2001

Meta-Analysis of Eye Movement Desensitization and Reprocessing Efficacy Studies in the Treatment of PTSD

Carole Alto
Seton Hall University

Follow this and additional works at: http://scholarship.shu.edu/dissertations
Part of the Psychoanalysis and Psychotherapy Commons

Recommended Citation
http://scholarship.shu.edu/dissertations/1293
META-ANALYSIS OF EYE MOVEMENT DESENSITIZATION AND
REPROCESSING EFFICACY STUDIES IN THE TREATMENT OF PTSD

BY

CAROLE ALTO

DISSertation Committee

Sandra Lee, Ph.D., Chair

Thomas Massarelli, Ph.D.

Olivia Lewis-Chang, Ph.D.

Submitted in Partial Fulfillment of the Requirements

For the Degree

Doctor of Philosophy

Seton Hall University

June, 2001
ABSTRACT OF THE DISSERTATION

Meta-Analysis of Eye Movement Desensitization and Reprocessing Efficacy Studies in the Treatment of PTSD

By CAROLE ALTO

Dissertation Chairperson: Sandra Lee, Ph.D.

Eye Movement Desensitization and Reprocessing (EMDR) is a relatively new psychological therapy used in the treatment of Posttraumatic Stress Disorder (PTSD) and other disorders. EMDR is unique in that it combines sets of therapist-administered eye movements in conjunction with protocol-driven cognitive and affective processing related to past trauma. EMDR has become a controversial technique for reasons including a lack of explanation for why it works and stunning claims made for its efficacy in the literature. Despite a large amount of research over the past decade, EMDR has not before been studied meta-analytically in its own right.

The present investigation used meta-analysis to examine the collection of EMDR PTSD studies available in the literature. The literature search resulted in a total of 21 studies, which met inclusion criteria. These primary studies in turn resulted in a collection of 118 effect sizes included in the analysis. Two separate analyses were conducted dependent on whether EMDR was compared to a no treatment control group or an alternative treatment control group.

In addition to an overall estimate of the efficacy of EMDR in the treatment of PTSD represented through an effect size, five sub-hypotheses were investigated. First, it was hypothesized that RMDR would be more efficacious with a non-combat population than with combat-related PTSD. The second sub-hypothesis was that there would be
significantly larger treatment effects associated with verbal report measures than with physiological outcome measures used in EMDR PTSD studies. Third, it was hypothesized that earlier EMDR studies would show larger treatment effects than more recent EMDR studies. The fourth sub-hypothesis concerned treatment dosage. It was hypothesized that there would not be significant differences based on the number of treatment sessions administered. Finally, it was hypothesized that the bilateral stimulation component of EMDR therapy would not contribute significantly to treatment effects.

The analysis consisted of generating effect sizes in the form of standardized difference scores on the various outcome measures. Effect sizes were then grouped according to independent variable categories and averaged together. Before testing for between-group differences, homogeneity testing was completed.

Results of the investigation were mixed, but overall supported the efficacy for EMDR in the treatment of PTSD. In terms of the overall efficacy of EMDR in the treatment of PTSD, moderate range effect sizes were found which meant that EMDR outperformed both no treatment and alternative treatment control groups. There were several hypotheses in which the two separate analyses did not reach agreement on whether or not differences were significant.

Support was found for the first hypothesis as EMDR showed larger treatment effects for a non-combat population than for combat related PTSD. There were mixed results for the second hypothesis as verbal report measures showed larger treatment effects in the EMDR/Control comparison, but not when EMDR was compared with alternative treatments. The third sub-hypothesis was not supported as there were no differences in treatment effects based on recency of study. Additionally, there were no
differences found based on the number of sessions administered, which supported the fourth hypothesis. Support was found for the final sub-hypothesis as the bilateral stimulation component was not found to significantly contribute toward treatment effects.

In light of the present findings, recommendations were made to better clarify and more accurately measure confounding issues in the complex assessment of EMDR in the treatment of PTSD.
Acknowledgments

I wish to thank my chairperson, Dr. Lee, and committee members Dr. Massarelli and Dr. Lewis-Chang for their contributions toward this project. Most of all, I thank them for encouragement and support when I needed it. I thank Dr. Foley for serving as the external reader. I thank Dr. Sadako Vargas for her statistical acumen and excellent teaching without which this ambitious undertaking would not have been completed. I thank my husband, George, whose support makes all things in my life possible.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>i</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>v</td>
</tr>
<tr>
<td>I. THE PROBLEM</td>
<td>1</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Trauma and PTSD</td>
<td>8</td>
</tr>
<tr>
<td>EMDR</td>
<td>15</td>
</tr>
<tr>
<td>Phases of Treatment</td>
<td>20</td>
</tr>
<tr>
<td>EMDR and Posttraumatic Stress Disorder</td>
<td>28</td>
</tr>
<tr>
<td>EMDR at Large</td>
<td>33</td>
</tr>
<tr>
<td>Summary</td>
<td>36</td>
</tr>
<tr>
<td>II. LITERATURE REVIEW</td>
<td>39</td>
</tr>
<tr>
<td>Introduction</td>
<td>39</td>
</tr>
<tr>
<td>EMDR with Combat Related PTSD</td>
<td>40</td>
</tr>
<tr>
<td>EMDR with Non-Combat Related PTSD</td>
<td>59</td>
</tr>
<tr>
<td>EMDR Research Reviews</td>
<td>96</td>
</tr>
<tr>
<td>III. METHODOLOGY</td>
<td>106</td>
</tr>
<tr>
<td>Introduction</td>
<td>106</td>
</tr>
<tr>
<td>Meta-Analysis</td>
<td>108</td>
</tr>
<tr>
<td>Advantages of Using Meta-Analysis</td>
<td>110</td>
</tr>
<tr>
<td>Criticisms and Limitations of Meta-Analysis</td>
<td>112</td>
</tr>
<tr>
<td>Steps in Meta-Analysis</td>
<td>116</td>
</tr>
<tr>
<td>Problem Formulation</td>
<td>117</td>
</tr>
<tr>
<td>Literature Search</td>
<td>118</td>
</tr>
<tr>
<td>Inclusion/Exclusion Criteria</td>
<td>120</td>
</tr>
<tr>
<td>Defining and Coding Variables</td>
<td>122</td>
</tr>
<tr>
<td>The Index of Effect Sizes</td>
<td>124</td>
</tr>
<tr>
<td>Statistical Analysis of Effect Sizes</td>
<td>128</td>
</tr>
<tr>
<td>Hypotheses Tested</td>
<td>131</td>
</tr>
<tr>
<td>Summary</td>
<td>124</td>
</tr>
<tr>
<td>IV. RESULTS</td>
<td>137</td>
</tr>
</tbody>
</table>
LIST OF TABLES

1. Journals Represented in the Meta-Analysis .............................................. 138

2. Descriptive Features of Studies Used in the Meta-Analysis ......................... 139

3. Summary of Pre-Selected Moderator Variables for EMDR/Control Group Comparison .............................................................. 159

4. Summary of Supplemental Moderator Variables Analyzed in EMDR/Control Group Comparison ...................................................... 162

5. Summary of Pre-Selected Moderator Variables for EMDR/Alternative Treatment Comparison ......................................................... 171

6. Summary of Supplemental Moderator Variables in EMDR/Alternative Treatment Comparison ......................................................... 174
LIST OF FIGURES

1. Proportion of Combat and Non-Combat Related PTSD Studies used in the Meta-Analysis ................................................................. 142
2. Effect Size Distribution for EMDR/Control Group Comparison ................. 152
3. Proportion of Self-Report, Therapist-Administered, and Physiological Measures .................................................................................. 156
4. Effect Size Distribution for EMDR/Alternative Treatment Comparison .......................................................... 164
5. Grand Means of EMDR/Control and EMDR/Alternative Treatment Comparisons ........................................................................ 165
CHAPTER I

THE PROBLEM

Introduction

Long before the practice of professional psychology, the human condition has suffered the effects of posttraumatic stress. Psychological effects from trauma come from myriad events ranging from natural disaster and unexpected loss to abusive child rearing practices. The harmful effects of posttraumatic stress can be pervasive and debilitating to an individual's functioning. For many, trauma leaves in its wake implications for the mind, body, and spirit.

As much variation as there is in the individual experience of trauma, there is as bountiful a supply of attempts at healing. Families have long attempted to help their members through home-style interventions while professional clinicians developed therapies like exposure treatment, psychoanalysis, and hypnosis to help resolve trauma related issues.

In 1989, Francine Shapiro happened upon an observation while walking through the park and processing a series of adverse events in her mind. She noticed that when moving her eyes in rapid succession there was some relief of presenting symptoms of her stress. Shapiro, a graduate psychology student at the time, continued to investigate the observation and followed up by developing a protocol to integrate rapid eye movements with a simultaneous processing of cognitive and affective experiences targeted at specific
memories. In a formal study (1989b), she found measurable outcomes with a sample of participants whose primary presenting problems centered around traumatic memories. Other studies followed resulting in the development of a physiologically based type of cognitive therapy known as Eye Movement Desensitization and Reprocessing (EMDR).

The past decade has seen a proliferation of studies on EMDR therapy in the treatment of trauma-related effects including Posttraumatic Stress Disorder. To date, EMDR is the single most studied therapy modality for the treatment of trauma (Shapiro & Forrest, 1997). There have been numerous investigations with a variety of trauma survivors and results have reported both strong treatment effects and no significant treatment effects. Among those found to support the benefit of EMDR (such as Shapiro’s original 1989b study), very large treatment effects have often been cited. Publications ranging from Newsweek to the Journal of Clinical and Consulting Psychology have examined the utility of EMDR and the questions underlying its seemingly too-good-to-be-true effects. No revolution goes uncontested and there has been as fervent an opposition to EMDR as there has been support for it in the clinical treatment literature.

Notably, EMDR has recently been found to be probably efficacious in the treatment of civilian PTSD according to the group of individuals formerly known as the APA Task Force on Promotion and Dissemination of Psychological Procedures (see Chambless, Baker, Baucom, Beutler, Calhoun, Crits-Christoph, Daiuto, DeRubeis, Detweiler, Haaga, Bennett Johnson, McCurry, Mueser, Pope, Sanderson, Shoham, Stickle, Williams, & Woody, 1998). Although in Update on Empirically Validated Therapies (Chambless et al., 1998) the group of researchers no longer published under the APA’s Task Force title, it is essentially the same body of researchers as published in the original 1995 article Training in and Dissemination of Empirically Validated
Psychological Treatments: Report and Recommendations (Chambless, Babich, Crits-Christoph, Frank, Gilson, Montgomery, Rich, Steinberg, & Weinberger, 1995). For the purpose of the present investigation therefore, the two will be considered the same body.

Inclusion on the list of empirically validated therapies is an important milestone as the task force uses specified research criteria in order to evaluate a treatment modality and is currently the premiere credentialing body within the American Psychological Association. In fact, few treatment modalities, even those commonly employed by professional mental health practitioners, have met the validation requirements of the task force. Despite inclusion on the most recent APA list of empirically validated therapies, EMDR continues to arouse controversy and debate in the literature whereas other empirically validated treatments do not.

There are presently dozens of published studies found in the research literature investigating the efficacy of EMDR for treating the effects of psychological trauma. But what does this collection of individual studies add up to? There are conflicting findings and inconsistent treatment effect sizes reported among studies. As the investigations accumulate so, too, does the debate and division over whether or not EMDR has shown itself to be an efficacious treatment for ameliorating the psychological effects associated with post-traumatic stress. Although two well-controlled clinical studies promulgated inclusion on the APA Task Force list of Empirically Validated Therapies for treating Posttraumatic Stress Disorder (Chambless, et. al, 1998), one of them (Wilson, Becker, & Tinker, 1995) did not even use a PTSD population but instead featured "psychologically traumatized individuals" (p. 928). Other respectable studies have found conflicting findings (see Journal of Anxiety Disorders January, 1999 special edition on EMDR). Further, EMDR has met the task force's list as probably efficacious for civilian PTSD but
not for combat-related PTSD, despite the publication of numerous well-controlled studies with this population. Does this mean that EMDR is not efficacious for combat-related PTSD?

The present study is twofold, designed to quantitatively answer the overall efficacy question of EMDR as treatment for PTSD, as well as several related hypotheses. Specifically, the efficacy question is: Is EMDR measurably effective in reducing the symptoms associated with PTSD? And if so, what is the magnitude of effect? Whereas the APA task force requires only two supporting studies meeting stringent research design criteria to deem a study as (probably) efficacious, there have been many more studies with both civilian and combat-related PTSD sufferers. There are numerous research reviews, summaries, and critiques, however, this question has only been addressed empirically once as one variable in a meta-analysis of multiple treatments.

With over 30,000 clinicians presently trained in using EMDR and many more thousands of PTSD survivors worldwide, the time has come to move beyond expert opinions, anecdotal evidence, and individual studies to draw sound conclusions about the efficacy of EMDR.

In addition to the general efficacy question, there are other key areas that warrant empirical investigation. They are:

1. Are there differences in the effect size of EMDR between combat-related PTSD and non-combat types of PTSD? It was hypothesized that there would be larger effect sizes found for non-combat related trauma survivors than for combat related trauma survivors. This is an important question for external validity as many of the frequently cited EMDR efficacy studies rely on a combat-PTSD population when generalizing to PTSD sufferers at large. Additionally, the fact that the APA Task Force
on Promotion and Dissemination of Psychological Procedure’s made a distinction for
including only EMDR for civilian PTSD on its list of empirically validated therapies
warrants further investigation. A measurable comparison between civilian and combat-
related PTSD would assess weather or not treatment differences do in fact exist between
these two populations. This has not been done to this point and is particularly relevant
given the task force’s distinction, a distinction that has not been made for other therapies.
Because of the distinction made by the task force, particularly given the fact that combat
related studies were available in 1998, it is hypothesized that EMDR will show larger
treatment effects for non-combat related PTSD.

2. What, if any, differences in effect size are found between single and multiple
EMDR sessions with PTSD diagnosed individuals? Obviously, it is imperative to
treatment planning to best estimate the dose response curve of treatment, yet this has
largely been ignored in the literature. This comparison was coded in the present analysis
to analyze a range of session clusters in order to compare effect sizes across treatment
lengths. It was hypothesized that there will be no difference between the range of
sessions administered. This hypothesis is based on the belief that, compared to other
treatments or no treatment, EMDR will be more effective as it is an active, solution-
focussed therapy. There are a number of studies that show effects even for single
sessions of EMDR, therefore, despite the fact that more sessions would be likely to result
in more progress, EMDR will still fare better than comparative groups on any range of
sessions.

3. Are there differences in the magnitude of effect on objective dependent
variables versus verbal report measures? Criticisms of early EMDR studies frequently
cited the use of self-report measures given by the client to the therapist as dependent
variables with inherent demand characteristics. The plethora of EMDR studies in the past
decade have since employed many different types of measures including standardized,
objective (i.e. physiological) and self-report indices to assess symptomatology of PTSD.
The present study compared effect sizes derived from objective, physiological indicators
to subjective measures reported by participants in order to see if there are differences
between the subjective experience of the participant and that which can be objectively
measured. It was hypothesized that there would be differences favoring the verbal
measures to physiological measures. This hypothesis was based on the EMDR protocol
which uses subjective feedback from the client to guide the treatment process. Since the
treatment requires continuation until symptom relief is achieved, it is reasonable to
expect that verbal posttest measures will show treatment effects. Physiological indices
were hypothesized to be less sensitive to subjective experience and therefore will show
smaller treatment effects.

4. Are there differences in the magnitude of effect between older EMDR studies
compared with more recent studies? There has been much criticism of early EMDR
studies particularly on the grounds of weak dependent measures, poor treatment fidelity,
and lack of blind evaluators (see Lohr, Kleinknecht, Tolin, & Barrett, 1995). Based on
early study limitations, many more recent studies attempted to minimize these
methodological flaws through research design improvements. A comparison of early
results with more recent results will help resolve the controversy over Shapiro &
Forrest's (1995) early claims regarding the efficacy of EMDR and lend probative value to
the current state of EMDR's efficacy in the psychological nomenclature. This is of
additional interest due to the fact that the APA Task Force on the Validation and
Dissemination of Psychological Therapies did not find EMDR to be an empirically
supported treatment in their first study (1995), but did in the more recent investigation (1998). Because of the controversy over stunning early claims of EMDR and the fact that it was not included on the 1995 list of empirically validated therapies, it was hypothesized that early studies will show larger treatment effects than more recent studies.

5. Are there differences in effect size between EMDR and the same therapy without the bilateral stimulation (that is, the eye movements) component? This is a most important question, particularly if strong effects are found in the general efficacy of EMDR in treating PTSD. If EMDR does work, then what specifically are the mechanisms of change? Determining whether or not the eye movements play a central role is imperative to answering this question. Early in the development of EMDR, Shapiro (1989a) identified the eye movements as critical to the success of EMDR. Soon afterward, other mechanisms of bilateral stimulation such as finger taps were recognized as effective alternatives. Over time and most recently, however, Shapiro (1999) specifically describes EMDR as an "integrated psychotherapy treatment incorporating aspects of many psychological orientations and one that makes use of a variety of bilateral stimuli besides eye movements" (p. 37). Accordingly, the term bilateral stimulation will be used herein to include the eye movement component as well as other forms of bilateral stimulation in EMDR therapy. Further, Shapiro (1999) stated that stressing the eye movements component of EMDR is incorrect and that the success of EMDR lies in many elements including imagery, cognitions, cognitive restructuring, and dosed exposure among others (pgs. 37-38).

While Shapiro (1999) continues to stress the importance of bilateral stimulation, others such as Devilly and Spence (1999), Lohr, Lilienfeld, Tolin, and Herbert (1999),
and Renfrey and Spates (1994) have concluded that the bilateral stimulation component does not contribute towards therapeutic change. A statistical analysis comparing primary studies that have tested for this as an independent variable will help to resolve the controversy. In order to begin to isolate and tease out the specific components of therapeutic change, an aggregate analysis of primary studies designed to assess the role of bilateral stimulation in general and eye movements in particular is necessary. It was hypothesized in the present study that the bilateral stimulation would not be found to contribute significantly toward treatment effects. In addition to the aforementioned studies, the structured and multi-faceted nature of manualized EMDR therapy is likely to result in therapeutic change regardless of the physiological eye movement intervention. In other words, EMDR’s components outside of the bilateral stimulation include the sustenance of sound therapy, which may be why some research has found that the bilateral stimulation was not necessarily the mechanism of change in EMDR.

While there has been ample advocacy from proponents and criticism from opponents of EMDR, the time has come to examine these issues empirically. Meta-analysis will accomplish this within the context of science instead of opinion. In this chapter, background about the effects of psychological trauma, the process and protocols of EMDR therapy, and controversial issues will be presented.

Trauma and Posttraumatic Stress Disorder

Trauma exists in many forms, including single catastrophic events or long enduring conditions like an abusive marriage or childhood. McFarlane and de Girolamo (in van der Kolk, McFarlane, & Weisaeth, 1996) divide traumatic stressors into three categories: Time-limited events like an accident or rape; sequential stressors that have a
cumulative effect, such as a career as an emergency service worker; and long-lasting
trauma exposure, such as combat or familial abuse (p. 132). Trauma is different from
tragedy and the DSM-IV (APA, 1994) specifies that the personal responses necessary for
trauma to qualify as PTSD must include “intense fear, helplessness, or horror” (p. 424).

Despite the context or form of the trauma, from catastrophic natural storms to date
rape, the event leaves its victim feeling powerless, frightened, and terrorized. The
trauma victim is left different from life before the trauma; one’s sense of the world and
core identity is shaken and damaged. Herman (1992) describes traumatic events as
“overwhelming the ordinary systems of care that give people a sense of control,
connection, and meaning...They are extraordinary because they overwhelm the ordinary
human adaptations to life” (p. 33).

While the idea of traumatic events underlying maladaptive psychological
phenomenon is not new, psychology’s acceptance of PTSD as a clinical syndrome has
only a 20-year history. Until publication of the DSM-III (1980), a variety of unspecified,
clustered criteria afflicted people under names like Shock or Combat War Syndrome.
Recognition of PTSD as a clinical disorder not only liberated its victims by recognizing
the validity of their pain, but it also has made more systematic research possible.

All traumas are not equal in their impact and all people do not experience trauma
in the same way; for some an innate resiliency allows the human spirit to transcend the
trauma. For others, however, trauma leaves in its wake a collection of debilitating effects
that can trap a person in their trauma for a lifetime. van der Kolk, McFarlan, and
Weisaeth (1996) emphasize that “Despite the human capacity to survive and adapt,
traumatic experiences can alter people’s psychological, biological, and social equilibrium
to such a degree that the memory of one particular event comes to taint all other experiences” (p. 4).

This latter group suffers from Posttraumatic Stress Disorder, an Axis I Anxiety Disorder diagnosis specified by the DSM-IV (APA, 1994) which includes four major criteria for diagnosis. The first criteria is exposure to a traumatic event that involved real or threatened serious injury/death accompanied by intense fear, helplessness, or horror. The second category of PTSD criteria includes persistent re-experiencing of the event. This typically takes the form of intrusive thoughts, dreams, or feelings. Additionally, there is a persistent avoidance of stimuli associated with the trauma and a diminished general responsiveness. This is experienced as feelings of detachment or decreased interest in previously enjoyed activities. The fourth cluster of symptoms consists of persistently increased arousal, such as insomnia, hypervigilance, or difficulty concentrating. This is like a generalized and continuous state of alert that the body and mind go into following the traumatic event, an inability to be truly at peace with one’s self. Finally, to meet PTSD criteria, these symptoms must endure at least one month and cause clinically significant impairment in important areas of functioning (pgs. 427-429).

There is much disputed in the DSM-IV classification of PTSD including its diagnosis as an anxiety disorder (see Brett in van der Kolk, 1996), the limited breadth of precipitating events deemed eligible for PTSD onset, and the scope of symptoms. Other controversies surrounding the PTSD diagnosis include the questions of whether PTSD is a normal response to abnormal circumstances; if there are personality variables related to one’s propensity for acquiring PTSD; and the prevalence of the disorder.

van der Kolk et al. (1996) emphasize that most people who are exposed to “dreadful experiences” do so without developing a psychiatric disorder (p. 3). Yet PTSD
is a relatively common disorder with estimates of its prevalence ranging from .5% of all men and 1.3% of all women to 24% of those exposed to a traumatic event (p. 140-141). Shalev (in van der Kolk et al., 1996) summarizes five factors identified in trauma research as possible predictors of posttraumatic stress disorder: Pretrauma vulnerability (including biological, personality, and environmental composition); magnitude of the stressor; preparation for the event; immediate and short-term responses; and post-trauma responses (p. 79-91). Research on these possible predictors are consistent with intuitive conclusions: The more intense the trauma the greater likelihood for PTSD onset; the more preparation for a traumatic event, the milder the effects of the trauma; and the stronger the immediate support, the milder the resulting symptoms. It is this latter factor that is of critical importance to mental health researchers and clinicians.

Herman (1992) summarizes the myriad symptoms of PTSD as falling into three main categories: Hyperarousal; intrusion; and constriction. Hyperarousal refers to the constant state of vigilance the body and mind are left with following a traumatic event. This includes a heightened state of perceptual alertness as well as increased psychobiological tension (i.e. chronic arousal of the autonomic nervous system). This is experienced as an unrelenting police state guarding the person in the aftermath of the event. Like all symptoms, frequency and intensity vary with the severity of the disorder. Regardless, it is an inability to maintain genuine and extended states of relaxation. Intrusion is the omnipresent reliving of the event through conscious and unconscious experience. This can occur in the form of intruding cognitions, obsessive thoughts, flashbacks, or dreams. These intrusions do not lose their emotional charge in the re-experiencing, even though they may be experienced as vividly as other memories or as vaguely as fragmented images. Constriction is the aspect of self that becomes numb and
shut down as a result of the trauma. Herman (1992) likens the state of surrender and powerlessness at the time of the trauma to the long-term effect of constriction. It is like an altered state of consciousness, a necessary detachment from the horrible, that results in an emotional numbing beyond the event itself. These three paramount symptoms combine and interact to leave the trauma victim feeling different than before, less complete and less integrated with a drastically compromised quality of life. Importantly, Herman’s (1992) model recognizes the continuum of effects of trauma ranging from the physical to cognitive to affective.

van der Kolk et al. (1996) added two important characteristics to the resulting information processing of PTSD victims: Compulsive re-exposure to the trauma and alterations to defense mechanisms and changes in personal identity (pgs. 10 and 14). The former refers to the tendency for PTSD sufferers to paradoxically reenact the trauma (e.g., the common phenomenon of physically abused children to engage in self-injurious behaviors). The latter refers to the way a trauma gets reintegrated into personality thereby changing it from the way it was before the trauma.

Herman’s (1992) elucidation of these hallmarks of PTSD is important because they highlight the range of emotional and physiological experience that is imperative to treatment of PTSD. Clearly, it is not simply a matter of clarifying and accepting the traumatic memory. The trauma leaves in its wake a battered array of ego functions and a worldview different from before the event took place. In a sense, the person is not the same person as before, and this is consistent with the self-reports of numerous survivors of trauma. The intensity and omnipresence of the three cardinal symptoms makes it easy to see how survivors frequently suffer from substance abuse, suicidal thoughts, and other serious disturbances of feeling and behavior following in the aftermath of the trauma.
Importantly, these effects reach beyond the psychological; there is a clear physiological experience associated with the symptoms. States of hypervigilance and arousal demand cooperation from the nervous system and place a premium on the functioning of a person’s psychobiology.

Unlike other normal life experiences, the traumatic memory is unable to become integrated in a healthy way into the sense of self and the tainting quality of it serves a destructive rather than adaptive purpose. While the mind and body arouse and constrict in an effort for self-preservation, it is this very arousal and constriction that serves to stonewall a return to healthy functioning. The psychobiology of self can not properly re-integrate and adapt when the body is in a heightened state of alert and emotionality is shut down. It is as if the body and mind are still in the combat foxhole; instead of striving for long-term health, there is an emphasis on getting through the day intact.

When this occurs in childhood the results are even more devastating because they interfere and severely compromise the growth of immature ego functions and the capacity to develop attachments (i.e., meaningful relationships). Therefore, in addition to the damage of the trauma to the person per se, the occurrence of a trauma can halt ongoing psychological development. For instance, part of the job of growing up includes learning appropriate social skills, the ability to function as part of a team, and skills like planning and attending among many others. In striving for self-preservation and homeostasis, the traumatized child can does not have the necessary developmental energy or opportunity to devote to these tasks. This helps to explain the sense of feeling fragmented and in some way not whole often reported among victims of PTSD.

Traumatic stress has been in existence since the beginning of humankind, and there is a certain amount of pain and tragedy that every person endures in a lifetime.
There is an old Guatemalan folk tale about a woman who to save her daughter must only find a mustard seed and one family who has not been befallen with a tragedy in their lifetime. Her hope quickly wanes as she travels from village to village only to learn that such a family does not exist. Loss in its myriad forms, the unexpected, and the tragic are part of the human condition; but this is not the sustenance of trauma.

Examination of the DSM-IV (1994) and reports of trauma victims speak to the horror and powerlessness experienced as part of the traumatic event. Contrast this with non-traumatic loss and there is a difference in the quality of experience as well as the intensity of experience. Such is the severity of experience that the mind and body are left with a self that is different than before. If Posttraumatic Stress Disorder is a mind-body experience, as suggested by the DSM-IV and confirmed by numerous studies (see van der Kolk et al. 1996), it is reasonable to conclude that healing occurs at the same level.

Psychotherapy in its many forms has long been involved in helping individuals come to terms with the damage evoked from trauma and has helped many people to gain insight and return to better functioning. However, traditional psychotherapy treatments have fallen short on two levels: for many the process is extensively long and there are a great many people who have not come to resolution through talk therapy alone. Eye Movement Desensitization and Reprocessing (EMDR) therapy has proven to have a place in the treatment of PTSD as an effective therapy for many whose symptoms presented as “untreatable” as well as in facilitating the psychotherapeutic process of typical PTSD patients.
EMDR

EMDR is predicated on the belief that traumatic memories are locked into an individual’s nervous system and as such, the effects of the trauma become locked into the psychological and physiological makeup of the person. By processing the trauma on a physiological level, both the emotional charge of the memory and the psychobiological effects of it are reduced and pre-trauma levels of functioning are once again possible. Shapiro & Forrest (1997) believe that EMDR “reconnects the stored event with the physical information-processing system of the brain” (p. 29) and this facilitates the healing process. In this way, Shapiro (1999 & Shapiro & Forrest, 1997) likens EMDR to an accelerated information processing that un-blocks the traumatic memory from being sufficiently processed by the brain and leads to an adaptive resolution (1999, p.4). The underlying trauma at the core of PTSD is brought to consciousness for therapeutic processing and the resolution consists of reintegrating the experience with new and more adaptive meaning, cognitions, and feelings. EMDR does not replace talk therapy; talk is a primary medium through which the process operates. Instead, EMDR combines a structured way of verbally processing the past and the present with a sensory intervention that summons the nervous system to join in the psychotherapeutic process.

EMDR involves a protocol-specific set of therapist induced eye movements (or other form of bilateral stimulation) that works in tandem with the client’s cognitive and affective experiences associated with their trauma. Essentially what happens is the PTSD client presents with a verbal account of the traumatic experience and specifies the thoughts and feelings experienced at the time of the event as well as the thoughts and feelings experienced in the present. As the client recounts the targeted experience, the
therapist uses her/his fingers to create a set of movements for the client to follow visually. While these eye movements are being elicited, the client continues to verbally process the traumatic experience with the therapist. For example, the therapist may ask the client to describe the scene, elaborate on the feeling and cognition associated with it, and then hold this scene or associated thoughts in their mind (e.g., for thirty seconds) as the client’s eyes continue to follow the therapist’s fingers. These integrated cognition, feeling, and eye movement interactions are called sets and sets continue until there is evident a softening of the traumatic memory and its associated cognitions and feelings. Evidence of this comes from client self-reported assessments throughout the entire EMDR process.

There are two major components that work in tandem with the bilateral stimulation. They are the client’s self-report about how disturbing the feeling associated with the trauma is and the level of validity a client places in the belief about themselves that they have come to believe since the trauma. For example, an extremely disturbing event such as shooting an enemy soldier may be reported as a 10 on a scale of disturbance. Similarly, a belief about oneself over committing this shooting may be “I am a cold-blooded person” and the validity of this belief might be deemed a seven by the client. Specifying the amounts of disturbance and the beliefs (and their level of validity) about oneself that result from the experience helps to tease out the cognitive and affective components of the trauma that contribute to the complex interweave of PTSD symptoms. In this way, irrespective of the bilateral stimulation, the EMDR therapy works toward clarifying the PTSD symptoms from the vague and ominous way they may be experienced.
In order to further specify and also assess these levels of experience, EMDR requires the use of specific scales quantified by the client. Shapiro & Forrest (1997) use the Subjective Units of Disturbance (herein referred to as SUDs) to assess how disturbing a feeling is to an individual. This self-report instrument developed by Joseph Wolpe (1958) uses a 0-10 scale to assess the range of disturbance from not at all disturbing (0) to the greatest disturbance imaginable (10). For instance, in recalling a particular incidence of abuse, a person might report a nine to reflect an extreme level of disturbance whereas the same memory might evoke a one, meaning almost no level of disturbance upon resolution of the trauma.

Similarly, Shapiro & Forrest (1997) use the Validity of Cognition Scale (herein referred to as VOC) for self-reports about how true a belief feels to a client. This is a scale developed by Shapiro (1997) that uses a seven point range from 1-7, one meaning completely false and seven equating to completely true. Part of the damage of trauma is its tendency to instill negative beliefs in its victims about themselves. For instance, trauma victims often report feeling that they are worthless, powerless, or deserving of maltreatment. They frequently attribute their own behavior as somehow responsible for the traumatic event or its aftermath. The structure of an EMDR protocol brings these beliefs to light and requires the client to not only articulate them, but feel them and also to assess their strength. For example, a victim of childhood abuse may state, “I really was a bad kid” and rate this on the VOC scale as a seven, meaning that it is very true to them. Upon resolution of the trauma, the belief “I really was a bad kid” may rate a one, indicating that the client no longer believes that she/he was a bad kid.

The use of ongoing assessment using these scales is paramount to EMDR because they provide the therapist with ongoing and acute feedback as to where the most critical
points of information specific to the individual are. Further, the essential material, though painful, is pulled to the surface allowing the alliance to work together toward resolution. Things happen quickly during EMDR and the reported levels of disturbance and belief provide fodder for processing in an intense and rapid way. An added benefit in the use of the scales is measurement for science as well as the individual. Inherent use of rating scales provides for the ability to measure progress within the individual, over time, and as a treatment modality as well.

As the EMDR session progresses, it is expected that decreased levels of disturbance (as reported by the client in SUD levels) and diminished levels of negative self-belief (assessed by the client on the VOC scale) will emerge. The therapist guides the client through a dual process of both visualizing the disturbing memory and specifying the personal information that the client has taken from the experience. Part of the process involves facilitating the client to identify and target erroneous negative beliefs about themselves, beliefs that may have been installed at the time of trauma and insidiously ingrained over years. Inherent in the plague of trauma is its ability to contribute negatively to one’s very worldview, affecting fundamental notions about the self and the world. These beliefs and the associated trauma become targeted during EMDR as the client continues to verbalize her/his experience while engaging in the eye movements. The direct and acute way that processing is facilitated in EMDR specifically challenges the rationality of these effects left from the trauma and removes their emotional charge.

The following case example illustrates how the therapist uses client feedback to continually target specific thoughts and feelings at the heart of the client’s PTSD symptoms. In this case vignette, Shapiro (from Shapiro & Forrest, 1997) works with a
war veteran who had been suffering with PTSD for 20 years. There were a total of five ninety-minute EMDR sessions over a nine week period. The primary symptoms reported by Eric were intrusive and obsessive thoughts about the war and Shapiro identified two specific combat memories in which Eric felt responsible for enemy deaths as the first targets of the EMDR.

S: Let’s start with the incident with the prisoner.
E: All right. We captured a Viet Cong (proceeds to tell specifics of the traumatic event).
S: Sum up your feelings about the event in a single sentence.
E: I should have done something.
S: Tell me (sic) on a 1-10 how guilty you feel about the event.
E: Nine.
S: Name a positive statement about how you would like to feel about the event.
E: I wasn’t responsible.
S: Fine. Now picture the scene and keep the words ‘I should have done something’ in mind. (Note: Shapiro felt that this would activate the aspect of the memory that most needed healing). *Eye movements begin in short, quick diagonals.* Between sets of eye movements, Eric is told to “Blank it out and take a deep breath.”
E: The picture is the same, but the feeling is down to a six or seven. I hear the sound of the M16 going off two feet from my head. I realize the effect it’s having.
*Two more sets.*
E: It’s a two now.
S: How do the words “I wasn’t responsible’ feel now?
E: I feel better about it. I wasn’t responsible but it still makes me feel sad. Someone’s still dying and I’m afraid the bad feelings will come back.
*Two more sets.*
E: Still a two. Someone’s died. I’m not responsible.
S: What would make it a one?
E: Nothing. Someone still died.
*Here, Eric has a spontaneous insight (memory) about just having woken up at the time of the event.*
E: There really wasn’t much I could do. I never thought about that part before (pgs. 39-40).

Another example taken from Shapiro and Forrest (1997) illustrates the range of emotions processed in a single EMDR session:

Ross kept following Bob’s fingers as he began to move step by step through the succession of emotions associated with the grief process: denial (“That couldn’t
have been him”); depression (“How could he do that to me?”); anger (“I could kill him for doing that”); bargaining (“If I had been a better kid”); and finally acceptance... (sic) ... By the time Ross and Bob finished doing EMDR on the rest of the kidnapping event, Ross made what proved to be a major and seemingly permanent shift in his feelings and attitude. Looking back at the time, after three years, Ross said, “It’s a shame that it happened, but it wasn’t my fault” (pgs. 198-199).

Phases of Treatment

There are eight phases of treatment in EMDR therapy. First, the therapist must take a client history and set up the treatment plan. A careful and thorough history is very important for all patients, and this is particularly true for those presenting with PTSD. Shapiro (1999) emphasizes the objectives of this phase as determining three key factors: What the past memories that set the dysfunction in motion are; the present triggers that bring up the disturbance; and the skills required for more adaptive future behavior (p. 8).

Another critical objective of this initial phase includes determining if the client will be an appropriate candidate for EMDR. This includes discussion regarding informed consent and assessment of ego strength and levels of dissociation, which may be present in the client presenting with PTSD symptomatology, and exploration of other preliminary issues. EMDR brings up much affective experience and the potential presence of a dissociative disorder or other vulnerabilities are crucial in treatment planning. See Carlson & Putnam (in Shapiro, 1999a, p. 110-127) for a more detailed account of assessment of dissociation.

Information garnered from the history is then used to formulate a treatment plan that according to Shapiro and Forrest (1997) “defines specific targets on which to use EMDR” (p. 51). Critical information to treatment planning includes the past event(s) at the heart of the problem, the present situations that cause distress (i.e., trigger situations for symptom manifestation or exacerbation), and identification of crucial skills and
behaviors the client needs to acquire for improved psychological function. This last item is especially important because one of the strengths of EMDR therapy is its emphasis on replacing negative, erroneous ways of thinking or feeling with more productive and adaptive behaviors. Although the intake and treatment planning are an important first phase of treatment, like initial phases in all modalities, the course of treatment is subject to change commensurate with client needs and emerging information and progress. This is particularly relevant to EMDR therapy because it is quite common for clients to spontaneously recall previously inaccessible memories related to traumatic events.

The second phase of treatment, client preparation, consists of a thorough explanation of the therapy to the client. This includes informing them of what to expect during and after treatment, explaining the theory of EMDR, and describing how it is done. It is during this phase that the clinician teaches the client a repertoire of relaxation techniques which may be necessary for some clients as they discuss the distressing and disturbing types of information that gets processed during sessions. Establishing a “safe place” is crucial toward this end and Level 1 training provides specific exercises to construct a safe place with the client prior to any processing. Part of the therapist’s responsibility throughout the often times alarming and disarming processing is to be sure that the client maintains a sense of safety and control. Much of what is processed is material that has been buried or avoided for a long time and it is distressing to the client, yet necessary for the process. During the preparation phase, the client is taught how to “maintain a dual awareness of first the distressing material from the past that is triggered by the eye movements and, second, the fact that they are now safe in the present” (Shapiro & Forrest, 1997, p. 52). EMDR is intense work and can only be effective if the client is both able and prepared to work with that level of intensity.
In addition to relaxation techniques, a system of communication about how to terminate or interrupt a point too intense is established. For instance, the client is instructed to raise a hand in the stop gesture if it is deemed necessary at any point in the therapy. This would provide the client with the necessary level of control and alert the therapist to the need to stop, which of course would take place immediately. Providing this kind of proactive system enables the clients themselves to control the dosage of material able to be metabolized without relying on clinical interpretation of a client’s threshold. Herman (1992) stresses the feelings of powerlessness that engulf the trauma survivor and this single example of the client-driven nature of EMDR is indicative of its clinical emphasis on client-centered control throughout the course of therapy.

Conducting a sensitive and complete preparation phase is essential, not only for understanding the treatment, but also for establishing a safe landscape of treatment and a trusting foundation for the therapeutic alliance.

Phase three, assessment, more specifically determines the detailed aspects of the targets to be subject to the EMDR process. Initially, the person selects a specific scene representing the memory. The client then formulates a statement that conveys a negative self-belief associated with the event. An example could be the client visualizing a scene of childhood sexual abuse and associating the statement, “I am helpless” with it. Next, the client selects a positive statement about her/himself that they would rather believe, such as “I have the ability to help myself” or “I am safe now.” The client must then assess how true the positive statement is felt to be using the VOC scale and evaluates the disturbance of the recalled image with the SUD scale. Queries are made to further identify the negative emotions and physical symptoms associated with the target and the SUD scale is used to rate these levels of disturbance as well. While all of this is going
on, there is an awareness of the levels of experience (i.e., the mind-body connection) being developed within the client. The specificity of the targets, though painful, are helpful in clarifying what for many have previously been experienced more vaguely, though also painfully.

Following phase three, which up to this point has been strictly verbal (with the possible exception of eye movements in the safe place exercise or introduction and preparation for the treatment), the therapist has a solid sense of the material to be processed and the client has both an understanding of what will occur in the therapy and a clarified sense of the sustenance of what has composed their particular experience of PTSD. Phase four, desensitization, operates on the client’s disturbing emotions by having the client visualize, verbalize, and hold the scene with associated emotions as the therapist initiates sets of eye movements or other forms of bilateral stimulation. Shapiro & Forrest (1997) summarize desensitization, “This phase encompasses all of the person’s responses (including memories, insights, and associations that may arise during the sessions) as the targeted event changes and its disturbing elements are resolved” (p. 53). It is here that the client faces the raw material of the disturbing images and puts into words the distressing memory and its related feelings. This can be very intense as the trauma is on some level re-experienced in the therapy and the clinician must be mindful to continually establish in the client some level of control; the dual awareness of being in the present while exploring the depths of the past.

The sustenance of this phase focuses on processing core trauma(s) much in the same way as other types of psychotherapy do. Unlike other treatment modalities however, EMDR provides a more structured and multifaceted scope of processing. Each phase is goal-specific and the therapeutic response targets the objective in the processing
of that phase. For example, the desensitization phase specifically addresses thoughts, emotions, and body sensations experienced as the memory is brought up in the client and the eye movements accompany its processing. The therapist begins the eye movements slowly and provides reassuring comments such as “That’s it” once or twice as the phase proceeds. Part of desensitization is aimed at the client holding the disturbing picture and its associated cognitive and affective experiences and the therapist will encourage this by a comment such as “Just notice it.” As the eye movements and processing continue during this phase, the client is then instructed to let the image and feelings go, thereby softening the memory through desensitization. Sets of eye movements facilitate this and the client and therapist revisit the original image with sets of eye movements until this is accomplished. Ongoing assessment via the Subjective Units of Disturbance Scale (SUD) is initiated by the therapist in order to have the client self-assess progress and provide the therapist with indicators about how the treatment is working in the present.

Installation, phase five, features the building of replacement behaviors (i.e., thoughts and feelings) to take the place of identified negative self-beliefs and emotions that have resulted from the traumatic experience. This is a crucial component of the therapy and part of the strength of EMDR because instead of merely exploring the traumatic memories and effects of PTSD, in this phase the client moves toward a more adaptive and productive mode of operating. At the time of trauma, the event often installs in its victim a negative and irrational self-belief that is strong and enduring. In fact, these negative beliefs become reinforced over time. The negative self-beliefs identified during the assessment phase are replaced by positive beliefs which are strengthened and ultimately installed as the eye movements and verbalizations are processed in this stage. An example would be a person victimized by a mugger whose
negative self-belief was “I am helpless.” After a thorough imaging and desensitization of the traumatic memory, this belief could be reprocessed as “I am in control now.” Concurrently, eye movements continue and the client assesses the strength of the installed belief using the VOC scale until there is considerable progress in assessed VOC levels. The objective of linking the desired positive cognition with the original memory is worked through by continued sets of eye movements accompanied by holding the images and related feelings with the cognition.

Of course, the goal is to establish movement from where someone was to where they wish to be in terms of the cognitive, affective, and physiological experience associated with the trauma and concomitant PTSD. In EMDR this is recognized through a high VOC score during the installation phase, however this does not always occur immediately and there can be any number of realistic factors that make a desired cognitive self-belief to be untrue. In such cases, there are several possibilities including working on the factors that make the desired self-belief untrue (e.g., a client is not in control because of a current marriage to an abusive spouse). EMDR does not aim to develop flawed or fantasy-like thinking in its clients and to install such variety of beliefs in a client would certainly be counterproductive. However, by processing the negative self-beliefs and identifying the desired state of being, the client can accomplish coming to terms with the trauma underlying their PTSD, developing a more positive self-belief about the event, and understanding the relationship between the PTSD and the current obstacles to improved functioning. The client can then be liberated to work in those related areas of functioning, in the former example that would include involvement in an abusive relationship.
Following the installation of positive beliefs comes the body scan phase. This occurs when the clinician has the client bring back up the originally targeted image in order to ascertain whether there are any remaining physical tensions or symptoms associated with it. If there are, further EMDR processing is indicated and the physical tensions are specifically targeted until relief is established. Once a positive physical feeling is established, this too is reinforced through additional eye movement set(s).

PTSD is comprised of interrelated symptoms of the mind and body and EMDR is purposely aimed at working on this interrelationship for improved functioning of the whole person. Targeting bodily symptoms of PTSD survivors is often an effective conduit for accessing underlying emotion-laden information that is not easily verbalized by many clients. Focusing on the physical often activates the associative network of memory and the interrelationship of the physical with the psychological. For example, a client could target the feeling of stomach nausea and activate the memory of his drunken father coming up the stairs. Therefore, like the treatment at large, this phase offers further opportunity to explore the PTSD effects at their origin as well as their present. A crucial underpinning of this phase of the process is in the idea of the therapy working towards a reintegration of the self. van der Kolk et al. (1996) describe trauma as leaving the PTSD survivor feeling fragmented and un-whole. By working at integrating body sensations with what’s going on in the mind, EMDR reintegrates these core experiences that contribute to one’s sense of self.

The seventh phase is closure and this pertains to individual sessions as well EMDR as a treatment process. It is important that the client end a session with composure and relief, even if processing of the trauma is not resolved in a given session. Closure consists of processing to the best extent possible for the session and also bringing
the client to temporary resolution until completion of the full therapy. For instance, a single memory will be targeted and processed until VOC and SUD levels are decreased without expanding to another image. Continued sets may be necessary in order to assure that the client does not leave the session while currently re-experiencing the trauma, as elicited by the treatment process, without bringing it to closure. Relaxation and visualization techniques may be employed during this phase in order to facilitate temporary resolution and reinforce client control over the processing until the next session.

The client is debriefed about possible effects to occur between sessions and equipped with a range of skills to help cope with effects of the treatment. It is common for EMDR to evoke additional memories and cognitions outside and between sessions and the client needs to both understand the possibility of this happening and be prepared for controlling it. In addition to reviewing the relaxation techniques discussed during the preparation phase, other strategies like journal writing and therapist contact are discussed. The goal during this phase is to help ensure that the level of client control established during the session is able to be maintained outside of the clinician's office, particularly in light of breakthroughs made through the therapy.

Reevaluation is the eighth and final phase of EMDR and this occurs at the beginning of each session subsequent to the first. This consists of confirming that the gains made in the level of disturbance experienced and associated self-beliefs, as well as physical sensations (as measured in SUD and VOC levels) remain. If there has been a weakening of effects, they are again targeted through the EMDR process. Additionally, other targets and any new areas that have emerged, such as associated memories recalled as a result of an earlier session, are processed. The sustenance of reevaluation is client
follow-up and this includes the overall integration established through the past EMDR as well as checking with the client about newly emerged cognitions, feelings, and body sensations that may have occurred. Behavioral shifts and newly emerged material, as well as emotional or cognitive blocks of previous material, can then be integrated into the treatment as a result of reevaluation. This builds into the treatment protocol the ability to process evolving and residual material while also maintaining focus on the assessed PTSD treatment goals.

**EMDR and Posttraumatic Stress Disorder**

Shapiro (1989b) has reported that EMDR has been found to be successful in treating trauma victims in as little as a single session. There has since been a differentiation made between victims of a single trauma and multiply traumatized individuals with respect to recommended treatment duration. Most recently Shapiro (1999) stated, “The successful treatment of PTSD of multiple-trauma victims, such as combat veterans and repeated molestation victims, usually requires longer treatment time than is the case for single-trauma clients, since many of the various traumata must be targeted separately” (p. 51). It is imperative to understand that while EMDR can work very effectively at targeting the traumatic memory and processing the irrational and negative cognitive beliefs installed at the time of trauma, all clients are different and the treatment should mirror the client rather than vice versa.

In addition to the negative cognitions, affective component, and body sensations the trauma leaves in its wake, trauma compromises the development of healthy ego function development. Therefore, developing healthy ego functioning and reintegrating the resolved trauma and newly acquired beliefs and skills can be primary goals for the
therapeutic alliance beyond the EMDR. Specifically damaged are the ability to form healthy attachments and the integration of a sense of self which has been fragmented by the PTSD. Examination of the DSM-IV symptoms of numbing of general responsiveness in particular (see diagnostic criteria C p. 428) highlight the difficulty PTSD survivors face in forming and maintaining significant object relations. This has been largely ignored in the literature. Primarily, the controversy over whether or not EMDR is efficacious in treating the more overt PTSD symptoms has taken center stage.

In the brilliant chapter “Reconnection” (pgs. 196-213), Herman (1992) describes the tasks of trauma victims in creating a new life after coming to terms with the traumatic past. Describing the tasks of a sexual abuse victim, the client has “Mourned the old self that the trauma has destroyed and must now develop a new self, new relationships, the survivor must reclaim her world. Emerging from an environment of total control, they feel simultaneously the wonder and uncertainty of freedom” (p. 196). At that point in the therapy, the maladaptive symptomatology may be cleared, however, the client is not necessarily ready to reintegrate into the world because much developmental ground may have been lost to the PTSD. Now the alliance must work towards reconnecting and adapting to the ‘new’ self. EMDR lays the foundation for such reconnection by moving the sense of control from the trauma that binds the victim into the control of the survivors themselves. There is a reintegration of the event into the sense of self with newfound beliefs and feelings accompanying the trauma. In this way, EMDR facilitates the psychotherapeutic process and helps resolve and relieve the Posttraumatic Stress Disorder. It does not, however, install ego functions that were not there before the trauma.

Herman (1992) provides a stage model of trauma recovery that includes three main phases: Safety; remembrance and mourning; and reconnection. The goal of a
successful recovery, according to Herman (1992) is “A gradual shift from unpredictable danger to reliable safety, from dissociated trauma to acknowledged memory, and from stigmatized isolation to restored social connection” (p. 155).

There are strong parallels between Herman’s (1992) stages and the eight phases of EMDR therapy. For instance, in the first stage, Herman (1992) stresses the therapeutic necessity of creating a safe environment, just as the preparation phase of EMDR strives for. She states (1992), “The second most common therapeutic error is premature engagement in exploratory work without sufficient attention to the tasks of establishing safety and securing a therapeutic alliance” (p. 173). Like Shapiro (1995), Herman stresses the “Need for a thorough evaluation of the patient’s current situation before agreement is reached on the focus of psychotherapy” (p. 174). It is in this stage that, according to Herman (1992), the patient tells the story of the trauma. Whereas Shapiro (1995) describes the therapist as a facilitator in this phase, Herman (1992) states, “The therapist plays the role of a witness and ally in whose presence the survivor can speak of the unspeakable” (p. 175).

EMDR’s middle phases bear great similarity to Herman’s middle stage of recovery called remembrance and mourning. Tasks of this middle phase of Herman’s (1992) model include reconstructing the story (p. 176), transforming the traumatic memory (p. 181), and mourning the traumatic loss (p. 188). It is easy to see the parallels between EMDR’s middle phase tasks of recalling the traumatic memory/image, installation of positive cognition, and processing of affective associations. Paramount to both is the importance of empowering the client during the middle phases of treatment. While Herman (1992) stresses the empowerment gained through telling the trauma narrative during therapy, EMDR empowers by building client control into the therapeutic
process. The client gains a power over her/his trauma by processing and regulating the targeted memories. Just as Shapiro (1995) notes the common occurrence of new memories coming to light during EMDR, so too, does Herman (1992). She notes, "As the patient experiences the full impact of facts she already knows, new recollections emerge spontaneously" (p. 184).

It appears that in accordance with this model, the latter phase of recovery is fodder for post-EMDR psychotherapy; that is, EMDR can facilitate the stages of safety and remembrance and mourning thereby setting the stage for reconnection to take place. It is interesting that the literature on trauma, including EMDR studies, consistently addresses the damage to normal ego function development following the trauma, yet the rebuilding of ego functions is conspicuously absent from treatment intervention studies. Reconnection, according to Herman (1992), is a time when the trauma victim builds that which has not been adequately constructed and rebuilds those functions of the self that have their basis in faulty development. It is here that the therapeutic alliance can work effectively in the more traditional modalities of psychotherapy, having used EMDR as a springboard past the trauma.

EMDR therapy is ideally suited for the treatment of PTSD because the central trauma makes the therapy relatively straightforward (Shapiro & Forrest, 1997, p. 99). By processing core issues in an acute way, maladaptive effects in the form of PTSD symptomatology dissipate. Further, the processing is multilevel: It works on the cognitive, physical, and affective and the past as well as the present. PTSD has a tendency to lock its victims into maladaptive functioning on many levels and it makes sense to treat it in an integrated way that recognizes these dimensions. Part of the phenomenon of PTSD is the mind’s tendency to dissociate from the pain of the trauma.
Shapiro & Forrest (1997) note that although this is a natural survival mechanism, dissociation from the trauma keeps one in the grips of PTSD (p. 143). EMDR resists this current and forces the necessary information processing to begin so that healing may take place.

The harm of PTSD is that it binds its victims in the present to a painful event of the past. From the onset of PTSD, body, behavior, thought, and feeling no longer have the same quality of functioning as they did before the trauma. The memory network established around the trauma subsumes all four of these components of functioning and like a ball and chain, they force the individual into a new way of functioning that restricts the normal range of mental health. EMDR works to disintegrate the chain trauma wraps around the individual with the goal of reintegration of positive affect and self-belief to promote a return to healthy functioning. EMDR does not erase the memory, nor does it try to. Rather, the goal is to replace erroneous negative associations of the trauma with adaptive ones so that the memory no longer causes PTSD symptomatology. By shattering the ties with PTSD, the survivor is able to live more fully in the present, despite having been a victim of a past trauma.

In fact, the basic “victimization” of the trauma is challenged not just by the results of EMDR, but by the very process. EMDR places the client in the driver’s seat of their recovery and puts the trauma in the back seat. Coates (1996) uses the expression, “A dance in which the client leads” (p. 2). The client is able to control the extent and intensity to which events are processed and is able to confront the traumatic memories and PTSD in full face rather than under cover of symptoms and treatment resistance. In EMDR, the client is responsible for their healing with the therapist assisting in the therapeutic process. Shapiro & Forrest (1997) describe this empowerment with a rape
survivor, “Going in and out of the experience of the assault reinforces the person’s knowledge that she is bigger than the trauma. The terrible images can be created and destroyed. The emotions can be felt and distanced from. The woman begins to regain control and perspective” (p. 143). The trick, says Shapiro, (in Butler, 1993) “is to keep the client anchored in present time while keeping the memories flowing” (p. 28). Rather than being at the mercy of the trauma, the EMDR affords the client the opportunity to be proactive, to control the processing rather than be controlled by the trauma.

EMDR at Large

Although EMDR was originally developed for trauma resolution, there are an increasing number of conditions for which EMDR has been found to be successful. There are different, specific protocols required for the various disorders treatable with EMDR and training is imperative for any clinician using this modality with any particular type of problem. Shapiro (1996) believes that the effectiveness of EMDR has to do with the brain’s information processing system and that the trauma becomes locked in the system causing malfunction. EMDR organizes and activates the information processing facilitating resolution of the event. Shapiro (1996) states, “Delayed learning takes place through the association of the targeted traumatic memory with neuropsychological networks which contain more appropriate information” (p. 144). This information processing model, therefore, is applicable to other clinical syndromes rooted in dysfunctional information processing.

In addition to varied populations, there appears to be a range of bilateral stimulation variations that have similar positive effects to standard EMDR protocols. Shapiro & Forrest (1997) cite the case of a 2½ year old boy, too hyperactive and
developmentally delayed to be able to follow the necessary eye movements. His therapist had the boy follow hard candy with his eyes and added alternating finger taps in a patty-cake style game (p. 103). Other variations include alternating hand taps on upturned palms, and delivering audio tones in alternating ears. The term bilateral stimulation used to describe this component of the treatment includes all of these variations in addition to the most widely used form which is the eye movements. In fact, successful variations in the basic method underlying EMDR has led Shapiro to propose EMDR as a bilateral stimulation therapy. According to Bender & Britt (1998), MRI studies are being conducted in order to test changes in the brain as a function of bilateral stimulation. van der Kolk (in Lott, 1998) is also engaged in neurological studies that use single photon emission computed tomography imaging to see if bilateral stimulation caused by eye movements produces measurable changes in the brain (p. 38). Recently, Levin, Lazgrove, and van der Kolk (1999) used neuroimaging techniques to identify the areas of the brain most active following EMDR.

Because EMDR is an intensive and sensitive therapy that deals with extremely raw and personal material, proper training and responsible supervision are absolutely essential, even for veteran psychologists. Training is available to licensed or certified masters level clinicians in the mental health field. The basic training is a two-part course provided by EMDR Institute sanctioned practitioners. The courses are a full weekend (Friday evening and full days Saturday and Sunday) in duration and the two parts are divided into Level I and Level II training. Each training type provides direct small group instruction, live and videotaped sessions, and supervised practice for students. It is recommended that the practitioner practice EMDR skills for a minimum of 20-30 sessions after Level I and before the more advanced training of Level II. Whereas Level
I training provides an overview in basic principles, practices, and applications to trauma, Level II offers a more specialized training with applications to a broader range of disorders. Currently, there is no required certification for EMDR therapy, however, the EMDR International Association (EMDRIA) offers certification through registered training.

Shapiro, originator of EMDR, feels strongly about the need for proper training and adherence to developed protocols. She cites the example of a PTSD sufferer traumatized by a rape who required hospitalization after a session with an untrained psychologist evoked the disturbing memory without attending to its emotional aftermath (Butler, 1993, p. 98). In a San Mateo Times article (da Sa, 1996), Shapiro stated this type of potential harm from untrained clinicians is real. “Without skillful facilitation, clients can become stuck in the activated disturbance state which re-traumatizes them and puts them at risk” (p. 3). For this reason, Shapiro and the EMDR Institute emphasize that coherent and successful completion of the treatment protocol-specific phases of EMDR is essential. Importantly, the APA task force’s findings are based on manualized EMDR treatment as administered by EMDR trained clinicians which means external validity relies upon similar training and treatment.

Shapiro remains convicted that the strength of EMDR is not to be taken lightly and not for amateur psychologists who are willing to use clients as part of an experiment. She states (in Butler, 1993) “training is crucial for learning how to keep a client from dissociating or cutting short the intense reliving of trauma that EMDR can precipitate” (p. 28). She further notes that incomplete EMDR or noncompliance with appropriate protocols can fill a client with disturbing images and sensations without bringing them to resolution. “I didn’t realize at first how much emotionally disturbing material it could
open up, and how dangerous it could be for people who suffered multiple traumas...If you stop someone in the middle of remembering a suicide, they may go home and attempt” (Butler, 1993, p. 28).

As of August 1999, over 30,000 clinicians throughout the world have been trained in EMDR. Additionally, over ten books and hundreds of research experiments and articles have been written. Currently, according to the EMDR Institute (1999), EMDR “Has had more published case reports and controlled outcome research to support it than any other method used in the treatment of Posttraumatic Stress Disorder” (p. 1).

Summary

EMDR therapy is a popular, but highly controversial treatment for ameliorating the symptoms of Posttraumatic Stress Disorder. EMDR is a protocol-specific eight-phase treatment process that incorporates therapist induced eye movements or other bilateral stimulation with client reported cognitive, affective, and physiological experiences. EMDR is predicated on the belief that the effects of traumatic experience become locked in the central nervous system and require some level of physiological intervention for resolution. Since its inception little more than a decade ago, EMDR has been used on thousands of clients ranging from adults to children with an ever widening application to clinical disorders beyond the treatment of trauma. Despite dozens of case study reports, group comparison research, and inclusion on the APA Task Force on the Validation and Dissemination of Validated Therapies (Chambless et al., 1998) there continues to be skepticism surrounding treatment effectiveness for EMDR.

Current controversies over EMDR include:

1. Conflicting findings regarding the efficacy of EMDR in treating PTSD.
2. The integrity of studies investigating EMDR.

3. The lack of theoretical explanation for how EMDR works.

4. Inconsistency regarding treatment dose response curves.

Additional controversies include the level of training required for clinicians and the types of dependent variables associated with its outcome measurement. Many in the field also question whether the bilateral stimulation component even has an integral role in the healing process (see Devilly, Spence, & Rapee, 1998; Pitman, Orr, Altman, Longpre, Poire, & Macklin, 1996; and Renfrey & Spates, 1994 among others). These conflicts and the backlash ensuing after EMDR’s inception and popularity have led to many individual studies led by both proponents and opponents of EMDR.

A review of the literature reflects this division among mental health researchers and clinicians. EMDR seems to bring out professional passions and this is evident in the number of clinicians currently being trained as well as through the low credibility statement given by Acierno, Hersen, Van Hasselt, and Tremont (1994) “To date, several uncontrolled case studies have found EMDR to be effective” (p. 287). Articles such as McNally’s (1999) “EMDR and mesmerism: A comparative historical analysis” and the Skeptical Inquirer’s “EMDR treatment: Less than meets the eye?” (Lilienfeld, 1996) further fuel the fire. The controversy continues with debate as recent as 1999 evident in rival articles of a special feature edition of The Journal of Anxiety Disorders devoted to EMDR (Jan., 1999) and a 2000 Dateline episode presenting the opinions of rival camps.

Contradictory findings about the efficacy of EMDR in the treatment of PTSD are reported in the literature. Whereas, for example, Wilson, Becker, and Tinker (1997) and Forbes, Creamer, and Rycroft (1994) reported significant improvement in PTSD symptoms, Devilly et al. (1998 and 1999) and Jensen (1994) did not. Further, there were
many research reviews that resulted in inconclusive generalizations about EMDR based on shortcomings of the studies available. Interestingly, some advocates of EMDR (such as Shapiro, 1996) tended to critique studies that did not find large treatment effects in results, while opponents such as Lohr et al. (1999) believe that EMDR should be subject to particular scrutiny since its originator claims it involves processes outside of extant treatments (p. 201).

An empirical investigation of the efficacy of EMDR is warranted. With the benefit of being the single most studied treatment modality for PTSD, a meta-analytic inquiry into the efficacy of EMDR in the treatment of PTSD will provide important and useful information for mental health practitioners. Further, a meta-analysis that explores specific treatment variables such as treatment duration, type of trauma, and the role of bilateral stimulation as well as research variables like types of dependent variables will contribute illuminating details toward resolving the controversy.
CHAPTER II
LITERATURE REVIEW

Introduction

EMDR is one of the most widely studied therapies in the psychological literature today and the most researched modality for treating PTSD (Shapiro & Forrest, 1997, p.247). Likely, this is due to a combination of clinical interest in the rapidity and outcomes of EMDR as well as the counterintuitive and controversial claims made by advocates of EMDR. The scope of interest in EMDR ranges from features on the 20/20 news magazine show (ABC News, 1994) to articles in popular magazines and newspapers like the New York Times (Paquette, 1997) to formal research documented in professional psychology journals. Research has taken place with diverse populations of trauma survivors including children with natural disaster-related distress (Chermob, Nakishima, Hamada, & Carlson, 1996) to Vietnam War veterans (Silver, Brooks, & Obenchain, 1995) to sexual assault victims (Rothbaum, 1997).

Since its inception marked by the publication of Shapiro’s two introductory articles (1989a & 1989b), EMDR has been the subject of numerous systematic investigations, including both group treatment comparison research and single case designs. Additionally, there were dozens of case studies, anecdotal reports, and critiques found in the literature. The present literature review is limited to systematic group comparison research as is salient to the efficacy of EMDR with PTSD diagnosed
individuals. While there are numerous studies investigating the efficacy of EMDR with a variety of traumatic memory variables, the present review is limited to the literature studying EMDR with PTSD diagnosed populations, as this will be the population of interest to the proposed meta-analytic study.

An interesting and dubious finding of the literature review was that many published studies purported to investigate EMDR in the treatment of PTSD when in fact the sample of participants were not diagnosed as having PTSD. This includes studies such as “The cognitive dismantling of EMDR treatment of Posttraumatic Stress Disorder” (Cusack & Spates, 1999) where of the 27 participants, only 18 met DSM-IV criteria for PTSD (p. 93). The title of such a study is incorrect and misleading and blurs the lines of external validity. The present investigation will avoid misinterpretation by excluding studies that do not specifically include and thus generalize to a PTSD diagnosed population.

EMDR with Combat-Related PTSD

Much of the efficacy literature investigated the use of EMDR with combat-related PTSD, particularly the earlier studies. Yet, as mentioned previously, the APA Task Force on the Validation and Dissemination of Psychological Procedures (1998) found EMDR to be probably efficacious in treating only civilian PTSD. Many of the combat-related PTSD investigations have been conducted with samples derived from the Veterans Affairs (VA) centers, most of which include veterans from the Vietnam War. Although Shapiro (1989b) included such veterans in her original study, she later cautioned (1999b) that war veterans primarily constitute a “multiply traumatized”
population and that such samples are not ideal for single or brief duration EMDR
treatment (p. 36).

Boudewyns, Stwertka, Hyer, Albrecht, and Sperr (1993) conducted one of the
earliest published EMDR efficacy studies of a combat-traumatized PTSD population.
Twenty chronic inpatient veterans were randomly assigned to either the EMDR,
exposure, or control conditions. Patients were combat veterans diagnosed with PTSD by
an interdisciplinary VA team according to DSM-III-R criteria and using the Clinician
Administered PTSD Scale (CAPS) to corroborate diagnosis. Exclusion criteria included
anyone who had taken psychotropic medication, alcohol, or other substances for a month
before the study began.

The EMDR condition consisted of two ninety-minute EMD(R) sessions; the
exposure consisted of imaginal exposure to the traumatic memory in what the authors
describe as similar to the EMD, but without the eye movements. The control condition
consisted of the standard SIPU (VA Special Inpatient PTSD Unit) milieu treatment. The
frequency and duration of the latter two treatments was unspecified. The treatment was
carried out over a two week period with posttesting approximately a week after the final
session. Dependent variables included the Clinician Administered PTSD Scale,
Subjective Units of Disturbance, the intrusion and avoidance subscales of the Impact of
Events Scale (IOE), and the Mississippi Scale for Combat-related PTSD. Additionally,
the physiological measures of heart rate, skin conductance, and hand temperature were
used at pre and posttest. There was also a therapist rating provided (and analyzed using
chi-square) of each participant in the EMD and exposure conditions as either a responder
or non-responder, though it is unclear why.
Boudewyns et al. (1993) found that compared to the latter two conditions, the EMDR condition resulted in greater decreased levels of disturbance on the client and therapist reported SUDs scale. However, when using this same instrument in response to a tape-recording of the traumatic memory there were no significant decreases for any of the treatment conditions. Chi-square performed on the therapists’ ratings of responders vs. non-responders showed a significantly greater number of responders in the EMD group. However, in a later research review, Lohr, et al. (1999) reported that these therapists were not blind to treatment conditions (p. 193). There were no significant differences found on any of the other therapist or self-report instruments including the Impact of Events Scale, the Mississippi Scale, or the CAPS. On all of the physiological measures, there were no significant treatment interactions indicating that although there was some difference between pre and posttest measures, no treatment was significantly more effective than another.

Strengths of the study include the use of psychological and physiological dependent variables, experienced and EMDR trained therapists, and comparison with both a standard control and an exposure therapy control to rule out the possibility of exposure-only treatment effects. The authors acknowledged several serious methodological flaws including insufficient treatment duration and inclusion of participants who were receiving some form of compensation for their continued diagnosed condition of PTSD. Other limitations include the small sample size and consequent low statistical power as well as the lack of fidelity checks. The study did not specify information regarding the blindness of therapists, there were no treatment fidelity safeguards documented, and specific screening/entrance criteria was not documented outside of the use of substances. It was noted by the authors that although participants
generally were multiply-traumatized, the treatment focused on the single most disturbing memory. However, it seems to that this is true of most studies reviewed and there is no evidence to suggest that this should prevent treatment effects as many of the participants Shapiro’s original 1989b study likely fit that description as well.

Two of these authors, Boudewyns and Hyer (1996) conducted another study more comprehensive than the initial pilot described above. Sixty-one VA patients diagnosed with PTSD via the DSM-III-R (and corroborated with the Structured Interview for DSM-III-R and review of client history) were randomly assigned to three groups. Treatment groups included the EMDR, exposure-control, and control. The exposure group as described in the study was the same treatment as the EMDR except there was no eye movement component (therefore this will qualify as EMDR w/o eye movements for inclusion in the present meta-analysis). Exclusion criteria included a co-existing diagnosis of organic mental disorder, schizophrenia, or antisocial personality disorder. Ten licensed psychologists trained in EMDR administered all treatments.

Treatment consisted of between five and seven sessions of EMDR, the same number of sessions of exposure, and eight sessions of the control. Control treatment was the standard VA milieu specified only as standard therapy. Prior to therapy, all participants were audiotaped describing their most disturbing traumatic memory. Dependent therapist and client-rated measures included the Clinician Administered PTSD Scale (CAPS), SUDs, intrusions and avoidance subscales of the Impact of Events Scale (IOES), and the anxiety scale of the Profile of Moods Scale (POMS). Physiological indices included frontalis EMG, skin conductance, heart rate, and blood pressure, however, for reasons addressed by the author, only heart rate was used in the analysis.
Results of heart rate analyses show significant decreases in both the EMDR and exposure only groups compared to the control, but no differences between these two groups themselves. The same pattern was found on the SUDs and on the POMs with the EMDR and exposure-only groups’ PTSD symptoms dropping significantly compared to controls, but no differences between these two groups. The IOE showed no significant change and was considered unreliable by the authors anyway as pretreatment differences made comparisons difficult. On the CAPs, all three treatment groups dropped significantly although there were no interactions indicating that no group was superior to another in the analysis. The authors believe, however, that this is an important clinical finding nonetheless as it shows positive change is possible with such a chronic population.

The Boudewyns and Hyer (1996) study describes itself as unrelated to current EMDR controversies and more interested in empirical inquiry into EMDR and this spirit was evidenced in the discussion section where the benefits of EMDR and future questions were offered. Other strengths included random videotaping with feedback on techniques to the therapists as a fidelity check. Additionally, the use of a standard control as well as a second comparison group designed to assess the role of eye movements vs. exposure and cognitive reprocessing only was beneficial. Discussion of the impact of attrition and participant tolerance for techniques was also helpful. Unfortunately, the nature of the population precluded screening for secondary gains as the participants are largely dependent on VA benefits.

Other limitations include sample size and statistical power, as was the case for the majority of studies located in the present literature review. Although the sample used was larger than most (N= 61), division into three groups nonetheless resulted in small
sized cells. Statistical power refers to a study’s sensitivity in detecting treatment effects that exist. That is, when a study is powerful it is better able to find differences accounted for by treatment effects than a study with low power. Statistically, this ability to avoid type II error is adversely affected by small cell size among groups. Therefore, if a study such as Boudewyns and Hyer (1996) has 61 participants divided into three groups, it has much less power to detect effects than if it were split into two treatment groups.

Jensen (1994) compared EMDR (N=13) with a control condition (N=12) in a study of 25 PTSD diagnosed Vietnam combat veterans. EMDR treatment consisted of a history taking session and two treatment sessions based on the manualized protocol. The control condition was reported to be “the same as the EMD/R except that control group subjects did not receive EMD/R treatment,” (p. 319) meaning that they served as a no treatment control. The author reported no significant differences on the Validity of Cognition (VOC), Goal Attainment Scale (GAS), Structured Interview for PTSD (SI-PTSD), and Mississippi Scale for Combat-related PTSD (MPTSD) measures between groups at post-test. The Subjective Units of Disturbance means did differ, with the EMDR group experiencing less in-session anxiety than controls.

There were strengths in the research design including random assignment, increased use of objective measures, use of an outside evaluator to monitor treatment fidelity, and an attempt to include goals of clinical significance beyond the common barometer of statistical significance (i.e. participants were assisted in developing a PTSD-related treatment goal). This appears to be the first EMDR study to address treatment fidelity through an outside rater. However, the study had serious methodological concerns, most acknowledged by the authors, including the small sample size, low power to detect potential effects, and brief treatment duration of two sessions (plus an additional
history taking session). The control group was aware that they were not in the experimental group and further, they were able to seek outside treatment during the study for their PTSD.

Most notable a threat to internal validity was the implementing therapists, two psychology interns whose education levels were unspecified by the author. The interns did have the two-day Level I training provided by Dr. Shapiro, however, the only reported implementation of EMDR was, “Prior to the study, these therapist also practiced the procedure by utilizing other interns as ‘clients’” (p. 317). Since the level of disturbing experiences focused on during EMDR training is expressly milder than that specified under the DSM-IV (APA, 1994) PTSD criteria A. (p. 427), it is quite possible that the implementing therapist-interns had no prior experience in working with Post-Traumatic Stress Disordered clients. Further, the Level I Training Manual (Shapiro, 1999) specifies utilizing EMDR with milder clinical presentations upon initial completion of the training (p. 57).

An outside monitor was employed to monitor treatment fidelity and criticized the amount of time that the therapists allowed in their clients’ active phase of problem resolution. Reported in Jensen (1994), “Negative results could not be used to criticize EMD/R as the clients may have received enough treatment to open up difficult areas, but without enough fidelity to the treatment to resolve these problems” (p. 321). It must be noted, however, that the independent rater used was H. Lipke, a known EMDR advocate. Further, Jensen argues that since the treatment therapists were properly trained, they may be representative of those practicing EMDR. In light of these compromises, results of the Jensen study must be interpreted with caution.
Pitman, Orr, Altman, Longpre, Poire, and Macklin (1996) conducted a well-designed dismantling study that compared EMDR with an eyes-fixed (but otherwise same) treatment on 17 PTSD diagnosed Vietnam veterans. The authors, like many others (see Renfrey & Spates, 1994; and Devilly et al., 1998; among others) were interested in assessing whether or not the bilateral stimulation contributed significantly to the treatment. Therefore, participants were administered both EMDR and a similar treatment which uses all components of EMDR with the exception of having the client maintain an “eyes fixed” position. The eyes-fixed treatment was identical to the EMDR condition with three exceptions: (a) the therapist used the hand movements but the client was instructed to maintain eyes fixed on a stationary point, (b) this was consistently monitored and corrected (if necessary) by the therapist, and (c) the participant was instructed to rhythmically finger tap during the therapists sets of hand movements.

Results show no significant differences between groups on the physiologic measures, contrary to the authors’ first hypothesis, which the Pitman et al. (1996) conclude, “Thus, the extent of emotional processing did not significantly differ in (sic) the conditions” (p. 425). These indices were thought to assess process variables, specifically extent of emotional processing, as they were monitored during the sessions. The authors hypothesized that there would be greater improvement on outcome measures in the EMDR group and this was not supported by the results, which showed stronger overall improvement in the eyes-fixed group. There was, however, a modest to moderate (23%) improvement across groups indicating that the EMDR and similar eyes-fixed treatments left participants in a better condition as a result of the therapy. The authors concluded that the bilateral stimulation did not contribute to treatment effects, “The finding that eye movements may be deleted from the EMDR procedure without loss of
emotional processing and therapeutic benefit necessarily falsifies neurological theories of the role of eye movements in EMDR’s mechanism of action” (p. 426). Pitman et al. (1996), who had done previous research with PTSD-diagnosed veterans further concluded, EMDR was “at least as efficacious as imaginal flooding proved to be in a previous study and was better tolerated by subjects” (p. 419). Further, the authors stated, “We would prefer EMDR over imaginal flooding in the treatment of chronic, combat-related PTSD” (p.428). Based on findings, however, the authors encouraged future research to delineate the specific components of the various psychotherapeutic approaches are integral to effects.

This study had several strong points, including the ability to conduct sessions up to 110 minutes which closely parallels the 90 minutes per sessions recommended by Shapiro (1997). Further, the six treatment sessions plus an additional preparatory session more closely resembles the landscape of a clinical treatment plan for a PTSD-diagnosed veteran than the two sessions allotted for each participant in the Jensen (1994) study. The three implementing therapists were doctoral level behavior therapists who had both Level I and Level II EMDR training in addition to clinical experience treating veterans. Treatment fidelity was assessed using an independent rater and results were examined quantitatively. The dependent measures included both objective measures, that is, physiologic indices like heart rate and skin conductance, and client, therapist completed assessments like the Subjective Units of Disturbance and Mississippi Scale for Combat-Related PTSD. Additionally, the authors attempted to include clinically meaningful outcome measures such as the number of intrusive combat memories experienced by individuals on assessment days. Methods to ascertain this data included using an
interval-programmed wrist watch and log for each participant to estimate and record intrusions experienced.

Despite the authors' attempts to address criticisms of previous EMDR studies in the research design, there were compromises to the Pitman et al. (1996) investigation. The sample size (n=17) was small and use of an independent evaluator showed some weakness in treatment fidelity, particularly in the eyes-fixed condition. More recent extensions of EMDR therapy have included alternate forms of physiological intervention (including finger tapping) so there is some question whether or not Pitman et al. (1996) were measuring a single or different construct. Shapiro (1999b) has since preferred to use the term bilateral or alternative stimulation to describe the physiological component to the EMDR process. In this way, the Pitman, et al. (1996) study seems more a component analysis than a crossover design. Also, physiological process variables were measured continuously and while this promoted measurement accuracy, there is some question as to the impact such devices had on participants’ therapeutic processing.

Carlson, Chemtob, Rusnak, and Hedlund (1996) similarly found no significant differences on physiologic measures among a small group of EMDR treated PTSD-diagnosed Vietnam veterans. Also similar, however, the authors did find what was termed, “very substantial clinical improvement for three (of four) veterans along a number of psychological dimensions that characterize the disorder” (p. 104). Though multi-modal measurements were documented on several widely accepted dependent measures, the study did not perform any significance testing and it is therefore unclear how the authors reached their conclusions. This study was conducted on four veterans and designed as a ABAA single-subject series design. Despite its potential as a well-planned and clinically relevant work, the omission of a comparison group make the
investigation difficult to interpret. This is unfortunate as the study had otherwise substantial strengths, particularly the utilization of 12 sessions averaging 60-75 minutes in duration.

In 1998, Carlson, Chmtoh, Rusnak, Hedlund, and Muraoka published a more comprehensive randomized controlled outcome group design utilizing the same 12 session treatment plan, but this time with a larger sample and comparison groups to allow for analysis of group differences. Thirty-five PTSD-diagnosed combat veterans were assigned to one of three conditions: EMDR; biofeedback-assisted relaxation; or control, which was routine clinical care (with wait list for experimental treatment). Participants were screened for psychosis and substance abuse, administered pretreatment measures, and then randomly assigned to one of the three conditions. The three implementing therapists were EMDR trained and two of the three were also versed in biofeedback. All three therapists implemented both experimental conditions. Twelve sessions of 60-75 minutes in duration were administered over six weeks, each with continuous physiologic measures including heart rate, thermal assessment, and skin conductance. Cognitive/behavioral assessment instruments were administered following treatment and again three months later. These included the Mississippi Scale for Combat-Related PTSD (MS-PTSD), Impact of Events Scale (IOES), PTSD Symptoms Scale (PTSD-SS), Beck Depression Inventory (BDI), Clinician-Administered PTSD Scale (CAPS), and the Spielberger State-Trait Anxiety Inventory (STAI).

Consistent with the two previously reported studies, results on the physiologic measures did not show significant differences between groups. While there was a main effect for lowered physiological arousal from pre-assessment to posttest among all three groups, there were no group/treatment interactions indicating that the changes were not
treatment specific. On the self-report and standardized symptom measures, however, there were significant differences found favoring a reduction in PTSD symptomatology for the EMDR group. This includes statistical differences on the MS-PTSD, PTSD-SS, and the IES. The EMDR group also showed statistically significant differences on the BDI, however, these were not maintained at the three month follow-up. The STAI did not show differences relative to treatment. Posttreatment differences between the two experimental groups on the CAPS showed decreased PTSD symptoms for the EMDR group. Whereas 7/9 of the biofeedback group who remained at follow-up continued to meet the criteria for PTSD, only 2/9 members of the EMDR group met the criteria three months after treatment.

Of course, these are small sized groups and this is certainly a limitation of the present study. A further limitation includes the lack of exclusionary criteria for secondary gains (particularly relevant when working with VA participants). The authors noted the potential for interviewer bias during the initial (non-blind) follow-up interview as well as the potential impact of the second therapist-monitor in the room during treatment sessions.

Positive aspects of the study are the utilization of up to 12 sessions of treatment, steps taken to ensure that the sample was representative of a chronic PTSD veteran population (see p. 5), and the use of audiotapes for treatment fidelity checks. Unfortunately, an independent evaluator was not utilized to assess the tapes, however, an additional therapist was used to monitor treatment and keep computerized count of eye movements. Training in both therapeutic techniques was adequate and using the same therapists for both conditions. The use of relaxation, a popular PTSD therapeutic technique, as a comparative treatment made for a meaningful and realistic comparison
with this population. The dependent variables employed represented a comprehensive continuum of indices and inclusion of eight physiological measures was impressive and important in light of the number of EMDR studies relying on purely verbal dependent variables. Finally, follow-up data were strong and the fact that a blind evaluator was used for the nine-month follow-up interviews lends confidence to the data.

The Carlson et al. (1996) study's discussion section commented on other studies' "clinically naïve restrictions on number of sessions of therapy" (p. 113), particularly specifying the misguidedness of such a treatment plan for combat-related PTSD. Yet, this appears commonplace in the literature and a study by Devilly, Spence, and Rapee (1998) employs this level of treatment. This study compared 51 PTSD-diagnosed war veterans using a mixed groups experimental design with three conditions: (a) EMDR, (b) EMDR without the eye movement component (termed REDDR and featuring use of a stationary flashing light to replace the eye movement component), and (c) standardized psychiatric support control group. Participants were screened for at least five exclusion criteria, including medico-legal claims, and a stratified random assignment was used (with some exception due to administrative complications). Considerable efforts were taken by the authors to document sample characteristics and assure that group variables were balanced across groups.

Dependent variables included standardized instruments: Beck Depression Inventory (BDI), Spielberger State and Trait Anxiety Scale (STAI), Mississippi Scale for Combat-related PTSD (M-PTSD), and Structured Interview for PTSD (PTSD-I); self-report measures: Personal Problem Definition Questionnaire (PPD), Subjective Units of Disturbance (SUD), and Credibility of Therapy Questionnaire (COT); and the physiological measures of heart rate and blood pressure. Assessments were administered
prior to treatment, at the end of the second treatment session, two weeks after the second session and again six months later (the latter two being postal administrations). After an initial history taking session, participants were seen for two treatment sessions of up to 90 minutes each. The same therapist administered both treatments and this person was trained by Francine Shapiro at Level II.

A variety of analyses were conducted on all of the pre and post measures and the authors concluded that “From pre to posttreatment, no statistically significant differences were found between the three conditions on any of the outcome measures” (p. 449). The three groups all showed improvement after treatment, with the EMDR group showing the most improvement, however there were no significant differences between groups found. By six month follow-up, treatment effects tended to wane. The authors concluded that the study did not provide support for the efficacy of EMDR with this population and that the bilateral stimulation component of EMDR was not likely to be integral to change.

Strengths of the study included the use of objective and subjective measures, comprehensive participant exclusion criteria, strong attempts to guard for potential secondary gains, and documented analyses of group equivalence on key dependent variables prior to treatment. Weaknesses included small sample size. While 51 participants is an initially impressive number, division into three groups puts power into peril and with only around 40% response rate at six month follow-up, only excessively large treatment effects would likely be detected at this interval. Further, and as noted by Shapiro (1999b), two sessions of treatment for combat-related PTSD is a very short treatment. The authors cite Shapiro’s initial (1989b) study as well as that of Wilson et al. (1995) where in one and three sessions respectively, large treatment effects were reported. These studies were not PTSD-specific, however, as the population of interest
were traumatized individuals who may or may not have met PTSD diagnostic criteria. Devilly et al. (1998) acknowledged the lack of independent evaluators or formal treatment fidelity measures as a limitation.

Perhaps the greatest compromise to the present study was the potential confound of experimental conditions. Devilly et al. (1998) hoped to assess whether or not the eye movements contribute toward treatment effects. However, the EMDR treatment minus the eye movements condition (REDDR) actually seemed quite similar to alternative forms of bilateral stimulation that are currently being investigated as variants of the EMDR protocol. If focusing on a flashing light instead of therapist hand movements has a similar neurological effect, and it seems reasonable that it could, then testing for differences between treatment groups is unlikely to find the main effects it set out to investigate.

Silver, Brooks, and Obenchain (1995) also conducted a study with war veterans diagnosed with PTSD. 100 Vietnam veterans in an inpatient VA Medical Center received either EMDR (as little as a single session of unspecified length), relaxation training (minimum three sessions), biofeedback (minimum three sessions), or control (standard VA program). The three experimental treatments were in addition to the standard psychiatric care received by the control and it appears that these extra treatments were optioned by participants. Patients were administered a set of scales called the Problem Report Form (PRF), an eight factor Likert-type self-report measure designed to assess the level of severity of Anxiety, Anger, Depression, Isolation, Intrusive Thoughts, Flashbacks, Nightmares, and Relationship Problems. This is not a standardized instrument, but the authors reported a .86 to .95 test, retest reliability.
The authors measured pre to posttreatment change by analyzing difference scores of the PRF on each of the eight factors across groups. Statistically significant differences were found favoring positive change in the EMDR condition on five of the eight factors: Anxiety, Isolation, Intrusive Thoughts, Nightmares, and Relationship Problems. None of the other groups showed statistical differences on any of the remaining factors.

This study evidences serious methodological flaws, most conspicuously the absence of a careful, thorough documentation of procedural information. There was no mention of participant exclusion criteria, nor specification of any of the experimental procedures, and little documentation of any attempts at controlling for validity. Random assignment was absent. Further, it was unclear whether or not the implementing therapist(s) were EMDR trained. At one point, the study stated, “When the study began, none of the clinical staff were trained in EMDR” (p. 339), but later on the page it stated, “EMDR was as taught by Shapiro.” Additionally, the number or level of administering therapists is unclear, except that they were at least supervised by the same experienced, doctoral level psychologist.

Further compromises included the small sample sizes of some groups. Group sizes were: Control (55); EMDR (13); Relaxation (9); and Biofeedback (6). No statistical weighting procedures were employed to address this in-equivalence. In supplementary analyses comparing more specified cross sections of groups on single factors (such as Biofeedback x EMDR), ANOVA cells had as little as two participants and obviously these are uninterpretable. Random assignment was not used which leaves open the possibility of client variables inherent in the self-selection of treatments causing confounding effects. There were no independent evaluators or any other specified
measure of treatment fidelity. Finally, the frequency and duration of treatments is virtually unreported and it is not clear why the EMDR group was permitted as little as one session while the other two experimental groups received a minimum of three sessions prior to comparisons. At the very least, this does not conform with the treatment protocol specified in the EMDR Level I Training Manual (Shapiro, 1999).

A recently published investigation by Rogers, Silver, Goss, Obenchain, Willis, and Whitney (1999) compares a single session of EMDR with exposure therapy in treating PTSD-diagnosed Vietnam Veterans. The study was designed to examine process variables related to EMDR rather than the long-term efficacy of the treatment. Twelve participants who were undergoing inpatient treatment for combat-related PTSD at a VA clinic were diagnosed with PTSD using DSM-III-R criteria confirmed by administration of the Clinician Administered PTSD Scale (CAPS). Participants were all male between the ages of 47-53 and all but one were reported to be taking some antidepressant medication during the study. Exclusion criteria included a comorbid dissociative disorder, personality disorder, psychosis, or previous treatment with exposure therapy or EMDR. After diagnostic confirmation by the CAPS interview, participants were randomly assigned to treatment conditions by a clinician blind to CAPS scores.

An independent, blind assessor conducted assessments, which included heart rate, blood pressure, SUD ratings, and the Impact of Events (IES) prior to treatment and a week after treatment. At the pretreatment assessment, participants were asked to describe a target memory which would also be used a week after the treatment at the posttreatment assessment. At posttreatment, participants were also asked to rate their therapy experience on a one to four scale. Treatment consisted of a single therapy session ranging from 60 to 90 minutes using EMDR “as described by Shapiro” (Rogers, et al.,
1999, p. 123). The exposure condition was based on the implosion procedure detailed by Lyons and Keane (as cited in Rogers et al., 1999, p. 122). Essentially, this consisted of a progressive muscle relaxation exercise with guided exposure to the traumatic memory takes place.

Dependent variables included the intrusion and avoidance subscales of the Impact of Events Scale (IES), the Subjective Units of Disturbance (SUDs), and the physiological indices of heart rate and blood pressure. Several process variables were also assessed for differences including within treatment SUDs levels (aimed at assessing comfort levels of the therapy), exposure times (culled from audiotaping sessions), and physiological reactivity (to the trauma memory). There is also a measure of self-monitored severity of intrusive re-experiencing measured by giving participants “forms for monitoring the severity of intrusive recollections, nightmares, and flashbacks related to the memory for a one-week period between their treatment and posttreatment assessment” (p. 122). However, no further details or properties of this measure are detailed.

Results showed the EMDR to have a significantly greater drop in SUD levels during treatment. Other process variables such as exposure time and treatment ratings showed no significant differences between the two groups. Posttreatment results show no significant differences between groups on the SUDs ratings. On the self-monitored severity of intrusive re-experiencing measure, the EMDR group showed significant decreases in intrusiveness after treatment whereas the control group did not (and this presumably considered established pre-treatment differences in the analysis). Both groups showed improvement on the Impact of Event Scale after treatment, although there was no differentiation between groups on the intrusion or avoidance subscales of the IES. The control group showed significantly greater physiological reactivity at posttreatment
as measured by heart rate and blood pressure in response to the targeted memory.

Rogers, et al. (1999) concluded that the rapid drop-off of SUD levels in the EMDR group is supportive of other research that found that “EMDR was more comfortable than more direct forms of exposure” (p. 128) and that the study suggests that there is a component other than exposure in EMDR that accounts for treatment effects.

Strengths of the study include pretreatment comparisons to assess for group differences unrelated to treatment which did occur on at least one variable with the EMDR group showing a significantly lower level of intrusive recollections than the control. Also strong was the use of physiological dependent measures as well as attempts to measure important process variables linked to the utility of treatment. The authors further sought to use clinically meaningful measures of change, such as the self-monitored measures of intrusion, though these were not documented adequately enough to assess their usefulness.

Weaknesses of the study include the small sample size (total n=12) as acknowledged by the authors with its related weak statistical power. Treatment integrity was not addressed in the investigation, although the authors noted that this is a preliminary study and that audiotapes will be used as a fidelity check in later analyses. However, should the EMDR be found to be non-compliant with the protocol, there obviously would be no corrections made for the current study. Importantly, secondary gains were not part of the exclusionary criteria and this is important for veteran participants who may likely receive benefits for continued disability. There is some logical question as well regarding the issue of previous treatment. Rogers et al. (1999) stated that exclusionary criteria included those with previous EMDR or exposure therapy. However, considering the chronic, inpatient sample with a long standing history of
PTSD, is it likely that these participants had no previous exposure therapy? Fortunately, random assignment was employed.

EMDR with Non-Combat Related PTSD

Unlike the previously described class of studies, non-combat, or civilian, EMDR studies earned categorization as probably efficacious in the treatment of Posttraumatic Stress Disorder (Chambless et al., 1998). This implies that the task force either found larger treatment effects or better designed and supportive studies within the non-combat published studies. Interestingly and not like the combat-related PTSD studies, it was more common for EMDR efficacy studies featuring a sexually traumatized population to feature the history of a sexually-related traumatic experience as the underlying variable rather than a related PTSD diagnosis. In other words, whereas many combat-trauma studies such as those outlined above included PTSD as a diagnostic criteria for participant inclusion in the study, the present literature search found only one controlled group study in which a PTSD diagnosis was part of the participant inclusion criteria.

Examples include the Wilson et al. (1995) investigation which combined a more general PTSD population, a population of individuals with “traumatic memories” such as Shapiro (1989b); or case study research such as Parnell (1994). Edmond, Rubin, and Wambuch (1999) conducted a controlled study of EMDR with adult survivors of childhood sexual abuse, however, participants were not necessarily diagnosed with PTSD. Similarly, Scheck, Schaeffer, and Gillette (1998) conducted a relatively large sample controlled study of EMDR with “traumatized women,” however, of the 60 participants, only 46% met criteria for PTSD. The absence of controlled sexual assault-related PTSD studies in the EMDR literature is interesting as the link between sexual
assault and PTSD has been well-documented (see Herman, 1992). Because these studies did not utilize a PTSD specific sample, the studies are not applicable to the present investigation.

Rothbaum (1997) conducted a well-controlled single group ABAA design with 21 adult female rape victims diagnosed with PTSD. This was one of the two sufficiently designed and controlled studies evidencing large enough treatment effects of EMDR for inclusion on the APA Task Force list of Empirically Validated Therapies (Chambless, et al., 1998). Participants were screened for drug or alcohol dependence and appropriateness for study by either the therapist or an independent assessor. Upon eligibility, participants attended an initial evaluation and were administered the PTSD Symptom Scale (PTSD-SS), the Treatment, Legal, and Drug Update Interview (TLDUI), Impact of Event Scale (IES), Rape Aftermath Symptom Test (RAST), Spielberger State and Trait Anxiety Inventory (STAI), Beck Depression Inventory (BDI), and the Dissociative Experience Scale (DES). The therapist completed the Assault Information Interview (AII) and Trauma History Checklist (THC) during the first treatment session. Following the initial evaluation, participants were randomly assigned to either the EMDR group or a wait-list control.

Treatment consisted of three 90-minute EMDR sessions administered by the author who received training at levels I & II and has expertise in treating PTSD. It is unspecified whether or not the wait-list control group (or the EMDR group for that matter) received any outside treatment during the investigation. Since participants in the control group received treatment after post-testing, follow-up assessment data are not available.
The EMDR group showed significant improvement on the PTSD Symptom Scale, Impact of Event Scale, and Beck Depression Inventory at posttest. There were no significant differences between groups found on the Rape Aftermath Symptom Test, Spielberger State-Trait Anxiety Inventory, or Dissociative Experience Scale despite larger decreases in fear, dissociative experiences, and anxiety in the EMDR group. The authors further reported that only 10% of the EMDR group continued to meet the full criteria for PTSD following treatment whereas 88% of the control group did. There were no post-hoc analyses reported specifically on the follow-up assessment. However, the data reported show lower mean levels of PTSD symptoms on the PSS and IES between posttest and follow-up, nominally higher BDI scores between posttest and follow-up, slightly higher STAI scores at follow-up, lower DES, and lower RAST scores between posttest and follow-up in the EMDR group.

This was a methodologically strong study as pre-treatment group differences were analyzed and found to be not significant. Treatment fidelity was adequately controlled for as an EMDR facilitator designated by Dr. Shapiro was enlisted to view and rate randomly picked treatment videotapes. Statistical analyses conducted on this independent assessor's ratings indicated that the EMDR treatment was deemed to be in acceptable conformity with the treatment protocol. Rothbaum (1997) also abided by the recommendations of Dr. Shapiro to utilize three 90-minute sessions (plus initial evaluation) in order “to allow for the consolidation of treatment effects” (p. 325). Further strengths include standard measures of PTSD and the inclusion of a follow-up diagnostic count of how many participants continued to meet PTSD criteria as this is a clinically meaningful measure.
Weaknesses of the study included small sample size with consequently low statistical power. The author addressed this as well as the fact that there was no comparative treatment studied. Additionally, more objective measures (i.e. physiological assessments) would have been valuable in order to see if there are pre-post group differences outside of self-report data in a PTSD-diagnosed sexual-assault related population (as there were none in the combat-related PTSD samples).

There were a variety of studies that focus on a mixed-origin PTSD-diagnosed population, meaning participants whose DSM-IV (APA, 1994) PTSD Criteria A (p. 424) include different categories of traumatic experience. While such investigations risk losing robustness in their generalizability, the broader-based inclusion criteria tend to create larger sample sizes with greater statistical power.

Such was the case with a controlled comparative study of EMDR with PTSD diagnosed individuals in an HMO setting. Marcus, Marquis, and Sakai (1997) studied 67 individuals in the Kaiser Permanente HMO in California who were referred to the study from their mental health providers after being diagnosed with PTSD according to the DSM-III-R. An independent evaluator also assessed for PTSD in order to meet inclusionary criteria and this included the Dissociative Experiences Scale and Dissociative Interview Schedule. Exclusionary criteria included psychosis, obsessive-compulsive or dissociative disorders; substance dependency; active suicidality; and current or pending litigation involvement. Participants taking and stabilized on psychotropic medication prior to the study were not excluded. Traumatic events of origin included physical and/or sexual assault, accident and injury, earthquake, and witnessing the death of a loved one.
Upon eligibility, participants were randomly assigned to either the EMDR or Standard Care groups using a blocked (sets of four) randomization procedure. Statistical analyses showed a significant difference on one of nine pretreatment assessments between groups and that was the Global Assessment of Functioning which was higher for the EMDR group. Participants in the control group were returned to their mental health providers for standard PTSD treatment as delivered by the professionals within the HMO. These professionals were described as "experienced clinicians using their preferred modalities for treatment of PTSD" (p. 308). The standard care milieu consisted of any one or more of the following: individual (cognitive, behavioral, or psychodynamic) psychotherapy; group therapy; medication; and in some instances day-treatment (one case) and brief inpatient hospitalization (two cases). Like the EMDR group, sessions were 50 minutes, however, unlike the EMDR group, there were no restrictions on the number of sessions administered over the course of the investigation.

EMDR treatment consisted of 50-minute sessions with three psychologists with Levels I & II training and varying levels of experience. In both the EMDR and SC groups, there were no set number of treatment sessions. Instead, individuals were treated, according to the authors, until they no longer met the DSM-III-R criteria for PTSD or the study ended. Assessments were administered at three points: Pretest, following three mental health visits, and posttreatment. Evaluations were conducted by an independent evaluator, however, this person was not blind to treatment conditions. Each evaluation included an interview by the evaluator from which scores on two measures were derived (Dissociative Experiences Scale and Dissociative Interview Schedule) and client self-report measures including the Modified PTSD Scale (MPTSD), Beck Depression Inventory (BDI), Spielberger State Trait Anxiety Inventory (STAI), Impact Of Events
Scale (IOS), Symptom Checklist-90 (SC-90), Positive Symptom Distress Scale (PSD), SUD, and Global Severity Index (GSI).

A multitude of analyses were made on the outcome measures, including differences between group means on the various dependent variables at the three evaluation points and measures of the rapidity of change by treatment. Overall, the EMDR group was found to have reduced levels of PTSD symptomatology at the three session and posttreatment intervals as measured by IED, MPTSD, BDI, STAI (anxiety subscale), SUD, and SCL-90. GAF scores were lower for the EMDR group at the second and third assessment intervals. No differences were found between groups on the Trait or State subscales of the STAI. Additionally, the EMDR group showed a significant difference in fewer individuals meeting PTSD criteria after treatment and in the utilization of fewer total number of sessions needed to complete treatment. On most measures, the greatest levels of change took place between pretreatment and the second assessment interval, which the authors conclude indicates that EMDR is rapidly effective.

The greatest strength of this study lies in its external validity. Throughout the design and implementation, the authors sought to study the efficacy of EMDR in treating PTSD in a realistic HMO setting. Compared with the vast majority of studies, the sample size was large, contributing better levels of statistical power to detect treatment effects. Additionally, the use of a standard care control group that mirrors typical clinical care for PTSD with an EMDR condition as would likely be practiced in most HMO settings made for a realistic and interpretable comparison. Multiple outcome measures that included some standardized measures and PTSD-specific assessments were sound and utilization of clinically meaningful information, such as the number of sessions likely to affect change for an individual, was helpful. Groups did not differ prior to treatment and the
study had thorough exclusionary criteria. The fact that the PTSD diagnosis was made by two separate parties lent confidence towards population validity.

The Type I vs. Type II dance, however, also represents some of the study’s weakness, particularly in that the 50-minute sessions are out of compliance with the 90 minute sessions recommended in the EMDR protocol. This was duly noted by the authors (and in a sense acts as an experimental variable) and the fact that treatment effects were greatest within the first three 50-minute sessions is an important finding for future studies in a managed care environment. Nonetheless, this condition combined with the fact that there were no treatment fidelity checks built into the study threatened its internal validity. A further compromise was the authors’ decision to not keep the independent evaluator blind to treatment condition (though it was specified that the evaluator was kept “naïve” to EMDR procedures and specific methods of standard care). The authors stated that this was necessary “due to the strong subject response to EMDR” (p.309). The independent evaluator, whose role was to score all assessments and conduct the three assessment interviews, was a graduate student. Like in the Jensen (1994) study, this could present potential confounds, particularly in this non-blind study where she was also an employee of three years with the HMO funding the study.

Montgomery and Ayllon (1994) conducted a multiple baseline dismantling study of six PTSD diagnosed participants, all whose trauma was of a civilian nature. In addition to testing EMDR’s efficacy across subjects on both objective, physiological measures and subjective, self-report measures, the study attempted to compare the bilateral stimulation component through its dismantling design.

Participants included six adults referred for outpatient treatment after being diagnosed with PTSD and the study offers a brief description of each participant,
including details of their trauma and past treatments. Inclusion criteria consisted of a PTSD diagnosis, however, the origin of these diagnoses is not reported. Exclusion criteria included screening for comorbid psychosis or personality disorder, substance abuse, serious cardiovascular problems, and suicidal homicidal ideation. Participants taking psychotropic medication or receiving outside psychotherapy were excluded, however, there was no screening for secondary gains reported.

Treatment consisted of three sessions of a non-saccade control which was described by the authors as including “All aspects of the EMD treatment procedure except for the saccadic eye movements” (p. 221). In the second phase of treatment, the participants received six sessions of EMD(R) and the authors document details of both phases of treatment. Sessions lasted for approximately 75 minutes each and all participants received all nine sessions with the exception of one whose self-reported units of disturbance (SUDS) were reduced to levels of one after a single EMD session.

Dependent variables included the Structured Interview for PTSD (SCID); Structured Interview for DSM-III Personality Disorders (SCID-II); Beck Depression Inventory (BDI); Anxiety Disorders Interview Schedule-Revised (ADIS-R); Fear Survey Schedule (FSS); Subjective Units of Disturbance (SUDs); and the physiological indices of blood pressure and heart rate. Additionally, the authors took the clinical measures of days with intrusive thoughts and dreams from participants as between-session data. T-tests for dependent samples were administered and results show statistically significant change across the verbal measures, but not on physiological indices. Because of the multiple baseline design, the authors were able to look at change across sessions and determined that in general, treatment effects “were achieved over the course of six EMD treatment sessions not the one or two previously reported by Shapiro” (p. 227). Finally,
results suggested that the bilateral stimulation component of EMD did significantly contribute to treatment effects as decreases in PTSD symptoms with the eye movements were significant whereas without the eye movements there were decreases, but generally not at a level of significance.

The Montgomery and Allyon (1994) study was ambitious in its goals and as a result provided information on a variety of levels. The dependent variables utilized were comprehensive and the investigation was well documented. The descriptive paragraph was helpful in getting a clinical picture of the presenting PTSD of participants and the authors further mentioned particulars about the participants at different points in the article. Other strengths included the multiple baseline assessments which allowed for across session measurement of change and the use of objective measures taken in conjunction with self-report indices. The six 75-minute sessions resulted in a more lengthy treatment than many studies and the measures over time helped to illustrate dose response curve trends. The dismantling design allowed for a comparison of the EMDR techniques with and without the eye movement component.

Unfortunately, dismantling designs also have inherent weaknesses, the foremost being that participants were subject to the identical EMD treatment, without eye movements, just before the EMD(R) treatment. Therefore, it can not be ruled out that the first phase of treatment did not have an effect on the second phase. Other weaknesses include the small sample size with consequently low statistical power. While the small sample size allowed for a personal, qualitative aspect of an otherwise quantitative study, the ambitious research questions and multiple dependent variables warranted a larger sample. An additional limitation in this early study is that the credentials of the
implementing therapists are not specified, nor is it delineated whether or not the administering therapists or evaluators were blind to treatment.

Lazgrove, Triffleman, Kite, McGlashan, and Rouseaullie (1998) conducted a two-part study with chronic PTSD-diagnosed individuals, the pre-pilot published at this time. There were different origins of traumatic experience, however, all were civilian and diagnoses were confirmed through administration on the Clinician Administered PTSD Scale (CAPS). Other exclusionary criteria included psychosis, personality disorders, severe depression, certain ophthalmologic problems, and the inability to contract for safety in tandem with the treatment. Seven of the eight participants completed the treatment which consisted of three 90-minute EMDR sessions with pretreatment, one week, and two month follow-up assessments. Outcome measures included standardized and non-standardized instruments which were administered by a non-treating assistant. Dependent measures included: Dissociative Experience Scale (DES); Clinician Administered PTSD Scale (CAPS); Impact of Events Scale-Revised (IOS); Beck Depression Inventory (BDI); Substance Use History on the Addiction Severity Index (ASI); Symptom Checklist-90-Revised (SC-90); and SUDS. Two Level II trained therapists administered the treatment sessions, all of which were videotaped and reviewed by Dr. Shapiro and judged to be authentic.

The authors reported significant differences on all measures between pretreatment and one week posttreatment. Most of these differences (8/12) remained significant two months after treatment. Additionally, none of the participants continued to meet DSM-III-R criteria for PTSD a week after treatment ended, nor two months later. The authors reported that the median number of endorsed symptoms for PTSD decreased to zero (Criterion B, C, and D).
This study was conducted as a pilot for a larger, more rigorously controlled randomized study, but has strengths in its own right. Treatment fidelity was insured by expert-reviewed videotapes of sessions and dependent variables included both standardized and clinically meaningful measures. Documentation of type of trauma and substance use history was helpful. A non-treating research assistant administered measures, however, it is not specified whether or not this person was blind to the nature of the treatment.

Small sample size is an obvious limitation of the study as is its lack of a comparison treatment. Exclusionary criteria did not include screening for secondary gains and there were several participants who were actively receiving concurrent psychiatric treatment at the time of the study. Although a range of PTSD-specific and related dependent variables were employed, there were no objective, physiological data utilized. Further, the authors did not report any outcome measurements beyond the median scores of dependent variables and this made for vague interpretation. The authors further noted the “open-label design” as well as the potential for non-specific treatment effects (since there was no comparison group) as limitations to the study (p. 607).

Vaughan, Wiese, Gold, and Tarrier (1994a) and Vaughan, Armstrong, Gold, O’Connor, Jenneke, and Tarrier (1994b) conducted two separate studies of EMD (as described in Shapiro’s protocol prior to 1995) with PTSD-diagnosed individuals in Australia. Vaughan, et al. (1994a) aggregated 10 consecutive case studies for paired-t-test comparisons (thereby organized as group comparison research) of three PTSD symptom categories. Participants represented a range of traumas and were referred for treatment at a specialized PTSD clinic. Of the three men and seven women (ages 18-45),
all but two had a long history of PTSD symptoms. Concurrent medication was permitted and no other inclusionary or exclusionary criteria are mentioned.

Participants were screened for assessed for PTSD symptoms before and after treatment, and at 2-3 months follow-up using the PTSD Structured Interview and the Hamilton Rating Scale for Depression. Interviews and assessments were conducted by one of the authors and it is not specified whether or not this investigator was blind to treatment issues. Treatment was administered by the chief author (K.V.) whose level of training is also unspecified in the present article, but said to be trained by Shapiro in another article of the same year (1994b). Treatment consisted of between one to six sessions for each participants and sessions ranged from 45-60 minutes. Most sessions were videotaped and the authors described applications of the EMD procedure in the procedures section.

The authors analyzed data by the three general PTSD categories of re-experiencing, avoidance, and hyperarousal. The analysis entailed paired t-tests on the Hamilton Rating Scale for Depression and “mean DSM-III-R symptom scores” (p. 536), however the authors did not describe procedures for categorizing symptoms on the dependent measures. In all three symptom areas, significant improvement was found between pre and posttreatment. Mean depression scores also showed significant improvement following treatment. Follow-up assessments taken between two and three months after treatment maintained significant gains in the areas of total re-experiencing and total avoidance, but not for total hyperarousal or depression scores. Calculation of mean SUD level before treatment v. after treatment was 8.7 and 2.3 respectively, however, statistical analysis was not performed. The authors provided multiple analyses for symptom change in each of the cases, noting interesting patterns particularly
regarding follow-up data. This will not be presented here as it is not relevant to the present analysis.

The effort to tie dependent variables to PTSD symptomatology was a strength of the study. Further, the authors provided detailed statistics to document specific PTSD symptom change as a result of treatment. Dependent variables attempted to avoid the early problem of relying solely on the SUDs and VOC measures. However, the study had a number of shortcomings and inadequate reporting of dependent variable information was among them. It was unclear how symptom cluster categories were derived and it was also unclear why SUD scores were not analyzed for significance (particularly since means were reported in the results section).

Based on detailed description of treatment implementation in the procedures section, it appeared that there was some departure between the EMD administered in the investigation and that in the manual. However, it is difficult to assess the degree to which this is acceptable relative to individual differences across therapists. Although videotapes were taken at some points in the investigation, there was no use of the tapes documented as treatment fidelity checks other than to count mean number of sets of eye movements. The examiner administering the assessments was different than the implementing therapist, however, it was not stated whether or not the assessor was blind to the treatment. Another weakness of the Vaughan et al (1994a) study concerned the fact that two of the cases did not adequately fit one of the DSM-IV’s PTSD criteria (in that the traumas occurred a month prior to the study). Since all symptomatology criteria as assessed by the Structured Interview for PTSD and the DSM-IV were fulfilled, and the participants were seeking treatment at a hospital PTSD clinic, the participants were included nonetheless, but caution should be exercised in external generalizability.
Vaughan, et al. (1994b) compared EMD to Image Habituation Training (IHT) and Applied Muscle Relaxation (AMR) in treating 36 PTSD-diagnosed outpatient participants. Individuals were screened for psychosis and personality disorder and diagnosis was confirmed through the PTSD Structured Interview (SI-PTSD). Following initial assessment, participants were randomly assigned to one of the three treatment groups or a wait-list control who was reassessed after two weeks and then provided with treatment. Treatment length was four sessions and the average session length was 50 minutes. Therapist characteristics were not reported except that therapists were trained by Dr. Shapiro in the EMD condition. A brief description of each treatment modality was reviewed and will not be reviewed here.

Dependent variables included analysis of categorical PTSD symptoms on the SI-PTSD, the Hamilton Rating Scale for Depression (HRSC), Anxiety Disorders Interview Schedule-Revised (ADIS-R), Spielberger State Trait Anxiety Inventory (STAI), Beck Depression Inventory (BDI), and the Impact of Events Scale (IES). All three treatments were found to be superior to the wait-list control. The authors reported that the lack of MANOVA interactions failed to demonstrate “consistent superiority of any one treatment over the other” (p. 288), however, they noted that the EMD group had a greater reduction of total PTSD symptoms (45% compared to 25% and 34%). It was further noted that the trend for clinical improvement favored the EMD condition over the more established treatments tested.

Strengths of this study are its inclusion of popular PTSD treatment comparison groups as well as a wait-list control. The authors attempted to make dependent variables clinically meaningful and did this by using standardized measures in a symptom category
analysis. This allowed for digestible interpretation of outcome measures and the ability to compare popular treatments in a meaningful way.

Limitations include the authors' acknowledgment that treatment length was arbitrarily assigned to four sessions and the duration does not comply with the 90 minute sessions recommended by Shapiro (1999). The IHT and AMR groups also had between 40-60 minutes of homework between sessions, but this is not part of EMD's treatment protocol. Participants were screened, but not for secondary gains. Sample size was small (between 11-13 participants in experimental groups) and power was low. There was some mention of an independent rater asked to guess treatment categories posttreatment (p. 288), however, no formal fidelity checks are documented.

Renfrey & Spates (1994) investigated EMD with 35 PTSD-diagnosed individuals of various trauma origins. In order to assess the role of bilateral stimulation, they compared EMD to two similar groups, an Automated EMD (AEM) condition and an AVA or visual fixation treatment that otherwise used the EMD protocol. Both experimental comparison groups were meant to differ from EMD only with respect to the visual task employed in the treatment. The AEM treatment consisted of focusing on two alternating lights and the AVA condition consisted of focusing on a light bar that flickered only periodically (to help maintain visual attention). Participants were screened for psychosis, active PTSD symptomatology, secondary gains, and adequate emotional stability to withstand the potential rigors of treatment. The CAPS was used to confirm PTSD diagnosis and participants were randomly assigned to one of the three treatment groups after an initial assessment. It was reported that groups were balanced in terms of descriptive factors (i.e. age, ethnicity) and groups ranged between seven and eight members. Dependent measures included heart rate, SUDS, Clinician Administered
PTSD Scale (CAPS), Impact of Events Scale (IES), and the Symptom Checklist Revised (SCL-90-R).

Treatment consisted of between two and six sessions, however, criteria is not specified beyond this except to say that the average length of treatment did not differ between groups. The two implementing therapists were trained in EMD and being the authors, were not blind to the study. Results show that while negative heart rate change scores did decrease for all conditions following treatment, there were no differences between groups. Similarly, SUD ratings went down, but with no between-group differences. Likewise, there were no differences on either measure between posttreatment and follow-up. Positive cognitions, as assessed by VOC ratings, increased significantly after treatment, but again there were no differences between groups. The authors concluded that the EMD and two alternative treatments all showed positive improvement in treatment of PTSD symptomatology. They further concluded that while EMD is an effective treatment for PTSD, its efficacy was not explained by the eye movements.

Screening procedures appeared to be thorough and random assignment was employed. The use of standardized and physiologic instruments in addition to the SUDS and VOC was comprehensive and facilitated a diagnostic-specific assessment of outcome goals. It was interesting and useful that the authors used the standardized instruments to collect a single measure of whether or not each participant met the DSM-III-R criteria for PTSD after treatment instead of reporting multiple, instrument-specific results.

This study with the two alternative groups used to assess EMDR’s bilateral stimulation component, attempted to more specifically investigate the function and contribution of therapist-induced eye movements. However, the alternative treatments
employed, particularly the AEM condition with its alternating flashing light, could be considered forms of bilateral stimulation rather than controls for them. This resulted in a confounding of condition which was the greatest weakness of the study.

Experimental groups ranged from seven to eight participants and this small size resulted in low statistical power. While on the one hand this division of the participant pool allowed for a more detailed comparison related to bilateral stimulation, on the other hand it resulted in groups too small for confident generalization. Other weaknesses included a lack of treatment fidelity checks, none being reported in the study.

Forbes et al. (1994) investigated EMDR with PTSD-diagnosed individuals with the purpose of using objective criteria as dependent measures since EMDR’s earliest studies tended to utilize self-report instruments. Eight PTSD-diagnosed participants were assessed for PTSD using the Structured Interview for PTSD (SI-PTSD) and Structured Clinical Interview for DSM-III-R (SCID-NP). Exclusion criteria included current alcohol or substance dependence, psychosis or organic mental disorder or depression severe enough to require immediate treatment. An ABAA design was used with a three month follow-up period. Two registered psychologists with EMDR training were used and dependent measures also included the Symptom Checklist Revised (SCL-90-R), Global Severity Index (GSI), Impact of Events Scale (IES), Beck Depression Inventory (BDI), SUDS, and a physiologic assessment of muscle tension as measured by EMG electrodes. Treatment consisted of four 90-minute EMDR sessions in addition to an initial assessment interview. Measures were grouped and reported as three categories of PTSD symptomatology: Overall PTSD symptoms, intrusion, avoidance, and hyperarousal.
The authors reported significant effects for each of the four categories with significant change occurring between pretreatment and immediate posttreatment, and no change between posttreatment and three-month follow-up. Half of the participants continued to meet PTSD criteria at posttreatment and follow-up. Various proportions of the participants continued to meet some of the diagnostic criteria for PTSD following treatment (see p. 116). Significant improvement was also found in depressive symptoms. The authors reported that "Consistent findings were obtained across clinical interview, self-report, and psychophysiological measures" (p. 118), however, there were no analyses offered on the EMG dependent variable due to complications.

This was a disappointment as a key purpose of the study was to test EMDR with more objective dependent measures. Small sample size is acknowledged as a limitation as is the lack of placebo or comparative treatment. Additionally, there is no mention of treatment fidelity checks. Inclusion of PTSD specific cluster symptoms were meaningful and unlike many studies, the investigation complied with the recommended 90 minute sessions. The authors included several descriptive sub-analyses and speculation, such as dose-response curves and participants’ ability to visualize a memory, which spur interesting questions. However, the small sample size and large-issue questions keep these questions unanswered.

The most recent between-groups study of non-combat related PTSD survivors treated with EMDR was by Devilly & Spence (1999) whose previous investigation was described in the preceding section. The authors compared an EMDR treated group to a Trauma Treatment Protocol (TTP) variant of Cognitive-Behavior Therapy, an important study, say the authors because the two treatments are "vying for supremacy without a controlled study actually comparing them" (p. 131).
Twenty-three self-referred (through media and other solicitation) participants were randomly assigned to either the EMDR or TTP groups. All participants met the DSM-IV diagnostic criteria for PTSD (as confirmed by the PTSD-Symptom Scale Self-Report), had no current psychosis or brain dysfunction, no reported suicidal ideation, nor ongoing medico-legal claim related to therapy. Participants were not receiving outside therapy for PTSD, however, some were using psychotropic medications monitored by a psychiatrist. Post assignment checks on such variables confirmed that there were no pretreatment differences between groups. Two implementing therapists, both with Level II EMDR training and at least one with experience in CBT as well, administered all treatment sessions.

Data were collected at pretreatment, posttreatment, two-week follow-up, and three-month follow-up points. Dependent measures included the trait portion of the Spielberger State-Trait Anxiety Inventory (STAI); Beck Depression Inventory (BDI); Revised Symptom Checklist-90 (SC-90); Subjective Units of Disturbance (SUDs); Personal Problem Definition Questionnaire (PPD); Civilian Mississippi Scale for PTSD (CMS); Impact of Event Scale (IES); PTSD Symptom Scale Self-Report (PSS-SR); PTSD Interview (PTSD-I); Credibility/Expectancy Questionnaire (CEQ); and the Distress Evaluation Scale for Treatment (DEVST).,

Treatment consisted of nine sessions for each condition and treatment sessions were approximately 90 minutes in duration. EMDR was conducted in accordance with manualized procedures. The CBT was detailed by the authors (p. 141) and essentially was comprised of a structured protocol integrating prolonged exposure, stress inoculation therapy, and in-depth cognitive therapy. Treatment fidelity measures included
videotaping with assessment by independent evaluators and statistical analyses performed on ratings.

Numerous analyses were conducted on the multiple assessment measures employed by the study. For purposes of brevity, results at posttreatment will be addressed with only generalizations made by the authors in reference to significant follow-up trends. Repeated measures of PTSD, including the IES, PSS-SR, and CMS, showed no main effect for treatment condition. However, there was a significant interaction for condition x time on PTSD symptom measures with TTP more effective in reducing PTSD symptomatology at all points with significant efficacy over time (i.e. reductions were greater on assessments and lasted relative to EMDR condition). A similar trend was found for the associated measures of distress. No significant main effects were found on the BDI or Y2. The three measures of SCL-G, PPD, and SUD all showed effects for time with the SCL-G showing an interaction for condition x time with TTP showing greater effects than EMDR, especially over time.

In an estimate of meaningful clinical change, the authors analyzed reductions in PTSD symptoms based on the PTSD-I and PSS-SR. Accordingly on the PTSD-I, 10/12 TTP participants (83%) no longer met PTSD criteria after treatment whereas 4/11 EMDR participants (36%) no longer met PTSD criteria. A similar trend was observed on the PSS-SR, however, the difference was not statistically significant. Chi-square analysis showed a higher percentage of participants in the TTP group improved on PTSD measures (PSS-SR, IES, CMS, and PTSD-I) compared to the EMDR group. Subsidiary analyses were performed showing no differences in treatment distress (DEVS-T) between groups, even though twice as many participants dropped out of the EMDR group. At
three-month follow-up, the TTP group showed greater effect sizes on both self-report PTSD measures and associated measures of pathology.

The authors concluded that TTP was more statistically and clinically significant than EMDR in exacting therapeutic change in PTSD symptomatology and more effective, though to a lesser extent, in associated psychopathology symptoms (p. 153). The authors stated that the superiority of TTP over EMDR “became even more pronounced from pretreatment to three-month follow-up” (p.153) and that this was with reasonable effect sizes and a high power. The authors question common assertions that EMDR is “kinder to patients” (p. 154) based on the higher EMDR drop-out rates and no differences found between groups on distress measures.

This study was one of the few to utilize the recommended 90-minute treatment sessions for EMDR and further, used a total of nine sessions per experimental group. Frequency and duration of treatment has been a common criticism made by Shapiro (1996). Other strengths include the detailed documentation of CBT procedures, independently rated videotapes of both therapies with quantitative difference analysis, adequate inclusion/exclusion criteria, and a variety of dependent variables. Overall, this investigation was a well-controlled, well-designed group difference study that compared two rival treatments instead of a single treatment with control. Follow-up data were useful and the high compliance at follow-ups (100% at both intervals) made for relatively robust follow-up comparisons. The authors further state that the use of a postal follow-up measure prevented interference by demand characteristics, a frequent criticism of dependent variables associated with EMDR research. Attempts at measuring clinically meaningful symptom change were employed and the focus on diagnostic symptoms was useful. Also useful was the additional effect size information reported. The use of
properly trained therapists was important, however, it is unclear why Therapist A treated a disproportionate number of participants compared to Therapist B (20 versus three).

While the use of multiple dependent variables is advantageous on many fronts, the 11 used here became confusing in the interpretation of their meaning. Grouping according to PTSD symptomatology and psychopathology symptoms was helpful; however, more helpful would be fewer self and therapist report measures and the inclusion of some physiological measures. Of course, larger sample size would be best, however, the authors calculated adequate statistical power notwithstanding. The authors discussed some study limitation caused by aspects of TTP that were altered from earlier studies.

Another recent study of the efficacy of EMDR with non-combat related PTSD was conducted by Levin, et al. (1999). Although this article was published as an early look at an upcoming within-group design study with only the results of a single participant reported, data from all participants was received in a personal communication with the authors (Levin, 01 & Hopper, 01) and was used in the present investigation.

Levin et al. (1999) were primarily interested in the neurological change associated with clinical change as a result of EMDR treatment and therefore used technologically advanced dependent variables such as Single Photon Emission Computed Tomography (SPECT) in addition to more common measures of PTSD symptoms. Although the article states that there were six subjects in the study, there were a total of 12 participants, however, not all participants had complete data sets for reasons documented by the authors. Participants were outpatient volunteers referred to a trauma clinic whose trauma included a range of events. PTSD diagnosis was corroborated by administration of the Clinician Administered PTSD Scale (CAPS). Exclusion criteria included
psychosis; a high score on the Dissociative Experience Scale (DES); ophthalmologic problems; and medical complications. Participants were screened and excluded for secondary gains and were permitted to remain in outside therapy or pharmacological treatment if it had been ongoing for a minimum of three months.

Treatment consisted of three to four 90-minute EMDR sessions administered by two senior psychologists with two and five years' experience with EMDR. It was not specified whether these were the primary researchers or if not. Additionally, an individualized script portraying the most traumatic memory subject to the treatment was developed for each participant. This was done with the assistance of a research assistant and was audiotaped and used at the points in assessment when neuroimaging, that is SPECT measurements, were taken.

Dependent variables included the SPECT data; intrusive, avoidant, and arousal measures of the CAPS; and heart rate (HR). Specific information regarding the SPECT equipment was provided by Levin et al. (1999, p. 165). Other measures used for baseline data and screening for comorbid diagnoses, but not included in outcome data include the DES; the Hamilton Rating Scale for Depression (HRS); Impact of Event Scale (IES); Traumatic Memory Inventory (TMI); and Rorschach technique.

Results suggest that the anterior cingulate gyrus and the left frontal lobe were areas of hyperactivity during the post-treatment EMDR relative to pretreatment levels of activity identified by SPECT images. Additionally, significant decreases in PTSD symptoms were found on the CAPS and heart rate showed a significant decrease after treatment.

The unique psychobiological perspective of this study is a contribution towards explaining the neurological underpinnings of the EMDR process. This is particularly
helpful in posing questions and directions for future research in an area of EMDR that
receives much criticism from opponents: That is, if EMDR does work, why does it
work? This study had other strengths as well including a very comprehensive screening
battery of both psychometric measures and participant information of a more qualitative
nature. Assessments in general were varied and included both verbal self-report and more
objective physiological measures. The research was also comprehensive in that two
baseline assessments on some measures were taken to assure reliability following a long
time interval after selection and before beginning treatment.

Also comprehensive in the study is the detailed description of the psychological
processing of traumatic experience laid out by the authors to better understand the
research objectives and role of neuroimaging. This not only explains the psychological
experience of trauma to the reader, but allows for understandable connections with the
conclusions drawn later in the discussion section. Treatment fidelity was facilitated by
videotaping sessions and having them reviewed by another senior EMDR clinician. It is
not known, however, if this was an independent evaluator.

Weaknesses of the study include the small sample size as is customary for such
expensive and time-consuming individualized techniques as neuroimaging. Additionally,
it was not specified who took assessments data and whether or not this person was blind
to the investigation. While the trauma script was detailed in its development and
purpose, it was unclear how exactly the script was incorporated into the manualized
nature of EMDR. Lastly, it is unclear why many of the baseline assessments, such as the
BDI and HRS were not used as posttest variables since these instruments measure PTSD
related symptoms and could have provided an even richer picture of the outcomes of the
participants.
There was an unpublished study completed in a primary care setting in the UK that compared EMDR to both exposure and a wait-list control. “A controlled comparison of EMDR versus Exposure plus cognitive restructuring versus waiting list in the treatment of Posttraumatic Stress Disorder” (Power, McGoldrick, & Brown, 2000) was acquired through a personal communication with one of the authors (McGoldrick, 2001). However, it was reported to be submitted to the Journal of Consulting and Clinical Psychology upon completion of remaining follow-up data (p. 22).

A total of 105 PTSD diagnosed participants were randomly assigned to three groups: EMDR; Exposure with cognitive restructuring; and a wait list control. After attrition, this resulted in posttesting groups of EMDR (28); Exposure (20); and Wait-list control (24). Following the individualized protocols of the respective therapies, the number of session varied between groups with a maximum of 10 sessions administered over a 10 week period. Number of sessions was used as a variable with the EMDR group requiring fewer treatment sessions (mean = 4.2) than the exposure plus restructuring group (mean = 6.4).

Implementing therapists consisted of both masters level and licensed (doctoral) level clinicians trained in both treatment modalities. Participant information was limited, but generally consisted of outpatient participants within a primary care network (more descriptive information will be available upon publication of the study). Blind raters were used at pre, post, and midpoint testing.

Dependent variables included: the Clinician-Administered PTSD Scale (CAPS); Impact of Events Scale (IES); a self-report version of the Structured Interview for PTSD (SI-PTSD); Montgomery Asberg Depression Rating Scale (MADRS); Hamilton Anxiety Scale (HAM-A); Hospital Anxiety and Depression Scale (HADS); and Sheehan
Disability Scale (SDS). In general, it was found that the EMDR and Exposure groups showed a significant decrease in PTSD symptoms following treatment over the wait list control which showed no change. However, there were no significant differences found between the EMDR and Exposure groups with the exceptions of shorter duration of treatment required for change in the EMDR group \( t = 2.7, p < .05 \) and a greater reduction in anxiety and depression symptoms for the EMDR group on the HADS assessment.

This study had many strengths, among them the combination and variety of self-report and clinician-administered dependent variables. The absence of physiologic data, however, was unfortunate. The authors did, however, provide indices of meaningful clinical change to supplement indices of statistical dependent variable change. The study had larger sample sizes than most studies located and as such had increased statistical power. Use of both a regarded alternative treatment group and a wait-list control made for good comparisons.

An added strength in this study was the manner in which results were reported. There were numerous tables illustrating the dependent variable descriptive statistics and analyses. Additionally, the authors broke down four sub-hypotheses (outside of the efficacy comparison) and re-addressed these upon completion of the analyses. These important and practical inquiries were: Which treatment works best? Which treatment works quickest? Which treatment produces maintained gains at follow-up? And what symptoms does each treatment preferentially address? The latter question was often neglected in other studies and despite finding no differences here, refers to variables like depression vs. anxiety vs. hyperarousal etc.

Weaknesses of the study are primarily attributed to the fact that the paper utilized was a summary report for the Scottish Home and Health Department and as such, omitted
some descriptive data that will be available in the ultimately published article.

Nonetheless, this included a lack of information regarding inclusion/exclusion criteria; other participant information; detailed procedural information; and discussion of study limitations and future directions. Additionally, there was no mention of treatment integrity measures taken beyond the use of blind assessors. As mentioned previously, dependent variables were limited to verbal measures.

There were a number of thesis and dissertation studies located that tested the efficacy of EMDR with a PTSD population. The vast majority of these were case study research and therefore inapplicable to the present investigation (Vandeusen, 1999; Coleman, G.L., 1999; and Reyes, 1999). There was one study found that empirically tested EMDR with PTSD diagnosed individuals in a between groups pre and posttest design.

Carpenter (1999) compared two groups of PTSD diagnosed residents of a California battered women’s shelter. Participants’ diagnosis was confirmed using a questionnaire and the Impact of Events Scale (IES) and randomly assigned to the EMDR plus shelter condition or the shelter-only condition. The Dissociative Experiences Scale (DES) was also used to both corroborate the PTSD condition and exclude potential participants with significantly high tendency toward dissociation. The shelter condition consisted of individual and group counseling provided as part of the residency and all participants received this. Therefore, the only reported difference was that the EMDR + shelter group additionally received EMDR.

EMDR was administered by a licensed clinical social worker with Level I EMDR training and “Experienced with treating battered women” (p. 21). Treatment consisted of a two hour history and pre-assessment session, twice weekly EMDR sessions were
administered up to seven total sessions with an average of four sessions. The average length of stay in the temporary shelter is 45 days and participants were assessed on dependent measures upon entry and departure.

Dependent variables included the IES, the State-Trait Anxiety Inventory (STAI), the Beck Depression Inventory (BDI), the Subjective Units of Disturbance Scale (SUDs), the Validity of Cognitions Scale (VOC), and the Conflict Tactics Scale (CTS). Comparative within-group pre posttest statistical analyses were conducted on all measures except the SUDs and VOC. The EMDR condition showed a significant decrease in PTSD symptoms on the IES, STAI-S, STAI-T, and BDI. Shelter-only participants also showed significant reduction of symptoms on the IES, STAI-S, and BDI, but not on the STAI-T.

The study had strengths including a detailed description of participant data, utilization of a trauma-specific setting related to the relevant precipitating trauma, confirmation of PTSD diagnoses with psychometric data, and provision of a realistic comparison group that received treatment typical for the presenting problem. Fidelity checks were made by audiotaping sessions and utilizing a licensed clinician to verify treatment fidelity.

The small sample size of five women per treatment group limits findings as does fact that participants were not matched resulting in an unequal amount of caucasian women in the EMDR group. It was not reported whether the implementing therapist and/or assessor was blind to treatment condition. Dependent variables were limited to verbal instruments and physiological data would have been useful to corroborate self-report findings.
There were a number of studies published that studied the efficacy of EMDR with survivors of trauma in general as opposed to PTSD specifically. This has often included individuals with a PTSD diagnosis, but has focused on populations suffering the effects of trauma without necessarily meeting DSM-IV diagnostic criteria for PTSD (Shapiro, 1989b; Wilson, Silver, Covi, & Foster (1996); Wilson, Becker, & Tinker (1995 and 1997); Edmond, Rubin, & Wambach, (1999); Grainger, Levin, Allen-Byrd, Doctor, & Lee (1997); Marquis, (1991); Puffer, Greenwald, & Elrod (in press); and Scheck, Shaeffer, & Gillette (1998)).

Shapiro’s original study (1989b) will be reviewed due to its importance, however, it is critical to note that unlike previously reviewed studies, the investigation did not feature a PTSD population, but rather, inclusionary criteria consisted of traumatic memory symptomatology. Twenty-two participants suffering from traumatic memories of varying origins volunteered for the treatment after being referred by a variety of mental health clinics. There were 5 males and 17 females ages 11-53 representing varied professions and education levels. Participants were randomly assigned to either treatment or control group.

Treatment consisted of an initial session in which all participants specified their presenting complaints and consented to treatment. The next instructions were to describe a memory from which relief was sought and to isolate a single picture that best represented that memory. Participants then answered which words about themselves or the incident best go with the picture and this typically took the form of statements like “I have no control” (p. 204). Next, participants were told to imagine the scene and the associated belief and to assign an SUD number to them. They were then asked for a preferred feeling and a new statement that would reflect this. Following this replacement
feeling, a VOC score was ascribed to it by the participant and recorded by the therapist. Finally, a brief instruction (p. 204) into the process is given to participants before implementing the two separate treatment conditions of EMD or control.

A single session of EMD treatment of unspecified duration comprised the treatment condition and it is unclear whether this was immediately following the initial session for all participants or occurred on a separate occasion. The procedure began with the patient instructed to visualize the traumatic scene, rehearse the negative statement and follow the investigator’s fingers as a series of 10-20 “voluntary, bilateral, rhythmic saccadic eye movements were elicited by the therapist. The therapist’s finger movements were in either a diagonal or horizontal sweeping motion. In the event of difficulty following the movements, a two handed approach appeared an adequate alternative.

After each set of movements, the participant was told to “Blank it out and take a deep breath.” They were asked to again bring up the picture, focus on anxiety generated, and give an associated SUD level. Related questions were sometimes asked at this point and if new associations were revealed, they were included with the original statement in the next set. After 3-12 sets, when the SUDs reached 0 or 1, the participant’s belief in the validity of cognition was queried using a VOC assessment. When the reported anxiety level was at zero or one, the EMD procedure was terminated. There were several cases where anxiety was sufficiently reduced, but the validity of cognition had not reached peak levels (of 6-7). In these cases, additional sets of eye movements were implemented to raise validity of cognition to the desired level. If a separate memory interfered with the desensitization of the targeted memory, then the additional memory was processed using the same procedure.
The control condition featured a modified flooding procedure that began with a detailed description of the trauma in order to match the groups on exposure to the traumatic memory. During the description of the trauma, the therapist interrupted seven times at approximately 1-1 1/2 minute intervals to ask for a new SUD level, paralleling similar SUD assessments done with the treatment group. Like the EMD group, participants were asked if the picture had changed or others come to mind and then continued with the description of the memory. After the 7th SUD reading, the statement was checked on the VOC scale and this completed the session. This was a delayed-treatment control design, therefore, the EMD procedure was administered to controls following completion of the study.

Follow-up procedures consisted of ½ hour interviews one and three months following the treatment session. This consisted of visualizing the targeted memory and rating anxiety using the SUDs scale. Additionally, the positive belief statement was also assessed with the VOC scale and related symptoms (such as intrusive thoughts and sleep disturbances) were reviewed with the therapist. This last category was not subject to any statistical analyses, however.

Results consisted of assessment in the areas of SUDs levels, VOC levels, pulse rate, and presenting complaints. Analysis of variance showed a highly significant pre-post drop in SUDs level for the treatment group with no pre-post change for the control group. When comparing the delayed treatment condition of the control, again significant drop in SUDs level occurred, suggesting a consistent interaction for the treatment intervention. These effects continued to be significant at three-month follow-up. The second dependent variable was the VOC scale used for the client to assess the extent to which positive beliefs incurred from the process were true. The same pattern emerged
with statistically significant gains in mean VOC scores for the treatment group and no significant difference for the control. Again, the delayed treatment control condition had similar gains following administration of the EMD treatment whereas the placebo condition did not. Effects for both follow-up intervals (one and three months) for both the treatment and delayed treatment conditions showed significant gains maintained.

Pulse rate was also used to monitor treatment effectiveness, however, for unknown reasons this measure was not treated adequately as a dependent variable and therefore did not contribute information towards results. In fact, pulse rate is not mentioned in the measurements portion of the methods section or in the procedural description where implementation of the SUDs and VOC are detailed. Mean pulse rate decreases were reported for both the treatment and delayed treatment control groups between session one and the one and three month follow-ups, however, inexplicably no pulse rate data were obtained for the placebo and delayed treatment conditions of the control group after session one.

Primary presenting problems were treated similarly in that they are reported descriptively without comparison analyses conducted for interpretation. Likewise, there was a section describing patterns that emerged in patient symptomatology as a result of the process as well as patterns regarding participant progressions over the course of treatment. While these last three categories did not act as dependent variables in that comparisons were not conducted, their purpose appears to be in providing clinically meaningful information regarding the processes involved in EMD.

Strengths of the study include these attempts to add clinical understanding to the experience of EMD and to link the effects with participant symptoms related to trauma. Random assignment and follow-up data were also strengths. The study suffered the
common limitation of small sample size and treatment duration. The latter is an interesting point because Shapiro has later criticized other studies for treating multiply traumatized participants in a short number of sessions (see Shapiro 1996 and 1999). The author generalized findings to the extent of "This makes the EMD treatment extremely efficacious as a single-session treatment for multiple rapes, molestations, and similar combat experiences" (p. 219). The sample itself in the study is not comprised of a PTSD population since five of its members were mental health professionals who joined the study to seek relief from the effects of traumatic memories without an established PTSD diagnosis. It is interesting that this original study is cited frequently in the PTSD literature.

The two major limitations of the study involve inadequate dependent variables and the use of a single assessment/treatment therapist who obviously is not blind to condition and originated the treatment model investigated. The potential for demand characteristics or experimenter bias is clear and Shapiro acknowledged this as a legitimate concern (p. 217). Further and related was the use of limited dependent measures which rely solely on client self-report, report which is delivered during the therapeutic encounter with the implementing therapist. No standardized objective measures were employed and it is unclear why the one physiological index partially employed was not utilized comprehensively. This would have been useful information to either corroborate or question the subjective data aggregated. In a review of the EMDR research, Cahill, Carrigan, and Fruch (1999) discuss a potential confound in the relationship between treatment duration and the "explicit contingency between reports of in-session fear reductions and the continuation of treatment" (p. 13). In essence, the criticism is that EMDR is administered until participants reported a decrease in SUDs and
an increase in VOC levels. The control, on the other hand, did not have a contingency between what they reported and when treatment would end. These authors expressed concern that the interruption to elicit SUD ratings occurred in controls every $1 - 1 \frac{1}{2}$ minutes which could result in considerably shorter session duration than the 90 minutes provided to the EMDR group.

Wilson, Becker and Tinker (1995) are frequently cited in the literature for an EMDR efficacy study with 80 traumatized individuals randomly assigned to EMDR or a wait-list control condition. This was the second investigation with adequate enough integrity ratings and large enough treatment effects to merit EMDR as “probably efficacious in the treatment of civilian PTSD” (p. 9) according to the APA Task Force for Psychological Interventions (Chambless, et al., 1998). While the study is methodologically strong and includes comprehensive follow-up data reported in a second study (Wilson, Becker, & Tinker, 1997), the population of interest targets psychologically traumatized individuals and not those diagnosed with PTSD. It is most interesting that considering the sample used, the study made the prestigious Chambless et al., (1998) list specifically for treating “civilian PTSD” (p. 9) when in fact PTSD was not a requisite condition. The authors noted that only 46% of participants met full criteria for PTSD and it is therefore interesting that the study has so often been cited in the PTSD literature. Data from this study was not partitioned into PTSD and non-PTSD groups on crucial dependent variables and therefore is not relevant to the present analysis.

Interestingly, the Wilson, Becker, and Tinker conducted a 1997 study which featured a 15 month follow-up of the same sample in the earlier study is entitled “Fifteen-month follow-up of eye movement desensitization and reprocessing (EMDR) treatment for PTSD and psychological trauma.” Unlike the first study, the follow-up (1997)
partitioned dependent variable data into two groups: Those meeting full PTSD criteria (n=32) and the remaining participants termed partial PTSD (n=34). This allowed for analysis and generalizability to the more specifically defined PTSD population as well as the calculation of treatment effect sizes for this group in the 15-month follow-up. For that reason, the follow-up study is relevant to the present analysis whereas the initial study is not.

In the initial study, Wilson et al. (1995) administered three 90-minute EMDR sessions to 80 participants randomly assigned to treatment or delayed treatment control conditions. Participants were initially pooled from announcements recruiting volunteers, screened through a telephone interview, and ultimately interviewed for inclusion/exclusion criteria. These requirements included specific and comprehensive medical, psychological, and legal exclusion criteria outlined in detail in the study (p. 929). The implementing therapists were all EMDR trained and an independent assessor who was not knowledgeable in EMDR was used to administer assessments. Outcome measures included the SUDs, the Impact of Events Scale (IES), the Spielberger Stait-Trait Anxiety Inventory (STAI) and the Symptom Check List (SCL-90-R). Additionally, PTSD diagnoses were assigned based on the Post-Traumatic Stress Disorder Interview (PTSD-I) and the SUDs and VOC were used as process outcome measures.

Results of the first study were threefold, analyzing treatment effectiveness of the EMDR vs. delayed treatment control; effectiveness for delayed treatment participants; and 90 day follow up data to assess treatment effects over time. Additional analyses were provided to assess related process and outcome variables. Statistical analysis showed decreases in presenting complaints for the EMDR group and increases in positive cognitions. Controls did not achieve change until after treatment with EMDR at which
time similar effects were found. Treatment effects were maintained at 90-day follow-up. Multiple regression analysis showed that variables such as type of trauma PTSD diagnosis, or nature of the memory predicted gain scores (although in the demographic analyses, married people showed higher gains).

The 1997 follow-up study by Wilson et al. assessed two aspects of the treatment: Maintenance of treatment effects and a comparison of long-term gains between those participants diagnosed with PTSD and those not meeting PTSD criteria. Of the original 80 participants, 15-month follow-up data were obtained for 66 of the original sample who either came to the study site for an interview and measures, or responded to the measures by mail. The same dependent measures as used in the 1995 study were used. Additionally, analyses were conducted on demographic variables, trauma characteristics, and the severity of trauma at three month follow-up to see if responders were different from follow-up participant data. There were mixed findings, summarized by “In summary, the non-responders were more likely to have higher scores on some of the outcome measures at the three month follow-up, they were more likely to have engaged in therapy prior to the study, and they earned less income” (p. 1050). The authors cautioned, however, that this be interpreted in light of the very high response rate (83%).

The five outcome measures were analyzed through MANOVA testing using PTSD (vs. partial diagnosis) at intake and measurement interval (of which there were four) as the independent variables. PTSD diagnosed participants scored significantly higher in the symptom direction on dependent variables. A significant effect for measurement time was reported with pretest measures showing a main effect, but no significant differences between the posttest, three-month follow-up, and 15-month follow-up scores on any of the measures. This suggests that the differences gained after
treatment were maintained 15 months later. There was no significant interaction between diagnosis and measurement time, meaning both the PTSD diagnosed participants and those who did not meet PTSD criteria maintained treatment effects equally well.

Group comparisons against a non-disturbed reference group were made in order to assess the clinical significance if treatment effects on dependent variables. For present purposes, however, the data are not as informative as the treatment effect sizes also reported by the authors. At 15 months following treatment the PTSD group estimated average effect size was 1.63 and the non-PTSD group's effect size was 1.54. These are large effect sizes according to the criteria set by Cohen. Further, of the 32 participants who met PTSD diagnostic criteria at pretest, only five continued to meet criteria at the 15 month follow-up. A 68% estimate of reduction in avoidance, intrusion, and hyperarousal symptoms were met by all participants at the 15-month follow-up as assessed by the PTSD-I. Analyses were performed to determine that no treatment differences were obtained as a result of some participants engaging in outside therapy between the three and 15-month assessment intervals.

The Wilson et al. (1997) study was the most comprehensive follow-up investigation of the efficacy of EMDR with PTSD-diagnosed individuals found in the literature. As with other follow-up studies, there is the inherent potential for bias due to the fact that responders may behave in ways different from non-responders with possible implications for the data, however, the authors employed extensive testing to assess potential differences. The partitioning of data into PTSD and non-PTSD diagnosed groups was very useful for both empirical understanding and external validity. Further, the detailed description of dependent variable data and availability of effect sizes facilitated meaningful interpretation of outcome measures. The use of varied measures
was also a strength, although physiological data would have made for even more comprehensive assessment.

Limitations included the fact that the lead author conducted follow-up interviews instead of utilizing a blind assessor and there was some (though a relatively small amount) participant attrition. Additionally, the authors acknowledge not investigating the effects of comorbidity on treatment effects. Lohr et al. (1999) caution against the follow-up data generalizability as there were no comparisons made with a control group.

EMDR Research Reviews

The proliferation of published EMDR studies has resulted in several narrative research reviews and critiques conducted to summarize the findings. Earlier reviews included Acierno, Hersen, Van Hasselt, and Tremont (1994) in which the authors expressed concern about methodological limitations precluding a stance on EMDR’s efficacy. After reviewing and critiquing both controlled group studies and case studies, the authors comment, “To date, therefore, EMDR has not been validated as an effective intervention for PTSD or any other disorder, and additional research employing larger samples and objective measures is definitely warranted” (pgs. 296-297). On the other hand, the authors stated that due to the methodological weaknesses present in the then published studies, “a failure to detect significant change following EMDR does not establish its lack of efficacy” (p. 296). Small sample sizes, treatment integrity, and dependent measures were primary among the methodological concerns.

Lohr, Kleinknecht, Tolin, and Barrett (1995) share similar concerns in their review of nine case studies and five single case designs, as well as eight group design studies published after the Acierno et al. (1994) critique investigating the efficacy of
EMDR. Specifically, the authors state that, “Psychophysiologic measures show little effect of treatment ... Eye movements do not appear to be an essential component of treatment, and there have been no substantial comparisons with other treatments. No studies have adequately controlled for nonspecific (placebo) effects of treatment” (p. 285). The authors noted that while case study information is compelling, there have been equivocal results in terms of treatment effect sizes.

In a more recent review, Lohr, Lilienfeld, Tolin, and Herbert (1999) specifically critiqued EMDR research regarding findings on specific versus nonspecific treatment factors. Essentially, the authors believe that the specific components of EMDR, particularly the bilateral stimulation and “modifications of attributions regarding the significance of the affective image” (p. 198) have not been shown with acceptable scientific rigor to be viable mechanisms of change. The authors stated that EMDR has not passed the strong tests of incremental validity necessary to regard it as a unique modality in its own right as opposed to another variant of extant treatments. Five experimental procedures and design elements are suggested in order to better assess the specific treatment components of EMDR such that it can be identified as something different that other exposure techniques.

In an evaluation of controlled PTSD research on EMDR, Shapiro (1996) reviewed a total of 12 EMDR Controlled Studies as well as over-viewing several related studies. It must be noted, however, that at least three of these studies (Shapiro, 1989b; Wilson, et al., 1996; and Wilson et al., 1995) did not study PTSD-diagnosed individuals specifically and therefore should not have been included in an article entitled “EMDR: Evaluation of controlled PTSD research.” Shapiro (1996) described and rated each study in terms of fidelity checks, number of sessions, dependent variables, participants (single v. multiple
trauma and whether or not participants were compensated for their disability), and type of research design. The author further compared some of the EMDR research with that of other treatment modalities such as stress inoculation training and other exposure techniques. It was concluded that "Within the context of extant PTSD research, the converging evidence provided so far would appear to justify recognizing EMDR as an empirically validated treatment for PTSD" (p. 214). The author further stated that more recent and methodologically sound studies have found large effects and that these studies have used similar methodology for PTSD studies utilizing other popular treatment modalities.

Several important points are brought up in the Shapiro (1996) review. Attrition rates appear to be lower for EMDR-treated participants, though this has yet to be systematically studied. This is a particularly important point in light of the stresses exposure and other types of trauma therapy imposes on its clients and warrants further study as a critical variable in its own right. Also, treatment efficacy research is more valuable when measures of clinical significance, in addition to statistical significance, are included in dependent variables. Further, size of treatment effects and number of sessions required to obtain treatment goals are important factors when evaluating the efficacy of any treatment modality.

A more recent review by Shapiro (1999b) focuses specifically on four investigations described as "independent, rigorously controlled studies" (p. 35) which feature "single trauma PTSD" (p. 36). The distinction between single episode trauma and multiply traumatized individuals such as is common of war veterans has become more pronounced in Shapiro's writings over the years. Shapiro (1999b) stated that these studies all used "acceptable standard measures and independent assessors demonstrated
that after the equivalent of three 90-minute sessions 84 to 100% of the single-trauma
subjects were no longer diagnosed with PTSD at posttest” (p. 37). Shapiro includes in
this review the studies of Marcus et al. (1997) and Rothbaum (1997) which were
reviewed previously in the present paper. However, also included are studies by Scheck
et al. (1998) and Wilson et al. (1995) which do not specifically feature a PTSD-diagnosed
sample thereby causing pause to the previous assertion to the focus asserted in the title.

In this review, Shapiro (1999b) critiqued previous reviews of the EMDR research
primarily for strict adherence to the earliest EMDR protocols that have been updated as
early as eight years ago and not commensurately updated in independent studies.
Examples include the consecutive use of SUDs reports by a client until no disturbance is
reported and attribution of treatment effects solely to the eye movements. The article is
more a useful statement for understanding the current state of EMDR than a research
critique as most of it is devoted to describing and updating Shapiro’s stance on EMDR.
EMDR is defended as “an integrated form of therapy incorporating aspects of many
traditional psychological orientations and one that makes use of a variety of bilateral
stimuli besides eye movements” (p. 37). Specifically addressed are the component
analysis studies which attempted to assess the role of bilateral stimulation in EMDR.
Shapiro (1999b) stated that most such studies are plagued by methodological flaws, and
“Thus the inescapable conclusion from an examination of all the EMDR components
analyses implemented so far is that no unambiguous determination of the weighting of its
individual components is yet possible” (p. 60).

DeBell and Jones (1997) reviewed seven experimental studies of EMDR, four of
which sampled a PTSD population. In addition to the issues of internal validity
highlighted elsewhere, the authors summarize important controversies surrounding
EMDR. This includes inconsistent information in the EMDR literature regarding the potential for harm with EMDR, training controversies, and the number of sessions required to facilitate change in clients. The authors concluded that “EMDR holds some promise for the treatment of PTSD, phobias, and perhaps some other anxiety-related disorders” and that the majority of studies supported some aspect of EMDR.

Recommendations for future EMDR research included standardizing pre and post-treatment assessment; more careful statistical analyses; increased use of objective measures; increased use of no treatment or delayed treatment control groups; confirmation of participant diagnoses through independent assessors; and more extensive research examining why EMDR procedures might work (p. 159).

Foa and Meadows (1997) reviewed several treatment modalities for PTSD including EMDR. Using what the authors termed “a gold standard criteria,” several frequently cited studies were reviewed and ranked in terms of how the gold standard criteria were met or not met. All of the EMDR studies treating PTSD were reviewed in the present paper. The authors concluded that the evidence on EMDR with PTSD is mixed and that “Methodological flaws rendered most though not all of these findings un-interpretable. The test of the efficacy of this much-discussed treatment awaits adequately controlled studies” (p.476).

Cahill, Carrigan, and Frueh (1999) wrote a recent critical review of the controlled outcome and dismantling research on EMDR which included treatment of PTSD and other disorders. Although the article narratively reviews EMDR research in general, sections of the critique are specific to EMDR with PTSD and trauma. The authors drew several conclusions including the following: EMDR appears to be effective in reducing at least some PTSD symptomatology; EMDR is at least as effective or more effective than
several non-validated treatments (such as active listening or relaxation training); and the dismantling study evidence provides very little evidence for a unique contribution of the eye movements or other bilateral stimulation component of EMDR, particularly for long-term effects. The authors pointed out that a major limitation previously ignored is the fact that until 1999 (see Devilly & Spence, 1999) EMDR had never been compared with an independently validated treatment for PTSD (p. 6).

The Cahill et al. (1999) review is useful in its organization as it looks at studies in terms of two questions: Does EMDR work? and How does EMDR work? These questions are addressed by categorizing published studies by comparisons with no treatment, validated treatments, or non-validated treatments. The available dismantling studies are used to address the latter question and this is further subdivided into population (i.e. dismantling studies of EMDR with PTSD studies). Suggestions to improve future research included more uniform treatment fidelity checks by outside raters with a quantitative analysis to assess the impact of treatment fidelity on outcomes. Finally, the authors concluded that the research on the eye movement component of EMDR consistently shows that bilateral stimulation does not account for effect sizes and therefore further research into the process and agents of EMDR is warranted.

Interestingly, there was only one meta-analytic research review that investigated EMDR as a treatment group in studying the efficacy of various treatment modalities for PTSD. Van Etten and Taylor (1998) examined 61 treatment outcome trials studying a variety of pharmacological and psychological therapies that were compared with controls. Controls were loosely defined and this meant that comparisons were made between both no treatment controls and alternative treatment controls. This is not an optimal structure for comparison as a treatment vs. no treatment can logically be expected
to have greater effect size than the same treatment vs. some other form of therapy (see Cooper, 1998 for a more detailed account of this issue).

EMDR was one of five psychological therapy conditions (others included behavior therapy, relaxation training, hypnotherapy, and dynamic therapy). There were a total of 11 EMDR studies analyzed using a confidence interval method of data analysis at the 90\% level. A number of comparisons were made including attrition rates, however, the research was primarily measured through symptom reduction variables. Dependent measures were organized (i.e. effect sizes generated) for the PTSD symptoms of intrusions, avoidance, and total PTSD symptoms. Anxiety and depression, two related symptoms, were also analyzed. Each of the five symptom categories were clustered as either self-report or observer-rated dependent measures and effect sizes were generated for aggregation and comparison.

Results varied with different pharmacological interventions rated superior to psychological therapies on some clusters and vice versa on others. There was inconsistency between psychological therapies, however, EMDR and behavior therapy generally showed larger effects than the other psychological therapies compared in the study. On self-report intrusion measures, drug and psychological therapies were both found in general to be equally effective in symptom reduction. EMDR was found to be superior to relaxation training and dynamic therapy, however, whereas behavior therapy had larger effects than one of the drug classes, EMDR did not. When compared with a different drug intervention, however, EMDR and behavior therapy were the only two psychological interventions found to be superior. On observer-rated measures of intrusions, EMDR fared better than behavior therapy. Only EMDR was found to be more effective than the most effective drug therapy as well as superior to controls.
On self-report measures of avoidance, Van Etten and Taylor (1998) reported that, “EMDR and behavior therapy were equally effective, but only EMDR was superior to relaxation and dynamic therapy” (p. 136). On the observer-rated measures of avoidance, these two conditions were also found to be equally effective, however, only EMDR was more effective than control conditions.

A third cluster was total PTSD symptoms and again EMDR and behavior therapy were found to be the most effective of the psychological therapies and about equally effective on self-report measures. EMDR was the only therapy, however, found to be significantly superior to all other psychological therapies. Conversely, on the observer-rated measures only behavior therapy was more effective than controls and was also found to be more effective than EMDR.

Self-report measures of anxiety, showed the most effective drug therapy (SSRIs) was “comparable” (p. 136) to behavior therapy and superior to EMDR. Again, these two psychological therapies had larger effects than the other three. Analysis on observer-rated measures of anxiety was not possible due to the small number of measures employed by primary studies. On depression, only self-report measures were analyzable for all conditions. Psychological and pharmacological interventions were generally comparable, and again the behavior and EMDR therapy conditions were equally effective. More effective than both of these conditions, however, was the SSRI drug intervention.

Follow-up data were provided by some, but not all of the primary studies. In terms of psychological interventions, the authors found behavior therapy and EMDR to maintain effects at follow-up points and generally equally so. Interestingly, effect size
increased in the EMDR condition from posttest to follow-up such that the effect size was equal to behavior therapy (previously the larger ES) at follow-up.

Van Etten and Taylor (1998) summarized that the psychological therapies were generally more effective than pharmacological interventions, and that SSRI and carbamazepine were the most effective of the drug therapies. Among psychological therapies, the authors stated, "behavior therapy and EMDR were the most effective with the two being generally equally efficacious, although behavior therapy was significantly more effective than all treatments on observer-rated total PTSD symptoms. No differences in comparative treatment efficacy were discernible between behavior therapy and EMDR across the specific symptom domains of PTSD" (p. 138). The authors further cited support for both EMDR and behavior therapies by noting their relatively lower attrition rates among studies (p. 139). Unfortunately, the study was not able to analyze variables specific to EMDR as it was a general comparison of the efficacy of therapies used to treat PTSD. Further, analysis of the 11 EMDR studies utilized by Van Etten and Taylor (1998) shows that at least one of these studies did not require a PTSD diagnosis among the sample.

Perhaps the opposite of meta-analysis is the case study. While case studies are criticized because of the obvious limitations in external validity and internal methodology, they provide valuable qualitative and clinical information. Numerous case studies have found positive effects for EMDR including McCann (1992); Blore, (1997); DeJongh and Broeke, (1998); Goldstein and Feske, 1994; and Greenwald, (1994) among others. A limitation of the majority of published cases of EMDR (as well as other treatment modalities) is the fact that single case designs are not employed. EMDR with its availability of baseline measurements and ongoing uniform assessment avails itself
easily to single case design research and this would provide methodologically sound data to complement the qualitative findings brought out by case studies. Several single case design studies were found in the literature, including Levin, Lazgrove, and van der Kolk (1999) and Kleinknecht and Morgan (1992). Each of these studies reported significant differences from pre to posttreatment in PTSD symptomatology following EMDR therapy.

Despite the controversy over the rapidity, theoretical underpinnings, and lack of neuroscientific explanation for EMDR, it continues to gain popularity with mental health practitioners and the public at large. Inclusion on the APA Task Force on Validation and Dissemination of Psychological Procedures list of empirically validated therapies as probably efficacious for treatment of civilian PTSD (Chambless et al., 1998) secures EMDR as a viable, sound therapeutic treatment modality for treating civilian type PTSD. The APA’s list of empirically validated treatments is based solely and discriminantly on a review of the research and is considered a leading authority for translating the plethora of studies, scientific criticisms, and research opinions into meaningful information for scholar-practitioners. An important research strength of EMDR highlighted by the task force’s 1995 report and recommendations for psychological procedures is its reliance on standardized treatment. Few psychotherapies, and particularly not the more psychodynamically oriented modalities, operate with the specific, protocol based process as does EMDR. Chambless et al. (1998) stress that “Standardization and precise definition of treatment through treatment manuals reduce methodological problems and lead to more specific clinical recommendations” (p. 4). It can be concluded with some confidence, therefore, that sound EMDR studies accurately measure the efficacy of EMDR rather than therapists’ variables.
CHAPTER III

METHODOLOGY

Introduction

The purpose of this study was to explore the accumulated evidence of the efficacy of EMDR in treating Posttraumatic Stress Disorder and to investigate the said variables associated with potential treatment effects. EMDR was operationally defined as the manualized treatment of Eye Movement Desensitization and Reprocessing as originated by Shapiro (1989b) and implemented by an EMDR trained therapist with a minimal of Level I training. PTSD is operationally defined as meeting requirements for diagnosis as stipulated by the DSM-IV (APA, 1994, pgs. 427-429). Essentially, this was a study of the studies which attempted to summarize and make sense out of the plethora of isolated primary studies. Instead of attempting this through the still common method of anecdotal narrative summary, the method used was meta-analysis.

While there were over a dozen narrative research summaries located in the literature, there were no meta-analyses located that featured EMDR as the primary independent variable of interest. Meta-analysis is a method of quantitatively aggregating the statistical findings of individual studies in order to estimate an overall effect size for a treatment. There are a variety of meta-analytic procedures that have been developed toward this end, each with its own advantages and disadvantages. The method selected
for a particular study is dependent on the assumptions of the data and the criteria employed by the examiner.

Although meta-analysis has become increasingly popular in the past two decades, its origins begin in 1904 when British mathematician Karl Pearson sought to make sense of the multiple findings testing the effectiveness of inoculation for typhoid fever (Hunt, 1997). A simple concept, Pearson calculated correlations for each study's findings and then averaged the correlations with rudimentary weighting provisions for each study. Decades later, statistician Leonard Tippett improved upon quantitative aggregation of individual studies by developing a method to combine the probability values of multiple studies. This made it possible to hypothesize how likely it was to determine if the results of the studies were due to chance as opposed to treatment effects.

Few scientists pursued meta-analysis with vigor until the latter part of the century when the social sciences began to publish studies at record numbers, thereby increasing the need to make sense of mass findings. In the early 1970's and independent of each other, Robert Rosenthal and Gene Glass began taking a new look at how to statistically summarize data. Glass presented his technique to the American Educational Research Association, spawning the modern age of meta-analysis. By 1994 nearly 400 meta-analyses were published in the scientific literature (Hunt, 1997) and to date over 3,000 meta-analyses can be found in published research.

Hunter and Schmidt (1990) highlight that meta-analysis was borne out of the need to make sense out of the vast empirical data of recent decades. For this reason, meta-analytic procedures are ideally suited for studying the efficacy of EMDR therapy for trauma survivors. Few treatment modalities have precipitated as many studies as EMDR yet presently there is no meta-analysis published to make sense out of the accumulated
individual studies. In this chapter, the definition of meta-analysis and its advantages and limitations will first be presented to explain the rationale for employing this technique in the present study. This will be followed by the proposed meta-analytic process for the present investigation with a description of each procedural stage.

Meta-Analysis

While statistical methods of aggregating findings of individual studies have existed for the past century, the term meta-analysis was introduced by Glass (1976). Glass, McGaw, and Smith (1981) describe meta-analysis “not as a technique, but as a perspective that uses many techniques of measurement and statistical analysis” (p. 21). Its essential character is that it is a statistical analysis of the summary findings of many empirical studies; its purpose to draw general conclusions. For the present study, meta-analysis was operationally defined as a statistical estimate of effect size calculated as an average from the effect sizes of primary studies. Primary studies were operationally defined as previously completed published or non-published empirical investigations that tested for treatment differences through formal research design.

A simple approach to data aggregation is to test the statistical significance of the combined results across primary studies. Basically, and in the same vane as individual studies, this method analysis the results to answer the question: Do the combined results reach statistical significance? This approach has become known as the combined-tests method and is quite limited in the richness of conclusion it is capable of drawing. Combined methods include the techniques of vote-counting, and summation procedures like Stouffer’s (reported in Cooper, 1998) Adding Z’s method.
Glass's meta-analytic method works toward a different goal which provides more useful and robust types of information. This technique estimates the effect sizes of individual studies in order to quantify treatment differences in relation to the independent variable. Using effect sizes derived from individual studies, an overall estimate of effect is calculated to illuminate the magnitude of an independent variable's effect across studies. There are several methods for estimating effect sizes, each a variation on the difference between means (i.e. treatment vs. control) divided by some form of standard deviation. The particular method depends on the purpose of the investigation, the nature of the construct under study, and the data reported by individual studies. Like the combined methods, this meta-analytic technique allows for aggregated significance testing in order to answer the question of EMDR's efficacy. Additionally, however, it has the advantage of estimating (and standardizing) the size of potential treatment effects. Through the effect size model, the present study allowed for an estimate of treatment effects, distribution of treatment effects, and examination of sub-variables (called moderator variables) found among the collection of primary studies employed.

The present study used the fixed effects model which Hedges and Olkin (1985) suggest when the effect sizes coming from the separate studies estimate a fixed population value rather than a random one. There is no universal criteria for using a fixed vs. random effects model, however, the general guideline is that an effect size is said to be fixed when there are no expected random variables that are likely to occur outside of things like sampling error. The fact that EMDR is a manualized treatment weighed heavily in this decision. Luborsky and DeRubeis (1984) note that in clinical efficacy studies, use of manualized treatments result in cleaner, more controlled studies by removing much of the variance associated with treatment implementation and
methodological interpretation. A further consideration promulgating use of the fixed effects model was that EMDR studies have been conducted within the past decade and as such have the advantage of the most modern scientific standards, thereby minimizing random effects. Further, since threats to validity will act as a coded variable, it was possible to estimate the role of error in overall findings.

Advantages of Using Meta-Analysis

The purpose of meta-analysis is to review the research literature and a key advantage of this method of review is that it does so empirically. Durlak and Lipsey (1991) summarize, "The distinctive features of meta-analysis compared to other review strategies are the quantitative representation of key research findings in the studies reviewed and the statistical analysis of the distribution of findings across studies and the relationship of study features to those findings" (p. 293). Whereas traditional literature/research reviews narratively, and consequently subjectively, represent the data, meta-analysis is able to mathematically summarize the findings of different studies on the same variable. Moreover, the direction, distribution, and the magnitude of effects can be examined empirically.

In addition to analysis of statistical significance for main effects, meta-analysis is capable of an even richer analysis; that is investigation of moderator variables. Beyond estimating overall effects of a treatment, meta-analysis has the ability to statistically analyze a subgroup of variables within an independent variable. Therefore, multiple independent variables/hypotheses can be tested from the same data, given the data is robust enough. This is very important to treatment efficacy studies because the investigator can go beyond the broad scope of a particular modality and determine what
specific factors contribute significantly to treatment effects. For example, in a meta-
analysis examining the relationship between class size and attitudes and instruction in the classroom, Smith and Glass (1981) were able to demonstrate effects on variables such as student self-concept, participation, and teacher morale in addition to the larger hypothesis regarding student achievement as a function of class size. This has important implications for understanding possible explanations for outcomes in treatment efficacy research.

A third advantage of meta-analysis is its ability to manage very large amounts of data. This helps the researcher to use multiple existing data to examine multiple independent and moderator variables as well as the interactions between them. Meta-
analysis has the capability of using multivariate techniques to analyze a sheer volume of data that qualitative methods could not manage nor is practical for individual studies. An example of the large scope of meta-analysis is illustrated by Lipsey (1993) who analyzed a total of 434 studies in a meta-analysis assessing the efficacy of psychological, educational, and behavioral treatment. Allowing for numerous individual studies has the advantages of reducing bias in the selection of studies to be summarized, enabling a large sample size, and increasing statistical power. Of course, meta-analytic techniques require selection decisions and judgments similar to primary studies and the researcher must employ fidelity safeguards throughout the process. However, careful documentation of selection criteria, coding techniques, and statistical procedures provides for replication and scrutiny of methods. The present study provided information in all of these areas for the purposes of treatment fidelity and replication.
Criticisms and Limitations of Meta-Analysis

Meta-analytic research currently appears to be the best method for synthesizing research. However, this method is not immune from criticism and there are legitimate grounds for skepticism on several fronts. The first and most enduring criticism is what has been referred to as the garbage in, garbage out problem. Essentially this refers to the fact that a meta-analysis is only as strong as the studies it is comprised of. Clearly, when a study has low levels of internal and external validity, its inclusion in another study will also suffer from these inadequacies. Eysenck (1978), a staunch critic of the technique, summarized the problem, “A mass of reports, good, bad, and indifferent, are fed into the computer in the hope that people will cease caring about the quality of the material on which the conclusions are based. ‘Garbage in-garbage out’ is a well-known axiom of computer specialists; it applies here with equal force” (p. 517). There are methods to control for the garbage in, garbage out problem, most often through rigorous inclusion criteria or differential weighting of individual studies within the analysis. Of course, limiting inclusion criteria presents a different problem, which is exclusion of potentially relevant studies that could contribute to overall findings.

Cooper (1998) described three approaches to addressing weaknesses inherent in the primary studies employed in the meta-analysis and the method to be employed in the present study is the mixed-criteria approach (p. 88). This approach combines two techniques, the first being the threats-to-validity approach which is based on the threats to internal and external validity first outlined by Campbell and Stanley in 1963 and developed further by others (i.e. Bracht & Glass, 1968 and Campbell & Cook, 1979). Basically, the researcher creates a list of potential threats to validity and assigns point values commensurate with the level of threat each study presents for each type of threat.
For example, in the Vaughan et al. (1994a) study, there were no treatment fidelity checks reported and this is a threat to internal validity. This would result in a low number of points for this particular validity issue and would adversely contribute to the study's overall points for internal validity. This method is advantageous in that it allows for the inclusion of many studies while concurrently addressing the limitations of each. The threats to validity criteria used in the present study will be based on those identified by Campbell and Stanley (1963) and listed in a reference table in the final chapter.

The second technique to address the garbage in garbage out problem in Cooper's (1998) description of the mixed-criteria approach is called methods-description method (p. 86). Essentially, this entails providing descriptions of each study's experimental design characteristics and dependent variable information. This information is coded so that analyses can be made for particular comparisons, such as between studies with a large sample size or comparing studies with comparative treatment groups with within-subjects designs. The present study coded for descriptive variables about each study’s research design as well as information related to the five moderator variables. Additionally, attrition rates for EMDR and comparative conditions were attempted to be coded to see if there was a difference favoring EMDR as has often been implied in the literature (see Devilly & Spence, 1999, p. 154). However, due to inadequate reporting among primary studies, analysis of this variable was not possible. Other variables nit amenable to analysis or conceived of as a result of the analysis results are addressed in chapter 4. Coding criteria are presented in Appendix C in chapter 5.

A second and related controversy in meta-analysis is what has come to be known as the file drawer problem, a term originated by Rosenthal (1979). This refers to publication bias for significant results. A problem faced by any meta-analysis begins
before the study is even started because instead of relying on controllable participants and research environments, the researcher relies on completed pieces of data. Most of this comes in the form of published data and inevitably most published data report significant findings, a research dilemma known as publication bias. To remedy this potential confound, meta-analytic researchers must often go to great lengths to find a more representative sample of studies. Typical methods include locating unpublished studies, dissertation investigations, or contacting researchers prominent in the relevant field in order to locate studies presented at case conferences and the like.

The present study reduced publication bias by conducting a thorough literature search that included methods to specifically seek unpublished studies. Primarily, this included contacts made to relevant researchers, professional associations, and computer database searches. A more detailed description of these techniques is presented in the Literature Search section.

The "apples and oranges" problem refers to the practice of aggregating studies with different or variant independent variables. This metaphor essentially argues that the diversity between studies makes comparison inappropriate. This dilemma varies with the construct under study, however its threat is to ignore potentially meaningful differences between, for example, treatment interventions. For instance, in the Smith and Glass (1981) class size study, the independent variable of size could reasonably be partitioned into five student intervals thereby allowing for comparisons between classes of 15, 20, 25, etc. However, the study of psychology is wrought with unclear lines between constructs including treatment and such partitioning is not always possible. This presents the statistical problem of having opposite direction results cancel each other out with potential for a zero treatment effect when in fact large but diverse effects may exist. This
has been a criticism of another and more famous Smith and Glass (1977) study called “Meta-analysis of psychotherapy outcomes.” Criticism centered around the aggregation of multiple methods of psychotherapy, that is multiple independent variables in their own right, into one construct. With such a broad based hypothesis there is the statistical risk of one psychotherapy modality having large effects and another having similar effects in the reverse direction with the result of an overall lack of effect for the construct of psychotherapy.

There are methods to circumvent the apples and oranges problem. Smith and Glass (1981) emphasize the importance of the research question itself and point to the utility of aggregating studies for a general question such as the title of the now infamous study. This holds true for the present investigation as an overall estimate of effect for EMDR in treating trauma is meaningful in its own right. A less common method is conducting independent analyses for variables prior to inclusion in the study, or more typically testing for the presence of modifier variables within the meta-analysis. The present study addressed a specific, manualized treatment for a specific diagnostic disorder and as such was not vulnerable to the problem of measuring different constructs. In order to further ensure the homogeneity of the moderating independent variables (that is, check for the problem of intercorrelation as a confound), certain statistical procedures known as homogeneity tests were employed and these methods are described further in the statistical procedures section.

A final major controversy surrounds the statistical soundness of meta-analysis regarding multivariate effect size data. By pooling all estimated effect sizes into a single analysis, the statistical assumption of independence is violated and results may be biased due to non-independence. Wolf (1986) summarized, “Multiple results from the same
study are often used which may bias or invalidate the meta-analysis and make the results appear more reliable than they really are because these results are not independent” (p. 14). There are methods for addressing the problem of multiple, non-independent results from the same study including performing separate analyses for different outcomes, averaging the multiple results from individual studies, or utilizing apriori limitations minimizing the number of effect sizes derived from a single study (such as including only two effect sizes from a given study).

Another way to address the dilemma of non-independence of multiple result data is to weight each effect size. There are various procedures for weighting and the present study will employ the method advocated by Cooper (1998). This involves weighting each effect size inversely proportional to the number of effects from that study. This allows for each study to contribute an equal unit of information and also provides for more measures than simply averaging effects. Thus, a more detailed analysis of more potential independent variables and moderators on different dependent variables is possible. The present investigation utilized weighting procedures to help control for the problem of data non-independence.

Steps in Meta-analysis

Like its primary study predecessors, a meta-analysis involves a series of well thought-out research steps each based on sound research design decisions. This means that like a primary study, the statistical analysis is an integral component of the investigation but by no means the sole sustenance of the study’s integrity. There are six major steps in a meta-analysis and in each one consideration of multiple issues is required
in order for logical decisions to be made. These steps will be described in the order in which they occur.

**Problem Formulation**

The first step in meta-analysis is formulation of the research question(s). Glass and Smith (1981) conclude that the primary purpose of meta-analysis is to summarize the girth of information available on a given topic. Therefore, the general purpose of meta-analysis is descriptive in intent. However, a purely descriptive meta-analysis has limited explanatory value. Durlak and Lipsey (1991) emphasize that a more investigative, issue-driven approach involves formulating a priori hypotheses to address the relationships between critical research variables. This allows for inferences regarding causal relationships (provided the use of experimental designs of the primary studies used investigated these relationships).

The problem in the present study was to empirically answer the controversy surrounding the efficacy of EMDR treatment in ameliorating the psychological effects resulting from traumatic experience. To date, this has been addressed in the previously described Van Etten and Taylor (1998) meta-analysis that compared EMDR with a total of nine other treatment conditions plus controls. The present study contributed additional and more current information regarding the efficacy question, however, it empirically tested independent variables specific to EMDR which the Van Etten and Taylor (1998) study did not address. Further, it did not collapse all studies as the Van Etten and Taylor (1998) investigation did, but analyzed EMDR against no treatment controls and alternative treatments separately for a more accurate estimate of effect. Specifically, the following research questions were tested: (a) Is EMDR efficacious in reducing the presenting psychological symptoms of trauma survivors and if so, what is the magnitude
of effect? (b) Are there differences in the magnitude of effect between combat-related trauma and survivors of non-combat-related traumas? (c) Are there differences in effect sizes between verbal report measures and objective, physiological indices? (d) Are there differences in the magnitude of effect between earlier and more recent studies? (e) Do the eye movements themselves play an integral role in contributing toward treatment effects? (f) Are there differences in effect sizes based on the number of EMDR sessions administered?

In formulating the research questions, all available experimental studies were examined to identify important questions, particularly those related to the controversies existing within the literature. All studies employing EMDR in the treatment of trauma were collected and after consideration of external validity issues, this population was further specified to include only studies whose participants were diagnosed with PTSD.

Literature Search

Data collection is an integral component of any study and this step is of paramount importance to a comprehensive meta-analysis. Durlak (in Grimm & Yarnold, 1997) describes, “The ultimate goal of a literature search is to obtain a representative and non-biased sample of relevant investigations” (p. 333). A broad and inclusive data collection increases the probability of uncovering unexpected and comprehensive findings. Cooper (1998) identified 15 different techniques used by research reviewers for retrieving research literature. Those relevant to treatment efficacy research will be used in the present study. The period of 1989, the year EMDR was originated, through the end of 1999 was covered in the search.
The first step in the literature search process included database computer searches using Proquest, Psychlit, Psychinfo, Medline, and OVID. Additionally, the world wide web outside of research databases was utilized. The key words used for the search included Eye Movement Desensitization and Reprocessing, EMDR, trauma, trauma treatment, trauma intervention, psychological trauma, Posttraumatic Stress Disorder, PTSD, and meta-analyses.

The second step in the literature search was a manual search and included the most recent Psychological Abstracts (unavailable on CD-Rom), and contacting relevant organizations such as the EMDR Institute and the EMDR International Association (EMDRIA) for additional materials including non-published studies. Finally, bibliographies from obtained studies were used to locate any remaining studies.

In order to further locate unpublished studies, with the purpose of reducing the file drawer problem, the following methods were employed: Utilizing the UMI Dissertation Abstracts; contacting EMDR researchers through the EMDRIA Listserv (LISTERV@MAELSTROM.STJOHNS.EDU); contacting EMDR researchers listed in the EMDR Level I Training Manual; and seeking unpublished papers presented at professional conferences by contacting professional associations related to the treatment of trauma (such as the International Society for Traumatic Stress, World Congress of Behavioral and Cognitive Therapies, and the American Psychological and Psychiatric Associations). This literature search was very successful as it resulted in personal communications with over a dozen prominent researchers in the areas of EMDR and PTSD and resulted in obtaining three unpublished primary studies.
Inclusion/Exclusion Criteria

There are advantages and disadvantages to both broad criteria for including the maximum number of studies feasible or a limited set of highly select studies. This is akin to the Type I vs. Type II tango faced in any research design for what is a gain from one perspective necessitates loss from a different standpoint. The present study attempted to employ high quality studies with representation from a variety of sources while coding for design type and integrity. Based on the recommendations of Cooper and Hedges, 1994), the following criteria was used before including a study into the data base: (1) studies must investigate the effect of EMDR on participants diagnosed with PTSD based on criteria in the DSM-III, DSM-III-R, or DSM-IV. The investigation did not include those whose presenting condition or complaint was surviving a trauma, traumatic memories, or any other trauma-related condition unless the diagnosis of PTSD was specified in the study. This criteria was aimed at enhancing clarity in the generalizability of findings. Unfortunately, this excluded a large number of studies while maintaining external validity.

EMDR therapy is a specified, manualized treatment therefore its definition is straightforward. Those studies using the EMDR treatment protocol as an independent variable were included. Criteria further included as little as a single session of EMDR and only studies employing a minimum of Level I trained EMDR therapists were used. While completion of this training does not guarantee treatment fidelity, it does fulfill the minimal requirements set by Shapiro (1995) and means that the implementing therapist has both knowledge and possession of the manual for treatment. Treatment fidelity was further addressed in the coding variables section as each study’s level of fidelity checks was rated and coded for analysis.
(2) Each study must have reported a comparison between at least two conditions. This may include two different groups or a single group with a pre and post-test design. Single case designs were not used in the present study.

(3) Each study must have reported the minimal quantitative data necessary for statistical analysis. This could be in the form of ready-to-use statistics or those amenable to transformation such as correlational data. Effect sizes can be calculated from t, F, or chi square tests or from raw data including means standard deviations, and the n of group sizes. When means or standard deviations of separate groups were not reported by primary studies, the following formula for the d index developed by Rosenthal (1994) will be used: 
\[ d = \frac{t}{\sqrt{df \text{ error}}} \] (Cooper, 1998, p. 129). Studies reporting less information than required by any of these means will necessarily be excluded.

(4) Dependent measures must have fallen into the broad category of trauma symptomatology. This included client self-report data, standardized measures, or physiological measures. In accordance with the mixed-criteria approach to coding employed in the present design, information on dependent measure type and integrity was coded to allow for analysis using type of measure as an independent variable.

(5) Each study must have met criteria for methodological soundness in terms of validity and reliability. As discussed previously, a good meta-analysis must balance the benefits of including methodologically superior studies with weaker, but potentially meaningful other studies. There are no absolute standards for this, therefore the present study utilized a general criteria for acceptance but also use methodological integrity as a coded variable. This strategy builds into the meta-analysis the ability to assess whether the degree of experimental rigor has any effect on outcomes. Ruling out methodological factors as an explanation for treatment effects was made possible by coding for
experimental quality (see Appendix D for rating form). Since the acceptance of EMDR as a legitimate therapy has been clouded by criticism of its research, this was particularly useful in settling the controversy.

Defining and Coding Variables

The third step in meta-analysis involves developing and executing coding procedures for translating the usable features of each study into quantitative data. Coding was operationally defined as categorizing each primary study according to a continuum of descriptive and hypothesis-driven independent variables. It is coding that enables the investigator to cull from a study the relevant information related to the hypotheses under investigation and put it into quantitative terms. There are many ways to code a study and the goal is to develop procedures that effectively reflect the essential features of the original study while enabling analysis of the hypotheses under investigation. Durlak (in Grimm & Yarnell, 1997) stated, “The aim is to develop a coding system that is general enough for all the studies being reviewed, but specific enough to capture unique study features” (p. 324). The result of coding is that separate analyses for the different variables coded allows one to assess the differences in the magnitude of effect for different independent variables; that is, the moderator variables under investigation.

The present meta-analysis employed the suggestions of Stock (in Cooper & Hedges, 1994) to code within the following six broad categories: report identification; setting; participants; methodology; treatment; and effect size (p. 127). A seventh suggested variable, process, was not used as it is not particularly relevant to manualized treatment. These will be briefly reviewed here; however, a comprehensive coding sheet used for all primary studies is included in Appendix 2.
Report identification included an assigned identification number for each study, author, year, geographic region, setting, and type of trauma of each study. Both the exact year and the two general categories of older and more recent studies were coded. The latter used the arbitrary cutoff of 1989-1994 as older studies and 1995-1999 as more recent studies. The idea is that the older category includes the first five years of EMDR research after its inception and the more recent category includes the most recent five years studied. Geographic region included either country of study, or if in the USA, the western, mid, and eastern USA. Setting was coded into the categories of inpatient, outpatient, mixed, or unknown. Participants were coded along several different categories including combat-related trauma vs. non-combat related trauma. Coding of this nature allowed for study of hypothesized moderator variables as well as other potential categories to serve as independent variables.

Methodology was coded by research design, sampling characteristics, and rating of methodological integrity. Methodological integrity featured the number of threats to internal validity within each study based on Campbell and Stanley's (1963) work and discussed earlier in this section. Treatment was coded in several ways, including the number of sessions and type of treatment employed (i.e. for the EMDR vs. similar treatment without eye movements variable). Effect size coding included information on dependent variables and the statistical information from each study. Dependent measures were coded into the categories of self-report vs. physiological measures. Statistical information included methods of analysis, scores used in analysis, type of test, and number of comparisons used in each study. Most studies featured more than a single comparison and thus each comparison was coded separately. Finally, a miscellaneous
column was available for coding of any potentially meaningful variables that were not conceived a priori. The coding form is presented in Appendix C.

The Index of Effect Sizes

The fourth step in meta-analysis is determining effect sizes. Effect size estimates provide meaningful information about both the direction of results and the magnitude of treatment differences. This is a major step in the meta-analysis as it mathematically quantifies dependent variable results into meaningful data points to be used in the aggregate analysis. There are several methods for deriving effect sizes but essentially they are all comprised of the difference between treatment group means divided by some form of the pooled standard deviation of groups. This allows for standardization across studies and enables the reviewer an index to compare results. Effect size in the present investigation were operationally defined as the difference score between two treatment conditions utilizing the equation of difference score divided by the pooled standard deviation of the two scores. A commonly used standard of categorization for the magnitude of effect originated with Cohen’s (1977) guidelines and are elaborated on by Lipsey (1990) where -.10 to .20 comprise small effects; .21 to .55 are moderate or medium effects; and .56 to 1.2 constitute large effects (p. 54). Vargas and Camilli (1997) point out that an effect size of one (1.0) means the average person in the treatment group exceeds 84% of the control group on some outcome measure, assuming a normal distribution.

Prior to analysis, it is necessary to determine the unit of analysis for study and some common examples include using samples, research settings, individual comparisons, or studies as units. The present investigation used the latter as the unit for
analysis, and a study is operationally defined as the difference between the dependent measures of one treatment condition vs. another. When multiple dependent measures were used, an effect size for each dependent measure was generated, with potential to yield many effect sizes per study. Effect sizes for each group were then averaged and that mean represented the results for each group. This was in the form of pre and posttreatment differences between EMDR and a control or alternative treatment group. However, when more than one treatment condition was compared in a given investigation, each treatment condition was compared with the EMDR condition for separate effect size calculations. Therefore, a given article may consist of two or more studies as units for analysis. For example, if a study compares EMDR vs. relaxation training vs. control, two studies (and therefore two effect sizes) were be generated for a given measure: One showing differences between EMDR and the control and another reflecting differences between EMDR and relaxation.

To estimate population effect sizes, the d index described by Hedges and Olkin (1985) was used. This metric is ideally suited to the present meta-analysis as it is typically used in group comparison research that employs t and F tests. The computational formula for simple d is:

\[ d = \frac{X_1 - X_2}{\text{pooled standard deviation}} \] (Cooper, p.128). When pre and posttesting are used the formula translates to:

\[ d = \frac{(X \text{ post} - X \text{ pre of exp. Gr}) - (X \text{ post} - X \text{ pre of control gr})}{\text{pooled s.d.}} \]

When only t tests are reported, the following formula will be used:

\[ D = t \left( \frac{1}{n \text{ of exp.}} + \frac{1}{n \text{ of cont.}} \right)^{1/2} \]

The control group standard deviation is commonly used as the estimate for the population variance in the calculation of effect sizes. This assumes, under random
sampling and assignment, that the two comparison groups share the same variance and this variance is an accurate estimate of the population variance. Hedges and Olkin (1985), however, have shown that there is bias in this estimator, particularly due to sampling error artifact. Since psychological efficacy studies typically employ small cells in comparative research, the present study utilized the pooled standard deviation calculation developed by Hedges and Olkin (1985). The formula is operationally defined by:

\[ \text{SD pooled} = \frac{(n \text{ of control} - 1)(\text{sd of control posttest}) + (n \text{ of exp.} - 1) \ (\text{sd of exp})}{\text{Total n for both groups} - \text{df}} \]

Calculation of \( d \), however, is not as simple as the formula suggests. Hedges and Olkin (1985) point out that sample effect sizes overestimate treatment effects in the population due to sampling error. This is minimized, but not removed by using the pooled standard deviation. When studies use very large sample sizes with sound statistical power, the bias is reduced. However, these authors recommend employing a correction factor and their formula was used in the present investigation due to the relatively small sample sizes used in the primary studies. This was calculated for each study following the calculation of \( d \) and using the formula:

\[ c(m) - 1 \times 3 / 4m - 1 \] where \( c(m) \) is the correction factor and \( m \) = the degrees of freedom equal to \( n \) experimental + \( n \) control – 2 using the pooled standard deviation.

Once the correction factor is found it is then multiplied by \( d \) and will reflect an effect that more closely parallels the population effect despite small sample size. This is referred to as the unbiased estimator of effect size (Hedges & Olkin, 1985, p. 81).

The third step in determining individual effect sizes concerns weighting the effect sizes. Although thus far potential bias due to primary study sample sizes has been
addressed, other potential confounds exist that could bias d. This is related to the levels of variance associated with each study and non-independence of multiple dependent variables. The latter dilemma is particularly relevant to the present investigation as most studies employ multiple dependent variables.

Whereas the correction factor corrects for sample size, weighting procedures have been developed by Hedges and Olkin (1985) in order to correct, or weight the study’s characteristics along these lines. Essentially, what these procedures do is allow for the studies to be differentially weighted in the aggregation of effect sizes based on each study’s characteristics in these areas. For example, studies with multiple comparisons will be weighted to control for non-independent data so that they can be aggregated fairly with studies that compare a single group with a control group on a single dependent variable.

The weighting procedure concerned studies with a large degree of variance. Of course, it is more desirable and generalizable to find low amounts of variance associated with a study so that the independent variable can be more closely associated with treatment effects. There is any given number of sources of variance in a study. Although the meta-analyst cannot control for these in the primary studies, weighting to minimize variance is an a priori method of having more precise, less variant studies contribute more than studies with high variance. The weighting is calculated on the corrected effect size so that variance is accounted for before the effect size is aggregated in the meta-analysis. Hedges & Olkin (1985) recommend the inverse of the variance of each d using the formula:

\[
W_i = \frac{2(n_i1 + n_i2)n_i1n_i2}{2(n_i1-n_i2)^2 + (n_i1)(n_i2)(d_i \text{ squared})}
\]
Where $i=\text{individual study and } d_i$ is the corrected ES. The variance weight obtained through this calculation is then multiplied by the corrected $d$.

The problem of multiple outcome measures addressed earlier can also be weighted for a more accurate estimate of effect. This is also done sequentially, after calculating $d$, multiplying it by the correction factor, and multiplying that by the first weight. Weighting for multiple outcome measures is simply multiplying by the reciprocal of the number of effect sizes per study so that each measure contributes equally to the final effect size of each study. In other words, if six dependent measures are used in one study, this weight would equal $1/6$ toward the total effect. Cooper (1998) points out that this allows for inclusion of all dependent variable information while at the same time reducing problems posed by non-independence of data. Thus, the formula is:

$$W_{ni} (ni=\# \text{ effects per study}) = 1 / ni.$$ 

This weight is then multiplied by the corrected, previously weighted $d$ to contribute a second weighting to help account for the effect of multiple, non-independent measures.

Statistical Analysis of Effect Sizes

The fifth step in meta-analysis involves statistically aggregating all of the effect sizes culled from the primary studies. This entails calculating an overall estimate of effect of the independent variable and determining whether or not that effect size is significantly different from zero at some pre-specified level of significance (.05 in this study). Next, the moderator variables' effect sizes are aggregated and tested for statistical significance as well as descriptive trends. Although there are many steps involved in this analysis, there are essentially two objectives which relate to the research
questions. The relevant information sought by the present investigation concerned an overall estimate of effect as well as estimates of effect for each moderator variable.

As stated previously, the fixed effect model was used which assumes that the effect sizes coming from individual studies estimate a fixed population value as opposed to a random one. Because EMDR is based on specific treatment protocols and uniform training (as opposed to a more broad-based therapy such as the category of psychodynamic therapy), and because specific diagnostic criteria are used, there should be little outside of sampling error to contribute to random effects. Therefore, it was reasonable to assume that the variability in study results should be explainable by the relationship between the independent variables and treatment effects. There always exists the possibility that threats to internal validity, particularly treatment fidelity, will blur the lines between the treatment intervention and effect sizes. This, however, was examined through the coding described earlier.

Upon computing an average effect size based on each study’s individual effect size, effect sizes for each independent variable (meaning the overall efficacy question and each moderator) were averaged to get an estimate of effect. Then, the total effect size estimates were calculated using the sum of the weighted d / the total weight. The best way to organize the data to expedite this process was to create a spreadsheet documenting by column the first effect size, correction, and weight so that summary data like sums and averages can be done stepwise. The example provided by Hedges and Olkin (1985, p. 164) will be used as the data organization technique of the present study.

Following calculation of the total effect sizes, the Q statistic, a homogeneity test, was calculated. This was the most interesting and informative part of the analysis akin to performing an F-test in primary studies. The term homogeneity refers to whether the
variance of each group of effect sizes, such as the effect sizes from the EMDR treatment group, is accountable by the independent variable or whether there are other influences responsible for effect size differences. Like descriptive statistics, effect sizes will likely vary somewhat due to things like sampling error and artifact. Once all of the corrected, weighted effects are generated, they are analyzed to test for within and between level variances along each of the different independent variables through homogeneity testing (the Q test). However, Q enables comparison of the observed variance with expected variance so that it can be determined at some pre-specified level (in the present study .05) whether the between group variance reflects true group differences. The Q test essentially determined whether or not the treatment group(s) truly estimated the underlying population value. “If homogeneity is found in each of the groups tested, then each group can be assumed to represent a population with a fixed effect size and the Q can be used to see if there are true differences between groups” (Durlak & Lipsey, 1991, p. 319). This metric is well-suited for fixed effect analysis and is widely used due to its interpretability. The operational definition for Q in the present study is expressed through the formula:

\[
Q_{\text{total}} = \frac{TWDS (\text{total}) - (TWD)}{TW} \text{ (where } T = \text{ total and } D = \text{ aggregated effect sizes)}.
\]

Basically, Q tests a collection of effect sizes to test for the null hypothesis. In the present case the null hypothesis would state that that all effect sizes are equal and therefore no treatment differences exist. Analyzing all effect sizes together makes this determination. Similarly, all of the effect sizes in each of the other independent variables tested the null hypothesis to see if moderating influences were operating and whether or not the null hypothesis for each moderator variable under consideration could be rejected.
Hypotheses Tested

In the present investigation, Q-testing was used to determine if statistical support could be found for the following sub-hypotheses:

1. There will be differences showing larger treatment effects for the EMDR treated group than for controls.

2. There will be differences favoring the effect sizes of EMDR in non-combat related studies compared to combat-related PTSD studies.

3. There will be no differences found between the treatment effects of studies employing a larger number of EMDR sessions compared to studies using a smaller number of sessions.

4. There will be differences in the magnitude of effect favoring verbal report dependent variables to physiological dependent measures.

5. Older EMDR studies will show a stronger magnitude of effect than more recent studies.

6. There will be no difference in the magnitude of effect between EMDR therapy and the same therapy without the bilateral stimulation component.

As stated previously, these hypotheses were assessed by looking at the EMDR vs. Control group studies and EMDR vs. Alternative treatment group studies separately. Results were based on whether or not significant differences in homogeneous data were found and the favored outcome was the tested variable with the larger effect size.

Despite the seeming straightforwardness of these procedures, there are many mitigating factors, both statistically and in terms of research design limitations, that can compromise findings and consequently interpretation. Forseeably, these included issues like outliers, covariance, and operative moderators not conceived of a priori. To some
extent, these were planned for within the present design. To some extent, the data
presented other challenges along the way.

It was likely that there would be heterogeneity, that is differences, in at least some
of the various effect sizes. The hope was that the independent variable, in this case a
treatment intervention, did something to reduce the overall symptomatology of PTSD
compared to untreated individuals. Based on past data such as the Van Etten and Taylor
(1998) meta-analysis, there was credible reason to expect this. However, it was also
possible that effect sizes would not be evenly distributed, such as large positive effects
from one study canceling out large negative effects from another study. If this were the
case, a mean effect size would not accurately portray the data. As Durlak and Lipsey (19
91) stated, “The basic challenge for analyses of this sort involves dealing appropriately
with the heterogeneity of the observed effect sizes and the differences among the studies
upon which those effect sizes are based” (p. 315).

The homogeneity test of course determined whether or not there were real
differences in effect size due to the various independent variables. In order to detect and
examine the effect size distribution, (and potential presence of outliers) however, a stem
and leaf plot and histogram illustrating the distribution of effect sizes was executed
following the calculation of effect sizes of all individual studies. Outlying data points
were identified and each study was examined closely in order to find relevant causal
information. Such information was also examined through coded variables such as
threats to internal validity, sample size, and the like. This included investigating new
moderator variables in order to account for heterogeneity or even just a lack of
differences. In the present investigation, these moderating influences were called
“supplementary moderators” since they were not included in the five sub-hypotheses.
When the data was significantly heterogeneous, meaning the variance was unable to be accounted for by sampling error, there were two options. The first and most practical will be to use the next level of homogeneity analysis for subgroups of effect sizes so that it is possible to find subgroups (and therefore other moderating influences not conceived apriori) that were homogeneous. Essentially, this level of analysis tests subgroups until a homogeneous subgroup is identified. This was necessary on only several occasions, each of which resulted in reorganizing (typically “collapsing”) the data.

Alternatively, it would have been possible to switch gears and turn to a random effects model of meta-analysis with related statistical analysis available to better suit the data. The sustenance of this model will not be reviewed here as the present study had based the choice of fixed effects model on thorough and sound research design supported in the previous sections. Fortunately, this more radical procedure was not required due to the strong levels of homogeneity reached in the analyses.

The sixth and final step in meta-analysis concerned conclusions and interpretations. This is the point in the investigation where the data were summarized to answer the research questions that prompted the study. Durlak and Lipsey (1991) highlight three main points for inclusion: Explaining non-significance, restricting conclusions, and uncovering study limitations.

Non-significant findings are important to research and the present study resulted in some differences between EMDR and controls that were not significant. There are a variety of possibilities for this, including that non-significance may represent the true state of affairs in the population or conversely, that the statistical power of the meta-analysis was too weak to detect true treatment effects. Like primary studies, power in
meta-analysis is determined largely by sample size and also by sensitivity of measures, size of treatment effects, and alpha level. Unfortunately for meta-analysis, sample size is determined by number of participating studies in the analysis, not by total participants. The present study addressed significant and non-significant difference scores in chapter 4 as both were considered important findings. Additionally, statistical trends were addressed since trends often have the potential to reach statistical significance given increased power.

Findings from the present study of course were restricted to the PTSD populations targeted. It is for this reason of generalizability that so many of the non-PTSD diagnosed studies were excluded. The final section of the results delineates this population more specifically after the collection of studies is complete. Finally, and again based on the collection of primary studies used in the analysis, limitations to both the database and the research were explored. From this and from information and influences that arose throughout the study, questions for future research were generated.

Summary

The hypotheses investigated in the present study concerned (a) the overall efficacy of EMDR therapy in reducing the symptoms of PTSD, (b) the effectiveness of EMDR for combat vs. non-combat related PTSD, (c) the influence of treatment duration on treatment effects, (d) the difference between client and therapist self-report measures and physiologic measures, (e) differences in effects between earlier and more recent EMDR intervention studies, and (f) the role of bilateral stimulation in treatment outcomes.
The present research design employed a fixed effects model of meta-analysis with methods for combining effects across studies. The literature review resulted in the collection of 36 primary studies of EMDR in treating trauma, 21 which met criteria for the present study. The literature review further revealed controversy regarding the efficacy of EMDR with some studies purporting to find large treatment effects and others finding no significant differences based on EMDR as an independent variable. EMDR is the most widely researched treatment for PTSD to date and there are charged debates in the literature regarding the claims of EMDR proponents on both the overall efficacy of EMDR and related variables around its treatment. The present meta-analysis contributed towards resolving current controversies by finding overall estimates of effect as well as exploring moderating influences (that is, the sub-hypotheses).

Based on findings obtained in the literature review regarding pertinent issues and variables in the therapy, coding sheets were developed in order to capture the important features of each study (see Appendix C). These variables were numerically coded in order to permit for analysis with key variables being tested as moderator variables. Additionally, a threats-to-validity sheet (see Appendix D) was developed based on Stanley and Campbell’s (1962) summary in order to further analyze the effects of study integrity on effect size.

A unit of study in the present investigation consisted of an experiment comparing EMDR treatment to another group on a dependent measure. This took the form of standardized difference scores between EMDR and a control group or an alternative treatment group. Therefore, a given article could produce more than one unit of study depending on how many dependent variables were used. Effect sizes were calculated for each dependent variable in each study using the d index of Hedges. The unit of variance
used in the formula was the pooled standard deviation as described by Hedges and Olkin (1985). Following the original calculation of effect sizes, sampling error was corrected for by calculating Hedges correction factor and multiplying it by the effect size. Next, this product was weighted to account for multiple outcome artifact. Following these procedures, each effect size was ready to use in the data analysis.

The effect sizes distribution was then examined using graphic organizers to identify outliers, distribution shape, and match of fit with proposed statistical procedures. When the data met the assumptions outlined in the study, the analysis continued with calculation of an average effect size for each independent variable. There were times, however, when data needed to be collapsed or partitioned to reduce heterogeneity. The Q test for Homogeneity was conducted to assess for homogeneity in each group and see if in fact significant between group differences exist and the null hypotheses could be rejected. The magnitude of effect was assessed using the arbitrary guidelines developed by Cohen (reported in Lipsey, 1990). Finally, conclusions and interpretations were made based on the collection of data.
CHAPTER IV

RESULTS

Description of Primary Studies Used in Meta-Analysis

Meta-analysis allows for different measures to be used as units of study and the present study employed the comparison of two groups, that is the difference score, as the unit of study. In adherence to selection criteria and procedures outlined in chapter 3, a total of 21 primary studies were employed. All but three data sets were derived from published articles. The Levin, et al. (1999) article featured outcomes presented as a case report when published. However, group data were obtained via personal communication with the primary research assistant, J. Hopper (personal communication, January 3, 2001) and were therefore included in the present investigation. Carpenter’s (1999) study was a well-conducted thesis that fit all inclusion criteria. Power, McGoldrick, and Brown (personal communication, February 27, 2001) was a controlled study conducted for the Scottish Home and Health Department which was reported by the authors to be submitted for publication in the Journal of Clinical and Consulting Psychology. All respective journals were from the areas of psychology, psychiatry, behavior therapy, or psychotherapy in general. Table 1 depicts represented journals.

The descriptive characteristics of each article included multiple variables such as region, year, publication source, and other study-specific information. A summary of these descriptive features is presented in Table 2.
Table 1

**Journals Represented in Meta-Analysis**

<table>
<thead>
<tr>
<th>Journal</th>
<th>Number of Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavior Therapist</td>
<td>1</td>
</tr>
<tr>
<td>Behavior Therapy</td>
<td>2</td>
</tr>
<tr>
<td>British Journal of Psychiatry</td>
<td>1</td>
</tr>
<tr>
<td>Bulletin of the Menninger Clinic</td>
<td>1</td>
</tr>
<tr>
<td>Clinical Psychology and Psychotherapy</td>
<td>1</td>
</tr>
<tr>
<td>Comprehensive Psychiatry</td>
<td>1</td>
</tr>
<tr>
<td>Journal of Anxiety Disorders</td>
<td>2</td>
</tr>
<tr>
<td>Journal of Behavior Therapy &amp; Experimental Psychiatry</td>
<td>4</td>
</tr>
<tr>
<td>Journal of Consulting and Clinical Psychology</td>
<td>1</td>
</tr>
<tr>
<td>Journal of Traumatic Stress</td>
<td>2</td>
</tr>
<tr>
<td>Psychotherapy</td>
<td>2</td>
</tr>
<tr>
<td>Unpublished</td>
<td>3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>21</strong></td>
</tr>
</tbody>
</table>
### Table 2

**Descriptive Features of Studies Used in Meta-analysis**

<table>
<thead>
<tr>
<th>Author</th>
<th>Comparison</th>
<th>Year</th>
<th># Sessions</th>
<th>Setting</th>
<th>Sample</th>
<th>Dep. Variables</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boudewyns &amp; Hyer</td>
<td>C</td>
<td>1996</td>
<td>5</td>
<td>Mixed</td>
<td>C</td>
<td>Objective</td>
<td>61</td>
</tr>
<tr>
<td>Boudewyns et al.</td>
<td>A</td>
<td>1993</td>
<td>2</td>
<td>Inpatient</td>
<td>C</td>
<td>Subjective</td>
<td>20</td>
</tr>
<tr>
<td>Carlson et al.</td>
<td>C</td>
<td>1996</td>
<td>12</td>
<td>Unknown</td>
<td>C</td>
<td>Subjective &amp;</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Objective</td>
<td></td>
</tr>
<tr>
<td>Carlson et al.</td>
<td>B</td>
<td>1998</td>
<td>12</td>
<td>Outpatient</td>
<td>C</td>
<td>Subjective</td>
<td>35</td>
</tr>
<tr>
<td>Carpenter</td>
<td>A</td>
<td>1999</td>
<td>4</td>
<td>Outpatient</td>
<td>N</td>
<td>Subjective</td>
<td>10</td>
</tr>
<tr>
<td>Devilly &amp;</td>
<td>A</td>
<td>1999</td>
<td>8</td>
<td>Outpatient</td>
<td>N</td>
<td>Subjective</td>
<td>23</td>
</tr>
<tr>
<td>Spence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Devilly et al.</td>
<td>B</td>
<td>1998</td>
<td>2</td>
<td>Unknown</td>
<td>C</td>
<td>Subjective</td>
<td>51</td>
</tr>
<tr>
<td>Forbes et al.</td>
<td>C</td>
<td>1994</td>
<td>4</td>
<td>Outpatient</td>
<td>N</td>
<td>Subjective &amp;</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Objective</td>
<td></td>
</tr>
<tr>
<td>Jensen</td>
<td>C</td>
<td>1994</td>
<td>2</td>
<td>Mixed</td>
<td>C</td>
<td>Subjective</td>
<td>25</td>
</tr>
<tr>
<td>Levin et al.</td>
<td>C</td>
<td>1999</td>
<td>3</td>
<td>Unknown</td>
<td>N</td>
<td>Subjective &amp;</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Objective</td>
<td></td>
</tr>
<tr>
<td>Marcus et al.</td>
<td>C</td>
<td>1997</td>
<td>3</td>
<td>Outpatient</td>
<td>N</td>
<td>Subjective</td>
<td>67</td>
</tr>
<tr>
<td>Montgomery &amp; Aiyon</td>
<td>A</td>
<td>1994</td>
<td>6</td>
<td>Outpatient</td>
<td>N</td>
<td>Subjective &amp;</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Objective</td>
<td></td>
</tr>
<tr>
<td>Pitman et al.</td>
<td>A</td>
<td>1996</td>
<td>10</td>
<td>Outpatient</td>
<td>C</td>
<td>Subjective &amp;</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Objective</td>
<td></td>
</tr>
<tr>
<td>Power et al.</td>
<td>B</td>
<td>2000</td>
<td>4.2</td>
<td>Outpatient</td>
<td>N</td>
<td>Subjective</td>
<td>72</td>
</tr>
<tr>
<td>Renfrey &amp;</td>
<td>A</td>
<td>1994</td>
<td>4</td>
<td>Outpatient</td>
<td>N</td>
<td>Subjective &amp;</td>
<td>23</td>
</tr>
<tr>
<td>Spates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Objective</td>
<td></td>
</tr>
<tr>
<td>Rogers et al.</td>
<td>A</td>
<td>1999</td>
<td>1</td>
<td>Inpatient</td>
<td>C</td>
<td>Subjective &amp;</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Objective</td>
<td></td>
</tr>
<tr>
<td>Rothbaum</td>
<td>C</td>
<td>1997</td>
<td>3</td>
<td>Outpatient</td>
<td>N</td>
<td>Subjective</td>
<td>18</td>
</tr>
<tr>
<td>Silver et al.</td>
<td>C</td>
<td>1995</td>
<td>1</td>
<td>Inpatient</td>
<td>C</td>
<td>Subjective</td>
<td>100</td>
</tr>
<tr>
<td>Vaughan &amp;</td>
<td>B</td>
<td>1994</td>
<td>3</td>
<td>Outpatient</td>
<td>N</td>
<td>Subjective</td>
<td>36</td>
</tr>
<tr>
<td>Armstrong et al.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vaughan et al.</td>
<td>C</td>
<td>1994</td>
<td>3</td>
<td>Mixed</td>
<td>N</td>
<td>Objective</td>
<td>10</td>
</tr>
<tr>
<td>Wilson et al.</td>
<td>C</td>
<td>1997</td>
<td>3</td>
<td>Outpatient</td>
<td>N</td>
<td>Subjective</td>
<td>32</td>
</tr>
</tbody>
</table>

N=642
**Note.**

Comparison: C represents control group comparison; A represents alternative treatment comparison; and B represents inclusion in both the control group comparison and alternative treatment comparisons (i.e. when a primary study compared EMDR to both a no treatment control and an alternative treatments).

Sample: C represents a sample with combat related PTSD and N represents a non-combat related PTSD sample.

# Sessions: Number of sessions reflects the minimal number of sessions specified, unless a computed average number of sessions were reported for a study. For instance, if number of sessions was reported as “between 3-5 sessions,” the number three was used in the analysis unless the authors specified an average, such as “with an average of 4.2 sessions.” All numbers of sessions were rounded to whole numbers.

Dependent Variables: Objective outcome measures were all physiologic indices, such as heart rate or skin conductance, which were assessed by absolute measurements from objective instruments. Subjective outcome measures included all verbal report indices, including client self-report or clinician administered verbally assessed instruments.

The total number of individual participants accounted for in the meta-analysis was 642. Because most studies did not report participant information beyond ranges for the group, separate partitioning of age, gender, and trauma variables beyond the combat relatedness was not possible.
Data Analysis

The intent of this study was twofold; first, to estimate the overall efficacy of EMDR therapy with a PTSD diagnosed population. Secondly, to answer specific questions related to efficacy, that is, to assess whether or not there are moderating variables that influence the degree of efficacy in EMDR. This included investigating variables within the actual treatment as well as the treatment with sub-populations within the group identified as having Posttraumatic Stress Disorder. The primary sub-populations of interest included those whose traumatic experience is combat-related and those whose PTSD is not rooted in combat experience. This was a critical research question since the APA Task Force on Psychological Interventions (reported independently by Chambless, et al., 1998) made a clear distinction when it credited EMDR as “probably efficacious in the treatment of civilian PTSD” (p. 9) but not for combat-related PTSD. Analysis of this variable quantifies the difference, if in fact, a difference exists. Figure 1 depicts the proportion of combat-related v. non-combat related studies used in the meta-analysis.
Figure 1. Proportion of combat and non-combat related PTSD studies used in the meta-analysis.
The other moderating variables investigated for their potential influence in the clinical and measurable effectiveness of EMDR as treatment, included whether or not there were differences in effect size based on the number of sessions, use of verbal vs. physiologic outcome measures, and recency of study. The latter variable is important considering the early claims made by EMDR research. An important part of testing the moderator variables was to see whether or not the bilateral stimulation component of EMDR uniquely contributed to its effectiveness. EMDR is comprised of a manualized treatment that combines elements from many types of therapies, but is most striking for its use of therapist-induced eye movements and other bilateral stimulation techniques. There were several studies in the past decade that investigated the role of the eye movements making testing of this possible at the meta-analytic level. In general, these potential moderators were investigated in light of current controversies in the literature.

After meeting selection criteria previously outlined (see chapter 3), studies were separated into two different data sets: EMDR versus no treatment control and EMDR versus alternative treatment comparison. The EMDR versus no treatment (EMDR/C) and EMDR versus alternative treatment (EMDR/AT) comparisons were analyzed separately. This is because of different expectations of results for the two comparisons. It is reasonable to expect larger differences when comparing EMDR with no treatment than when comparing EMDR with other types of treatment administered to participants. Indeed it was surprising to see that the only other meta-analysis found that included EMDR (among other treatments) did not partition data this way to avoid bias. However, when a single investigation compared EMDR with both an alternative treatment group and a no treatment control, each comparison was included in the respective analysis since the unit of study was the difference scores between treatment and control.
Once a study was included in either the EMDR/C or EMDR/AT analysis groups, it was coded according to a number of variables related to the research questions under investigation. In addition to the moderator variables, descriptive study characteristics were coded to yield additional data of interest and also to allow for sub-analyses that may be relevant as either potential confounds or moderators not conceived of a priori. See coding form located in Appendix C.

After a study was coded, all data for that study was inputted into a spreadsheet using Microsoft Excel. There were two spreadsheets created, one for the EMDR/C data and the other for the EMDR/AT data. Once the data were included in the spreadsheets, any given column, that is any designated study variable, were eligible to become an independent variable of the meta-analysis. For instance, one of the moderators of interest was whether or not the participants were civilian or combat-related in their PTSD. Therefore, there was a column in which each piece of data was labeled as either combat or non-combat and an analysis could be run according to this as an independent variable. However, there were over a dozen descriptive columns, which meant that multiple comparisons could be made on variables such as location of study, setting of study, or specific types of measures.

MetaWin (Rosenberg, Adams, & Gurevitch, 2000) was the statistical software program used to analyze the data contained in the spreadsheets. The program was designed specifically for meta-analyses. SPSS (1996) was also used as needed to make minor statistical calculations for input and for graphical representations of the data. This chapter will describe the EMDR/C and EMDR/AT results separately, however, the procedures utilized were the same. As documented in chapter 3, each study contributed the means and standard deviations of dependent variables employed. When a study
instead reported other usable data such as t-statistics, these were transformed into
difference scores. The unit of study was the difference score found between pre and
posttest scores of the two different groups.

This resulted in an effect size representing each unit of study. The effect sizes
were represented by Hedges's d and were calculated for each study and then added as a
separate column on the spreadsheet. Importantly, effect sizes were weighted using the
formula provided in chapter 3 such that each study contributed equally to the analysis
irrespective of the total number of outcomes a study provided. This was important in the
present investigation as it reduced the bias created by multiple outcome measures. For
instance, if an author reported three scores from the Impact of Event Scale that were used
as discrete units, the measurement weight would in effect divide this total number (i.e.,
1/3 for each subtest) to consider this ratio in the total analysis.

After effect sizes were calculated, an average was computed to represent the
estimate of overall efficacy of EMDR in treating PTSD (again, this included an EMDR/C
statistic and an EMDR/AT statistic). This average effect size for each collection of
studies represented an answer to the first question of the present investigation: Is EMDR
effective in the treatment of PTSD and if so, what is the magnitude of effect? The Q test
was used as both the summary statistic for the collection of effect sizes and the statistical
index used to determine whether or not true differences exist between subsets of effect
sizes. In the Q Test, Q Total (Qt) assesses whether or not a data set is homogeneous,
and the present investigation used a .05 probability value for this. Q Between (Qb) was
the index used to determine whether or not significant differences were found between
comparison groups, again, at the .05 probability level. Thus, when Qb is found to be
significant at the .05 probability level, then there are differences between groups that can be attributed to the independent variable of interest.

In terms of interpretation, when homogeneity is found, then the single estimated effect size is a good representative of the collection of effect sizes. Thus, if homogeneity is found, then the single estimated effect size is a good representative of the collection of effect sizes. When used to summarize specific independent variables like the sub-hypotheses tested, Qb was used to test whether or not a true difference exists related to the independent variable of interest. This can be interpreted similarly to ANOVA in that homogeneity reflects accountability from the independent variable of interest. When Q finds effect sizes to be homogeneous and Qb finds between group differences, it can be estimated that there are true differences in the population accounted for by the independent variable.

Description of Pre-Selected Moderator Variables

Study data were coded along several lines, including descriptive participant information and research design variables. Year ranged from 1994-1999 with an additional coding of either older study or more recent study. The older category consisted of the years 1990-1994, the first five years of EMDR research, and the more recent category constituted of the next five years of study. Setting was coded threefold, as either taking place in an inpatient setting, with outpatient participants, or with a mixture of in and outpatient study participants. Several studies did not provide this information and were coded as "unknown." The number of sessions was coded as a discrete number and this was done in one of three ways dependent on how data were reported. When studies utilized a set number of sessions, this number was entered. If a study only
reported a range of sessions, such as in instances when different participants required varying amounts of treatment, the minimum number of sessions was coded. If, however, such studies reported the mean number of sessions across the group, then that number was coded instead of the minimal number of sessions. An additional coding for moderator variable analysis arbitrarily divided the number of sessions into the following categories: Small (1-3 sessions); Moderate (4-7 sessions); and Large (8-12 sessions).

Dependent variable type was initially simply coded as either a verbal measure or an objective measure. However, these categories were found to be highly heterogeneous and therefore the final analysis utilized the three categories of self-report measures, therapist-administered measures, and objective measures. The first two categories tap into the participant's subjective experience of PTSD related symptomatology, either as assessed by the person or the evaluating therapist. Objective measures, on the other hand, reflect data measured by objective, absolute or ratio scales such as temperature or heart rate. This was considered a useful category because of the abundance of criticism in the EMDR literature regarding the potential for demand characteristics in the SUD and VOC assessments as well as scales administered to participants. There was an uneven distribution of dependent variable type and so a supplemental moderator analysis was conducted in order to categorize dependent variables along a continuum of symptoms. This is discussed in further detail in the supplemental moderators section of the present chapter.

Trauma type was also dichotomized, using the categories of either combat-related trauma or non-combat related trauma. Figure 1 depicted the ratio of combat and non-combat studies analyzed. When studies included both, the non-combat category was used to more accurately reflect the range of trauma experienced by participants. A more
detailed sub-categorization was desired, however, very few PTSD studies featured more specific sub-populations.

A final category of mention concerns the coding for internal validity ratings. The purpose of this variable was to ascertain whether or not perceived study integrity has any association with effect size. A frequent criticism of meta-analysis is the garbage in–garbage out problem discussed in chapters two and three. The present study maintained integrity along these lines in two ways: Appropriate and sufficiently stringent inclusion and exclusion criteria and coding included studies to rate for within-study internal integrity. Using the internal integrity rating system in Appendix 3, each study was awarded a number of points representing its degree of internal validity. The basis of the list was the eight classes of extraneous variables delineated by Campbell and Stanley (1963). Other internal validity concepts included in the rating system were controls for demand characteristics (as highlighted by Borg & Gall, 1989, p.193) and controls related to the implementing clinicians (described in Waltz, Addis, Koerner, & Jacobson, 1998). These quantitative internal validity ratings were then analyzed to see if there were differences in effect size associated with the strength of primary studies' internal validity.

Statistics of Total Effect Sizes

Effect Size Analysis

There were 51 individual effect sizes computed for the EMDR vs. Control comparisons and 67 for the EMDR vs. Alternative Treatment comparisons for a total aggregate of 118 effect sizes. Figures 2 and 4 show the respective frequency distributions for effect sizes found in all of the studies. Each effect size was corrected for sample bias and the collection of effects for each data set were then weighted for number
of outcomes per study using the formulas presented in chapter 3. Therefore, from this point forward the term effect size refers to the corrected and weighted standardized difference score.

**Pooling Effect Sizes**

In order to account for each study's variance, effect sizes were weighted according to the formula presented in chapter 3 such that studies with smaller variance had larger weight in the aggregate and studies with a large amount of variance received less weight. Additionally, study weight helped control for multiple comparisons made with an experimental group. Essentially this meant that the study weight was inversely proportional to the number of comparisons made with a single group.
Outliers

After calculation of effect sizes, a boxplot was created to examine outliers. Since graphical depiction via the boxplot revealed only two extreme scores, it was not necessary to use a statistical rule to identify and omit outliers. Instead, these were excluded from the data sets simply based on a lack of reasonable position within the distribution. This included one score from the Jensen (1994) study that showed an aberrant standard deviation, possibly a misprint. The other outlier was from the Rogers et al. (1999) investigation and was omitted because it was so extreme as to be considered "renegade data," particularly when compared descriptively to the same dependent variable information provided by other studies. The index used was the SUD, however, the drastic decrease from a pre-treatment mean of 7.4 to a posttreatment mean of .4, particularly since the control groups scores rose in comparison, violated statistical assumptions necessary for analysis. Generally, however, since the effect sizes from both comparisons were found to be homogeneous, outliers were not prevalent.

EMDR vs. Control Group Analysis

The average effect sizes for the EMDR vs. Control are presented in Figure 2. The distribution of effect sizes is organized according to descending values and shows a relatively even distribution of effect sizes. Individual effect sizes ranged from -2.77. Most effect sizes fall to the right of zero indicating positive treatment effects for the EMDR group compared to the no treatment control. The average of all EMDR vs. Control studies was .50 which was found to be significantly different from zero at alpha = .05. This is represented in Figure 2 by the small line midway through the effect size values. According to Cohen (1977) this is in the medium range of effects. Bootstrap
confidence interval of .20-.81 for the grand mean does not include zero and Q was found to be homogeneous indicating that the effect size is significantly different than zero and the average accurately represents the distribution of effect sizes.
Figure 2. Effect size distribution for EMDR / Control Group comparison

Note:

Figure 2 includes 51 individual effect sizes, 90% confidence intervals of each effect size, and the grand mean of effect sizes represented by small line midway through figure.
Moderator Variables

Five potential moderator variables were conceived of apriori as potential determinants for the efficacy of EMDR in the treatment of PTSD. Those moderator variable research questions were:

1. Is EMDR more effective in the treatment of combat-related trauma vs. non-combat related trauma?

2. Are there differences in effect sizes between verbal report measures and physiological indices of change?

3. Are there differences in effect sizes between earlier and more recent studies?

4. Does the bilateral stimulation contribute significantly toward treatment effects?

5. Are there differences in effect sizes based on the number of treatment sessions administered? The hypothesis and results of each moderator variable analysis will be discussed individually.

In addition to these treatment-based variables, it is possible that research design variables could have a moderating influence on the distribution of effect sizes. Additionally, there were some variables outside of the major hypotheses that were analyzed because they were both interesting and relevant to clinical significance. For this reason and as outlined in chapter three, a number of research design characteristics were coded and analyzed for influence. The supplementary treatment and research design moderators will be discussed separately for each collection of studies.

Support was found for the first sub-hypothesis, which was the prediction that larger treatment effects would be shown for non-combat related studies than for combat related PTSD studies. Type of trauma was analyzed by performing Q testing for studies coded as combat related trauma compared to studies coded as non-combat related trauma.
Q total (Qt) was considered significant when the value exceeded chi square critical values based on the number of studies analyzed. Qt was found to be homogeneous indicating that the data for both groups represented the total distribution for effect sizes. The moderator effect for type of trauma was found to be significant, with Qb=12.31, (p.<.05). Non-combat results showed a large average d=.87 and combat effect sizes = -.04. This result was consistent with the hypothesis that EMDR would show larger effects in the treatment of non-combat related PTSD than for combat related PTSD. Most studies did not report type of trauma more specifically than whether or not it was combat-related. Therefore, further analysis to determine source of heterogeneity of this categorical was not possible.

The next sub-hypothesis tested was the prediction that greater effects would be found on verbal dependent measures than on physiologic outcomes. In the initial homogeneity test for differences between the effect sizes found on verbal report measures and objective dependent measures used in primary studies, heterogeneity remained after analysis. This was likely due to the combination of a skewed ratio of verbal measures to objective measures and the large amounts of variance associated with the broad category of verbal dependent variables. Data were then collapsed into the categories of self-report (d=.62), therapist-administered (d=1.01), and objective (d=.72) measures. Although the three-way analysis distributed individual effects evenly enough to reach homogeneity, inclusion of only two objective dependent measures must be considered in interpretation. This proportion is illustrated in Figure 3. These findings supported the second hypothesis in that the physiologic measures were significantly lower in their effects than the verbal report measures, but again, with only two sets of objective effects this conclusion must be considered preliminary until further replication with larger data sets are completed by
future studies. Data were homogenous and Qb was significant (p<.05). Post hoc analysis showed significant differences between the objective measures with both the self-report and therapist administered measures. This was consistent with results reported across the primary studies and may be due to the lack of sensitivity of physiological indices as well as the more generic nature of this type of dependent variable in measuring PTSD symptoms.
Figure 3. Proportion of Self-Report, Therapist-Administered, and Physiological Measures.
An additional descriptive analysis was completed to assess whether or not there were differences between outcomes that specifically measured PTSD symptoms (d=.95) vs. those that measured psychiatric symptoms in general (d=.61) vs. "other" (d=.24). The latter included physiological measures as well as general process measures. Data were found to be homogeneous and Qb significant (p<.05). Post hoc testing indicated significant differences favoring PTSD-specific category than the category of "other." This is an important finding that suggests that EMDR treatment for PTSD resulted in symptom reduction in the large range of effects for symptoms specific to PTSD.

It was hypothesized that earlier EMDR studies in the treatment of PTSD would show larger treatment effects than more recent studies. Recency of study showed data were again homogeneous, however, contrary to the hypothesis, no between group differences were found (Qb=1.13, p=.29). As predicted, however, the trend showed greater effect sizes for earlier studies (d=.82) than for more recent studies (d=.46). It was then hypothesized that there would be differences between individual years and this was found to be the case. Differences occurred between the years 1998 and 1996 (p<.05) with larger effects shown for the year 1998 (d=.50) and between effect sizes found for the years 2000 and 1996 (p<.05) with larger effects shown for the year 2000 (d=1.26).

Session size was tested in order to see if there was a main effect for the dose response curve of EMDR treatment. It was hypothesized that there would be no differences between treatment effects based on whether s short or longer number of EMDR sessions was administered. Comparisons were made between effects found for a small number of sessions administered and for a large number of sessions because the effect size distribution for moderate administration was found to be heterogeneous. Separate analysis for moderate session size data found uncontrollable confounds causing
the lack of homogeneity, specifically that all of these effects were restricted to non-combat participants and self-report dependent measures. $Q_b$ was not significant for the large vs. small comparison and this supported the hypothesis that there would be no differences found between smaller and larger numbers of treatment sessions ($Q_b=1.19$, $p=.55$). Effect sizes were nearly identical for large number of sessions ($d=.59$) and small number of sessions ($d=.60$) and both respective confidence intervals showed that the effects were different from zero indicating that despite a lack of significance between categories, the groups differed from controls both showing medium effect sizes over the control group.

Results of main effect analysis for the general efficacy hypothesis and the four apriori sub-hypotheses are presented in Table 3. Note: the fifth sub-hypothesis was relevant only to alternative treatment comparisons and was therefore not included in the first data set.
Table 3
Summary of Pre-Selected Moderator Variables for EMDR/Control Group Comparison

<table>
<thead>
<tr>
<th>Variable</th>
<th>Effect Size (d)</th>
<th># of Effects (n)</th>
<th>Bootstrap Confidence Interval*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trauma Type</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combat-Related</td>
<td>-.04</td>
<td>20</td>
<td>-.39-.48*</td>
</tr>
<tr>
<td>Non Combat Related</td>
<td>.87</td>
<td>31</td>
<td>.69-1.06</td>
</tr>
<tr>
<td><strong>Dependent Variable Type</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Report</td>
<td>.62</td>
<td>35</td>
<td>.48-.78</td>
</tr>
<tr>
<td>Therapist-Administered</td>
<td>1.01</td>
<td>14</td>
<td>.77-1.23</td>
</tr>
<tr>
<td>Objective</td>
<td>-.72</td>
<td>2</td>
<td>-.96--.49*</td>
</tr>
<tr>
<td><strong>Recency</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Newer</td>
<td>.46</td>
<td>42</td>
<td>.14-.80</td>
</tr>
<tr>
<td>Older</td>
<td>.82</td>
<td>9</td>
<td>.41-1.26</td>
</tr>
<tr>
<td><strong>Sessions Administered</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large</td>
<td>.59</td>
<td>8</td>
<td>.31-.89</td>
</tr>
<tr>
<td>Medium</td>
<td>.31</td>
<td>11</td>
<td>-.18-.79*</td>
</tr>
<tr>
<td>Small</td>
<td>.60</td>
<td>32</td>
<td>.47-.76</td>
</tr>
</tbody>
</table>

**Note**

Probability value: p<.05

*When bootstrap confidence interval includes zero in range, effect size is not considered significantly different from zero.
Supplemental Moderator Variables

Outside of the major moderators hypothesized before data collection, there were a number of variables of interest based on data provided by the primary studies. These variables were analyzed using the Q test for homogeneity. Results of those analyses found to be homogeneous are reported beginning with the analysis of setting of study.

The setting of study was of interest due to potential clinical significance. Settings analyzed were outpatient, mixed (outpatient and inpatient), and unknown (as some studies did not report this information). Unfortunately, there were few inpatient settings employed and none in the EMDR vs. Control data comparison. The three groups were found to be homogeneous and there was a significant difference (Qb=17.48, p<.05) favoring the outpatient group (d=.84). The unknown group showed a medium effect (d=.41) and the mixed inpatient and outpatient group was not significantly different from zero (d=-.45, bootstrap confidence interval=-.80-.30). The fact that there were no exclusively inpatient groups studied, coupled with the fact that studies did not distinguish between single and multiply traumatized samples made it impossible to assess for the severity of trauma as a moderating influence in treatment effects. This is an important limitation of the current EMDR research that will be addressed in chapter 5.

It was not possible to test for the professional affiliation of the primary researcher of each study as this was not reported and could not be reliably coded for all studies. This would have been an interesting variable given the criticisms and controversies between opponents and proponents of EMDR. The only related variable possible to test in the present investigation was journal type. Journal type categories with effect sizes included anxiety/trauma specific journals (d=.59); behavioral psychology (d=.61); clinical psychology (d=.15); and unpublished studies (d=1.26). Data were homogeneous
and significant differences were found between journal types ($Q_b=10.06$, $p=.02$). Post hoc analysis found significant difference between the unpublished studies and effects found in the clinical psychology journals only. These results were within reasonable expectations in that unpublished studies are frequently associated with poor validity and such validity typically yields exaggerated effects.

The final variable tested for moderating influence was related to research design. Internal validity ratings, as described in chapter three and reproduced in Appendix 3, was analyzed in order to determine if a study’s judged research integrity had a moderating influence on treatment effects. Data were found to be homogeneous and there was a significant difference ($Q_b=9.68$, $p<.05$) found between studies with high internal validity ($d=-.24$) and studies with medium internal validity ($d=.71$). This data was consistent with the idea described in the research literature that stronger internal validity controls within a study are associated with lower treatment effects. Results for the supplemental moderator variables found are presented in Table 4.
Table 4  
Summary of Supplemental Moderator Variables Analyzed in the EMDR/Control Comparison.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Effect Size (d)</th>
<th># of Effects (n)</th>
<th>Bootstrap Confidence Interval*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Setting</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outpatient</td>
<td>.84</td>
<td>39</td>
<td>.69-.99</td>
</tr>
<tr>
<td>Mixed</td>
<td>-.45</td>
<td>7</td>
<td>-.77-.32*</td>
</tr>
<tr>
<td>Unknown</td>
<td>.41</td>
<td>5</td>
<td>.35-.47</td>
</tr>
<tr>
<td><strong>Journal Type</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety/Trauma</td>
<td>.59</td>
<td>8</td>
<td>.30-.93</td>
</tr>
<tr>
<td>Behavioral Psychology</td>
<td>.61</td>
<td>14</td>
<td>.35-.82</td>
</tr>
<tr>
<td>Clinical Psychology</td>
<td>.15</td>
<td>20</td>
<td>-.23-.65*</td>
</tr>
<tr>
<td>Unpublished Study</td>
<td>1.26</td>
<td>8</td>
<td>1.13-1.40</td>
</tr>
<tr>
<td><strong>Internal Validity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>.71</td>
<td>39</td>
<td>.58-.86</td>
</tr>
<tr>
<td>High</td>
<td>-.24</td>
<td>12</td>
<td>-.61-1.26*</td>
</tr>
</tbody>
</table>

*When bootstrap confidence interval includes zero in range, effect size is not considered significantly different from zero.

*Note.*

Probability value: p<.05
EMDR vs. Alternative Treatment Comparison

The average effect size distribution for EMDR v. Alternative Treatment studies are presented in Figure 4. Effect sizes ranged from -.90 – 1.96. The fact that more negative effect sizes were found on some of the standardized difference scores indicated that in at least some cases, the alternative treatment modalities were associated with better results than the EMDR group. This is not a surprising finding since the EMDR/Control comparison compared some treatment with no treatment and the EMDR/Alternative comparison compared EMDR with some often used and well regarded treatments for PTSD. It was also not surprising that this data set resulted in a smaller average effect size (d=.43) than the EMDR/Control comparison because it was reasonable to expect smaller outcomes when comparing two treatments with each other than when comparing one treatment with no treatment. According to Cohen’s guidelines, this average effect size is considered in the medium effect size range and since a d of .43 is interpreted as having that much more effect than the comparison group, this showed sizable effects for EMDR therapy. Data were homogenous (Q=17.01, p=1.0) and bootstrap confidence intervals did not include (.31-.56) indicating that the average effect size was both different from zero. A summary of the grand means found in both the EMDR/Alternative treatment comparison and the EMDR/Control comparison is presented in Figure 5.
Figure 4. Effect Size Distribution for EMDR / Alternative Treatment Comparison

Note.

Figure 4 includes 67 individual effect sizes, 90% confidence intervals, and grand mean represented by small line midway through figure.
Figure 5. Grand Means of EMDR/Control and EMDR/Alternative Treatment Comparisons.
Moderator Variables

The first moderator variable analyzed was type of trauma, that is, combat related trauma vs. non-combat related trauma. It was hypothesized that EMDR would show a larger magnitude of effect for non-combat related trauma than for combat related trauma. Data were found to be homogeneous indicating that the average effect sizes represented the collections of individual effect sizes. Although results followed the trend of the EMDR/Control comparison (combat related trauma d=.35 and non-combat related trauma d=.50), there was no significant difference found favoring either group (Qb=.24, p=.63). Thus, although EMDR was found to have a larger effect for non-combat trauma compared to combat trauma, the difference was not considered significant at the .05 level and therefore the hypothesis not supported. Although both effect sizes were in the medium range according to Lipsey's (1990) elaboration of Cohen's (1977) guidelines, .35 is considered the minimum effect present in the medium range while .55 is considered the upper limit of the same range (p. 54). Neither group's main effect included zero in bootstrap confidence intervals indicating that both scores differ from zero. So while trends were maintained in conjunction with the EMDR/Control comparison, when tested against other treatments, EMDR effects were not significantly different within a combat and non-combat related trauma group.

A similar result was found for the type of dependent variables employed in the EMDR/Alternative treatment comparison. It was hypothesized that there would be larger treatment effects found among verbal measures than among physiologic dependent measures. This time, there were a total of seven objective measure effects, which meant an overall more even distribution, which allowed for more confidence in findings. Aggregate effect sizes were again found to be representative of the distributions of effect
sizes (Qw=.67-12.9, p>.05), however there were no differences found between groups (Qb=.07, p=.96). This meant that effect sizes derived from the three types of dependent variables, self-report (d=.41); therapist-administered (d=.47); and objective (d=.52) did not differ from each other appreciably. All bootstrap confidence intervals did not include zero indicating that on all three types of outcome measures the results were significantly different from zero reflecting true effects from treatment. This finding did not support the hypothesis that physiological dependent variables would yield smaller effects than verbal report measures and in fact showed that medium size treatment effects were found on physiological indices.

An additional analysis was conducted to see whether there were differences in the EMDR/Alternative treatment comparison on the symptom cluster of the dependent variables. Again, data were homogeneous, but without significant differences between groups (Qb=.93, p=.63). Interestingly, on the PTSD symptom cluster the effect was smallest (d=.28) while general psychiatric symptoms and “other” symptom measures showed medium effects (d=.59 and d=.50 respectively). This was not consistent with the EMDR/Control group comparison, which found strong effects for PTSD-specific symptoms.

It was hypothesized that older studies would show larger treatment effects than more recent EMDR studies in the treatment of PTSD. Recency data coded by splitting studies into the categories of newer and older based on the median of years for EMDR studies were found to be homogeneous, but there was no difference between the average effect size of earlier studies compared to newer studies (Qb=.49, p=.48). This was contrary to the hypothesis that the early studies would show significantly larger effects based on the early claims for EMDR. However, trends did show larger effects for older
studies (d=.66) than for more recent studies (d=.39) despite the differences not reaching significance. A post hoc analysis was conducted on this variable to determine whether the specific years of study showed any differences and there were no differences found based on the year of study (data homogeneous with Qw= 14.12, p=.58).

The final sub-hypothesis was the prediction that the bilateral stimulation component of EMDR would not contribute significantly to treatment effects. Results analyzing the potentially moderating influence of the bilateral stimulation component in EMDR therapy were consistent with the original hypothesis that EMDR without the bilateral stimulation component would not significantly differ from EMDR. However, although the analysis was consistent with the hypothesis, it can not be concluded based on the present investigation that the bilateral stimulation component does not significantly contribute toward change. Limitations in the ability of primary studies to accurately measure this component preclude the ability of the present investigation to draw conclusions on this variable. This will be addressed in further detail in chapter 5, however, the very underlying presumption that a manualized treatment can be measured without a component of the protocol must be considered in reviewing results. This comparison was made in several different ways in an attempt to estimate the contribution of eye movements to the overall effect size of EMDR. First was by using the Q test to compare difference scores between EMDR and those studies that used all other components of EMDR except for the bilateral stimulation. Homogeneity was reached, however there was no significant difference found between EMDR and EMDR treatment that attempted to remove the bilateral stimulation component (Qb=.15, p=.70). Even without the eye movements, however, EMDR showed medium size effects (d=.35).

Since the bootstrap confidence interval did not include zero (.16-.55), results suggested
that when studies attempted to remove the bilateral stimulation from the EMDR protocol, effects were still significantly different from zero.

The last pre-selected moderator variable analyzed was the number of treatment sessions administered and as previously addressed, this was categorized as large, medium, and small. It was hypothesized that there would be no differences between the treatment effects for a larger administration of sessions than for a smaller number of sessions. Data were again homogeneous at the .05 level, and there were no significant differences found between groups (Qb=5.45, p=.07). Like in the EMDR/Control comparison, the hypothesis was supported as the data did not show a greater effect for more sessions.

However, because the chi square probability value was so close to the requisite .05 needed to reach significance in finding between group differences, an additional exploratory analysis was conducted in the interest of determining if significance would be obtained with increased statistical power. This was not necessary as data were homogeneous, however, calculation of a descriptive index known as Orwin’s fail-safe n was an interesting adjunct statistic since the Qb result was so close to the p value required for finding significant between group differences. Since the large session effects were of a negligible size (d=.04), the differences were likely to be between this group and the medium session size group which showed large effects (d=.85). Because these groups were ranked next to each other, collapsing into two dichotomous variables would skew the data thereby resulting in heterogeneous, or un-representative data different statistic was calculated to shed light on how to better measure this variable.

Orwin’s fail-safe n represents the additional number of studies that would be required for a data set to have enough statistical power to reach significance in
consideration of its current values. Orwin’s n was found to be 79.2 meaning that approximately 79 studies with the same results would yield enough power to have these values reach significance. Thus, there were no significant differences between groups and the fact that an additional 79 studies with similar results would be required to result in a significant difference favoring a large treatment dosage lends confidence to the present finding that despite its closeness, Qb shows no differences between groups. Results for the pre-selected moderator variables found are presented in Table 5.
Table 5

Summary of Pre-Selected Moderator Variables Analyzed in EMDR/ Alternative Treatment Comparison.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Effect Size (d)</th>
<th># of Effects</th>
<th>Bootstrap Confidence Interval*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trauma Type</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combat-Related</td>
<td>.35</td>
<td>31</td>
<td>.21-.48</td>
</tr>
<tr>
<td>Non Combat-Related</td>
<td>.50</td>
<td>37</td>
<td>.28-.69</td>
</tr>
<tr>
<td><strong>Dependent Variable Type</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Report</td>
<td>.41</td>
<td>46</td>
<td>.25-.56</td>
</tr>
<tr>
<td>Therapist-Administered</td>
<td>.47</td>
<td>15</td>
<td>.18-.71</td>
</tr>
<tr>
<td>Objective</td>
<td>.52</td>
<td>7</td>
<td>.18-.79</td>
</tr>
<tr>
<td><strong>Recency</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Newer</td>
<td>.39</td>
<td>57</td>
<td>.25-.52</td>
</tr>
<tr>
<td>Older</td>
<td>.66</td>
<td>11</td>
<td>.35-.96</td>
</tr>
<tr>
<td><strong>Sessions Administered</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large</td>
<td>.04</td>
<td>29</td>
<td>-.15-.23*</td>
</tr>
<tr>
<td>Medium</td>
<td>.85</td>
<td>20</td>
<td>.70-1.03</td>
</tr>
<tr>
<td>Small</td>
<td>.48</td>
<td>19</td>
<td>.33-.66</td>
</tr>
<tr>
<td><strong>Eye Movements</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMDR</td>
<td>.43</td>
<td>67</td>
<td>.31-.55</td>
</tr>
<tr>
<td>EMDR w/out Eye Movements</td>
<td>.35</td>
<td>21</td>
<td>.17-.55</td>
</tr>
</tbody>
</table>

**Note.**

Probability value: p<.05

*When bootstrap confidence interval includes zero in range, effect size is not considered significantly different from zero.
Supplemental Moderator Variables in EMDR/Alternative Treatment Comparison

A better comparison for treatment setting was obtained in the EMDR/Alternative treatment comparison than the EMDR/Control comparison because of the inclusion of studies using alternative treatments within an inpatient setting. Additionally, no EMDR/Alternative treatment primary studies included both inpatient and outpatients in the same sample allowing for a clearer comparison of participants. Setting effect sizes were again found to be homogeneous and there were no significant differences found between groups (Qb=.10, p=.95). Although the groups did not significantly differ from each other, the inpatient group showed a large range average effect size (d=.62). The outpatient and “unknown” groups each yielded a medium range effect size of .42. This finding is not surprising as the more severe symptoms of an inpatient population would likely be more sensitive to treatment effects measured by quantified variables than persons with more nominal symptoms. Additionally, it is reasonable to assume that most inpatient PTSD participants are part of a more chronic population and as such have been exposed to other treatment modalities and these treatments obviously did not ameliorate PTSD in that the participants were hospitalized at the time of the primary studies.

Journal type was analyzed and data were found to reach homogeneity. No differences were found on the treatment effects reported by various journal types (Qb=3.89, p=.27). Trends similar to the EMDR/Control comparison emerged with the largest effects being reported by unpublished studies (d=.75) and the second largest found in behavioral psychology publications (d=.60). Psychiatric journals showed small effects with an average of .07. Interestingly and unlike the previous comparison, anxiety/trauma-specific publications yielded a small sized average effect of .14. Perusal of the primary studies in this small latter category (as cell sizes were split between four
groups) showed that the 22 individual effect sizes were from two articles in the same special EMDR edition of The Journal of Anxiety Disorders. This journal edition attempted to represent studies from “rival” camps and this may have impacted the overall effect in this category. Conversely, the EMDR/Control comparison had greater variety of sources in this category.

The final descriptive supplemental moderator explored was related to research design and this was an analysis of the internal validity ratings for each study (criteria presented in Appendix D). Data for the groups were found to be homogeneous and there were no significant differences found between groups. This was true when internal validity was categorized three ways (high, medium, and low validity ratings) as well as when data were collapsed two ways as done in the EMDR/Control analysis. Trends were found to be similar to the EMDR/Control analysis with the higher ranked studies showing smaller effects (d=.07) and the lower ranked studies showing large effect sizes (d=1.19). Medium internal validity was found to have a medium effect size (d=.41). While the medium and low internal validity categories showed effects significantly different from zero, the high internal validity effects included zero in bootstrap confidence intervals (−.13-.29) indicating that effects were not significantly different from zero for studies with strong internal validity. The trend for smaller effect sizes commensurate with higher internal validity ratings is consistent with the EMDR/Control analysis and the hypothesis for this independent variable. Results for the supplemental moderator variables are presented in Table 6.
### Table 6
Summary of Supplemental Moderator Variables Analyzed in the EMDR/Alternative Treatment Comparison

<table>
<thead>
<tr>
<th>Variable</th>
<th>Effect Size (d)</th>
<th># of Effects</th>
<th>Bootstrap Confidence Interval* (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Setting</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outpatient</td>
<td>.42</td>
<td>57</td>
<td>.27-.57</td>
</tr>
<tr>
<td>Inpatient</td>
<td>.62</td>
<td>6</td>
<td>.21-1.47</td>
</tr>
<tr>
<td>Unknown</td>
<td>.42</td>
<td>5</td>
<td>.35-.49</td>
</tr>
<tr>
<td><strong>Journal Type</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety/Trauma</td>
<td>.14</td>
<td>22</td>
<td>-.11-.39*</td>
</tr>
<tr>
<td>Behavioral Psychology</td>
<td>.60</td>
<td>15</td>
<td>.46-.79</td>
</tr>
<tr>
<td>Psychiatry</td>
<td>.07</td>
<td>13</td>
<td>-.14-.29*</td>
</tr>
<tr>
<td>Unpublished</td>
<td></td>
<td>18</td>
<td>.56-.96</td>
</tr>
<tr>
<td><strong>Internal Validity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>1.19</td>
<td>7</td>
<td>.91-1.56</td>
</tr>
<tr>
<td>Medium</td>
<td>.41</td>
<td>48</td>
<td>.27-.54</td>
</tr>
<tr>
<td>High</td>
<td>.07</td>
<td>13</td>
<td>-.13-.29*</td>
</tr>
</tbody>
</table>

**Note.**
Probability value: p<.05

*When bootstrap confidence interval includes zero in range, effect size is not considered significantly different from zero.
CHAPTER V
CONCLUSIONS AND RECOMMENDATIONS

Summary

This study investigated the efficacy of Eye Movement Desensitization and Reprocessing therapy in treating Posttraumatic Stress Disorder. The method used was meta-analysis and EMDR was compared to both no treatment control group studies and studies that utilized alternative treatment modalities. Because it was reasonable to expect that comparison of EMDR with no treatment would yield larger effects than if comparing with a viable treatment, separate comparisons were made for each collection of studies. Summaries with interpretations for each of these comparison analyses will be presented for each hypothesis. These summaries will present the integrated findings of both the EMDR/Control group and EMDR/Alternative treatment comparisons together in relation to each sub-hypothesis. Following, the limitations of the present investigation will be addressed and recommendations for future studies explored.

Overall Efficacy of EMDR in the Treatment of PTSD

When EMDR was compared with the no treatment control condition, the overall average effect size of .50 was found and this is considered in the medium range of effects. This analysis was based on 51 individual effect sizes that Q testing found to accurately represent the distribution of effect sizes. This can be interpreted to mean that
the average effect size of .50 represents the data in the primary studies adequately and therefore variance associated with the collection of effects did not confound the data. This is a notable effect size and is interpreted to mean that EMDR treatment showed .50 more effect on average in treating PTSD diagnosed participants than did no treatment. Considering a normal distribution, an effect size of 1.0 is interpreted as the average person in the treatment group exceeding 84% of the people in the control group on the dependent variable used to measure outcome (Vargas and Camilli, 1997). Therefore, a .50 effect indicates that the average person in receiving EMDR treatment exceeded 67% of the participants receiving no treatment.

Because data were found to be significant and there was also a significant difference found between groups, the hypothesis that EMDR would be effective in the treatment of PTSD was supported and the magnitude of effect was found to be medium at .50. Data for EMDR/Alternative Treatment comparison were similar and these results with an average effect size of .43 showed medium range effects for EMDR, even when compared to other treatments commonly used in the treatment of PTSD. This data was homogeneous and between group differences were significant which suggests that these data estimate true effects in the population. This is a very important finding because it shows that EMDR resulted in superior quantitative differences when compared with other treatments such as relaxation therapy, exposure therapy, and the general category of other cognitive-behavioral treatment modalities. This was consistent with the findings of Van Etten and Taylor (1998) which found EMDR to yield stronger treatment effects than most alternative therapies. These two analyses which tested a total of 119 effect sizes indicated conclusively that EMDR is efficacious in the treatment of PTSD. The fact that significant heterogeneity, or variance, did not exist in either comparison bolsters
confidence in these findings. Figure 5 in chapter four illustrates the grand means of the EMDR/Control group and EMDR/Alternative treatment comparisons.

**Moderator Variables**

Separate analyses were conducted in order to explain sources of variation and this included the apriori moderator variable sub-hypotheses as well as analyses of research design characteristics and other treatment-related variables. Since the overall aforementioned efficacy comparisons yielded homogeneous data, examination of moderating influences was technically not necessary because the homogeneity indicated that the treatment conditions accounted for group differences. However, moderator variable investigation was useful for clinical significance reasons. By sub-dividing the collections of effect sizes into different groups, it was possible to see which groups fared best with EMDR and in which groups there were not significant differences when compared with controls. This allowed for interpretations on different clinical populations, such as type of trauma and those receiving a small number of sessions vs. those who received a larger number of treatment sessions. Of course, generalizations are only possible if the variance is low enough for the collection of studies to achieve homogeneity and all testing did reach homogeneity in the final analysis.

The first sub-hypothesis of interest was the prediction that there would be differences in the magnitude of effect favoring survivors of non-combat related trauma to survivors of combat related trauma. It was hypothesized that there would be differences favoring the non-combat group since EMDR received empirical support for this based on two studies included on the 1998 list of empirically validated therapies (Chambless, et al., 1998). Interest in this variable was borne out of the APA Task Force’s findings
(Chambless, et al., 1998) that EMDR was "probably efficacious in the treatment of civilian PTSD" (p. 9), but without empirical validation noted for non-civilian PTSD. The fact that non-civilian sample studies were available at that time made this a particularly curious distinction.

Results of the present investigation showed equivocal support for the hypothesis: EMDR showed a significant effect in treating non combat-related PTSD compared to combat related PTSD in the EMDR/Control comparison, but no significant difference between these groups was found when alternative treatments were used in the comparison. The trends of both the significant and non-significant difference findings, however, show greater effects of EMDR in treating non-combat related PTSD than combat related PTSD, consistent with the findings of Chambless et al. (1998) that EMDR shows empirical validation for civilian PTSD. Thus, the hypothesis that EMDR would be more effective in treating non-combat PTSD was supported when compared to a no treatment control, but not significantly so among the standardized difference scores of EMDR compared to alternative treatments.

These results were reasonable and consistent with trends found in the literature (see chapter 2). As discussed previously, it is reasonable to expect larger effects when a treatment was compared to no treatment than when the same treatment was compared to other treatments. This was found to be the case for the present moderator analysis. Similarly, it is reasonable to find smaller effects when comparative treatments were administered to a treatment resistant population, as combat veterans are frequently considered. McFarlane and Girolamo (in van der Kolk, et al., 1996) note the differential treatment expectation for chronic PTSD compared with acute or "uncomplicated PTSD" (p. 138). Carlson et al. (1996) echo this concern stating that, "The nature and severity of
the treated disorder must be taken into account. The chronic and severe nature of PTSD in combat veterans renders the disorder difficult to treat under any conditions” (p. 113).

The fact that EMDR showed larger effects consistently in the two comparisons, even though in the EMDR/Alternative treatment comparison the differences were not considered statistically different, lends support to the hypothesis.

The second sub-hypothesis investigated the question: Are there differences in the effect sizes found on verbal report measures compared to physiological measures? It was hypothesized that there would be differences and that verbal report measures would show larger effect sizes than physiological dependent variables in measuring PTSD symptomatology. This question was borne out of the often cited criticism that EMDR studies (particularly the earlier EMDR studies) suffered from use of dependent variables with inherent demand characteristics. Equivocal support was found for this hypothesis as differences favoring verbal measures were found in the EMDR/Control comparison, but no significant differences were found in the EMDR/Alternative treatment comparison. The greatest equivocation for any conclusions for the former comparison is the fact that there was an inadequate number of dependent objective measures to draw any sound conclusions. The fact that the bootstrap confidence interval was large and included zero corroborates the need for caution against generalizations.

Although large effects for verbal indices were found on the first comparison, medium effects for verbal dependent measures were found when compared to alternative treatments. Further, larger effects on physiological measures were found for the alternative treatments than for EMDR in the second comparison. This was surprising, particularly in light of the fact that many alternative treatments utilized relaxation models of treatment. However, several explanations may be accountable for this finding.
Ideally, this finding reflected a more decreased state of hypervigilance as a result of receiving EMDR treatment. This explanation is consistent with the basic tenet of EMDR therapy which underlines the importance of treating psychotherapy clients on a mind-body level instead of just treating the mind. Alternatively is the fact that physiological measures are not PTSD-specific, but instead measure the physiological effects associated with PTSD. For instance, increased heart rate is associated with the DSM-IV symptom cluster of hypervigilance but increased heart rate is not in itself a measure of PTSD whereas a verbal measure such as the Clinician-Administered PTSD Scale more directly targets PTSD-specific symptoms.

The third sub-hypothesis investigated was the question: Are there differences in the magnitude of effect between earlier and more recent studies? It was hypothesized that earlier studies would show larger treatment effects. This hypothesis was borne out of the stunning effects reported among some earlier studies and the backlash against the efficacy that ensued in the later psychological nomenclature. For instance, McCann (1992) and Spector and Huthwaite (1993) reported remarkable successes, even within a single session of EMDR. Later research reviews, however, such as by Acierno et al. (1994) and Cahill et al. (1999) harshly criticized the soundness of much of EMDR’s early claims. Homogeneous results in both comparisons found that there were no significant differences between older and newer studies. Thus the hypothesis was not supported which corroborates the efficacy reported for EMDR in early studies. Earlier studies did show larger effects than later studies for both groups, but even the later studies showed that treatment effects were significantly different from zero. This supports the efficacy of EMDR in treating PTSD across the decade of 1990-2000 without significant bias for time of study.
The fourth sub-hypothesis investigated was the question: Does the bilateral stimulation component of EMDR therapy contribute significantly to treatment effects? It was hypothesized that the eye movements would not contribute significantly to the efficacy of EMDR. There were several reasons behind this hypothesis. Many of the EMDR studies, including those conducted by researchers who showed enthusiasm for EMDR, reported no significant differences when the eye movement component was removed from the treatment in dismantling designs. Secondly, there have been no conclusive explanations found as to what the mechanisms for change are in EMDR therapy. The third and most important reason for this hypothesis related to this author’s belief in the integrity of the overall treatment protocol specified by the manualized treatment of EMDR. Independent of the bilateral stimulation, EMDR consists of an active processing by the participant on both a cognitive and affective level. The trauma survivor is re-exposed to the trauma in manageable doses in a safe environment. The participant processes at his/her own rate of emotional processing and develops replacement beliefs and emotions associated with the traumatic experience. In all respects, this is considered good psychological therapy and to this author, seems to not replace widely used exposure techniques, but supplements these techniques.

Although results of the present investigation supported the hypothesis that bilateral stimulation (in the present case, eye movements) does not significantly contribute to the overall efficacy of EMDR, research design limitations preclude acceptance of this hypothesis. The meta-analysis found that there was no significant difference between the group administered EMDR and the group whose therapy attempted to remove the bilateral stimulation component so in essence, these groups could be considered part of one larger group. Put another way, the lack of a significant
difference suggested at face value that the eye movements were not a significant
differential component of the larger treatment. This was an interesting finding and is
consistent with the findings reported by the primary authors, such as Pitman et al. (1996);
Renfrey and Spates (1994); and Montgomery and Allyon (1994).

This moderating variable analysis, however, had many limitations which preclude
acceptance of the hypothesis and generalizations about the importance of the eye
movements. From a research design point-of view, the idea of measuring a manualized
therapy without one specific component has inherent danger. One of the problems was
that while the person was not following therapist-induced eye movement sets, there were
other things happening for each participant and these are not necessarily controlled within
the experimental setting. This was particularly problematic in the present investigation as
several researchers used alternatives of the eye movements such as a light tracking bar,
and visual focus on an automated light, and focus on a stationary light. The Renfrey and
Spates (1994) study also included having participants rhythmically tap fingers and it was
not detailed whether or not this was in an alternating movement which in turn could
easily meet the criteria of bilateral stimulation. Given EMDR’s recent attention toward
alternative forms of bilateral stimulation, it can not even be said that these experimental
groups (that is, EMDR without the eye movements) were conducting therapy different
from EMDR. Shapiro (1999) pointed this out in her review of the eye movement
component research literature, “Some component analyses have used as their placebo
conditions alternative stimuli that have actually been used by successful practitioners of
EMDR for many years as effective substitutes for the eye movements” (p. 60).

It is argued here, however, that regardless of whether this component can ever
truly be measured (and this may only happen when medical technology can make it
possible), the findings are still good for EMDR. The EMDR treatment protocol is comprehensive yet specific to the trauma; it is solution focused; and it is described by many as more tolerable than other forms of trauma therapy. These are components of a good therapy administration and considering the medium range effect sizes found in the present study, EMDR is clearly alleviating much of the symptomatology of PTSD survivors sampled here. This in itself is clinically significant. As stated previously in forming this hypothesis, the fact that EMDR still showed impressive treatment effects without the eye movement component speaks to its importance as a treatment modality for PTSD.

The final sub-hypothesis investigated was the question: Are there differences in effect size based on the number of EMDR sessions administered? It was hypothesized that there would be no difference in the effect sizes associated with a small, medium, and large range of sessions administered. This hypothesis belies conventional wisdom that the more treatment given, the more effects will be gained. However, the present hypothesis was based specifically on the EMDR protocol of treatment which regulates the continuance of therapy commensurate with the progress reported by the client during the treatment process. As described in chapter 1, the sets of eye movements and related cognitive and affective processing continue until desired results are achieved. This was the reason that many of the studies illustrated in Table 2 reported an average number of sessions administered rather than a fixed number of administrations. In fact, it is valid to say that studies that did not operate in this client-driven manner were not true to the EMDR manualized protocol (this will be further addressed in the limitations section). The obvious exception to this would be studies that provided an insufficient number of sessions such as the single administration studies. However, based on the overall
efficacy reported in the literature, the EMDR group was still hypothesized to fare better than both no-treatment or alternative treatments offered in the same dosages. Therefore, it was conceived that the dosage administered would not significantly bear on effects in comparison to other treatments. This hypothesis was supported, as both comparison groups showed homogeneous results with no significant between group differences based on range of treatment sessions. This is an important finding because it suggests that the medium range effects favoring EMDR to both no treatment controls and alternative treatments are consistent for both very short and longer treatment interventions.

Supplemental Moderator Variables

Outside of the overall efficacy question and the five sub-hypotheses tested, several descriptive independent variables were analyzed to determine if there was any moderating influence on the efficacy of EMDR.

Setting of treatment, that is whether conducted in an inpatient or outpatient basis, must be considered a limited finding. Although both data comparisons showed homogeneity, there were too many primary studies that did not report this information adequately (coded as "unknown") and others that mixed an inpatient and outpatient sample in the same study. The result of this was that the EMDR/Control comparison did not include any inpatient effects and the EMDR/Alternative treatment comparison did not have a "mixed" category of effect sizes. This was unfortunate because this categorization would clearly have clinical importance if valid generalizations could be generated. Nonetheless, the data that were tested were homogeneous reflecting that estimates among the population could be made with confidence. Consistent effects were shown in the medium range for outpatient treatment and the group categorized as "unknown."
mixed group of studies analyzed in the EMDR/Control comparison showed negative
effects for the EMDR group, however, this was based on a small number of effects (n=7)
and none available for the EMDR/Alternative treatment comparison). Inpatient treatment
also consisted of a small number of effect sizes in the medium range of effect (n=6 in the
EMDR/Alternative treatment comparison and none in the EMDR/Control comparison).
This analysis lends equivocal support for the efficacy of EMDR in treating different
level-of-service populations. The fact that no group showed significant differences and
all groups (except the dubious category of “mixed”) showed treatment effects consistent
with the overall efficacy findings of EMDR support the efficacy of EMDR across
treatment settings. Support must be considered preliminary due to the small number of
studies analyzed for this variable. It is recommended that future primary study research
maintain more pure experimental settings so that a more sensitive estimate of effect as
well as external validity is possible.

Journal type was examined to see if there were any moderating influences
associated with the type of journal the primary studies originated from. This was
conceived of during the investigation when certain journals, such as those from
behavioral psychology, seemed to more commonly include EMDR studies that yielded
positive results. Although there was a trend for larger effects in the behavioral categories
of both comparisons, this group did not significantly differ from any of the other groups.
This indicated that EMDR did not show significantly larger effects dependent on the
publication source. In fact, there was no significant difference found between any of the
groups in the EMDR/Alternative treatment comparison and only one between group
difference found on the EMDR/Control group comparison. This difference occurred in
the post hoc comparison of unpublished studies with clinical psychology publications,
which showed significantly larger effects for the former group. Examination of the data found that the negative effect for EMDR coming from objective measures covaried with the clinical psychology variable categorized under journal type. Therefore, there is potential for interaction un-analyzable by the present investigation due to the small number of objective measures. Alternatively, it is possible that the clinical psychology variable, which included prestigious publications like the Journal of Clinical and Consulting Psychology, were associated with more rigorously controlled studies. Such caliber of investigations, in turn, are associated with smaller effect sizes.

The final descriptive moderator investigated related to research design quality. In order to assess the so-called garbage in-garbage out problem discussed in chapter 3, studies were rated for internal validity (see Appendix D). The reasoning behind this examination was the conventional wisdom that speculates that stringently controlled, high integrity studies typically result in smaller effects found for experimental groups. The reason for this is that the more rigorously an experiment is controlled, the less variance is associated with measurement and the more true effects are. Lipsey (1990) notes that poorly controlled studies show variance, or heterogeneity, such as through outlying scores, that often result in less sensitive, less representative scores. This was found to be the case, however, only significantly so in the EMDR/Control group comparison. This comparison yielded smaller effects for more internally valid studies than for those found in the medium range of internal validity (there were no studies considered to be in the low range of validity in this comparison). Both comparisons showed that in the high range of internal validity, treatment effects were in fact not significantly different from zero. In the EMDR/Alternative treatment comparison, however, there were no significant differences accounted for by internal validity ratings.
The trend, however, was maintained with better quality studies yielding smaller effects and studies with lower internal validity ratings showing smaller effects.

The internal validity analysis was consistent with the statistical generalization discussed in the literature that speaks to the tendency for more liberally controlled studies to show greater variance. In the present study, this was only true for the EMDR/Control group comparison. The fact that this was true only when EMDR was compared with a no treatment control may not be coincidental. This finding was consistent with the argument made by many that testing an experimental group with a no treatment control is in itself a less robust comparison than comparing one treatment modality with another. As stated previously, it is reasonable to expect larger differences for a treatment/no treatment comparison than when two bona fide treatments are compared.

Areas of internal validity weakness found among the primary studies included a lack of physiological dependent variables in many studies, small sample size, and consequently poor statistical power to detect effects. Of course, the lower the power the less sensitive a study’s ability to find effects and only 1/21 studies even included power details in its design. The problem of power in social science research, therefore, was not avoided in the primary studies included in the present investigation.

Two other frequent areas of weakness noted in the primary studies concerned blindness and treatment fidelity. In many of the studies, the implementing administrator of dependent variables was not reported to be blind. Related, about half of the studies did not report methods implemented to quantify or ensure treatment fidelity. Of course, it is possible that these conditions were controlled for, but without adequate reporting the internal validity rankings would still be low.
There was only one meta-analysis found in the literature that included EMDR in the treatment categories. Van Etten and Taylor (1998) included EMDR as one of many treatments investigating the treatment of PTSD, and this investigation included pharmacological as well as psychotherapeutic interventions (see chapter 2 for a detailed review of this study). Results of the present investigation were consistent with those reported by Van Etten and Taylor (1998), but smaller. In terms of overall PTSD, Van Etten and Taylor (1998) reported a grand mean effect size of 1.24 on self-report measures and .69 on observer-rated measures. The latter category, however, did include zero in its confidence interval. Results of the two meta-analyses are not directly comparable as Van Etten and Taylor (1998) reported means instead of d and did not report separate analyses for no treatment controls and alternative treatment controls which, obviously, has potential to skew data as the data are based on separate distributions.

Limitations and Recommendations

The present study was limited by both the nature of meta-analysis at large as well as problems unique to the present investigation. Despite holding prominence as the most studied psychological therapy in treating Posttraumatic Stress Disorder (Shapiro, 1999), stringent inclusion criteria kept the number of studies relatively small. Part of the reason for this was a preponderance of published EMDR studies that did not include a PTSD-specific sample but instead featured participants who experienced trauma in general. The present investigation's intention of generalizing to a PTSD population therefore precluded use of studies whose criteria did not include a PTSD diagnosis. Indeed it was surprising that even in the prestigious APA Task Force (Chambless et al., 1998) selection of studies earning inclusion on the list of "Probably efficacious in the treatment of
civilian PTSD” (p. 9), one of the two studies included did not require a PTSD-specific sample as did the present analysis.

Additionally, there were a number of primary studies that were excluded for other reasons such as research design characteristics or publication in a foreign language. Such limitations further narrowed the pool of resources. A greater number of studies would have been useful in providing for a more detailed investigation as well as greater statistical power for detecting treatment effects. This particularly limited testing for moderator variables since each time an independent variable is split into categories, the cell size and consequent statistical power is reduced by the number of categories used.

Other inadequacies of the participating primary studies included limited reporting of critical research information, most notably standard deviations and posttest sample sizes. For instance, Lazgove, Triffleman, Kite, McGlashan, and Rounsaville (1998) conducted a frequently cited study using a PTSD-specific population but did not report the statistical information necessary for meta-analytic study. The data was no longer available two years after its publication from the primary author (personal communication, November 27, 2000). Although much data used in the present study was ultimately retrieved through personal communications with the study authors, important data was necessarily omitted that could have potential value toward both the overall efficacy question and the investigation of moderator variables. It is recommended that future studies report the means, standard deviations, and all cell size n’s to allow for inclusion in meta-analytic studies. This would allow for quantitative reviews replacing, or at least supplementing, the preponderance of narrative/anecdotal reviews.

Limitations in the description of participant information such as type of trauma would have allowed for a more robust analysis as well. The limitation of primary studies
in including specific trauma types was of particular import as this variable includes clinically important information. For example, the only study in an exhaustive literature search that specifically included PTSD survivors of sexual assault was the Rothbaum (1997) study, despite the estimate that rape victims constitute the largest number of PTSD survivors in the country (p. 318). While others such as Edmond, Rubin, and Wambach (1999) also investigated the efficacy of EMDR with sexually traumatized participants, the study did not feature a PTSD diagnosis and therefore could not be generalized to a PTSD population as is the present meta-analysis. While some studies included descriptions of the trauma of origin, such as Vaughan et. al (1994), data were not partitioned into different outcome groups which necessitated the more generic generalization of PTSD survivors. A recommendation for future EMDR PTSD studies is to limit studies to the trauma of origin so that the effects of EMDR could be compared between specific PTSD groups such as survivors of natural disasters, sexual abuse, or victims of violent crimes to name a few.

The primary studies in the present investigation suffered from many of the common flaws in research designs of social science experiments. Of paramount importance was the small sample size employed, particularly when samples were further divided into treatment group categories. While meta-analysis has the benefit of aggregating multiple samples within single statistical tests, it does not change the effect sizes found within any individual study. Consequently, if a study has low power and therefore poor sensitivity in detecting treatment effects, this will only contribute the effect sizes found given the conditions of power.

Treatment fidelity, an ongoing problem in social science research, was a concern of the present investigation. While studies reported various attempts to ensure treatment
integrity for EMDR, that was often not the case for the alternative treatments described. Thus even if the EMDR condition maintained treatment fidelity, if no measures were reported to ensure fidelity of the other treatment, for example applied muscle relaxation training, then there is not necessarily a fair comparison between treatment effects.

Furthermore, several studies (such as Pitman et. al, 1996 and Renfrey & Spates, 1994) used trauma scripts created in the initial session with participants which is somewhat different than the targeted images discussed with the client at interview specified in the first phase of the EMDR protocol. It must be acknowledged, however, that the manualized nature of EMDR coupled with mandatory reported adherence to manualized procedures documented in the present study’s inclusion criteria, promoted treatment fidelity as much as possible.

The nature of dependent variables was a limitation in the present investigation. Unlike the absolute scores available to other sciences, the overlap of measurement was inescapable in the present study. Even when dependent measure categories were collapsed due to heterogeneity confounds, important information was lost. Further, there were too few objective effect sizes found in the literature. While many studies criticized the use of non-standardized self-report measures prevalent in EMDR studies, few actually counteracted this by using physiological dependent variables. The small n of objective measures limited conclusions and generalizations about physiological treatment effects in the present investigation.

In addition to the limitations imposed by the characteristics of primary studies, the present investigation suffered from inherent limitations as well. First, and like many of the primary studies, small sample size was a limitation. Although inclusion of 21 primary studies was typical for meta-analyses in the social sciences, this relatively small
sample size did not afford high power, particularly when partitioned for examination of moderator variables. Meta-analysis was originally developed for use in the so-called "hard sciences" such as medicine and agriculture and it is common in these fields to find hundreds of effect sizes available for analysis. As the dose response curve analysis of the EMDR/Alternative treatment comparison showed through calculation of Orwin's fail-safe n, multiple additional studies bearing identical findings can result in statistically significant differences. The question about whether or not this would be clinically meaningful is a different matter.

In a personal communication with the author, Shapiro (November, 2000) suggested a limitation that the present study was unable to control for due to both the number of studies available and the descriptive limitations of those studies. This is best explained as the potential for an interaction confound between number of sessions administered and nature of trauma, specifically those participants considered "multiply traumatized" and those whose PTSD is based upon a single traumatic incident. In other words, the present study did not investigate covariance associated with the independent moderator variables. For instance, Shapiro (personal communication, November, 1999) noted that many primary studies with severe, or multiply traumatized samples utilized a small number of treatment sessions which would likely not result in large treatment effects. Due to limitations in reporting found in primary studies described above and n cell sizes too small to provide for statistically feasible partitioning (as well as the tendency for studies to aggregate the trauma groups together), this was not possible. While the present study attempted to estimate effects of independent variables such as combat vs. non-combat related PTSD and compare effects for the range of sessions, it was not possible to test these independent variables together. Such an analysis would
provide clinically relevant information and as the number of EMDR PTSD studies increases, this type of investigation is recommended for future meta-analytic review.

The present investigation did not examine treatment effects for follow-up data and follow-up data would have provided rich and clinically meaningful information. Foa and Rothbaum (in van der Kolk, et al., 1990) in fact cautioned against generalizing any conclusions from PTSD studies that do not provide follow-up data as symptom assessment beyond immediate post-testing is necessary to accurately gauge the long-term effects of treatment. Although there was adequate data provided in the primary studies to assess follow-up results, the inclusion of such information was beyond the scope of the present investigation. Further, primary studies were not consistent in the intervals used to provide follow-up assessments. Additional confounds beyond the sustenance of the present study involve the attrition rates, lack of follow-up data across studies, and use of uncontrolled assessment settings (such as mail-back questionnaires). It is recommended that future meta-analyses investigate follow-up data when sufficient dependent variables are available to assist in the control of these potential confounds. Along these lines, there were no studies found that tested long-term effects of EMDR such as after more than one year. Now that EMDR has a decade of history in treating PTSD, studies analyzing effects would be extremely important in evaluating the efficacy of EMDR.

Another limitation of the present study that was not able to be ameliorated despite attempts to do so was the measurement of some “clinically significant” outcomes as opposed to statistically significant dependent variables (see Kazdin, 1998, chapter 25). This class of dependent variables refers to the important but less concrete effects of clinical change that result from an intervention. An example relevant to the present paper was the inability to include the clinically significant outcome of the number of
participants who no longer met diagnostic criteria for PTSD following completion of a primary study. Clearly, this is clinically meaningful. However, because this data is not amenable to quantitative analysis (i.e. lack of standard deviations) it was not included in the present investigation. When possible, clinically significant measures amenable to meta-analysis such as client documented number of intrusive thoughts per week were included and coded under the dependent measure category of “other.” However, a more simplistic meta-analytic design such as the vote-counting method would be useful in future research to better estimate this variety of data.

Clinical Implications

The present study demonstrated the efficacy of EMDR in the treatment of PTSD. Overall, treatment effects were in the moderate range and for a person suffering the harmful effects of PTSD, such effect size can translate into meaningful improvement. This was generally consistent with the meta-analytic findings of Van Etten and Taylor (1998) which found EMDR to be among the most effective psychological therapies and reported large effect sizes relative to control conditions. As such, this provides support for mental health practitioners to administer EMDR to PTSD survivors. While there were mixed findings related to the sub-groups studied, trends suggested a positive impact from EMDR even when statistically significant differences between groups were not found. Importantly, in almost every variable tested, positive treatment effects were detected for the EMDR group. The only exceptions to this was the near-zero effect of - .04 for the combat related trauma group in the EMDR/Control comparison and the physiological measures group in the same comparison which was based on a mere two studies. The fact that almost all treatment effects were positive for EMDR suggests that
participants were in better condition after the treatment than they were prior to getting EMDR. Clinically, this provides support for the use of EMDR with PTSD diagnosed clients.

The present study indicated that EMDR was particularly efficacious with non-combat related traumas. This provides support for EMDR as a treatment modality with this population and is consistent with the conclusions drawn by Chambless et al. (1998) in citing EMDR therapy as "probably efficacious in the treatment of civilian PTSD" (p. 9). However, moderate effects were also found for combat survivors in at least one of the data sets, which provides partial support for the use of EMDR with a combat-related population. Further studies in this area are needed to better assess the clinical utility of EMDR with combat-related PTSD participants since findings were mixed. However, the quantitative trends found in the present study suggest that EMDR was generally more effective than other treatments commonly utilized with this chronic population.

In terms of clinical research, there was no statistically significant evidence to suggest a positive bias from early EMDR studies. While it was true that effects found in early studies were consistently larger, these effects were not significantly different from more recent EMDR studies. This implies that in terms of time of study, the earlier studies should not be treated differently than recent studies as part of the research base for EMDR in the treatment of PTSD. There is frequent criticism about the early claims made for EMDR and this finding argues against that criticism since statistically there were no differences between the early and later findings for EMDR.

Related, there were no consistent differences found among effect sizes found on the dependent variables used in EMDR studies. This too was consistent with the findings of Van Etten and Taylor (1998), however, only self-report and therapist-administered
indices were used in that investigation. Clinical research for EMDR has been criticized for reliance on verbal measures and a lack of physiological measures. The present study did find a lack of objective measures among studies when compared with the volume of self-report and therapist-administered verbal indices. However, the present study also found that physiological indices showed a tendency toward extreme values, both positive and negative. While note was made of the limitation of this finding due to small sample size, the lack of consistency calls into question the utility of physiological dependent variables. Further research into the area of physiological measurement of PTSD is needed to determine the clinical utility of this variety of outcome before clinical research can be judged on its usage.

It was very interesting, and supportive of the hypothesis, that there was no main effect found for the number of sessions administered. To properly interpret this requires the understanding that like all analyses in the investigation, this sub-analysis compared EMDR with another group (either control or alternative treatment). This finding means that the treatment effects found for EMDR were generally consistent regardless of whether there were a large or small number of sessions administered. Clinically, this suggests that the implementation of EMDR can be supported in both short and longer-term treatment plans for individuals diagnosed with PTSD. Obviously and as stated previously, a limitation of this finding is that potential interactive effects were not studied in the present investigation or in any of the literature found in the literature review. An example would include the potential interactive effect for multiply traumatized veterans and small number of sessions. However, in general, the present study provided support for administration of both a small or a larger number of EMDR sessions for the treatment of PTSD.
It is important to appreciate the clinical implications of manualized treatment in general and of EMDR in particular. Every study employed in the present investigation included a minimal of Level I EMDR training. Therefore, it is necessary to limit generalizations to EMDR delivered by clinicians trained at a minimum of Level I and imperative for clinicians to receive this training prior to attempting EMDR therapy on clients. Since only studies featuring properly trained clinicians have been tested, the moderate range treatment effect sizes found are restricted to the implementation of EMDR by EMDR Institute recognized training. Clinically, this means that any clinician considering EMDR for treating a PTSD diagnosed client must receive proper training before administering the treatment. Likewise, EMDR must be implemented according to proper protocol in order to have a realistic expectation that treatment effects similar to those found in the present investigation will be achieved.

Future Directions

In addition to suggestions for future research related suggested thus far, other areas for further investigation prompted by the present investigation include the following:

1. Are there differences in effect size related to the efficacy of EMDR in the treatment of specific non-combat related traumas? EMDR was shown to have larger effects with non-combat related PTSD than for combat related PTSD in the present study. This spawns the question of whether or not categories subsumed within the non-combat population show important clinical differences as a result of EMDR therapy.

2. What, if any, are the physiological underpinnings of EMDR therapy? Although results were not directionally consistent, physiological indices of change were most
variant of the three categories of dependent variables, suggesting that to some extent, physiological change was measured as a result of treatment. Future research would benefit the study of EMDR by utilizing modern technological assessment of therapeutic change, such as Positron Emission Scans (PET) or neuroimaging, to more accurately assess neurological change and explain the causality of change. The Levin et al. (1999) study attempted this, but published only a single case instead of a group study of the total participants. Additionally, neurologically based investigations would contribute more conclusively to current speculations regarding neurological explanations for EMDR.

3. Is there a better way to more accurately measure PTSD? A consistent theme throughout the literature search for the present paper was criticism of dependent measures used to assess PTSD symptomatology in EMDR research. PTSD is a complex and multi-symptom disorder and despite some useful indices, there is a lack of outcome measures for the more complex symptoms of PTSD such as the ability to modulate affect. Even more concretely, some of the tangible symptoms specified in the DSM-IV (1994) like exaggerated startle response or difficulty sleeping were not measured by any of the primary studies used in the present analysis. Development of more precise PTSD-specific assessments, specifically measures that could avoid demand characteristics, would assist efficacy research in the area of PTSD and also spare clients the burden of excessive measures (as needed at present) for thorough research.

4. Are there differences in the magnitude of effect for different forms of bilateral stimulation utilized in variants of manualized EMDR? All of the primary studies employed in the present investigation used eye movements as the primary mode of
bilateral stimulation, however, Shapiro (1999) noted that there are alternative forms of bilateral stimulation being used in EMDR therapy. Research investigating alternative methods of bilateral stimulation such as finger taps or alternating sounds would be useful in evaluating the effects of a variety of bilateral stimulation.

5. Recommendations for future research include cross-cultural studies of EMDR with PTSD as well as multicultural studies that include culturally diverse samples. Related, it would be interesting and clinically relevant to ascertain whether or not the gender or culture of the implementing therapist and study participants has any association with the magnitude of effect. In order for such findings to be made, primary research must be more careful to both thoroughly describe participant and therapist characteristics as well as design studies to specifically measure these variables.

6. A final category of future variables for recommended study concerns the potential for interaction effects among various classes of independent variables both included and not included in the present study. Previously mentioned was the potential for interaction between level of traumatic experience (i.e. single vs. multiply traumatized individuals) and level of dosage (i.e. small vs. large number of treatment sessions). There are, however, numerous potentially interactive effects possible that were beyond the limits of the present investigation. Some examples recommended for future study include the independent variables of type of dependent measures used and length of treatment; matching for gender or other descriptive variables between therapist and participant; type of setting and dosage of treatment; and extent of training and experience of implementing therapist with intensity level of PTSD among others.
7. One important potential interaction virtually ignored by studies is the potential for interaction between pharmacological agents and EMDR treatment. Most studies did describe whether or not psychotropic medication was part of inclusion criteria, but none found quantified this variable in any way. Solomon, Gerrity, and Muff (1992) completed an empirical review of pharmacological and psychological therapies for PTSD, but did not analyze the interaction between such interventions. There would be much clinical value if future studies could assess the effects of combining different commonly used medications in the treatment of PTSD in tandem with EMDR therapy.
REFERENCES


EMDR Institute (1999). About EMDR. Newsletter of the EMDR Institute (available from EMDR Institute, PO Box 51010, Pacific Grove, CA 93950).


APPENDIX A

List of Articles Used in the Meta-Analysis


restructuring versus waiting list in the treatment of Posttraumatic Stress Disorder.

Unpublished raw data.


APPENDIX B

Study Inclusion Criteria
I. Sample

A. PTSD diagnosis

B. Sample size minimum of four participants per group

II. Research Design

A. Group comparison research

B. Use of group means (or other descriptive statistics capable of being transformed into standardized difference scores) as data for analysis

C. Minimum internal validity of .5 on Internal Validity Ratings

III. Other

A. Utilization of EMDR manualized treatment procedures

B. Inclusion of all statistical data necessary for meta-analysis (means, standard deviations, and group n's)
APPENDIX C

Coding Sheet
Study ID#

Comparison #:

I. Identification

Author(s):

Title:

Publication:

Year: Volume Pages

II. Background of Study

1. Publishing status: (1) Published (2) Unpublished (this category includes dissertations in UMI not published elsewhere)

2. Publication type:

   (1) Journal (2) Book (3) Dissertation (4) Conference/Convention Presentation (5) Other

3. Study Setting:

   (1) Hospital (inpatient) (2) Private Practice (3) Neither hospital nor private practice or unspecified.

4. Location of Study: W. USA (defined as follows: states west of North Dakota through Texas); Mid USA (states east of and including North Dakota through Texas through Michigan and Ohio south through and including Alabama); and E. USA (includes all states east of Ohio). All other countries list separately:
(1) West USA (2) East USA (3) Other USA (4) Outside of USA

5. Recency of Study:
List year of study.

III. Participant Information

1. PTSD Diagnostic Descriptor:
   (1) Combat-related (2) Non-combat related (or mixed/unspecified type of trauma)

2. Chronicity:
   (1) Single event trauma (2) Multiply traumatized (3) Mixed or unspecified chronicity of trauma

IV. Research Design.

1. Experimental Design:
   (1) Single group pre/post-test design (2) Two group pre/post-test design (3) Two group post-test design (4) Three group design (with or without pretest) (5) More than three groups in design

2. Follow-up Data:
   (1) No follow-up data beyond post-test (2) One set of follow-up data (3) Two sets of follow-up data (4) More than two sets of follow-up data
APPENDIX D

Internal Validity Ratings
Internal Validity

<table>
<thead>
<tr>
<th>Condition</th>
<th>Point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Participant Inclusion/Exclusion Criteria</td>
<td></td>
</tr>
<tr>
<td>A. PTSD diagnosis</td>
<td></td>
</tr>
<tr>
<td>No description of diagnosis derivation or use of previous diagnosis</td>
<td>0</td>
</tr>
<tr>
<td>Therapist clinical assessment</td>
<td>1</td>
</tr>
<tr>
<td>Formal measures to diagnose PTSD</td>
<td>2</td>
</tr>
<tr>
<td>B. Exclusion criteria</td>
<td></td>
</tr>
<tr>
<td>No description of participant exclusion criteria</td>
<td>0</td>
</tr>
<tr>
<td>Exclusion criteria based on pre-existing psychosis, personality disorder, substance abuse, or other comorbid psychological condition</td>
<td>1</td>
</tr>
<tr>
<td>Exclusion criteria includes screening for secondary gains</td>
<td>2</td>
</tr>
<tr>
<td>2. Group Status</td>
<td></td>
</tr>
<tr>
<td>A. Assignment</td>
<td></td>
</tr>
<tr>
<td>No description of assignment procedures</td>
<td>0</td>
</tr>
<tr>
<td>No random assignment</td>
<td>0</td>
</tr>
<tr>
<td>Random assignment</td>
<td>1</td>
</tr>
<tr>
<td>Stratified random assignment</td>
<td>2</td>
</tr>
<tr>
<td>Matched assignment</td>
<td>2</td>
</tr>
<tr>
<td>B. Pre-experimental group status</td>
<td></td>
</tr>
</tbody>
</table>
No formal assessment to assess group differences prior to treatment ..........0
Descriptive analysis of group equivalence prior to treatment ...............1
Formal measures documenting pre-experimental group equivalence (or use of statistics that account for pre-experimental differences) .....................2
Statistical weighting/correcting for lack of pre-experimental equivalence ....2

3. Implementing Therapist

A. Blindness

Implementing therapist(s) not blind to treatment ..................................0
Existence of blindness not reported ......................................................0
Not blind, but assessed by other than Tx administrator or primary researcher.1
Assessor blind to condition and/or study ............................................2

B. Credentials (minimum inclusion criteria for study was Level I EMDR training) *

Not specified ..................................................................................0
Education below doctorate level psychologist, unless Masters level with five or more years experience .........................................................0
Masters level with five or more years psychotherapy experience .............1
Doctoral level psychologist ..................................................................1
Doctoral level psychologist with three or more years psychology experience or licensed psychologist ..........................................................2
* Add one point to any of these ratings if implementing therapist is additionally trained at Level II (which will assume the prerequisite of 30 hours EMDR practice between Levels I and II training as specified in the manual).

4. Treatment Integrity

A. Fidelity to manualized procedures

No measures specified to ensure treatment integrity .................................................. 0
Measures taken within primary research team to ensure Tx fidelity .................. 1
Use of outside/independent evaluator to assess for Tx fidelity .......................... 2

B. Dependent Measures: If two of following categories are used, add points allotted for each together (maximum 3 points).

Use of only non-standardized client self-report measures ................................. 0
Use of standardized verbal report measures ....................................................... 1
Use of objective (physiologic) measures .............................................................. 2

C. Sample Size: This reflects cell size for each group in analysis.

Group size (either experimental, control, or both) five or less per group ...... 0
Group size (either experimental, control, or both) 6 - 11 per group .............. 1
Group size (either experimental, control, or both) 12 - 20 per group ............ 2
Group size (either experimental, control, or both) greater than 20 .............. 3

D. Statistical Power

No a priori design specifications to increase power ............................................ 0
Alpha level adjusted appropriately (i.e. for unilaterality) ..........1

A priori effect size estimates moderate to large based on previously reported findings ...........................................................................................................2

Group size per cell 40 or larger .................................................................3

Internal Validity Rating computed by assigning relevant points to each study, summing, and dividing by the total number of categories scored. This number will either be nine or ten as every category must be scored. Within groups design can not be scored on criteria 2A (group assignment), therefore code as “n/a” and divide total by nine.

Note.

Categories one, two, and four based on Campbell & Stanley’s (1963) delineation of “Factors jeopardizing internal and external validity” (p. 5). Category three based on Borg & Gall’s (1989) discussion of effects of observer bias and experimenter contamination on internal validity (pgs. 186-200), as well as recommendations regarding clinician competence developed by Waltz et al. (1998, pgs. 481-482).