

Towards Scalable Training: Narrowing the Research-Practice Gap in the Treatment of Eating Disorders

Suzanne M. Bailey-Straebler

Balliol College

Department of Psychiatry
Warneford Hospital
Oxford
OX3 7JX

Thesis submitted for the degree of Doctor of Philosophy
at the University of Oxford

Trinity Term 2015

Abstract

Towards Scalable Training: Narrowing the Research-Practice Gap in the Treatment of Eating Disorders

Suzanne M. Bailey-Straebler

Balliol College, University of Oxford, DPhil, Trinity Term 2015

Empirically supported treatments (ESTs) now exist for a variety of psychological disorders; however, few individuals have access to these treatments and even fewer receive them in well delivered form. This has been termed the research-practice gap. It is likely that a combination of factors contribute to individuals not receiving good quality ESTs. One major reason is the limited availability of effective training in these treatments. Although many therapists wish to learn such treatments, they seldom have the opportunity as training relies on scarce expert resources and is costly. Furthermore, relatively little is known about the effectiveness of this method or how best to train clinicians: despite having evidence-based treatments, there are no evidence-based trainings.

This dissertation examined one example of an EST - enhanced cognitive behavior therapy for eating disorders (CBT-E) – with the overarching aim of evaluating both existing, and commonly accepted, training methods, as well as, newly developed more scalable ones. How best to train clinicians in CBT for eating disorders has not been investigated previously. The Kirkpatrick training evaluation framework was adopted to guide the studies. Chapter One provided an overview of the research-practice gap with a particular emphasis on the obstacles faced in training therapists. Chapter Two reviewed the literature on training in ESTs and highlighted gaps in the research evidence and areas for improvement in future

studies. An important conclusion was that, although studies varied in design and the precise form and content of the training investigated, results were mostly consistent in indicating that knowledge and skills tended to improve following training. However, the outcome measures used to assess training were often poorly described with unknown psychometric properties. Perhaps most importantly the lack of clearly defined competence cut-points made interpretation difficult. In addition, much of the training investigated had limitations in terms of scalability. Chapters Three, Four and Five, aimed to overcome some of these difficulties and provided a series of studies investigating training in CBT-E. Chapter Three employed qualitative methods to investigate trainees' reaction to conventional workshop and more scalable web-based training and found that although trainees enjoyed training, they had a variety of reasons for not planning to implement the treatment as learned. Chapters Four and Five evaluated the impact of different forms of training on knowledge and skill acquisition respectively. Training in CBT-E was associated with increases in knowledge especially when paired with supervision or scalable guidance, which proved feasible and acceptable to clinician trainees. The results for skill acquisition were less clear, but the new scalable online training was associated with therapists achieving competence. Finally Chapter Six discussed the broader implications of the work and highlighted areas for future research.

Acknowledgements

First and foremost I am immensely grateful to my supervisor, Professor Zafra Cooper. She has guided me through every step of the work presented in this thesis with unparalleled expertise, encouragement and support. The work could not have been completed without her. I am also incredibly indebted to Professor Christopher Fairburn without whom this thesis would have been impossible. I must thank him for taking me on as a research trial therapist, giving me the space to conduct this research and for allowing me to investigate his training.

My sincerest thanks go to my colleagues at the Centre for Research on Eating Disorders/ Dissemination at Oxford for all of their help and support. In particular I would like to thank Marianne O'Connor for spending endless hours sifting through various databases with me and for keeping my spirits up at all times (in and out of work!). Special thanks are due to Beth Edwards and Erin Charlton for skillfully providing web-guidance. I would also like to thank Rebecca Murphy for her unending encouragement and friendship. My deepest thanks are also owed to Kristin Bohn, Mara Catling, Fjola Helgadottir, Kate Muse, Sarah Knowles, Katy Sivyer and Elizabeth Vanner for their extremely helpful ideas and support.

I am enormously thankful to The Wellcome Trust for making my work over the past nine years possible. Special thanks are also due to The Beck Institute for awarding me a student fellowship, which allowed me to learn about their dissemination projects and to undertake my own training in Cognitive Therapy for Depression. I am especially appreciative to Kamryn Eddy, Jenny Thomas and Debbie Franko of the Eating Disorders Clinical and Research Program at Massachusetts General Hospital for awarding me a Summer Research

Fellowship which allowed me to write substantial sections of this thesis under their expert guidance.

I am tremendously grateful to my family who, in spite of the 3,500 miles of distance separating us, have provided a constant source of encouragement and support. I would especially like to thank my mom and dad for always being my biggest supporters along with Kevin, Jeanie, Jeff, Dougie and Margie.

I would also like to thank my 'constant companion' who has kept me company every moment of the past nine months. We cannot wait to meet you. Finally I would like to thank Christian Bailey for entering my life nearly 16 years ago. His presence fills me with joy on a daily basis. I cannot thank him enough for his belief in me.

Table of Contents

CHAPTER ONE 1.0

1.1	INTRODUCTION	1.1
1.2	THERAPIST TRAINING	1.2
1.2.1	Current methods of training and their problems	1.3
1.2.2	The need for scalable training	1.6
1.2.3	Therapists' concerns and attitudes about ESTs and their implementation	1.10
1.2.4	Assessment of training outcome	1.11
1.3	CBT FOR EATING DISORDERS	1.12
1.4	THE KIRKPATRICK MODEL FOR TRAINING EVALUATION	1.16
1.5	CHAPTER SUMMARY	1.18

CHAPTER TWO 2.0

2.1	INTRODUCTION	2.1
2.1.1	Search strategy for review of training studies	2.1
2.1.2	Inclusion and exclusion criteria	2.2
2.1.3	Overview summary of the literature	2.2
2.2	ASSESSMENT OF THE REACTION TO TRAINING: LEVEL ONE	2.5
2.2.1	Studies of trainees' reaction	2.6
2.2.2	Summary of findings	2.10
2.2.3	Discussion	2.11
2.2.4	Conclusions	2.13
2.3	ASSESSMENT OF THE EFFECTS OF TRAINING ON KNOWLEDGE: LEVEL TWO	2.14
2.3.1	Studies of knowledge acquisition	2.14
2.3.2	The effect of training on knowledge acquisition	2.16
2.3.3	Summary of findings	2.18
2.3.4	Discussion	2.18
2.3.5	Conclusions	2.20
2.4	ASSESSMENT OF THE EFFECTS OF TRAINING ON SKILL ACQUISITION: LEVEL THREE	2.21
2.4.1	Part 1: skill acquisition assessed by recordings of performance	2.22
2.4.2	The effect of training on skills	2.25
2.4.3	Summary of findings	2.31
2.4.4	Discussion	2.31
2.4.5	Part 2: assessment of skill acquisition by self-report and chart review	2.35
2.4.6	The effects of training on skills via self-report or chart review	2.37
2.4.7	Summary of findings	2.39
2.4.8	Discussion	2.39
2.4.9	Conclusions	2.41
2.5	ASSESSMENT OF THE EFFECTS OF TRAINING ON PATIENT OUTCOME: LEVEL FOUR	2.41
2.5.1	Description of studies	2.42
2.5.2	Summary of findings	2.45
2.5.3	Discussion	2.45
2.5.4	Conclusions	2.46
2.6	CHAPTER SUMMARY	2.46

CHAPTER THREE 3.0

3.1 OVERVIEW.....	3.1
3.2 THE QUALITATIVE APPROACH.....	3.2
3.2.1 <i>Differences between qualitative research and quantitative research.....</i>	<i>3.2</i>
3.2.2 <i>Strengths and weaknesses of qualitative research.....</i>	<i>3.10</i>
3.2.3 <i>Rigor in qualitative research.....</i>	<i>3.10</i>
3.2.4 <i>Summary.....</i>	<i>3.15</i>
3.3 AIMS OF STUDY 1	3.15
3.4 METHOD	3.15
3.5 RESULTS.....	3.23
3.5.1 <i>Themes.....</i>	<i>3.26</i>
3.6 DISCUSSION.....	3.36
3.6.1 <i>Reactions to two methods of learning CBT-E: similarities and differences.....</i>	<i>3.37</i>
3.6.2 <i>Summary.....</i>	<i>3.43</i>
3.7 CONCLUSIONS	3.44

CHAPTER FOUR 4.0

4.1 OVERVIEW.....	4.1
4.2 INTRODUCTION	4.1
4.3 OVERALL AIM.....	4.2
4.4 STUDY 2A: WORKSHOP TRAINING.....	4.3
4.4.1 <i>Method</i>	<i>4.5</i>
4.4.2 <i>Results.....</i>	<i>4.10</i>
4.4.3 <i>Discussion</i>	<i>4.14</i>
4.5 STUDY 2B: WORKSHOP TRAINING PLUS CLINICAL SUPERVISION	4.18
4.5.1 <i>Method</i>	<i>4.18</i>
4.5.2 <i>Results.....</i>	<i>4.20</i>
4.5.3 <i>Discussion</i>	<i>4.22</i>
4.6 STUDY 2C: WEB-CENTERED TRAINING.....	4.24
4.6.1 <i>Method</i>	<i>4.24</i>
4.6.2 <i>Results.....</i>	<i>4.27</i>
4.6.3 <i>Discussion</i>	<i>4.34</i>
4.7 STUDY 2D: WORKSHOP TRAINING PLUS CLINICAL SUPERVISION AND WEB-CENTERED TRAINING	4.36
4.7.1 <i>Method</i>	<i>4.36</i>
4.7.2 <i>Results.....</i>	<i>4.37</i>
4.7.3 <i>Discussion</i>	<i>4.39</i>
4.8 CONCLUSIONS	4.40

CHAPTER FIVE 5.0

5.1 OVERVIEW.....	5.1
5.2 INTRODUCTION	5.1
5.3 OVERALL AIM.....	5.3
5.4 STUDY 3A: WORKSHOP TRAINING PLUS CLINICAL SUPERVISION	5.3
5.4.1 Method	5.3
5.4.2 Results.....	5.10
5.4.3 Discussion	5.12
5.5 STUDY 3B: WEB-CENTERED TRAINING	5.14
5.5.1 Method	5.14
5.5.2 Results.....	5.16
5.5.3 Discussion	5.18
5.6 STUDY 3C: WORKSHOP TRAINING PLUS CLINICAL SUPERVISION AND WEB-CENTERED TRAINING	5.19
5.6.1 Method	5.20
5.6.2 Results.....	5.21
5.6.3 Discussion	5.23
5.7 CONCLUSIONS	5.24

CHAPTER SIX 6.0

6.1 OVERVIEW.....	6.1
6.1.1 Summary of findings.....	6.1
6.1.2 Broader relevance and future directions	6.5
6.1.3 Future directions.....	6.6

REFERENCES..... 7.0

INDEX of TABLES 8.0

INDEX of FIGURES 9.0

INDEX of APPENDICES 10.0

APPENDICES..... 11.0

1 CHAPTER ONE

Overview of the Research-Practice Gap

1.1 Introduction

Considerable progress has been made in the past 30 years in the development of effective evidence supported psychological treatments for a variety of psychiatric disorders (Shafran et al., 2009). As a result, empirically supported treatments¹ (ESTs) are now recommended as first lines of treatment by the National Institute for Health and Care Excellence (NICE) and the American Psychological Association (APA) for among others, anxiety disorders, depression and eating disorders (e.g., APA, 2005; NICE, 2004). Despite this, only a small minority of patients receive ESTs (Berry & Haddock, 2008; Dobson & Beshai, 2013; Ehlers, Gene-Cos, & Perrin, 2009; Simmons, Milnes, & Anderson, 2008; Wolitzky-Taylor, Zimmermann, Arch, De Guzman, & Lagomasino, 2015). This discrepancy is frequently described as the ‘research-practice gap’ or as the Institute of Medicine puts it more forcefully, “not just a gap, but a chasm”(Institute of Medicine, 2001, 2015). Furthermore it is widely acknowledged that there is a great need to decrease the burden of mental illness (Kazdin & Blase, 2011) and that patients show a strong preference for psychological treatments over other types of treatment (McHugh, Whitton, Peckham, Welge, & Otto, 2013). In response to these problems and in order to improve patient access, government agencies, professional organizations, prominent clinical researchers and insurance companies have all voiced the urgent need for improved dissemination and implementation of ESTs from research settings to

¹ Despite there being a difference between empirically supported and evidenced based treatments these terms are often used interchangeably in the literature. Some (e.g., Lilienfeld et al., 2013; Thyer & Pignotti, 2011) have argued for a clearer distinction to be made between these terms and urge consistency in their use. As much of the literature have used these terms interchangeably such adherence was not possible and both terms will be used.

community settings (e.g., Institute of Medicine (IOM), 2001; National Institute of Mental Health (NIMH), 1999).

In discussing the research-practice gap, many researchers (e.g. Cooper & Bailey-Straebler, 2015; Fairburn & Patel, 2014; McHugh & Barlow, 2010, 2012; Shafran et al., 2009) have suggested that the limited availability of training for therapists² is an obstacle to narrowing this gap. More specifically they regard limited knowledge about how to effectively train therapists and about how to increase the availability of training as the most important factors impeding effective dissemination of ESTs. Indeed, despite the number of ESTs available, relatively little research has been conducted on how best to train therapists in these treatments (McHugh & Barlow, 2010; Rakovshik & McManus, 2010).

This introduction outlines the problems faced by current methods of training in meeting the challenge of becoming more robustly evidence-based and more widely available. These problems will be considered in greater detail in relation to a particular EST, enhanced cognitive behavior therapy (CBT-E) for eating disorders, which will serve as the focus for the exploration and evaluation of training in this thesis. The framework adopted for guiding the evaluation of various forms of training in CBT-E will be described together with a brief outline of the studies evaluating reported in this thesis.

1.2 Therapist Training

Training clinicians to competently carry out evidence-based psychological treatments presents a range of interconnected difficulties discussed in detail below.

In brief, the current ideal model of training is based on the method used to train

² The terms therapist, trainee and clinician are used interchangeably in this thesis to denote someone who provides psychological treatment.

therapists in research trials. In the case of trials, adequate levels of competence are established indirectly by virtue of patient outcome- indeed the trials produce the evidence supporting the treatments. However, the method is time consuming and costly (making it not scalable) and results in too few training opportunities.

Relatively little information exists about how best to train clinicians, particularly in a scalable way or on what the content of the training should be. A further issue that needs to be addressed in the implementation of ESTs is therapists' attitudes towards ESTs and their use: even if therapists receive effective training, we do not know if they intend to rigorously apply this training. A related but separate issue concerns the lack of adequate and consistently used measures of training outcome. If trainings are to be evaluated, reliable and valid measures must be used.

1.2.1 Current methods of training and their problems

"Quality assurance in the preparation of mental health professionals is assumed more often than verified" (Alberts & Edelstein, 1990, p. 497).

Early investigations on the topic of how best to train therapists concluded that the techniques of good therapy training were not clear. Even more concerning, these studies found that the current methods did not ensure either therapist competence or adequate delivery of treatment (Alberts & Edelstein, 1990; Binder, 1993; Ford, 1979). Unfortunately, over twenty years later with numerous ESTs developed, the quality and content of training continues to lack standardization (Beidas & Kendall, 2010; Herschell et al., 2010) as does the measurement of training outcome (e.g., Fairburn & Cooper, 2011; Shafran et al., 2009). Instead of one effective, evidence-based and scalable method we have one main approach: training

provided as part of randomized controlled trials and two diluted versions of this: the so called ‘gold standard’ of training, and training as usual.

1.2.1.1 *Current approaches*

Treatment trial standard

Randomized controlled trials (RCTs) serve as the basis for testing many ESTs. Therapists involved in these trials are usually highly motivated and committed to the notion of ESTs. They receive an initial intensive training (Roth, Pilling, & Turner, 2010), which typically lasts around six months. Training includes the treatment of several pilot patients, and it involves a demanding supervision process, with video or audio recordings of patient sessions being audited in order to achieve high levels of adherence to the treatment protocol (e.g., Hollon et al., 1992; Stangier, Schramm, Heidenreich, Berger, & Clark, 2011). Recordings are also rated by independent assessors to ensure that high levels of skill and adherence are met. This level of supervision is usually maintained until the end of the trial. Although this form of training creates capable and competent therapists as judged by patient outcome, there are limits to its usefulness. The main shortcoming of this model is that it is obviously impossible to provide training and supervision of this intensity on a scale sufficient to meet the need for widespread treatment. Simply put, there are not enough experts to provide this kind of supervision and the cost involved in such training makes it largely prohibitive.

The ‘gold-standard’

Outside of research trials the so called ‘gold standard’ training in ESTs is a combination of attendance at a short (~2 days) didactic workshop, study of a treatment manual, and ongoing clinical supervision (Insel, 2009; Weissman et al.,

2006), usually from an expert in the treatment concerned. However, this ‘gold standard’ method of training has only recently been empirically investigated and results have been mixed, at least in part due to a lack of consistency and clarity in the measurement of training outcome (discussed below) (for detailed reviews see: Beidas & Kendall, 2010; Herschell et al., 2010). Such ambiguous results have led some researchers (McHugh, Murray, & Barlow, 2009) to conclude that there is no current ‘gold standard’ for training in ESTs. Even so, it is worth noting that this method has been successful in organizations which are able to 1) insist that employees participate in such training and 2) have adequate resources to pay for ongoing supervision that includes evaluation of ‘live’ performance in the form of rated³ therapy recording. Examples of such programs include those implemented by the Veterans Health Administration (VHA) and the Increasing Access to Psychological Therapies (IAPT) program (Clark, 2012; Walser, Karlin, Trockel, Mazina, & Barr Taylor, 2013). However, these impressive large scale programs require a level of resources that cannot be provided throughout the mental health sector. ‘Gold standard’ training therefore remains outside the reach of most clinicians.

The usual standard

Only a small minority of therapists receive training as part of an RCT or what has been described as the ‘gold standard’. The reality for many therapists is that they will be given a treatment manual and/ or be offered attendance at a brief training workshop. They will not generally be provided with expert supervision and frequently will receive no training at all in ESTs. This is despite the generally accepted view, as expressed by Kendall and colleagues, that “we cannot expect therapists to

³ Such rating of therapy recordings is not typically considered a part of typical clinical supervision, but is seen more frequently in treatment trial training.

implement empirically supported techniques without appropriate training and supervision” (Kendall, Settipani, & Cummings, 2012, p. 108). This view has been shaped by preliminary investigations assessing the workshop-only method of training (referred to as the “train and hope” method (Stokes & Baer, 1977) due to its lack of follow-up or assessment of skill acquisition). These investigations suggest that treatment manuals and workshops are insufficient for attaining the requisite mastery of knowledge, skills and sustained clinician behavior change (Herschell et al., 2009; Lopez et al., 2011). Indeed, in their extensive review of therapist training, Beidas and Kendall (2010) found that although this method of training improved knowledge scores it “was not enough to produce proficient change in therapist adherence, competence, and skill” (p. 20).

1.2.2 The need for scalable training

As described above while there is a need for better empirical support for existing methods of training the more pressing issue is that few therapists receive training at all. Training opportunities therefore need to be greatly increased. For, while therapists have a desire to undertake more training, they commonly cite both a lack of training availability (Cook, Biyanova, & Coyne, 2009) and the high cost of training workshops (Stewart, Stirman, & Chambless, 2012) as barriers preventing them from doing so. At least three possible routes exist for scaling up training: increasing the availability of the current methods that are effective (described above), training therapists more efficiently by teaching only the active ingredients or common elements of treatment, and/ or developing and testing new methods of potentially more scalable training such as ‘train the trainer’ methods and web-based training.

1.2.2.1 Increase the current training methods

Accepting that the style of intensive training used for RCT therapists is effective, it makes sense that these methods should be offered on a much larger scale. As mentioned above, such a solution is possible for well-resourced programs such as the VHA in the USA and the IAPT program in the UK (see McHugh & Barlow, 2010 for a detailed review of such programs). Clinician and patient outcomes from these programs on adherence and reductions in illness severity are promising for a number of disorders (Clark et al., 2009; Jolley et al., 2015; Karlin et al., 2012; Karlin, Trockel, Taylor, Gimeno, & Manber, 2013; Walser et al., 2013). However, as also mentioned above, both the VA and IAPT systems are not scalable because they rely on large financial resources (provided by US and UK governments respectively) and because of the scarce resource of expert trainers and supervisors.

1.2.2.2 Increase training efficiency

Treatments for psychiatric illness have traditionally been disorder specific. Therapists have therefore had to learn multiple distinct evidence-based treatments in order to treat a variety of illnesses. In recent years some researchers have called for a reconsideration of this. Newer approaches include treatments that focus on common underlying psychopathology such as the transdiagnostic treatment of eating disorders (Fairburn, Cooper, & Shafran, 2003) and unified protocol (UP) for emotional disorders (Ellard, Fairholme, Boisseau, Farchione, & Barlow, 2010). Others (Chorpita, Daleiden, & Weisz, 2005) advocate training in the use of shared evidence-based practice elements (for example, treatment interventions such as relaxation and graded exposure) that could be used for a variety of disorders. The emerging evidence for the effectiveness of these approaches is promising and potentially

reduces training times by not requiring that the clinician learn multiple treatments. For example, Fairburn and colleagues' transdiagnostic approach to treating eating disorders is an effective single treatment for three different diagnostic categories of eating disorder (Fairburn et al., 2015). Equally the unified protocol allows therapists to learn one set of skills to treat a broad variety of 'emotional disorders' such as anxiety and mood disorders (Farchione et al., 2012).

1.2.2.3 Devise new methods of training

Train the trainer

The train the trainer (TTT) model is a training method widely utilized in a number of fields (e.g., Orfaly et al., 2005). Trainers are typically non-experts at a senior level with a good deal of experience (though not always). In the field of mental health, trainers are first trained by experts (workshop and supervision) to proficiency. They then train others in their particular service via a workshop with accompanying clinical supervision and day to day support in implementation. Utilizing TTT methods would alleviate the problem of scarcity involved in relying on training from experts only. However, little empirical research has been done to evaluate the effectiveness of TTT with training in psychological treatments. Preliminary results are promising (Karlin et al., 2010; Martino, Ball, et al., 2011)⁴ but the method requires further investigation. Moreover, this method of training would take far too long to create a meaningful impact on the current widespread need for psychological treatment (Fairburn & Patel, 2014).

⁴ Results for the Karlin et al., study are difficult to classify as the supervision provided was not by the 'trainer' as the TTT model suggests, but by an expert in the therapy (making it not a true investigation of TTT).

Web-based training

An alternative form of training currently receiving increased attention is web-centered training. Online training has already intensified in many areas. For example distance learning (via the internet), is frequently now offered in schools and universities (for a detailed review see: Means, Toyama, Murphy, Bakia, & Jones, 2009). Professional training has also seen an increased reliance on online training, for example in the area of law enforcement personnel training (Schmeeckle, 2003). Online or web-based training in psychological treatment has the potential to alleviate some of the difficulties faced by conventional training such as cost and limited availability (Khanna & Kendall, 2015). Increased access to training would be possible through online methods, as those who wished to receive training could be located anywhere as long as they had internet access. Although such methods might seem less participatory than face-to-face instruction, aspects of e-learning such as the use of active instructional methods (e.g., practice exercise and simulations) and multimedia feedback provided via text, video and audio (Clark & Mayer, 2011; Cucciare, Weingardt, & Villafranca, 2008) may compensate for this loss by adding interactivity. E-learning may also have a number of advantages over live training. Web-centered training could easily be personalized and therefore allow for trainees to complete the training at their own pace. Training could also be pitched at the appropriate level for each therapist and adapted throughout the training based on progress assessments. And the formal assessment of skills and knowledge could be integrated within a web-centered training program to evaluate the effectiveness of the training. Encouragingly, recent investigations have found that web-centered training is well liked by therapists (Becker & Jensen-Doss, 2013).

1.2.3 Therapists' concerns and attitudes about ESTs and their implementation

Even when therapists have received training in ESTs, they do not always implement them in clinical practice or they employ them partially, sometimes omitting aspects of the treatment or adding procedures and strategies from other treatments or both. There is a growing literature on this topic, which investigates therapists' reasons for only partial implementing ESTs (e.g., Gallo & Barlow, 2012; Lilienfeld, Ritschel, Lynn, Cautin, & Latzman, 2013; Nelson, Steele, & Mize, 2006; Rapp et al., 2010) and studies the characteristics of therapists who do not implement ESTs or express objections (e.g., Turner, Tatham, Lant, Mountford, & Waller, 2014). Clearly therapists' reluctance is a potentially important barrier to patients' not only receiving ESTs but receiving ESTs that are well delivered. Generally the solutions suggested to ameliorate this problem concerning implementation are to provide further training and ongoing supervision (e.g., Lilienfeld, Ritschel, Lynn, Cautin, et al., 2013; Shafran et al., 2009; Waller, Stringer, & Meyer, 2012). While crucial questions remain concerning what sorts of further training should be provided and when it is most appropriate to provide it, this is primarily an issue of implementation and not the main focus of the thesis. However, this issue does suggest that evaluations of training must consider how therapists can use and interpret ESTs in a way that is consistent with the evidence, rather than simply focusing on acquiring knowledge of strategies and procedures (Cooper & Bailey-Straebler, 2015). At the very least, trainees' views about how they plan to use the training they receive needs to be part of evaluations aimed at establishing how best to train clinicians in ESTs.

1.2.4 Assessment of training outcome

In order to empirically investigate training and to move towards achieving evidence-based training immediately, researchers must confront the issue of how outcome should be defined and measured. Until recently this topic had been relatively neglected but burgeoning interest in training research has led to renewed attention being paid to this question. Typically assessment of outcomes focuses on *adherence*, defined as how closely the therapist follows a particular protocol (Perepletchikova, Treat, & Kazdin, 2007), and the therapists' ability to demonstrate overall *competence*. Competence, however, is a rather loosely used term with a number of overlapping definitions. One definition provided by Roth and Pilling is "the knowledge and skills required to practise therapy in a competent manner" (Roth & Pilling, 2008, p. 129). A later refinement by McHugh and Barlow identifies competence as the ability to "administer a treatment skillfully and with fidelity" (McHugh & Barlow, 2010, p. 74). Others (Fairburn & Cooper, 2011) offer the term *therapy quality* as an alternative that covers "the extent to which a psychological treatment was delivered well enough for it to achieve its expected effects". Broadly speaking, this means that therapists will be, "doing the right things, well" (p. 374). Regardless of how one defines these terms, adherence and competence have been generally measured through direct assessments of knowledge and/or skill. A number of studies have additionally used a variety of patient outcomes (e.g., medication compliance, treatment attendance, quality of life scores and symptom severity) as indirect methods of assessing training outcome. This is based on the assumption that a clear and probably linear relationship between therapist competence and patient outcome exists. The problem with this notion is that it overlooks the impact that

individual patient circumstances have on outcome. Furthermore, patient outcome does not provide a feasible and practical way of determining whether training has been successful in anything other than large treatment services.

A number of researchers have drawn attention to the lack of adequate and consistently used training outcome measures. They have particularly emphasized the paucity of readily interpretable estimates measuring the percentage of clinicians who become competent as a result of training (e.g., Fairburn & Cooper, 2011; McHugh & Barlow, 2010; Simons, Rozek, & Serrano, 2013). The lack of these measures makes both interpreting the results of previous training evaluations and future investigations difficult.

1.3 CBT for Eating Disorders

CBT for eating disorders provides a good example of many of the issues discussed above. It is considered to be the leading EST for bulimia nervosa (BN) (Agras, Walsh, Fairburn, Wilson, & Kraemer, 2000; American Psychiatric Association, 2006; Fairburn et al., 1991; Hay, 2013; National Institute for Health and Care Excellence, 2004; Shapiro et al., 2007) and binge eating disorder (BED) (Hay, 2013; Kass, Kolko, & Wilfley, 2013). Recent research shows that in its enhanced transdiagnostic form (CBT-E) (see Table 1.1 and: Fairburn et al., 2003) it can be extended to treat the majority of eating disorder cases (BN, eating disorder NOS, BED) (Fairburn et al., 2009, 2015), and that it can be used with somewhat encouraging results for both adolescents and adults with anorexia nervosa (Dalle Grave, Calugi, Doll, & Fairburn, 2013; Fairburn et al., 2013). Its comparative efficacy has been evaluated in adults with bulimia nervosa compared to psychoanalytic psychotherapy (Poulsen et al., 2014), interpersonal psychotherapy (Agras et al.,

2000; Fairburn et al., 2015), and integrative cognitive affective-therapy (Wonderlich et al., 2014).

Studies in community samples with community therapists administering treatment have achieved similar results for treatment completers, but reported more treatment drop outs (Byrne, Fursland, Allen, & Watson, 2011; Knott, Woodward, Hoefkens, & Limbert, 2014; Turner, Marshall, Stopa, & Waller, 2015; Waller et al., 2014). Outcomes from these studies provide evidence that CBT works outside of research settings.

In addition to being the recommended frontline treatment for BN and BED, CBT-E has two other beneficial aspects which might assist its more widespread use. The treatment has been described in detail in a widely available therapist's guide (Fairburn, 2008) and it is transdiagnostic in scope. As a result, therapists need to learn only this one treatment for eating disorders as opposed to several DSM defined disorder specific treatments.

Table 1.1 A brief description of CBT-E

Enhanced cognitive behavioral therapy for eating disorders • offers short term focused psychological treatment (20 sessions over 20 weeks, plus one initial assessment session and one review session) • concentrates on the current maintaining mechanisms of an individual's eating disorder	
The treatment has four stages:	
Stage One	• therapist and the patient work collaboratively to better understand the patient's eating problem. Time is taken to establish a regular, stable pattern of eating and to develop weekly in-session weighing of the patient. Personalized psycho-education is offered. • sessions are held twice-weekly over four weeks • sessions 0-7
Stage Two	• progress is reviewed and plans are discussed together with the patient regarding moving onto stage three • two weekly sessions • sessions 8-9
Stage Three	• content focuses on the remaining processes that are maintaining the eating problem. This stage of the treatment usually involves addressing the patient's concerns about shape and weight; dietary rules; the influence of her/his mood on eating; and teaching the patient to identify her/ his eating disorder mindset (or frame of mind). • eight weekly sessions • sessions 10-17
Stage Four	• maintain changes made in treatment and minimize the risk of relapse • three fortnightly sessions • sessions 19-20 • A review session is typically held some months following the treatment.

1.3.1 The research- practice gap in eating disorders

Despite the existence of an evidence-based treatment for the majority of eating disorders there is evidence to suggest that few patients receive it (reviewed here: Hart, Granillo, Jorm, & Paxton, 2011). Furthermore, even when they do receive CBT there is evidence to suggest that it is not well implemented (e.g., Tobin, Banker, Weisberg, & Bowers, 2007; von Ranson, Wallace, & Stevenson, 2013). This situation appears to be caused by a combination of “questionable” training quality (von

Ranson et al., 2013, p. 336), a lack of effective training (Cook et al., 2009; Harvey & Gumport, 2015; Stewart, Chambless, & Baron, 2012) and therapists' concerns about treatment.

A lack of effective training

As with other disorders, there is no evidence-based training available in CBT for eating disorders. The main reason for this is that no research has been conducted into how best to train therapists or to investigate if current training is effective (Fairburn & Cooper, 2011). Moreover, a great need for newer scalable training has been identified (Fairburn & Patel, 2014). Evaluations of both current and scalable training for the treatment of eating disorders requires development, as does the establishment of measures of therapist competence following training (Cooper & Bailey-Straebler, 2015). Thus, CBT-E is the ideal case for evaluating and improving training in eating disorders.

Therapists' concerns about treatment

To establish an effective method of training in CBT-E, the concerns of therapists must be considered. Few therapists offer CBT as their primary approach when treating patients with eating disorders (e.g., Tobin et al., 2007; von Ranson et al., 2013; Wallace & von Ranson, 2011). Instead, treatments with no evidence base, most frequently an 'eclectic' approach or some combination of, among others, narrative therapy, addiction based models, feminist therapy, hypnotherapy and solution focused therapy are offered (Ressler & Brooks, 2014; von Ranson & Robinson, 2006; von Ranson et al., 2013). Many therapists also provide psychodynamic psychotherapy (Mussell et al., 1999), despite there being only limited evidence to suggest its effectiveness (Zipfel et al., 2014). And some rely on

treatments that have been found to be less effective than CBT, such as psychoanalytic psychotherapy (Poulsen et al., 2014).

Those who do offer CBT are not implementing the treatment as suggested (e.g., Tobin et al., 2007; von Ranson & Robinson, 2006; Wallace & von Ranson, 2011; Waller et al., 2013), which creates the problem that “the label ‘CBT’ is not a reliable indicator of the therapy that is being offered” (Waller et al., 2012, p. 171).

Therapists have many reasons to adhere or not adhere to treatment protocols. These reasons typically tend to be based on: concerns about damaging the therapeutic alliance (e.g., Waller et al., 2013); a belief that treatment trial patients are not as complex as those seen outside of trials (e.g., Simmons et al., 2008; Tobin et al., 2007); and views that treatment manuals are too rigid (e.g., von Ranson et al., 2013). Recent investigations have sought to explore if decisions to use or not use CBT as a complete treatment package are related to individual personality characteristics or demographics of therapists (Turner et al., 2014; Wallace & von Ranson, 2011; Waller et al., 2013). However, therapists’ behavior is not completely understood and requires further investigation.

1.4 The Kirkpatrick Model for Training Evaluation

Training evaluation is an essential ingredient to informing “which...training interventions to stop, modify, continue, or improve” (Kaufman, Keller, & Watkins, 1995, p. 8). While several frameworks for evaluating training have been described, one of the most influential and thorough is the Kirkpatrick model. This model was first described in a series of papers in 1959 and 1960 (Kirkpatrick, 1959a, 1959b, 1960a, 1960b) and was subsequently updated on several occasions, the most recent of which was in 2006. The Kirkpatrick model for evaluating training has long been

considered the benchmark in the fields of business, management, personnel, and human resources development and as such it has been widely used and cited in training evaluation program (Alvarez, Salas, & Garofano, 2004). The model suggests evaluation on four levels:

1. Assessment of the *reaction* to training
2. Assessment of the *effects* of training
3. Assessment of *transfer* of knowledge into skills
4. Assessment of end *results*.

These four stages capture well the various aspects that need to be evaluated when investigating the effects of training clinicians to implement a complex empirically supported psychological treatment such as CBT-E for eating disorders. First, trainees need to receive the training favorably, believe they can implement it and plan to do so. Hence their reaction must be assessed. Second, they need to acquire both theoretical and applied knowledge of the treatment, what they should do and how it should be done. For the researcher, this involves measuring and analyzing the effects of training. Third, this newly gained knowledge needs to be translated into the ability to deliver the treatment with an adequate degree of skill. Measuring therapists' behavior change thus involves assessing their increased capability to deliver the treatment effectively. Last, the trainee needs to show the end results of their training: i.e., that they continue to implement the treatment with fidelity over time and deliver improvements in the standard of patient care and ultimately patient outcomes.

In the present thesis the Kirkpatrick framework is used to evaluate both current and newly developed scalable training methods in CBT-E for eating disorders. As noted, there is a great need for evidence-based treatment to be more widely

available and neither current training nor scalable alternatives training have been previously investigated in the field of eating disorders. This thesis reports on three main studies which were devised to investigate training. Following from this introduction, Chapter Two reviews the current status of evidence-based training in empirically supported psychological treatments. Chapter Three describes Study 1, which assesses therapists' reaction to training using qualitative methodology (Level 1). Chapter Four reports Study 2 using quantitative methods to evaluate the effect of the traditional and new online training on knowledge and applied knowledge (Level 2). In Chapter Five, Study 3 reports the effect of the different forms of training on skill acquisition (Level 3). Chapter Six discusses the findings, their implications for training generally and eating disorders in particular and highlights future research directions. The assessment of results on patient outcomes (Level 4) is beyond the scope of this thesis.

1.5 Chapter Summary

Few individuals with psychiatric illness receive disorder specific recommended treatment. A major reason for this is a lack of trained clinicians. In order to provide better trained therapists and to bridge the research-practice gap or "chasm", training needs to be evaluated, both conventional workshop based and more innovative and scalable forms. Evaluations must include views from therapists about the training and treatment as well as valid assessments of knowledge and skill acquisition as a result of training. Training in CBT-E is ideal for an investigation of this type: it is the leading recommended treatment, yet few individuals receive it and no research on training therapists in CBT-E currently exists. Furthermore, results from investigating training in CBT-E may be applicable to training more generally.

2 CHAPTER TWO

Literature Review of Training Studies



Kirkpatrick's 4 Levels of Evaluation

2.1 Introduction

In Chapter One it was suggested that the lack of readily available training was a major contributor to patients not receiving evidence-based treatment. In this narrative review of the existing studies training outcome is considered in relation to each of the four levels (reaction, knowledge, skills, and outcome) outlined in the Kirkpatrick model.

2.1.1 Search strategy for review of training studies

The aim of this literature review was to evaluate studies that investigated methods of training therapists working in clinical practice settings in evidenced-based psychological treatment. Only those studies that included at least some aspect of training evaluation as defined by Kirkpatrick (Kirkpatrick & Kirkpatrick, 2006) were included. In order to identify appropriate papers numerous databases were searched including: Google Scholar, Medline, PsychInfo, PsycNet, PubMed and Web of Science. Search terms included the following words, plus various combinations of these words: ‘adherence’; ‘assessment’; ‘behavior change’; ‘clinical skill’; ‘competence’; ‘community clinician’; ‘dissemination’; ‘education/ continuing education’; ‘effectiveness studies’; ‘empirically supported’; ‘evaluation’; ‘evidenced-based training’; ‘implementation’; ‘learning’; ‘online training’; ‘psychological treatment training’; ‘psychotherapy training’; ‘satisfaction’; ‘therapist training’; ‘training’; ‘training strategies’; ‘workshop’. The reference lists of all relevant articles were then considered (‘snowball’ method of retrieval) and further studies were retrieved. ‘Related articles’ suggested by the databases were also retrieved and considered. As this field is currently expanding, these searches were frequently re-run to ensure that the latest studies were included. The final date of checking was:

September 10, 2015. Two systematic reviews (Decker, Jameson, & Naugle, 2011; Rakovshik & McManus, 2010) were helpful in identifying recent therapist training studies. In the former, the training studies reviewed focused mainly on community clinicians, while the latter was broader in scope and also considered the training of therapists conducting research trials (excluded from this review).

2.1.2 Inclusion and exclusion criteria

As mentioned, the literature in the general area of training research is expanding rapidly. For the purpose of this review only studies of direct relevance to the topic of the thesis were included, that is training mental health clinicians in specific evidence supported treatments. Therefore studies focusing on isolated interventions (albeit ones that form part of a treatment (e.g., Ruzek et al., 2014)), year-long generic postgraduate training programs (e.g., Barnfield, Mathieson, & Beaumont, 2007), year- long treatment specific training (e.g., Siqueland et al., 2000), programs offering training in generic skills (e.g., Bennett-Levy, Hawkins, Perry, Cromarty, & Mills, 2012; Rakovshik et al., 2013; Westbrook, McManus, Clark, & Bennett-Levy, 2012), training for non-mental health professionals in psychological interventions (i.e. lay people, primary care practitioners) (e.g., Chagnon, Houle, Marcoux, & Renaud, 2007; Maunder, Milne, & Cameron, 2008), training in treatments that currently lack an evidence base (e.g., Kanter, Tsai, Holman, & Koerner, 2013), and evaluation of training for research therapists as part of a research trial (efficacy study) (e.g., Hollon et al., 1992) were not included.

2.1.3 Overview summary of the literature

As discussed, training has been evaluated directly, by examining trainee outcomes and indirectly by evaluating patient outcome. Direct trainee outcomes

tend to focus on what might broadly be termed trainees' competence or ability to provide the treatment to the required standard (Fairburn & Cooper, 2011). In the case of evidence-based psychological treatments this is generally understood to include having theoretical and applied knowledge of the treatment and the skill to apply this knowledge in providing the treatment. As noted, assessing training outcome indirectly relies on the assumption that better trained therapists will produce better treatment outcome, which many would argue is the ultimate goal of any training program. For the purposes of this review and ease of discussion, the studies reviewed have been divided into those that primarily investigate trainee outcomes (therapist reaction, knowledge and skill) summarized in Table 2.1 (uncontrolled studies) and Table 2.2 (RCTs) and those that focus on patient outcome summarized in Table 2.3. Each of these tables is located at the end of this chapter.

(i) Direct training (trainee) outcomes

In total, sixty-five studies evaluating training in ESTs and reporting therapist outcomes were identified (see Table 2.1 and Table 2.2). The majority of these were uncontrolled studies reporting changes as a result of training (N=39) with a growing number, particularly in recent years, being randomized controlled trials (N=26). Most studies focused on substance use disorder. Other disorders including depression, anxiety disorders, and borderline personality disorder were also considered but there were relatively fewer studies of these disorders. No studies examined training in treatments for eating disorders. The workshop trainings (N=54) all contained some element of didactic teaching though several differences were noted regarding level

of interactivity¹, whether time was allocated for manual reading, and whether subsequent clinical supervision was offered. Some studies explored the usefulness of computer based and online training in place of workshops or in comparison to workshops. The vast majority of studies evaluating training examined at least two of the Kirkpatrick levels while a minority included all the levels of evaluation.

(ii) Indirect training (patient) outcomes

Eighteen studies investigated the impact of training on patient outcomes (Level 4 of the Kirkpatrick hierarchy) (see Table 2.3). Most of these studies investigated patient outcome in addition to direct training outcomes (e.g., knowledge, skill) and are also featured on Table 2.1 and/ or Table 2.2), while four studies examined patient outcome on its own. Nearly all studies were uncontrolled (N=16) with only two being RCTs. Again, treatments for substance use disorders were the most well investigated with other studies focusing on treatments for depression, post-traumatic stress disorder, insomnia, serious and persistent mental illness, and autism spectrum disorders. There were no studies of training in eating disorder treatments. All training was provided through in-person workshops and varied in terms of interactivity and the inclusion of clinical supervision.

In the review that follows individual studies are discussed in terms of their contribution to the various levels of evaluation detailed by the Kirkpatrick framework. The discussion will focus on distilling research findings about the outcome of training as regards therapists' reaction and evaluation of training (level one) the knowledge that they gain (level 2) the skills they acquire (level 3) and

¹ Interactive learning is the use of experiential or active learning (doing rather than watching) such as the use of role play and small group exercises. Such interactivity is a component of adult learning principles, which are claimed to be a vastly superior teaching method for adults (Knowles, Holton III, & Swanson, 2005).

patient outcomes reported (level 4). As mentioned earlier some studies have evaluated training on multiple levels and appear in multiple sections.

2.2 Assessment of the reaction to training: Level One

“Failure to attend to the information and concerns of stakeholders clearly is a kind of flaw in thinking or action that too often and too predictably leads to poor performance, outright failure or even disaster” (Bryson, 2004, p. 23).



Therapists are agents of EST implementation serving as the essential bridge between research and practice. Yet their input has only recently been sought in determining the effectiveness of training. One way of including the views of therapists regarding the training they receive is to assess their response to training. The measurement of ‘reaction’ has the potential to provide insight into how useful, engaging and effective therapists find training. By better understanding therapists’ views of training and its usefulness in guiding them about them how to implement treatment some of the problems highlighted in the previous chapter about the quality of ESTs they subsequently offer patients may begin to be addressed. The Kirkpatrick Model specifies five requirements for assessing reaction (Kirkpatrick & Kirkpatrick, 2006): clarity about information sought; use of feedback questionnaires to assess such information; adequate design of feedback questionnaire so that

feedback can be tabulated; anonymous feedback; and finally a method of encouraging and obtaining additional feedback and comments not covered by the form, often best done by interviewing attendees. The studies reviewed here are evaluated in the light of these requirements and provide the background for Study 1 reported in Chapter Three.

2.2.1 Studies of trainees' reaction

Twenty-nine studies (see Table 2.1 and Table 2.2) sought to investigate therapist reaction or feedback to training in ESTs. Before 2005 ($n=3$, 8.6%), relatively few investigations included trainee feedback with most studies ($n=18$, 51.4%) having been completed since 2011. Feedback has been sought after training in a variety of different kinds of treatment and has been measured using surveys and free text entry (described in detail below). While it was almost always assessed immediately after the training finished, two studies also requested satisfaction ratings at a follow-up date (at three and six months respectively) (Baer et al., 2009; Bartholomew, Joe, Rowan-Szal, & Simpson, 2007).

In roughly half of the studies feedback was obtained from uncontrolled studies, with the rest being RCTs. The majority of the uncontrolled studies were of didactic workshops ($N=6$) (Bartholomew et al., 2007; Ekers, Dawson, & Bailey, 2013; Jensen-Doss, Hawley, Lopez, & Osterberg, 2009; Karlin et al., 2012; Petry, Alessi, & Ledgerwood, 2012; Vismara, Young, Stahmer, Griffith, & Rogers, 2009), four studies were of interactive workshops (Russell, Silver, Rogers, & Darnell, 2007; Seng, Prinz, & Sanders, 2006; Vismara, Young, & Rogers, 2013; Walser et al., 2013), three were of online training (Kobak, Craske, Rose, & Wolitsky-Taylor, 2013; National Crime Victims Research and Treatment Center, 2007; Worrall & Fruzzetti, 2009) and one offered

stepped training that started with online training and finished with a workshop (Martino, Canning-Ball, Carroll, & Rounsaville, 2011). The majority of the RCTs (n=6) compared various forms of didactic workshop training to each other (Baer et al., 2009; Chu, Carpenter, Wyszynski, Conklin, & Comer, 2015; DeViva, 2006; Kolko et al., 2012; Luoma & Vilardaga, 2013; Morgenstern, Morgan, McCrady, Keller, & Carroll, 2001), five compared online or computer based training to workshop training (Beidas, Mychailyszyn, et al., 2012; Dimeff et al., 2009; Dimeff, Harned, Woodcock, & Skutch, 2015; Gega, Norman, & Marks, 2007; McDonough & Marks, 2002), three compared different forms of online training (Harned et al., 2014; Harned, Dimeff, Woodcock, & Skutch, 2011; Hubley, Woodcock, Dimeff, & Dimidjian, 2015); one investigated an interactive workshop compared to a purely didactic workshop (Herschell et al., 2009), and one compared workshop training to treatment manual (Martino, Ball, et al., 2011). Roughly one third (n=11) of all of the studies included supervision.

The studies varied greatly in size with a range from 10 trainees (Ekers et al., 2013) to 3,558 trainees studied by the National Crime Victims Research and Treatment Center (NCVC, 2007). Although most studies examined reaction in addition to other training outcomes (i.e., skills), in two studies reaction was the only outcome reported (Bartholomew et al., 2007; Ekers et al., 2013). Reactions to training have been most well investigated in relation to treatments for substance use, with others treatments studied being anxiety disorders, depression, trauma, disruptive disorders, borderline personality disorder and autism spectrum disorder. Reaction to training has not been investigated in the field of eating disorder treatment training.

(i) Feedback surveys

Nearly all studies relied on anonymously filled in feedback surveys to collect information on trainees' reaction to training. For all studies listed in Table 2.1 and Table 2.2, Likert-based rating scales were used. In most cases few details regarding the exact content of the feedback questions were available; however, the areas that were most frequently assessed were satisfaction and relevance/ acceptability. In addition topics including, how the training compared to other training; satisfaction with the trainer; extent of coverage of core material; and a desire to learn more were also measured in some studies (e.g., Baer et al., 2009; Gega et al., 2007; Martino, Canning-Ball, et al., 2011; Martino, Ball, et al., 2011; Petry, Alessi, Ledgerwood, & Sierra, 2010). As can be seen from Table 2.1 and Table 2.2 results from feedback surveys were consistently positive across a wide range of training despite different content, style, and length of training.

A limited number of studies reported data on the psychometric properties of the feedback scales used. Those that reported on reliability (e.g., Bartholomew et al., 2007; Dimeff et al., 2009; Ekers et al., 2013; Harned et al., 2011; Hartzler, Baer, Dunn, Rosengren, & Wells, 2007; Luoma et al., 2007) had Cronbach's alpha scores of at least 0.72, suggesting at least good internal consistency of the measures. Issues of validity of the measures were rarely reported. A notable exception to this was Kobak and colleague's (2013) use of the validated System Usability Scale (SUS) to measure satisfaction with technical aspects of an online training of CBT for anxiety disorders (Bangor, Kortum, & Miller, 2008).

While some studies assessed reactions to just one form of training (e.g., Ekers et al., 2013) a number were RCTs compared different types of training (i.e., online

training vs live workshop training or workshop vs reading a treatment manual) (e.g., Beidas, Edmunds, Marcus, & Kendall, 2012; Dimeff et al., 2009; Gega et al., 2007; Harned et al., 2014). Although the results were always positive for the various methods of training, differences and preferences between different forms of training did emerge. For example, Dimeff and colleagues (2009) found higher levels of satisfaction with learning objectives and presentation style in both online training (OLT) and instructor led training (ILT) as compared to written manual only training (TM). However, in a replication study by Dimeff and colleagues (2015), satisfaction was higher in ILT than either OLT or TM. Similarly inconsistent results have been found in other studies. Baer et al (2009) found higher levels of satisfaction with training that included role-plays and feedback compared to training that was purely didactic, whereas Herschell (2009) discerned no differences in satisfaction when comparing an interactive workshop to a purely didactic workshop.

(ii) Free text feedback

Few studies included the possibility of free text responses or open ended questions, although there were some exceptions to this (e.g., Karlin et al., 2012; Morgenstern et al., 2001), with an increased number of studies doing so in recent years (16.1% of sample, n=5) (e.g., Karlin et al., 2013; Kolko et al., 2012; Vismara et al., 2013). Generally, on average, three such questions were placed at the bottom of a feedback form. All five studies used such questions to inquire which aspects of the training therapists liked the most. In response, therapists mentioned enjoying training activities, requested illustrative examples (Kolko et al., 2012), video examples, group exercises, and the opportunity to ask questions (Vismara et al., 2011). One study however, (Karlin et al., 2012) did not report the therapists'

responses to these open questions. In addition to questions about satisfaction, Morgenstern and colleagues (2001) also investigated therapists' beliefs regarding the treatment described in the training, in this case CBT for substance use disorders. Although feedback, from the Morgenstern study, yielded positive views of the training, a number of clinicians felt that their clients would not benefit from CBT and that an eclectic approach would be better. Additionally, they believed that CBT lacked elements required for recovery and that the manual was too rigid to address patient needs.

(iii) Qualitative interviews

No studies (see Table 2.1 and Table 2.2) included qualitative interviews despite such interviews being highlighted by Kirkpatrick as the best way to fully explore reaction to training. In the broader psychological training literature one study by Krawitz and colleagues (2001) investigated therapist views of training in general skills when working with individuals with personality disorders through qualitative interviews (n=7, 0.77% of the total therapists trained). The detailed findings of this study were published in an unobtainable local source²; however, some themes were discussed in a subsequent publication (Krawitz, 2001). Therapists found that the workshop instilled hope, emphasized the importance of the work, provided a theoretical background for the treatment and led to a shared language being adopted by clinicians.

2.2.2 Summary of findings

The thirty-four training studies reviewed used similar measures to assess trainees reaction to the training they had received and reported obtaining strikingly

² Two email requests were sent to the author which remain unanswered

similar results. Therapist reaction to training was almost exclusively collected through anonymously completed surveys (DeViva, 2006; Ekers et al., 2013; Luoma et al., 2007) using Likert-based scales. The findings suggest overwhelming support from therapists who report that they enjoy training and find it useful. Feedback appears to be favorable regardless of training method (workshop, book or online) but when different forms of training are compared directly some forms of training receive relatively more favorable reviews than others; however, this finding was not consistent (e.g., Beidas, Edmunds, et al., 2012; Dimeff et al., 2009).

2.2.3 Discussion

The growing interest in including therapists' reaction to training is encouraging. Therapists bridge the research-practice gap and failure to seek out their views about how training is received may result in missing an opportunity to increase our understanding about why implementation following training is so inconsistent and, often, poor. Although the use of Likert-based feedback scales provides easily quantifiable and immediate data, the value of the data produced is, nevertheless, not clear. Despite favorable responses from therapists, few go on to deliver the treatments as described whether because they do not learn enough to do so or are not persuaded to do so is not clear. Although some researchers have suggested that therapist satisfaction and interest in training is likely to lead to a potential increase in implementation (Bartholomew et al., 2007), it is not certain that this is always the case. For example, Luoma and colleagues (2007) found that satisfaction, (even high ratings) did not predict self-reported adoption of skills at any point. Likewise, high satisfaction rates do not seem to be associated with high levels of fidelity in implementing treatment (Vismara et al., 2009). In a similarly interesting

finding by Seng and colleagues (Seng et al., 2006), satisfaction rates were high following a workshop amongst all participants regardless of whether they completed the training or not. However, only those who completed the training were likely to go on to deliver the treatment as measured by self-report. As Richards and colleagues (2011, p. 119) have argued:

“...measures...can be difficult to validate. Participants may find a workshop interesting or well-constructed but this may not translate into knowledge change; participants may demonstrate demand effects in questionnaire completion and may not have had their interest stimulated.”

Similarly, others have found that feedback from trainees in a variety of areas is an unreliable predictor of either knowledge acquisition or subsequent use of the treatment as it is heavily influenced by other factors including speaker charisma (Shevlin, Banyard, Davies, & Griffiths, 2000). While speaker charisma may have an indirect impact on learning, there is little evidence to suggest that this assumption is true.

Only a handful of studies included space for more in-depth information to be provided via free text entry (e.g., Kolko et al., 2012) and none included fully open assessments such as qualitative interviews. As currently designed, feedback forms may not be capable of capturing sufficiently nuanced information of the kind that might relate to subsequent knowledge and behavior change. More detailed open – ended feedback, particularly in the form of qualitative interviews would perhaps enable a fuller exploration of therapists’ reactions to training and their implementation plans. Currently a number of unanswered questions remain:

- Why do therapists like training?
- What does a favorable response actually mean?

- If reactions are always overwhelmingly positive why are treatments not later implemented?
- If they do intend to implement the training, will it be as suggested?
- What barriers do therapists perceive to using the training?

Given these unanswered questions, a qualitative inquiry is warranted as such methods are particularly helpful when little is known about a phenomenon of interest (Merriam & Tisdell, 2015). In-depth interviews provide the possibility of conducting initial explorations, moving towards explanations and giving voice to those whose views are rarely heard (Sofaer, 1999), such as trainee therapists.

2.2.4 Conclusions

Given our present state of knowledge about clinicians' reactions to training, the use of qualitative methods to inform our understanding of why therapists like training, what they gain from it, and what their intentions and reservations are concerning future implementation is warranted. Once there is a better understating of their reactions, the creation of informative and valid feedback surveys will perhaps be possible.

2.3 Assessment of the effects of training on knowledge: Level Two

"The assessment of a therapist's knowledge of a treatment...is central to the evaluation of therapist competence" (Fairburn & Cooper, 2011, p. 374).



While a basic pre-requisite of a training program is that it is well received by trainees, training in psychological treatment clearly aims to improve trainees' ability to provide the treatment. As such, competence - often conceptualized as constituting both theoretical knowledge of the treatment and skill in implementing it - has been identified as a training outcome of major importance (Cooper et al., 2015; Fairburn & Cooper, 2011; McHugh & Barlow, 2010; Muse & McManus, 2013). This corresponds to both the learning (knowledge acquired) and transfer (skills) levels in the model. Level two, "learning", in the Kirkpatrick model aims to assess the "resulting change in knowledge" following training (Kirkpatrick & Kirkpatrick, 2006). Kirkpatrick suggests that specific learning objectives or aims of the training be clearly defined in advance and that reliable instruments with clear scoring are used. The following section focuses on studies which assessed knowledge as a training outcome and constitutes the background for Study 2 described in Chapter Four, which investigates trainees' knowledge as an outcome of various methods of CBT-E training.

2.3.1 Studies of knowledge acquisition

Twenty-four training studies investigated changes in therapist knowledge following training in ESTs (see Table 2.1 and Table 2.2). As with the studies of trainees' reaction to training, the number of investigations evaluating knowledge as a training outcome has increased in recent years. Only one study was completed prior to 2000, with nearly half (n=10) of the studies published since 2011. The majority of assessments covered straightforward content (n=11) (e.g., Harned et al., 2011) - 'knowing that'; a small number (n=4) only assessed application of knowledge or 'knowing how' (e.g., Miller & Mount, 2001) and many assessed both (n=9) (e.g.,

Kobak et al., 2013; Sholomskas et al., 2005). Most of the of assessments were designed by study researchers (e.g., Beidas, Mychailyszyn, et al., 2012; Harned et al., 2014) specifically for the study, though occasionally better known, previously published measures were used (e.g., Miller & Mount, 2001). Questions were multiple choice tests, true or false tests and/ or free text (described below).

Most studies used a simple pre-post design assessing knowledge on two occasions only (e.g. Beidas et al., 2012; Gega et al., 2007; Herschell et al., 2009) (n=8 check). Five studies assessed knowledge on more occasions, with three studies having one additional follow up (Baer et al., 2004; Dimeff et al., 2015; Hawkins & Sinha, 1998), one having follow up assessments at three points: six, 12 and 18 months later (Kolko et al., 2012), and another including multiple baseline assessments (Herschell et al., 2009). Approximately half of the studies assessed knowledge change following a single form of training without a comparison group (e.g., Petry et al., 2012); the other half were RCTs comparing different forms of training (e.g., live remote training compared to online didactic (Kobak et al., 2013). The method of content delivery varied with most being purely didactic workshops (n=10), six being completely online (Harned et al., 2014, 2011; Hubley et al., 2015; Kobak et al., 2013; National Crime Victims Research and Treatment Center, 2007; Weingardt, Cucciare, Bellotti, & Lai, 2009), five containing at least one arm of computer based or online training (Dimeff et al., 2009, 2015; Gega et al., 2007; McDonough & Marks, 2002; Sholomskas & Carroll, 2006), four containing some type of interactivity in addition to didactic lecture (Miller & Mount, 2001; Miller, Yahne, Moyers, Martinez, & Pirritano, 2004; R. Richards et al., 2011; Shafer, Rhode, & Chong, 2004), and three directly comparing didactic workshop to interactive

workshop (Beidas, Edmunds, et al., 2012; Herschell et al., 2009; Sholomskas et al., 2005). Despite the general recommendation for training to include supervision, most studies did not include clinical supervision (n=14), while two included supervision in only one or some of the training arms (Miller et al., 2004; Sholomskas et al., 2005).

Sample sizes of the studies varied greatly with the smallest study having as few as 15 therapists (Petry et al., 2012) and the largest as many as 3,358 (National Crime Victims Research and Treatment Center, 2007) trainees. While most studies investigated changes in knowledge as well as changes in at least one other area (i.e., reaction, skills), three studies examined knowledge alone (Baer et al., 2004; Hawkins & Sinha, 1998; Weingardt et al., 2009). Knowledge was most frequently assessed following training in treatments for substance use disorders (n= 8), with treatments for anxiety, depression, borderline personality disorder, trauma and childhood maltreatment also being investigated. Knowledge has not been evaluated in eating disorders treatment training.

2.3.2 The effect of training on knowledge acquisition

(i) Multiple choice tests and true or false tests

Knowledge was most commonly assessed using multiple choice questionnaire tests (MCQs) (79.2%, n=19) (e.g., McDonough & Marks, 2002; Petry et al., 2012; R. Richards et al., 2011). As with most forms of knowledge assessment the multiple choice tests were specifically developed to cover specific content relevant for the particular study (e.g., Kolko et al., 2012).

In a small proportion of studies traditional MCQs were used in conjunction with (12.5%, n=3) true or false (T/F) test items (e.g., Beidas, Barmish, & Kendall, 2009; Hawkins & Sinha, 1998).

(ii) Free text response

While less common than multiple choice tests, some studies also assessed knowledge using a variety of free text entry responses, either in addition to MCQs or instead of MCQs. These included one study that used an open ended question test (4.2%, n=1 (Herschell et al., 2009), one that required a response to a clinical case scenario (4.2%, n=1, Gega et al., 2007), and four which used free responses to written statements (16.7%, n=4 (Baer et al., 2004; Miller & Mount, 2001; Miller et al., 2004; Shafer et al., 2004). At present, open ended questions or free text responses require human marking or rating.

Results

Results from MCQs, T/F and free text assessments, suggest that knowledge improves following training regardless of training type (interactive workshop, didactic workshop, online training) (e.g., Beidas, Edmunds, et al., 2012; Dimeff et al., 2015; Harned et al., 2014; Kolko et al., 2012), apart from treatment manual only where results were inconsistent (Dimeff et al., 2009; e.g., Miller et al., 2004) (see Table 2.1 and Table 2.2). Improvements were not always sustained at follow-up (Baer et al., 2004; Shafer et al., 2004). There is some evidence to suggest that certain forms of training appear to improve knowledge more than others. For example, training with a treatment manual only improved knowledge to a limited extent compared to either treatment manual plus computer based training (Sholomskas & Carroll, 2006) or online training (Dimeff et al., 2009, 2015); and expert clinical supervision provided greater gains in knowledge when compared to training without supervision (Kolko et al., 2012; Miller et al., 2004; Sholomskas et al., 2005). Trainings which were computer led such as CD-ROMs (Gega et al., 2007) and complete online

training packages (Dimeff et al., 2009, 2015), appeared to offer benefit over instructor led workshops; however, this finding was not consistent (e.g., Beidas, Edmunds, et al., 2012).

2.3.3 Summary of findings

Twenty-three training studies investigated changes in therapist knowledge following training. The method of assessment and the results obtained were fairly uniform across the studies. Knowledge was generally assessed pre- and post-training through MCQ and T/F questionnaires, many of which lack psychometric data supporting their use. In addition few provide methods to evaluate the significance of knowledge changes (e.g., competence cut-points, benchmarking). This is a particular problem when using specially designed tests, because almost none of them have been psychometrically evaluated for validity or provide competence cut-points.

In summary all studies found that knowledge scores improved following training no matter what training method used but direct comparisons suggested relatively greater improvements with some forms of training than others. However, the comparative findings are not consistent.

2.3.4 Discussion

Knowledge acquisition is a key component in establishing training effectiveness. While the emergence of training studies investigating changes in knowledge scores is encouraging, a number of limitations are evident. The main one which will be discussed below is the lack of psychometrically sound outcome measures with competence cut-points³. The other limitations discussed briefly here are the lack of controlled studies, the narrowness of current investigations in terms

³ A competence cut-point demarcates those who achieved the minimum necessary knowledge to adequately implement a treatment and those who have not.

of treatments and disorders investigated, and the lack of follow up assessments. Roughly half of the studies do not compare the training to anything else. In a field where so little is known research of this type is warranted. However, without a comparison group the findings are restricted. Additionally the range of the investigations has mainly focused on the area of substance abuse with few studies in other areas. All treatment trainings require evaluation to ensure that therapists are being equipped with the knowledge necessary to competently implement treatment. Finally, only a handful of studies assessed knowledge at a follow-up point (Baer et al., 2004; Dimeff et al., 2009, 2015; Kolko et al., 2012). Without including a follow-up evaluation it is difficult to know if training has lasting results.

Measurement of training outcome

Research investigating provider knowledge of EBPs as a potential barrier to dissemination is significantly limited by a lack of instrumentation.
(Stumpf, Higa-McMillan, & Chorpita, 2009, p. 50)

One of the major problems in evaluating the outcome of training is the lack of reliable and valid measures of training outcome. As a result, the majority of measures are developed by researchers for individual studies. Unfortunately few studies report details of test development or their psychometric properties of these measures. Therefore, despite the consistent finding that knowledge improves following training, it is difficult to interpret the significance of this increase. A few of the MCQ studies included in this review reported 'good' internal consistency scores of the measures with Cronbach's alphas of at least $\alpha=.73$. Yet, the majority did not report on the reliability of the assessment measure (e.g., Dimeff et al., 2009; Petry et al., 2012; R. Richards et al., 2011). Likewise inter-rater reliability was adequate to

excellent (ICC range .51-.99) for the free text responses, when reported. Validity of the measures was rarely reported, with only one study (Hawkins & Sinha, 1998) specifically describing the development and validity of the measure used. While many of the measures have been developed by experts in the field, indicating that they do have high levels of face and content validity (evidence that the test content is relevant and adequately covers the construct being assessed) few studies provide evidence that standardized test development methods (see American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014; Downing & Haladyna, 2006) have been used.

A further and possibly more pressing problem is that even if knowledge scores increase on properly developed measures, little is known about competence cut-points of these measures. While several studies did include cut-points of a type: “clinically significant criterion” (Beidas, Edmunds, et al., 2012, p. 662), “passing” score (Petry et al., 2012, p. 295), or Herschell et al's (2009) ‘mastery’ score. Little detail was provided on how these cut-points were established. Without such benchmarks it is difficult to interpret what the changes in knowledge indicate. Without determining if knowledge scores are related to competence, the meaning of the scores is difficult to interpret and they tell us little about trainees’ ability to use the knowledge of treatment effectively (McHugh & Barlow, 2010).

2.3.5 Conclusions

The investigation of knowledge acquisition after training in ESTs is an expanding area of research. At present the interpretation of results is limited by the lack of well described reliable and valid methods of measurement and the absence

of competence benchmarks. Additionally the range of treatments covered by knowledge acquisition studies is limited.

2.4 Assessment of the effects of training on skill acquisition: Level Three



In addition to ensuring that trainees acquire theoretical and applied knowledge of the treatments they learn, training must also ensure that therapists have the skills to implement the treatments they have learned. Assessing behavior or skill acquisition (SA), involves demonstrating the transfer of the knowledge acquired of the treatment into practice at the required level. Skills may be assessed immediately after training, at which time it would provide an indication of a therapist's ability to provide the treatment (competence). When assessed some months following training, in the workplace, it also provides information about ongoing ability to implement the treatment. As might be expected, Kirkpatrick stresses that the evaluation of SA will only deliver useful results if the research design is robust and reliable and valid measures are employed to assess outcome. It is also worth noting that if therapists react poorly to training and/ or knowledge scores are low one might consider whether measuring behavior is any longer relevant.

This section of the training review focuses on studies that assessed skill acquisition and, in some cases, also implementation. It provides the background for Study 3 (described in Chapter Five) evaluating skill acquisition as a result of two methods of training CBT-E for eating disorders.

SA, following training is commonly assessed by rating therapy recordings of live sessions with actual patients or role-played sessions or part of sessions with standardized patients (SP) (discussed in detail below). Self-report ratings of intervention used by therapists, reports from patients or clients about interventions and chart reviews are also used, especially when measuring implementation over a period of time. Studies that measured behavior change using therapy recordings (actual or simulated) to assess skills immediately following training and/ or later in the workplace are discussed in part one of this review while those that relied on self-report or chart reviews will be described in part two.

2.4.1 Part I: skill acquisition assessed by recordings of performance

Thirty-seven studies (see Table 2.1 and Table 2.2) assessed immediate skill acquisition and/ or longer term implementation after training by means of actual or simulated therapy recordings. As with other aspects of training assessment, interest in skill acquisition has grown in recent years. Only two studies (Carroll, Nich, & Rounsaville, 1998; Crits-Christoph et al., 1998) were conducted pre 2000, with a rapid increase after 2010 (n=14) (e.g., Beidas et al., 2012; Petry et al., 2012; Walser et al., 2013). The majority of studies evaluated SA through the use of therapy recordings with actual patients (n=24), with a growing number using role-play sessions or parts of sessions with standardized patients (n=14). And one using both RPs and SPs (Martino, Ball, et al., 2011). Ratings of the recordings of these sessions,

both in terms of the rating scales employed and the raters using them varied considerably. While some studies used psychometrically sound measures (e.g., Karlin et al., 2012; Kobak et al., 2013) others developed study specific measures (e.g., Dimeff et al., 2015) with limited information to support their use. Raters varied between clinician experts (e.g., Crits-Christoph et al., 1998), students (e.g., DeViva, 2006), study authors (Simons et al., 2010), and professional coding teams such as the one set-up at the University of New Mexico to rate recordings using the Motivational Interviewing Skill Code (e.g., Schoener, Madeja, Henderson, Ondersma, & Janisse, 2006). Many studies did not report any details about the rating scales used, nor did they provide details of the raters or their training.

Half of the studies assessed behavior change following a single form of training without a comparison group; the other half were RCTs comparing different forms of training or training and a control group. Most studies of skill acquisition compared pre- training to post- training (e.g. DeViva, 2006; Herschell et al., 2009; Petry et al., 2012; Sholomskas & Carroll, 2006) with over half (n=15) including at least one follow-up assessment. Six studies examined skill only at post training (Brookman-Frazee, Drahota, & Stadnick, 2012; Crits-Christoph et al., 1998; Martino, Canning-Ball, et al., 2011; Schoener et al., 2006; Vismara et al., 2009; Walser et al., 2013), a number rated all therapy sessions (e.g., Carroll et al., 2006; Martino, Ball, Nich, Frankforter, & Carroll, 2008, 2009) and two assessed skills at three time points: post- workshop training but before the start of clinical supervision and then again post- clinical supervision (Karlin et al., 2012, 2013).

A variety of training was offered. Most were delivered via expert led workshops (n=27), eight contained at least one comparison arm of computerized or

online training (Beidas, Edmunds, et al., 2012; Beidas, Mychailyszyn, et al., 2012; Dimeff et al., 2009, 2015; Larson et al., 2013; Martino, Canning-Ball, et al., 2011; Sholomskas & Carroll, 2006; Sholomskas et al., 2005), six included a treatment manual only comparison arm (Dimeff et al., 2009, 2015; Herschell et al., 2009; Larson et al., 2013; Sholomskas & Carroll, 2006; Sholomskas et al., 2005), three investigated training via a treatment manual plus videos (Martino, Ball, et al., 2011; Miller et al., 2004; Moyers et al., 2008), one included a train the trainer comparison arm (Martino, Ball, et al., 2011), and one consisted of purely online training (Harned et al., 2014). The level of interactivity within the training was also variable. The majority of the training involved interactive workshops (n=20), fifteen were purely didactic, and two studies directly compared interactive workshops with didactic workshops. Clinical supervision was included in the majority of the studies (n=19), with seven RCTs directly investigating the effects of supervision by offering it in only some of the comparison arms.

Study sample sizes ranged from five, in a training study of CBT for depression (Hepner et al., 2011) to 334 for ACT training with counselors working at the Veterans Administration (VA) (Walser et al., 2013). Skill acquisition following training was mainly studied in relation to treatments for substance use disorders (n=22); other treatments studied included those for anxiety disorders, depression, insomnia, borderline personality disorder, autism spectrum disorder and childhood maltreatment. There were no studies of eating disorder treatment training.

2.4.2 The effect of training on skills

(i) Standardized 'Patients'

A number of studies (n=14), particularly those that assessed behavior shortly after training, relied on role-plays with standardized patients (SP)⁴. Widely used in the general medical and nursing training literature (e.g., Barrows, 1993; Cant & Cooper, 2010), SP encounters involve trained 'patients' who portray the disorder of interest.

Studies varied greatly in a number of ways in which encounters with SPs were conducted. The main differences included the kind of actors used, the degree of training they were given, the extent to which their reactions were constrained by a script, the length of the role-play, the content covered, the ratings scales employed and the training and experience of the raters.

The actors and actor training

In many studies the actor was a student or research assistant (e.g., Beidas, Edmunds, et al., 2012; DeViva, 2006; Dimeff et al., 2015), although in others expert clinicians were used (e.g. Miller et al., 2004; Sholomskas et al., 2005). Actor training for these roles was generally fairly extensive. For example, Baer and colleagues (Baer et al., 2009) required that their SPs undergo ten hours of training to accurately represent individuals with substance use disorders. The training consisted of a didactic component about the clinical problem to be portrayed, viewing video simulations, and practice with two of the study authors.

⁴ A distinction lies between the terms simulated patient and standardized patient. 'Simulated' means that the actor is comparable in every way to a real patient, apart from actually having a disorder. Various actors could play the same simulated patient in different ways giving the role flexibility and variability. 'Standardized' instead implies that the actors will play the same patient in very similar ways, with little variation. All standardized patients are simulated, but the reverse is not always the case. Few of the studies included in this review reported whether they were using standardized or simulated patients, or appeared to use the terms interchangeably. As the term standardized patient is more commonly used to assess a trainee, that is the term used here.

In a number of studies, actors were required to follow scripts specifying exactly what they should say and how they should react in order to increase internal validity. DeViva and colleagues (2006) provided actors who were role-playing individuals undergoing a 12-step program for substance use with a detailed script and then monitored the actors throughout the study to check on adherence to the script. In addition they had raters use a checklist to verify and rate actors consistency with the script. Others, however (e.g., Dimeff et al., 2009) preferred guidelines that instructed SPs to interact and follow the therapist flexibly.

The length of the role-plays and the content covered within them was not always provided in the research reports of these studies. Studies that did include such detail reported that role-plays (often 1-3 in number) tended to cover limited components of the treatment, usually about two or three different techniques lasting about 15-20 minutes each (Beidas et al., 2009; Dimeff et al., 2009, 2015). Some of the role-plays were designed to be conducted in person (e.g., Herschell et al., 2009; Miller & Mount, 2001) with others to be completed over the phone (e.g., Dimeff et al., 2015; Harned et al., 2014).

Rating scales for assessing recorded interactions with SPs

Rating scales to assess the content and quality of encounters with SPs were frequently created by the study researchers. Generally they were designed to measure adherence and competence according to criteria identified by treatment experts as essential for providing good quality treatment. For example, Beidas and colleagues (Beidas et al., 2009; Beidas, Mychailyszyn, et al., 2012) devised the treatment specific Adherence and Skill Checklist to measure therapist fidelity to CBT for childhood anxiety. Adherence was assessed by the presence of six CBT

competences (the identification of somatic symptoms, the identification of anxious cognitions, the use of relaxation, the use of coping thoughts, the use of problem solving, and positive reinforcement). To assess skill, raters were required to assess collaborative style, the use of understandable language and therapist confidence. Others studies relied on adapted versions of psychometrically sound measures with predetermined competence cut-points such as the Yale Adherence and Competence Scale (YACS) (Sholomskas & Carroll, 2006; Sholomskas et al., 2005), a 55-item scale used to assess therapists delivering substance use treatment covering six treatment areas (assessment, general support, goals of treatment, clinical management, twelve step facilitation, and CBT).

The raters and rating protocol

The raters in studies using SPs consisted of a diverse group with different backgrounds and with varying levels of rater training. For example, Dimeff and colleagues (2009) created a “coding team” of two DBT experts (defined as ‘gold standard’ coders) and one, less experienced, primary rater. While the primary rater, trained to reliability, rated all of the tapes the two experts rated a random sample to check reliability (10%). DeViva (2006) employed a similar system but used undergraduate and graduate psychology students. Several studies did not describe the raters in any detail other than to say they were university ‘staff’ (e.g., Baer et al., 2009; Miller & Mount, 2001) or ‘experts’ (defined as the study author and two research assistants) involved with the training (e.g., Petry et al., 2012). Others provided no description of the raters at all (e.g., Beidas, Edmunds, et al., 2012).

Reliability and validity of measures

Despite some studies not fully describing the development of their measures, inter-rater reliability was nearly always reported with most being in the good to excellent category (range= ICC-.44-.98) for all ratings. Validity data and competence cut-points were, however, rarely reported, and even when competence cut-points were determined, the percentage of therapists reaching the competence cut-point was not always stated (see Table 2.1 and Table 2.2).

(ii) Real Patients (RP)

The majority of studies (n=24) used ratings from recorded therapy sessions with actual patients. As these sessions are actually part of a treatment they may perhaps be thought to provide a more accurate picture of what therapists are actually doing in their clinical practice. Karlin and colleagues utilized therapy recordings in two different treatment training studies, depression (CBT-D)(2012) and insomnia (CBT-I) (2013), both of which examined therapist training within the Veterans Administration (VA). In the CBT-D study 167 therapists (76% of original sample) submitted at least 10 full session recordings covering various stages of treatment and for the CBT-I study, nearly all therapists (n=94, 92%) submitted all six treatment session recordings.

Rating scales for assessing therapy sessions with RPs

Compared with the rating scales utilized in the studies with SPs, the rating scales used for rated interventions with RPs tended to be more widely used and published. For example, the Cognitive Therapy Scale (CTS) or Cognitive Therapy Rating Scale (CTRS) was used in several studies (e.g., Crits-Christoph et al., 1998; Karlin et al., 2012; Simons et al., 2010). The CTS/CTRS is a generic scale extensively

used to assess therapist adherence and competence in a wide variety of CBT treatment (J. Young & Beck, 1980; J. E. Young & Beck, 1988).

Similar to the scales created for use with SPs, some rating scales for therapy recordings with RPS were created by researchers for evaluating specific training. For example, to assess therapist adherence and competence in implementing motivational interviewing (MI) the Independent Tape Rater Scale (ITRS) was adapted from the YACS by Ball et al., (unpublished manuscript 2002). The items cover MI areas such as fundamental skills advanced skills, MI inconsistent interventions, and general substance abuse counselling skills and the scale has an established cut-point.

The raters and rating protocol

Ratings of complete therapy sessions often require experienced researchers or expert clinicians. For example, to evaluate implementation of CBT for depression, Simons and colleagues (2010) had all session video tapes rated by both a therapist with over 30 years of experience and two advanced graduate students. The two graduate students had each received two years of CBT training and were treating clients with CBT under the supervision. Each tape was rated independently by the three raters after which an item by item assessment was conducted. Any discrepancies were discussed until a group consensus was met. Blindness of raters to treatment arm and time point in treatment was not discussed. Some studies used less experienced raters, although these raters had still received extensive training. For instance, studies which used the Motivational Interviewing Treatment Integrity scale (MITI) (e.g. Baer et al., 2009; Miller et al., 2004) trained student raters for at least 50 hours. To rate study recordings, raters had to be deemed 'proficient' by scoring similarly to 'gold standard' raters and to prevent drift, raters met weekly and

coded a recording together. Additionally, double coding of 10% of the sample occurred every 100 tapes. Raters were blind to assessment time and training arm.

Reliably and validity of measures

Despite some studies not reporting on the reliability of their measures, those that did reported inter-rater reliability that was mostly in the good to excellent category (range= ICC-.51-.99). Validity and competence cut-points were also not always reported (Table 2.1 and Table 2.2).

Results

In general results suggest that competence and skills improve to some extent following training. However, studies do not provide conclusive evidence on which forms of training are best. RCTs that directly compared various types of training delivered conflicting outcomes in a number of areas. For example, in terms of skill acquisition, some studies found computer training to be equally effective to in person training (Dimeff et al., 2009, 2015) while others found computer training to be less effective (Beidas, Edmunds, et al., 2012); and while some found that the use of a treatment manual was able to improve skills equally to attendance at workshop training (Beidas et al., 2009; Dimeff et al., 2009, 2015), others found manual learning to be ineffective at increasing skills (Martino, Ball, et al., 2011; Miller et al., 2004; Sholomskas & Carroll, 2006; Sholomskas et al., 2005). The importance of clinical supervision was also somewhat inconsistent. While there was a consensus among most studies that supervision increased skills more than training without supervision (Beidas, Edmunds, et al., 2012; Harned et al., 2014; Miller et al., 2004; J. L. Smith et al., 2012), one study found no benefit of additional supervision (Moyers et al., 2008). One area that seemed to be more constant was: level of interactivity. All studies

agreed that no differences were discernible between training with differing levels of interactivity (e.g., Beidas, Edmunds, et al., 2012; Herschell et al., 2009).

Studies that investigated the number of therapists who achieved competence produced inconsistent results. Some studies found that, despite scores increasing, the increases were not clinically significant as none or only few therapists reached competence thresholds (e.g., Beidas et al., 2009; Beidas, Edmunds, et al., 2012). Others nevertheless found that over 50% of therapists met competence levels following some forms of training (Karlin et al., 2012; Sholomskas et al., 2005; Walser et al., 2013).

2.4.3 Summary of findings

Thirty-seven training studies investigated changes in therapists' applied skills via therapy recordings or role-plays following training. The assessments were either completed from recordings of role-played encounters with standardized patients or recordings of actual therapy sessions with real patients. Most studies saw some improvement of skills following training; however, in many cases it is unclear if this improvement was 'good enough' to equate to proficiency or competence. Additionally it remains unclear what types of training offer the most benefit in terms of improvements in skills. Information from the available RCTs provided inconsistent results as regards the best training formats (i.e., workshops, online, book based), the optimum level of interactivity and the role of supervision.

2.4.4 Discussion

While it is generally accepted that assessing skill acquisition is essential in evaluating outcome, there is no consensus on the best way to do so. Rating

recordings of clinical interventions with either SPs (role-plays) or RPs is the most common method.

Benefits and limitations of use SPs and RPs

There are a number of benefits to using SPs. By ensuring that trainees interact with a 'patient' who presents identical symptoms, SPs provide standardized, systematic, and quantifiable measure to all therapists being assessed (Adamo, 2003). Such standardized assessment can guarantee that one patient is not more difficult than another, making it easier to compare trainees. This form of assessment also ensures that important concepts and techniques of a treatment are covered as actors can play a variety of patients depicting various points in, and interventions in, treatment, which allows for a comprehensive evaluation. The use of SPs for assessment also has benefit for patients. Issues of confidentiality (having a third party listening to the session) are not relevant when relying on actors as opposed to real patients. Furthermore, the use of SPs protects patients from receiving potentially questionable care from a therapist who has not yet been assessed for competence in a particular method.

Rating therapy sessions with RPs nevertheless offers a number of benefits when compared with role-plays with SPs. These benefits are primarily the result of the authenticity of the interaction. Unlike with SPs where 'sessions' are artificial, recordings with RPs are real therapy sessions conducted as part of a treatment. They thus allow for a more accurate glimpse into what therapists are actually doing in their clinical practice. Additionally therapists taking part in the sessions have a working knowledge of the patient and a history to build from.

Despite these advantages each has its limitations. Using SPs may produce specific difficulties because therapists do not have a relationship or history with the client to draw on. This causes sessions to feel inauthentic, resulting in therapists not behaving as they would with conventional patients. For example, the preparatory clinical vignette may describe a treatment intervention that occurs late in treatment. The therapist is then required to ‘make-up’ what they imagined has happened in previous sessions, without having the actual experience of working with this client to draw upon. Equally, unless assessed in the clinic as a mystery patient⁵ (e.g., Giesen et al., 2007; Luck & Peabody, 2002) role-plays with SPs are only *suggestive* of how a therapist may actually behave in practice. A therapist may be able to perform the task or the intervention for the test, but may choose to implement the technique differently when working with a real client. Concern has also been expressed over rating empathy using SPs (Epstein et al., 2005). An empathic stance is considered an important therapeutic skill; however, it is not clear that role-plays are capable of adequately capturing this multifaceted concept (e.g., Brenner, 2009). For example, Talente and colleagues (Talente & Wilson, 2007) investigated the use of SPs when assessing the interpersonal skills of medical interns. They found that prior knowledge of the encounter being with an actor rather than a real patient had a negative impact on communication. As the interns became familiar with the test approach, they stopped engaging emotionally with the ‘patients’.

Equally, therapy recordings with RPS suffer from a number of limitations.

Primarily, there are problems with submission of the recordings. Apart from a few

⁵ Akin to a mystery shopper or secret shopper, defined as an actor who is employed to test a service in shops and businesses by pretending to be a normal customer, a mystery patient is an actor who tests healthcare services by pretending to be a patient.

examples (Karlin et al., 2012, 2013) where all therapy recordings were submitted, therapists typically choose which recording to submit (e.g., Larson et al., 2013). This may lead to recordings suffering from selection bias and thus not being representative of actual practice. Researchers have also found it difficult to get therapists to submit their tapes. For example, in a study by Shafer and colleagues examining the impact of live teleconference training in motivational interviewing, only 30% of therapists requested to submit tapes did so (Shafer et al., 2004). Additionally, it is likely that when therapists are aware of the session being taped their behavior is different to how it might otherwise be. Another limitation regards the severity of each patient's illness which can vary greatly. Some individuals will be more difficult to treat than others (i.e., those with comorbid disorders); as a result the therapist's performance may be influenced by the patient rather than the therapist's skill. This makes systematic comparisons between therapists difficult as the content of the tapes (i.e., procedures and treatment techniques used) may no longer be similar. The rating scales are further limited by raters often needing to be very experienced clinicians, which causes issues of scalability. Finally, patients may be reluctant to give consent for their sessions to be recorded and assessed.

Differences in the methods used (such as using SPs or RPs, and different rating scales and raters) make comparing the results of various types of training with respect to skill acquisition difficult. Nevertheless, the general findings from studies using both SPs and RPs to assess skill acquisition are consistent, with most studies seeing an improvement in skills following training. As with the findings with regard to knowledge acquisition, there are, however, inconsistencies in the findings when methods of training are compared. No clear consensus emerges about which

method of training is best in terms of imparting skills to trainees. Results on the benefit of additional clinical supervision were also inconsistent. While most studies found it to improve skills further, this was not always the case in studies, especially in studies that used real patient assessments. These inconsistencies may be a reflection of the methodological variations described.

The scales used to rate role-played interactions with SPs often lack validity data and competence cut-points are rarely available whereas those with used to rate RPs have tended to be more psychometrically sound and to have cut-points. Even these methods have been subject to criticism.

It is worth noting here that both methods of assessing skill acquisition are severely limited in terms of scalability. While they may be useful in research studies to establish what methods of training work best, they could not easily be used as methods of establishing that trainees had met the requisite standards in large scale training initiatives and their expense and reliance on expertise might well limit further evaluation of new methods of training.

2.4.5 Part 2: assessment of skill acquisition by self-report, patient report and chart review

Twenty-four studies (see Table 2.1 and Table 2.2) used therapist self-report, patient and/ or caregiver report or chart review to assess implementation following training in ESTs. This was usually done in addition to measuring another training variable (i.e., knowledge), but a large proportion (n=8) examined self-report and/ or chart review only (e.g., Fadden & Vale, 1997; Sanders, Prinz, & Shapiro, 2009).

As with the previous sections, research in this area has increased dramatically over the past 10 years. Two studies (Carroll et al., 1998; Fadden & Vale, 1997) were

conducted prior to 2000 with the rest from 2005 or later. Therapist self-report was the most common (n=18) type of assessment, although the exact content of the self-report questionnaires or interviews and method of collection varied between studies. Some studies collected the data through telephone interviews (e.g., Sanders et al., 2009; Seng et al., 2006), while many mailed out paper surveys (Harned et al., 2014; Luoma & Vilardaga, 2013); some questions focused solely on adherence (e.g., Henggeler et al., 2008) while others included questions on implementation (e.g., Beidas, Mychailyszyn, et al., 2012). Two studies examined adherence as reported by either care-givers (Schoenwald, Carter, Chapman, & Sheidow, 2008) or by both care-givers and patients (Henggeler, Chapman, et al., 2008). Two studies relied on chart review (Jensen-Doss, Cusack, & de Arellano, 2008; Liddle et al., 2006) (described below).

Most studies (n=14) assessed reported SA following a single form of training without a comparison group (i.e., workshop only); the rest were RCTs comparing different forms of training. The delivery of the training was fairly split with approximately half being expert led purely didactic trainings and the other half including some interactivity. A small number of studies (n=5) included a fully online version of the training (e.g., Dimeff et al., 2015; Harned et al., 2014) and one investigated the effectiveness of training delivered via CD-ROM (Sholomskas & Carroll, 2006). Clinical supervision was included in roughly half of the studies (n=14), with two RCTs investigating the effects of additional supervision specifically (Henggeler, Sheidow, Cunningham, Donohue, & Ford, 2008; Luoma et al., 2007). The number of therapists trained in each study varied greatly with a range from 11 (Carroll et al., 1998) to 611 (Sanders et al., 2009). Changes in behavior following

training have been most well investigated in relation to treatments for substance use, while treatments for anxiety disorders, depression, trauma, schizophrenia, insomnia, borderline personality disorder, disruptive disorder, trauma, and childhood maltreatment were also studied.

2.4.6 The effects of training on skills via self-report or chart review

(i) Self-report

Twenty-two studies investigated the effect of training on skills implementation through the use of self-report following training. Self-report allows for quick administration to a large number of therapists and, when done with check marks or other easily ratable scoring, self-report is also extremely scalable (e.g., Henggeler, Chapman, et al., 2008; Sanders et al., 2009; Schoenwald et al., 2008; Seng et al., 2006). Self-report measures mainly assessed therapists' views of implementation and sometimes included ratings of therapists' adherence to treatment guidelines following training as well as patients' or caregivers' views. Implementation was typically assessed via mailed surveys (e.g., Jensen-Doss et al., 2009) or questionnaires (e.g. Lopez & Basco, 2011; Schoenwald et al., 2008). Some studies also examined adherence ratings instead of, or in addition to, implementation rates (e.g., Karlin et al., 2013).

(ii) Chart review

Two studies, Jensen-Doss et al. (2008) and Liddle et al. (2006) used the process of chart review to assess implementation. Chart review involves collecting all of a patient's therapy notes and systematically assessing the treatment. As a result, chart reviews offer detailed insight regarding therapy session content that otherwise is hard to capture and would be missed. Jensen-Doss and colleagues (2008) coded all

clinic visits for three months prior to the workshop and 3 months following a training workshop. Liddle and colleagues' (2006) coded themes from all clinic progress notes pre-and post-training.

Results

Overall therapists report increase use of treatment interventions following training. However, some studies found that this increase was only seen in 40% of therapists (Sanders et al., 2009; Walser et al., 2013). Results from RCTs are mixed, with studies finding improved rates of use and adherence following some forms of training but not others (i.e., active training versus control training or training with and without supervision) (Harned et al., 2014; Luoma et al., 2007) and other studies finding no difference between training formats (Dimeff et al., 2009, 2015; Sholomskas & Carroll, 2006). One study generated inconsistent results between caregiver and youth reports of the same therapists following training (youth reported significant changes in therapist technique following interactive training as compared with purely didactic while care-givers reported no differences) (Henggeler, Sheidow, et al., 2008). With only two studies examining SA via chart review, any generalizations are impossible to make. Both studies examined the use of therapy interventions following training; one finding no change (Jensen-Doss et al., 2008), and the other finding significant changes (Liddle et al., 2006).

Reliability and validity of measures

Fewer than half of the studies described the development or psychometric properties of their measures. Those that did tended to report Cronbach's alpha scores ranging from .82 - .99, suggesting good to excellent internal consistency. None reported on measurement validity. In fact, developers of one of the measures

noted that the measurement required future examination of the “correspondence between therapist reports and...observable therapist behavior in sessions” (Weersing, Weisz, & Donenberg, 2002, p. 178). For the chart-review studies, inter-rater reliability (Cohen’s kappa ranged from 0.67-1.0) was only reported by Jensen-Doss and colleagues (2008).

2.4.7 Summary of findings

Twenty-three training studies investigated changes in therapists’ applied skills and implementation via self-report and/ or chart review following training. Nearly all assessments were therapist self-report with two being report by patient and/ or caregiver and two being chart reviews. The majority of studies found an increased use of interventions following training, but it is worth noting that this was not always the case (e.g., Henggeler, Sheidow, et al., 2008; Jensen-Doss et al., 2009; Walser et al., 2013).

2.4.8 Discussion

As has been demonstrated, skill acquisition is a key component in establishing training effectiveness. Assessment of SA through self-report and/ or chart review provides a way to evaluate therapist implementation. Each method has its own benefits and drawbacks. A major limitation for self-reporting is that retention can often be low. Richards and colleagues (2011), for example, sent out self-report questionnaires to 73 therapists attending training in ACT and, although all of the therapists had filled in both pre- and post- training questionnaires, only 24 (33%) completed implementation sheets one year later. Similarly, Lopez and colleagues (2011) were only able to retrieve follow-up surveys from 33.9% (n=20) of therapists, despite phone call reminders. As described above, it is also unclear how accurate

these self-report measures are. For instance, patient youths and caregivers (Henggeler, Sheidow, et al., 2008) report opposing findings when asked the same questions about therapy content. Self-reporting may likewise suffer from response bias because therapists may feel under pressure to report inaccurately high levels of intervention use and may not be the best judge of their abilities. The evidence on self-efficacy ratings is contradictory with some finding that therapists *overestimate* their ability (Brosan, Reynolds, & Moore, 2008) and others finding that therapists *underestimate* their ability (McManus, Rakovshik, Kennerley, Fennell, & Westbrook, 2012) as compared to observable outcome measures. Finally it is unknown how closely correlated reported intervention use is with actual proficient implementation as despite having good face validity and often good to excellent internal consistency the validity of these measures is not known. Likewise, assessment by chart review faces several limitations. Chart reviews are extremely time-consuming and rely on therapists not only writing detailed progress notes but also writing them in such a way that techniques will be accurately coded by raters. Chart reviews that instead rely on counting session numbers or extra sessions (i.e., family therapy sessions in addition to individual sessions) (Liddle et al., 2006) offer a faster assessment. This is especially the case if electronic medical charts are used as they can easily be totaled by computer. The latter form, however, do not offer the detail of the former and, by merely counting session numbers, do little to indicate skill at implementation.

While in general many studies saw increased use of interventions following training, the results are inconsistent. This may be due to some of the problems listed above such as poor retention rates and difficulty of the method. These inconsistencies are further complicated by the unknown relationship between self-

report, patient report, care giver report and actual implementation. Chart reviews may offer a better indication of this, but as yet only a very small number of studies exist.

2.4.9 Conclusions

It is essential to determine if training is able to improve skills and to ensure transfer of those skills to clinical practice. At present a variety of methods aim to assess this (role-plays, therapy recordings, self-reports, chart reviews). Each has its strengths and limitations. The creation of valid measures with pre-determined cut-points for competence is essential for meaningful interpretation of results and will greatly enhance future evaluations of SA. The greatest challenge to the assessment of skill is the unscalable nature of the assessments. Without creating more scalable methods, their use is greatly limited. Finally, there is also a need to conduct such evaluations for other treatment trainings; currently none exist in the field of eating disorders training.

2.5 Assessment of the effects of training on patient outcome: Level Four



Ensuring good or improved patient outcome is the ultimate goal of EST dissemination. Quite naturally it has been assumed that successful training will lead to better outcomes. But, as described earlier, measuring patient outcomes is an

indirect means of assessing training outcome as it is influenced by a number of factors unrelated to level of training of the clinician offering treatment. The fact that the relationship between therapist expertise and patient outcome may not be straightforward does not undermine such outcome being a part of the overall assessment of training. If improvements in patient outcome are observed following training, then it may be reasonable to infer that knowledge and skills transfer has taken place. However, even if knowledge and skills transfer has taken place, therapists may fail to implement the knowledge and skills they have acquired for a variety of attitudinal, institutional and organizational reasons such as their beliefs about the suitability of the treatment for the patients they see, the policies of the setting in which they work and time constraints (as briefly described in Chapter One).

Patient outcome equates to the final level of Kirkpatrick's training evaluation - the assessment of tangible outcomes or "results", described also as the full impact of the training back in the trainees' workplace. Such impact is considered difficult and time-consuming to measure as by its very nature it has to overcome the difficulties of carrying out careful research in real world clinical settings.

A brief overview of studies that have examined patient outcomes will be provided in order to complete the review of current knowledge concerning training in ESTs. Due to time frame and resources needed to assess patient outcome this level of training evaluation is outside the scope of this thesis.

2.5.1 Description of studies

Eighteen studies investigated change back in the trainees' place of work following training in ESTs (see Table 2.3). As with previous sections, studies prior to

2000 are rare (McDonel et al., 1997; Strosahl, Hayes, Bergan, & Romano, 1998), with the majority (n=11) of reports published after 2005. 'Results', in the form of patient outcome were assessed in a variety of ways typically including symptoms (i.e., scores on a depression scale), functional outcomes (i.e., quality of life scores), and/ or treatment variables (i.e., attendance at therapy or days in treatment). Assessing symptom changes was the most common, with all studies investigating these. The number of patients treated and therapists participating in the studies varied greatly with some including hundreds of patients (e.g. Schoenwald, Sheidow, & Chapman, 2009 n=1979; Walser et al., 2013, n=745), and large numbers of therapists (n=429 and n=391, respectively) while others had small patient samples (Cook, Walser, Kane, Ruzek, & Woody, 2006, n=18) and few therapists (n=4). Nearly all studies were uncontrolled with only two being RCTs (Carroll et al., 2006; Miller et al., 2004). Training was either delivered via expert led workshops or was provided online. The delivery of the training varied: most were interactive workshops (n=14), three were purely didactic and one study offered a didactic training workshop first followed by interactive training (Strosahl et al., 1998). Clinical supervision was included in the majority of the studies (n=14), with one RCT investigating the effects of supervision by offering it in two of the comparison arms (Miller et al., 2004). Again patient outcome following therapist training has been most investigated in relation to treatments for substance use disorders, with other disorders studies being depression, post-traumatic stress disorder, insomnia, serious and persistent mental illness and autism spectrum disorders.

(i) Symptom related measures

Changes in symptom severity was the most commonly assessed indicator of patient improvement and was measured in all studies (n=18). Severity was often measured directly by assessing change in symptom related behavior. For instance, treatments for substance use disorders assessed abstinence from drug and alcohol use via urine toxicology screen and breath alcohol assays (e.g., Liddle et al., 2006; Martino et al., 2008) and treatments for autism measured changes in associated problem behaviors (Social Skills Improvement System (SSIS)) (Brookman-Frazee et al., 2012). Symptoms were also measured using well established, psychometrically tested measures, such as the Beck Depression Inventory II (BDI-II) (Walser et al., 2013).

(ii) Measures of functional outcome

Patients' general functioning was also evaluated in a number of studies (n=10) usually through various psychometrically sound quality of life measures such as the World Health Organization Quality of Life Brief (WHOQOL-BREF) (e.g., Karlin et al., 2013; Walser et al., 2013) and Global Assessment of Functioning scale (GAF) (McDonel et al., 1997).

(iii) Measures of treatment variables

Treatment variables were also considered as a measure of training impact (n=10). Examples of these included: days in treatment/ treatment attendance as assessed by self-report and caregiver report (Carroll et al., 2006; Martino et al., 2008) or therapist (Brookman-Frazee et al., 2012); self-reported medication compliance checklists (McDonel et al., 1997); and therapeutic alliance (Karlin et al., 2012; Walser et al., 2013).

Results

Results were fairly consistent across studies with many finding a reduction in symptoms (e.g., Martino et al., 2008; Russell et al., 2007; Simons et al., 2010; Walser et al., 2013), improved global functioning (e.g., Walser et al., 2013) and treatment variables (e.g., McDonel et al., 1997) from pre- to post-treatment following training. However, this was not always the case (e.g., Lopez et al., 2011; Miller & Mount, 2001; Miller et al., 2004). In fact, Lopez and colleagues (2011) found that patients treated prior to the training had better outcomes than those treated following the training.

2.5.2 Summary of findings

This review evaluated eighteen training studies which investigated changes in patient outcomes following therapist training. Common to all studies was the measurement of symptom related outcomes. Most studies found a reduction in patient symptoms following training, but this was not consistent. Some studies also investigated functional outcomes such as the patient's quality of life and adherence to treatment variables. All studies saw improvements on these domains following training.

2.5.3 Discussion

Patient outcome is an important aspect of assessing the success of wide scale dissemination and implementation programs. The ultimate aim of such programs is, after all, to provide good quality treatment so that better outcomes can be achieved for patients. However, as a proxy for training outcome its use is more limited. Patient outcome is an indirect method of assessing the immediate effects of training and is influenced by a number of factors other than the expertise of the clinician providing

treatment (Fairburn & Cooper, 2011). Furthermore the relationship between competence and patient outcome has been difficult to establish. A recent review by Branson and colleagues (2015) investigating the relationship between competence and patient outcome in IAPT found that in general no relationship appears to exist between competence and patient outcomes. Nevertheless, patients receiving treatment for anxiety from the 10% most competent therapists improved more than those receiving treatment from other therapists. Additionally, patients treated by the weakest 10% of therapists experienced a deterioration in their symptoms. This suggests that the relationship may be complex and that it requires further investigation.

2.5.4 Conclusions

The ultimate aim of training clinicians in ESTs is to ensure that patients receive good quality treatment and outcomes of the kind backed by the research evidence. Thus while patient outcome may be useful in evaluating the value of dissemination and implementation efforts and may play a role in establishing the predictive validity of newly developed competence measures, relying on it as an outcome of success in training is problematic for the reasons discussed above. It is also worth noting that as a way of assessing the competence of people who have been trained it is also not practical.

2.6 Chapter summary

This review of the training outcome literature suggests that therapists are interested in learning new treatments and react favorably to training in general. Current methods of training (both traditional workshop based and more innovative online training) improve knowledge to some extent, but without the use of valid

assessments measures with clearly defined competence cut-points, these results are difficult to interpret. However, training has been less consistent at improving skill, subsequent implementation of treatment or indeed patient outcomes.

Although the recent interest in training evaluation is encouraging, this review highlighted several potential limitations of the research to date. Few studies described their evaluation methods in detail. For example, no studies included specific information regarding the assessment of therapists' reaction to training (level one). Many studies neglected to provide details of the procedures used when devising knowledge tests (level 2). Only a limited number of studies included information about how therapy encounters (with real or standardized patients) were coded. Specifically, little information was provided about the raters' level of expertise or the training they were offered (level 3).

Besides the lack of detail about assessment methods, few studies used reliable and valid assessment measures. Often study specific scales needed to be created with the result that little is known about how these measures perform and even less is known about what level of score would indicate competence. Even when known measures were used, some of these measure generic skills rather than domain specific competence⁶ (Barber, Sharpless, Klostermann, & McCarthy, 2007) and, in some cases, their reliability has not been adequately demonstrated. Thus current methods of training require more robust support. Perhaps even more importantly, most of the current methods even if effective are not suitable for widespread use. There is thus a need to develop and test new methods of more scalable training to meet the widespread need for psychological treatment; and a

⁶ The therapist's ability to implement specific types of treatment (i.e., CBT-E).

need to understand how such training compares with existing methods of training.

Finally training in the area of eating disorders has not been studied.

The studies evaluating training in CBT-E for eating disorders reported in the following chapters aimed to systematically investigate various methods of training using the Kirkpatrick framework while learning from, and avoiding, the methodological limitations highlighted in this literature review. Although the studies focused on CBT-E many of the findings are likely to be more generally applicable.

3 CHAPTER THREE

Study 1: Therapists' reaction to training: A qualitative study



Kirkpatrick's 4 Levels of Evaluation

3.1 Overview

The study presented in this chapter focuses on understanding trainees' evaluation of training in CBT-E, provided in either the conventional workshop format or in a new online format. This information is critical as therapists are a key part of decreasing the research-practice gap; and no such study has been previously undertaken. As discussed in Chapter Two, current investigations of therapist reaction to training have been informed by post-training surveys which are of limited value and have left a number of unanswered questions (see Chapter Two, section 2.2). To address these limitations a qualitative method of inquiry was therefore ideal.

3.2 The qualitative approach

Qualitative research is a method of inquiry that focuses on the in-depth exploration of human experience. The approach has long been employed in the social sciences to understand experiences through exploring and interpreting the meanings people ascribe to them (Denzin & Lincoln, 1998; Merriam & Tisdell, 2015). It has the ability to generate hypotheses and as a result is often considered a prerequisite for quantitative methods in areas where little is known. In the mid-1990s, the importance of qualitative studies in medical research was highlighted when the British Medical Journal devoted an entire journal series the topic. By running such a series the editors hoped that: "...by making clear what these methods entail, how they are used, and how they can be evaluated, they...will be viewed as valuable tools in the methodological tool box of health and health services research" (Pope & Mays, 1995, p. 42). More recently, the value of qualitative methods in psychological research and its particular relevance for dissemination research was

clearly emphasized when it was suggested as a “...way to bridge the divide between research and practice” (Kazdin, 2008, p. 154).

3.2.1 Differences between qualitative research and quantitative research

The dominant approach in scientific research is a quantitative one. Generally a priori theories guide such research, which involves hypothesis testing, operational definitions, carefully controlled investigations, and statistical analysis (Kazdin, 2003). By contrast, qualitative research, is hypothesis generating, allowing data to inform and generate theories, and involves looking for emerging themes. While both quantitative and qualitative methods seek to understand phenomena, their goals and methods of sampling, data collection and, data analysis differ (see Table 3.1).

(i) Goals

Qualitative research endeavors to describe, interpret and gain further understanding of phenomena in their day to day settings. It is ideal for the exploration of practices as they occur in dynamic social settings rather than in controlled environments. Quantitative methods aimed at testing hypotheses in controlled conditions may be of limited value when researchers have not yet established which variables are the most relevant for further study. Therefore, the knowledge gathered through qualitative research can provide the basis for formulating hypotheses which may subsequently be investigated with quantitative methods. In terms of its goals qualitative research seems ideally suited to understanding therapists’ response to training

(ii) Sampling

The differences in the sampling techniques of qualitative and quantitative methods reflect the divergent goals of the two methods. Qualitative methods focus

on sampling subjects that are part of a particular group or have a shared experience of interest to the researcher rather than obtaining representative samples capable of yielding results that may be generalized. Deliberate seeking of subjects who fit the criteria of interest is often termed *purposive* sampling (Giacomini & Cook, 2000). In purposive or purposeful sampling “...individuals, groups, etc., are selected according to their (expected) level of new insights for the developing theory...” (Flick, 2009, p. 65) and provide an in-depth understanding of the group’s shared experience (Barbour, 2001; Kuzel, 1992).

Within the broad notion of purposive sampling, 24 different sampling schemes have been identified (Onwuegbuzie & Leech, 2007). These schemes include examples of both random (probability) sampling (i.e., stratified random sampling; cluster random sampling) and non-random sampling (i.e., critical case sampling; snowball sampling and politically important sampling). Non-random purposive sampling is particularly useful in collecting “information-rich” cases to “...reflect the average person, situation or instance of the phenomena of interest” (Merriam & Tisdell, 2015, p. 78). Therefore, non-random purposive sampling was chosen for the current study to ensure that interviews were conducted with trainees from a range of professional backgrounds.

Consonant with the aims of qualitative research sample size is determined by reaching theoretical saturation - this is regarded as the point when the addition of further data does not produce new information (Bowen, 2008; Merriam & Tisdell, 2015, p. 183). It has been suggested that in the case of interviews, as used in the present study, approximately 12 is generally sufficient to gather and detect themes with a relatively homogenous groups (Guest, Bunce, & Johnson, 2006). This is in

marked contrast to the use of power calculations to determine, in advance, the sample size necessary to detect a previously specified difference between two groups at a statistically significant level (Howell, 2013).

(iii) Design and Data collection

Qualitative research employs firsthand accounts of phenomena of interest that are interpreted by the researcher rather than data (expressed in numeric form) gathered through more objective measurement instruments where the aim is to avoid researcher bias as far as possible. In qualitative research the researcher is the instrument (Golafshani, 2003; Kazdin, 2003). The value and credibility of the work relies on the researcher taking careful steps to ensure reliability and validity (discussed below) rather than reliance on valid and reliable measuring tools.

Both verbal and visual methods may be used to collect qualitative data (Flick, 2009). Verbal data is generally obtained from interviews, either individual or group or both, sometimes called narratives (i.e., “tell me your story...”) or focus groups respectively. Interviews can be open (using topic areas without specific questions), semi-structured, problem centered, or ethnographic (where the researcher lives with the subjects of interest collecting stories through conversation and observation (Reeves, Kuper, & Hodges, 2008)). However, in each case they have two steps; first understanding the subject’s overall view and then probing deeper. In this study open interviews which started with completely open-ended questions and only followed by more structured questions after interviewees had expressed their views were used.

Visual data methods include the use of observation, ethnography, photographs, and film. Through the researcher actively participating in the

interaction, observational and ethnographic methods aim to gain an insider's knowledge of an event. The use of photos and film, or image-based data is classified as secondary observation (Mason, 2005) as the researcher is not actually present.

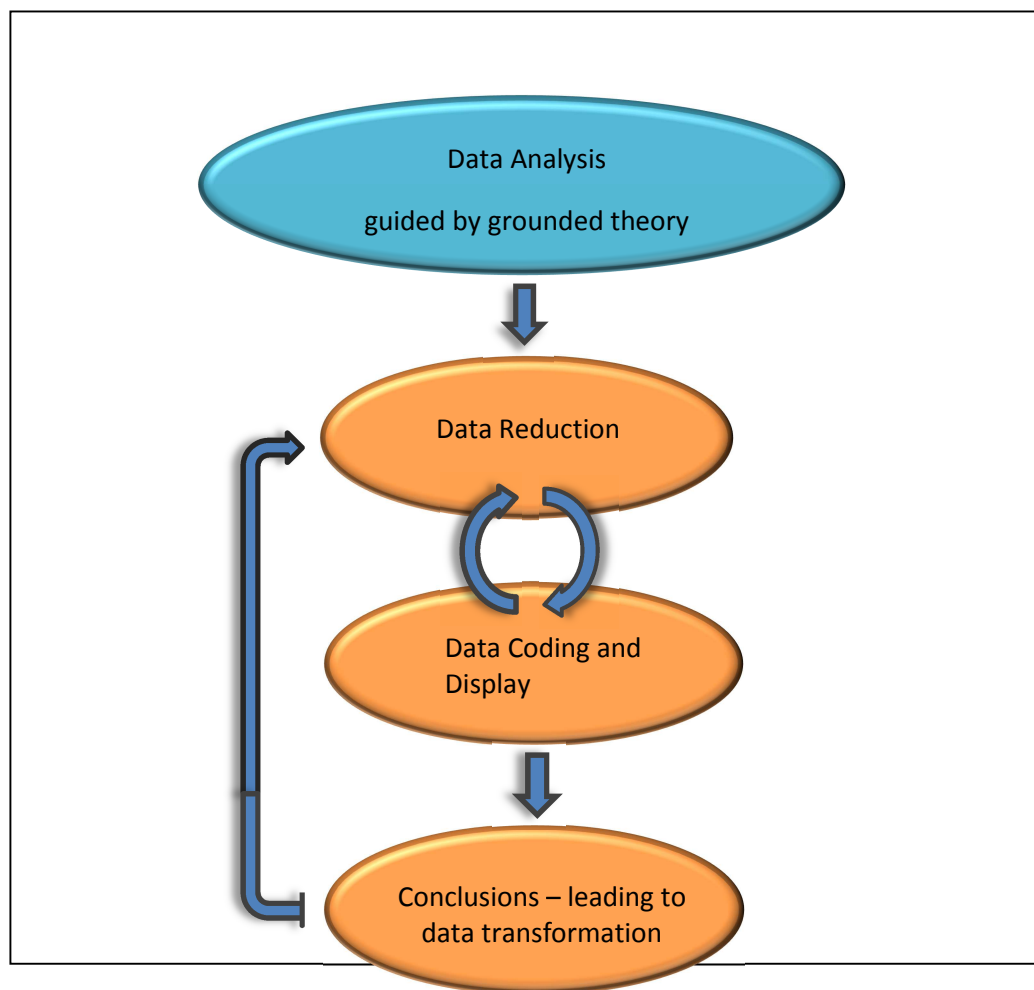
(iv) Data analysis

Since qualitative analysis uses stories, narratives and full text, "...essentially the words are the data" (Kazdin, 2003, p. 328). As a result qualitative data analysis has been described as the "most complex and mysterious of all of the phases of a qualitative project" (Thorne, 2000, p. 68). Despite this, most qualitative research can be described as involving four steps: data collection (as already discussed); data reduction; data coding and display and the forming of conclusions (Miles, Huberman, & Saldana, 2013) (see Figure 3.1). These steps are part of an iterative process that may be repeated several times throughout the data collection and analysis phases.

The conceptual framework or theoretical perspective adopted by the researcher largely determines decisions about data analysis. One of the most widely adopted and well-developed frameworks is that of grounded theory. Derived from the philosophy of symbolic interactionism (Chenitz & Swanson, 1986), grounded theory provides guidelines for collecting and analyzing data in order to build theoretical frameworks and theories (Charmaz, 2008; Merriam & Tisdell, 2015). It is guided by two principles: the theory should emerge from the data rather than being forced on the data (Benoliel, 1996); and the analysis should follow an iterative process (Miles et al., 2013). While both qualitative and quantitative researchers have biases and prejudices, these potentially undermine qualitative research as it relies so heavily on the interpretations of the researchers. Grounded theory attempts to ameliorate this problem of researcher bias by insisting that all interpretations or

meanings be 'grounded' in the data. Grounding is achieved by closely following the four-step process for data collection and analysis as outlined above. This process allows new hypotheses to be explored and areas of relevance to be highlighted with new ideas being re-tested against the data and not the researcher's preconceived ideas.

Figure 3.1 Steps of data analysis



Data reduction

Qualitative research tends to produce large quantities of information. The first step in analysis involves narrowing or down reducing the material. For interview data, this is achieved though first transcribing the interviews, then coding, looking for themes, and repeating the process multiple times.

Data coding

Coding begins once interviews are transcribed; most researchers prefer to code as they transcribe to ensure that the interview is able to capture what they are interested in. Often this preliminary coding is able to enhance interview data through new topics being added or increasing researcher vigilance on a particular topic. Coding represents "...the operations by which data are broken down, conceptualized and put back together in new ways...It is the central process by which theories are built from data"(Corbin & Strauss, 2015, p. 157). It involves pulling out passages of text from the transcriptions deemed to be relevant. As a method, coding is approached in distinct ways influenced by underlying differences in methodology (Gläser & Laudel, 2013). One can either decide to code based on pre-existing theories, such as examining the text for areas of interest, or one can inspect the text more generally looking for a wider group of potential points of relevance.

There are two main methods of coding (each relating to a different qualitative methodology): theoretical, and thematic. Theoretical coding, which develops a grounded theory, typically consists of three steps: 1) open coding (having no preconceived idea of what patterns may emerge from the data); 2) axial coding (a refining of the categories resulting from open coding); and 3) selective coding (using all codes created from steps one and two and re-reading/ re-coding again) (Gläser & Laudel, 2013; Ruppel & Mey, 2015). These categories are somewhat fluid and a researcher will move back and forth between the different types, often with a greater emphasis on open coding at the start and more selective coding towards the end of the coding process (Flick, 2009). Thematic coding, part of thematic analysis, is often found in comparative studies where groups and themes are identified in

advance by the research question (Fereday & Muir-Cochrane, 2006). Thematic analysis, unlike grounded theory analysis, does not set out to develop a theory but instead looks for pre-determined patterns and concepts within the data. Theoretical coding is particularly relevant when investigating a new area where it would not be possible to already have potential themes defined in advance.

Data display

Data display or 'cleaning' is started after the data are initially reduced (as described above). It consists of further organizing and compacting the data with the main goal being to bring meaning to the data collected. Often this is done through the use of figures and diagrams which are able to succinctly show patterns and emerging themes between codes (Verdinelli & Scagnoli, 2013). It is important to note that data reduction is cyclical in nature often requiring multiple rounds of data reduction followed by data display followed by data reduction.

Data conclusions

Finally, after cycles of data reduction and display, conclusions can be drawn from the data. Examining emerging themes and patterns allows researchers to understand data in-depth. Such understanding is then used to view the data from a new perspective and may serve to reduce it further. It is important that data are examined several times in this way in order for analysis to be sufficiently rigorous. This process ultimately leads to *data transformation*, when information is condensed, clustered, sorted and linked (Gherardi & Turner, 2002).

(v) Analysis with computer software

The development and use of software packages uniquely tailored to qualitative data has revolutionized the field by improving the rigor of analysis (Kelle,

1995; Pope, Ziebland, & Mays, 2000). These packages store, organize and allow efficient retrieval of data. Interviews can be transcribed directly into the program thus streamlining the transcription and coding procedure. The programs are further able to perform sophisticated analyses using various algorithms, display the data in interesting ways such as word frequency maps and uncover patterns and themes that may otherwise have gone unnoticed. In the present study one of the most widely used package, NVivo (T. Richards, 2002) (originally NUD*IST (Non-numerical Unstructured Data Indexing, Searching, and Theorizing) (T. Richards & Richards, 1994) was used.

Table 3.1 Differences between qualitative and quantitative research

	Qualitative Research	Quantitative Research
Goals	describe, interpret and gain further understanding of particular phenomena and generate hypotheses	- test theories and hypotheses - quantify group differences and establish group means - allow generalizations and quantitative descriptions
Sampling	purposive (choosing subjects who fit the criteria of interest)	statistical sampling (e.g., random assignment, group equivalence and matching) emphasis on statistically representative samples
Data collection	- open ended interviews or direct observation to describe thoughts, attitudes, views and feelings - researcher as the instrument - words, actions, attitudes	- systematic controlled investigations to obtain objective representative data - objective measurement tools numerically expressed data
Data analysis	- inductive- building of ideas from the data-hypothesis generating - categorization of data into pattern (coding themes, words are the data) - interpretation by researcher	- deductive-hypothesis testing - operationalized by scores on standardized measures more objective based on statistical significance
Results	- understand participants experience - new hypotheses	- establish generalizations /causality - enable predictions - support/fail to support hypotheses

3.2.2 Strengths and weaknesses of qualitative research

Qualitative research is particularly well suited to research where the goal is to gain an in-depth understanding of the participant's experience, especially when this is poorly understood (Kazdin, 2003). As such it is ideal for providing information about clinicians' experience of training and their intentions about subsequent implementation.

The major weakness of qualitative research is related to the very nature of this type of research. As discussed the method relies heavily on researcher interpretation which may be thought to be unduly contaminated by investigator bias and therefore not replicable.

Qualitative methods of inquiry are also limited by the considerable amount of time this type of research requires. Creating, administering, transcribing, coding and analyzing interview data is extremely time consuming and potentially produces far fewer results than quantitative methods.

3.2.3 Rigor in qualitative research

A number of steps may be taken to ensure rigor and minimize researcher bias in qualitative research (e.g., Kazdin, 2003; Krefting, 1991; Morse, Barrett, Mayan, Olson, & Spiers, 2002). The discussion about how to do so is ongoing. Basically two approaches have been taken: some have suggested that qualitative research needs to conform more to quantitative standards; while others have suggested that different standards more appropriate to qualitative methods should be defined.

In their landmark work, Lincoln and Guba (1985) suggested that qualitative research should not be forced to meet the standards for quantitative research. Instead they argue that qualitative research should meet the criterion of

“trustworthiness” consisting of: credibility, transferability, dependability, and confirmability (Lincoln & Guba, 1985). Credibility, akin to internal validity, is seen as being reached when “an accurate description or interpretation of human experience that people who share the same experience would immediately recognize” (Krefting, 1991, p. 216). It is achieved by the researcher examining transcripts not only on an individual level but also comparing them to all other study participants in order to establish shared themes (Thomas & Magilvy, 2011). Transferability, which is roughly parallel to external validity, is frequently achieved by ensuring that all groups of interest are included (based on demographic data) in a sample. Dependability, the analogue of reliability is achieved by audit trails involving detailed description of the steps involved in the research. Finally it is suggested that replicability follows from achieving credibility, transferability and dependability.

However, there is a growing body of researchers (e.g., Morse et al., 2002) who argue the creation of a parallel system for rigor perpetuates the view that qualitative research cannot be considered as rigorous as quantitative research. They strongly advocate a shared terminology, but argue that reliability and validity must be achieved through methods unique to qualitative research. These methods put the emphasis on the researcher consistently scrutinizing their methods and themselves rather than waiting for a reviewer to evaluate the data they have generated. The debate is complex and further details are beyond the scope of this thesis. Following Morse et al’s (2002) recommendation, the present study employed the accepted and well known terms of reliability and validity - terms that remain widely used by qualitative researchers in spite of the ongoing debates.

(i) Reliability

Described by Cohen and Crabtree (2008), the strategies most commonly used to ensure reliability in qualitative research involve the use of peer review, external audits, and what is termed procedural reliability (Flick, 2009).

As might be expected peer review involves checking ideas, interpretation and understanding of particular phenomena with colleagues not connected with the research (Patton, 1999).

Similar to the peer review, an external audit involves an independent analysis to ensure that findings are supported by the data. This process is sometimes simply termed 'multiple coding' (Barbour, 2001). As with peer review it is thought that greater objectivity is gained by the reviewer or auditor being independent of the researcher or the research team. In the case of interview studies like the present one, this external auditing is frequently accomplished by having an independent person code the transcribed interviews in addition to the researcher.

Procedural reliability refers to the providing of a detailed description of methods used. Flick (2009) argues that in addition to being able to ascertain "...where the subject's statement ends and the researcher's interpretation begins..." (pp.176) high levels of documentation can vastly improve the reliability of qualitative research. Researchers are encouraged to carefully organize and store all aspects of the study including: details of the study design, methods of data collection, method of data analysis, investigator meeting (especially those where important decisions were made) notes or minutes, research tapes, transcripts of interviews, reflexive journals, software analysis, preliminary coding logs, coder classification schemes and results. It is suggested that by making the research procedure transparent it is easier

to assess what was done and therefore more likely that another investigator would be able to repeat an investigation and achieve similar findings.

(ii) Validity

Validity has received more attention than reliability within qualitative research (e.g. (Flick, 2009; Onwuegbuzie & Leech, 2007). Showing that a researcher's interpretation accurately reflects the actual events, views, etc., that are being studied is an onerous task. The techniques of triangulation, theoretical sampling, comparative analysis, member checking, and adhering to the principles of grounded theory are typically relied on to increase the validity of the researcher's interpretation of the data.

Triangulation is the use of multiple sources of data, procedures and perspectives to support emerging conclusions (Cohen & Crabtree, 2008; Kazdin, 2003). Four types of triangulation have been described. *Data triangulation* is the use of multiple sources of data. *Investigator triangulation* relies on the use of multiple investigators, with the purpose of reducing individual investigator bias. *Theory triangulation* is employed where various theories and hypotheses are considered and compared side by side to explain a phenomenon. Finally *methodological triangulation* relies on using quantitative methods and qualitative methods (e.g., combining a questionnaire with a semi-structured interview) (Flick, 2009). Investigator triangulation was planned for this study by having both multiple investigators and multiple coders.

The constant comparative method, a type of theoretical coding (detailed above) first described by Glaser (1969) and later refined by Strauss and Corbin (1994), is another way to increase validity in qualitative research. The method

consists of four steps: 1) comparing incidents applicable to each category, 2) integrating categories and their properties, 3) delimiting the theory, and 4) writing the theory (Glaser, 1969, p. 220) and is achieved when a researcher compares their codes over and over, continually revising them. This allows for the checking of previous codes, the elimination of unnecessary codes, and the development of further codes. Of particular interest are any statements or codes that appear to be outliers or extreme cases. Looking for contrasting cases (Flick, 2009; Gerhardt, 1986), ensures that the investigator is always challenging their own dominant interpretations and considering other ideas. Given that theoretical coding also forms the basis for grounded theory, it was the primary coding form used in the present study.

Additionally, member-checking is the process whereby all aspects of the research, in particular the findings, are examined by research subjects who participated in the study. There are objections to this technique including individuals' ability to recall what they said and, more importantly, individuals being unable to recognize their own individual contribution once material has been organized into new higher-order themes or insights (Cohen & Crabtree, 2008). Due to these objections it was decided to not seek verification of the results from participants.

Finally, it is essential that the investigator adheres to the principles of whichever analysis procedure they are using (in this case grounded theory). Since grounded theory requires researchers to begin with a research topic of interest rather than a hypothesis (McGhee, Marland, & Atkinson, 2007) it is essential that the researcher examines all relevant information gathered rather than squeezing

concepts into rigid preconceived criteria. In addition to following the four-step process described above, adherence to the method is further ensured through transparency at all stages of the research. Many researchers achieve this by including exact quotes from participants as done in this study and/ or complete interview transcripts. This allows readers to assess the validity of the researcher's conclusions.

3.2.4 Summary

Qualitative research has different goals and methods compared to quantitative research which lend themselves to studying topics where little is known such as therapists' reaction to training. Despite its strengths, its main weakness is that it relies on researcher interpretation. There are a number of ways to overcome this weakness and to improve reliability and validity. The present study investigates therapist reaction to training using interviews in a purposively sampled group of clinician trainees.

3.3 Aims of Study 1

- To explore therapists' views about learning CBT-E
- To compare their views of two different forms of training: attendance at a conventional workshop or participation in a new form of web-centered training

3.4 Method

(i) Overall Design

Exploratory qualitative interviews designed to elicit trainees' reactions and attitudes to training were conducted with two groups of therapists receiving training in CBT-E. The first group of trainees (Group One) attended conventional workshop training and the second (Group Two) received a new form of online training, termed web-centered training. All interviews were carried out by the researcher and coding

was completed by the researcher and an independent coder. Data analysis was informed by grounded theory and used Computer Assisted Qualitative Data Analysis Software (CAQDAS), specifically NVivo.

Ethical approval was obtained from the University of Oxford's Central University Research Ethics Committee (REF:MSD/IDREC/C1/2012/10).

(ii) The two forms of training

Workshop training

Workshop training represented the typical or conventional training that many therapists receive in evidence-based psychological treatment. As described in Chapter One, section 1.3.1, this style of training, when combined with clinical supervision, is considered the 'gold standard' of training. The CBT-E workshop training was provided by a leading expert in CBT-E. The two day workshop was approximately 12 hours in length and was mostly didactic in presentation style. Didactic material was supplemented with three previously recorded demonstrations of clinical interventions involving the presenter acting as a therapist and an actor role-playing a patient. The training covered the core-material of the treatment, focusing on patients who are not significantly underweight (that is have a BMI of >17.5). The training mirrored the four stages of the treatment and provided trainees with information about: assessing patients and engaging them in treatment, achieving early change; tackling the main mechanisms maintaining the eating disorder; reviewing progress; ending treatment and planning to maintain progress.

Web-centered training

Web-centered training is a new form of training. The training material (similar to the workshop content) was presented online. Trainees were given access to a CBT-E training website and provided with additional web-guidance (see below) over the telephone or Skype. The online training package is intended to be used in a modular fashion, with therapists viewing a segment of the website, while simultaneously treating a patient using the interventions described. To work through the online training took approximately nine hours. The majority of the content was delivered didactically (video based, oral presentation accompanied by PowerPoint slides as in a workshop) by the same presenter as the conventional workshop. However, additionally there were short clinical demonstration videos (“minis”); and interactive exercises, in which the therapist was asked to enter text or answer questions based on the material presented. As with the workshop, the online training material covers the same CBT-E core-material for patients who are not underweight. The course material, as with the workshop, follows the four stages of treatment.

Web-guidance

Web-guidance describes a form of support that aims to provide the trainee with encouragement and practical support in using the training materials provided by the website. This form of guidance is not supervision. Expertise resides in the online training program and not with the guide who does not necessarily need to be a clinician, let alone an expert in the treatment. This model of training is potentially much more scalable than most forms of therapist training as once the training materials exist support can be provided by a range of people who have become

familiar with the content of the training site. In this study web-guidance was provided by four guides – two were non-clinicians with no expertise in CBT-E, while for comparison purposes two were clinicians, expert in CBT-E. Despite their CBT-E expertise the clinicians were functioning as guides, not supervisors.

Each web-guidance session was standardized and included the following: 1) setting an agenda; 2) inquiring about any technical difficulties with the website; 3) inquiring about progress with the training website; 4) discussing any difficulties the trainees had experienced in using the treatment with their patients; and 4) summary and arranging the next session.

The main role of those who provide web-guidance is to motivate trainees to complete the online training and to encourage them to apply the treatment they are learning with fidelity. This is unlike the role of a clinical supervisor who provides expert advice about how to implement treatment. Thus both non-clinician guides and clinician guides referred trainees back to the training website to obtain information about the treatment. The clinicians were able to refer to their clinical experience when suggesting that trainees follow the treatment as described.

Web-guidance sessions were delivered via telephone or Skype on an individual basis. Each session lasted up to 30 minutes. Therapists were offered up to 12 guidance sessions over a 6 month period. Guidance sessions roughly followed the pattern of treatment appointments and were offered weekly for the first four weeks and fortnightly for the remaining 16 weeks.

(iii) Sample recruitment

All therapists who attended training were eligible to participate. Both samples were recruited purposefully using non-random sampling. Non-random

purposive sampling, as described above (section 3.2.1), ensures that an in-depth understanding of phenomena of interest will be gained by focusing on those individuals who have specifically experienced the event of interest (i.e., training in CBT-E).

Workshop training: Group one

All therapists currently providing psychological treatment and attending the training workshops on CBT-E were eligible to participate. The primary researcher attended two separate CBT-E workshops in Reading, UK and Monaghan, Ireland and described the study to the therapist trainees during breaks in the workshop. Therapists were given an information sheet (see Appendix 1.2) and if interested asked to provide their email address to be contacted to participate in the study. Three days after attending the workshop attendees who agreed to be contacted were asked to formally give consent via an electronic consent sheet (see Appendix 1.3). All therapists who provided consent were scheduled for an interview within two weeks.

Web-centered training: Group two

All therapists participating in the web-centered training on CBT-E were eligible. Recruitment was purposive to ensure the participation of members of each of the professional groups receiving training. Those who expressed an interest were sent an online consent form (see Appendix 1.3) and an interview was arranged within the following two weeks.

Recruitment for both groups was planned to continue until theoretical saturation was achieved, as described earlier (section 3.2.1). To ensure that

saturation was truly met, it was also planned to recruit trainees from each professional group in case these groups offered different or new perspectives.

(iv) The researcher

The relationship between the researcher and the interviewees is important in all research and potentially even more so in qualitative research (Kvale, 2006). The primary researcher's role differed with regard to the workshop and web-centered training groups. When interviewing trainees in the workshop group the researcher's role was simply one of information gatherer independent of the training provided by the workshop. As she was one of the clinician web-guides her role was much more complex in gathering information about web-centered training. For this reason the numbers of interviews planned with clinicians she guided was kept to a minimum, consistent with ensuring the participation of all professional groups. Procedures for extra vigilance concerning this blurring of roles and its impact were planned, such as a content analysis by an independent party to examine if the interviews conducted with the therapists she guided contained different topics or seemed overly positive in comparison to the rest.

The researcher attended two training courses offered at the University of Oxford on qualitative research: "Introduction to Qualitative Research Methods" and "Introduction to Qualitative Interviewing".

(v) The interview schedule

The interview was designed specifically for this study (see Appendix 1.4) with guidance from experts at the Health Economics Research Centre at the University of Oxford. The interview went through numerous versions and was piloted on several psychological therapists (n=3) prior to its use. The therapists included one CBT

therapist and two clinical psychologists who had attended two day trainings for other treatments in the past. Based on their feedback the interview was revised. Few adjustments were needed apart from the addition of two topic areas (‘views on supervision’ and ‘use of outcome measures’). The pilot also gave the researcher a chance to familiarize herself with this style of interviewing.

Topic areas to be covered (see Table 3.2) rather than specific questions were used to ensure that participants could express their views without being led or confined by questions. Interviewees were invited to give their views with the researcher only prompting with open ended questions when necessary (i.e., when the topic had not already been covered. The final section contained some specific questions to collect demographic details.

<i>Table 3.2 The main topic areas covered by the interview schedule</i>	
	• Prior general training • Experience of training/learning treatments • Evidence-based treatment/treatment manuals/protocols etc. • The treatment (CBT-E) • The workshop/ training - format of the presentation - relevance of presentation to interviewee’s clinical work - interest in material - effectiveness as a method of learning • Perception of competence • Likelihood of implementation • Perceived barriers to implementation-internal or external • Supervision • Use of outcome measures

(vi) Interview procedure

All interviews were conducted via telephone¹, Skype, or in person. At the start of the interview, a general description of the study was read out (Appendix 1.4). Participants were reminded that the interview would be digitally recorded and later transcribed. No reimbursements or incentives of any kind were offered.

(vii) Data handling and coding

Interviews were digitally recorded, transcribed verbatim, coded and analyzed using a Computer Assisted Qualitative Data Analysis Software (CAQDAS) program, NVivo. The recordings, transcripts, themes and codes are all stored in the NVivo software package currently held on an internal network drive at the University of Oxford. Table 3.3 outlines the process that was followed for data coding.

The analysis was guided by grounded theory, as described earlier. First all interviews were initially coded, using open coding. Two coders, the researcher (SBS) and an external independent coder, independently coded the interviews to reduce any bias in interpretation. Each identified and subsequently discussed emerging themes and patterns and any apparent discrepancies (axial coding). All discrepancies were resolved by either reaching consensus or creating a new code (if necessary). Constant comparative analysis was used to examine the data and to create a thorough list of themes and codes that adequately represented the data. Once a consensus was reached all previously coded transcripts required re-coding to ensure that the new code was not missed (selective coding). This process was carried out at least three times for each interview.

¹ Although a concern to some in the field of qualitative interviews, telephone interviews have been shown to produce data of equal quality to those conducted in person (e.g., Crippa et al., 2008; Midanik & Greenfield, 2003).

Table 3.3 Summary of the data coding process

1. Discussions were held with the primary investigator's supervisor to ensure that all areas were adequately covered, to decide whether or not to include new areas, and to assess the quality of the interviews
2. Preliminary coding was carried out after each interview using NVivo software. Each line of text was explored and coded for any themes of potential importance.
3. An independent (external) coder coded all of the interviews to increase reliability and validity as described above.
4. Discussions were subsequently held between the primary investigator and secondary coder to uncover any additional codes or discrepancies.
5. A final version of codes was developed based on these discussions
6. All interviews were then recoded using the new agreed upon codes.
7. Finally all codes were analyzed and stored using NVivo to assess for frequency of topic.

3.5 Results

(i) Participants

In total 35 therapists who received training in CBT-E (n=17 workshop and n=18 web-centered training) participated in this study. Demographic details are summarized in Table 3.4.

Workshop training: Group one

The sample consisted of 17 therapists of varying backgrounds. The majority of the sample was female 82.4% (N=14), which was consistent with the composition of the workshop training group. The mean age was 42.8 (range= 27-57). The majority (N=8, 47.1%) of therapists had been working clinically for 5- 10 years. Therapists came from three main professional groups CBT therapists (N=5, 29.4%), clinical psychologists (N=5, 29.4%), and nurses (N=4, 23.5%). Of the remaining three, two were counselling psychologists (11.8%) and one was a psychoanalyst (N=1, 5.9%).

The average length of interview was 50 minutes. Most interviews were conducted over the phone (n=13), while two were conducted in person and one via Skype.

Web-centered training: Group two

The sample contained 18 therapists of varying backgrounds. The majority of the sample was female 88.9% (N=16), which was consistent with the composition of the web-centered training group. The mean age was 42.8 (range= 27-57). The majority (N=10, 62.5%) of therapists had been working clinically for over 10 years. The majority of therapists were either clinical psychologists (N=6, 33.3%) or nurses (N=6 33.3%), with other professional groups including occupational therapists (N=2, 11.1%), social workers (N=2, 11.1%), psychiatrist (N=1, 5.6%), and nutritionists (N=1, 5.6%).

The average length of interview was 45 minutes. Nearly all interviews were conducted via the telephone (n=16), with two being conducted over Skype. In total three (16.7%) therapists who were guided by the primary investigator also completed an interview. An independent assessor blind to both therapist and guide examined a collection of transcripts including the three of specific concern here for any differences in content and found none.

There were no significant differences between the two groups of therapists in terms of age, $X^2(2, N=35) = 2.18, p = .38$, gender, $X^2(1, N=35) = .31, p = .58$; or years providing psychological treatment, $X^2(2, N=35) = 4.24, p = .12$.

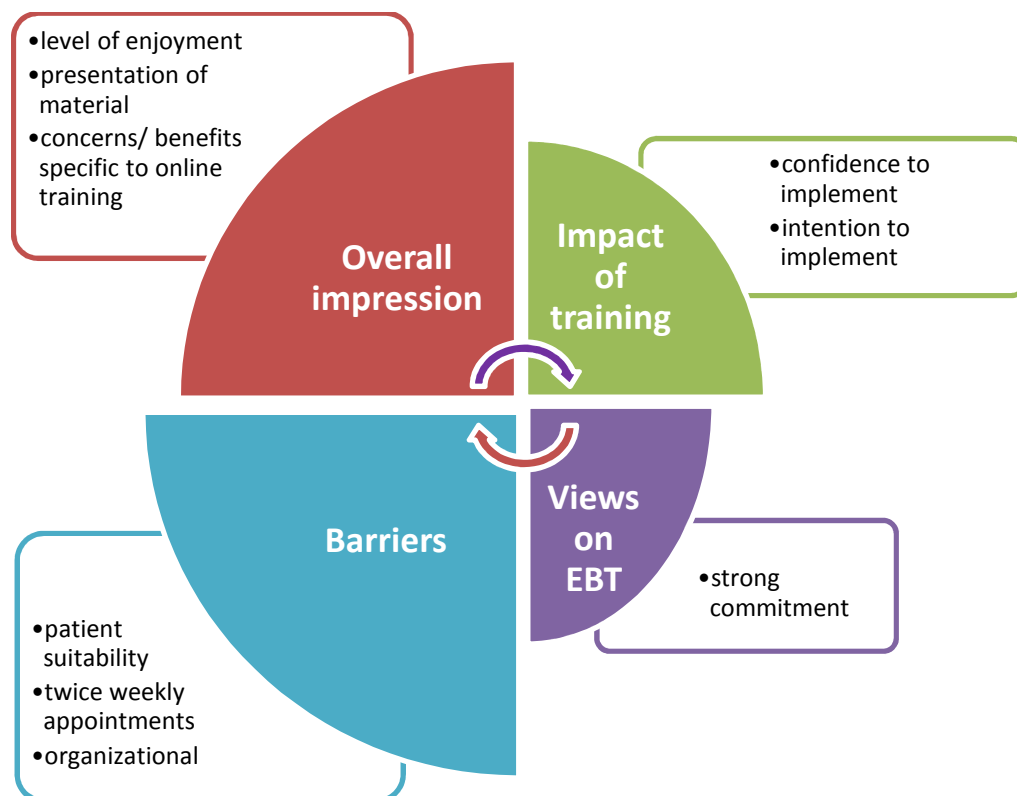
Table 3.4 Demographic details of therapists

	Post-Workshop training (N=17)	Post- Web-centered training (N=18)	Total (N=35)
Gender			
-female	14 (82.4%)	16 (88.9%)	30 (85.7%)
-male	3 (17.6%)	2 (11.1%)	5 (14.3%)
Age			
-20-29	3 (17.7%)	1 (6.3%)	4 (11.4%)
-30-39	9 (52.9%)	8 (44.4%)	17 (48.6%)
-40+	5 (29.4%)	9 (50%)	14 (40%)
Years providing psychological treatment			
-1-4 years	5 (29.4%)	2 (12.5%)	7 (20%)
-5-10 years	8 (47.1%)	6 (37.5%)	14 (40%)
->10 years	4 (23.5%)	10 (62.5%)	14 (40%)
Professional background			
-CBT therapist	5 (29.4%)		5 (14.3%)
-Clinical psychologist	5 (29.4%)	6 (33.3%)	11 (31.4%)
-Counselling psychologist	2 (11.8%)		2 (5.7%)
-Nurse	4 (23.5%)	6 (33.3%)	10 (28.6%)
-Psychoanalyst	1 (5.9%)		1 (2.9%)
-Occupational therapist		2 (11.1%)	2 (5.7%)
-Social worker		2 (11.1%)	2 (5.7%)
-Psychiatrist		1 (5.6%)	1 (2.9%)
-Nutritionist		1 (5.6%)	1 (2.9%)

3.5.2 Themes

Data from both groups despite the different methods of training can be seen as falling into four broad categories generally consistent with the aims of the investigation (overall impression, impact of the training, barriers to implementation, views on evidence-based training; see Figure 3.2). As depicted by slice size in Figure 3.2, 'Barriers to implementation' was the most commonly mentioned category. Within these categories several themes emerged. Themes are based on data from both groups combined that is all thirty-five participants. A remarkable degree of uniformity emerged across the groups, with few, but important differences related to the method of training.

Figure 3.1 Main qualitative categories and themes



(i) Overall impression

As expected, therapists tend to enjoy attending training. Both workshop and web-centered therapists were generally favorable in their assessment. Many therapists endorsed views similar to the following:

*"I found it [the workshop] very useful, very informative; I thought we got a lot of important information."*¹ -workshop trained therapist, Nurse, age: 32

*"I think it's easier than trying to read material in a book and an easier way to understand it [the treatment] and kind of multi-sensory, if you like. I think it's a better way to do training."*² -web-centered trained therapist, Clinical Psychologist, age: 47

Presentation of the material

Views on the best way to present material were similar for the two groups. In the workshop training only a minority of therapists preferred the didactic lecture style as this therapist described:

*"... being able to sit back and listen...and absorb it [the training]...I am very comfortable with that kind of style [didactic lecture]...when...the person delivers...essentially what they are there to deliver...and not...veer off too much...to the detriment of the information that they are trying to get across...that style suits me..."*³ -workshop trained therapist, CBT Therapist, age: 55

However, most workshop trained therapists wished for more interactivity (both in the form of role-plays with fellow therapists and the opportunity to brainstorm service specific concerns). The following comment was typical:

*"I would have wanted...more experiential learning, having information to read is great...but actually putting that into practice as well is useful, because doing it and using different types of methods, it is more likely to be retained, and taken forward."*⁴ -workshop trained therapist, Clinical Psychologist, age: 39

Although some therapists mentioned that their own anxieties would have prevented them from being able to participate fully with increased interactivity had it been available. One therapist expressed this well:

“...some workshops...divide people up into groups and you...are expected to participate and I find that uncomfortable with people I do not know...at those times...I am unable to take in new information...”⁵ -workshop trained therapist, Counselling Psychologist, age: 46

Similarly, therapists in the web-centered training preferred the interactive elements in the web-training rather than the more didactic video lecture format. The following quote illustrates this well:

“Those things [MCQs, interactive exercises, acted demonstrations] really jumped out rather than just the video and him talking - anything with a different medium.”⁶ -web-centered trained therapist, Social Worker, age: 41

Therapists in both groups reported that they found the acted clinical demonstrations to be particularly enjoyable and useful. One therapist expressed this viewpoint succinctly:

“I think seeing it [the treatment] in action is always helpful...”⁷ -workshop trained therapist, Nurse, age: 43

The workshop training contained three demonstrations and therapists often mentioned that they would have preferred more. Several workshop therapists also requested shorter clips feeling that the clips shown, one of which lasts 30 minutes, were too long. The web-centered training contained more clinical demonstration videos of much shorter length (approximately 3-5minutes each). Web-centered trained therapists frequently commented on how much they enjoyed the videos with none expressing any adverse comments.

Concerns specific to online training

A number of therapists mentioned having reservations about online training.

One described the disappointment of receiving online training well:

*"Initially when I heard we were getting the training online, I thought: "Oh, that's kind of disappointing", but I was actually very encouraged and I did find it really good. I said to people that were being sceptical, that you should absolutely do it, it's definitely worthwhile..."*⁸ -web-centered trained therapist, Social Worker, age: 41

While another therapist expressed concerns over the technology itself:

*"All of that [online training] took me a little bit of time, someone else maybe they would have had it spot on, but I'm afraid for me it took a little bit of time to get used to... I kept thinking: "am I pressing the right buttons?""*⁹ -web-centered trained therapist, Nurse, age: 50

Additionally, therapists who received online training reported missing some aspects of in-person training. The following comments from two separate therapists illustrate this:

*"The down side is that it's [web-centered training] not as interactively based as you can interact with a person when you are attending a training event."*¹⁰ -web-centered trained therapist, Clinical Psychologist, age: 47

*"In terms of motivation and attention mine is probably better if I go along to an actual training."*¹¹ -web-centered trained therapist, Clinical Psychologist, age: 47

Technical problems with the training website were initially a major concern. Many of the therapists used health service networked computers with out of date software that they were not able to update independently. In addition the broadband capacity was patchy in a number of areas. The frustration caused by technical failings cannot be overlooked. One polite therapist commented:

*"I know how to use a computer but when things aren't working out or straight-forward I get a little perplexed"*¹² -web-centered trained therapist, Occupational Therapist, age: 45

While another said:

*"...the technical requirements need to be hammered out before [offering the training to others]...The problem of the thing freezing –you're just bored and frustrated out of your mind."*¹³ –web-centered trained therapist, Nurse, age: 33

Benefits specific to online training

Some positive aspects of web-centered training that were mentioned

frequently included the ability to re-watch segments, and take the training at the therapists' own pace. For example two therapists mentioned:

*"With it [training] being online, you could go back and listen again, whereas when it's live you get one shot and if you get distracted - as you can easily do – at least with the online you can go back again and get it a second time."*¹⁴ - web-centered trained therapist, Clinical Psychologist, age: 48

*"So from a practical clinical point of view...it was a brilliant program in terms of being able to look at it in the evenings..."*¹⁵ -web-centered trained therapist, Nurse, age: 29

Also, the practical aspect of less travel was mentioned as this therapist commented:

*"You have to travel to the UK or USA to get this kind of training. Which obviously has drawbacks, there is rarely training like this accessible..."*¹⁶ -web-centered trained therapist, Nurse, age: 29

The addition of web-guidance to the online training material (obviously not a feature of the workshop training) was rated extremely positively by therapists. The following quote represented a common view:

*"The phone conversations I found really helpful. Absolutely they were invaluable."*¹⁷ -web-centered trained therapist, Clinical Psychologist, age: 47

Therapists frequently commented that the guides offered reassurance and motivation to persevere with the treatment and training as this therapist commented:

*"I think it's [web-guidance] brilliant...because it encourages you to keep going with the course as well, because...you are going to have the phone call and you want to have the stuff done cause if there is anything that comes up. You want to have the work done before you have the chat with your supporter."*¹⁸
-web-centered trained therapist, Clinical Psychologist, age: 47

While trainees did not comment on whether their guides were clinicians or not, the concept of web-guidance as opposed to clinical supervision was confusing for some therapists. One therapist pointed out:

*"To clarify, when you say "web-guidance", what do you mean? The supervision?"*¹⁹ –web-centered trained therapist, Psychiatrist, age: 51

With some therapists expressing concern that mistaking the two could be problematic, as this therapist did:

*"...there's a risk that web-guidance is mistaken for proper clinical supervision and that's what's actually happened...it [web-guidance] has been mistaken for actual supervision and that's a dangerous territory."*²⁰ –web-centered trained therapist, Nutritionist, age: 36

(ii) Impact of the training

Confidence

Almost all therapists in both the workshop training and web-centered training described feeling more confident in their ability to treat eating disorder patients with CBT-E following the training. For example therapists said:

*"I think for my own confidence it [the training] really has consolidated a lot of things. And has helped me personally with that."*²¹ –workshop trained therapist, Counselling Psychologist, age: 46

*"I've gone from maybe a '2', in terms of confidence, no not even a '2', maybe a '1' in terms of confidence before having done all of this [training] and pored over everything. Now I'd say maybe a '7'."*²² -web-centered trained therapist, Nurse, age: 38

At the same time many therapists, in both forms of training, expressed concern about the ability of others undergoing the training, although none placed themselves in this less-able category. This web-centered trained therapist explained it well:

*"I just would be very concerned that people could go on this training and assume that they can just do it because they've done this video training with the book. I think it does a disservice to...say that and that's my worry really."*²³ -web-centered trained therapist, Nurse, age: 29

Many therapists in the workshop trained group but not the web-centered trained group, felt that the training was too short in length. This therapist captured the sentiment of others by offering:

*"...you are talking about delivering a therapy that lasts 20 weeks, then you need at least double that!"*²⁴ -workshop trained therapist, Clinical Psychologist, age: 43

Implementation

Despite their favorable impression of the training and their perceived improved ability to use the treatment, very few of those in the workshop group planned to change their practice as a result of training or implement the treatment as described during training. For example, one therapist acknowledged:

*"I couldn't, hand on heart, say that I am fully doing CBT-E...I am sort of doing it, I feel that I am doing a bit of it, but I don't think, I have pulled it all together, where I can honestly say I am delivering it the way that...it needs to be delivered..."*²⁵ -workshop trained therapist, Counselling Psychologist, age: 46

Most in the workshop group believed that the treatment was not suitable for their patients as this therapist explained:

"I guess there was (sic.) loads of times that I felt: "well that [the treatment] is great to hear, but actually it doesn't really fit with the patient group..."²⁶
-workshop trained therapist, Clinical Psychologist, age: 39

Specifically many therapists felt that the treatment could only be used with a small minority of their patients. Another therapist reported:

"I suppose I would be able to use the whole CBT intervention with the manualized components...I guess with about 15 to 20 percent of the people that I see..."²⁷ -workshop trained therapist, Nurse, age: 31

Many therapists believed that their patients would require the blending of two or more approaches and planned to do this. For example:

"I have had quite a lot of training in EMDR and IPT and solution focused things and...other things...like re-scripting, so...they do creep in when you are trying to think of different strategies to keep the whole thing in place and to really help the patient..."²⁸ -workshop trained therapist, CBT Therapist, age: 55

"Obviously...anything that you can use from your experience is going to be a benefit to the client."²⁹ -workshop trained therapist, Clinical Psychologist, age: 42

In contrast, many therapists who received web-centered training intended to use the complete treatment with their clients as this therapist indicates:

"I think personally I would use it [the treatment] as it is and for two reasons: one...you cover all the major issues in this [treatment] and two, if you stick to the plan...you should have everything covered that you need to cover."³⁰
-web-centered trained therapist, Nurse, age: 37

(iii) Barriers

Patient suitability

A variety of perceived barriers to using the treatment were mentioned by therapists in both forms of training. The most frequent one was that the treatment was unsuitable for many patients routinely seen in clinical practice. Therapists, particularly those in the workshop training group, described their patients as more complex and as having more co-morbidity than those seen in the treatment trials that provided the evidence for CBT-E. A number of therapists believed that the exclusion criteria for trial participation are too rigid. In addition many therapists expressed concern about patients with personality disorders. The following examples capture these problems well:

“...it is hard...to be strict and manualized with really complex patients...”³¹
- workshop trained therapist, Nurse, age: 43

“I guess it [the difficulty with generalization] is just the research trials and who the research is done on, how they are selected.”³² -workshop trained therapist, Clinical Psychologist, age: 29

“This model works really well, but it’s not going to be suitable for the kind of complex people that we’re seeing.”³³ -web-centered trained therapist, Social Worker, age: 54

“I do think that a lot of people with eating problems have significant personality problems. I don’t think the little module on events, moods and eating, or whatever it is covers it. I think it [the treatment] over simplifies that piece of it.”³⁴ -web-centered trained therapist, Social Worker, age: 41

Twice-weekly appointments

Another barrier that was commonly mentioned was the strong recommendation in CBT-E for twice-weekly appointments during the first month of

treatment². Therapists from both training groups listed practical difficulties with meeting twice-weekly that included part-time work, travel distance for both the therapist and the client, school/ work commitments for the patient, and organizational pressures to see more clients. Comments like the following were common:

*“...how are we going to juggle that [twice weekly appointments] with the targets that we have to meet?”*³⁵ -workshop trained therapist, CBT Therapist, age: 31

*“I work in a rural setting so I am moving around from clinic to clinic...twice weekly [sessions]...trying to organize that with clinics would be tough.”*³⁶ -web-centered trained therapist, Social Worker, age: 41

Patient motivation

Therapists in both groups mentioned concerns regarding patient motivation and difficulties with engaging patients, as this therapist commented:

*“We have to consider doing lots of motivational work at the beginning as well which can take up a portion of the treatment...”*³⁷ -workshop trained therapist, Clinical Psychologist, age: 48

Many clients believed that the client examples in the training were not similar in terms of motivation to the kinds of patients they met with. The following comment was typical:

*“Well, of course, that would be great if my clients were fully compliant, like the ones described in the training, but, you do get clients who aren’t compliant and I would have liked to see a bit more about how to trouble shoot when you weren’t getting anywhere.”*³⁸ -web-centered trained therapist, Nurse, age: 37

² This level of frequency is intended to assist with engagement and motivation of the patient. Additionally, it has been shown that progress made in the first few weeks of CBT for eating disorders is predictive of later treatment success (Agras et al., 2000; Fairburn, Agras, Walsh, Wilson, & Stice, 2004).

Organizational barriers

A number of organizational barriers were mentioned particularly by therapists in the workshop training. The obstacles included lack of support from the service, and long waiting lists. For example:

*"...there would be some resistance within the team for us to work in a different way..."*³⁹ -workshop trained therapist, Nurse, age: 59

*"...but by the time they [patients] have been sitting on a waiting list for a year...Their motivation is really different and they may not be wanting to engage in treatment..."*⁴⁰ -workshop trained therapist, Nurse, age: 43

Therapists in both forms of training mentioned a lack of ongoing supervision as a potential barrier as these two therapists said:

*"...what we really need now is supervision where you can actually relate what is going on and get feedback from somebody who is experienced and working in that area...but our supervision, when we get it, would be from someone who hasn't been trained in CBT-E herself..."*⁴¹ -workshop trained therapist Clinical Psychologist, age: 37

*"...the supervision is the piece that...I will miss really...I know you can't [continue to offer web-guidance] but...for the future that is something that's got to be considered. You know, there is no supervision offered in this service"*⁴² -web-centered trained therapist, Nutritionist, age: 36

(iv) View of evidence-based treatments

Nearly all therapists in both training groups expressed a strong commitment to evidence-based treatments. The following quotes illustrate this well:

*"I think it is unethical to abandon the most robust therapies just because you would like to do something else. You need to try the things with evidence first. You need to have a pretty good reason as to why you would be deviating."*⁴³ -workshop trained therapist, Clinical Psychologist, age: 55

*"I think it is really important that the treatment you use has an evidence-base, otherwise you are...doing things for other reasons. I don't know. I don't see why there would be any problem with evidence-based treatment."*⁴⁴ -web-centered trained therapist, Nurse, age: 53

3.6 Discussion

This study aimed to investigate the reactions that therapists have to training in CBT-E. Two different methods of training were investigated, a standard training typical of that usually offered and a new innovative online method. Trainees' attitudes and reactions were explored using a qualitative method for two reasons: qualitative inquiry is the method of choice when the phenomena of interest are poorly understood, as is in the case of therapist training, and post- training feedback forms have been shown to have questionable utility (as argued previously). Qualitative methods have not previously been used to study the reaction to training in evidence-based psychological treatments and CBT-E training specifically has not been previously studied.

3.6.1 Reactions to two methods of learning CBT-E: similarities and differences

Therapist reactions to training related to four broad categories: impression, impact, barriers, and views on evidence-based training. Themes within these categories were similar for both groups, but their views relating to these themes sometimes varied with training method.

Impression

With regard to enjoyment, therapist reactions to both workshop training and web-centered training were largely extremely favorable. Therapists tended to find the material useful and described enjoying the time they had spent learning.

Overall, most therapists preferred a mix of presentation style with discussion, role-play, questions and video demonstrations in addition to didactic lecture material. Most in the workshop training would have preferred more interactive features, and those in the online training commented how much they liked the more

numerous interactive features in this training format. The acted clinical demonstrations were a favored component for both groups. It seems that interactive components, made training more enjoyable and hence possibly more effective.

Differences emerged with regard to views about initial perception of training, the length of training, and the ease of accessing training. Despite subsequently feeling positive about the web-centered training, a number of therapists in the web-centered group described initial feelings of disappointment that the training would be online and would not involve interaction with a presenter. Some also mentioned being fearful of the technology. Most of these concerns were short lived and dissipated with experience using the website. Trainees appreciated the benefits of web-centered training, including less travel, the ability to work in their own time and the option of participating in web-guidance.

For some, a number of technical problems were a major drawback of the web-centered training. While these problems of course can potentially be overcome through the use of up-to-date computers and software and reliable internet connections, they do need careful consideration in situations where infrastructure might be outdated or absent. Attention to such technical issues is essential when considering the widespread dissemination of training necessary to meet the need for treatment not only in developed countries, but also in low and middle income countries. In these situations it may be that very simple online components, downloadable materials (as opposed to streaming live video) and written materials may be better.

For those who received workshop training this format was obviously familiar, but they believed that the training over two full days was not long enough- a view not shared by those in web-centered training.

Impact

Most therapists in both groups described confidence in their ability to use the treatment. However, while confident in their own abilities they tended to doubt their colleagues abilities. This is puzzling but perhaps in line with inconsistent research findings about how accurate trainees are at judging their own performance (e.g., Brosan et al., 2008; McManus et al., 2012).

Despite favorable reactions to training from both groups, only those in the web-centered training intended to change their practice. It was clear that those in the workshop group had little intention of using the interventions as a complete protocol while those in web-based training expressed a desire to do so. This is surprising in the case of the workshop group who did also express confidence in their ability to use the treatment, but is perhaps related to trainees in the web-centered group having had the opportunity to practice with a patient while receiving guidance.

Barriers

Both groups frequently mentioned that they thought the treatment was not suitable for the patients that they treated in routine practice. Although in both forms of training trainees are briefly told that the research that underpins CBT-E is based on research trials with relatively few exclusion criteria (Fairburn et al., 2009, 2015) and that similar results have been found in community-based studies (Byrne et al., 2011; Knott et al., 2014; Turner et al., 2015) this information appears not to convince

trainees or perhaps is not conveyed in enough detail to them. Therapists still describe their patients as being more complex, having more comorbidities and being less motivated than patients seen in research trials. For this reason they intend from the outset not to offer the treatment as described but to combine it with other treatments or in some cases to offer an entirely different treatment. This is consistent with previous research (outlined in Chapter One) (e.g., Turner et al., 2014; Wallace & von Ranson, 2012). It seems that training in the future needs to devote more time and resources to understanding and better directly addressing these concerns.

A number of organizational barriers were also mentioned by both groups. These included resistance from the team, no clinical supervision, and long waiting lists. Finally there were some practical, CBT-E specific barriers. The main one mentioned was twice-weekly sessions. Few therapists believed they would be able to offer sessions at such frequency. The main obstacles concerned part-time work, rural locations, travel times for the clients and the therapists, and employer targets concerning numbers of patients seen. It may be helpful to include a discussion of these barriers in training either with the presenter in didactic workshops or on online forums with web-centered training. Perhaps allowing time for addressing these issues, possibly within a problem solving or brainstorming framework will help generate helpful ways to overcome these barriers. For example, may not consider that it might be possible to arrange telephone or Skype therapy sessions instead of face to face sessions to overcome the problems of travel.

Views on evidence-based treatment

Trainees in both groups had positive views of evidenced-based practice, with many believing that it would be unethical to offer anything else. This might seem inconsistent with the report of many that they use a combination of treatments. However, it is perhaps more likely that researchers and clinicians interpret the notion of evidence base differently. While researchers view “evidence-based” as implementing treatment with fidelity to a treatment protocol (e.g., McHugh et al., 2009; Waller, 2009), therapists see evidenced-based treatment as using any part of treatment that has some basis in evidence. This is hardly surprising as the literature in this area tends to use the two terms, EST and evidence-based practice largely interchangeably, even though there is a fairly clear distinction to be made (Lilienfeld, Ritschel, Lynn, Cautin, et al., 2013; Lilienfeld, Ritschel, Lynn, Brown, et al., 2013). Again some further discussion about these issues may help including greater clarity about when it may be appropriate to follow a protocol with fidelity and when a more evidence-based practice approach should be adopted.

Strengths and limitations

(i) Strengths

The rigor of qualitative research is frequently questioned. In this case every attempt was made to follow all the recommended guidelines for qualitative research and a number of steps were taken to ensure the validity of the study. Prior to starting the research the study was planned with guidance from the Health Economics Research Centre at University of Oxford (HERC). With regard to reliability, peer review was established through regular meetings with an academic located at the HERC who was outside of the specific field of research but who had advanced

knowledge of qualitative research. This allowed the independent checking of ideas and interpretation of the codes. To ensure that codes were accurately created multiple coding was employed with two investigators (one researcher and another, independent of the project) analyzing the transcripts for codes and themes. The use of multiple coders also allowed for investigator triangulation to be met. The creation of codes followed the four steps outlined by Glaser (1969) as described previously and all transcripts were read and coded a minimum of three times to ensure that no codes were missing or overlooked. Procedural reliability was met through precise documentation of all interviews, transcripts, codes and data storage (Table 3.6 shows exactly where in each transcript a corresponding code is located on the university network). Strict adherence to grounded theory and comparative analysis helped to ensure validity. The study also conformed to the checklist for qualitative research produced by the National Institutes of Health and Care Excellence (NICE). This list, presented in Table 3.5 shows how the study met these guidelines.

(ii) Limitations

The study does have several limitations. Apart from the limitations inherent in the qualitative method, there were several study specific concerns. The first and most important is the lack of data triangulation. For practical reasons it was not possible to collect data by a complementary method. Without the ability to compare and contrast information collected by more than one method an important component of validity may be compromised.

A second limitation was the researcher's role. With regard to the workshop training group the researcher's role was more neutral. For the web-centered training

the role of the researcher was more complicated with her role being both a clinician web-guide and also an interviewer. This blending of the roles is problematic given the importance of the role of the researcher in qualitative inquiry.

A third limitation concerns the composition of the two training groups. As can be seen from Table 3.4, the samples are equivalent in many respects but differ significantly with regard to professional background. Some caution is needed in comparing the results as diverse groups may react somewhat differently and one cannot therefore necessarily attribute any differences to training method alone. The training therapists received also differed in one important way other than in their method of presentation. Web-centered training included web-guidance and the opportunity for therapists to practice what they had learned with the online training. This makes it difficult to know whether any observed differences are attributable solely to the mode of presentation or to the guidance or both. However, since the main interest was to understand reactions to different types of training (where each method is offered as a complete training) this does not seem such a problem. Of course it could be argued that workshops should be followed by ongoing supervision, but in practice this rarely happens and the only choice for those wanting training may be a workshop.

3.6.2 Summary

Several potentially important findings emerged from this study. First, therapists found web-centered training to be acceptable. Despite some initial concerns about online training nearly all therapists enjoyed and completed the training. Second, trainees particularly valued the interactive components of the web-centered training. Third, web-guidance was seen to be a useful addition to the online

training and therapists were not concerned about the clinical qualifications of their guide. Fourth, trainees also valued workshop training and preferred training to have more interactive components. Fifth, those who received web-centered training planned to implement treatment, while workshop attendees did not. Of course whether they actually do so requires further investigation.

This investigation also highlighted areas of training that require improvement. First, trainees widely held views that evidence-based treatments are not suitable for the patients they see in routine clinical practice need to be addressed more successfully. Second, the different views that researchers and therapists hold regarding what constitutes evidence-based treatment also needs to be addressed within training.

3.7 Conclusions

Therapists are essential in bridging the research-practice gap and yet research focusing on the views that clinicians have about training, has been relatively neglected. This study shows that new methods of training are well received by therapists. Of particular importance is that, although clinical supervision is seen as indispensable by therapists, it may be possible to offer a less time consuming alternative that does not rely on clinical expertise. The implication of these findings is that more scalable web-based training may be an alternative to conventional training. This clearly needs further investigation. In particular, it needs to be established whether web-centered training is capable of producing competent therapists with the knowledge and skills to implement treatment well.

Table 2.1 Direct Training Outcome Studies – Uncontrolled Studies

Study	Disorder and Treatment	N=	Outcome domain	Study design	Assessment measure	Reliability	Results
(Baer et al., 2004)	Substance Use Motivational Interviewing (MI)	22	2) Knowledge	WS	Helpful Responses Questionnaire (HRQ)	$\alpha = .70-.84$	Increase in knowledge scores. Scores declined from post-training to follow-up both remained higher than baseline. No competence cut-point
(Bartholomew et al., 2007)	Substance Use Multiple Interventions	300	1) Reaction	WS	Feedback survey (Likert scale)	$\alpha = .72-.90$	Overall satisfaction high
(Beidas et al., 2009)	Anxiety	20	2) Knowledge	TM→WS	Knowledge questionnaire (MCQ)	$\alpha = .76$	Training increased therapist knowledge. No competence cut-point
	Cognitive Behavioral Therapy (CBT)		3) Skills		1 role-play (SP)	adherence (ICC=.98); competence (ICC=.63-.83)	No difference in the two training conditions. No participants reached competence cut-point. Competence cut-point determined.

Table 2.1 Direct Training Outcome Studies – Uncontrolled Studies

Study	Disorder and Treatment	N=	Outcome domain	Study design	Assessment measure	Reliability	Results
(Brookman-Frazee, Drahota, & Stadnick, 2012)	Autism Spectrum Disorder AIM HI	13	3) Skills	IWS+	1) review of treatment planning 2) 3 required in session therapist behaviors (RP)	ICC= .77-.83	Positive skills scores post- training. Assessment only occurred at post-. No competence cut-point.
(Carroll et al., 1998)	Substance Use CBT or Contingency Management (CM)	11	4) Skills self-report and/or chart review	WS+ for both training conditions CBT or CM	CBT therapist behaviors (self-report)	Not described	Increased use of CBT/ CM interventions post- training.
(Crits-Christoph et al., 1998)	Substance Use Multiple Treatments	65	3) Skills	WS+	a) Cognitive Therapy Scale b) Penn Adherence and Competence Scale for Supportive – Expressive Therapy c) Adherence and Competence Scale for Addiction Counselling (RP)	a) α = .92 b) α = .85-.89 c) α = .84-.87	Competence improved in cognitive therapy only, not supportive expressive dynamic therapy or individual drug counselling. Competence cut-point determined.
(Ekers et al., 2013)	Depression Behavioral Activation (BA)	10	1) Reaction	WS	Feedback survey (Likert scale)	α = .99	Overall satisfaction high.

Table 2.1 Direct Training Outcome Studies – Uncontrolled Studies

Study	Disorder and Treatment	N=	Outcome domain	Study design	Assessment measure	Reliability	Results
(Fadden & Vale, 1997)	Schizophrenia Behavioral family therapy (BFT)	86	4) Skills self-report and/or chart review	IWS+	Implementation of BFT (self-report)	Not described	More than half of the sample reported seeing families more frequently and offering BFT
(Hartzler et al., 2007)	Substance Use MI	23	3) Skills	WS+	Motivational Interviewing Treatment Integrity (MITI) scale (SP)	ICC= .44- .93	Skills increase on MITI. Competence cut-point determined.
(Hawkins & Sinha, 1998)	Borderline Personality Disorder DBT	109	2) Knowledge	WS+ (option of additional further training)	Knowledge questionnaire (MCQ and T/F)	Not described	Increase in knowledge scores. Additional training (led to higher scores. No competence cut-point.
(Henggeler, Chapman, et al., 2008)	Substance use CM	430	4) Skills self-report and/or chart review	WS plus adherence phone calls	Use of cognitive-behavioral and monitoring aspects of CM (self-report)	CBT = .89, monitoring = .99	Reported increased use of CM interventions

Table 2.1 Direct Training Outcome Studies – Uncontrolled Studies

Study	Disorder and Treatment	N=	Outcome domain	Study design	Assessment measure	Reliability	Results
(Hepner et al., 2011)	Depression CBT	5	3) Skills	IWS+	Adapted version of Cognitive Therapy Adherence Competence Scale – all sessions recorded and selected at random (RP)	Adjusted κ = .57-.68	Both adherence and competence above cut-points post-training. Competence cut-point determined.
(Jensen-Doss et al., 2008)	Trauma Trauma Focused CBT	66	4) Skills self-report and/or chart review	WS (some watched recorded workshop videos)	Technique usage (chart review)	κ = .15– 1.0	No changes in use of TF-CBT despite clinicians stating that the TF-CBT was the most effective treatment option.
(Jensen-Doss, Hawley, Lopez, & Osterberg, 2009)	7 different ESTs for children	197	1) Reaction	WS+ (not standardized)	Feedback survey (Likert scale)	α = .74	Overall satisfaction high. Therapists requested more supervision post- training.
			4) Skills self-report and/or chart review		Percent of clients with whom therapists used protocols (self-report)	Not described	Reported increased use of ESTs interventions post- training
(Karlin et al., 2012)	Depression CBT	220	1) Reaction	WS+	Feedback survey (Likert scale and open ended questions)	Not described	Overall satisfaction high.
			3) Skills		Cognitive Therapy Revised Scale on one self-selected tape (RP)	α = .59-.77	Increase in skills. Competence cut-point determined. Therapists reaching competence threshold raising from 41% to 87%

Table 2.1 Direct Training Outcome Studies – Uncontrolled Studies

Study	Disorder and Treatment	N=	Outcome domain	Study design	Assessment measure	Reliability	Results
(Karlin et al., 2013)	Insomnia CBT-Insomnia	182	3) Skills	IWS+	Competency rating scale all sessions (RP)	$\alpha = .93$	Increase in skills. Competence cut-point determined.
			4) Skills self-report and/or chart review		Application of the course content (self-report)	Not described	Reported use of CBT-I post- training with 47% of patients.
(Kobak et al., 2013)	Anxiety CBT	39	1) Reaction	OLT+	Feedback survey (Likert scale)	$\alpha = .82$; $\alpha = .90$; $\alpha = .87$	Overall satisfaction high.
			2) Knowledge		Knowledge questionnaire (MCQ)	$\alpha = .73$	Increase in knowledge scores. No competence cut-point.
			3) Skills		Yale Adherence Competence Scale- 2 role-plays in CBT *not just for evaluation purposes practice and feedback were given during training (RP)	ICC= .88	Increase in skills. Average scores being above the competence cut-point.
(Lewis & Simons, 2011)	Depression CBT	24	4) Skills self-report and/or chart review	IWS+	Implementation of interventions (self-report)	Not described	Reported 'some' increased use of CBT interventions post-training.
(Liddle et al., 2006)	Substance Use multidimensional family therapy	10	3) Skills	Extended WS+	Therapist behavior rating scale (RP)	α range .68–.77. ICC-range .90–.99	Increase in skills. No competence cut-point.

Table 2.1 Direct Training Outcome Studies – Uncontrolled Studies

Study	Disorder and Treatment	N=	Outcome domain	Study design	Assessment measure	Reliability	Results
			4) Skills self-report and/or chart review		Session content focus (session review)	Not described	Reported increased use of MDFT techniques.
(Martino et al., 2008)	Substance Use Motivational Enhancement Therapy (MET)	35	3) Skills	IWS+	Independent Tape Rater Scale (RP)	ICC = .57–.98	Increase in skills. Competence cut-point determined.
(Martino, Ball, Nich, Frankforter, & Carroll, 2009)	Substance Use MET	14	3) Skills	IWS+	Independent Tape Rater Scale (RP)	ICC = .66–.99	Skills use was mixed- advanced strategies used less frequently than fundamental strategies. Therapists rated themselves consistently higher for strategy use than the ITRS suggested. Competence cut-point determined.
			4) Skills self-report and/or chart review		Therapist Session Checklist MI-consistent behaviors (self-report)	Not described	Reported increased use of MET interventions post- training.
(Martino, Canning-Ball, et al., 2011)	Substance Use MI	26	1) Reaction	Stepped-OLT→WS→WS+	Feedback survey (Likert scale)	Not described	Overall satisfaction high.
			3) Skills		Independent Tape Rater Scale- self-selected client tape (RP)	ICC = .87-.88	Positive scores. Competence cut-point determined.

Table 2.1 Direct Training Outcome Studies – Uncontrolled Studies

Study	Disorder and Treatment	N=	Outcome domain	Study design	Assessment measure	Reliability	Results
(Miller & Mount, 2001)	Substance Use MI	2	2) Knowledge	IWS+	Helpful Responses Questionnaire	test–retest reliability $r = .45$; $\alpha = .92$	Increase in knowledge. No competence cut-point
		22	3) Skills		Motivational Interviewing Skill Code (MISC) (SP)	Unclear	Increase in skills, not mainlined at F/U. No competence cut-point.
(National Crime Victims Research and Treatment Center, 2007)	Trauma	3,558	1) Reaction	OLT	Feedback survey (Likert scale)	Not described	Overall satisfaction high.
	Trauma focused CBT		2) Knowledge		Knowledge questionnaire (MCQ)	Not described	Increase in knowledge scores. No competence cut-point
(Petry et al., 2012)	Substance Use CM	15	1) Reaction	WS+	Feedback survey (Likert scale)	Not described	Overall satisfaction high.
			2) Knowledge		Knowledge questionnaire (MCQ)	Not described	Increase in knowledge scores. Required a score of 80% to indicate 'passing' unclear if this indicates a competence cut-point.
			3) Skills		≥3 role-plays on CM (12 item checklist) (SP)	IRR= .92	Increase in skills. Competence cut-point determined.

Table 2.1 Direct Training Outcome Studies – Uncontrolled Studies

Study	Disorder and Treatment	N=	Outcome domain	Study design	Assessment measure	Reliability	Results
(R. Richards et al., 2011)	Depression	73	2) Knowledge	IWS	Knowledge questionnaire (MCQ)	Not known	Increase in knowledge. No competence cut-point
	Acceptance and Commitment Therapy (ACT)		4) Skills self-report and/or chart review		Extent to which practice have been directly or indirectly impacted following workshop (self-report)	Not described	83% reported the workshop influenced their work; 90% intended to use ACT in the future. All said they intended to integrate ACT with another approach.
(Russell et al., 2007)	Trauma EMDR	175	1) Reaction	IWS	Feedback survey (Likert scale)	Not described	Overall satisfaction high.
(Sanders et al., 2009)	Child Behavioral Problems Triple P	611	4) Skills self-report and/or chart review	IWS	Use and maintained use of intervention phone interview (self-report)	Not described	Not using the intervention (37.15%), initiating intervention use (62.85%), maintaining intervention use over time (58.27%)

Table 2.1 Direct Training Outcome Studies – Uncontrolled Studies

Study	Disorder and Treatment	N=	Outcome domain	Study design	Assessment measure	Reliability	Results
(Schoener et al., 2006)	Substance use MI	10	3) Skills	IWS+	Motivational Interviewing Skill Code (MISC) (RP)	Not described	Increase in skills. No competence cut-point.
(Schoenwald et al., 2008)	Child Behavioral Problems Multisystemic therapy (MST)	429	4) Skills self-report and/or chart review	Extended IWS+	MST Therapist Adherence Measure-revised (care giver rated adherence)	Not described	Not clear if care givers reported increases in adherence following training.
(Seng et al., 2006)	Child Behavioral Problems Triple P	579	1) Reaction	IWS	Feedback survey (Likert scale)	Not described	Overall satisfaction high. No differences between training completers and non-completers.
			4) Skills self-report and/or chart review		Use of intervention (self-report)	Not described	67.6% of training completers continue to use intervention; only 37.8% who do not complete training endorse use; significant difference on χ^2 analysis
(Shafer et al., 2004)	Substance Use MI	145	2) Knowledge	IWS	Knowledge questionnaire (MCQ) and Helpful Responses Questionnaire	Not described	Increase in knowledge. No clinically significant. Scores dropped at F/U. Scale contains a minimum score of 3 unclear if this is related to competence.

Table 2.1 Direct Training Outcome Studies – Uncontrolled Studies

Study	Disorder and Treatment	N=	Outcome domain	Study design	Assessment measure	Reliability	Results
			3) Skills		Motivational Interviewing Skill Code (RP)	Not described	No change across training, possibly due to ceiling effect or small sample size (n=9). No competence cut-point.
(Sholomskas et al., 2005)	Substance Use CBT	78	2) Knowledge	TM,OLT,WS+	Knowledge questionnaire (MCQ)	Not described	Increase in knowledge. WS+ out performed TM and OLT. TM performed the worst. No competence cut-point.
		78	3) Skills		Yale Adherence and Competence Scale- 3 role-plays (RP)	ICC= .83-.87	Increase in skills across conditions. WS+ condition improves skills more than TM. Competence cut-point determined. Percentage of clinicians meeting this criterion was 15% for the manual only condition, 48% for the Web condition, and 54% for the seminar plus supervision condition.
(Simons et al., 2010)	Depression CBT	12	3) Skills	IWS+	Cognitive Therapy Revised Scale – three self- selected video recordings (RP)	All tapes double-coded; consensus achieved	No increase in generic skills Increase in CBT skills. Competence cut-point determined.
(Vismara et al., 2013)	Autism Spectrum Disorder	24	1) Reaction	IWS	Feedback survey (Likert scale and open ended questions)	Not described	Overall satisfaction high
(Vismara et al., 2013)	ESDM		3) Skills		Fidelity scale ratings (RP)	ICC= .88	Increase in skills. Maintained at follow-up. Competence cut-point determined.

Table 2.1 Direct Training Outcome Studies – Uncontrolled Studies

Study	Disorder and Treatment	N=	Outcome domain	Study design	Assessment measure	Reliability	Results
(Vismara et al., 2009)	Autism Spectrum Disorder ESDM	10	1) Reaction	WS+ (some via teleconference)	Feedback form (Likert scale and open ended questions)	Not described	Overall satisfaction high
(Walser et al., 2013)	Depression Acceptance and Commitment Therapy	334	1) Reaction	IWS+	Feedback form (Likert scale)	Not described	Overall satisfaction high
			3) Skills		ACT Core Competency Rating Form tapes selected at random (RP)	Not described	Increase in skills. Competence cut-point determined.
			4) Skills self-report and/or chart review		a) number of patients treated with depression b) % of patients receiving ACT consistent with protocol c) therapists belief about efficacy of ACT d) implementation specific questions (self-report)	Not described	39% of therapists reported using protocol consistent ACT with their patients.
(Worrall & Fruzzetti, 2009)	Borderline Personality Disorder DBT	56	1) Reaction	OLT	Feedback survey (Likert scale)	Not described	Overall satisfaction high

Table 2.1 Direct Training Outcome Studies – Uncontrolled Studies

Note.

DISORDER AND TREATMENT:

AIM HI- An Individualized Mental Health Intervention for children with Autism Spectrum Disorder, **EMDR**- Eye Movement Desensitization and Reprocessing, **ESDM**- Early Start Denver Model, **PCIT**- Parent-Child Interaction Therapy, **Triple P**- Positive Parenting Program

DESIGN:

Primary abbreviations:

(⁺)= plus supervision, **IWS**=interactive workshop, **OLT**= online treatment, **TM**= training manual, **WS**= workshop

Other abbreviations:

$\frac{1}{2}$ **WS**= 3hrs, **ALTC**= alternative control workshop, **CB**= computer based, **CTT** =context tailored training with role-play and feedback, **IWSF**= interactive workshop plus feedback, **OLT+ME**= online treatment plus motivational enhancement, **OLT+ME+LC**= online treatment plus motivational enhancement plus learning community, **TAU**= training as usual, **TMV**= training manual plus training videos, **WLC**= waitlist control,

Assessment Measure: **MATCH**- matching alcoholism treatments to client heterogeneity, **MCQ**- multiple choice questionnaire, **RP**- real patient, **SP**- standardized patient, , **T/F**- true or false questionnaire

Table 2.1 Direct Training Outcome Studies – Uncontrolled Studies

Table 2.2 Direct Training Outcome Studies – Randomized Controlled Trials

Study	Disorder and Treatment	N=	Outcome domain	Study design	Assessment measure	Reliability	Results
(Baer et al., 2009)	Substance Use MI	123	1) Reaction	WS, CTT	Feedback survey (Likert scale)	$\alpha = .79-.93$	Overall satisfaction high. No difference between groups. At follow-up greater satisfaction in CTT condition on all domains.
			3) Skills		Motivational Interviewing Treatment Integrity scale (SP)	ICC = .67	Increase in skills. WS showing greater gains at post-training and CTT showing greater gains at follow-up. Competence cut-point determined
(Beidas, Edmunds, et al., 2012)	Childhood Anxiety CBT	115	1) Reaction	WS+, OLT+, IWS+	Feedback survey (Likert scale)	Not described	Overall satisfaction high. WS+ and IWS+ rated higher than OLT+
			2) Knowledge		Knowledge questionnaire (MCQ and T/F)	$\alpha = .76$	Increase in knowledge. Increase was not different between training arms. A score of 70% meant 'trained to criterion' unclear how this was established.
			3) Skills		1 role-play (SP)	adherence (ICC=.98); skill (ICC=.92)	Increase in skills. IWS+ improved skills more than OLT+. Competence cut-point determined.
(Beidas, Mychailyszyn, et al., 2012)	Childhood Anxiety CBT	17	3) Skills	WS+, OLT+, IWS+	3 role-plays (SP)	adherence (ICC=.98); skill (ICC=.92)	Increase in skills. Competence cut-point determined.

Table 2.2 Direct Training Outcome Studies – Randomized Controlled Trials

Study	Disorder and Treatment	N=	Outcome domain	Study design	Assessment measure	Reliability	Results
			4) Skills self-report and/or chart review		Implementation of interventions- Identification and Treatment of Anxious Youth (self-report)	Not described	Reported increased use of CBT interventions.
(Carroll et al., 2006)	Substance use MI	37	3) Skills	IWS, ALTC (alternative control workshop)	Yale Adherence and Competence Scale all session recordings (RP)	ICC = .81-.94	Increase in skills. IWS better than ALTC. Competence cut-point determined.
(Chu et al. 2015)	Childhood Anxiety CBT	35	1) Reaction	WS→ 3 different forms of extended training	Feedback survey (Likert scale)	$\alpha = .90$	Overall satisfaction high. No differences between groups.
			2) Knowledge		Knowledge questionnaire (MCQ)	Not described	Decrease in knowledge. No differences between groups.
			4) Skills self-report/ or chart review		Use of skills (SR)	$\alpha = .74$	Reported increase in CBT skills. No differences between groups.
(DeViva, 2006)	Substance use 12-step	60	1) Reaction	WS, 1/2WS	Feedback survey (Likert scale)	Not described	Overall satisfaction high. Differences between training groups not examined.

Table 2.2 Direct Training Outcome Studies – Randomized Controlled Trials

Study	Disorder and Treatment	N=	Outcome domain	Study design	Assessment measure	Reliability	Results
			3) Skills		2 role-plays (SP)	ICC = .85	Increase in skills. No difference between groups. No competence cut-point.
(Dimeff et al., 2009)	Borderline Personality Disorder Dialectical Behavior Therapy (DBT)	150	1) Reaction	WS, OLT, TM	Feedback survey (Likert scale)	Objectives $\alpha = .86$, training style and methods $\alpha = .95$	Overall satisfaction high. Higher satisfaction with learning objectives and presentation style in OLT and ILT than in TM.
			2) Knowledge		Knowledge questionnaire (MCQ)	Not described	Increase in knowledge. OLT better than WS or TM. Change maintained at follow-up. No competence cut-point.
			3) Skills		1 role-play in DBT phone coaching (SP)	ICC= .70-.77	Increase in skills. No differences between conditions. No competence cut-point.
			4) Skills self-report and/or chart review		Use of skills, use of treatment manual (self-report)	$\alpha = .97-.98$	Reported increased use of DBT interventions post- training. No differences across training condition in skills use. Differences across training conditions regarding training materials use with those in ILT and TM using training materials more than those in OLT.
(Dimeff et al., 2015)	Borderline personality Disorder DBT	172	1) Reaction	WS, OLT, TM	Feedback survey (Likert scale)	Not described	Overall satisfaction high. Higher satisfaction with ILT, than OLT and TM.
			2) Knowledge		Knowledge questionnaire (MCQ and matching)	Test-retest reliability $r = .71$	Increase in knowledge. OLT better than WS or TM. Change maintained at follow-up. No competence cut-point

Table 2.2 Direct Training Outcome Studies – Randomized Controlled Trials

Study	Disorder and Treatment	N=	Outcome domain	Study design	Assessment measure	Reliability	Results
			3) Skills		3 role-plays in DBT phone coaching (SP)	ICC= .91 and ICC= .87	Increase in skills. No differences between groups. No competence cut-point.
			4) Skills self-report and/or chart review		Use of skills, use of treatment manual (self-report)	$\alpha = .93-.97$	Reported increased use of DBT interventions post- training. No differences across training conditions.
(Gega et al., 2007)	Panic/ phobia Exposure Therapy	92	1) Reaction	WS, CD-ROM	Feedback survey (Likert scale)	Not described	Overall satisfaction high. No differences between groups.
			2) Knowledge		Knowledge questionnaire (MCQ and case scenario response)	Not described	Increase in knowledge. No differences between groups. No competence cut-point.
(Harned et al., 2011)	Anxiety Exposure Therapy	46	1) Reaction	OLT, OLT+ME, OLT placebo	Feedback survey (Likert scale)	(1) $\alpha = .83-.91$; (2) $\alpha = .82-.88$	Overall satisfaction high. OLT and OLT+ME higher than OLT placebo.
			2) Knowledge		Knowledge questionnaire (MCQ)	Not described	Increase in knowledge. OLT and OLT+ME better than OLT placebo. No competence cut-point.
			4) Skills self-report and/or chart review		Application of course content (self-report)	Not described	Reported increased use of exposure interventions post-training. OLT and OLT+ME using skills more than those in OLT placebo. Those in OLT placebo reported no change in skills use.
(Harned et al., 2014)	Anxiety Exposure	182	1) Reaction	OLT, OLT+ME, OLT+ME+LC	Adapted version of the e-learner satisfaction survey (Likert scale)	$\alpha = .90-.95$	Overall satisfaction high. OLT+ME+LC higher than OLT+ME.

Table 2.2 Direct Training Outcome Studies – Randomized Controlled Trials

Study	Disorder and Treatment	N=	Outcome domain	Study design	Assessment measure	Reliability	Results
	Therapy		2) Knowledge		Knowledge questionnaire (MCQ)	Not described	Increase in knowledge. No differences between groups. No competence cut-point.
			3) Skills		Role-play of applying exposure therapy (SP)	ICC=.85	Increase in skills. OLT+ME+LC better than OLT. No competence cut-point.
			4) Skills self-report and/or chart review		Application of course content (self-report)	$\alpha = .85-.90$	Reported increased use of exposure interventions post-training. No differences between training groups.
(Henggeler, Sheidow, et al., 2008)	Substance use CM	30	4) Skills self-report and/or chart review	WS, WS+	Use of cognitive-behavioral and monitoring aspects of CM (patient report and care giver report)	CBT = .89, monitoring = .99	Youth reports= increase use in WS+ more than WS only. Caregiver reports = no difference between groups.
(Herschell et al., 2009)	Child Abuse PCIT	42	1) Reaction	TM→ WS, IWS	Feedback survey (Likert scale)	Not described	Overall satisfaction high. No difference between groups.
			2) Knowledge		Knowledge questionnaire (open-ended question quiz)	ICC= .99	Increase in knowledge after TM. No differences between WS and IWS. No competence cut-point.
			3) Skills		Role-play child-directed interaction (SP)	IRR= 85% ICC= .98	Increase in skills. No difference between groups. No competence cut-point.

Table 2.2 Direct Training Outcome Studies – Randomized Controlled Trials

Study	Disorder and Treatment	N=	Outcome domain	Study design	Assessment measure	Reliability	Results
(Hubley et al., 2015)	Depression BA	46	1) Reaction	OLT, OLT placebo	Feedback survey (Likert scale)	$\alpha = .81-.94$	Overall satisfaction high. Differences between training groups not examined
			2) Knowledge		Knowledge questionnaire (written response to client vignette)	ICC=.88-.91	Increase in knowledge. OLT better than OLT placebo. No competence cut-point.
(Kolko et al., 2012)	Trauma Alternatives for Families CBT	182	1) Reaction	WS+, TAU	Feedback survey (Likert scale and 2 open ended questions)	Not described	Overall satisfaction high. Differences between training groups not examined.
			2) Knowledge		Knowledge questionnaire (MCQ)	Not described	Increase in knowledge in WS+ only, maintained at follow-up. No competence cut-point.
			4) Skills self-report and/or chart review		a) teaching processes b) abuse-specific techniques c) gen psych. Techniques (self-report)	a) $\alpha = .82$ b) $\alpha = .89$ c) $\alpha = .94$	Reported increased use of treatment interventions post-training for WS+ with a, b, and c and TAU with a and b, but not c.
(Larson et al., 2013)	Substance Use CBT	127	3) Skills	TM+, OLT+	Rating scale 2 different patients at each time point (RP)	Moderate (57%) agreement	Increase in skills, but not significantly. No difference between groups. Competence cut-point determined.

Table 2.2 Direct Training Outcome Studies – Randomized Controlled Trials

Study	Disorder and Treatment	N=	Outcome domain	Study design	Assessment measure	Reliability	Results
(Luoma et al., 2007)	Substance Use Group Drug Counselling	27	1) Reaction	WS, WS+	Treatment Acceptability Rating Scale (TARS), 6 point LS (1-6)	$r = .83$	Overall satisfaction high. Differences between training groups not examined
			4) Skills self-report and/or chart review		Adoption of procedures and strategies (self-report)	$\alpha = .92$	Reported increased use of exposure intervention, significantly higher levels of adoption in WS+.
(Martino, Ball, et al., 2011)	Substance Use MI	92	1) Reaction	WS, TT, TMV	Feedback survey (Likert scale)	Not described	Overall satisfaction high. TMV better than TT.
			3) Skills		Independent Tape Rater Scale one self-selected client tape (SP) and one role-play tape of MI (RP)	ICC= .87- .88	Increase in skills. Competence cut-point determined.
(McDonough & Marks, 2002)	Phobia/Panic Exposure Therapy	37	1) Reaction	WS, CB	Feedback survey (Likert scale)	Not described	Overall satisfaction high. WS better than CB.
			2) Knowledge		Knowledge questionnaire (MCQ)	Not described	Increase in knowledge. No differences between groups. No competence cut-point.

Table 2.2 Direct Training Outcome Studies – Randomized Controlled Trials

Study	Disorder and Treatment	N=	Outcome domain	Study design	Assessment measure	Reliability	Results
(Miller et al., 2004)	Substance Use MI	140	2) Knowledge	IWS, IWSF, IWS+, IWSF+, TMV	Helpful Responses Questionnaire	r = .93	Increase in knowledge for all but TMV group. IWS+ and IWSF+ showed further increase in scores. No competence cut-point.
			3) Skills		Motivational Interviewing Skill Coding (MISC) (RP)	ICC= .51-.99	Increase in skills for all but TMV group. The improvement was only sustained in follow-up in the IWS+ and IWSF+ groups. No competence cut-point.
(Morgenstern et al., 2001)	Substance Use CBT	29	1) Reaction	Extended WS+, ALTC	Feedback survey (Likert scale and open-ended questions)	Correlation between all satisfaction and attitudinal questions: mean r= .77	Overall satisfaction high. Differences between training groups not examined. Most frequent themes: CBT is an effective approach and valuable to learn, training broadened the clinician's skill set; supervision with videotaped sessions is helpful
			3) Skills		MATCH Video Tape Rating Scale (RP)	ICC= .80-.87	Increase in skills. Competence cut-point determined. 90% of therapists were judged to have reached at least adequate treatment skillfulness.
(Moyers et al., 2008)	Substance Use MI	116	3) Skills	IWS, IWS+, TMV	Motivational Interviewing Treatment Integrity Scale (RP)	ICC= .43-.93	Increase in across all conditions. No difference between groups. Supervision did not lead to increased gains. Competence cut-point determined.

Table 2.2 Direct Training Outcome Studies – Randomized Controlled Trials

Study	Disorder and Treatment	N=	Outcome domain	Study design	Assessment measure	Reliability	Results
(Sheidow, Donohue, Hill, Henggeler, & Ford, 2008)	Substance Use Family based CM	27	3) Skills	IWS, IWS+	All sessions audio taped and rated at random using a rating scale (RP)	Absolute agreement ks = .91–.95	Details of comparison not fully covered in the text.
(Sholomskas & Carroll, 2006)	Substance Use 12 step	26	2) Knowledge	TM plus CB, TM	Knowledge questionnaire (MCQ)	Not described	Increase in knowledge. TM plus CB better than TM alone. No competence cut-point.
			3) Skills		Yale Adherence and Competence Scale - 5 role-plays (RP)	Mean ICC= .93	Increase in 12-step skills. TM plus CB better than TM alone. Cut-score of four = score required for certifying a therapist. Competence cut-point determined.
			4) Skills self-report and/or chart review		Number of clients seen, frequency of concept use (self-report)	Not described	Therapists introduce CBT to an average of 13.3- 18.7 clients in the last three months. No differences between training conditions
(J. L. Smith et al., 2012)	Substance Use MI	97	3) Skills	IWS, IWS plus live remote supervision, IWS plus tape based supervision	Motivational Interviewing Treatment Integrity scale (SP)	ICC= .62	Increase in skills. Both forms of IWS+ better than IWS. No differences between two IWS+. Competence cut-point determined.
(Weingardt et al., 2009)	Substance Use CBT	1147 147	2) Knowledge	OLT+, OLT+ with emphasis on adherence	Knowledge assessment (MCQ)	Not described	Increase in knowledge. No differences found between the groups. No competence cut-point.

Table 2.2 Direct Training Outcome Studies – Randomized Controlled Trials

Note.

DISORDER AND TREATMENT:

AIM HI- An Individualized Mental Health Intervention for children with Autism Spectrum Disorder, **EMDR**- Eye Movement Desensitization and Reprocessing, **ESDM**- Early Start Denver Model, **PCIT**- Parent-Child Interaction Therapy, **Triple P**- Positive Parenting Program

DESIGN:

Primary abbreviations:

(⁺)= plus supervision, **IWS**=interactive workshop, **OLT**= online treatment, **TM**= training manual, **WS**= workshop

Other abbreviations:

$\frac{1}{2}$ **WS**= 3hrs, **ALTC**= alternative control workshop, **CB**= computer based, **CTT** =context tailored training with role-play and feedback, **IWSF**= interactive workshop plus feedback, **OLT+ME**= online treatment plus motivational enhancement, **OLT+ME+LC**= online treatment plus motivational enhancement plus learning community, **TAU**= training as usual, **TMV**= training manual plus training videos, **WLC**= waitlist control,

Assessment Measure: **MATCH**- matching alcoholism treatments to client heterogeneity, **MCQ**- multiple choice questionnaire, **RP**- real patient, **SP**- standardized patient, , **T/F**- true or false questionnaire

Table 2.2 Direct Training Outcome Studies – Randomized Controlled Trials

Table 2.3 Indirect Training Outcome Studies

Study authors	Tx	N=	Training type	Symptom/problem-related outcomes (method)	Functional outcomes (method/scale)	Treatment variables (method)	Results
(Brookman-Frazee et al., 2012)	Childhood Autism	13	IWS+	Child problem behaviors (SSIS)	Observed level of in session parent involvement	Parent/child treatment attendance	Retention was high (92% completion). Parent involvement in session was high. Decreases in overall parent-reported child problem from baseline to 5 month f/u.
(Carroll et al., 2006)	Motivational Interviewing	37	<u>RCT</u> - IWS,ALTC	Drug use (SR; UTS; BAA); psychosocial (ASI); HIV risk (HRBS); intent to change (URICA)	n/a	Days in treatment, sessions (SR; CR)	Clients receiving MI more likely to be enrolled at 28-day F/U and to have received more treatment sessions
(Cook et al., 2006)	Substance Use	4	IWS	Drug use (UTS); PTSD symptoms (PCL-M)	Quality of life (QOLI)	n/a	Improvements on all measures
(Karlin et al., 2012)	CBT for depression	182	WS+	Depression (BDI-II)	Quality of life (WHOQOL)	Therapeutic alliance (WAI-SR)	Improvements on all measures
(Karlin et al., 2013)	CBT for insomnia	94	IWS+	Insomnia (ISI) Depression (BDI-II)	Quality of life (WHOQOL)	n/a	Improvements on all measures
(Liddle et al., 2006)	Substance Use	10	Extended WS+	Drug use (SR; UTS); CBCL (caregiver and youth self-report)	Out-of-home placement after program release (SR) at 9-month follow-up	n/a	Study phase differences on self-reported and UTS drug abstinence between baseline and implementation phases, baseline and durability phases. No phase differences on frequency of use. Significant time effect for externalizing and internalizing behaviors on youth and parent report. Improvement in number of clients living at home
(Lopez et al., 2011)	CBT for depression	59	IWS	Common emotional and behavioral problems (OPSFS)	n/a	n/a	Findings suggest either no differences pre / post- or even that youth treated prior to the workshop had better outcomes.

Table 2.3 Indirect Training Outcome Studies

Study authors	Tx	N=	Training type	Symptom/problem-related outcomes (method)	Functional outcomes (method/scale)	Treatment variables (method)	Results
(Martino et al., 2008)	Motivational enhancement Therapy	37	IWS+	Drug use (SR; UTS; BAA); motivation during session (OR)	n/a	Days in treatment (SR; CR)	MET clients more likely to sustain reduced use at 12-week F/U; in-session motivation change and negative drug screens during treatment positively associated with MET therapists' adherence and competence
(McDonel et al., 1997)	Assertive Community Treatment	13	WS+	Symptoms (BPRS)	Quality of life (LSC); functioning (ILOF, GAF); vocational/legal (SR)	Medication compliance (SR checklist)	No difference by treatment condition at 1-year F/U; differences on ILOF, LSC, BPRS at 2-year follow-up
(Miller & Mount, 2001)	Motivational Interviewing	22	IWS+	In-session change in motivation (OR): change talk and resistance talk	n/a	n/a	Ratio of change to resistance talk did not improve
(Miller et al., 2004)	Motivational Interviewing	140	<u>RCT</u> -IWS, IWSF, IWS+, IWSF+, TMV	In-session change in motivation (OR)	n/a	n/a	Improvement of change talk; at follow-up, only training group with feedback and coaching maintained changes; other groups no change
(Russell et al., 2007)	EMDR for trauma	8	IWS+	Distress during EMDR (SUDS); acceptance of cognitions (VOC); PTSD (IES-R); depression (BDI)	n/a	Number of EMDR sessions (CR)	Improvement on all symptom domains
(Schoenwald, Sheidow, Letourneau, & Liao, 2003)	Contingency Management	217	IWS+	CBCL (caregiver report)	VFI caregiver report	Discharge decision (TR)	Increase in CBCL, VFI. Improved rate of discharge decisions.
(Schoenwald et al., 2009)	Contingency Management	429	IWS+	Criminal charges during post-treatment (JAC)	n/a	n/a	Decrease in youth charges per year; therapist adherence associated with reduced criminal charges
(Simons et al.,	CBT for	12	IWS+	Anxiety (BAI), depression (BDI)	n/a	n/a	Improvements in BDI and BAI scores

Table 2.3 Indirect Training Outcome Studies

Study authors	Tx	N=	Training type	Symptom/problem-related outcomes (method)	Functional outcomes (method/scale)	Treatment variables (method)	Results
2010)	Depression						
(Strosahl et al., 1998)	Acceptance and Commitment Therapy	18	WS→IWS+	Presenting problem (SR); severity, acceptance rating on 5-point LS (SR)	Coping ratings on 5-point LS (SR, TR)	Service usage, modality, diagnosis, termination (TR)	Control clients greater on acceptance post-treatment. No differences between groups on acceptance; ACT therapists' clients higher on coping; ACT therapists and clients more likely to mutually agree on termination.
(Vismara et al., 2009)	Childhood Autism	10	IWS+	Child social communication behaviors (1) functional verbal utterances, (2) imitative behaviors (3) social initiations (4) child attention	Behaviors were coded	Reduction or increase in desired behaviors	Improvement in (1), (3) and (4), not (2).
(Walser et al., 2013)	Acceptance and Commitment Therapy	391	IWS+	Depression (BDI-II)	Quality of life (WHOQOL)	Therapeutic alliance (WAI-SR), ACT specific patient self-report process measures (AAQ-II)	BDI-II scores decreased. WHOQOL, AAQ-II, and therapeutic alliance increased.

ASI- addiction severity index; **BAA**- breath alcohol assay; **BAI**- beck anxiety inventory; **BDI**- beck depression inventory; **BPRS**- brief psychiatric rating scale; **CBCL**- child behavior checklist; **CR**- caregiver report; **GAF**- global assessment of functioning; **HRBS**- HIV risk-taking behavior scale; **IES-R**- impact of events scale-revised; **ILOF**- Indiana level of functioning scale; **ISI**-insomnia severity index; **JAC**- state and county juvenile justice agencies and court; **LSC**- life satisfaction checklist; **OPSFS**- Ohio problem severity ad functioning scales; **OR**- observer rating; **PCL-M**- PTSD checklist-military version; **QOLI**- quality of life index; **SR**- self-report; **SSIS**- social skills improvement system; **SUDS**- subjective units of distress; **TR**- therapist report; **URICA**- University of Rhode island change assessment scale; **UTS**- urine toxicology screen; **VFI**- Vanderbilt functioning index; **VOC**- validity of cognitions scale; **WAI-SR**- the working alliance inventory short revised; **WHOQOL**- world health organization quality of life brief.

Table 2.3 Indirect Training Outcome Studies

4 CHAPTER FOUR

Study 2: The Effect of Training on Therapist Knowledge



Kirkpatrick's 4 Levels of Evaluation

4.1 Overview

This chapter addresses the effect of various methods of training on knowledge acquisition, the second level of training evaluation in the Kirkpatrick framework. It describes a series of studies evaluating clinician trainees' knowledge after three different forms of training: conventional workshop training, workshop training followed by expert supervision and web-centered training. As discussed previously, workshop alone training tends to be all that is available to clinicians, while workshop training followed by supervision, although rarely available, is considered the 'gold standard' for training. Web-centered training is an innovative form of training combining aspects of both workshop and supervision in a scalable form, potentially capable of reaching many more clinicians.

All of the studies report on changes to knowledge following training in CBT-E. Study 2a focuses on conventional workshop training, Study 2b assesses workshop training plus clinical supervision, Study 2c evaluates web-centered training and Study 2d reports on a preliminary comparison, in terms of knowledge acquisition, between conventional workshop training and web-centered training.

4.2 Introduction

The qualitative findings described in the previous chapter suggest that therapists enjoyed and valued training in CBT-E independent of the delivery method and also felt confident of their ability to use the treatment. However, it cannot be assumed that their positive assessment indicates that their knowledge of the treatment had actually increased, let alone that they had attained what might be thought to be the required level of knowledge (for competence) of the treatment.

Similarly, although web-centered training appeared to offer some additional benefits such as reduced cost, the ability of completing training in trainees' own time and more reporting by therapists of the intention to subsequently implement the treatment, the effect of the training on knowledge acquisition is not known.

This chapter examines training in terms of the second level of Kirkpatrick's model: trainee therapists' *learning*. It does so by measuring changes in knowledge and applied knowledge as a result of training. Knowledge, for the purposes of these studies is regarded as theoretical – knowing the details of the particular treatment (CBT-E), that is its strategies and procedures - and applied, knowing how and when to use these. When assessing knowledge, Kirkpatrick suggests that the following two steps are taken: learning objectives are specified in advance, and reliable assessments that offer meaningful measurement are used for assessment.

As discussed in Chapter Two, a number of studies have examined knowledge acquisition as an outcome of training and the number is increasing rapidly. While many of these studies (see Table 2.1 and Table 2.2) found that knowledge scores improved following training, it was argued that they needed to be interpreted with some caution. Most current studies relied on study specific measures with unknown psychometric properties and no established competence cut-points (McHugh & Barlow, 2010). Thus while such measures may have face validity reports of increased knowledge after training remain difficult to interpret.

4.3 Overall Aim

The overall aim of the studies reported in this chapter is to evaluate the effect of various methods of training in CBT-E on knowledge and applied knowledge of the treatment. Despite some overlap in methods and trainees studied, each

individual study has particular specific aims and methods and so will be presented individually to ensure clarity.

4.4 Study 2a: Workshop training

Aims of Study 2a:

- To assess and compare knowledge acquisition in a cross-sectional study of three groups of trainees pre-workshop training, post-workshop training and post-workshop plus expert supervision
- Establish in those who received training, how many were able to reach a pre-defined cut score for competence

4.4.1 Method

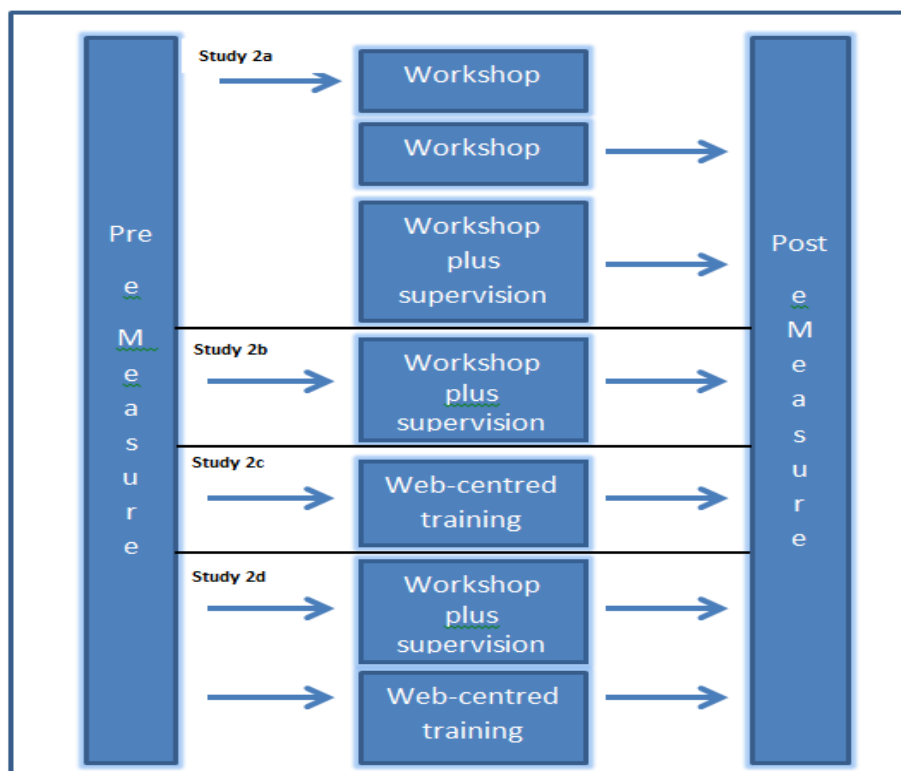
(i) Design

This study investigated the effects of training on knowledge using a cross-sectional design¹. Three groups at different points (see Figure 4.1) in training were compared in terms of their knowledge of CBT-E and the number of trainees meeting competence cut-points:

- 1) Therapists pre-workshop training
- 2) Therapists post-workshop training
- 3) Therapists post-workshop training and expert clinical supervision.

¹ The benefit of such a design allows for the identification of any associations within a population and phenomena of interest and thus frequently serves as a first step before more rigorous randomized controlled trials (Mann, 2003). As each group (pre-workshop, post-workshop, and post-workshop plus clinical supervision) were sampled at one time point, sample matching was not possible.

Figure 4.1 Design of studies 2a, 2b, 2c, and 2d



(ii) Recruitment

All therapists who provided psychological treatment for individuals with eating disorders were eligible to take part regardless of professional background. There were no exclusion criteria. Three groups of trainees were recruited at different stages of CBT-E training: prior to attendance at a standard two day CBT-E workshop, after attending such a workshop, and after having received workshop training followed by clinical supervision.

Pre-workshop training: Group one

Prior to attending a CBT-E workshop open to all therapists, therapists were contacted via email by the expert trainer. The email described the aims of the study (see Appendix 2.2). Volunteers were offered a complimentary copy of either

Cognitive Behavioral Therapy and Eating Disorders or Overcoming Binge Eating if they were prepared to complete the knowledge assessment.

Post-workshop training: Group two

Volunteers were recruited by the expert presenter during the workshop. The study was explained by the presenter and interested volunteers were encouraged to email the study administrator if they were willing to take part and complete a post-workshop knowledge assessment. As with group one, those who completed the online knowledge assessment were offered a complimentary copy of either *Cognitive Behavioral Therapy and Eating Disorders or Overcoming Binge Eating*.

Workshop training plus clinical supervision: Group three

A sub-group of therapists from the workshops were offered and agreed to additional clinical supervision as part of training initiatives run by their respective NHS services. After the completion of 16 weeks of fortnightly expert clinical supervision they completed the online knowledge measure.

(iii) The training

A detailed description of workshop training can be found in Chapter 3, (section 3.4). A brief summary of its main features can be found in Table 4.1.

Trainees in group 3 also received expert clinical supervision following their workshop training. Supervision was provided by four experts in CBT-E, all of whom had been treatment trial therapists or supervisors (three clinical psychologists and one psychiatric nurse practitioner). Supervision groups were arranged based on geographical location with the average group size of four trainees per group (range: 2-6). Supervision took place via telephone, Skype, video conference, or face to face.

Trainees were expected to treat at least one case, preferably two, while following the CBT-E treatment guide. Each supervision group received 16 sessions of supervision over an eight month period. Supervision session length varied by the number of therapists in the group. Smaller groups of two therapists received one hour of supervision every other week and larger groups received two hours every other week. Supervision was standardized following a ‘supervision manual’ (see Appendix 2.3) specially created for this purpose and involved the following: setting an agenda; reviewing patients’ progress in treatment; teaching of treatment interventions as appropriate; and summarizing and arranging the next meeting. Supervisors had regular weekly meetings to discuss and resolve any difficulties that occurred in providing supervision to trainees.

Table 4.1 Brief description of workshop training	
CBT-E Workshop Training	
<i>Style</i>	• Presented by international expert • Two day in person training • Approximately 12 hours in length • Didactic in presentation • Use of slides • 3 previously recorded clinical demonstrations
<i>Content</i>	• Core material of treatment • Appropriate for patients with BMI of >17.5 • Contains four stages consistent with treatment.

(iv) Assessment measure

The electronic measure (eMeasure)

The measure used to assess changes in knowledge for this chapter is the enhanced cognitive behavioral treatment electronic measure or CBT-E eMeasure.

The eMeasure is a treatment specific 22-item multiple choice questionnaire designed to assess knowledge of CBT-E (Cooper et al., 2015). It was developed following recommended test development procedures (described in detail here: American Educational Research Association et al., 2014; Downing & Haladyna, 2006) to ensure validity and has been demonstrated to have adequate psychometric properties (Cooper et al., 2015). Importantly the measure has a defined cut-point to benchmark competence. Knowledge is assessed by a combination of questions related to the content of the treatment, its stages and strategies, and a greater number of questions assessing how and when to apply particular interventions. These latter questions are mostly in the form of short clinical vignettes where the trainee is asked to select the most appropriate next step. The questionnaire uses a binary scoring system with trainees being awarded one mark for a correct answer and zero for an incorrect one. Individual tests take between 30 and 40 minutes to complete. Scoring is done automatically by the software. There are three equivalent versions of the eMeasure so as to allow assessment at multiple time points (further details on the development of the eMeasure can be found in Appendix 2.1).

(v) Procedure

Trainees were able to complete the eMeasure through logging in with individualized log in details (see Figure 4.2 and Appendix 2.4), to an online system hosted by ExamOnline (www.intelligentassessment.com). The system offers secure online hosting and the software to create knowledge assessments with diagrams and visual material. It is also able to score tests and store the data. The system is user friendly with few technical problems reported by users.

Figure 4.2 Log in page for eMeasure

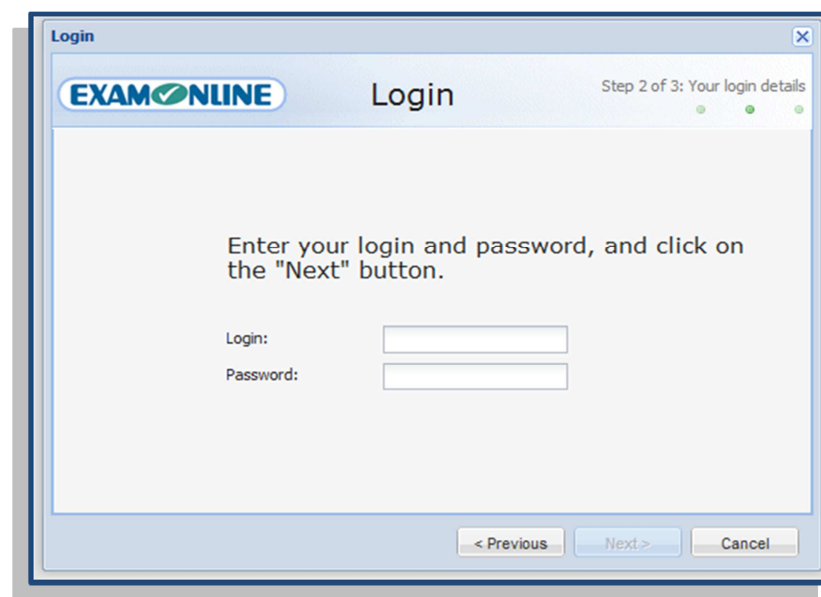


Figure 4.2 shows a screenshot of the EXAMONLINE Login page. The window is titled "Login" and features the EXAMONLINE logo in the top left corner. The main heading is "Login". A progress indicator in the top right corner shows "Step 2 of 3: Your login details". The central text instructs the user: "Enter your login and password, and click on the 'Next' button." Below this instruction are two input fields: "Login:" and "Password:". At the bottom of the window, there are three buttons: "< Previous", "Next >", and "Cancel".

Once recruited all three groups of therapists were assessed in the same way. Therapists provided their email addresses and were then sent an online consent form via email. Once consent was received a link to the electronic measure (eMeasure) was emailed to them together with brief instructions about how to access it and complete the eMeasure (see Appendix 2.4). Therapists were given two weeks to complete the assessment. The link to the measure could only be used once. Trainees were given an hour to complete the assessment, as all items can be comfortably answered in this time without any time pressure (see above).

As described, each group of therapists (pre-workshop, post-workshop and post-workshop plus expert clinical supervision) completed one eMeasure at the time they were assessed.

Demographic data was collected separately from all participants using an electronic survey hosted by Qualtrics (www.qualtrics.com). A link to the survey was emailed to all participants after completing the eMeasure.

Ethical approval for the study was obtained from the Oxford University Central Research Ethics committee: MSD/IDREC/C1/2012/43.

(vi) Data analysis plan

First the data were checked for normality and to ensure that the assumptions of the planned statistics were not violated. The data (eMeasure scores) for all three training groups, regardless of time-point, was roughly normally distributed with similar distribution seen for each group when examined separately (i.e., skew < 2.0 and kurtosis < 9.0 Schmider, Ziegler, Danay, Beyer, & Bühner, 2010). Visual inspections were conducted on histograms, normal Q-Q plots, and box plots. The assumptions of homogeneity of variance (Levene's test $F(2,336) = .082, p = .921$) were not violated and therefore significance was measured using the ANOVA test statistic with post-hoc analysis to see where any differences lay.

To assess for knowledge differences across the three groups (pre-workshop training, post-workshop and post-workshop plus clinical supervision) a one-way analysis of variance (ANOVA) was used. ANOVAs of this type are used to determine significant differences between three or more independent (unmatched) groups (Haslam & McGarty, 2014, pp. 273–277). The outcome of interest was test score on the eMeasure. Post-hoc comparisons using the Scheffe test examined pairwise difference among group means. Due to differences in group sizes the traditional Tukey HSD test was inappropriate, the Scheffe test is recommended as an alternative post-hoc test when group sizes are different (Scheffe, 1953). Effect size of the ANOVA was measured using eta squared. For eta squared a large effect size is over 0.14, a moderate effect size is 0.06 and a small effect size is 0.01 (Lakens, 2013).

To explore the number of therapists able to reach the pre-defined competence cut-point, a Pearson's chi-square test of competence by training group was used (Peck, Olsen, & Devore, 2016, pp. 602–603). Post-hoc analyses of further chi-square tests were conducted to identify the location of differences.

4.4.2 Results

(i) Sample characteristics

In total 339 therapists took part in this study: 62 (18.3%) therapists were assessed pre-workshop training; 218 (64.3%) were assessed post-workshop; and 59 (17.4%) were assessed post-workshop and supervision. 88.2% completed at least some demographic information with 67.8% providing complete information.

Demographic details are presented in Table 4.2.

Pre-workshop training: Group one

95% of the group provided at least some demographic information with complete information being obtained from 54.8%. As can be seen in Table 4.2, of the 62 (18.3% of the total, groups 1-3) therapists whose knowledge was assessed pre-training, the majority of those who provided data were female (n=49/59, 83.1%). The mean age was 43.3 (range= 26-61). Half of those who provided data (n=18/36, 50%) had been working clinically for 10 years or less. Therapists came from a variety of professional backgrounds with nursing being the most well represented (n=7/36, 19.4%), followed by clinical psychology (n=6/36, 16.6%). A large proportion identified their background as 'other' (n=17/36, 47.2%), although few supplied further details.

Workshop training: Group two

87% completed at least some demographic information with 67% providing complete information. As can be seen in Table 4.2, of the 218 (64.3% of the total,

groups 1-3) whose knowledge was assessed post-workshop training, the majority of those who provided data were female (n=168/189, 88.9%). The mean age was 43 years (range=24-68). The majority of those who provided data (n=92/171, 53.8%) had been working clinically for 10 years or less. Therapists came from nine different professional backgrounds with clinical psychology being the most well represented (n=59/171, 34.5%), followed by nursing (n=26/171, 15.2%).

Workshop training plus clinical supervision: Group three

86.4% completed at least some demographic information with 67% providing complete information. Of the 59 therapists (17.4% of the total, groups 1-3) whose knowledge was assessed post-workshop plus clinical supervision, the majority of the sample were female (n=42/51, 82.4%). The mean age was 47 (range=29-62). The majority of those who provided data (n=31/44, 70.5%) had been working clinically for 10 years or more. Therapists came from seven different professional backgrounds with clinical psychology being the most well represented (n=12/47, 25.5%), followed by nursing (n=9/47, 19.1%).

Table 4.2 Demographic details of therapists Study 2a

Demographic details ²	Pre-workshop training (N=62)	Post-workshop (N=218)	Workshop plus supervision (N=59)
Gender			
Female	49 (83.1%)	168 (88.9%)	42 (82.4%)
Male	10 (17.0%)	21 (11.1%)	9 (17.6%)
Age			
20-29	3 (6.7%)	13 (9.0%)	1 (2.6%)
30-39	15 (33.4%)	46 (31.7%)	9 (23.1%)
40-49	11 (24.4%)	44 (30.3%)	12 (30.8%)
50-59	15 (33.4%)	34 (23.4%)	16 (41.1%)
60+	1 (2.2%)	8 (5.5%)	1 (2.6%)
Years providing psychological treatment			
0 years		4 (2.3%)	
1-5 years	6 (17.6%)	47 (27.5%)	6 (13.6%)
6-10 years	12 (35.3%)	41 (24.0%)	11 (25%)
11-15 years	3 (8.8%)	25 (14.6%)	9 (20.5%)
16-20 years	7 (20.6%)	26 (15.2 %)	6 (13.6%)
21-25 years	3 (8.8%)	14(15.8%)	6 (13.6%)
25-30 years	1 (2.9%)	11 (8.2%)	4 (9.1%)
30+ years	2 (5.9%)	3 (1.8%)	2 (4.5%)
Current role			
Clinical psychologist	6 (16.6%)	59 (34.5%)	12 (25.5%)
Nurse	7 (19.4%)	26 (15.2%)	9 (19.1%)
Psychological trainee	1 (2.8%)	15 (8.8%)	2 (4.3%)
Social worker	1 (2.8%)	15 (8.8%)	4 (8.5%)
Psychiatrist	2 (5.6%)	9 (5.3%)	4 (8.5%)
Dietician	0	7 (4.1%)	0
Ed therapist	2 (5.6%)	13 (7.6%)	6 (12.8%)
Occupational therapist	0	9 (5.3%)	0
Other	17 (47.2%)	18 (10.5%)	10 (21.3%)

There were no significant differences between the two groups of therapists in terms of age ($\chi^2(10, N=229) = 8.89, p=.54$), gender ($\chi^2(2, N=229) = 2.29, p=.318$), or professional background ($\chi^2(4, N=119) = 5.20, p=.267^3$). There was a statistically

² Numbers on the table vary due to missing data and may not always add up to 100%.

³ The professional backgrounds of clinical psychologist, nurse and 'other' were included in this analysis.

significant difference between the groups for years providing psychological treatment ($\chi^2(16, N=249) = 25.40, p=.05$). With those in the workshop plus clinical supervision group having more years of clinical experience on average.

(ii) Outcome analysis

Knowledge scores

The mean eMeasure score was 6.77 ($SD=4.07$) pre-workshop; 6.64 ($SD=3.96$) post-workshop; and 12.11 ($SD=4.11$) post-workshop plus clinical supervision. Results from the ANOVA suggest no differences between pre-workshop to post-workshop ($p > .05$). There was, however, a statistically significant difference in scores following up to 16 sessions of clinical supervision, $F(2,336) = 45.18, p < .001$, with a large effect size ($\eta^2=.21$) (illustrated by Figure 4.3 and detailed in Table 4.3). Post-hoc exploration via Scheffe scores (described above) revealed significant pairwise difference between baseline (i.e., group 1) and training without supervision (i.e., group 2) as compared to training with supervision (i.e., group 3), $p < .01$. Some outliers were identified by SPSS in the samples; however, as these outliers were genuine results it was decided to retain them in the sample.

Table 4.3 Outcome of one-way ANOVA for eMeasure scores by training group

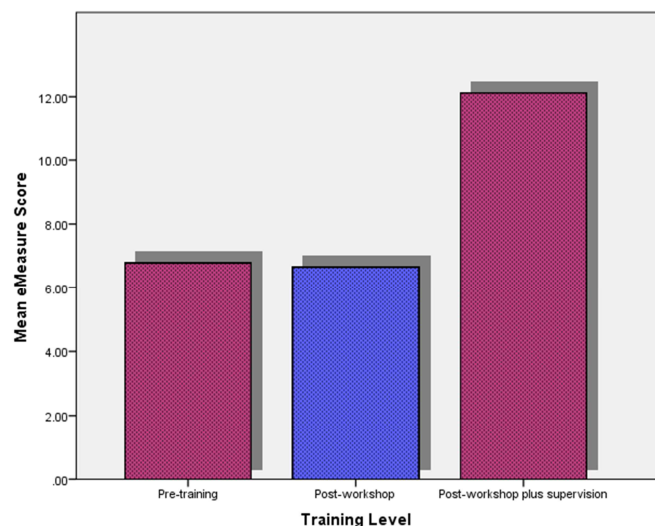
	Pre-workshop	Post-workshop	Post-workshop plus clinical supervision	F	df	p	η^2
	N=62 M (SD)	N=218 M (SD)	N=59 M (SD)				
eMeasure score	6.77 (4.07)	6.64 (3.96)	12.11 (4.11)	45.18	2, 336	<.001	.21

Trainees reaching competence cut-point

Thirty-two (55.2%) therapists in the post-workshop plus supervision group achieved a score at or above the competence cut-point (Cooper et al., 2015) on the

eMeasure as compared to 8 (14.5%) and 22 (10.4%) in pre-training and post-workshop respectively. The ability to reach the competence cut-point was significantly different between the groups ($\chi^2(2, N=339) = 67.57, p < .001$). Post-hoc analysis did not find statistically significant differences between groups one and two ($\chi^2(1, N=280) = .60, p = .438$), but did find statistically significant differences between groups 1 and 3 ($\chi^2(1, N=121) = 31.24, p < .001$), and groups 2 and 3 ($\chi^2(1, N=277) = 56.07, p < .001$).

Figure 4.3 eMeasure scores across three groups



4.4.3 Discussion

Aims of Study 2a:

- To assess and compare knowledge acquisition in a cross-sectional study of three groups of trainees pre-workshop training, post-workshop training and post-workshop plus expert supervision
- Establish in those who received training, how many were able to reach a pre-defined cut-score for competence

This cross-sectional study examined scores on a test of CBT-E knowledge across three groups of trainees at different stages of training. Trainees about to

attend a conventional training workshop, those who had completed workshop training and those who had both workshop training and expert clinical supervision (the current 'gold standard') were assessed on a treatment specific knowledge test with good psychometric properties and predefined competence cut-points.

The main strengths of the study were the use of the psychometrically sound eMeasure and the inclusion of training seeking therapists from of a variety of professional groups who treat eating disorders (a group not previously studied).

No significant differences were found in knowledge scores between those who had not yet attended a training workshop and those who had completed one. There was a significant difference, as compared to these two groups, for those whose attendance at a workshop was combined with subsequent clinical supervision. This latter finding is of particular concern. Although the current perceived 'gold standard' of training includes clinical supervision, for many this is not available and as a result workshop training is all they receive. Clearly this finding requires replication.

The finding that those who have had clinical supervision score better in terms of knowledge acquisition than those who have not had such supervision is consistent with other investigations (e.g., Beidas et al., 2012; Miller et al., 2004); however, the finding that a workshop alone was not associated with knowledge scores any higher than those who had not yet received training is inconsistent with previous research (e.g., Gega, Norman, & Marks, 2007; Herschell et al., 2009; Hawkins & Sinha, 1998). Though it should be noted that those studies used a pre-post design and tested the same therapists. There are at least three possible explanations for this inconsistency. The first relates to the nature of the material being taught; the second to the

manner of its presentation; and the third to the manner in which outcome was measured. The first explanation may be related to the content of CBT-E. CBT-E is a complex treatment comprised of numerous techniques and as a result may be more difficult to learn than others. Many trainings that have seen changes in knowledge have assessed treatments that could be argued are less complex than CBT-E. For example, Herschell et al., (2009) trained therapists in Parent-Child Interaction Therapy (PCIT). PCIT focuses on teaching parent-child interaction skills and problem solving. Skills training and problem solving are the main focus of every session and unlike CBT-E there are few other components. However, without any previous CBT-E training research we simply do not know if it is harder to learn. The second reason to explain this apparent inconsistency may be to do with how the training is presented. Some training researchers have suggested (e.g., Bennett-Levy & Padesky, 2014; Cross et al., 2011) that workshop training would be improved by following the guidelines of adult learning principles (the use of experience, observation, reflection, and planning) and have pointed to research suggesting that these techniques are correlated with increased knowledge retention in adults (Knowles, Holton III, & Swanson, 2005). More specifically the inclusion of such techniques in the teaching of components or elements of ESTs and generic CBT skills has been found to be beneficial for reported increased use of skills, knowledge and over all competence (Bennett et al., 2007; Bennett-Levy, McManus, Westling, & Fennell, 2009; Bennett-Levy & Padesky, 2014). Such findings have led many (see Table 2.1 and Table 2.2) to incorporate interactivity into their training. Though it should be noted that to date only one training study has specifically compared an interactive workshop with a regular workshop and found no differences between the two forms of training in

terms of knowledge acquisition (Herschell et al., 2009). While the CBT-E workshops provide therapists with the opportunity to observe the treatment being implemented via three clinical demonstrations, the training does not contain experiential activities such as role-play or reflection, nor does it include a planning component. It is therefore possible that a different method of presentation of the workshop material or another teaching style might result in trainees acquiring more information. Third, since very little is known about the knowledge measures used to assess outcome in most of the studies published, it is difficult to compare the results of studies.

This study has several limitations worth bearing in mind when drawing conclusions. As described a cross-sectional design was used. While this design allowed differences in knowledge scores to be observed between groups with different training experiences, it does not allow any causal attributions. So while an association was observed between receiving clinical supervision (following workshop training) and both higher knowledge scores and greater numbers of therapists meeting competence cut offs, it cannot be concluded that this was a result of the supervision they received. It is not possible to be certain that there were not systematic differences between the groups that might account for the differences observed in knowledge scores. For example it might be the case that those volunteers supported by their employers to receive supervision had higher knowledge scores than the other two groups before beginning training or that they were more motivated to receive training and therefore perhaps more diligent in learning the treatment. Additionally, despite, chi-square analysis, of those who provided information, suggesting that the three groups were fairly similar on all

demographics; they did differ on years providing psychological treatment. Trainees in the supervision group had on average more years of experiences than those in the other two groups. Finally, the present study was exploratory, recruiting as it did volunteers from those taking part in training all of whom were recruited in different ways. As a result care should be exercised in generalizing findings to all therapists undertaking training. It was not possible to ascertain whether the volunteers who agreed to take part differed from those who did not.

The results of the present study suggest that workshop training is associated with knowledge scores no different from those who have not yet received training, while only training that includes supervision from experts over an extensive time period is associated with higher scores and higher numbers of trainees reaching a defined competence cut-point. These findings should, however, be treated with caution.

4.5 Study 2b: Workshop training plus clinical supervision

Aims of Study 2b:

- In a pre-post design assess change in knowledge, following workshop training plus expert clinical supervision
- Establish how many participants were able to reach a predefined cut score for competence after 'gold standard' training

4.5.1 Method

(i) Design

This study employed a pre-post quasi-experimental design⁴ assessing CBT-E knowledge in a group of therapists receiving workshop training plus expert supervision ('gold standard' training). Therapists were assessed before attending the

⁴ Similar to experimental research but lacks random assignment and a control comparison

training workshop and again after 16 sessions of supervision over 8 months (see Table 4.1). Knowledge scores and the percentage of trainees meeting the competence cut-points were compared before and after training.

(ii) Recruitment

Therapists were offered a training workshop plus expert clinical supervision as part of a training initiative in several UK counties. All therapists who provided psychological treatment for individuals with eating disorders were included, regardless of professional background. There were no exclusion criteria. All therapists agreed to take part in the training and to complete the knowledge assessment measure before and after training. No additional incentives were provided. One group of trainees was recruited and assessed at both time points.

(iii) The training

A detailed description of workshop training plus clinical supervision can be found in section 4.4.1, of this chapter.

(iv) Assessment measure and Procedure

All therapist trainees completed the knowledge assessment before and after training. The online eMeasure (see above) was used following the same procedure described in Study 2a.

Demographic data was collected from participants using an electronic survey hosted by Qualtrics (www.qualtrics.com) as in Study 2a. However, to ensure a higher level of complete information than in the previous study, trainees were asked to complete the survey before they began the assessment and training. As before, ethical approval was obtained from the Oxford University Central Research Ethics committee: MSD/IDREC/C1/2012/43.

(v) Data analysis plan

First the pre- and post- eMeasure and change scores were checked for normality. The data for the group were normally distributed for pre-training and post-training eMeasure scores and for change scores (Schmider et al., 2010). A Shapiro-Wilks' test ($p > .05$) (Shapiro & Wilk, 1965; Razali & Wah, 2011) and a visual inspection of histograms, normal Q-Q plots, and box plots all further suggested that the data were approximately normally distributed. Therefore to investigate differences in eMeasure score from pre-training to post-training in attendees who received workshop training plus clinical supervision paired samples t-tests were used. This type of t-test is used when observations from two time points for the same or matched participants exist (Haslam & McGarty, 2014, pp. 199–213), in this case pre- and post- training eMeasure scores. Cohen's d is typically used to calculate effect sizes for t-tests and so was used for this analysis. Effect sizes (d) are considered large at .80, medium at .50, and small at .20 (Lakens, 2013).

To explore the number of therapists able to reach the pre-defined measure competence cut-point, a McNemar's chi-square test of competence by time (pre-post- training) was used (Peck et al., 2016, pp. 574–582).

4.5.2 Results

(i) Sample characteristics

Twenty therapists were put forward from four different UK counties to participate in the training. Data from one therapist was lost due to technical malfunction of the website. The final analyses were conducted on the 19 therapists who had both pre- and post-training data.

All therapists completed at least some demographic information with 79% providing complete information. The majority of the total sample was female (n=16, 84.2%) (see Table 4.4). The mean age was 41.5 (range=26-61). Most therapists were qualified as either clinical psychologists (n=9, 47.4%) or nurses (n=8, 42.1%), the remaining two (10.5%) were social workers. Just over half of the sample who provided data (n=8/15, 53.3%) had been working clinically for more than 10 years.

Table 4.4 Demographic details of therapists Study 2b

Demographic details ⁵	Workshop plus supervision (Study 2b) (N=19)
Gender	
Female	16 (84.2%)
Male	3 (15.8%)
Age	
20-29	1 (5.3%)
30-39	6 (31.6%)
40-49	5 (26.3%)
50-59	6 (31.6%)
60+	1 (5.3%)
Years providing psychological treatment	
1-5 years	1 (6.7%)
6-10 years	6 (40.0%)
11-15 years	3 (20.0%)
16-20 years	2 (13.3%)
21-25 years	1 (6.7%)
25-30 years	2 (13.3%)
30+ years	
Current role	
Clinical psychologist	9 (47.4%)
Nurse	8 (42.1%)
Social worker	2 (10.5%)
Psychiatrist	0
Family therapist	0
Occupational therapist	0
Dietician	0

⁵ Numbers on the table vary due to missing data and may not always add up to 100%.

(ii) Outcome analysis

Knowledge scores

On average therapists gained 4.42 (SD=3.08) points on the eMeasure from pre-training to post-training. Mean eMeasure total scores increased significantly pre- to post-workshop plus clinical supervision, from 6.95 (SD=3.22) to 11.37 (SD= 3.25), $t(18) = 6.26$, $p < .001$, $ES(d) = 2.87$, 95% confidence intervals (CI) (1.59, 4.15) (see Table 4.5).

Table 4.5 Outcome of paired samples t-test on eMeasure score following workshop plus clinical supervision

	Pre-training M (SD)	Post-training M (SD)	t	df	P	d	95% CI
eMeasure score (N=19)	6.95 (3.22)	11.37 (3.25)	6.26	18	<.001	2.87	1.59, 4.14

Trainees reaching competence cut-point

Ten trainees (52.6%) reached the competence cut-point (Cooper et al., 2015) on the eMeasure following training compared with one before completing training (5.2%). There was a significant difference between pre-and post- training in terms of the numbers of trainees reaching competence ($p = .004$).

4.5.3 Discussion

Aims of Study 2b:

- In a pre-post design assess change in knowledge, following workshop training plus expert clinical supervision
- Establish how many participants were able to reach a predefined cut-score for competence after 'gold standard' training

This study investigated the effect of conventional workshop training plus expert clinical supervision on CBT-E knowledge. A strength of this study was the use of the pre-post design so eliminating the concern about pre-existing differences

between groups. Trainees who received this 'gold standard' form of training showed a significant improvement from baseline gaining on average 4.42 points on the CBT-E knowledge measure. Additionally, using the pre-defined competence cut-scores (Cooper et al., 2015) over half (52.6%) of training completers were able to achieve (knowledge) competence. These findings are consistent with studies that report that workshop training plus clinical supervision in other psychological treatments improve knowledge scores (Kolko et al., 2012; Miller et al., 2004; Sholomskas et al., 2005). This is the first time such an investigation has been conducted on treatment for eating disorders.

There are three main limitations of this study: the small sample size; the lack of a control group; and the possibly biased or non-representative nature of the sample. For these reasons it is difficult to conclude that supervision is required in order for therapists to acquire sufficient knowledge for competence in CBT-E. A future study that includes a large sample of therapists randomized to receive training by a variety of different means (e.g. with and without supervision) might well be warranted. However, given that much conventional training requires scarce expertise, another but by no means mutually exclusive option, is to develop and investigate more scalable options for training.

4.6 Study 2c: Web-centered training

Aims of Study 2c:

- In a pre-post design assess change in knowledge, following a new form of online web-centered training
- Establish how many participants were able to reach a predefined cut score for competence after web-centered training
- Compare the effects on knowledge and reaching a competence cut-point of having a clinician versus or a non-clinician guide during web training

4.6.1 Method

(i) Design

This study employed a pre-post quasi-experimental design (as described above) assessing CBT-E knowledge in a group of therapist receiving web-centered training. Therapists were assessed before starting the training and again after up to 12 sessions of web-guidance over 6 months (see Figure 4.1). Knowledge scores and the percentage of trainees meeting the competence cut-points were compared before and after training.

(ii) Recruitment

Therapists were recruited through an initiative of the Irish government national health care system, Health Services Executive (HSE), to increase training in psychological therapies. All therapists who provided psychological treatment for individuals with eating disorders were included, regardless of professional background. There were no exclusion criteria.

Therapists who agreed to participate were required to complete a knowledge assessment pre-training and post-training. A certificate of completion of training was provided once therapists completed a post-training knowledge assessment. No additional incentives were offered.

(iii) The training

A detailed description web-centered training can be found in Chapter 3, section 3.5. A brief description is found in the box below, Table 4.6.

Table 4.6 Brief description of web-centered training	
CBT-E Web-centered Training	
Style	• Presented by international expert • Approximately 9 hours in length. • All material presented online • Didactic in presentation –video based lecture style • Use of PowerPoint slides • 12 brief acted clinical demonstrations • Interactive exercises • Completed alongside treating a patient • Web-guidance - up to 12 web-guidance sessions lasting 30 minutes
Content	• Core material of treatment • Appropriate for patients with BMI of >17.5 • Contains four stages consistent with treatment.

(iv) Assessment measure and Procedure

All therapist trainees completed the knowledge assessment before and after training. The online eMeasure (see above) was used following the same procedure described in Study 2a.

Demographic data was collected at the start of the training on all participants via Qualtrics software as described for Study 2b. As before, ethical approval was obtained from the Oxford University Central Research Ethics committee: MSD/IDREC/C1/2012/43.

(v) Data analysis

First the data were checked for normality and to ensure that the assumptions of the planned statistics were not violated. The data were normally distributed for pre-training and post-training eMeasure scores, and change scores (as described above). A Shapiro-Wilks' test ($p > .05$) and a visual inspection of histograms, normal Q-Q plots, and box plots all suggested that the data were approximately normally distributed for these three variables (pre-post and change scores). Therefore to determine differences in eMeasure score from pre-training to post-training, paired samples t-tests were used with Cohen's d to determine effect size (rationale described above).

Additionally to investigate the impact of guide-type (clinician or non-clinician) on test score a split-plot⁶ repeated measures analysis of variance (SPANOVA) was used. A SPANOVA is useful when examining mean differences between two or more independent groups assessed at more than one time point (Huck & McLean, 1975). The analysis incorporates both a repeated measures effect and a between groups effect. For this study the independent groups or between subjects variable was guide-type while the dependent variable or within subjects variable was time (as measured by pre-training test score and post-training test score). However, as there is some controversy about the usefulness of SPANOVAs (Dimitrov & Rumrill, 2003) an independent samples t-test based on measuring test score gain was also conducted. Independent samples t-test are used determine if there are statistically significant differences between the means of two unrelated groups (Howell, 2013, pp. 206–211).

⁶ Split-plot repeated measure analysis of variance (SPANOVA) is often referred to as a mixed-design ANOVA (Jones & Nachtsheim, 2009).

For the SPANOVA the assumptions of homogeneity of variance (Levene's test $F(1,91) = .695, p = .407$ pre-test and $F(1,91) = 1.213, p = .274$ post-test) and the equality of covariance matrices (Box's test $p = .732$) were not violated. Therefore significance was measured using Wilks' Lambda. As described above, effect sizes for the SPANOVA were determined using eta squared and Cohen's d was used for the t-test.

To explore the number of therapists able to reach the pre-defined competence cut-point, by guide-type, a Pearson's chi-square test of competence was used.

4.6.2 Results

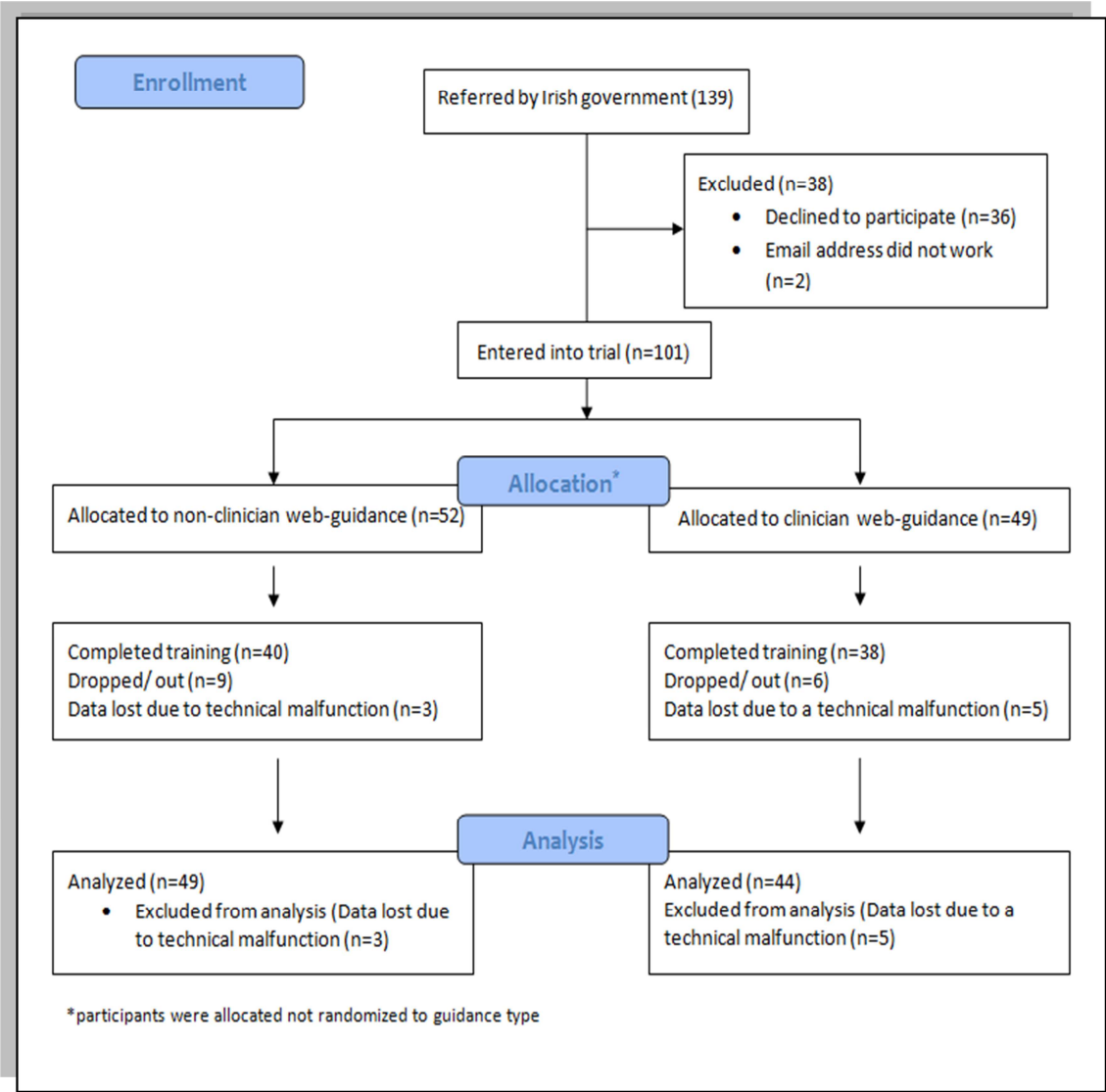
(i) Sample characteristics

One hundred and thirty nine therapists (see consort style diagram Table 4.7) were initially referred to our center by the HSE Irish government as part of a national training scheme. Of those, 101 went on to be allocated (based on geographic location⁷) to receive either non-clinician web-guidance ($N=52$) or clinician web-guidance ($N=49$) (guide-type is explained in detail in the previous chapter: section 3.4). In total, 15 (14.9%) therapists dropped out of training after completing an assessment. Data from a further 8 (7.9%) therapists were lost due to a technical fault (our server did not capture the scores on the eMeasure although we could see that the eMeasure had been completed). Data from the 15 therapists who dropped out were carried over, last observation carried forward (LOCF). LOCF data is often viewed as a conservative method of assessment and appropriate to use when sample sizes are not considerable in size (Hamer & Simpson, 2009). Therefore the

⁷ Due to a concern that clinicians at the same location might discuss the training with each other and potentially confound any results of the impact of guide-type, we decided to allocate the clinicians based on location rather than to randomize them. To overcome this difficulty and limitation, in a subsequent study, we have randomized with geographical location included as a factor.

final analyses were conducted on a total of 93 trainees, 49 in the non-clinician web-guidance group and 44 in the clinician web-guidance group. As the 15 dropouts were included in the analyses, the analyses will be considered ‘intention to train’ analyses.

Table 4.7 CONSORT style diagram for Web-centered training



All of the therapists provided complete demographic data. The majority of the sample was female (n=84, 90.3%) (see Table 4.8). The mean age was 40.1 (range=25-58). The main professions represented were clinical psychologists (n=34,

36.6%), nurses (n=33, 35.5%), social workers (n=12, 12.9%) and occupational therapists (n=8, 8.6%). More than half of the sample (n=51, 54.8%) had been working clinically for 10 years or less.

No differences were found between completers and non-completers on age ($t(91) = 1.17, p=.245$), gender ($\chi^2(1) = 1.92, p = .166$), professional background ($t(91) = 6.20, p=.537$) and years working clinically ($t(91) = 6.16, p=.539$). However, the non-completers were more likely to not have a patient while completing the training ($t(91) = 10.54, p<.001$). The average number of guidance sessions was 8.7 (SD=4.00), with the majority of therapists having at least half of the offered sessions (n=75 (80.6%). Fifty-six therapists (60.2%) used 10 or more guidance sessions. Therapists guided by clinician guides were more likely to use at least half of the sessions as compared to sessions with non-clinician guides ($\chi^2(1, N=93) = 5.64, p = .018$).

Table 4.8 Demographic details of therapists Study 2c

Demographic details	Web-centered training (Study 2c) (N=93)
Gender	
Female	84 (90.3%)
Male	9 (9.7%)
Age	
20-29	10 (10.8%)
30-39	41 (44.1%)
40-49	26 (28%)
50-59	16 (17.2%)
60+	
Years providing psychological treatment	
1-5 years	20 (21.5%)
6-10 years	31 (33.3%)
11-15 years	19 (20.4%)
16-20 years	10 (10.8%)
21-25 years	4 (4.3%)
25-30 years	5 (5.4%)
30+ years	4 (4.3%)
Current role	
Clinical psychologist	34 (36.6%)
Nurse	33 (35.5%)
Social worker	12 (12.9%)
Psychiatrist	3 (3.2%)
Family therapist	1 (1.1%)
Occupational therapist	8 (8.6%)
Dietician	2 (2.2%)

There were no significant differences between the group with regards to guide-type in terms of age ($\chi^2(31, N=93) = 27.03, p = .671$), gender ($\chi^2(1, N=93) = .03, p = .856$), professional background ($\chi^2(6, N=93) = 4.84, p = .565$), or years providing psychological treatment ($\chi^2(6, N=93) = 5.01, p = .532$).

(ii) Outcome analysis

Knowledge scores

On average therapists gained 4.96 (SD=3.89) points on the eMeasure from pre-training to post-training. Mean eMeasure total scores increased significantly pre- to post-web-centered training, from 5.31 (SD=2.76) to 10.28 (SD=4.10), $t(92) = 12.31$, $p < .001$, ES (d) = 2.55, 95% CI (2.01, 3.10) (see Table 4.9). When study dropouts are removed (N=15) training completer analysis shows that post-training scores rise approximately an additional point to 11.32 (SD=3.4).

Table 4.9 Outcome of paired samples t-test on eMeasure score following web-centered training

	Pre-training <i>M</i> (SD)	Post-training <i>M</i> (SD)	<i>t</i>	df	<i>p</i>	<i>d</i>	95% CI
eMeasure score (N=93)	5.31 (2.76)	10.28 (4.10)	12.31	92	<.001	2.55	2.01, 3.10

The independent-samples t-test based on eMeasure gain score found that average gain scores revealed no significant differences based on guide-type (clinician or non-clinician), $t(91) = 1.53$, $p = .130$, ES (d) = .32, 95% CIs (-0.09, 0.73) (see Table 4.10).

Table 4.10 Outcome of independent samples t-test of change score by guide-type following web-centered training

	Pre-post change score <i>M</i> (SD)	<i>t</i>	df	<i>p</i>	<i>d</i>	95% CI
Non-clinician guide (N=49)	4.39 (3.58)	1.53	91	.130	.32	-0.09, 0.73
Clinician guide (N=44)	5.61 (4.16)					

The SPANOVA which split the groups into non-clinician and clinician examined repeated measures scores of eMeasure (pre- and post- training) found a significant effect for time, Wilks' Lambda = 0.37, $F(1,91)=155.38$, $p<.001$, with a large effect size ($\eta^2=.63$); however, unlike the independent samples t-test above for change score, the SPANOVA also found a significant effect of guide-type, $F(1,91)=7.252$, $p=.008$, with a small effect size ($\eta^2=.07$). Importantly, the SPANOVA also found no significant interaction between time (pre-post eMeasure score) and guide-type, $F(1,91) = 2.33$, $p=.130$, with a small effect size ($\eta^2=.03$) (see Table 4.11). Figure 4.4 (below) provides some indication to explain the significant effect for guide-type. Therapists who were guided by clinicians had higher pre training scores than those guided by non-clinicians. The slopes of the lines however, do not differ significantly. The main results for the SPANOVA, interaction between time and guide, were in line with the independent-samples t-test of change score, but with the additional finding that guide-type was significant. As a result, a further investigation was required to ensure that a true interaction was not missed.

Figure 4.4 Pre/Post eMeasure score by guide-type

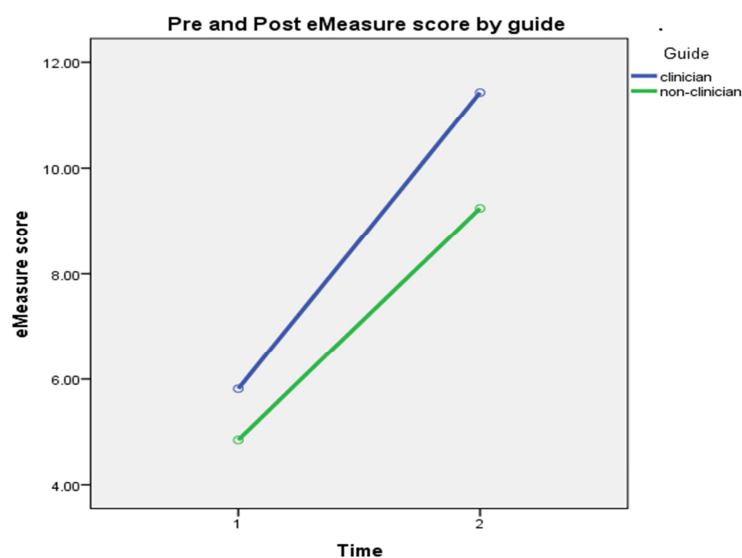


Table 4.11 Split-plot ANOVA for eMeasure scores time by guide-type

	Pre-training M (SD)	Post-training M (SD)	F	df	p	η^2
Non-clinician guide (N=49)	4.86 (2.68)	9.24 (3.69)	2.33	1,91	.130	.03
Clinician guide (N=44)	5.82 (2.80)	11.43 (4.23)				

To further explore the interaction between time and guide-type an additional analysis was conducted: an analysis of co-variance (ANCOVA). An ANCOVA is used when it is likely that an additional variable is likely to influence the outcome (Bentler & Bonett, 1980). In this case pre-score and post-score on the eMeasure were correlated ($r=.412$) indicating that pre-score was a covariate. Pre-score had a positive impact on post-score with therapists who received higher pre-scores also achieving higher post-scores. The ANCOVA was able to determine if there was a statistically significant difference between guide-type on post-score eMeasure, controlling for pre-eMeasure score. Results of the ANCOVA were similar to the independent samples t-test of change score, indicating that guide-type did not have a significant effect on post-score after controlling for pre-score, $F(9,70) = 0.58$ $p=.654$, with a very small effect size ($\eta^2=.007$). Notably, the ANCOVA suggested that pre-score on the eMeasure did have a significant effect on post-score, $F(12,13) = 2.74$ $p=.041$, with a strong effect size ($\eta^2=.72$) (Levene's test was not violated $p=.063$) (see Table 4.12). Moreover, post-hoc comparisons indicated that each pairwise difference (of guide-type) was not significant, $p=.109$, further suggesting that post-eMeasure scores did not vary significantly by guide-type.

Table 4.12 ANCOVA for post-training eMeasure scores time by guide-type

	Post-training eMeasure score* <i>M</i> (SE)	F	df	p	η^2
Non-clinician guide (N=49)	9.50 (.63)	.58	(9,70)	.654	.007
Clinician guide (N=44)	11.15 (.63)				
* Covariates appearing in the model are evaluated at the following values pre-training total score = 5.31.					

Trainees reaching competence cut-point

Thirty-four trainees (43.6% of training completers), met the pre-defined competence cut-point on the eMeasure, compared with one therapist before training. The number of those who reached competence was significantly different dependent on guide-type, $\chi^2(1, N = 93) = 6.51, p = .011$. More therapists in the clinician guided group reached competence compared to those guided by non-clinicians. This difference is likely explained by virtue of the fact that the clinician guided group started out with a higher test score (as explained above).

4.6.3 Discussion

Aims of Study 2c:

- In a pre-post design assess change in knowledge, following a new form of online web-centered training
- Establish how many participants were able to reach a predefined cut-score for competence after web-centered training
- Compare the effects on knowledge and reaching a competence cut-point of having a clinician versus or a non-clinician guide during web training

This study investigated the effect of a novel online training for CBT-E on knowledge: web-centered training. It is the first study of this form of training for therapists learning psychological treatments. Several findings of note emerged. First,

web-centered training was highly acceptable to therapists with a low drop-out rate (14.9%). Second, web-centered training was effective at improving eMeasure score, with therapists gaining nearly 5-points on average ($M=4.96$, $SD=3.89$). Additionally, using pre-defined competence cut-scores (Cooper et al., 2015) over 40% of training completers were able to achieve an eMeasure score above the competence cut-point. Third, and perhaps most striking, this change in score was found regardless of whether the therapists received guidance from an expert clinician or a non-clinician guide. The first two of these findings are largely consistent with previous research suggesting that online training formats are effective and acceptable (e.g., Dimeff et al., 2009, 2015; Harned et al., 2014; Sholomskas & Carroll, 2006), especially when combined with supervision (e.g., Herschel et al., 2010). This study is however, the first time that non-clinicians have been used to offer a form of ‘supervision’ or guidance. The present study suggests that effective web-centered training (in terms of knowledge acquisition) could be supervised by non-clinicians. If these results were to be replicated it appears that new scalable methods of training are indeed possible. This is a potentially important development, given that a lack of expert clinical supervisors has been identified by therapists as a barrier to treatment implementation (e.g., Addis, Wade, & Hatgis, 1999). Indeed, a recent training study by Dimeff and colleagues (Dimeff et al., 2015) again stressed the need for developing “effective methods of making supervision and consultation more widely available” pp.294. Web-centered training with guidance provided by non-clinicians may offer a viable alternative.

The results of this study are limited by the relatively small sample size and the lack of a control group. Furthermore therapists were not randomized to clinician

or non-clinician guides but assigned as a matter of convenience, although few differences were found between the groups. Finally, the sample of trainees studied were volunteers from a group of trainees put forward by one Health Authority and as such may not be representative of all trainees seeking training. A more robust study in which therapists are randomized to receive either guidance from a clinician or from a non-clinician and including a non-guided comparison is warranted. In the meantime while the results are promising they should be viewed as preliminary.

4.7 Study 2d: Workshop training plus clinical supervision and Web-centered training

Aims of Study 2d:

- Compare the effects of workshop plus supervision training and web-centered training both on trainees' knowledge acquisition and the number who reach a competence cut-point

4.7.1 Method

(i) Design

Using a pre-post quasi-experimental design (described above), this study compared two groups of therapists: those who completed the knowledge assessment pre- and post- conventional workshop plus expert supervision training and those who were similarly assessed pre- and post- web-centered training (see Figure 4.1).

(ii) Recruitment

Recruitment for the two groups has been described previously (see Study 2b and Study 2c).

(iii) The training

Both forms of training have been described previously (see Study 2b and Study 2c).

(iv) Assessment measure and Procedure

The assessment measure used and procedure followed for both forms of training has been described previously (see Study 2b and Study 2c).

(v) Data analysis plan

The distribution of the data for Study 2d are described above for the corresponding groups in Study 2b and Study 2c. The data (pre-post e-measure scores and change scores) for both groups together were also normally distributed.

Therefore to determine differences of the impact of training type (workshop plus supervision or web-centered) on knowledge, an independent samples t-tests and a SPANOVA were used. The independent samples t-tests measured test score gain by training group and the SPANOVA defined the between subjects variable as 'training type' while the dependent within subjects variable as 'time' (as measured by pre-training test score and post-training test score). As described above, Cohen's *d* (for the t-tests) and eta squared (for the ANOVA) were used to determine effect size.

For the SPANOVA the assumptions of homogeneity of variance (Levene's test $F(1,110) = .841, p = .361$ pretest and $F(1,110) = 1.259, p = .264$ post-test) and the equality of covariance matrices (Box's test $p = .347$) were not violated. Therefore significance was measured using Wilks' Lambda.

To explore the number of therapists able to reach the competence cut-point post-training, by training type, a Pearson's chi-square test of competence was used.

4.7.2 Results

(i) Sample characteristics

As described above for the corresponding groups in Study 2b and Study 2c (see Table 4.2 and Table 4.4).

No differences were found between therapists who completed the workshop plus supervision training and those who completed the web-centered training on age ($\chi^2(3) = 2.38, p = .497$), gender (Fisher's exact test, $p = .331$), professional background ($\chi^2(6) = 5.76, p = .451$) and years providing psychological treatment ($\chi^2(7) = 12.96, p = .073$).

(ii) Outcome analysis

Knowledge scores

The independent samples t-test of change score found no difference between the two groups (workshop or web-centered) $t(110) = .58, p = .566$, ES (d) = .15, 95% CIs (-0.35, 0.64). (Figure 4.5 illustrates this, details are in Table 4.13).

Table 4.13 Outcome of independent samples t-test of change score by training type

	Pre-post change score M (SD)	t	df	p	d	95% CI
Workshop plus clinical supervision (N=19)	4.41 (3.08)	.58	110	.566	.15	-0.35, 0.64
Web-centered training (N=93)	4.97 (3.89)					

The SPANOVA, which split the groups into those who received workshop training plus clinical supervision and those who received web-centered training examined repeated measures of eMeasure scores (pre-post), found a significant effect for time, Wilks' Lambda = 0.53, $F(1,110) = 97.84, p < .001$, with a large effect size ($\eta^2 = .47$). Additionally, the SPANOVA found no significant interaction between

time (pre-post eMeasure score) and training type $F(1,110) = 3.4, p=.07$, with a small effect size ($\eta^2=.03$) (see Table 4.14).

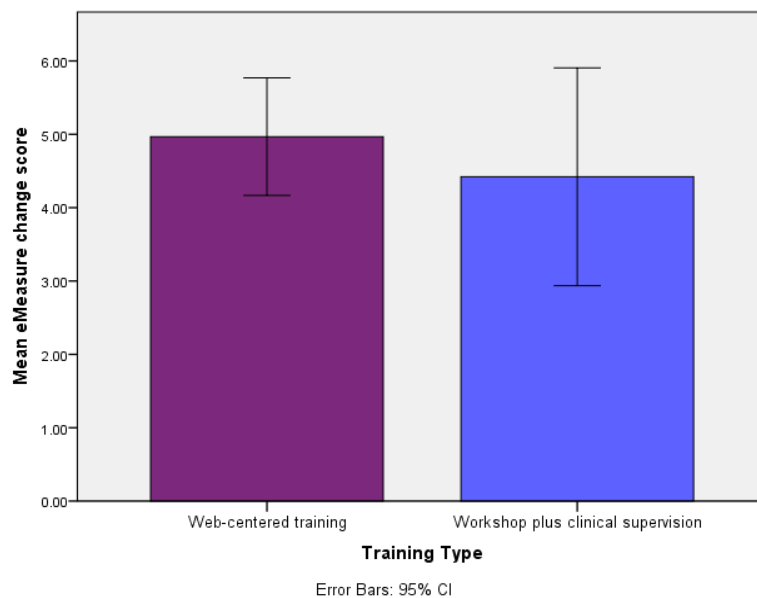
Table 4.14 Split-plot ANOVA for eMeasure scores time by training type

	Pre-training <i>M</i> (SD)	Post-training <i>M</i> (SD)	<i>F</i>	<i>df</i>	<i>p</i>	η^2
Workshop plus clinical supervision (N=19)	6.95 (3.22)	11.37 (3.25)	3.4	1, 110	.07	.03
Web-centered training (N=93)	5.31 (2.76)	10.28 (4.11)				

Trainees reaching competence cut-point

As described above, 10 trainees (52.6%) in the workshop plus clinical supervision group and 34 trainees (43.6%) in the web-centered training group, met the pre-defined competence cut-point on the eMeasure following training. There were no significant differences between training type in terms of the numbers of trainees reaching competence following training, $\chi^2(1, N = 112) = 1.71, p = .191$.

Figure 4.5 eMeasure change scores by training type



4.7.3 Discussion

Aims of Study 2d:

- Compare the effects of workshop plus supervision training and web-centered training both on trainees' knowledge acquisition and the number who reach a competence cut-point

The final study in this chapter explored the effects on knowledge of two different methods of training, both of which included some element of follow-up 'supervision' in using the treatment previously learned. This is the first comparison of these two forms of training.

The key finding of this study was that no significant differences were found between conventional 'gold standard' training (workshop plus expert clinician supervision) and a new form of web-centered training (online training with guidance from either a clinician or non-clinician). Both forms of training increased knowledge scores by an average of 4.6 points, with just over 40% in web-training and just over 50% of those in workshop training reaching the competence cut-point. This suggests that scalable online training for CBT-E offers a viable alternative to traditional workshop based training with expert clinical supervision.

The results of this comparison must be treated with caution. The comparison was performed by comparing two separate training groups, recruited in different ways and not randomized to receive one or other form of training. An additional limitation is the sample size. In the workshop plus clinical supervision group the sample size was small (n=20), limiting the interpretation of the results. A future study consisting of adequate numbers of therapists randomized to various training types is warranted.

4.8 Conclusions

Individuals with eating disorders rarely receive evidence-based treatment and even when they do, the treatments are not always delivered to a high standard. As previously discussed this is likely due, in part, to a lack of effective training for therapists. Yet, the effectiveness of training in all psychological treatments to impart the required knowledge to therapists' remains insufficiently studied. The four studies described add to the growing body of literature investigating the effectiveness of training generally and are the first to do so for CBT-E.

The main findings from the studies presented are that, with respect to CBT-E knowledge, attendance at a training workshop in CBT-E alone is not associated with any greater knowledge than non-attendance. Guidance of some form appears to be associated with greater knowledge, with those receiving expert supervision and web-based guidance both showing an increase in knowledge as a result of training. More importantly these forms of training are associated with therapists reaching competence. Notably, this study suggests that *expert* clinical supervision may not be essential and that non-clinician guides can also offer appropriate guidance. Additionally, this study highlights that web-based training offers a potentially effective, scalable alternative to traditional in person workshop training. However, the findings so far relate only to knowledge and change of knowledge. Whether this knowledge is translated into more skilled behavior on the part of the therapist still needs investigation.

5 CHAPTER FIVE

Study 3: The Effect of Training on Therapist Behavior



Kirkpatrick's 4 Levels of Evaluation

5.1 Overview

This chapter concerns the assessment of skills acquisition or the transfer of knowledge into observable skill, which is the third level of Kirkpatrick's model of training evaluation. It describes a series of studies evaluating therapists' skills after workshop plus supervision training and web-centered training. As discussed workshop training plus expert clinical supervision is considered to be the 'gold standard' of training outside of research trials. Web-centered training is a new scalable online alternative to workshop training and incorporates a form of scalable supervision termed "web-guidance".

All of the studies presented in this chapter report on changes to skill acquisition following training. Study 3a focusses on workshop training plus clinical supervision, Study 3b evaluates web-centered training and Study 3c reports on a preliminary comparison of workshop training plus clinical supervision and web-centered training.

5.2 Introduction

The findings from the previous chapter suggest that workshop training plus clinical supervision and web-centered training are equally capable of improving therapists' knowledge of CBT-E such that they reach a competence cut-point. Building on these findings, this chapter will investigate if training is also able to improve therapists' skills. As has been described, a combination of knowledge and skill are considered essential elements of competent therapy delivery (Cooper & Bailey-Straebler, 2015; Fairburn & Cooper, 2011; Muse & McManus, 2013) and hence are the main outcomes of interest when evaluating training. 'Skill' for the purposes of these studies is regarded as a therapist's ability to appropriately implement CBT-E interventions to a standard that it likely to achieve

outcomes similar to those reported in the literature from treatment trials (e.g., Byrne et al., 2011; Fairburn et al., 2015; Turner et al., 2015).

As defined by Kirkpatrick, transfer of knowledge into skill is ideally measured once a training attendee has returned to his or her workplace. This allows investigators to assess not only if skills were learned, but also if they are adequately being implemented. However, if this is not feasible or practical, measuring transfer of knowledge into skills can be assessed immediately or shortly after training. The transfer of learning thus encompasses two stages: acquiring the ability to demonstrate a particular skill and then continuing to use it for the longer-term.

As described in Chapter Two, the assessment of skill following training is a growing research area with just over 50 studies investigating this topic in some way. As discussed, the studies that assessed skill did so in a variety of ways, using role-played sessions, live recordings of actual sessions, therapist self-report, patient or care-giver report and chart review. Therapist behavior was evaluated by a variety of raters and rating scales with different degrees of reliability and validity. Despite this variability, many results tended to show improvement of skills following training (see Table 2.1 and Table 2.2). Results were less uniform when examining competence. Few studies included competence cut-points within their measures and results from those that did ranged from no therapists reaching the cut-point (e.g., Beidas et al., 2009) to nearly all therapists achieving competence in more intensive training involving expert supervision and ratings of therapy recordings (e.g., Walser et al., 2013).

5.3 Overall Aim

The main aim of the studies presented in this chapter is to evaluate the effect of various methods of training in CBT-E on skill acquisition. As in the previous chapter each study has specific aims and methods and so for clarity will be presented separately.

5.4 Study 3a: Workshop training plus clinical supervision

Aims of Study 3a:

- In a pre-post design assess change in skill, following workshop training plus expert clinical supervision
- Establish how many participants were able to reach a predefined cut-score for competence after 'gold standard' training

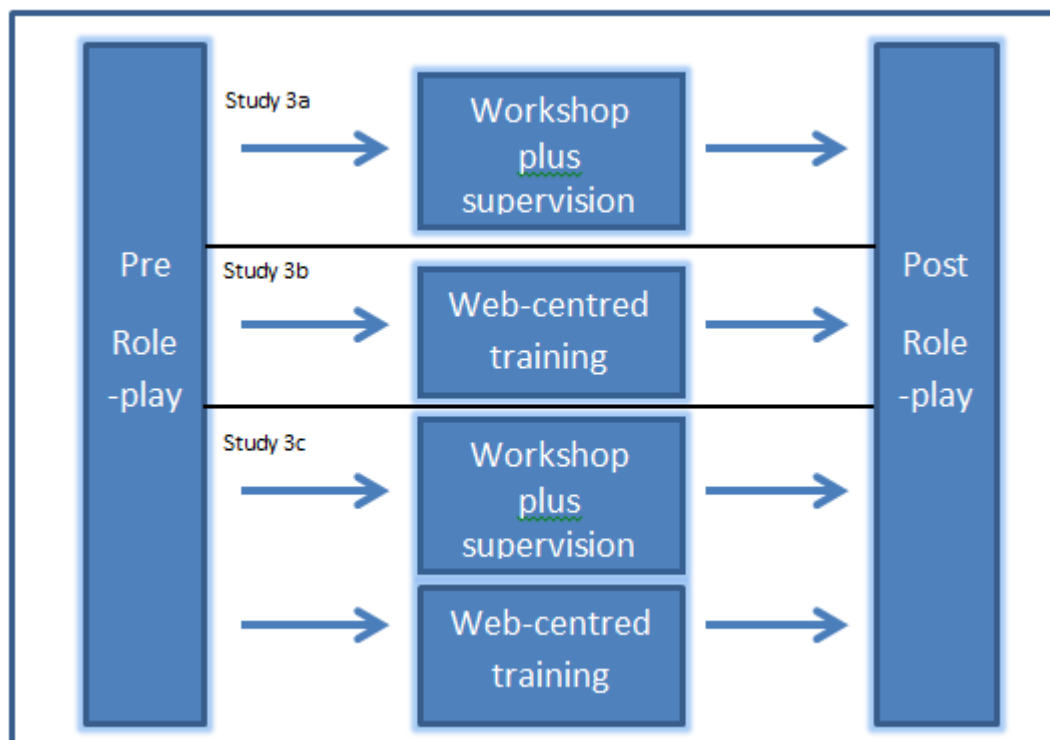
5.4.1 Method

(i) Design

This study employed a pre-post quasi-experimental design¹ assessing CBT-E skills in a group of therapists receiving workshop plus expert supervision ('gold standard' training). Therapists were assessed before attending the training workshop and again after 16 sessions of supervision over 8 months (see Figure 5.1). Skill score and the percentage of trainees meeting the competence cut-points were compared before and after training using a newly developed standardized patient role-play measure.

¹ As described in Chapter Four a quasi-experimental design is similar to experimental research but lacks random assignment and a control comparison.

Figure 5.1 Design of Studies 3a, 3b and 3c



(ii) Recruitment

Therapists were offered a training workshop plus expert clinical supervision as part of a training initiative in several UK counties. All therapists who provided psychological treatment for individuals with eating disorders were included, regardless of professional background. There were no exclusion criteria. All therapists agreed to take part in the training and to complete the skills assessment measure before and after training. No additional incentives were provided. One group of trainees was recruited and assessed before and after workshop training plus clinical supervision.

(iii) The training

A detailed description of workshop training plus clinical supervision can be found in Chapter 4, section 4.4.1.

(iv) Assessment measure

Role-play Measure

The CBT-E role-play measure used in this study was created to assess therapists' skill at implementing various CBT-E treatment techniques. At present no specific measure exists to assess CBT-E that could be used either for rating live therapy sessions or for standardized patient encounters. Existing generic therapy skills measures such as the CTS are not suitable for assessing limited domain intervention competence, the outcome of interest for establishing whether a trainee has the ability to deliver CBT-E in a manner likely to obtain expected results (Barber et al., 2007; Muse & McManus, 2013).

The decision to use role-plays with standardized patients as opposed to therapy recordings with real patients was taken by the developers in order to standardize the 'patients' trainees were asked to treat. By doing this it was possible to avoid the problems encountered with live sessions where patient difficulty cannot be controlled or indeed very successfully measured, as described in Chapter Two. Further advantages of this method included the possibility of sampling a wider range of trainee behavior with the use of several short standard acted clinical scenarios; and it allowed trainees to be more easily assessed before and after training (Fairburn & Cooper, 2011).

The content covered by the measure was agreed by a group (N=7) of international CBT-E expert clinicians. Ten interventions covering all the important stages of treatment emerged from this consultation process (see Table 5.1). With the proportion of interventions selected in each stage matching the time spent on them in therapy (Stage one- 4 interventions, Stage two- 0 interventions, Stage three- 5 interventions, Stage four- 1 intervention). Clinical scenarios were devised to cover the main aspects of each intervention focusing on two typical presentations of patients with bulimia nervosa. Patient

A represents a quiet, anxious patient whose eating is generally very rigidly controlled and whose weight is at the low end of the healthy range, while patient B is talkative, easily distracted with a fairly chaotic eating pattern and a higher (healthy) weight. The measure thus contains 20 scenarios and covers the main CBT-E interventions for two “typical” patients.

Table 5.1 Clinical scenario topics for the role-play measure

- 1) The formulation
- 2) Self-monitoring
- 3) Regular eating
- 4) Motivation
- 5) Weekly weighing
- 6) Body checking
- 7) Residual binges
- 8) Food rules
- 9) Feeling fat
- 10) Staying well

Each scenario involves a brief clinical vignette describing the patient, the stage and progress in treatment and specifies the trainee’s task, which is to spend up to 8 minutes ‘treating’ the simulated patient in the way they think most appropriate. Each role-play encounter lasts up to a total of 13 minutes (this includes five minutes of preparation time during which the therapists familiarize themselves with the clinical vignette).

Each individual trainee assessment was designed to consist of three scenarios drawn from the 20 potential scenarios. The simulated encounters were originally designed to take place in face to face settings, but have also been used over the telephone and via Skype video links. Subsequent ratings are performed by trained raters, but not expert clinicians, on audio recordings of the encounters using a specially designed rating scale.

Rating Scheme

The rating scale was devised by a group (N=5) of CBT-E experts listening to the performances of therapists of various degrees of expertise (judged globally by two expert clinicians) and extracting the core features they covered when completing the 10 interventions. These were then converted into scenario specific checklists used to guide raters in making their ratings. Consensus was also achieved about what, in the case of each scenario, constituted a complete absence of these relevant CBT-E procedures and what was involved in limited, moderate and consistent applications of the core procedures respectively. The role-plays are rated on a 7 point 'CBT-E procedure' scale, with a number of anchor points: a score of zero indicating a complete absence of CBT-procedures; 2 indicating limited application; 4 moderate application and 6 consistent application. A rating of 4 involving the application of most of the scenario specific features is taken to meet minimum competence in the intervention. In addition raters are required to make two further yes /no ratings regarding whether the trainee used a CBT-E consistent approach and whether the trainee did anything likely to undermine treatment. They make these judgements with the help of a 15 item CBT-E approach checklist.

Data on the reliability and validity of this measure exist but are yet to be published, although preliminary investigation suggests that it performs reasonably well. For example face validity is suggested by a sample of 68 therapists who completed the role-play measure and rated the role-plays as feeling "genuine" with 74.6% categorizing their performance as either moderately or closely resembling their actual behavior in therapy sessions. This rating did not differ for those who completed the role-play in-person, over the phone or over Skype.

Raters

Four research assistant coders were trained to reliability. Training was very minimal so as not to undermine potential scalability. Raters were introduced to the measure and given a brief summary of the core interventions to be rated. They subsequently completed ratings for (N=12) demonstration tapes. Ratings were compared with those of experts and any discrepancies were discussed and resolved. In total across the three studies presented in this chapter, 132 individual role-plays were rated. Inter-rater reliability for the role-plays was established by calculating intraclass correlation coefficients (ICC) for all four 'blind' raters (Bartko, 1966). Raters were 'blind' to the trainees' stage in training and training-type. Each role-play was rated by at least two raters. Inter-rater reliability was performed on all role-plays and was high across the four raters (mean ICC = .82, range = .76-.88).

Actors

In nearly all (88.6%) of the standardized patient scenarios the patient was played by one person familiar with the treatment and the typical presentation of patients with eating disorders (the current author). As a result no special training was involved with the patient being portrayed in accordance with instructions for each scenario. The advantage of one actor was that there was very little variation in 'patient' difficulty.

(v) Procedure

All therapists in this training were sent an online consent form via email. Once consent was received, a mutually convenient time to conduct the role-play was arranged. Role-plays were conducted before trainees attended the workshop and then repeated after the end of eight months clinical supervision. The majority of role-plays were done face to face in the trainee's own office. Two of the role-plays were conducted over Skype.

Demographic data was previously collected from participants using an electronic survey hosted by Qualtrics (www.qualtrics.com) as in Study 2b. As before, ethical approval was obtained from the Oxford University Central Research Ethics committee: MSD/IDREC/C1/2012/43.

Assessment procedure

At each time-point, three scenarios were chosen by a computer based random number generator. To begin the assessment, the actor read out the initial introductory information regarding the role-play and what would happen. Then the actor read out a brief introduction to the specific scenario and handed both role-play specific information and background information about the patient to the trainee. The trainee was then given up to five minutes to read the introductory information and familiarize themselves with any of the provided background information. At the end of the five minutes the trainee ‘treated’ the ‘patient’ on the given intervention for up to eight minutes. No feedback was given at the end of the role-play and actor read out a brief introduction to the next scenario, and later the third.

All role-plays were digitally recorded and time was kept on a timer visible to both the trainee and the actor. At a later date role-plays were rated by ‘blind’ raters described above.

(vi) Data analysis plan

First the data were checked for normality and to ensure that the assumptions of the planned statistics were not violated. The data (role-play measure scores) for the workshop plus clinical supervision group was sufficiently normally distributed for pre-training role-play and post-training role-play, post-training role-play competence scores and change scores for conducting the analysis (i.e., as described, skew < 2.0 and kurtosis < 9.0 Schmider, Ziegler, Danay, Beyer, & Bühner, 2010). A Shapiro-Wilks’ test ($p > .05$) and a visual inspection of

histograms, normal Q-Q plots, and box plots further suggested that the data were approximately normally distributed. Therefore to investigate differences in role-play measure score pre-to post- training, paired samples t-tests were used. As described in Chapter 4, a paired samples t-test can be used when observations from two time points for the same or matched participants exist (Haslam & McGarty, 2014, pp. 199–213). Cohen's d was used to calculate effect sizes for the t-tests, as detailed in Chapter Four.

To explore the number of therapists able to reach the pre-defined measure competence cut-point, a McNemar's chi-square test of competence by time (pre- post- training) was used (Peck et al., 2016, pp. 574–582).

5.4.2 Results

(i) Sample characteristics

Twenty-one therapists were put forward from four different UK counties to participate in the training. All therapists completed at least some demographic information with 71.4% providing complete information. The majority of the total sample was female (n=17, 81%). The mean age was 43.3 (range= 26-61). Just over half of those who provided detailed information (N=8/15, 53.3%) had been providing psychological treatment for 10 years or more. Clinical psychology (N=9/20, 45%) and nursing (N=9/20, 45%), were the most well represented professional backgrounds (see Table 5.2).

Table 5.2 Demographic details for participants in Study 3a

Demographic details	Workshop plus supervision (Study 3a) (N=21)
Gender	
Female	17 (81.0%)
Male	4 (19.0%)
Age	
20-29	1 (5.3%)
30-39	6 (31.6%)
40-49	5 (26.3%)
50-59	6 (31.6%)
60+	1 (5.3%)
Years providing psychological treatment	
1-5 years	1 (6.7%)
6-10 years	6 (40.0%)
11-15 years	3 (20.0%)
16-20 years	2 (13.3%)
21-25 years	1 (6.7%)
25-30 years	2 (13.3%)
30+ years	0
Current role	
Clinical psychologist	9 (45.0%)
Nurse	9 (45.0%)
Social worker	2 (10%)
Psychiatrist	0
Family therapist	0
Occupational therapist	0
Dietician	0

(ii) Outcome analysis

Role-play scores

On average therapists gained 0.21 (SD= 0.94) points on the role-play measure from pre-training to post-training. Mean role-play measure scores changed from pre- to post-workshop plus clinical supervision, from 3.34 (SD=1.08) to 3.55 (SD= 0.94). The change was

not significant $t(20) = 1.00$, $p = .329$, 95%, with a small effect size (ES (d) = 0.44), 95% confidence intervals (CIs) (-0.43, 1.30) (see Table 5.3).

Table 5.3 Outcome of paired samples t-test on skills assessment following workshop plus clinical supervision

	Pre-training M (SD)	Post-training M (SD)	t	df	p	d	95% CI
Role-play measure score (N=21)	3.34 (1.08)	3.55 (0.94)	1.00	20	.329	.44	-0.43, 1.30

Trainees reaching competence cut-point

Eight therapists (38.1%) reached the competence cut-point prior to training and nine therapists (42.9%) reached competence following training. This difference was not significant ($p = 1.00$).

5.4.3 Discussion

Aims of Study 3a:

- In a pre-post design assess change in skill, following workshop training plus expert clinical supervision
- Establish how many participants were able to reach a predefined cut-score for competence after 'gold standard' training

This study investigated the effect of conventional workshop training plus expert supervision in CBT-E on trainees skills using standardized role-plays to assess their performance. The findings indicate that training did not significantly improve therapists' skills or increase the number achieving competence. This is surprising given that other studies investigating training plus supervision have generally found improvements in skills following training (e.g., Beidas, Edmunds, et al., 2012; Smith et al., 2007). A possible explanation for this apparent inconsistency may be a ceiling effect. Therapists' pre-training scores on the role-play measure were high (3.34 (SD=1.08)); however, as the scale went up to a score of '6' it seems unlikely that there was not enough room to show change. Other

reasons to account for this inconsistency, as detailed in Chapter Four (Study 2a) and so only briefly described here, are likely related to the complex nature of the material being taught; the manner of its presentation; and the open-ended style of the assessment measure. As has been argued, unlike many of the other treatments which have been assessed as part of a training evaluation, CBT-E is comprised of numerous interventions which may be difficult to learn over a two day workshop, even with clinical supervision. Additionally, it may be that this training requires more interactivity based on adult learning principles. Finally, the assessment measure used in this study is far more complex than other domain specific assessments. For example, Beidas and colleagues (2012) had trainees complete one role-play assessment focused on a single intervention: an exposure task. This is rather different to the 10 interventions (times two types of patients) presented in the CBT-E role-play measure. Further testing is required to determine if the role-play measure used in this study is inherently more difficult than others currently used.

This study has several limitations. Its principal limitations were a small sample size; the lack of a control group; and possibly the biased or non-representative nature of the sample. For these reasons it is difficult to make any firm conclusions about the results or to generalize results. To better determine the effects of workshop training plus clinical supervision, future studies should randomize a sufficient number of therapists to different forms of training. Finally, the study is limited by the use of an outcome measure that does not at present have published reliability and validity data. As described earlier, there are currently no established measures to assess for skills acquisition following training in CBT-E. Therefore a newly created measure was used. Although the role-play measure was developed systematically, has promising face validity and the inter-rater reliability in this study was good, its psychometric properties are still being investigated. Additionally, while

the role-play measure is potentially able to assess trainees' competence or ability to provide CBT-E, it does not assess whether they actually do provide it to real world patients in their everyday practice nor does it provide information about longer term implementation of CBT-E.

5.5 Study 3b: Web-centered training

Aims of Study 3b:

- In a pre-post design assess change in skill, following a new form of online web-centered training
- Establish how many participants were able to reach a predefined cut-score for competence after web-centered training

5.5.1 Method

(i) Design

This study employed a pre-post quasi-experimental design (as described above) assessing CBT-E skills in a group of therapist receiving web-centered training. Therapists were assessed before starting the training and again after up to 12 sessions of web-guidance over 6 months (see Figure 5.1). Skills scores and the percentage of trainees meeting the competence cut-points were compared before and after training.

(ii) Recruitment

Therapists were recruited through an initiative of the Irish government national health care system, Health Services Executive (HSE), to increase training in psychological therapies (described in Chapter Four, section 4.4.1). A subgroup of therapists participating in the training volunteered to complete a role-play measure. All therapists who provided psychological treatment for individuals with eating disorders were included, regardless of professional background. There were no exclusion criteria. No incentives were offered.

(iii) The training

A detailed description of web-centered training can be found in Chapter 3, section 3.5.

(iv) Assessment measure and Procedure

All therapist trainees completed the role-play assessments before and after training. The standardized patient role-play measure described above was used (section 5.4.1).

To increase scalability of this method of assessment, all role-plays were conducted over the phone or Skype. While Skype, had already been piloted, using the telephone was also piloted with two therapists who did not view the method as problematic. Electronic versions of all necessary documents were held on ExamOnline and accessed by therapists in the same way described for the eMeasure in Chapter Four.

Demographic data were collected at the start of the training on all participants via Qualtrics software as described for Study 3a. As before, ethical approval was obtained from the Oxford University Central Research Ethics committee: MSD/IDREC/C1/2012/43.

(vii) Data analysis plan

First the data were checked for normality and to ensure that the assumptions of the planned statistics were not violated. The data for the group were sufficiently normally distributed for pre-training role-play, post-training role-play, post training role-play competence scores and change scores (as described above). A Shapiro-Wilks' test ($p > .05$) and a visual inspection of histograms, normal Q-Q plots, and box plots all suggested that the data were approximately normally distributed. Therefore, as described above, to investigate differences in role-play measure score pre-to post- training, paired samples t-tests were used. Cohen's d was used to calculate the effect size.

As above, to explore the number of therapists able to reach the pre-defined measure competence cut-point, a McNemar's chi-square test of competence by time (pre- post-training) was used (Peck et al., 2016, pp. 574–582).

5.5.2 Results

(i) Sample characteristics

Twenty-three therapists volunteered to complete a pre-training and post-training role-play. All therapists completed the demographic information. The majority of the sample was female (n=19, 82.6%). The mean age was 38.6 (range= 25-52). Nearly 70% (n=16, 69.4%) of the sample had been providing psychological treatment for over 10 years. Clinical psychology (N=13, 56.5%) and nursing (N=7, 30.4%), were the most well-represented professional backgrounds (see Table 5.4).

Table 5.4 Demographic details for participants in Study 3b

Demographic details	Web-centered training (Study 3b) (N=23)
Gender	
Female	19 (82.6%)
Male	4 (17.4%)
Age	
20-29	5 (21.7%)
30-39	8 (34.8%)
40-49	7 (30.4%)
50-59	3 (13.0%)
60+	
Years providing psychological treatment	
1-5 years	1 (4.3%)
6-10 years	6 (26.1%)
11-15 years	3 (13.0%)
16-20 years	5 (21.7%)
21-25 years	5 (21.7%)
25-30 years	0
30+ years	3 (13.0%)
Current role	
Clinical psychologist	13 (56.5%)
Nurse	7 (30.4%)
Social worker	1 (4.3%)
Psychiatrist	0
Family therapist	0
Occupational therapist	2 (8.7%)
Dietician	0

(ii) Outcome analysis

Role-play scores

On average therapists gained 0.70 (SD= 1.02) points on the role-play measure from pre-training to post-training (see Table 5.5). Mean role-play measure score increased significantly pre- to post-web-centered training, from 3.22 (SD= 0.97) to 3.92 (SD= 0.93), $t(22) = 3.28$, $p = .003$, with a large effect size (ES (d) = 1.37), 95% CI (0.46, 2.28) (see Table 5.5).

Table 5.5 Outcome of paired samples t-test on role-play measure score following web-centered training

	Pre-training M (SD)	Post-training M (SD)	t	df	p	d	95% CI
Role-play measure score (N=23)	3.22 (0.97)	3.92 (0.93)	3.28	22	.003	1.37	0.46, 2.28

Trainees reaching competence cut-point

Seven therapists (30.4%) reached the competence cut-point prior to training and 14 therapists (60.9%) reached competence following training. The number of therapists reaching competence following training increased significantly ($p=.016$).

5.5.3 Discussion

Aims of Study 3b:

- In a pre-post design assess change in skill, following a new form of online web-centered training
- Establish how many participants were able to reach a predefined cut-score for competence after web-centered training

This study investigated the effect of a novel form of training in CBT-E, web-centered training, on therapists' skill. Web-centered training significantly improved scores on a role-play measure from pre- to post- training. Additionally, over 60% of therapists met the pre-defined competence cut-point on the role-play measure following training, as compared to 30% at baseline.

Both of these findings are of note. While the first result is consistent with previous research suggesting that online training formats are acceptable and effective at improving skill (e.g., Dimeff et al., 2009, 2015; Harned et al., 2014; Sholomskas & Carroll, 2006), this is the first study to show this improvement using a scalable method of supervision ('web-guidance'). It is also the first study of its kind in the field of online training for eating

disorders. In addition this study not only demonstrated significant change in trainees' skills but was also able to show that they were able to meet a predefined competence threshold.

The results of this study, however, are limited by several factors. First, there was no control group. Robust conclusions are difficult to draw as it cannot be assumed that the change in therapist skills can be attributed to the training. It is always possible that other factors in the therapist's environment may have played a role. For example additional training or peer supervision may have had an impact on a change in therapists' skills, although to our knowledge this did not happen. Second, the sample size was small (N=23), and third the therapists represented a self-selected subgroup of volunteers and so results may not generalize to all therapists who receive web-centered training in CBT-E. Finally, a further limitation is the assessment measure. As described above, the full psychometric properties of the measure are still being studied. Also, a measure of this kind can only serve as a proxy for how a therapist might behave in routine clinical setting.

Despite the limitations of this study, the preliminary support for web-centered training is encouraging and suggests that further study within an adequately powered randomized controlled study is warranted.

5.6 Study 3c: Workshop training plus clinical supervision and Web-centered training

Aims of Study 3c:

- Compare the effects of workshop plus supervision training and web-centered training both on trainees' skill acquisition and the number who reach a competence cut-point

5.6.1 Method

(i) Design

Using a pre-post quasi-experimental design (described above), this study compared two groups of therapists: those who completed a skills assessment pre- and post-conventional workshop plus expert supervision training and those who were similarly assessed pre- and post- web-centered training (see Figure 5.1).

(ii) Recruitment

Recruitment for the two groups has been described previously (see Study 3a and Study 3b).

(iii) The training

Both forms of training have been described previously (see Study 3a and Study 3b).

(iv) Assessment measure and Procedure

The assessment measure used and procedure followed for both forms of training has been described previously (see Study 3a and Study 3b).

(v) Data analysis plan

Distribution of data as described above for the corresponding groups in Study 3a and Study 3b. The data (pre-post role-play measure scores and change scores) for both groups together were also normally distributed. Therefore an independent samples t-test, and a SPANOVA were used to determine differences in the impact of training type (workshop plus clinical supervision or web-centered) on skills acquisition. The independent samples t-test examined role-play measure change score by training group. The SPANOVA defined the between subjects variable as 'training type' while the dependent within subjects variable was 'time' (as measured by pre-training role-play score and post-training role-play score).

As described in Chapter 4, independent samples t-test are used determine if there are statistically significant differences between the means of two unrelated groups (Howell,

2013, pp. 206–211). A SPANOVA is useful when examining mean differences between two or more independent groups assessed at more than one time point (Huck & McLean, 1975). The analysis incorporates both a repeated measures effect and a between groups effect. Effect sizes for the t-test used Cohen's d and eta squared for the ANOVA.

For the SPANOVA the assumptions of homogeneity of variance (Levene's test $F(1,42) = 1.12, p = .296$ pre-role-play, and $F(1,42) = 0.004, p = .948$ post-role-play and the equality of covariance matrices (Box's test $p = .917; p = 9.26$, respectively) were not violated. Therefore significance was measured using Wilks' Lambda.

To explore the number of therapists able to reach the competence cut-point, by training type, a Pearson's chi-square test of competence was used.

5.6.2 Results

(i) Sample characteristics

As described above for the corresponding groups in Study 3a and Study 3b (see Table 5.2 and Table 5.4).

No differences were found between therapists who completed the workshop plus supervision training and those who completed the web-centered training on age (Likelihood ratio, $p = .125$), gender (Fisher's exact test, $p = .596$) and professional background (Likelihood ratio, $p = .250$). The groups did differ significantly on years providing psychological treatment (Likelihood ratio, $p = .012$). Web-centered trained therapists on average were a more clinically experienced group than those in workshop plus supervision training.

(ii) Outcome analysis

Role-play scores

The independent samples t-test of role-play change score found no significant differences between the two training groups, $t(42) = 1.61$, $p = .115$, ES (d) = .15, 95% CI (-0.35, 0.64) (see Table 5.6).

Table 5.6 Outcome of independent samples t-test of role-play measure change score by training type

	Pre-post change score <i>M</i> (<i>SD</i>)	<i>t</i>	<i>df</i>	<i>p</i>	<i>d</i>	95% CI
Workshop plus clinical supervision (<i>N</i> =21)	.22 (.94)	1.61	42	.115	.15	-0.35, 0.64
Web-centered (<i>N</i> =23)	.70 (1.02)					

The SPANOVA which split the groups by training-type and then examined repeated measures of role-play scores (pre- and post-training) found a significant effect for time, Wilks' Lambda = 0.82, $F(1,42) = 9.24$, $p = .004$, with a large effect size ($\eta^2 = .18$). Additionally, the SPANOVA found no significant interaction between time (pre-post eMeasure score) and training type $F(1,42) = 2.72$, $p = .106$, with a moderate effect size ($\eta^2 = .06$) (see Table 5.7).

Table 5.7 Split-plot ANOVA for role-play measure scores time by training type

	Pre-workshop <i>M</i> (<i>SD</i>)	Post-workshop <i>M</i> (<i>SD</i>)	<i>F</i>	<i>df</i>	<i>p</i>	η^2
Workshop plus clinical supervision (<i>N</i> =21)	3.34 (1.08)	3.55 (.94)	9.24	1,42	.106	.06
Web-centered training (<i>N</i> =23)	3.22 (.97)	3.92 (.93)				

Trainees reaching competence cut-point

As described above, nine trainees (42.9%) in the workshop plus clinical supervision group and 14 trainees (60.9%) in the web-centered training group, met the pre-defined

competence cut-point on the role-play assessment following training. The number of those who reached competence was not significantly different between the two training types, $\chi^2(1, N = 44) = 1.43, p = .232$. It should be noted that as the chi-square assessed differences in competence at post-training only, and not pre-post change for both groups, interpretation of this result is difficult.

5.6.3 Discussion

Aims of Study 3c:

- Compare the effects of workshop plus supervision training and web-centered training both on trainees' skill acquisition and the number who reach a competence cut-point

The final study in this chapter offered a preliminary comparison of the effects of conventional workshop plus supervision training ('gold standard') and web-centered training on skill acquisition. The main finding of this exploratory comparison is that training does improve skill; however, this is regardless of training type. No differences in skill acquisition, as measured by role-play change score or raw score, were detected with the two forms of training. This exploratory outcome is consistent with the literature of RCTs finding no differences between online and in-person trainings (Dimeff et al., 2009, 2015) with respect to skill acquisition. This finding is particularly noteworthy as the two trainings vary dramatically in terms of scalability and cost. The workshop training relied on expert clinical supervisors trained in CBT-E of which there are few. By contrast, web-centered training included therapists guided by non-clinician guides (described in Chapter Three, section 3.4).

When considered alongside the results for Study 3a and Study 3b (above), the results here may appear inconsistent and this is likely due to a number of limitations, inherent in

exploratory studies. The quasi-experimental design means that these two groups were not randomized to training type and although significant sample differences between the groups were found on only one variable (years providing psychological treatment), therefore comparisons between the groups should only be considered preliminary. Additionally, the groups were recruited in different ways; all of those in the workshop plus supervision group were required to complete the assessment, while those in web-training were volunteers. A further limitation is the small sample size. With roughly 20 therapists in each group the sample is underpowered to detect differences. The non-significant findings and yet moderate effect sizes provide an indication that the study, as currently powered, is insufficient. This investigation might allow future research to determine the appropriate number of therapists that need to be studied so as to have adequate power to detect differences between the groups (Sullivan & Feinn, 2012). As such, future research should randomize a well-powered sample of therapists to receive workshop plus supervision or web-centered training.

5.7 Conclusions

Investigations into the impact of training in psychological treatment on therapists' skill acquisition is an emerging research area. As detailed in Chapter Two, section 2.4, while most training produces some improvements, the results are inconsistent and, as few report on competence cut-point, the results are difficult to interpret. The three studies presented here add to the growing body of literature that investigates the effectiveness of training generally and are the first to do so for CBT-E. The main finding from the studies presented here is that web-centered training is able to improve therapists' skills and ability to reach a predefined competence cut-point. Web-centered training might offer an effective, scalable alternative to traditional in person workshop training at improving therapist skill.

6 CHAPTER SIX

Conclusions and Implications

6.1 Overview

In this final chapter the findings of the three main reported studies are summarized and their conclusions are reviewed. The broader relevance of the findings are discussed and areas for further research are highlighted.

6.1.1 Summary of findings

The overarching aim of the research presented in this thesis was to investigate the effects of both current and more novel and scalable training in CBT-E. CBT-E was chosen as an example of an evidence-based psychological treatment which is not widely available to those who might benefit from it, in part due to the lack of easily accessible training. Furthermore, to date, there has been no research systematically investigating the effects of training in CBT-E. This first exploratory investigation of training was guided by a training evaluation framework commonly used in the field of business but also increasingly employed in the psychological training literature (e.g., Dimeff et al., 2015; Harned et al., 2014; Westbrook et al., 2012): the Kirkpatrick Method. Following Kirkpatrick's framework three broad areas were investigated: reaction to training, and the impact of training on knowledge, and skill acquisition. Given the exploratory nature of the work and the practical difficulties involved in systematically assessing training, to a large extent, each of these areas was assessed separately. Further investigation of the inter relationships between trainees' attitudes to training and implementation, knowledge and skill acquisition and its measurement was beyond the scope of the present investigation; as was the study of longer term implementation of evidence-based treatments in routine clinical settings.

Chapter One

The first chapter described the research-practice gap in the area of evidence-based psychological treatments, and suggested that a lack of effective and scalable training was likely a major contributing factor. While an effective method of training currently exists, this intensive method used to train trial therapists (generally comprising didactic teaching, treating several pilot patients, measures to ensure strict adherence to treatment protocols and mandatory and detailed supervision), is not available to most clinicians. There are simply not enough supervisors to provide the training and the financial burden of training of this type would be too great. It was therefore decided that the training therapists are typically offered, either workshops alone or, more rarely, workshops plus supervision, should be evaluated alongside more scalable methods that are potentially capable of reaching many more therapists. The case for investigating one particular EST, CBT-E, using the Kirkpatrick Model for training evaluation, was made.

Chapter Two

In Chapter Two the literature on current training studies of ESTs with community clinicians was reviewed. The review of studies was guided by the Kirkpatrick framework, which assessed: therapists' reactions to training; knowledge acquisition; skill acquisition; and the impact that training had on patient outcomes. The studies varied considerably in terms of study design, content of the training, number of participants, and the training outcome measures employed, making comparisons difficult. Despite these methodological problems, the results were fairly consistent. Overall the studies indicated that therapists liked training and that training appeared to improve knowledge. Results were less consistent, though still largely positive, with regards to skill acquisition and patient outcomes. However, as described the robustness of these positive findings were particularly

limited by the reliance on superficial feedback forms to assess reaction and assessment measures without adequate psychometric data. In particular, clearly defined competence cut-points to assess the extent or value of reported knowledge and skill gains were lacking.

The review highlighted the need for in-depth assessments of therapist reactions to training and the use of valid scales with well described competence cut-points to assess knowledge and skills.

Chapter Three

Chapter Three presented the first of three main studies that formed the core of this thesis. Study 1 was an exploratory study that employed formal qualitative research methods. In-depth interviews with individual therapists were conducted to assess their reactions (level one of Kirkpatrick's model) to either workshop training or web-centered training. A number of findings resulted from this investigation. Most important was that therapists found web-centered training (a scalable method of training) to be acceptable regardless of whether their web-guide was a clinician or a non-clinician. Furthermore this study revealed that although therapists had a strong belief in their ability to carry out treatment following training, they did not intend to implement the treatments as suggested. In addition, this study emphasized that therapists' definition of evidence-based treatment may differ from researchers' definition.

Chapter Four

Study 2, consisting of a number of sub-studies (2a, 2b, 2c, and 2d), assessed the next level of training proposed by the Kirkpatrick framework: learning or knowledge acquisition related to training. Theoretical and applied knowledge of CBT-E was assessed using the validated online CBT eMeasure. A number of significant findings emerged from these series of studies. Study 2a found that compared to those who had received either no training or

workshop only training, those who had received workshop training combined with expert supervision had higher scores on the knowledge measure. Study 2b specifically investigated the impact of clinical supervision combined with workshop training on changes in therapist knowledge. It found significant increases in both knowledge scores and the percentage of those meeting competence cut-points after training as compared to pre-training. Study 2c investigated knowledge and competence changes after a course of web-centered training. Three important findings resulted from this study. One, therapists found training to be valuable with less than 15% dropping out. Two, training resulted in increases in knowledge and the percentage of those who achieved pre-defined competence cut-points as compared to pre-training baseline. Of particular interest was the third finding that changes in knowledge and competence rates were not dependent on whether trainees had a clinician or non-clinician guide. Study 2d involved an exploratory comparison between workshop training plus clinical supervision and web-centered training. Results indicated that there were no differences between the methods in terms of knowledge acquisition.

Chapter Five

Study 3, consisting of sub-studies 3a, 3b and 3c, investigated whether training also resulted in skill acquisition (Level three of Kirkpatrick framework). Using a newly developed CBT-E role-play measure trainees' behavior was assessed through standardized simulated therapy encounters. Study 3a investigated pre-post skill and competence changes after a conventional workshop training plus clinical supervision. No differences were detected in skill or in the percentage of those meeting competence cut-points after training as compared to pre-training. Study 3b concerned changes in therapist skill following a course of web-centered training. Training resulted in a significant increase in both skills and in the percentage of those who achieved competence. Study 3c offered a preliminary exploration

comparing workshop training plus clinical supervision and web-centered training. Results indicated that there were no differences between the two training groups in terms of skill acquisition.

The strengths and limitations of all of the studies have been covered in each of the respective chapters.

6.1.2 Broader relevance and future directions

(i) Therapists' views about training

Generally therapists enjoyed training. Their positive reactions are consistent across a variety of different types of training, including face to face workshops, online formats and web-guidance. Yet, despite their support for evidence-based treatments and their stated confidence in their ability to use the treatments they have learned, from the outset they do not generally intend to fully implement these. Their intention to only partially implement treatment appears to be related to their views, and in some cases misconceptions, regarding patient suitability and evidence-based treatments rather than their confidence in their abilities or indeed training. Although this finding relates specifically to therapists training in CBT-E, there are reports of therapists having similar objections in other treatment areas (e.g., Gallo & Barlow, 2012; Stewart, Stirman, et al., 2012). If this is indeed a general problem, it may be possible to change intentions and even behavior by paying more attention to and addressing these concerns and misconceptions during training.

(ii) The effect of training

The exploratory studies indicated that both conventional workshop plus expert supervision training and novel web-based training, even with non-clinician guides, were associated with gains in knowledge, while the effects of training through workshop alone were less clear. This finding, which highlights the importance of supervision is in line with

previous research (Kolko et al., 2012; e.g., Miller et al., 2004). While studies of training in other areas (e.g., Dimeff et al., 2015; R. Richards et al., 2011) have reported knowledge gains after workshop training alone, it should be noted that this was not directly assessed in the studies reported in this dissertation.

Web-centered training in CBT-E was associated with pre- to post- increases in skill to a degree that therapists' reached a predefined competence level. While no other study has used this form of web-based training with its particular form of scalable 'supervision', this investigation raises an interesting question of whether this form of training might be more generally applicable in order to achieve wider scale dissemination of evidence-based treatments.

6.1.3 Future directions

The work presented in this thesis is best viewed as exploratory research. As such there are a number of areas in which the work could be expanded. And, indeed, some of these avenues for further work have already been undertaken as part of the work of the Centre for Research on Dissemination (Credo) (described below).

Firstly, regarding the qualitative studies described in Chapter Three it would be helpful to know if changes to training could deal with some of the mismatches between trainers goals and trainees' expectations. For example, trainers (in-person or online) could, as part of the training, include a section that addresses 'common concerns'. This part of training might include discussions on issues such as patient complexity – i.e., the perceived differences between patients seen in research trials and those seen in clinical practice. It might also be possible to persuade therapists not to prejudge who is likely to respond to treatment but instead only make changes to the evidence-based treatment once it has been given a good trial and proves not to be effective. Trainers might also be able to discuss the

differences between implementing entire treatment packages versus only specific components of treatment. Discussions on this topic may allow trainers to demonstrate that the latter is no longer an EST. And by so doing, therapist drift, when a clinician engages in “therapy-interfering behaviors” (Waller, 2009, p. 120) often with the intention to help their patient, may be avoided. Additionally, trainers would be able to clearly state that one training does not equate to a therapist being equipped with the skills to deliver competent therapy. Therapists should be encouraged to view training workshops without supervision as only an introduction to the topic. Such changes to training might subsequently be tested in a variety of ways, possibly both qualitatively and quantitatively with longer term follow-up studies or studies of patient outcomes.

Using the detailed information obtained in the qualitative study it has been possible to create a semi-structured interview regarding therapists’ reactions and intentions following training. We are currently testing it out on current trainees. Based on findings from the semi-structured interview we plan to create more useful feedback forms than have previously been available, to assess how therapists view training. This would allow much faster and easier data collection on a much larger scale.

Secondly, the studies on knowledge and skill acquisition associated with different forms of training are all exploratory, involving cross-sectional designs or quasi-experimental designs without control or comparison groups. Additionally, the trainees investigated in these studies were recruited by different methods and not randomized to training methods. Hence well-powered randomized trials are required. There are, however, interesting issues about the design of such studies. When evaluating new scalable methods of training non-inferiority or equivalence designs (Greene, Morland, Durkalski, & B. Frueh, 2008) may be more appropriate than superiority designs. Given that some forms of conventional training

appear to be successful in achieving acceptable levels of competence, it may be that scalable training need only be shown to be equivalent. Indeed, from a public health point of view, even if scalable training produces less good outcomes, it may still deliver net gains as more therapists can be trained because of its greater accessibility. In addition interesting questions arise as to the most appropriate comparison groups when assessing training methods. While some 'gold standard' training attempts to approximate the training provided as part of RCTs (e.g., Karlin et al., 2012; Walser et al., 2013) and therefore is akin to training which has produced the evidence-base for psychological treatments, it could be argued that since this method cannot be adapted to meet the scale of the training need, it does not form a useful comparison group when investigating how best to train therapists.

If web-centered training proves to be an effective method of training, the question of whether it can be made even more scalable arises. As a result of the preliminary positive findings on the effect of web-centered training, the group at Credo is now conducting a RCT comparing web-centered training with non-clinician guides and independent web-centered training. This study will potentially provide a replication of the results for non-clinician guides and also examines a much more scalable unguided version of the training. This study also includes a six month follow-up assessment which is vital to better understanding the longer-term impact of training.

Thirdly, further work is required in relation to the measurement of training outcome. While the role-play measure clearly requires additional psychometric testing, it is worth noting that it is a time consuming method of assessing therapist skills' and requires trained raters. Credo is in the process of investigating the relationship between the online knowledge measure (which is scalable) and the skill measure. This will not only have practical benefits in terms of measuring competence in training studies as well as for the

routine assessment of training outcome, but it may throw further light on the relationship between theoretical and practical knowledge of CBT-E and its skillful practice.

Fourthly, the studies presented in this thesis deal only with the dissemination of training. The implementation of the training, particularly as it relates to patient outcomes requires investigation.

Finally, it is important that training studies like those described in this thesis are conducted for other ESTs. The majority of training available for therapists has not been evaluated and certainly not with adequate measures. Such evaluations will serve as a first step in reducing the research-practice gap.

7 REFERENCES

- Adamo, G. (2003). Simulated and standardized patients in OSCEs: achievements and challenges 1992-2003. *Medical Teacher*, 25(3), 262–270.
doi:10.1080/0142159031000100300
- Addis, M. E., Wade, W. A., Health, B., & Hatgis, C. (1999). Barriers to dissemination of evidence-based practices: Addressing practitioners' concerns about manual-based psychotherapies. *Clinical Psychology: Science and Practice*, 6, 430-441.
doi:10.1093/clipsy.6.4.430
- Agras, W. S., Crow, S. J., Halmi, K. A., Mitchell, J. E., Wilson, G. T., & Kraemer, H. C. (2000). Outcome predictors for the cognitive behavior treatment of bulimia nervosa: data from a multisite study. *The American Journal of Psychiatry*, 157(8), 1302–1308.
doi:10.1176/appi.ajp.157.8.1302
- Agras, W. S., Walsh, B. T., Fairburn, C. G., Wilson, G. T., & Kraemer, H. C. (2000). A multicenter comparison of cognitive-behavioral therapy and interpersonal psychotherapy for bulimia nervosa. *Archives of General Psychiatry*, 57, 459–466.
doi:10.1001/archpsyc.57.5.459
- Alberts, G., & Edelstein, B. (1990). Therapist training: A critical review of skill training studies. *Clinical Psychology Review*, 10, 497–511. doi:10.1016/0272-7358(90)90094-Q
- Alvarez, K., Salas, E., & Garofano, C. M. (2004). An integrated model of training evaluation and effectiveness. *Human Resource Development Review*, 3(4), 385–416.
doi:10.1177/1534484304270820
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington DC: American Educational Research Association.
- American Psychiatric Association. (2006). Treatment of patients with eating disorders. *American Journal of Psychiatry*, 163(7 Suppl), 4–54.
- Baer, J. S., Rosengren, D. B., Dunn, C. W., Wells, E. A., Ogle, R. L., & Hartzler, B. (2004). An evaluation of workshop training in motivational interviewing for addiction and mental health clinicians. *Drug and Alcohol Dependence*, 73, 99–106.
doi:10.1016/j.drugalcdep.2003.10.001
- Baer, J. S., Wells, E. A., Rosengren, D. B., Hartzler, B., Beadnell, B., & Dunn, C. (2009). Agency context and tailored training in technology transfer: A pilot evaluation of motivational

- interviewing training for community counselors. *Journal of Substance Abuse Treatment*, 37(2), 191–202. doi:10.1016/j.jsat.2009.01.003
- Ball, S. A., Martino, S., Corvino, J., Morganstern, J., & Carroll, K. M. (2002). *Independent tape rater guide*. Unpublished Manuscript, Department of Psychiatry, Yale University, New Haven, USA.
- Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*, 24, 574–594. doi:10.1080/10447310802205776
- Barber, J. P., Sharpless, B. A., Klostermann, S., & McCarthy, K. S. (2007). Assessing intervention competence and its relation to therapy outcome: A selected review derived from the outcome literature. *Professional Psychology: Research and Practice*, 38(5), 493–500.
- Barbour, R. S. (2001). Checklists for improving rigour in qualitative research: A case of the tail wagging the dog? *British Medical Journal*, 322, 1115–1117. doi:10.1136/bmj.322.7294.1115
- Barnfield, T. V., Mathieson, F. M., & Beaumont, G. R. (2007). Assessing the development of competence during postgraduate cognitive-behavioral therapy training. *Journal of Cognitive Psychotherapy*, 21(2), 140–147. doi:10.1891/088983907780851586
- Barrows, H. (1993). An overview of the uses of standardized patients for teaching and evaluating clinical skills. *Academic Medicine*, 68, 443–451.
- Bartholomew, N. G., Joe, G. W., Rowan-Szal, G. A., & Simpson, D. D. (2007). Counselor assessments of training and adoption barriers. *Journal of Substance Abuse Treatment*, 33, 193–199. doi:10.1016/j.jsat.2007.01.005
- Bartko, J. J. (1966). The intraclass correlation coefficient as a measure of reliability. *Psychological Reports*, 19(1), 3–11. doi:10.2466/pr0.1966.19.1.3
- Becker, E. M., & Jensen-Doss, A. (2013). Therapist attitudes towards computer-based trainings. *Administration and Policy in Mental Health and Mental Health Services Research*, 41(6), 845–854. doi:10.1007/s10488-013-0522-4
- Beidas, R. S., Barmish, A. J., & Kendall, P. C. (2009). Training as usual: Can therapist behavior change after reading a manual and attending a brief workshop on cognitive behavioral therapy for youth anxiety? *Behavior Therapist*, 32(5), 97–101.
- Beidas, R. S., Edmunds, J. M., Marcus, S. C., & Kendall, P. C. (2012). Training and consultation to promote implementation of an empirically supported treatment: A randomized trial. *Psychiatric Services*, 63(7), 660–5. doi:10.1176/appi.ps.201100401

- Beidas, R. S., & Kendall, P. C. (2010). Training therapists in evidence-based practice: A critical review of studies from a systems-contextual perspective. *Clinical Psychology: Science and Practice*, 17(1), 1–30. doi:10.1111/j.1468-2850.2009.01187.x
- Beidas, R. S., Mychailyszyn, M. P., Edmunds, J. M., Khanna, M. S., Downey, M. M., & Kendall, P. C. (2012). Training school mental health providers to deliver cognitive-behavioral therapy. *School Mental Health*, 4, 197–206. doi:10.1007/s12310-012-9074-0
- Bennett, G. A., Moore, J., Vaughan, T., Rouse, L., Gibbins, J. A., Thomas, P., ... Gower, P. (2007). Strengthening motivational interviewing skills following initial training: A randomised trial of workplace-based reflective practice. *Addictive Behaviors*, 32, 2963–2975. doi:10.1016/j.addbeh.2007.06.013
- Bennett-Levy, J., Hawkins, R., Perry, H., Cromarty, P., & Mills, J. (2012). Online cognitive behavioural therapy training for therapists: Outcomes, acceptability, and impact of support. *Australian Psychologist*, 47, 174–182. doi:10.1111/j.1742-9544.2012.00089.x
- Bennett-Levy, J., McManus, F., Westling, B. E., & Fennell, M. (2009). Acquiring and refining CBT skills and competencies: Which training methods are perceived to be most effective? *Behavioural and Cognitive Psychotherapy*, 37, 571–83. doi:10.1017/S1352465809990270
- Bennett-Levy, J., & Padesky, C. A. (2014). Use it or lose it: Post-workshop reflection enhances learning and utilization of CBT skills. *Cognitive and Behavioral Practice*, 21(1), 12–19. doi:10.1016/j.cbpra.2013.05.001
- Benoliel, J. Q. (1996). Grounded theory and nursing knowledge. *Qualitative Health Research*, 6(3), 406–428. doi:0803973233
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88(3), 588–606. doi:10.1037/0033-2909.88.3.588
- Berry, K., & Haddock, G. (2008). The implementation of the NICE guidelines for schizophrenia: Barriers to the implementation of psychological interventions and recommendations for the future. *Psychology and Psychotherapy*, 81, 419–436. doi:10.1348/147608308X329540
- Binder, J. L. (1993). Is it time to improve psychotherapy training? *Clinical Psychology Review*, 13, 301–318. doi:10.1016/0272-7358(93)90015-E
- Bowen, G. A. (2008). Naturalistic inquiry and the saturation concept: A research note. *Qualitative Research*, 8(1), 137–152. doi:10.1177/1468794107085301

- Brandt, J. (1991). The hopkins verbal learning test: Development of a new memory test with six equivalent forms. *Clinical Neuropsychologist*, 5(2), 125–142. doi:10.1080/13854049108403297
- Branson, A., Shafran, R., & Myles, P. (2015). Investigating the relationship between competence and patient outcome with CBT. *Behaviour Research and Therapy*, 68, 19–26. doi:10.1016/j.brat.2015.03.002
- Brenner, A. M. (2009). Uses and limitations of simulated patients in psychiatric education. *Academic Psychiatry*, 33, 112–119. doi:10.1176/appi.ap.33.2.112
- Brookman-Frazee, L. I., Drahota, A., & Stadnick, N. (2012). Training community mental health therapists to deliver a package of evidence-based practice strategies for school-age children with autism spectrum disorders: A pilot study. *Journal of Autism and Developmental Disorders*, 42, 1651–1661. doi:10.1007/s10803-011-1406-7
- Brosan, L., Reynolds, S., & Moore, R. G. (2008). Self-evaluation of cognitive therapy performance: Do therapists know how competent they are? *Behavioural and Cognitive Psychotherapy*, 36, 581–587. doi:10.1017/S1352465808004438
- Bryson, J. M. (2004). What to do when stakeholders matter. *Public Management Review*, 6(1), 21–53. doi:10.1080/14719030410001675722
- Byrne, S. M., Fursland, A., Allen, K. L., & Watson, H. (2011). The effectiveness of enhanced cognitive behavioural therapy for eating disorders: An open trial. *Behaviour Research and Therapy*, 49, 219–26. doi:10.1016/j.brat.2011.01.006
- Cant, R. P., & Cooper, S. J. (2010). Simulation-based learning in nurse education: Systematic review. *Journal of Advanced Nursing*, 66(1), 3–15. doi:10.1111/j.1365-2648.2009.05240.x
- Carroll, K. M., Ball, S. A., Nich, C., Martino, S., Frankforter, T. L., Farentinos, C., ... Woody, G. E. (2006). Motivational interviewing to improve treatment engagement and outcome in individuals seeking treatment for substance abuse: A multisite effectiveness study. *Drug and Alcohol Dependence*, 81, 301–312. doi:10.1016/j.drugalcdep.2005.08.002
- Carroll, K. M., Nich, C., & Rounsaville, B. (1998). Utility of therapist session checklists to monitor delivery of coping skills treatment for cocaine abusers. *Psychotherapy Research*, 8, 307–320. doi:10.1080/10503309812331332407
- Chagnon, F., Houle, J., Marcoux, I., & Renaud, J. (2007). Control-group study of an intervention training program for youth suicide prevention. *Suicide & Life-Threatening Behavior*, 37(2), 135–144. doi:10.1521/suli.2007.37.2.135
- Charmaz, K. (2008). Grounded theory. In J. A. Smith (Ed.), *Qualitative psychology: A practical guide to research methods*. Thousand Oaks, CA: Sage Publications.

- Chenitz, C. W., & Swanson, J. M. (1986). *From practice to grounded theory: Qualitative research in nursing*. Menlo Park, CA: Addison-Wesley.
- Chorpita, B. F., Daleiden, E. L., & Weisz, J. R. (2005). Identifying and selecting the common elements of evidence based interventions: A distillation and matching model. *Mental Health Services Research*, 7(1), 5–20. doi:10.1007/s11020-005-1962-6
- Chu, B. C., Carpenter, A. L., Wyszynski, C. M., Conklin, P. H., & Comer, J. S. (2015). Scalable options for extended skill building following didactic training in cognitive-behavioral therapy for anxious youth: A pilot randomized trial. *Journal of Clinical Child & Adolescent Psychology*, (September), 1–10. doi:10.1080/15374416.2015.1038825
- Clark, D. M. (2012). The English improving access to psychological therapies (IAPT) program: History and progress. In R. K. McHugh & D. H. Barlow (Eds.), *Dissemination and implementation of evidence-based psychological interventions* (pp. 61–77). New York: Oxford University Press.
- Clark, D. M., Layard, R., Smithies, R., Richards, D. A., Suckling, R., & Wright, B. (2009). Improving access to psychological therapy: Initial evaluation of two UK demonstration sites. *Behaviour Research and Therapy*, 47, 910–20. doi:10.1016/j.brat.2009.07.010
- Clark, R. C., & Mayer, R. E. (2011). *E-Learning and the Science of Instruction (Third Edition)*. San Francisco: Pfeiffer, A Wiley Imprint.
- Cohen, D. J., & Crabtree, B. F. (2008). Research in health care: Controversies and recommendations. *Annals Of Family Medicine*, 6, 331–339. doi:10.1370/afm.818.INTRODUCTION
- Cook, J. M., Biyanova, T., & Coyne, J. C. (2009). Barriers to adoption of new treatments: An internet study of practicing community psychotherapists. *Administration and Policy in Mental Health and Mental Health Services Research*, 36(2), 83–90. doi:10.1007/s10488-008-0198-3
- Cook, J. M., Walser, R. D., Kane, V., Ruzek, J. I., & Woody, G. (2006). Dissemination and feasibility of a cognitive-behavioral treatment for substance use disorders and posttraumatic stress disorder in the Veterans Administration. *Journal of Psychoactive Drugs*, 38, 89–92. doi:10.1080/02791072.2006.10399831
- Cooper, Z., & Bailey-Straebler, S. (2015). Disseminating evidence-based psychological treatments for eating disorders. *Current Psychiatry Reports*, 17(3), 1–9. doi:10.1007/s11920-015-0551-7
- Cooper, Z., Doll, H., Bailey-Straebler, S., Kluczniok, D., Murphy, R., O'Connor, M. E., & Fairburn, C. G. (2015). The development of an online measure of therapist competence. *Behaviour Research and Therapy*, 64, 43–48. doi:10.1016/j.brat.2014.11.007

- Corbin, J., & Strauss, A. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (4th ed.). Thousand Oaks, CA: Sage Publications.
- Crippa, J. A. S., de Lima Osório, F., Del-Ben, C. M., Filho, A. S., da Silva Freitas, M. C., & Loureiro, S. R. (2008). Comparability between telephone and face-to-face structured clinical interview for DSM-IV in assessing social anxiety disorder. *Perspectives in Psychiatric Care*, 44(4), 241–247. doi:10.1111/j.1744-6163.2008.00183.x
- Crits-Christoph, P., Siqueland, L., Chittams, J., Barber, J. P., Beck, A. T., Frank, A., ... Woody, G. (1998). Training in cognitive, supportive-expressive, and drug counseling therapies for cocaine dependence. *Journal of Consulting and Clinical Psychology*, 66(3), 484–492. doi:10.1037//0022-006X.66.3.484
- Cross, W. F., Seaburn, D., Gibbs, D., Schmeelk-Cone, K., White, A. M., & Caine, E. D. (2011). Does practice make perfect? A randomized control trial of behavioral rehearsal on suicide prevention gatekeeper skills. *Journal of Primary Prevention*, 32, 195–211. doi:10.1007/s10935-011-0250-z
- Cucciare, M. A., Weingardt, K. R., & Villafranca, S. (2008). Using blended learning to implement evidence-based psychotherapies. *Clinical Psychology: Science and Practice*, 15, 299–307. doi:10.1111/j.1468-2850.2008.00141.x
- Dalle Grave, R., Calugi, S., Doll, H. a., & Fairburn, C. G. (2013). Enhanced cognitive behaviour therapy for adolescents with anorexia nervosa: An alternative to family therapy? *Behaviour Research and Therapy*, 51(1), R9–R12. doi:10.1016/j.brat.2012.09.008
- Decker, S. E., Jameson, M. T., & Naugle, A. E. (2011). Therapist training in empirically supported treatments: A review of evaluation methods for short- and long-term outcomes. *Administration and Policy in Mental Health*, 38, 254–86. doi:10.1007/s10488-011-0360-1
- Denzin, N. K., & Lincoln, Y. S. (1998). *Strategies of qualitative inquiry*. Thousand Oaks, CA: Sage Publications.
- DeViva, J. C. (2006). The effects of full-day and half-day workshops for health care providers in techniques for increasing resistant clients' motivation. *Professional Psychology: Research and Practice*, 37(1), 83–90. doi:10.1037/0735-7028.37.1.83
- Dimeff, L. A., Harned, M. S., Woodcock, E. A., & Skutch, J. M. (2015). Investigating bang for your training buck: A randomized controlled trial comparing three methods of training clinicians in two core strategies of dialectical behavior therapy. *Behavior Therapy*, (46), 283–295. doi:10.1016/j.beth.2015.01.001
- Dimeff, L. A., Koerner, K., Woodcock, E. A., Beadnell, B., Brown, M. Z., Skutch, J. M., ... Harned, M. S. (2009). Which training method works best? A randomized controlled trial

- comparing three methods of training clinicians in dialectical behavior therapy skills. *Behaviour Research and Therapy*, 47, 921–30. doi:10.1016/j.brat.2009.07.011
- Dimitrov, D. M., & Rumrill, P. D. (2003). Pretest-posttest designs and measurement of change. *Work*, 20, 159–165.
- Dobson, K., & Beshai, S. (2013). The theory-practice gap in cognitive behavioral therapy: Reflections and a modest proposal to bridge the gap. *Behavior Therapy*, 44(4), 559–567. doi:10.1016/j.beth.2013.03.002
- Downing, S. M., & Haladyna, T. M. (2006). *Handbook of test development*. Mahwah, NJ: Routledge.
- Ehlers, A., Gene-Cos, N., & Perrin, S. (2009). Low recognition of post-traumatic stress disorder in primary care. *London Journal of Primary Care*, (4), 36–42.
- Ekers, D. M., Dawson, M. S., & Bailey, E. (2013). Dissemination of behavioural activation for depression to mental health nurses: Training evaluation and benchmarked clinical outcomes. *Journal of Psychiatric and Mental Health Nursing*, 20, 186–192. doi:10.1111/j.1365-2850.2012.01906.x
- Ellard, K. K., Fairholme, C. P., Boisseau, C. L., Farchione, T. J., & Barlow, D. H. (2010). Unified protocol for the transdiagnostic treatment of emotional disorders: Protocol development and initial outcome data. *Cognitive and Behavioral Practice*, 17(1), 88–101. doi:10.1016/j.cbpra.2009.06.002
- Epstein, R. M., Franks, P., Fiscella, K., Shields, C. G., Meldrum, S. C., Kravitz, R. L., & Duberstein, P. R. (2005). Measuring patient-centered communication in Patient-Physician consultations: Theoretical and practical issues. *Social Science and Medicine*, 61, 1516–1528. doi:10.1016/j.socscimed.2005.02.001
- Fadden, I., & Vale, A. (1997). Implementation of family interventions in routine clinical practice following staff training programs: A major cause for concern. *Journal of Mental Health*, 6(6), 599–612.
- FAIMER. (n.d.). No Title. Retrieved June 3, 2013, from <http://www.faimer.org/>
- Fairburn, C. G. (2008). *Cognitive Behavior Therapy and Eating Disorders*. New York: Guildford Press.
- Fairburn, C. G., Agras, W. S., Walsh, B. T., Wilson, G. T., & Stice, E. (2004). Prediction of outcome in bulimia nervosa by early change in treatment. *American Journal of Psychiatry*, 161(12), 2322–2324.
- Fairburn, C. G., Bailey-Straebler, S., Basden, S., Doll, H. A., Jones, R., Murphy, R., ... Zafra Cooper. (2015). A transdiagnostic comparison of enhanced cognitive behaviour therapy

- (CBT-E) and interpersonal psychotherapy in the treatment of eating disorders. *Behaviour Research and Therapy*, 70, 64–71. doi:10.1016/j.brat.2015.04.010
- Fairburn, C. G., & Cooper, Z. (2011). Therapist competence, therapy quality, and therapist training. *Behaviour Research and Therapy*, 49, 373–8. doi:10.1016/j.brat.2011.03.005
- Fairburn, C. G., Cooper, Z., Doll, H. A., O'Connor, M. E., Bohn, K., Hawker, D. M., ... Palmer, R. L. (2009). Transdiagnostic cognitive-behavioral therapy for patients with eating disorders : a two-site trial with 60-week follow-up. *American Journal of Psychiatry*, 166, 311–319.
- Fairburn, C. G., Cooper, Z., Doll, H. A., O'Connor, M. E., Palmer, R. L., & Dalle Grave, R. (2013). Enhanced cognitive behaviour therapy for adults with anorexia nervosa: A UK-Italy study. *Behaviour Research and Therapy*, 51, R2–8. doi:10.1016/j.brat.2012.09.010
- Fairburn, C. G., Cooper, Z., & Shafran, R. (2003). Cognitive behaviour therapy for eating disorders: A “transdiagnostic” theory and treatment. *Behaviour Research and Therapy*, 41, 509–528. doi:10.1016/S0005-7967(02)00088-8
- Fairburn, C. G., Jones, R., Peveler, R. C., Carr, S. J., Solomon, R. A., Connor, M. E. O., ... Hope, R. A. (1991). Three psychological treatments for bulimia nervosa. *Archives of General Psychiatry*, 48, 463–469.
- Fairburn, C. G., & Patel, V. (2014). The global dissemination of psychological treatments: a road map for research and practice. *The American Journal of Psychiatry*, 171(5), 495–8. doi:10.1176/appi.ajp.2013.13111546
- Farchione, T. J., Fairholme, C. P., Ellard, K. K., Boisseau, C. L., Thompson-Hollands, J., Carl, J. R., ... Barlow, D. H. (2012). Unified protocol for transdiagnostic treatment of emotional disorders: A randomized controlled trial. *Behavior Therapy*, 43, 666–678. doi:10.1016/j.beth.2012.01.001
- Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis : A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods*, 5, 80–92. doi:10.1063/1.2011295
- Flick, U. (2009). *An Introduction to Qualitative Research* (4th ed.). London, UK: Sage Publications.
- Ford, J. D. (1979). Research on training counselors and clinicians. *Review of Educational Research*, 49(1), 87–130. doi:10.3102/00346543049001087
- Gallo, K. P., & Barlow, D. H. (2012). Factors involved in clinician adoption and nonadoption of evidence-based interventions in mental health. *Clinical Psychology: Science and Practice*, 19, 93–106. doi:10.1111/j.1468-2850.2012.01276.x

- Gega, L., Norman, I. J., & Marks, I. M. (2007). Computer-aided vs. tutor-delivered teaching of exposure therapy for phobia/panic: Randomized controlled trial with pre-registration nursing students. *International Journal of Nursing Studies*, 44, 397–405. doi:10.1016/j.ijnurstu.2006.02.009
- Gerhardt, U. (1986). Verstehende strukturanalyse: Die konstruktion von idealtypen als analyseschritt bei der auswertung qualitativer forschungsmaterialien. In G. Soeffner (Ed.), *Sozialstruktur und Soziale Typik* (pp. 31–83). Frankfurt, DE: Campus Verlag.
- Gherardi, S., & Turner, B. (2002). Real men don't collect soft data. In A. M. Huberman & M. B. Miles (Eds.), *The qualitative researcher's companion*. London, UK: Sage Publications.
- Giacomini, M. K., & Cook, D. J. (2000). Users' guides to the medical literature XXIII. Qualitative research in health care. *Journal of the American Medical Association*, 284(3), 357–362.
- Giesen, P., Ferwerda, R., Tijssen, R., Mokkink, H., Drijver, R., van den Bosch, W., & Grol, R. (2007). Safety of telephone triage in general practitioner cooperatives: do triage nurses correctly estimate urgency? *Quality & Safety in Health Care*, 16, 181–184. doi:10.1136/qshc.2006.018846
- Glaser, B. G. (1969). The constant comparative method of qualitative inquiry. In G. J. McCall & J. L. Simmons (Eds.), *Issues in participant observation* (pp. 216–228). Reading, MA: Addison-Wesley.
- Gläser, J., & Laudel, G. (2013). Life with and without coding: Two methods for early-stage data analysis in qualitative research aiming at causal explanations. Retrieved from <http://www.qualitative-research.net/index.php/fqs/article/view/1886/3528>
- Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report*, 8(4), 597–606.
- Greene, C., Morland, L., Durkalski, V., & B. Frueh, F. (2008). Noninferiority and equivalence designs: Issues and implications for mental health research. *Journal of Trauma Stress*, 21(5), 433–439. doi:10.1002/jts.20367.Noninferiority
- Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough?: An experiment with data saturation and variability. *Field Methods*, 18(1), 59–82. doi:10.1177/1525822X05279903
- Haladyna, T. M., Downing, S. M., & Rodriguez, C. (2014). A review of multiple-choice item-writing guidelines for classroom assessment, *Applied Measurement in Education*, 15(3), 309–333. doi:10.1207/S15324818AME1503

- Haladyna, T. M., Downing, S. M., & Rodriguez, M. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309–333.
- Hamer, R. M., & Simpson, P. M. (2009). Last observation carried forward versus mixed models in the analysis of psychiatric clinical trials. *American Journal of Psychiatry*, 166, 639–641. doi:10.1176/appi.ajp.2009.09040458
- Harned, M. S., Dimeff, L. A., Woodcock, E. A., Kelly, T., Zavertrnik, J., Contreras, I., & Danner, S. M. (2014). Exposing clinicians to exposure: A randomized controlled dissemination trial of exposure therapy for anxiety disorders. *Behavior Therapy*, 45, 731–744. doi:10.1016/j.beth.2014.04.005
- Harned, M. S., Dimeff, L. A., Woodcock, E. A., & Skutch, J. M. (2011). Overcoming barriers to disseminating exposure therapies for anxiety disorders: A pilot randomized controlled trial of training methods. *Journal of Anxiety Disorders*, 25, 155–163. doi:10.1016/j.janxdis.2010.08.015
- Hart, L. M., Granillo, M. T., Jorm, A. F., & Paxton, S. J. (2011). Unmet need for treatment in the eating disorders: a systematic review of eating disorder specific treatment seeking among community cases. *Clinical Psychology Review*, 31(5), 727–35. doi:10.1016/j.cpr.2011.03.004
- Hartzler, B., Baer, J. S., Dunn, C., Rosengren, D. B., & Wells, E. (2007). What is seen through the looking glass: The impact of training on practitioner self-rating of motivational interviewing skills. *Behavioural and Cognitive Psychotherapy*, 35, 431–445. doi:10.1017/S1352465807003712
- Harvey, A. G., & Gumport, N. B. (2015). Evidence-based psychological treatments for mental disorders: Modifiable barriers to access and possible solutions. *Behaviour Research and Therapy*, 68, 1–12. doi:10.1016/j.brat.2015.02.004
- Haslam, S. A., & McGarty, C. (2014). *Research methods and statistics in psychology*. Thousand Oaks, CA: Sage Publications.
- Hawkins, K. A., & Sinha, R. (1998). Can line clinicians master the conceptual complexities of dialectical behavior therapy? An evaluation of a State Department of Mental Health training program. *Journal of Psychiatric Research*, 32, 379–384.
- Hay, P. (2013). A systematic review of evidence for psychological treatments in eating disorders: 2005–2012. *The International Journal of Eating Disorders*, 46(5), 462–469. doi:10.1002/eat.22103
- Heise, D. (1969). Separating reliability and stability in test-retest correlation. *American Sociological Review*, 34(1), 93–101. Retrieved from <http://www.jstor.org/stable/2092790>

- Henggeler, S. W., Chapman, J. E., Rowland, M. D., Halliday-Boykins, C. A., Randall, J., Shackelford, J., & Schoenwald, S. K. (2008). Statewide adoption and initial implementation of contingency management for substance abusing adolescents. *Journal of Consulting and Clinical Psychology, 76*(4), 556–567. doi:10.1037/0022-006X.76.4.556.Statewide
- Henggeler, S. W., Sheidow, A. J., Cunningham, P. B., Donohue, B. C., & Ford, J. D. (2008). Promoting the implementation of an evidence-based intervention for adolescent marijuana abuse in community settings: Testing the use of intensive quality assurance. *Journal of Clinical Child and Adolescent Psychology, 37*, 682–689. doi:10.1080/15374410802148087
- Hepner, K. A., Hunter, S. B., Paddock, S. M., Zhou, A. J., & Watkins, K. E. (2011). Training addiction counselors to implement CBT for depression. *Administration and Policy in Mental Health, 38*(4), 313–23. doi:10.1007/s10488-011-0359-7
- Herschell, A. D., Kolko, D. J., Baumann, B. L., & Davis, A. C. (2010). The role of therapist training in the implementation of psychosocial treatments: A review and critique with recommendations. *Clinical Psychology Review, 30*, 448–466. doi:10.1016/j.cpr.2010.02.005
- Herschell, A. D., McNeil, C. B., Urquiza, A. J., McGrath, J. M., Zebell, N. M., Timmer, S. G., & Porter, A. (2009). Evaluation of a treatment manual and workshops for disseminating, parent-child interaction therapy. *Administration and Policy in Mental Health, 36*(1), 63–81. doi:10.1007/s10488-008-0194-7
- Hollon, S. D., DeRubeis, R. J., Evans, M. D., Wiemer, M. J., Garvey, M. J., Grove, W. M., & Tuason, V. B. (1992). Cognitive therapy and pharmacotherapy for depression. *Journal of Consulting and Clinical Psychology, 49*, 774–781. doi:10.1037/0022-006X.59.1.88
- Howell, D. C. (2013). *Fundamental statistics for the behavioral sciences* (8th ed.). Belmont, CA: Wadsworth Publishing.
- Hubley, S., Woodcock, E. A., Dimeff, L. A., & Dimidjian, S. (2015). Disseminating behavioural activation for depression via online training: Preliminary steps. *Behavioural and Cognitive Psychotherapy, 43*, 224–238. doi:10.1017/s1352465813000842
- Huck, S., & McLean, R. (1975). Using a repeated measures ANOVA to analyze the data from a pretest-posttest design: A potentially confusing task. *Psychological Bulletin, 82*(4), 511–518.
- Insel, T. R. (2009). Translating scientific opportunity into public health impact: A strategic plan for research on mental illness. *Archives of General Psychiatry, 66*(2), 128–133.

- Institute of Medicine. (2001). *Crossing the quality chasm: A new health system for the 21st century*. Washington DC: National Academy Press. Retrieved from <http://www.nap.edu/books/0309072808/html/>
- Institute of Medicine. (2015). *Psychosocial interventions for mental and substance use disorders: A framework for establishing evidence-based standards*. Washington DC: National Academy Press. Retrieved from <http://www.nap.edu/catalog/19013/psychosocial-interventions-for-mental-and-substance-use-disorders>.
- Jensen-Doss, A., Cusack, K. J., & de Arellano, M. A. (2008). Workshop-based training in trauma-focused CBT: An in-depth analysis of impact on provider practices. *Community Mental Health Journal*, 44, 227–244. doi:10.1007/s10597-007-9121-8
- Jensen-Doss, A., Hawley, K. M., Lopez, M., & Osterberg, L. D. (2009). Using evidence-based treatments: The experiences of youth providers working under a mandate. *Professional Psychology: Research and Practice*, 40(4), 417–424. doi:10.1037/a0014690
- Jolley, S., Garety, P., Peters, E., Fornells-Ambrojo, M., Onwumere, J., Harris, V., ... Johns, L. (2015). Opportunities and challenges in Improving Access to Psychological Therapies for people with severe mental illness (IAPT-SMI): Evaluating the first operational year of the South London and Maudsley (SLaM) demonstration site for psychosis. *Behaviour Research and Therapy*, 64, 24–30. doi:10.1016/j.brat.2014.11.006
- Jones, B., & Nachtsheim, C. J. (2009). Split-plot design: what, why, and how. *Journal of Quality Technology*, 41(4), 340–361.
- Kanter, J. W., Tsai, M., Holman, G., & Koerner, K. (2013). Preliminary data from a randomized pilot study of web-based functional analytic psychotherapy therapist training. *Psychotherapy*, 50(2), 248–55. doi:10.1037/a0029814
- Karlin, B. E., Brown, G. K., Trockel, M., Cunnig, D., Zeiss, A. M., & Taylor, C. B. (2012). National dissemination of cognitive behavioral therapy for depression in the Department of Veterans Affairs health care system: Therapist and patient-level outcomes. *Journal of Consulting and Clinical Psychology*, 80(5), 707–18. doi:10.1037/a0029328
- Karlin, B. E., Ruzek, J. I., Chard, K. M., Eftekhari, A., Monson, C. M., Hembree, E. a, ... Foa, E. B. (2010). Dissemination of evidence-based psychological treatments for posttraumatic stress disorder in the Veterans Health Administration. *Journal of Traumatic Stress*, 23(6), 663–73. doi:10.1002/jts.20588
- Karlin, B. E., Trockel, M., Taylor, C. B., Gimeno, J., & Manber, R. (2013). National dissemination of cognitive behavioral therapy for insomnia in veterans: Therapist- and patient-level outcomes. *Journal of Consulting and Clinical Psychology*, 81(5), 912–7. doi:10.1037/a0032554

- Kass, A. E., Kolko, R. P., & Wilfley, D. E. (2013). Psychological treatments for eating disorders. *Current Opinion in Psychiatry*, 26(6), 549–55. doi:10.1097/YCO.0b013e328365a30e
- Kaufman, R., Keller, J., & Watkins, R. (1995). What works and what doesn't: Evaluation beyond Kirkpatrick. *Performance and Instruction*, 35(2), 8–12.
- Kazdin, A. E. (2003). *Research Design in Clinical Psychology* (4th ed.). Needham Heights, MA: Allyn and Bacon.
- Kazdin, A. E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *The American Psychologist*, 63(3), 146–59. doi:10.1037/0003-066X.63.3.146
- Kazdin, A. E., & Blase, S. L. (2011). Rebooting Psychotherapy Research and Practice to Reduce the Burden of Mental Illness. *Perspectives on Psychological Science*, 6(1), 21–37. doi:10.1177/1745691610393527
- Kelle, U. (1995). *Computer-aided qualitative data analysis: Theory, methods and practice*. London, UK: Sage Publications.
- Kendall, P. C., Settapani, C. A., & Cummings, C. M. (2012). No need to worry: the promising future of child anxiety research. *Journal of Clinical Child and Adolescent Psychology*, 41(1), 103–15. doi:10.1080/15374416.2012.632352
- Khanna, M. S., & Kendall, P. C. (2015). Bringing technology to training: Web-based therapist training to promote the development of competent cognitive-behavioral therapists. *Cognitive and Behavioral Practice*, 22(3), 291–301. doi:10.1016/j.cbpra.2015.02.002
- Kirkpatrick, D. (1959a). Techniques for evaluating training programs. *Journal of the American Society for Training and Development*, 75(11), 3–9.
- Kirkpatrick, D. (1959b). Techniques for evaluating training programs: Part 2-learning. *Journal of the American Society for Training and Development*, 13(12), 21–26.
- Kirkpatrick, D. (1960a). Techniques for evaluating training programs: Part 3-behavior. *Journal of the American Society for Training and Development*, 14(1), 13–18.
- Kirkpatrick, D. (1960b). Techniques for evaluating training programs: Part 4-results. *Journal of the American Society for Training and Development*, 14(2), 28–32.
- Kirkpatrick, D., & Kirkpatrick, J. (2006). *Evaluating training programs* (3rd ed.). San Francisco, CA: Berrett-Koehler Publishers, Inc.
- Knott, S., Woodward, D., Hoefkens, A., & Limbert, C. (2014). Cognitive behaviour therapy for bulimia nervosa and eating disorders not otherwise specified: Translation from

- randomized controlled trial to a clinical setting. *Behavioural and Cognitive Psychotherapy, First View*, 1–14. doi:10.1017/S1352465814000393
- Knowles, M. S., Holton III, E. F., & Swanson, R. A. (2005). *The Adult Learner* (Sixth.). San Diego, CA: Elsevier.
- Kobak, K. A., Craske, M. G., Rose, R. D., & Wolitsky-Taylor, K. (2013). Web-based therapist training on cognitive behavior therapy for anxiety disorders: A pilot study. *Psychotherapy (Chic)*, 50(2), 235–247. doi:10.1016/j.biotechadv.2011.08.021.Secreted
- Kolko, D. J., Baumann, B. L., Herschell, A. D., Hart, J. A., Holden, E. A., & Wisniewski, S. R. (2012). Implementation of AF-CBT by community practitioners serving child welfare and mental health: A randomized trial. *Child Maltreatment*, 17(1), 32–46. doi:10.1177/1077559511427346
- Krawitz, R. (2001). Borderline personality disorder: foundation training for public mental health clinicians. *Australasian Psychiatry*, 9(1), 25–28. doi:10.1046/j.1440-1665.2001.00308.x
- Krefting, L. (1991). Rigor in qualitative research: The assessment of trustworthiness. *American Journal of Occupational Therapy*, 45(3), 214–222.
- Kuzel, A. J. (1992). Sampling in qualitative inquiry. In B. F. Crabtree & W. L. Miller (Eds.), *Doing Qualitative Research* (pp. 31–44). London, UK: Sage Publications.
- Kvale, S. (2006). Dominance through interviews and dialogues. *Qualitative Inquiry*, 12(3), 480–500. doi:10.1177/1077800406286235
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 1–12. doi:10.3389/fpsyg.2013.00863
- Larson, M. J., Amodeo, M., Locastro, J. S., Muroff, J., Smith, L., & Gerstenberger, E. (2013). Randomized trial of web-based training to promote counselor use of cognitive behavioral therapy skills in client sessions. *Substance Abuse*, 34(2), 179–87. doi:10.1080/08897077.2012.746255
- Levant, R. F. (2005). *Report of the 2005 presidential task force on evidence-based practice*. American Psychological Association. Retrieved from <https://www.apa.org/practice/resources/evidence/evidence-based-report.pdf>
- Lewis, C. C., & Simons, A. D. (2011). A pilot study disseminating cognitive behavioral therapy for depression: Therapist factors and perceptions of barriers to implementation. *Administration and Policy in Mental Health and Mental Health Services Research*, 38, 324–334. doi:10.1007/s10488-011-0348-x

- Liddle, H. A., Rowe, C. L., Gonzalez, A., Henderson, C. E., Dakof, G. A., & Greenbaum, P. E. (2006). Changing provider practices, program environment, and improving outcomes by transporting multidimensional family therapy to an adolescent drug treatment setting. *The American Journal on Addictions, 15*, 102–12. doi:10.1080/10550490601003698
- Lilienfeld, S. O., Ritschel, L. a, Lynn, S. J., Brown, A. P., Cautin, R. L., & Latzman, R. D. (2013). The research-practice gap: bridging the schism between eating disorder researchers and practitioners. *The International Journal of Eating Disorders, 46*(5), 386–94. doi:10.1002/eat.22090
- Lilienfeld, S. O., Ritschel, L. A., Lynn, S. J., Cautin, R. L., & Latzman, R. D. (2013). Why many clinical psychologists are resistant to evidence-based practice: Root causes and constructive remedies. *Clinical Psychology Review, 33*, 883–900. doi:10.1016/j.cpr.2012.09.008
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Wiley.
- Lopez, M. A., & Basco, M. R. (2011). Feasibility of dissemination of cognitive behavioral therapy to Texas community mental health centers. *The Journal of Behavioral Health Services & Research, 38*(1), 91–104. doi:10.1007/s11414-009-9209-8
- Lopez, M. A., Osterberg, L. D., Jensen-Doss, A., & Rae, W. A. (2011). Effects of workshop training for providers under mandated use of an evidence-based practice. *Administration and Policy in Mental Health, 38*(4), 301–12. doi:10.1007/s10488-010-0326-8
- Luck, J., & Peabody, J. W. (2002). Using standardised patients to measure physicians' practice: validation study using audio recordings. *British Medical Journal, 325*, 679. doi:10.1136/bmj.325.7366.679
- Luoma, J. B., Hayes, S. C., Twohig, M. P., Roget, N., Fisher, G., Padilla, M., ... Kohlenberg, B. (2007). Augmenting continuing education with psychologically focused group consultation: Effects on adoption of group drug counseling. *Psychotherapy: Theory, Research, Practice, Training, 44*(4), 463–469. doi:10.1037/0033-3204.44.4.463
- Luoma, J. B., & Vildardaga, J. P. (2013). Improving therapist psychological flexibility while training acceptance and commitment therapy: A pilot study. *Cognitive Behaviour Therapy, 42*(1), 1–8. doi:10.1080/16506073.2012.701662
- Mann, C. . (2003). Observational research methods. Research design II: cohort, cross sectional, and case-control studies. *Emergency Medicine Journal, 20*, 54–60.
- Martino, S., Ball, S. A., Nich, C., Canning-Ball, M., Rounsaville, B. J., & Carroll, K. M. (2011). Teaching community program clinicians motivational interviewing using expert and

- train-the-trainer strategies. *Addiction*, 106, 428–441. doi:10.1111/j.1360-0443.2010.03135.x
- Martino, S., Ball, S. A., Nich, C., Frankforter, T. L., & Carroll, K. M. (2008). Community program therapist adherence and competence in motivational enhancement therapy. *Drug and Alcohol Dependence*, 96, 37–48. doi:10.1016/j.drugalcdep.2008.01.020
- Martino, S., Ball, S., Nich, C., Frankforter, T. L., & Carroll, K. M. (2009). Correspondence of motivational enhancement treatment integrity ratings among therapists, supervisors, and observers. *Psychotherapy Research*, 19(2), 181–93. doi:10.1080/10503300802688460
- Martino, S., Canning-Ball, M., Carroll, K. M., & Rounsaville, B. J. (2011). A criterion-based stepwise approach for training counselors in motivational interviewing. *Journal of Substance Abuse Treatment*, 40(4), 357–65. doi:10.1016/j.jsat.2010.12.004
- Mason, P. (2005). Visual data in applied qualitative research: Lessons from experience. *Qualitative Inquiry*, 5(3), 325–346.
- Mauder, L., Milne, D., & Cameron, L. (2008). Pilot evaluation of brief training in CBT for primary care practitioners. *Behavioural and Cognitive Psychotherapy*, 36, 341–351. doi:10.1017/S1352465808004220
- McDonel, E. C., Bond, G. R., Salyers, M., Fekete, D., Chen, A., McGrew, J. H., & Miller, L. (1997). Implementing assertive community treatment programs in rural settings. *Administration and Policy in Mental Health*, 25(2), 153–173. doi:10.1023/A:1022286921362
- McDonough, M., & Marks, I. M. (2002). Teaching medical students exposure therapy for phobia/panic - Randomized, controlled comparison of face-to-face tutorial in small groups vs. solo computer instruction. *Medical Education*, 36, 412–417. doi:10.1046/j.1365-2923.2002.01210.x
- McGhee, G., Marland, G. R., & Atkinson, J. (2007). Grounded theory research: Literature reviewing and reflexivity. *Journal of Advanced Nursing*, 60(3), 334–342. doi:10.1111/j.1365-2648.2007.04436.x
- McHugh, R. K., & Barlow, D. H. (2010). The dissemination and implementation of evidence-based psychological treatments. A review of current efforts. *The American Psychologist*, 65(2), 73–84. doi:10.1037/a0018121
- McHugh, R. K., & Barlow, D. H. (2012). Training in evidence-based psychological interventions. In R. K. McHugh & D. H. Barlow (Eds.), *Dissemination and implementation of evidence-based psychological interventions* (pp. 43–60). New York: Oxford University Press.

- McHugh, R. K., Murray, H. W., & Barlow, D. H. (2009). Balancing fidelity and adaptation in the dissemination of empirically-supported treatments: The promise of transdiagnostic interventions. *Behaviour Research and Therapy*, 47, 946–953. doi:10.1016/j.brat.2009.07.005
- McHugh, R. K., Whitton, S. W., Peckham, A. D., Welge, J. A., & Otto, M. W. (2013). Patient preferences for psychological vs. pharmacological treatment of psychiatric disorders: A meta-analytic review. *Journal of Clinical Psychiatry*, 74(6), 595–602. doi:10.1016/j.biotechadv.2011.08.021.Secreted
- McManus, F., Rakovshik, S., Kennerley, H., Fennell, M., & Westbrook, D. (2012). An investigation of the accuracy of therapists' self-assessment of cognitive-behaviour therapy skills. *British Journal of Clinical Psychology*, 51, 292–306. doi:10.1111/j.2044-8260.2011.02028.x
- Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2009). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. Washington DC: US Department of Education Office of Planning, Evaluation, and Policy Development Policy and Program Studies Service. doi:10.1016/j.chb.2005.10.002
- Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative Research: A guide to design and implementation*. San Francisco, CA: John Wiley & Sons, Ltd.
- Midanik, L. T., & Greenfield, T. K. (2003). Telephone versus in-person interviews for alcohol use: Results of the 2000 National Alcohol Survey. *Drug and Alcohol Dependence*, 72(3), 209–214. doi:10.1016/S0376-8716(03)00204-7
- Miles, M. B., Huberman, A. M., & Saldana, J. M. (2013). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Thousand Oaks, CA: Sage Publications.
- Miller, W. R., & Mount, K. A. (2001). A small study of training in motivational interviewing: Does one workshop change clinician and client behavior? *Behavioural and Cognitive Psychotherapy*, (29), 457–471.
- Miller, W. R., Yahne, C. E., Moyers, T. B., Martinez, J., & Pirritano, M. (2004). A randomized trial of methods to help clinicians learn motivational interviewing. *Journal of Consulting and Clinical Psychology*, 72(6), 1050–1062. doi:10.1037/0022-006X.72.6.1050
- Morgenstern, J., Morgan, T. J., McCrady, B. S., Keller, D. S., & Carroll, K. M. (2001). Manual-guided cognitive-behavioral therapy training: A promising method for disseminating empirically supported substance abuse treatments to the practice community. *Psychology of Addictive Behaviors*, 15(2), 83–88. doi:10.1037//0893-164X.15.2.83

- Morse, J. M., Barrett, M., Mayan, M., Olson, K., & Spiers, J. (2002). Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods*, 1(2), 13–22. doi:10.1063/1.2011328
- Moyers, T. B., Manuel, J. K., Wilson, P. G., Hendrickson, S. M. L., Talcott, W., & Durand, P. (2008). A randomized trial investigating training in motivational interviewing for behavioral health providers. *Behavioural and Cognitive Psychotherapy*, 36, 149–162. doi:10.1017/S1352465807004055
- Muse, K., & McManus, F. (2013). A systematic review of methods for assessing competence in cognitive-behavioural therapy. *Clinical Psychology Review*, 33(3), 484–499. doi:10.1016/j.cpr.2013.01.010
- Mussell, M. P., Crosby, R. D., Crow, S. J., Knopke, A. J., Peterson, C. B., Wonderlich, S. A., & Mitchell, J. E. (1999). Utilization of empirically supported psychotherapy treatments for individuals with eating disorders: A survey of psychologists. *International Journal of Eating Disorders*, 27(2), 230–237.
- National Crime Victims Research and Treatment Center. (2007). *TF-CBTWeb: First year report*. Charleston, SC: Medical University of South Carolina.
- National Institute for Health and Care Excellence. (2004). *National Clinical Practice Guideline: Eating disorders- core interventions in the treatment and management of anorexia nervosa, bulimia nervosa and related eating disorders*. National Institute of Clinical Excellence/ Gaskell and The British Psychological Society.
- National Institute of Mental Health. (1999). *National Institute of Mental Health: Bridging Science and Service. A Report by the National Advisory Mental Health Council's Clinical Treatment and Services Research Workshop*. Rockville, MD.
- Nelson, T. D., Steele, R. G., & Mize, J. A. (2006). Practitioner attitudes toward evidence-based practice: Themes and challenges. *Administration and Policy in Mental Health*, 33(3), 398–409. doi:10.1007/s10488-006-0044-4
- Onwuegbuzie, A. J., & Leech, N. L. (2007). Validity and qualitative research: An oxymoron? *Quality and Quantity*, 41(2), 233–249. doi:10.1007/s11135-006-9000-3
- Orfaly, R. A., Frances, J. C., Campbell, P., Whittemore, B., Joly, B., & Koh, H. (2005). Train-the-trainer as an educational model in public health preparedness. *Journal of Public Health Management Practice*, 6, 123–127. doi:10.2105/AJPH.30.12.1466
- Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. *Health Services Research*, 34, 1189–1208.
- Peck, R., Olsen, C., & Devore, J. L. (2016). *Introduction to statistics and data analysis* (5th ed.). Boston, MA: Cengage Learning.

- Pereplechikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology, 75*(6), 829–841. doi:10.1037/0022-006X.75.6.829
- Petry, N. M., Alessi, S. M., & Ledgerwood, D. M. (2012). A randomized trial of contingency management delivered by community therapists. *Journal of Consulting and Clinical Psychology, 80*(2), 286–298. doi:10.1016/j.biotechadv.2011.08.021.Secreted
- Petry, N. M., Alessi, S. M., Ledgerwood, D. M., & Sierra, S. (2010). Psychometric properties of the Contingency Management Competence Scale. *Drug and Alcohol Dependence, 109*(1-3), 167–174. doi:10.1016/j.drugalcdep.2009.12.027
- Pope, C., & Mays, N. (1995). Reaching the parts other methods cannot reach: An introduction to qualitative methods in health and health services research. *British Medical Journal, 311*, 42–45. doi:10.1136/bmj.311.6996.42
- Pope, C., Ziebland, S., & Mays, N. (2000). Analysing qualitative data. *Nurse Researcher, 320*, 114–116. doi:10.1192/bjp.109.458.163
- Poulsen, S., Lunn, S., Daniel, S. I. F., Folke, S., Mathiesen, B. B., Katznelson, H., & Fairburn, C. G. (2014). A randomized controlled trial of psychoanalytic psychotherapy or cognitive-behavioral therapy for bulimia nervosa. *The American Journal of Psychiatry, 171*(1), 109–16. doi:10.1176/appi.ajp.2013.12121511
- Rakovshik, S. G., & McManus, F. (2010). Establishing evidence-based training in cognitive behavioral therapy: A review of current empirical findings and theoretical guidance. *Clinical Psychology Review, 30*(5), 496–516. doi:10.1016/j.cpr.2010.03.004
- Rakovshik, S. G., McManus, F., Westbrook, D., Kholmogorova, A. B., Garanian, N. G., Zvereva, N. V., & Ougrin, D. (2013). Randomized trial comparing internet-based training in cognitive behavioural therapy theory, assessment and formulation to delayed-training control. *Behaviour Research and Therapy, 51*(6), 231–239. doi:10.1016/j.brat.2013.01.009
- Rapp, C. A., Etzel-Wise, D., Marty, D., Coffman, M., Carlson, L., Asher, D., ... Holter, M. (2010). Barriers to evidence-based practice implementation: Results of a qualitative study. *Community Mental Health Journal, 46*(2), 112–8. doi:10.1007/s10597-009-9238-z
- Razali, N., & Wah, Y. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of Statistical Modeling and Analytics, 2*(1), 21–33.
- Reeves, S., Kuper, A., & Hodges, B. D. (2008). Qualitative research methodologies: ethnography. *British Medical Journal, 337*, 512–514. doi:10.1136/bmj.a1020

- Ressler, A., & Brooks, G. (2014). The Renfrew difference: Investing in recovery for your client [Webinar]. USA: The Renfrew Center.
- Richards, D. A., & Borglin, G. (2011). Implementation of psychological therapies for anxiety and depression in routine practice: Two year prospective cohort study. *Journal of Affective Disorders*, 133, 51–60. doi:10.1016/j.jad.2011.03.024
- Richards, R., Oliver, J. E., Morris, E., Aherne, K., Iervolino, A. C., & Wingrove, J. (2011). Acceptance and commitment therapy training for clinicians: An evaluation. *The Cognitive Behaviour Therapist*, 4, 114–121. doi:10.1017/S1754470X11000043
- Richards, T. (2002). An intellectual history of NUD*IST and NVivo. *International Journal of Social Research Methodology*, 5(3), 199–214. doi:10.1080/13645570210146267
- Richards, T., & Richards, L. (1994). QSR NUD*IST, version 3.0. London, UK: Sage Publications.
- Roth, A. D., & Pilling, S. (2008). Using an evidence-based methodology to identify the competences required to deliver effective cognitive and behavioural therapy for depression and anxiety disorders. *Behavioural and Cognitive Psychotherapy*, 36, 129–147. doi:10.1017/S1352465808004141
- Roth, A. D., Pilling, S., & Turner, J. (2010). Therapist training and supervision in clinical trials: Implications for clinical practice. *Behavioural and Cognitive Psychotherapy*, 38(3), 291–302. doi:10.1017/S1352465810000068
- Ruppel, P. S., & Mey, G. (2015). Grounded theory methodology—Narrativity revisited. *Integrative Psychological and Behavioral Science*, 49, 174–186. doi:10.1007/s12124-015-9301-y
- Russell, M. C., Silver, S. M., Rogers, S., & Darnell, J. N. (2007). Responding to an identified need: A joint department of defense/departments of veterans affairs training program in eye movement desensitization and reprocessing (EMDR) for clinicians providing trauma services. *International Journal of Stress Management*, 14(1), 61–71. doi:10.1037/1072-5245.14.1.61
- Ruzek, J. I., Rosen, R. C., Garvert, D. W., Smith, L. D., Sears, K. C., Marceau, L., ... Stoddard, A. M. (2014). Online self-administered training of PTSD treatment providers in cognitive-behavioral intervention skills: Results of a randomized controlled trial. *Journal of Traumatic Stress*, 27, 703–711. doi:10.1002/jts.
- Sanders, M. R., Prinz, R. J., & Shapiro, C. J. (2009). Predicting utilization of evidence-based parenting interventions with organizational, service-provider and client variables. *Administration and Policy in Mental Health and Mental Health Services Research*, 36, 133–143. doi:10.1007/s10488-009-0205-3

- Scheffe, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika*, 40(4), 87–104. doi:10.1093/biomet/40.1-2.87
- Schmeeckle, J. M. (2003). Online training: An evaluation of the effectiveness and efficiency of training law enforcement personnel over the internet. *Journal of Science and Education Technology*, 12(3), 205–260.
- Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust?: Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology*, 6(4), 147–151. doi:10.1027/1614-2241/a000016
- Schoener, E. P., Madeja, C. L., Henderson, M. J., Ondersma, S. J., & Janisse, J. J. (2006). Effects of motivational interviewing training on mental health therapist behavior. *Drug and Alcohol Dependence*, 82(3), 269–75. doi:10.1016/j.drugalcdep.2005.10.003
- Schoenwald, S. K., Carter, R. E., Chapman, J. E., & Sheidow, A. J. (2008). Therapist adherence and organizational effects on change in youth behavior problems one year after multisystemic therapy. *Administration and Policy in Mental Health and Mental Health Services Research*, 35, 379–394. doi:10.1007/s10488-008-0181-z
- Schoenwald, S. K., Sheidow, A. J., & Chapman, J. E. (2009). Clinical supervision in the treatment transport: Effects on adherence and outcomes. *Journal of Consulting and Clinical Psychology*, 77(3), 410–421. doi:10.1037/a0013788.Clinical
- Schoenwald, S. K., Sheidow, A. J., Letourneau, E. J., & Liao, J. G. (2003). Transportability of multisystemic therapy: Evidence for multilevel influences. *Mental Health Services Research*, 5(4), 223–239. doi:10.1023/A:1026229102151
- Seng, A. C., Prinz, R. J., & Sanders, M. R. (2006). The role of training variables in effective dissemination of evidence-based parenting interventions. *International Journal of Mental Health Promotion*, 8, 20–28. doi:10.1080/14623730.2006.9721748
- Shafer, M. S., Rhode, R., & Chong, J. (2004). Using distance education to promote the transfer of motivational interviewing skills among behavioral health professionals. *Journal of Substance Abuse Treatment*, 26(2), 141–148. doi:10.1016/S0740-5472(03)00167-3
- Shafran, R., Clark, D. M., Fairburn, C. G., Arntz, A., Barlow, D. H., Ehlers, A., ... Wilson, G. T. (2009). Mind the gap: Improving the dissemination of CBT. *Behaviour Research and Therapy*, 47(11), 902–9. doi:10.1016/j.brat.2009.07.003
- Shapiro, J. R., Berkman, N. D., Brownley, K. a, Sedway, J. a, Lohr, K. N., & Bulik, C. M. (2007). Bulimia nervosa treatment: a systematic review of randomized controlled trials. *The International Journal of Eating Disorders*, 40(4), 321–36. doi:10.1002/eat.20372

- Shapiro, S., & Wilk, M. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3), 591–611.
- Sheidow, A. J., Donohue, B. C., Hill, H. H., Henggeler, S. W., & Ford, J. D. (2008). Development of an audio-tape review system for supporting adherence to an evidence-based treatment. *Professional Psychology: Research and Practice*, 39(5), 553–560. doi:10.1111/j.1600-6143.2008.02497.x.Plasma
- Shevlin, M., Banyard, P., Davies, M., & Griffiths, M. (2000). The validity of student evaluation of teaching in higher education: Love me, love my lectures? *Assessment & Evaluation in Higher Education*, 25, 397–405. doi:10.1080/713611436
- Sholomskas, D. E., & Carroll, K. M. (2006). One small step for manuals: Computer-assisted training in twelve-step facilitation. *Journal of Studies on Alcohol*, 67, 939–945.
- Sholomskas, D. E., Syracuse-Siewert, G., Rounsaville, B. J., Ball, S. A., Nuro, K. F., & Carroll, K. M. (2005). We don't train in vain: a dissemination trial of three strategies of training clinicians in cognitive-behavioral therapy. *Journal of Consulting and Clinical Psychology*, 73(1), 106–115. doi:10.1037/0022-006X.73.1.106
- Simmons, A. M., Milnes, S. M., & Anderson, D. A. (2008). Factors influencing the utilization of empirically supported treatments for eating disorders. *Eating Disorders*, 16(4), 342–354. doi:10.1080/10640260802116017
- Simons, A. D., Padesky, C. A., Montemmarano, J., Lewis, C. C., Murakami, J., Lamb, K., ... Beck, A. T. (2010). Training and dissemination of cognitive behavior therapy for depression in adults: a preliminary examination of therapist competence and client outcomes. *Journal of Consulting and Clinical Psychology*, 78(5), 751–760. doi:10.1037/a0020569
- Simons, A. D., Rozek, D. C., & Serrano, J. L. (2013). Wanted: Reliable and valid measures for the science of cognitive behavioral therapy dissemination and implementation. *Clinical Psychology: Science and Practice*, 20(2), 181–194. doi:10.1111/cpsp.12033
- Siqueland, L., Crits-Christoph, P., Barber, J. P., Butler, S. F., Thase, M., Najavits, L., & Onken, L. S. (2000). The role of therapist characteristics in training effects in cognitive, supportive-expressive, and drug counseling therapies for cocaine dependence. *The Journal of Psychotherapy Practice and Research*, 9(3), 123–130.
- Smith, J. L., Amrhein, P. C., Brooks, A. C., Carpenter, K. M., Levin, D., Schreiber, E. A., ... Nunes, E. V. (2007). Providing live supervision via teleconferencing improves acquisition of motivational interviewing skills after workshop attendance. *The American Journal of Drug and Alcohol Abuse*, 33(1), 163–168. doi:10.1080/00952990601091150
- Smith, J. L., Carpenter, K. M., Amrhein, P. C., Brooks, A. C., Levin, D., Schreiber, E. A., ... Nunes, E. V. (2012). Training substance abuse clinicians in motivational interviewing

- using live supervision via tele-conferencing. *Journal of Consulting and Clinical Psychology*, 80(3), 450–464. doi:10.1038/nature11130.Reduced
- Sofaer, S. (1999). Qualitative methods: what are they and why use them? *Health Services Research*, 34, 1101–1118.
- Stangier, U., Schramm, E., Heidenreich, T., Berger, M., & Clark, D. M. (2011). Cognitive therapy vs interpersonal psychotherapy in social anxiety disorder: A randomized controlled trial. *Archives of General Psychiatry*, 68(7), 692–700. doi:10.1001/archgenpsychiatry.2011.67
- Stewart, R. E., Chambless, D. L., & Baron, J. (2012). Theoretical and practical barriers to practitioners' willingness to seek training in empirically supported treatments. *Journal of Clinical Psychology*, 68(1), 8–23. doi:10.1002/jclp.20832
- Stewart, R. E., Stirman, S. W., & Chambless, D. L. (2012). A qualitative investigation of practicing psychologists' attitudes toward research-informed practice: Implications for dissemination strategies. *Professional Psychology: Research and Practice*, 43(2), 100–109. doi:10.1016/j.biotechadv.2011.08.021.Secreted
- Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis*, (2), 349–367.
- Strauss, A., & Corbin, J. (1994). Grounded theory methodology: An overview. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research*. Thousand Oaks, CA: Sage Publications.
- Strosahl, K. D., Hayes, S. C., Bergan, J., & Romano, P. (1998). Assessing the field effectiveness of acceptance and commitment therapy: An example of the manipulated training research method. *Behavior Therapy*, 29, 35–64. doi:10.1016/S0005-7894(98)80017-8
- Stumpf, R. E., Higa-McMillan, C. K., & Chorpita, B. F. (2009). Implementation of evidence-based services for youth. *Behavior Modification*, 33(1), 48–65.
- Sullivan, G. M., & Feinn, R. (2012). Using effect size - or why the P value is not enough. *Journal of Graduate Medical Education*, 4, 279–82. doi:10.4300/JGME-D-12-00156.1
- Talente, G., & Wilson, J. F. (2007). A potential problem with standardized patient exam validity, 30(1), 64–74.
- Thomas, E., & Magilvy, J. K. (2011). Qualitative rigor or research validity in qualitative research. *Journal for Specialists in Pediatric Nursing*, 16, 151–155. doi:10.1111/j.1744-6155.2011.00283.x
- Thorne, S. (2000). Data analysis in qualitative research. *Evidenced Based Nursing*, 3, 68–70. doi:10.1136/ebn.3.3.68

- Thyer, B. A., & Pignotti, M. (2011). Evidence-based practices do not exist. *Clinical Social Work Journal*, 39, 328–333. doi:10.1007/s10615-011-0358-x
- Tobin, D. L., Banker, J. D., Weisberg, L., & Bowers, W. (2007). I know what you did last summer (and it was not CBT): A factor analytic model of international psychotherapeutic practice in the eating disorders. *International Journal of Eating Disorders*, 40(8), 754–757. doi:10.1002/eat
- Turner, H., Marshall, E., Stopa, L., & Waller, G. (2015). Cognitive-behavioural therapy for outpatients with eating disorders: Effectiveness for a transdiagnostic group in a routine clinical setting. *Behaviour Research and Therapy*, 68, 70–75. doi:10.1016/j.brat.2015.03.001
- Turner, H., Tatham, M., Lant, M., Mountford, V. A., & Waller, G. (2014). Clinicians' concerns about delivering cognitive-behavioural therapy for eating disorders. *Behaviour Research and Therapy*, 57, 38–42. doi:10.1016/j.brat.2014.04.003
- Verdinelli, S., & Scagnoli, N. (2013). Data Display in Qualitative Research. *International Journal of Qualitative Methods*, 12, 359–381.
- Vismara, L. A., Young, G. S., & Rogers, S. J. (2013). Community dissemination of the early start denver model: Implications for science and practice. *Topics in Early Childhood Special Education*, 32(4), 223–233. doi:10.1177/0271121411409250
- Vismara, L. A., Young, G. S., Stahmer, A. C., Griffith, E. M., & Rogers, S. J. (2009). Dissemination of evidence-based practice: Can we train therapists from a distance? *Journal of Autism and Developmental Disorders*, 39, 1636–1651. doi:10.1007/s10803-009-0796-2
- Von Ranson, K. M., & Robinson, K. E. (2006). Who is providing what type of psychotherapy to eating disorder clients? A survey. *The International Journal of Eating Disorders*, 39(1), 27–34. doi:10.1002/eat.20201
- Von Ranson, K. M., Wallace, L. M., & Stevenson, A. (2013). Psychotherapies provided for eating disorders by community clinicians: Infrequent use of evidence-based treatment. *Psychotherapy Research*, 23(3), 333–43. doi:10.1080/10503307.2012.735377
- Wallace, L. M., & von Ranson, K. M. (2011). Treatment manuals: Use in the treatment of bulimia nervosa. *Behaviour Research and Therapy*, 49(11), 815–20. doi:10.1016/j.brat.2011.09.002
- Wallace, L. M., & von Ranson, K. M. (2012). Perceptions and use of empirically-supported psychotherapies among eating disorder professionals. *Behaviour Research and Therapy*, 50(3), 215–22. doi:10.1016/j.brat.2011.12.006

- Waller, G. (2009). Evidence-based treatment and therapist drift. *Behaviour Research and Therapy*, 47, 119–127. doi:10.1016/j.brat.2008.10.018
- Waller, G., Gray, E., Hinrichsen, H., Mountford, V., Lawson, R., & Patient, E. (2014). Cognitive-behavioral therapy for bulimia nervosa and atypical bulimia nervosa: Effectiveness in clinical settings. *The International Journal of Eating Disorders*, 47(1), 13–17. doi:10.1002/eat.22181
- Waller, G., Mountford, V. A., Tatham, M., Turner, H., Gabriel, C., & Webber, R. (2013). Attitudes towards psychotherapy manuals among clinicians treating eating disorders. *Behaviour Research and Therapy*, 51(12), 840–4. doi:10.1016/j.brat.2013.10.004
- Waller, G., Stringer, H., & Meyer, C. (2012). What cognitive behavioral techniques do therapists report using when delivering cognitive behavioral therapy for the eating disorders? *Journal of Consulting and Clinical Psychology*, 80(1), 171–5. doi:10.1037/a0026559
- Walser, R. D., Karlin, B. E., Trockel, M., Mazina, B., & Barr Taylor, C. (2013). Training in and implementation of acceptance and commitment therapy for depression in the Veterans Health Administration: Therapist and patient outcomes. *Behaviour Research and Therapy*, 51, 555–63. doi:10.1016/j.brat.2013.05.009
- Weersing, V. R., Weisz, J. R., & Donenberg, G. R. (2002). Development of the therapy procedures checklist: A therapist-report of measure of technique use in child and adolescent treatment. *Journal of Clinical Child and Adolescent Psychology*, 31(2), 168–180. doi:10.1207/S15374424JCCP3102_03
- Weingardt, K. R., Cucciare, M. A., Bellotti, C., & Lai, W. P. (2009). A randomized trial comparing two models of web-based training in cognitive-behavioral therapy for substance abuse counselors. *Journal of Substance Abuse Treatment*, 37(3), 219–227. doi:10.1016/j.jsat.2009.01.002
- Weissman, M. M., Verdelli, H., Gameroff, M. J., Bledsoe, S. E., Betts, K., Mufson, L., ... Wickramaratne, P. (2006). National survey of psychotherapy training in psychiatry, psychology, and social work. *Archives of General Psychiatry*, 63(8), 925–34. doi:10.1001/archpsyc.63.8.925
- Westbrook, D., McManus, F., Clark, G., & Bennett-Levy, J. (2012). Preliminary evaluation of an online training package in cognitive behaviour therapy: Satisfaction ratings and impact on knowledge and confidence. *Behavioural and Cognitive Psychotherapy*, 40, 481–490. doi:10.1017/S1352465811000701
- Wolitzky-Taylor, K., Zimmermann, M., Arch, J. J., De Guzman, E., & Lagomasino, I. (2015). Has evidence-based psychosocial treatment for anxiety disorders permeated usual care in community mental health settings? *Behaviour Research and Therapy*, 72, 9–17. doi:10.1016/j.brat.2015.06.010

- Wonderlich, S. A., Peterson, C. B., Crosby, R. D., Smith, T. L., Klein, M. H., Mitchell, J. E., & Crow, S. J. (2014). A randomized controlled comparison of integrative cognitive-affective therapy (ICAT) and enhanced cognitive-behavioral therapy (CBT-E) for bulimia nervosa. *Psychological Medicine*, 44(3), 543–53. doi:10.1017/S0033291713001098
- Worrall, J. M., & Fruzzetti, A. (2009). Improving peer supervision ratings of therapist performance in dialectical behavior therapy: An internet-based training system. *Psychotherapy Theory, Research, Practice, Training*, 46(4), 476–479.
- Young, J., & Beck, A. (1980). Cognitive therapy scale: Rating manual. Unpublished Manuscript, Department of Psychiatry, University of Pennsylvania, Philadelphia, USA. doi:10.1016/S0005-7894(70)80030-2
- Young, J. E., & Beck, A. T. (1988). *Revision of the cognitive therapy scale*. Unpublished Manuscript, Department of Psychiatry, University of Pennsylvania, Philadelphia, USA.
- Zipfel, S., Wild, B., Groß, G., Friederich, H.-C., Teufel, M., Schellberg, D., ... Herzog, W. (2014). Focal psychodynamic therapy, cognitive behaviour therapy, and optimised treatment as usual in outpatients with anorexia nervosa (ANTOP study): Randomised controlled trial. *Lancet*, 383, 127–37. doi:10.1016/S0140-6736(13)61746-8

8 INDEX of TABLES

TABLES

Table 1.1 A brief description of CBT-E.....	1.14
Table 2.1 Direct Training Outcome Studies – Uncontrolled Studies	2.49
Table 2.2 Direct Training Outcome Studies – Randomized Controlled Trials.....	2.61
Table 2.3 Indirect Training Studies.....	2.71
Table 3.1 Differences between qualitative and quantitative research	3.9
Table 3.2 The main topic areas covered by the interview schedule	3.21
Table 3.3 Summary of the data coding process	3.23
Table 3.4 Demographic details of therapists.....	3.25
Table 3.5 Study 1 with NICE guidance on qualitative research	3.45
Table 3.6 Location of quotes used in Study 1.....	3.48
Table 4.1 Brief description of workshop training	4.6
Table 4.2 Demographic details of therapists Study 2a.....	4.12
Table 4.3 Outcome of one-way ANOVA for eMeasure scores by training group.....	4.13
Table 4.4 Demographic details of therapists Study 2b.....	4.21
Table 4.5 Outcome of paired samples t-test on eMeasure score	4.22
Table 4.6 Brief description of web-centered training	4.25
Table 4.7 CONSORT style diagram for Web-centered training	4.28
Table 4.8 Demographic details of therapists Study 2c.....	4.30
Table 4.9. Outcome of paired samples t-test on eMeasure score following web-centered training .	4.31
Table 4.10 Outcome of independent samples t-test of change score	4.31
Table 4.11 Split-plot ANOVA for eMeasure scores time by guide-type	4.33
Table 4.12 ANCOVA for post-training eMeasure scores time by guide-type	4.34
Table 4.13 Outcome of independent samples t-test of change score by training type	4.38
Table 4.14 Split-plot ANOVA for eMeasure scores time by training type	4.39
Table 5.1 Clinical scenario topics for the role-play measure.....	5.6
Table 5.2 Demographic details for participants in Study 3a.....	5.11
Table 5.3 Outcome of paired samples t-test on skills assessment	5.12
Table 5.4 Demographic details for participants in Study 3b	5.17
Table 5.5. Outcome of paired samples t-test on roleplay	5.18
Table 5.6 Outcome of independent samples ttest.....	5.22
Table 5.7 Split-plot ANOVA for role-play measure scores time by training type	5.22

9 INDEX of FIGURES

FIGURES

Figure 3.1 Steps of data analysis.....	3.6
Figure 3.2 Main qualitative categories and themes	3.26
Figure 4.1 Design of studies 2a, 2b, 2c, and 2d	4.4
Figure 4.2 Log in page for eMeasure.....	4.8
Figure 4.3 eMeasure scores across three groups.....	4.14
Figure 4.4 Pre/Post eMeasure score by guide-type	4.32
Figure 4.5 eMeasure change scores by training type	4.39
Figure 5.1 Design of Studies 3a, 3b and 3c.....	5.4

10 INDEX of APPENDICES

Appendix 1.1	Web-centred training in CBT-E – Therapist Information Sheet.....	11.0
Appendix 1.2	Study 1 – Qualitative Interviews- Therapist Information Sheet.....	11.2
Appendix 1.3	Study 1 – Qualitative Interviews- Email Consent Form	11.3
Appendix 1.4	Study 1 – Qualitative Interviews- Individual Interview.....	11.4
Appendix 2.1	Study 2 – Further Details on the Development of the eMeasure.....	11.10
Appendix 2.2	Study 2 and 3 – Recruitment Email.....	11.14
Appendix 2.3	Study 2 – CBT-E Supervision Manual for Supervisors.....	11.15
Appendix 2.4	Study 2 – ExamOnline Log-in Details Email.....	11.20

11 APPENDICES

APPENDIX 1.1

WEB-CENTRED TRAINING IN CBT-E

Therapist Information Sheet

Web-Centred Training

Web-centred training is a new and scalable way of training therapists to deliver a psychological treatment. It centres on the use of a comprehensive therapist training website. This includes an accessible and authoritative description of the treatment interspersed with multiple clinical illustrations and exercises.

The website is designed to be viewed on a wide range of equipment (tablets, laptops and desktop computers) and it should work in most settings.

Patients / Clients

It is best if you practise implementing the treatment as you work your way through the website. Ideally you should have two training patients / clients, one ready to start immediately and the other ready to start in about four weeks. These patients / clients should be aged 16 or over, have an eating disorder as their primary diagnosis, and a BMI over 18.5. *Clinical responsibility for your patients' management will remain with you and your local team.*

Guidance

Web-centred training is accompanied by support from a “guide” based in Oxford. The guide will be an expert on the training website. There will be twelve 30-minute telephone (or Skype) -based training sessions designed to help you get the most out of the website. These will be at a time that suits you. The sessions will be weekly at first and then every two weeks or so.

Each of the support sessions will include the following:

- i. Review of your progress viewing the website
- ii. Review of your progress implementing the treatment
- iii. Planning the next steps.

These sessions will be recorded. This is so that we can progressively improve upon the training that we are providing. The recordings will be destroyed within six months of completing the training. They will not be shared with anyone outside the core CREDO team.

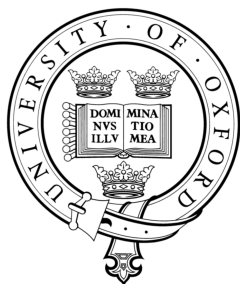
Questionnaires

The training is accompanied by completion of a brief online questionnaire before starting, at the end of the training and six months later.

Problems

If you need to be in touch at any stage please email: credo2@medsci.ox.ac.uk We will get back to you within one working day.

INFORMATION SHEET



Study Title:

Understanding Therapist's Views of Training

Investigator: Suzanne Straebler
DPhil candidate in Psychiatry

Ethical approval: MSD/IDREC/C1/2012/10

Contact details: Department of Psychiatry
Warneford Hospital
Oxford OX3 7JX
Suzanne.straebler@psych.ox.ac.uk
01865 226 387

APPENDIX 1.2

REASONS WHY WE ARE DOING THIS STUDY

At present, a number of evidence-based treatments have been developed to treat a variety of psychological disorders. However, few sufferers have access to these treatments at least in part because training for practitioners in these methods is limited. When training is available it generally takes the form of a workshop presented by an expert in the treatment. Such training has not been evaluated and little is known about how attendees perceive and use this training. This study aims to better understand how clinicians perceive and use the training provided by workshops.

WHAT PARTICIPATION INVOLVES

The study involves participating in a qualitative research interview lasting between 30-40 minutes. During the interview you will be asked about the CBT-E training you will have recently attended. The interview will take place in the two weeks following the workshop, at a time that suits you. It can be conducted at your home, in the Department of Psychiatry at Oxford, over the telephone or via skype. The interview will be recorded to make sure no information is lost and to maintain interview quality. The content of the recordings will be kept confidential and no identifying information will be used. Transcripts will be made of each of the recordings which will then be analyzed using qualitative data analysis software. Emerging themes from the data will be used to generate hypothesis to help improve training. When the study is over the recordings and transcripts will be destroyed.

Appendix 1.3



WARNEFORD HOSPITAL

OXFORD

OX3 7JX

DEPARTMENT OF PSYCHIATRY

CONSENT FORM via EMAIL Therapists' Views of Training

Principal Researcher: Suzanne Straebler, Supervisor: Professor Zafra Cooper

Study Purpose: The study aims to better understand therapists' reactions to training using qualitative interviews.

Please tick box

I confirm that I have read and understand the information sheet dated October 1, 2012 for the above study. I have had the opportunity to consider the information, ask questions and have had any questions I asked answered satisfactorily.

☐

I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason and without any penalty for doing so.

☐

I understand that personal data will be stored and identified using only a number code and will be accessed only by members of the research team and will be destroyed after the study has been completed.

☐

I understand that data collected during the study may be looked at by individuals from the University of Oxford and other regulatory authorities, for the purpose of audit and monitoring, and where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records.

☐

I understand that the research will be written up as a thesis which will be deposited both in print and online in the University archives. The thesis will be published with open access, meaning available to every internet user.

☐

I understand that the study has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee (reference number MSD-IDREC-C1-2012-10).

☐

I understand how to raise a concern or make a complaint.

☐

I agree to take part in the above study.

☐

Name of Participant

Date

Appendix 1.4

Qualitative Study Individual Interview Time Point One Version- Two 1.4.12

In-depth interview focusing on therapist's views of recently attended workshop or web-centered training

Preamble

Thank you for agreeing to participate in this interview today. Before we begin I would like to briefly tell you about this project and what we will do during the interview. I am mainly interested in understanding more about how therapists learn evidence-based treatments and how we might be able to improve the learning of these treatments. During the interview I would like to ask you about your thoughts of the training workshop/ online CBT-E training you just completed. The interview will not contain many yes or no questions; but is more of an opportunity to gather as many of your thoughts, views and opinions about the workshop as possible. Do not hesitate to talk about components of the online training that were unclear, difficult to understand or likely to be impossible to implement in your everyday practice. Likewise, please be sure to include your thoughts on parts of the training that were useful and will likely be a benefit to your everyday practice. None of the comments obtained will be attributed to any one person and the transcript of our conversation will be filed under an anonymous identifier. Do you have any questions before we start?

Check list of domains

- *Professional training*
- *Experience of treating patients with eating disorders*
- *Specific training for ED treatment prior to workshop*
- Prior general training
- Training/learning treatments
- Evidence-based treatment/treatment manuals/protocols etc.
- The treatment (CBT-E)
- The online training
 - format of the presentation
 - relevance of presentation to interviewee's clinical work
 - interest in material
 - effectiveness as a method of learning?
 - Web-guidance
- Perception of competence
- Likelihood of implementation
- Perceived barriers to implementation-internal or external
- Supervision

- Use of outcome measures
- Anything else about the workshop and experience?

Note to interviewer

With the exception of the first three italicized questions that can be asked formally, the interviewer should adopt the following style/method

- Order is not important
- Let participant talk freely and encourage them to continue by using nonverbal cues if face to face/ 'noises of acknowledgement' if over the phone. Noises of acknowledgement include- "mmm", "hmm", "go on", "uh huh", "can you tell me more", "can you explain that further". Please be careful to keep these noises as neutral as possible.
- Repeating/summarizing what the participant has said can be helpful- "so you think..."
- Once the participant does not volunteer any more information, introduce topics by saying "could you tell me what you think about..." Or just "I would like to introduce the topic of..."

Some suggestions on how to introduce a domain area

- Training and specific eating disorders training
 - "A good place to start would be if you could tell me about your current practice of treating patients with eating disorders; how many years have you done so"
 - "Have you had any specific training on the treatment of eating disorders prior to this workshop? If so, what has it consisted of?"
- Prior training
 - "Could you tell me about any prior training you have received?"
- Training/learning treatments
 - "Could you tell me a bit more about any specific training in evidence-based treatments that you have received (manual, workshop, etc.)?"
- Evidence-based treatment/treatment manuals/protocols etc.
 - "Could you tell me what you think about evidence-based treatments in general?"
 - "And what about treatment manuals and treatment protocols?"
- CBT-E
 - "Could you tell me a bit about your views of CBT-E?"
- Thoughts about the workshop/ training

- “Could you tell me your thoughts on the workshop you recently attended?”
- Format of presentation
 - “Could you tell me your views about the format of the presentation (slides/ lecture/ clips)?”
- Relevance
 - “Could you tell me your views about the relevance of the trainings to your current practice?”
- Interest
 - “What are your views about the interest level of the training?”
 - “How interesting was the training”
- Effectiveness
 - “What did you think of the style of training?”
 - “Could you tell me about how effective you believe this style of training is?”
 - “Could you tell me your thoughts on what might be added/ removed to make learning more effective?”
- Perception of competence
 - “What do you think you have learned? / gained?”
 - “What are your thoughts about doing the treatment?”
 - “Could you tell me about your confidence in your ability to do the treatment”
 - “Could you tell me about your feelings on your competence to use the treatment now?”
 - “Could you tell me a bit more about your attitude to using the treatment?”
 - “Could you tell me how you would use the treatment – as a whole treatment or part/ blending with your current way of working?”
 - “What else would be helpful in improving your feelings of competence at delivering CBT-E?”
 - “Could you tell me you are likely to do help to increase your feelings of competence to deliver the treatment?” (read manual, get supervision, etc.)
- Likelihood of implementation
 - “What are your views about using CBT-E?”
- Barriers to implementation
 - “What are your views about any difficulties using CBT-E?”
- Supervision
 - “What are your views about supervision?”

- “Could you tell me about the support and supervision you are likely to be offered (if any)?
- Use of outcome measures
 - “Could you tell me your thoughts on the usefulness of outcomes measures (self-report questionnaires; structured interviews)?”
- Anything else about the workshop and experience?
 - “Is there anything else you would like to add?”

Additional, more specific, questions if needed

Topic- Thoughts

- What will you take away from today?
- What specifically did you learn more about that you did not know?
- What would you have liked to learn more about?
- What specific areas would you have liked more information on?

Topic- Format

- If you were to redesign the workshop format what might you change?
- What would you have like more of?
- What would you have liked less of?
- What did you find helpful?
- What was less useful?

Topic- Relevance

- How well did you feel that the workshop provided what you wished to learn?
- Was it pitched at the appropriate level?
- How relevant was the training when considering your current patients?

Topic- Interest

- How well did the workshop keep your interest?
- Were there some parts that were less/ more interesting than others?

Topic- Perception of Competence

- Do you think you could put into practice what you have learned about the treatment?
- What do you plan to do to help increase your levels of competence? Buy manual/ attend supervision?

Topic- Implementation

- How likely is it that you will put into practice what you have learned?
- How much or little of the day's training will you use in your clinical practice?
- What specific techniques might you like to use?
- What specific techniques do you feel would not fit in with your work?

Topic- Barriers

- What might stop you from using the training in your clinical practice?
- Would your views about the suitability of the treatment for your patients stop you from using it?
- Would there be problems using the treatment in the setting in which you work? e.g., offering 20 sessions, offering twice weekly sessions

Topic- Outcome Measures

- Do you use outcome measures?
- Which ones do you typically use?

Topic- Supervision

- What provision for supervision is available to you?
- Who will provide the supervision (eating disorders specialist, other)?
- How frequently will you receive supervision?
- Will some of your sessions be listened to?

Types of interview questions with examples (Kvale p. 133-135)

Introducing questions:

“Can you tell me about....?”

“Do you remember an occasion when...?”

“What happened in the episode mentioned?”

Follow-up questions:

Direct questioning of what has just been said, nodding, “mm”, repeating significant words

Probing questions:

“Could you say something more about that?”

“Can you give a more detailed description of what happened?”

“Do you have further examples of this?”

Specifying questions:

“What did you think then?”

“What did you actually do when you felt a mounting anxiety?”

“How did your body react?”

Indirect questions:

Projective questions such as ‘How do you believe other pupils regard the competition of grades?’

Appendix 2.1 - Further Details on the Development of the eMeasure

Development of the eMeasure

To ensure that the eMeasure was able to accurately assess knowledge as described by Kirkpatrick (Chapter 4), it was essential to create a reliable assessment tool with clearly defined test objectives and competence benchmarks. To increase the likelihood of creating a valid and usable assessment measure we followed good test development strategies as described in the Standards for Educational and Psychological Testing (American Educational Research Association et al., 2014). Consistent with the *Standards*, Downing and Haladyna (2006; 2014) suggest 12-steps that should be followed for effective test development. In addition to meeting these steps which are outlined below, we also sought advice from the Foundation for the Advancement of International Medical Education and Research (FAIMER). FAIMER's goal is to "improve the health of communities by improving health professions education" (FAIMER, n.d.). The organization is dedicated to improving the education of medical professionals and offers expertise in a number of areas including assessment development. Prior to developing the eMeasure, a two day workshop was held with a FAIMER researcher who provided invaluable information on good test creation.

(i) Overall plan

According to Downing and Haladyna (2006), an overall plan is the starting point for the development of any test. During this step decisions about which concepts to test, which administration methods to use (paper, computer), what type of questions to use (MCQ, true false, essay) are necessary. Step one "places a systematic framework on all major activities associated with the test development project, makes explicit many of the most important *a priori* decisions, puts the entire project on a realistic timeline, and emphasizes test security and quality control issues from the outset" (Downing & Haladyna, 2006, p. 6). Our need for scalable methods made an online computer-based test the most suitable. Additionally, as we intended to use the test as a repeat measure we developed three versions of the test. Creating three tests with roughly equivalent or "parallel" questions allowed us to reduce any "carryover effect" where participants simply remember the questions and then are able to look up the answers before the next test (Brandt, 1991). Several of the questions were identical over the three exams, this also allows for test-retest reliability assessment (Heise, 1969).

(ii) Content definition

Clearly defined content is essential to creating a valid test. Experts in CBT-E were consulted to develop the areas of content. A decision was reached that item content would be exclusively informed by the CBT-E manual (Fairburn, 2008). This would allow for full coverage of topics and ensure that all content was available to trainees.

Stage	Intervention Topics
pre	Indications and contraindications for CBT-E
	Structure of treatment
1	Creating the formulation
	Establishing real-time self-monitoring
	In session weighing
	Establishing regular eating
	Education about eating problems
2	Involving significant others
	Review of progress
3	Over-evaluation of shape and weight: • self-evaluation and its consequences • addressing shape checking and avoidance • addressing feeling fat • exploring origins of over-evaluation • controlling ED-Mindset
	Dietary restraint, rules and controlling eating: • addressing dietary rules • addressing food avoidance • over-evaluation of control of eating
	Events, moods and eating: • addressing event-related changes in eating • addressing residual binges • addressing mood-related changes in eating
4	Ending well: • concerns about ending treatment • ensuring progress is maintained • phasing out treatment procedures • minimizing risk of relapse
post	Post-treatment review

(iii) Test specifications

To ensure that the manual was adequately covered and that the most detailed and longest sections had a higher proportion of related test questions a *test blueprint* (Table 1 Appendix) was developed. A blueprint outlines the proportion of test questions that should be allocated to each section. In addition we specified what type of question each was (theoretical, applied; Table 2 Appendix adapted from: Cooper et al., 2015).

(iv) Item development

The creation of questions or ‘items’ that can measure important content at the correct cognitive level is considered the most difficult challenge for test developers (Haladyna et al., 2014). As described in Chapter Two, knowledge measures in previous research have tended to be multiple choice questionnaires (MCQs), true or false, free text response, or self-report. For the eMeasure we chose to use the multiple choice questionnaire format for several reasons. MCQs are the most widely used item format

and have been shown to be valid measures of knowledge (Haladyna, Downing, & Rodriguez, 2002). Additionally, MCQs offer an objective measure to score which is fast and efficient.

Item creation and writing is an area where expert guidance is suggested: “Without specific training, most novice item writers tend to create poor-quality, flawed, low-cognitive-level test questions that test unimportant or trivial content” (Downing & Haladyna, 2006, p. 11). To ensure the quality of the items all item writers underwent a two day training workshop with FAIMER (described above). Final versions of the questions were also sent to FAIMER for comments.

Table 2 Appendix- Coverage and type of question

Content area/topic	% of items	Application & problem solving %	Theoretical knowledge %
Preparation for treatment	5	0	5
Stage 1	45	30	15
Stages 2 & 3	45	35	10
Stage 4 and review	5	0	5
Totals	100	65	35

Adapted from Cooper et al., 2015

(v) Test design and assembly

This step concerns the quality of the test regarding formatting, errors in grammar and spelling, and in the case of MCQs, proportion of answers that are the same (i.e. balancing how many ‘a’, ‘b’, or ‘c’ responses there are). If there are serious flaws in the test design the interpretation of the results may not be deemed reliable. Much time was taken to examine the test for errors of this kind. At times putting the questions into the online format required focus and diligence. Answer options were balanced proportionally.

(vi) Test production

Test production mainly focusses on security. To create an effective and valid test, multiple versions will have been created with many reviewers. All of this can lead to test questions being leaked which would give test takers prior knowledge of the questions and invalidate the test. To avoid any security concerns all draft versions of the test were held in a single secure University of Oxford repository. Reviewers of the test items were asked to delete old versions and not to save versions to their own computers. The online test administrator ExamOnline (described below), was hosted on a secure server owned by University of Oxford.

(vii) Test administration

The validity of a test can be improved by standardization of administration. Often tests are given at exactly the same time under similar conditions to remove any additional confounding information. The tests are supervised by observers to check for cheating

or sharing of information. A limitation of our study was that we asked respondents to use the 'honor system' when taking the test. We requested that they did not use any books or notes while sitting the exam. As therapists took the test in the place and time of their choosing and from a great distance away, we were not able to have invigilators present. In future, should the eMeasure become part of a licensure examination, it may be useful to host these tests in testing centers. Testing centers are able to offer additional security in terms of monitoring for the use of books and notes and provide invigilators.

(viii) Scoring test responses

Appropriate scoring of a test ensures the test has a valid meaning. It allows those administering the test to believe with some certainty that the test is measuring what they intended it to and that the scores are meaningful (Downing & Haladyna, 2006). The ExamOnline package automatically computed all scores, allowing for a high level of quality control.

(ix) Passing scores

How one determines a passing score or cut-score is extremely important for the validity of the measure. A detailed explanation of cut-score for the CBT-E measure can be found here: Cooper et al., 2015.

(x) Reporting test results

For many tests the reporting of results is guided by factors such as the need to retake the exam in the cases of failing scores, testing score challenges, timeliness and fairness. For the eMeasure we decided in advance that test scores would not be generally released. On some occasions trainees did ask for their score and we emailed it to them along with the cut-score. The decision to not release scores was made on the premise that we were testing the training and scores of the group with individual scores not being the main focus.

(xi) Item banking

Item banking again refers to issues of security. Once items have been developed and the tests are finalized it is important to securely store them for future use. All of our questions and final test versions are either stored for use on the ExamOnline website or in the University's backed-up server. Both systems offer high levels of security.

(xii) Test technical report

The final step is to actually test the exam. Is the exam psychometrically rigorous? Although this step is often overlooked it is the only way to truly know if the test is able to measure what it purports to and if there are any threats to validity. All three versions of the eMeasure have undergone extensive analysis which are reported elsewhere (Cooper et al., 2015). These analyses suggest that the eMeasure is a sound and valid measure for CBT-E. In addition the analyses have shown that all three versions of the eMeasure are in fact parallel in content.

By following these 12 steps we were able to create a useful and valid measure for CBT-E with a competence cut-point.

Appendix 2.2 - Recruitment Email

Email sent to interested participants- eMeasure and role-play measure.

Many thanks for your email offering to help us with our work improving training in psychological treatments. We are writing to provide you with further information.

As you know, we are currently working on developing and testing new ways of training therapists to implement psychological treatments. As part of this we have developed two measures designed to assess people's knowledge of CBT-E and their understanding of how to implement the treatment with clients.

Firstly, we would like to invite you to complete an online questionnaire. The questionnaire will take about 30 minutes to complete and has to be done in one go. Secondly, we are also very keen to find some people who would be willing to volunteer to help us with our new role-play measure assessing the implementation of CBT-E. In this role-play you would be asked to take the role of a therapist, and a member of our team would role-play being a client with an eating disorder. To do this, we would need to meet up face-to-face for an hour. We would be very grateful if you would let us know the answer to the following questions:

1. Are you willing and able to complete the online questionnaire? If so, we will send you a link to the online questionnaire at some point in the next four weeks and simply ask that you complete it
2. Are you willing and able to consider taking part in our role-play measure? If so, please could you provide us with a contact telephone number so that we could ring you to discuss the practicalities?

Thank you very much for agreeing to help us with our work on developing new ways of training therapists to implement psychological treatments.

Would you be willing to help? If so, please respond to this email by saying YES

It takes about 30 minutes and has to be completed in one single session.

Best wishes

Appendix 2.3 - CBT-E Supervision Manual for Supervisors

Preparation for Supervision	
Setting up connection (2-3 months in advance)	Supervisor needs to discuss with team best mode of communication for supervision sessions: • Ideally video conference calls via internet (through IP address) [video conferencing through ISDN lines has shown to be too disruptive] • If the former is unavailable or trust does not permit IP video conferencing use Skype (only with wired connection); participants need to ensure their trust allows Skype if using trust computers; CREDO will provide camera for Skype • Important to test equipment and connection before start of supervision – supervisors to liaise with a nominated IP person at participants' site (in case of IP video conferencing) • If video conferencing or Skype impossible, may need to go ahead with telephone contact (trainees need to have a speaker phone).
Competence assessments	Ensure that all trainees have completed a competence assessment (role-play measure and e-measure) before and after training (workshop plus supervision). Post-training assessment to be completed after 6-8 months of Supervision in order to fit in with timeline of research.
Discuss Practicalities of Supervision (2 months in advance)	1. <u>Supervisor needs to inform trainees about practicalities of CBT-E supervision:</u> • Send out information leaflet for participants (saved here: O:\RESEARCH AND GENERIC TREATMENT INFORMATION\CBT-E Supervision) and ask them to read it! • Summarise main practicalities in email to participants: - Training and SV in focussed, 20-week version of CBT-E - 1-2 hour fortnightly sessions, depending on size of group (2-3 participants: 1 hour; 4 participants: 1.5 hrs; 5-6 participants: 2 hours) - Maximum time frame: 8 months; i.e., 16 sessions - Number of participants: min 2; max 6 - All participants need to have one case ready to start treatment with within 2 weeks of starting Supervision and ideally take on a second case after 2 months of Supervision - Type of case: stress that for training purposes it is

	best to choose someone with typical BN or BN-EDNOS, ideally with the following characteristics: - BMI between 19 and 30 (stress that tx not evidence-based for BMI > 40) - not depressed - ready to start and commit to treatment (i.e., not away in first 4 weeks if tx) - can attend sessions at CBT-E frequency - Training/supervision is based on “Cognitive Behavior Therapy and Eating Disorders” by CGF; stress that in order to be competent in the treatment, participants will need to familiarise self with book. It is suggested participants read the introduction section of the book (up to page 44) before the first supervision session. 2. <u>Supervisor needs to liaise with trainees to set up a fixed day and time for Supervision</u> • Stress that need to commit to at fortnightly intervals • Can be flexible around holiday periods (e.g., Christmas) • Ideally the first meeting should be face-to-face to maximise engagement
Final Preparations before Start of Supervision (2 weeks in advance)	Supervisor to remind participants via email of the following: • Participants need to establish one case ready to start treatment (or who has recently started) and after two months, one or two more • There will be approximately 8 months of supervision, 1 ½ hours fortnightly; encourage participants to block out diary time • Remind participants to read introduction section of book • Encourage participants to email supervisor (and rest of group if helpful) a brief summary of the client they are bringing to supervision (a bullet point type list of current eating disorder features, history of eating disorder, personal and other relevant problem history and personal circumstances) before first meeting. Give option to bring above along to first meeting. Remind participants to remove any identifying information for reasons of confidentiality. Supervisor and participants need to check whether technology for supervision sessions is working (i.e., have a practice run for Skype or video conferencing)

Supervision Sessions	
First session	Agenda for first meeting: 1. Introductions 2. General issues regarding supervision 3. Cases 4. Teaching of treatment interventions as appropriate General issues regarding supervision to be discussed in first meeting: • What is CBT-E Supervision? ⇒ Applied, case-based training; i.e., training in and help with implementing CBT-E as described in manual ⇒ Case management that is specific to CBT-E (i.e., frequency of contact with referrer, treatment practicalities), but not generic/local case management ⇒ Clinical responsibility remains with local team • What is the structure of CBT-E supervision sessions? ⇒ Every meeting starts with setting an agenda ⇒ The meeting time will be divided into two main parts: 1) Discussion of cases – every participant has the opportunity to discuss their client and recent progress; 2) Discussion of treatment interventions/general treatment issues that have arisen from discussion of cases. Discussion of cases will generally take longer than the teaching section of the supervision session. ⇒ If the supervision session lasts longer than 1 hour, a short break (5-10 minutes) can be scheduled in (offer this to participants). • Implementation of the treatment ⇒ The materials needed to deliver CBT-E can be downloaded (free of charge) from http://www.psychiatry.ox.ac.uk/research/researchunits/credo. ⇒ Familiarising oneself with “Cognitive Behavior Therapy and Eating Disorders” is essential for becoming competent in delivering the treatment ⇒ Treatment interventions will be discussed in Supervision sessions as and when appropriate ⇒ Supervisor to raise issue of possibility of having to adapt treatment to suit service guidelines (e.g., unable to see clients twice per week, involvement of dietician etc) and be understanding. Point out however that every adaption of treatment will change the treatment delivered (i.e.,

	therefore not evidence-based anymore). Check current service practice regarding weighing arrangements and use of dietician. • Discussion of Cases ⇒ Participants are encouraged to send a brief email summarising their client's progress before each supervision session to the supervisor (and other participants if helpful) ⇒ Ongoing treatment material (such as formulation or weight graph) can be emailed or faxed through to the supervisor in advance of the session. ⇒ Confidentiality: Patients should only be referred to by their first names/treatment material should not reveal identity ⇒ Encourage participants to come prepared to their supervision session and summarise clients succinctly • Recording of Supervision sessions ⇒ Obtain consent for recording and explain purpose
Following Sessions	• Re-scheduling of sessions ⇒ If participants need to re-schedule a session, try to be flexible but remind participants of maximum time frame of SV sessions (8 months). ⇒ If participants are slow to get back in touch regarding the re-scheduling of a session, do not chase them. • Difficulties implementing CBT-E ⇒ If service guidelines do not allow for implementation of CBT-E as described in manual, discuss deviations from evidence-based treatment and its implications. Also discuss risk of conducting more 'generic' / 'CBT-informed' Supervision, rather than CBT-E SV. Discuss with participants whether possible to bend guidelines for one training case? • Practicalities of SV sessions ⇒ Keep a note of who is attending each session (without it being obvious to participants) ⇒ Point out that it is important to start on time – latecomers can join but it will not be possible to repeat information already discussed. ⇒ If participants in different locations (phone SV or Skype) – need to all start at same time – if not present at start time, cannot join later, as too disruptive for SV session. • Content of Case discussions: ⇒ Keep focus on implementation of CBT-E – e.g., what session/stage of tx is client in, are the appropriate CBT-E procedures being implemented? ⇒ Keep referring to book, ask participants to read relevant

	sections of book between sessions • Give a lot of praise and be positive

Appendix 2.4 - ExamOnline Log-in Details Email

Dear

We are delighted to welcome you to our new web-centred training in Enhanced Cognitive Behaviour Therapy (CBT-E), a leading evidence-based treatment for eating disorders. The training programme was developed by the Centre for Research on Eating Disorders at Oxford (CREDO) which is part of Oxford University Department of Psychiatry. We have a special interest in developing effective methods of training clinicians in psychological treatments, in particular CBT-E.

All trainees are being asked to complete an online questionnaire on CBT-E before they begin the programme. Given that at this stage you will not have received the training, we do not expect you to have much, or indeed any, knowledge of the treatment. As part of our ongoing evaluation of the training programme, it is nevertheless important that we assess knowledge at this point. You will be asked to complete another questionnaire once you finish the programme.

Below is the link to our CBT-E knowledge questionnaire. Please complete this by It takes about 30 minutes and has to be finished in **one single session**. Once you have completed the questionnaire we will be in touch to get you started on the training programme.

We look forward to working with you.

Best wishes

NB: Your responses will be treated in the **strictest confidence and will be anonymised**.

TO ACCESS THE QUESTIONNAIRE PLEASE CLICK ON THIS LINK.

<http://credolearning.com/org/psychiatry-ox-ac-uk/>

Please remember that once you log in, **you will not be able to re-enter the questionnaire**.

Please click on 'candidate' and provide the following information:

Login: CBT0200...

Password: CBT0200....

