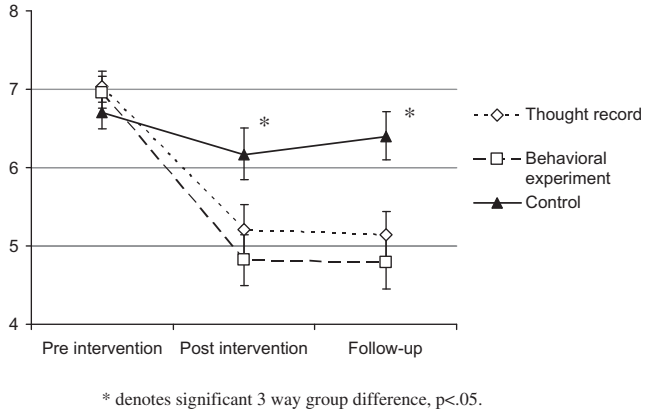

Fig. 1. Mean belief rating for 'not washing your hands after going to the toilet will make me ill' for the three groups (behavioral experiment vs thought record vs control) at pre-intervention, post-intervention and follow-up.

The ANOVA for the related belief "Not washing your hands after going to the toilet will make *others* ill" revealed a significant main effect of Time, $F(2,84) = 40.59$, $p < .001$, partial $\eta^2 = 0.33$ and a significant Group × Time interaction, $F(3.28,137.86)$[5] $= 3.67$, $p < .05$, partial $\eta^2 = 0.08$. As shown in Fig. 2, neither TR nor BE group differed significantly from controls immediately after the intervention, whereas at follow-up, the BE group's ratings were significantly lower than the control group's, $t(57) = 2.53$, $p < .016$,[5] but the TR group was not, $t(57) = -1.95$, $p =$ ns.

3.3.2. Anxiety ratings

The ANOVA for the rating of anxiety about going to the toilet without washing your hands revealed a significant main effect for Time, $F(2,84) = 4.73$, $p < .01$, partial $\eta^2 = 0.05$ and a significant Group × Time interaction, $F(4,168) = 3.12$, $p < .05$, partial $\eta^2 = 0.07$. Although follow-up tests were unable to clarify the meaning of this interaction, inspection of the pattern of means in Fig. 3 suggests that it was carried by a decrease in both intervention groups (particularly the BE group) compared to controls, which showed a slight increase.

3.3.3. Behavior ratings

The ANOVA for ratings of the likelihood of engaging in the target behavior (having to go to the toilet without washing your hands) in the subsequent week revealed a significant main effect of Time, $F(2,84) = 5.65$, $p < .01$, partial $\eta^2 = 0.06$, and a significant Group × Time interaction, $F(4,168) = 2.44$, $p < .05$, partial $\eta^2 = 0.05$.

As shown in Fig. 4, the one way ANOVA comparing groups at post-intervention revealed a non-significant trend toward group differences ($p = .08$). However, subsequent contrasts did not reveal any significant differences. At follow-up the difference between the TR and controls remained non-significant, $t(59) = 1.66$, $p =$ ns, but the difference between the BE and control groups approached significance, $t(59) = 2.29$, $p < .03$.[6]

[5] Threshold for Bonferroni corrected p values.

[6] Note that the one way ANOVA at follow-up failed to reach significance. Threshold for Bonferroni corrected significance was $p < .016$.

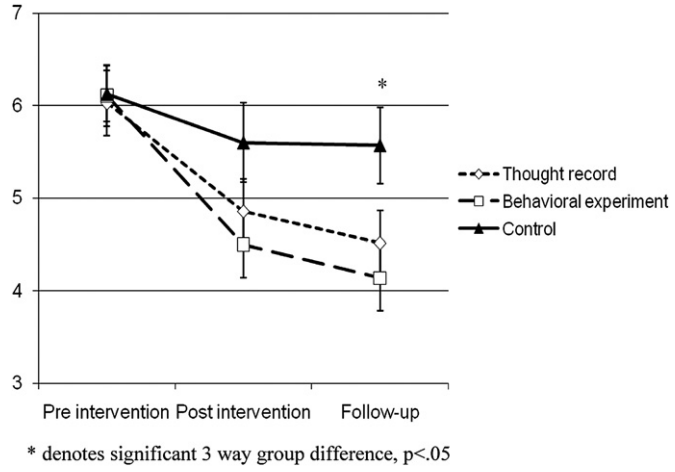

Fig. 2. Mean belief rating of 'not washing your hands after going to the toilet will make other people ill' for the three groups (behavioral experiment vs thought record vs control) at pre-, post-intervention and follow-up.

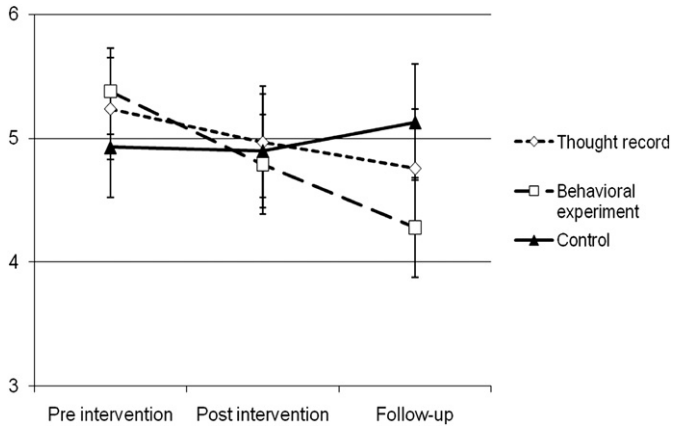

Fig. 3. Mean rating of anxiety in the relevant situation for the three groups (behavioral experiment vs thought record vs control) at pre-intervention, post-intervention and follow-up.

3.3.4. Standardized measures

3.3.4.1. OBQ-44. The results of repeated measures Group × Time ANOVA on OBQ-44 scores showed a significant main effect of Time $F(2,170) = 19.04$, $p < .001$ and a significant Group × Time interaction, $F(3.70,157.36)$[7] $= 2.66$, $p < .05$, partial $\eta^2 = 0.06$. Follow-up tests did not reach significance, however inspection of the pattern of means (Fig. 5) suggests that the interaction reflected a reduction in scores for both intervention groups, relative to controls.

3.3.4.2. IBI. The results of repeated measures ANOVAs of Group (BE, TR, Control) × Time (pre-intervention, post-intervention, follow-up) on IBI scores showed a significant main effect of time but no significant Group × Time interaction, $F < 1.19$, $p > .32$.

4. Discussion

The aim of the present study was to examine the effect of a single session intervention using either thought records (TRs) or behavioral experiment (BE). To overcome the limitations of previous studies, the present study included a larger sample size, a control group to control for the passage of time and provide a baseline to compare the active interventions to, and focused on a belief that is widely held in the general population so as to be able to keep the interventions constant across participants. The primary question of interest was to determine whether the interventions were effective in reducing participants' endorsement of beliefs, anxiety and behaviors related to the target belief that 'not washing your hands after going to the toilet will make you ill'. Results showed that both TR and BE had a beneficial therapeutic impact in comparison to the control condition on beliefs. Separate contrasts with the control condition showed a different pattern of findings according to the type of intervention. The BE intervention was effective in reducing endorsement of the target belief (that not washing your hands after going to the toilet would make you ill) immediately following the intervention and at follow up. In contrast the TR intervention group only differed significantly from controls on the target belief at follow up. BE, but not TR, was effective at reducing the same belief as applied to others (that not washing your hands after going to the toilet would make *others* ill) at follow-up. Again at follow-up, there was a non-significant trend ($p = .03$, Bonferroni adjusted significance level $p = .16$) for the BE to increase the likelihood of participants engaging in the behavior. On anxiety ratings, and the standardized symptom measure OBQ-44, significant interactions suggested that both

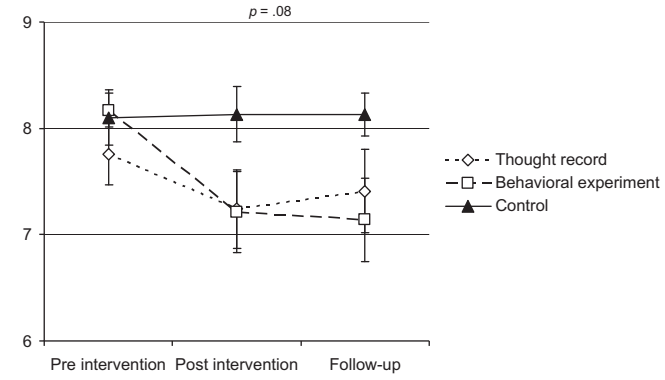

Fig. 4. Mean ratings of likelihood of engaging in the target behavior in the next week for the three groups (behavioral experiment vs thought record vs control) at pre-intervention, post-intervention and follow-up. (Note that ratings are reversed meaning lower ratings indicate higher likelihood of the behavior).

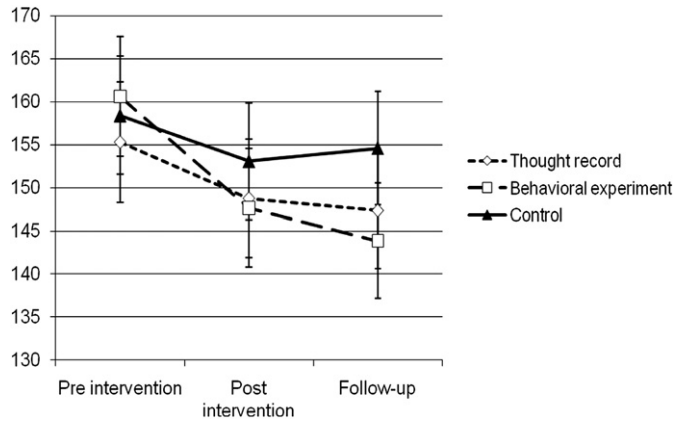

Fig. 5. Mean scores on Obsessional Beliefs Questionnaire for the three groups (behavioral experiment vs thought record vs control) at pre-intervention, post-intervention and follow-up.

[7] For the OBQ-44 the assumption of sphericity had been violated for the main effect of time χ^2 (2) = 7.04, $p = .03$. Therefore degrees of freedom were corrected ($\varepsilon = 0.93$).

interventions were beneficial compared to controls, although the effects were insufficiently strong to allow clarification by follow-up testing.

Our data suggest that while both interventions conferred significant benefit in this non-clinical sample, the BE intervention may have had some advantages over the TR intervention. Improvement in belief ratings about becoming ill appeared to emerge sooner and generalize from the self to others in the BE group compared to controls, but not in the TR group compared to controls. Similarly the stronger effect of the BE intervention resulted in a near significant difference from controls for behavior ratings. The superiority of the BE intervention could not be attributed to differences in the credibility of the interventions or the therapeutic alliance as the TR intervention was perceived to be marginally more credible and therapeutic alliance did not significantly differ between the two active interventions. Overall the study's findings are consistent with recent studies demonstrating the efficacy of BEs over TR interventions (e.g., Bennett-Levy, 2003; Tang & Harvey, 2006) and highlight the potential benefits of using BEs in CBT (McMillan & Lee, 2010).

Others have suggested that BEs may be the most effective CBT technique for changing beliefs (Chadwick et al., 1996; Waller, 2009; Wells, 1997). BEs involve experiential learning and should therefore be more effective in aiding new learning or producing belief change than verbal interventions that focus on rational augmentation of beliefs. Current theoretical opinion is in line with this. For example in Teasdale's (1997) framework there is little value in only exposing the logical flaws in thought processes: to do so is merely to focus at the propositional ('head') level without engaging emotional processes (Longmore & Worrell, 2007). By using verbal techniques of examining the evidence for and against a particular belief TRs would impact only this propositional level. In contrast, by utilizing experiential learning BEs can impact the implicational code ('heart' level), where belief and symptom change need to occur. Similarly Kolb's (1984) theory of experiential learning suggests that maximum learning requires four stages: planning, experiencing, observing and reflecting. Effective BEs are designed to guide the patient through these four stages (Rouf, Fennell, Westbrook, Cooper, & Bennett-Levy, 2004) whereas TRs neglect the experiential component. The current findings are therefore consistent with the suggested advantage in engaging experiential learning to change beliefs, and with information processing theories that propose this.

While the results suggest an advantage of BEs over TRs on the target belief and a similar belief related to others, effects on other measures were weak. There are a number of possible explanations for this. First, we administered a comparatively small dose of the interventions (30 min). This may explain why post-intervention and follow-up means although all in the expected direction (i.e., most symptomatic in the control group, then the TR intervention and the lowest in the BE group) were not strong enough to survive follow-up testing. Second our non-clinical sample unsurprisingly showed a low level of clinical symptoms at baseline meaning that the scope for reduction of scores on standardized measures was small. Third, there was little opportunity for the intervention to impact beyond the session. One potent mechanism for CBT interventions to take effect is via the patient deploying the technique outside the therapy session to alleviate distress across a range of relevant situations in their life. In the current study using a brief intervention and a non-treatment-seeking sample, left little chance that the technique would be applied outside of the session itself. Future studies may want to use higher dose interventions of several sessions, with clinical patient samples, and to better mirror the clinical reality of attempting to use the technique not only to change a specific belief within the session but also to teach the patient the technique for use independently.

4.1. Limitations

The use of a non-clinical sample is a significant limitation of the current study and further comparisons in clinical samples are needed to establish the relative efficacy of BEs and TRs. Similarly, the use of graduate students to carry out the interventions is a limitation and future studies could employ experienced CBT therapists and ensure interventions are monitored for competence as well as integrity. Nevertheless, the use of preliminary analog studies to test phenomena of potential clinical importance is a widely used and important strategy, with some advantages. These include the ability to control experimental conditions more precisely and economical use of clinical samples to examine only promising research directions. Future studies could also employ control conditions that specifically control for the non-specific effects of therapeutic interventions (patient expectations, therapeutic alliance). While these effects may have accounted for the differences from control in the current study, it is important to acknowledge that they could not account for the difference *between* the two active conditions.

A further limitation of the current study is the use of non-standardized measures (belief, anxiety and behavior ratings). However, this parallels CBT practice where the impact of a therapeutic intervention within sessions is typically assessed by belief and emotion ratings. However, it is worth noting that there are ways in which the current study fails to mirror clinical practice. First, the one-session intervention was highly standardized and brief which is not representative of the typical course of 12–20 1 h sessions of CBT. Second, no homework task was set between the intervention and the follow-up, and third, the outcome measure involved ratings of hypothetical behavior and in situation anxiety, rather than actual ratings from, for example, a behavioral assessment task. In addition, while the one week follow-up period confirmed the short-term stability of treatment gains, no conclusions can be drawn about longer-term effects. To better mirror clinical reality future studies should employ higher dose interventions, delivered by trained therapists, with longer follow-up periods, in treatment-seeking clinical populations, with homework tasks, and use of a broader range of assessment including assessing actual behavior and anxiety in the situation.

4.2. Conclusions

To summarize, the present study is preliminary and results should be interpreted cautiously in the light of the limitations of the study's design. However the data provide evidence that while both BEs and TRs are effective in achieving therapeutic change, BEs may be more effective than TRs, in that change occurs earlier and generalizes further. This is consistent with the view that BEs are the most effective technique for achieving belief change in CBT. CBT is known as an evidence-based therapy (Clark, 2004; Salkovskis, 2002) and a multitude of studies have demonstrated its efficacy for a range of psychological disorders. However, the majority of CBT efficacy research is of limited use to the clinician once a patient has been accepted for CBT and there is a paucity of research evaluating what goes on within CBT protocols. Thus the task of choosing a focus or technique within CBT frequently remains a decision that the clinician makes in the absence of empirical data. Results from this study suggest that, when all else is equal, clinicians may be well advised to favor BE over TR interventions. Future work could seek to clarify whether there are particular types of beliefs (e.g., conditional or unconditional) or disorders that respond differentially to these two core CBT techniques or whether particular patient characteristics, such as learning style, determine patients' responses to different CBT techniques.

Acknowledgments

The authors are grateful to Dr Gavin Clark for assisting with the testing of participants and to Kate Muse for assistance with recruitment of participants. None of the authors have any conflicting interests or funding sources affected by the publication of the paper.

References

Andersson, G., & Cuijpers, P. (2008). Pros and cons of online cognitive behavioural therapy. *British Journal of Psychiatry, 193*, 270–271.
Beck, A. T. (1976). *Cognitive therapy and the emotional disorders*. New York: International Universities Press.
Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: The Guilford Press.
Becker, C. B., Zayfert, C., & Anderson, E. (2004). A survey of psychologists' attitudes towards and utilization of exposure therapy for PTSD. *Behavior Research and Therapy, 42*, 277–292.
Bennett-Levy, J. (2003). Mechanisms of change in cognitive therapy: the case of automatic thought records and behavioural experiments. *Behavioural and Cognitive Psychotherapy, 31*, 261–277.
Bennett-Levy, J., Butler, G., Fennel, M., Hackmann, A., Mueller, M., & Westbrook, D. (Eds.). (2004). *Oxford guide to behavioural experiments in cognitive therapy*. Oxford: Oxford University Press.
Bridges, K. R., & Sanderman, R. (2002). The irrational beliefs inventory: cross cultural comparisons between American and Dutch samples. *Journal of Rational-Emotive and Cognitive-Behavior Therapy, 20*, 65–71.
Borkovec, T. D., & Nau, S. D. (1972). Credibility of analogue therapy rationales. *Journal of Behavior Therapy and Experimental Psychiatry, 3*, 257–260.
Chadwick, P. D. J., Birchwood, M. J., & Trower, P. (1996). *Cognitive therapy for delusions, voices and paranoia*. Chichester: Wiley.
Clark, D. M. C. (2004). Developing new treatments: on the interplay between theories, experimental science and clinical innovation. *Behavior Research and Therapy, 42*, 1089–1104.
Devilly, G. J., & Borkovec, T. D. (2000). Psychometric properties of the credibility/expectancy questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry, 31*, 73–86.
Dobson, K. S., & Khatri, N. (2000). Cognitive therapy: looking backward, looking forward. *Journal of Clinical Psychology, 56*, 907–923.
Drankiewicz, D., & Dundes, L. (2003). Hand-washing among female college students. *American Journal of Infection Control, 31*, 67–71.
Elkin, I., Shea, T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., et al. (1989). National institute of mental health treatment of depression collaborative research program: general effectiveness of treatments. *Archives of General Psychiatry, 46*, 971–983.
Gibbs, N. A. (1996). Nonclinical populations in research on obsessive–compulsive disorder: a critical review. *Clinical Psychology Review, 16*, 729–773.
Greenberger, D., & Padesky, C. A. (1995). *Mind over mood: Change how you feel by changing the way you think*. New York: Guilford Press.
Howell, D. C. (2010). *Statistical methods for psychology*. Belmont, California: Duxbury Press.
Jarrett, R. B., & Nelson, R. O. (1987). Mechanisms of change in cognitive therapy of depression. *Behavior Therapy, 18*, 227–241.
Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice-Hall.
Koopmans, P. C., Sanderman, R., Timmerman, I., & Emmelkamp, P. M. G. (1994). The irrational beliefs inventory (IBI): development and psychometric evaluation. *European Journal of Psychological Assessment, 10*, 15–27.
Lester, K., Mathews, A., Davison, P., Burgess, J., & Yiend, J. (2011). Modifying cognitive errors promotes cognitive well being: a new approach to bias modification. *Journal of Behavior Therapy and Experimental Psychiatry, 42*, 298–308.
Longmore, R. J., & Worrell, M. (2007). Do we need to challenge thoughts in cognitive behavior therapy? *Clinical Psychology Review, 27*, 173–187.
Mathews, A., Ridgeway, V., Cook, E., & Yiend, J. (2007). Inducing a benign interpretational bias reduces trait anxiety. *Journal of Behavior Therapy and Experimental Psychiatry, 38*, 225–236.
McGuire-Snieckus, R., McCabe, R., Catty, J., Hansson, L., & Priebe, S. (2007). A new scale to assess the therapeutic relationship in community mental health care: STAR. *Psychological Medicine, 37*, 85–95.
McManus, F., Clark, D. M., Grey, N., Wild, J., Hirsch, C., Fennell, M., et al. (2009). A demonstration of the efficacy of two of the components of cognitive therapy for social phobia. *Journal of Anxiety Disorders, 23*, 496–503.
McMillan, D., & Lee, R. (2010). A systematic review of behavioral experiments vs. exposure alone in the treatment of anxiety disorders: a case of exposure while wearing the emperor's new clothes? *Clinical Psychology Review, 30*, 467–478.
Moe, C. L., Christmas, W. A., Echols, L. J., & Miller, S. E. (2001). Outbreaks of acute gastroenteritis associated with Norwalk-like viruses in campus settings. *Journal of American College Health, 50*, 57–66.
Mortberg, E., Clark, D. M., Sundin, O., & Wistedt, A. (2007). Intensive group cognitive treatment and individual cognitive therapy vs. treatment as usual in social phobia: a randomized controlled trial. *Acta Psychiatrica Scandnavica, 115*, 142–154.
Newport, F., & Strausberg, M. (2001). *Americans' belief in psychic and paranormal phenomena is up over last decade*. Princeton: Gallup News Service.
Obsessive Compulsive Cognitions Working Group (2005). Psychometric validation of the obsessive belief questionnaire and interpretation of intrusions inventory: part 2, factor analyses and testing of a brief version. *Behavior Research and Therapy, 43*, 1527–1542.
O'Donohue, W., & Szymanski, J. (1993). Change mechanisms in cognitive therapy of a simple phobia: logical analysis and empirical hypothesis testing. *Journal of Rational-Emotive and Cognitive-Behavior Therapy, 11*, 207–222.
Pocock, S. J. (1983). *Clinical trials: A practical approach*. Chichester: Wiley.
Roth, A., & Fonagy, P. (2004). *What works for whom? A critical review of psychotherapy research*. New York: Guilford Press.
Rouf, K., Fennell, M. J. V., Westbrook, D., Cooper, M., & Bennett-Levy, J. (2004). Devising effective behavioural experiments. In J. Bennett-Levy, G. Butler, M. Fennell, A. Hackmann, M. Mueller, & D. Westbrook (Eds.), *Oxford guide to behavioural experiments in cognitive therapy* (pp. 21–58). Oxford: Oxford University Press.
Salkovskis, P. M. (2002). Empirically grounded clinical interventions. Cognitive-behaviour therapy progresses through a multi-dimensional approach to clinical science. *Behavioural and Cognitive Psychotherapy, 30*, 3–9.
Salkovskis, P. M., Hackmann, A., Wells, A., Gelder, M. G., & Clark, D. M. (2007). Belief disconfirmation versus habituation approaches to situational exposure in panic disorder with agoraphobia: a pilot study. *Behavior Research and Therapy, 45*, 877–885.
Schulte, D., & Eifert, G. H. (2002). What to do when manuals fail? The dual model of psychotherapy. *Clinical Psychology: Science & Practice, 9*, 312–328.
Shapiro, D., & Startup, M. (1992). Measuring therapist adherence in exploratory psychotherapy. *Psychotherapy Research, 2*, 193–203.
Stangier, U., Heidenreich, T., Peitz, M., Lauterbach, W., & Clark, D. M. (2003). Cognitive therapy for social phobia: individual versus group treatment. *Behavior Research and Therapy, 41*, 991–1007.
Tang, N. K. Y., & Harvey, A. G. (2006). Altering misperception of sleep in insomnia: behavioral experiment versus verbal feedback. *Journal of Consulting and Clinical Psychology, 74*, 767–776.
Teasdale, J. D. (1997). Clinically relevant theory: integrating clinical insight with cognitive science. In P. M. Salkovskis (Ed.), *Frontiers of cognitive therapy: The state of the art and beyond* (pp. 26–47). New York: Guilford Press.
Tolin, D. F., Brady, R. E., & Hannan, S. (2008). Obsessional beliefs and symptoms of obsessive–compulsive disorder in a clinical sample. *Journal of Psychopathology, Behavior and Assessment, 30*, 31–42.
Waller, G. (2009). Evidence-based treatment and therapist drift. *Behavior Research and Therapy, 47*, 119–127.
Warmerdam, L., Smit, F., van Straten, A., Riper, H., & Cuijpers, P. (2010). Cost-utility and cost-effectiveness of internet-based treatment for adults with depressive symptoms: randomized trial. *Journal of Medical Internet Research, 12*(5), e53.
Wells, A. (1997). *Cognitive therapy of anxiety disorders: A practice manual and conceptual guide*. Hoboken, NJ: John Wiley & Sons Inc.
Wells, A., Clark, D. M., Salkovskis, P. M., Ludgate, J., Hackmann, A., & Gelder, M. (1995). Social phobia: the role of in-situation safety behaviors in maintaining anxiety and negative beliefs. *Behavior Therapy, 26*, 153–161.
Westbrook, D., Kennerly, H., & Kirk, J. (2007). *An introduction to cognitive behavior therapy: Skills and applications*. London: Sage Publications.
Westbrook, D., & Kirk, J. (2005). The clinical effectiveness of cognitive behavior therapy: outcome for a large sample of adults treated in routine practice. *Behavior Research and Therapy, 43*, 1243–1261.