
The Relationship between Efficacy and Methodology in Studies Investigating EMDR Treatment of PTSD



Louise Maxfield

*Psychology Department, Lakehead University,
Thunder Bay, Canada*



Lee Hyer

University of Medicine and Dentistry at New Jersey

The controlled treatment outcome studies that examined the efficacy of EMDR in the treatment of posttraumatic stress disorder have yielded a range of results, with the efficacy of EMDR varying across studies. The current study sought to determine if differences in outcome were related to methodological differences. The research was reviewed to identify methodological strengths, weaknesses, and empirical findings. The relationships between effect size and methodology ratings were examined, using the Gold Standard (GS) Scale (adapted from Foa & Meadows, 1997). Results indicated a significant relationship between scores on the GS Scale and effect size, with more rigorous studies according to the GS Scale reporting larger effect sizes. There was also a significant correlation between effect size and treatment fidelity. Additional methodological components not detected by the GS Scale were identified, and suggestions were made for a Revised GS Scale. We conclude by noting that methodological rigor removes noise and thereby decreases error measurement, allowing for the more accurate detection of true treatment effects in EMDR studies. © 2002 John Wiley & Sons, Inc. *J Clin Psychol* 58: 23–41, 2002.

Keywords: EMDR; PTSD; meta-analysis; methodology

The results attained in treatment outcome studies are intrinsically related to the methods employed to evaluate outcome. Research design, type of measures, sample selection and size, treatment delivery, and assessment are elements that may influence treatment out-

The data in this article were presented at the annual meetings of the Association for the Advancement of Behavior Therapy, Toronto, November 1999, and the International Society for Traumatic Stress Studies, Miami, November 1999.

Correspondence concerning this article should be addressed to: Louise Maxfield, Psychology Department, Lakehead University, 955 Oliver Road, Thunder Bay, ON, P7B 5E1, Canada; e-mail: jlmaxfie@flash.lakeheadu.ca.

come. Kazdin (1994) states that methodological differences may result in studies reaching different conclusions about treatment efficacy. Research design can even interfere with the detection of true treatment effects (Kazdin & Bass, 1989).

The controlled treatment outcome studies that examined the efficacy of Eye Movement Desensitization and Reprocessing (EMDR) in the treatment of posttraumatic stress disorder (PTSD) have yielded a range of results, with the efficacy of EMDR varying across studies. The current study sought to investigate whether these differences in outcome were related to methodological factors by examining differences in design features and by determining if higher ratings of methodological rigor predicted treatment effect sizes.

Van Etten and Taylor (1998) included EMDR studies in a meta-analysis that compared the effect sizes of 61 treatment outcome trials from 39 studies of chronic PTSD. This meta-analysis examined the comparative efficacy of treatments, using pharmacotherapies, psychological therapies (behavior therapy, EMDR, relaxation training, hypnotherapy, and dynamic therapy), and control conditions (pill-placebo, wait-list controls, supportive psychotherapies, and non-saccade EMDR control). The meta-analysis indicated that behavior therapy, serotonin reuptake inhibitors, and EMDR were the most effective treatments for PTSD. The authors observed that EMDR treatment used significantly fewer sessions than behavior therapy (4.6 vs. 14.8) and took significantly less time (3.7 vs. 10.1 weeks). Although instructive about the possible efficacy of EMDR, Van Etten and Taylor's (1998) meta-analysis did not provide information to explain the range of results achieved in individual EMDR studies.

The aim of the present article was to take the investigation a step further and to determine what methodological factors might have produced variability in the EMDR efficacy results, and to come to a conclusion with regard to the aggregate evidence for EMDR's effectiveness in the treatment of PTSD. We begin with a review of the research studies to identify methodological strengths and weaknesses and empirical findings. This is followed by an investigation in which the relationships between methodology ratings and effect size are examined for each study. Methodology is operationalized using the Gold Standard (GS) Scale (modified from Foa & Meadows, 1997). Three methodological factors that appeared to influence outcome and that were not detected by the GS Scale are identified. Suggestions are made for a Revised Gold Standard Scale. These findings have important implications for EMDR and PTSD research, and for treatment outcome research in general.

A Review of EMDR Studies Investigating Efficacy in the Treatment of PTSD

EMDR is hypothesized to facilitate the accessing and processing of traumatic memories and to bring these to an adaptive resolution, indicated by desensitization of emotional distress, reformulation of associated cognitions, and relief of accompanying physiological arousal. During EMDR the client focuses on emotionally disturbing material in sequential doses while simultaneously focusing on an external stimulus. Therapist directed eye movements are the most commonly used external stimulus but a variety of other stimuli including hand-tapping and aural stimulation are often used (Shapiro, 1991, 1994b, 1995). This dual (external/internal) focus is combined with frequent brief periods of focusing on new associations as they arise. According to Hyer and Brandsma (1997) and Fensterheim (1996), EMDR is a complex multicomponent, multistaged process, blending many elements of other effective therapies into a comprehensive treatment protocol. Shapiro (1995, 1999) maintains that EMDR, with its brief exposures to associated material, external/internal focus, and structured therapeutic protocol, is a distinctly different form of therapy.

Because of its claims for rapid effective treatment, EMDR has been subjected to many empirical tests and to much scientific scrutiny. Since Shapiro's (1989) original study, there have been 12 controlled randomized studies that investigated the efficacy of EMDR with participants diagnosed with PTSD. The following section summarizes the methodology and outcome of each study.

Boudewyns and Hyer (1996) sought to evaluate the addition of EMDR and an EMDR analogue with eyes closed (EC) to standard group therapy. They randomly assigned 61 combat veterans to one of three conditions: EMDR+group, EC+group, or standard group therapy. Every participant received 8 sessions of group therapy, with the EMDR and EC conditions also receiving 5 to 7 treatment sessions of either EMDR or EC. These veterans had chronic PTSD and were considered multiply disordered. Subjects in the EC+group and EMDR+group conditions showed superior improvement on mood and physiological measures compared to group therapy controls. Participants in all three conditions improved significantly on a structured interview measuring PTSD symptoms, with no group differences. The EMDR and EC sessions focused on one or two of the many traumatic events experienced by each veteran, and did not attempt to process the complete trauma history. This was done because of the chronicity of the veterans and the large, potential number of traumas. This choice appears to limit any generalization of effects (Fairbank & Keane, 1982), and may have focused the EMDR treatment unnaturally, preventing patients from exploring other memories when this was clinically indicated. Other limitations include concurrent group treatment received by participants. Treatment fidelity was assessed as variable, based on reviews of video-tapes of randomly selected sessions. A blind trained independent interviewer conducted pre- and posttests. This study indicates that the addition of EMDR or EC to group treatment may improve outcome.

Carlson, Chemtob, Rusnak, Hedlund, and Muraoka (1998) randomly assigned 35 Vietnam combat veterans with PTSD to a wait list control, or to 12 treatment sessions of biofeedback relaxation, or EMDR. At posttreatment, the EMDR group had significantly lower scores on instruments measuring PTSD and depression than the wait list. At three-month follow-up, EMDR had significantly lower scores than the biofeedback relaxation group on measures of PTSD and self-reported symptoms. Both treatment groups and the wait list control showed significant improvement on physiological measures with no differences between groups, and with the decrease in physiological arousal maintained at three-month follow-up. The strengths of this study include provision of a full course of treatment to combat veterans, and comparison with another treatment. It controls for the often neglected variable of therapist allegiance (Hollon, 1999), as the non-EMDR subjects received the treatment to which the therapist had allegiance. Although it is not known whether biofeedback relaxation therapy has any unique effect beyond nonspecific therapeutic effects, it has been found effective in the treatment of PTSD (Peniston, 1986; Peniston & Kulkosky, 1991). The only posttreatment assessment conducted by a trained blind independent assessor was at nine-month follow-up, and it was completed by 9 of the 10 EMDR subjects. It confirmed the maintenance of treatment effects, with 78% of the EMDR subjects no longer meeting diagnostic criteria for PTSD.

Lee and colleagues (Lee & Gavriel, 1998; Lee, Gavriel, Drummond, Richards, & Greenwald, 2002) randomly assigned 24 civilian subjects with PTSD to Stress Inoculation Training with Prolonged Exposure (SITPE) or EMDR. After serving as their own controls during a wait list period, participants were provided with seven 60-minute treatment sessions. Measures were collected at pre- and posttreatment and at three-month follow-up. Both EMDR and SITPE were found to be effective, with significant improvement on PTSD and depression measures. At follow-up 83% of the EMDR subjects and 75% of the SITPE subjects no longer met PTSD criteria. The only difference found

between groups was on the Intrusion subscales of the PTSD measures with the EMDR group showing significantly greater improvement. The trained interviewer was not blind or independent. Fidelity checks were satisfactory for both treatments. Strengths of this study include its comparison to an empirically validated treatment, SITPE. This study indicates that EMDR and SITPE may be fairly equivalent in treatment effectiveness. The authors point out that EMDR may be more efficient by not requiring homework assignments. EMDR required an average of 3 hours of homework, SITPE required 28 hours.

Marcus, Marquis, and Sakai (1997) compared EMDR to "Standard Kaiser Care" (SKC) in an outpatient HMO. SKC consisted of individual therapy (cognitive, psychodynamic, or behavioral). Sixty-seven individuals with PTSD were randomly assigned to EMDR or SKC treatment and received an unlimited number of 50-minute treatment sessions. EMDR participants attained symptom reduction with significantly greater rapidity and had significantly fewer treatment sessions than SKC participants. EMDR produced significantly lower scores than SKC, after three sessions and at posttreatment, on measures of PTSD symptoms, depression, and anxiety. After three sessions, 50% of the EMDR participants no longer met criteria for PTSD, compared to 20% of the SKC group. At posttreatment, 77% of the EMDR group (including 100% of the single-trauma victims) no longer met criteria for PTSD compared to 50% of the SKC group. Limitations of this study include the numerous statistical analyses without Bonferroni corrections. The independent interviewer (with unspecified training) was not blind to treatment condition due to participant response. The therapists had pre-assessed treatment fidelity. An unspecified number in each group had medication-related supervision appointments. Strengths of this study include its high external validity: This study indicates that EMDR may be superior to the wide variety of treatments used in an HMO setting. However, the heterogeneous nature of SKC therapies precludes specific knowledge of their effectiveness for PTSD treatment, thus limiting the conclusions that can be drawn.

In a study evaluating treatment process and outcome, Rogers et al. (1999) provided a single session of EMDR or Exposure therapy treatment. The session focused on the most distressing identified combat memory and used measures designed to be sensitive to change on that one treated memory. Twelve combat veterans with PTSD were randomly assigned to treatment condition. Both groups significantly improved on a trauma measure (as it was applied to that particular memory). A posttest measure in which participants monitored the severity of intrusive recollections for the one memory showed a significant decrease for the EMDR group compared to the Exposure group. Pre- and postassessments were done by an independent blind assessor. Treatment fidelity was not reported, outcome measures considered of self-reports collected by a trained assessor. As the purpose of the study was a comparison of therapeutic process, not of treatment efficacy, the results must be considered in this context. This study illustrates the importance of using sensitive and appropriate measures when only one traumatic memory is targeted in subjects with multiple traumas.

Rothbaum (1997) randomly assigned 18 adult female rape victims with PTSD to four 90-minute sessions of EMDR or a wait-list control group. The self-report scores of the EMDR participants on PTSD and depression scales showed a mean decrease of more than two standard deviations at posttreatment, which was significantly superior compared to wait-list controls. The mean scores of the EMDR group on other self-report measures decreased to within normal range, but, due to low power, were not significantly different from the wait-list group. At posttreatment, 90% of the subjects in the EMDR group no longer met full criteria for PTSD compared to 12% of the wait-list group. Results were evaluated by a trained blind independent assessor using structured interviews and self-report measures. Rothbaum was the only therapist for participants in the

initial EMDR condition, but other therapists were added for the delayed treatment condition, which followed the same pattern of recovery, mitigating against the confound of therapist effects. This study is limited by the wait-list design, which controls for repeated testing and the passage of time, but does not control for nonspecific factors such as therapeutic alliance, expectations, or placebo effects. Other limitations include the small sample size, and the provision of concurrent treatment to 3 EMDR and 2 wait-list participants.

Scheck, Schaeffer, and Gillette (1998) compared EMDR to an active listening (AL) control with a group of 60 traumatized young women who were engaging in high risk behavior such as sexual promiscuity, runaway behavior, or substance abuse. Seventy-seven percent were diagnosed with PTSD using a structured interview. The women received two 90-minute treatment sessions, and had a homework assignment of journal writing. Both AL and EMDR resulted in significant improvement on all measures, which included measures of PTSD, depression, anxiety, and self-concept. The effects of EMDR were significantly greater than that of AL on all measures except self-concept. Treatment gains were maintained at three-month follow-up for both groups. Posttreatment measures were collected by a trained independent blind assessor, but did not include a structured diagnostic interview. Limitations of this study include the lack of posttreatment assessment of PTSD diagnosis, provision of only two treatment sessions, and lack of treatment integrity ratings. Strengths of this study include the evaluation of EMDR with a high-risk population. These results indicate that EMDR is superior to a condition that controls for some of the nonspecific effects of treatment such as attention, therapeutic rapport, and active listening.

Vaughan et al. (1994) assigned 36 civilian participants, 78% of whom were diagnosed with PTSD, to EMDR, imaginal exposure (IHT), applied muscle relaxation training (AMR), and wait-list conditions. Those participants without PTSD failed only to meet Criterion C, which requires three avoidance symptoms. Three to five treatment sessions were administered, with daily homework assigned to the IHT and AMR groups only. The IHT group listened daily for 60 minutes to an audio-taped description of their trauma, and recorded thoughts and feelings. All treatments led to significant decreases in depression and PTSD symptoms for subjects in the treatment groups as compared to those on the wait list. A comparison between treatment groups found a significantly greater reduction at posttreatment for the EMDR group on PTSD intrusive symptoms. At follow-up, 70% of participants in all treatment groups no longer met PTSD diagnostic criteria. Blind independent assessments were conducted by a trained assessor at pretreatment, posttreatment, and at three-month follow-up. Limitations of this study include the lack of treatment fidelity ratings, limited number of treatment sessions, and different amounts of treatment received by the groups with additional daily homework time in the AMR and IHT groups. Strengths of this study include the comparison of EMDR with an exposure treatment, previously found effective for PTSD in a case series (Vaughan & Tarrier, 1992). However, IHT was not conducted in accordance with standard exposure designs, and may not be representative of CBT exposure therapies.

Wilson, Becker, and Tinker (1995) randomly assigned a sample of 80 traumatized individuals to wait-list or EMDR conditions. Fifty-four percent of participants did not receive a PTSD diagnosis; of these, 75% met four of Criteria A, B, C, D, and E. Subjects received three 90-minute sessions of EMDR. At posttreatment and three-month follow-up, significant differences were found between EMDR and wait-list groups on measures of PTSD symptoms, depression, and anxiety. This improvement was clinically significant, with means for all measures moving into normal range. When treatment was provided to the wait-list group after three-month follow-up, treatment effects were replicated,

with significant effects for all measures. A linear regression analysis indicated that treatment gains did not vary as a function of pretreatment symptom severity or PTSD diagnosis. There was assessed treatment fidelity and a trained blind independent assessor administered all self-report measures. This study is limited by the lack of a structured interview to assess posttreatment diagnosis and the wait-list design. In a 15-month follow-up study (Wilson, Becker, & Tinker, 1997), 32 of the original 37 subjects with PTSD were interviewed by an independent assessor. The assessment was not blind as all subjects had received EMDR treatment by this time. There was an 84% reduction in PTSD diagnosis compared to pretreatment. Because this design does not control for influences during the 15-month period, such as other treatment or spontaneous remission, it is not possible to conclude that the maintenance of posttreatment outcome resulted solely from EMDR treatment effects.

Deville, Spence, and Rapee (1998) assigned 51 combat veterans with PTSD, in non-random blocks to one of three conditions: Standard Psychiatric Support at other settings (SPS), two sessions of EMDR, or two sessions of an EMDR variant (REDDR) in which subjects concentrated on a stationary flashing light. Most participants were also receiving other concurrent mental health treatment. At posttreatment all groups showed significant improvement on measures of PTSD, depression, anxiety, and problem coping. There were no differences between the three groups. Measures of reliable change indicated that 67% of the EMDR group, 42% of the REDDR group, and 10% of the SPS group were reliably improved. Forty-six percent of the veterans did not mail back their follow-up measures, and the authors note a diminishing of treatment effect over time. This study lacked random assignment, with only one therapist and subjects assigned in nonrandom blocks. The trained assessor was not blind or independent. Treatment delivery did not follow standard protocols: Deficiencies include inaccurate instructions, rating the negative cognition, repetitions of the negative cognition during treatment, and excessive use of SUD ratings (see Shapiro, 1995). Other limitations include different assessment procedures at pre- and posttest, participants receiving concurrent mental health treatment, and an insufficient number of sessions for multiply traumatized veterans.

Deville and Spence (1999) compared a CBT variant developed by Devilly, Trauma Treatment Protocol (TTP) to EMDR. TTP is a treatment package combining elements of CBT, Stress Inoculation Training, and Prolonged Exposure. Although these components have been individually found effective for PTSD treatment, their combination and modification in TTP have not been empirically assessed. Twenty-three civilian subjects with PTSD were assigned in nonrandom blocks to eight sessions of either EMDR or TTP. After the initial session, which provided an explanation of the treatment, 31% of EMDR subjects dropped out of treatment before receiving any EMDR. Both EMDR and TTP were significantly effective on all measures. TTP was significantly more effective than EMDR on combined PTSD measures, and a scale of global function. At three-month follow-up, scores on a self-report PTSD measure indicated that 58% of the TTP subjects no longer met PTSD criteria on a self-report measure compared to only 18% of the EMDR group. Improvement was maintained with TTP, but worsened with EMDR. Limitations of this study include the high dropout rate, large number of statistical analyses done with no Bonferroni correction for Type I error, change in assessment procedures at post test, lack of an independent blind assessor, and partial randomization. One therapist provided treatment to every participant in the TTP condition, and to most EMDR participants. Although treatment integrity was rated as high by an independent EMDR therapist, description of the technique (see description above in Devilly et al., 1998) indicates a lack of conformity to standardized procedures, with multiple deficiencies in treatment delivery.

Jensen (1994) randomly assigned 25 Vietnam combat veterans suffering from PTSD to a wait-list condition or two sessions of EMDR. Most of the veterans were receiving concurrent treatment. No difference was found between groups at posttreatment. Instead of improvement, the condition of the veterans actually deteriorated. To assess outcome, the researchers used a diagnostic instrument, insufficient to measure the small amount of change that may have been achieved in two sessions. Other limitations include poor treatment fidelity, lack of a trained blind independent assessor, insufficient number of sessions, and concurrent treatment. The wait-list condition was confounded by informing participants that no treatment would be provided and encouraging them to seek treatment elsewhere.

An Investigation of the Relationship between Methodology and Efficacy

The evidence appears to indicate that the EMDR studies with better methodology were also the studies with better outcomes. It was hypothesized that more rigorous methodology would be significantly correlated with treatment effect size. This investigation was conducted to assess the relationship between demonstrated efficacy and methodological rigor.

Method

Inclusion Criteria. There have been 17 controlled research studies that examined the use of EMDR with PTSD. This analysis evaluated all 12 controlled studies that investigated the efficacy of EMDR in the treatment of PTSD, including one study (Deville & Spence, 1999) that purported to be a randomized study, but in which treatment was provided in nonrandom blocks. Not included in this analysis are the two controlled studies (Boudewyns, Stwertka, Hyer, Albrecht, & Sperr, 1993; Shapiro, 1989) that did not provide data for effect size calculation. Also not included are the three component studies (D.L. Wilson, Silver, Covi, & Foster, 1996; Pitman et al., 1996; Renfrey & Spates, 1994) that only investigated EMDR's mechanism of action. Such studies control for only one aspect (eye movements) of a complex process. They cannot be used to evaluate treatment efficacy since the control condition may still contain the possible effective mechanism, which could be focused attention, distraction, stimulation of an orienting response, bilateral activation, or rhythmic activity (Shapiro, 1995).

Procedure

The 12 controlled studies were reviewed. Where information in the studies was incomplete or unclear, the researchers were contacted to ensure accuracy. The methodological rigor of each study was assessed by three raters who applied Foa and Meadow's (1997) gold standards (GS) to pre- and posttreatment methods. One rater was the first author; the other raters were blind to the hypotheses in this study. Pre-post effect sizes and comparison effect sizes were calculated. Regression analyses were conducted to determine if there was a significant relationship between outcome and methodology.

Measures

Effect Size (ES). Pre-post ESs were calculated for the primary PTSD measure used in each study using Cohen's *d* statistic (the difference of the pretreatment mean minus post-

treatment mean, divided by the pooled standard deviation). Comparison ESs were also calculated (the difference of post-treatment means, divided by the pooled standard deviation).

Gold Standard (GS) Scale. To evaluate methodology in treatment outcome studies, Foa and Meadows (1997) developed a set of seven GSs. These include the following: (a) "Clearly defined target symptoms," so that appropriate measures can be employed to assess improvement, with specifications of inclusion and exclusion criteria; (b) "Reliable and valid measures," with good psychometric properties; (c) "Use of blind evaluators," other than the treatment provider, to collect assessment measures; (d) "Assessor training," with demonstrated interrater reliability; (e) "Manualized, replicable, specific treatment programs," to ensure consistent and replicable treatment delivery; (f) "Unbiased assignment to treatment," either random assignment to conditions, or stratified sampling, with treatment delivered by at least two therapists; (g) "Treatment adherence," evaluated by treatment fidelity ratings.

In our application of these GS criteria to the EMDR studies, we developed the Gold Standard (GS) Scale, with seven items each representing the corresponding GS. A 3-point Likert scale was used for each item: a study that fully met the criteria was rated at "1"; a study that partially met the criteria was rated at "0.5"; and a study that did not meet the criteria received "0" (see Table 1). The total possible score on the GS Scale is 7.

Results

Efficacy and Methodology.

Methodology. The GS Scale scores ranged from 3.5 to 6.5 with a mean of 5.42 and a SD of 1.10 (see Table 2). There were two methodological factors that performed less adequately across the studies. These were GS #3, blind independent assessment, and GS #7, treatment fidelity. Both these GSs had mean of .54 (range 0–1). The means of the other GSs ranged from .71 (GS #4) to 1.0 (GS #5). Three methodological deficits were identified in the methodological review, but were not detected by the GS Scale. These include participants receiving concurrent treatment, insufficient course of treatment, and the use of self-report instruments *only*. In a later section, we discuss the utility of creating additional GSs to assess these elements.

Efficacy. Pre-post ESs for the EMDR conditions ranged from $-.50$ to 2.22, with a mean of 1.23 and a standard deviation (SD) of 0.79 (see Table 2). EMDR compares favorably to the ESs found in Kazdin and Bass's (1989) meta-analysis of 85 psychological treatment outcome studies. Kazdin and Bass reported that the average ES for treatment comparisons was 0.50, and for wait-list comparisons it was 0.85. When EMDR was compared to wait-list (Rothbaum, 1997; Wilson et al., 1995) the comparison ES was 2.17. When compared to five non-CBT treatments (psychotherapy, relaxation, etc.), the average comparison ES was 0.81. EMDR was compared to CBT/exposure in four studies with an average comparison ES of 0.45. Although it appears that EMDR may be slightly superior to CBT, the wide variation in the outcome of these four studies ($SD = .78$) prevents any conclusions and makes apparent the need for methodologically sound comparison studies.

Relationship between ES and GSs. The regression analysis between the GS Scale score and pre-post ES resulted in a significant correlation for the EMDR condition,

Table 1
Gold Standard (GS) Scale
 (adapted from Foa & Meadows, 1997)

GS #1	Clearly defined target symptoms. 0: no clear diagnosis, symptoms not clearly defined .5: not all subjects with PTSD, clear defined symptoms 1: all subjects with PTSD
GS #2	Reliable and valid measures. 0: did not use reliable and valid measures .5: measures used inadequate to measure change 1: reliable, valid, and adequate measures
GS #3	Use of blind independent assessor. 0: assessor was therapist .5: assessor was not blind 1: assessor was blind and independent
GS #4	Assessor reliability 0: no training in administration of instruments used in the study .5: training in administration of instruments used in the study 1: training with performance supervision, or reliability checks
GS #5	Manualized, replicable, specific treatment. 0: treatment was not replicable or specific 1: treatment followed EMDR training manual, Shapiro 1995
GS #6	Unbiased assignment to treatment. 0: assignment not randomized .5: only one therapist, OR other semi-randomized designs 1: unbiased assignment to treatment
GS #7	Treatment adherence 0: treatment fidelity poor .5: treatment fidelity unknown, or variable 1: treatment fidelity checked & adequate

$R = .67$, $F(1, 10) = 7.98$, $p = .02$, but not for control conditions (see Figure 1). Because the result of this omnibus test was significant for the EMDR condition, it was considered appropriate to do further regression analyses between ES and each GS item to explore the relationship between GS and ES. To increase power, and because only one direction of relationship made sense, one-tailed tests were used for these secondary analyses (see Table 3). There were significant correlations with GS #7 (treatment fidelity), $R = .79$, and GS #4 (trained reliable assessor), $R = .54$.

Of the 12 studies, nine (Boudewyns & Hyer, 1996; Carlson et al., 1998; Lee et al., 2002; Marcus et al., 1997; Rogers et al., 1998; Rothbaum, 1997; Scheck et al., 1998; Vaughan et al., 1996; S.A. Wilson et al., 1995, 1997) were above the mean GS Scale score of 5.5. These nine studies all found EMDR effective. Three studies (Devilley & Spence, 1999; Devilly et al., 1998; Jensen, 1994) were below the GS mean; they found EMDR noneffective or minimally effective (see Table 4).

In the nine more rigorous studies, the EMDR pre-post ESs ranged from 0.67 to 2.22, with an average ES of 1.57 (see Table 4). The control conditions had an average pre-post ES of 0.70. Six of these nine studies calculated the decrease in PTSD diagnosis in the EMDR groups, which was substantial, ranging from 61% to 90%. In the three studies with less rigorous methodology, the EMDR pre-post ESs had a mean of 0.21, and the

Table 2
Gold Standards and Effect Sizes for Each of the Controlled EMDR PTSD Studies

STUDY	Gold Standard Scale Scores								Effect Sizes		
	Total	GS#	GS#	GS#	GS#	GS#	GS#	GS#	EMDR	Control	Comparison
		1	2	3	4	5	6	7			
Boudewyns & Hyer	6.5	1	1	1	1	1	1	.5	0.67	.38	.46
Carlson et al.	5.5	1	1	0	.5	1	1	1	1.53	1.24	1.44
Devilley et al.	3.5	1	1	0	.5	1	0	0	0.37	-.01	-.01
Devilley & Spence	3.5	1	1	0	.5	1	0	0	0.75	1.81	-.67
Jensen	4.0	1	.5	0	.5	1	1	0	-0.50	-1.0 (wl)	1.01 (wl)
Lee et al.	5.5	1	1	0	.5	1	1	1	2.07	1.55	.62
Marcus et al.	6.0	1	1	.5	1	1	1	.5	2.03	1.13	.76
Rogers et al.	6.0	1	1	1	.5	1	1	.5	1.10	.20	1.13
Rothbaum	6.5	1	1	1	1	1	.5	1	2.22	.56 (wl)	2.85 (wl)
Scheck et al.	6.0	.5	1	1	1	1	1	.5	1.60	.52	.76
Vaughan et al.	6.0	.5	1	1	1	1	1	.5	1.38	.60	.69
S.A. Wilson et al.	6.0	.5	1	1	.5	1	1	1	1.49	.19 (wl)	1.50 (wl)
Means	5.42	.875	.958	.542	.708	1.00	.792	.542	1.23	0.60	0.88

Note. Gold Standard Scale: #1, Diagnosis; #2, Measures; #3, Blind independent assessor; #4, Trained reliable assessor; #5, Manualized treatment; #6, Randomization; #7, EMDR Treatment Fidelity. Rating of each gold standard: 1—criteria for the standard fully met; 0.5—criteria partially met; 0—criteria not met. Effect size = Cohen's *d*. "wl" = wait-list control.

control conditions had an average pre-post ES of 0.27. Only one of these studies calculated the decrease in PTSD diagnosis in the EMDR groups; this was minimal, 36%.

Rating. During the rating process, it was discovered that method information was incomplete for some studies, making it impossible to rate some items. Study authors were

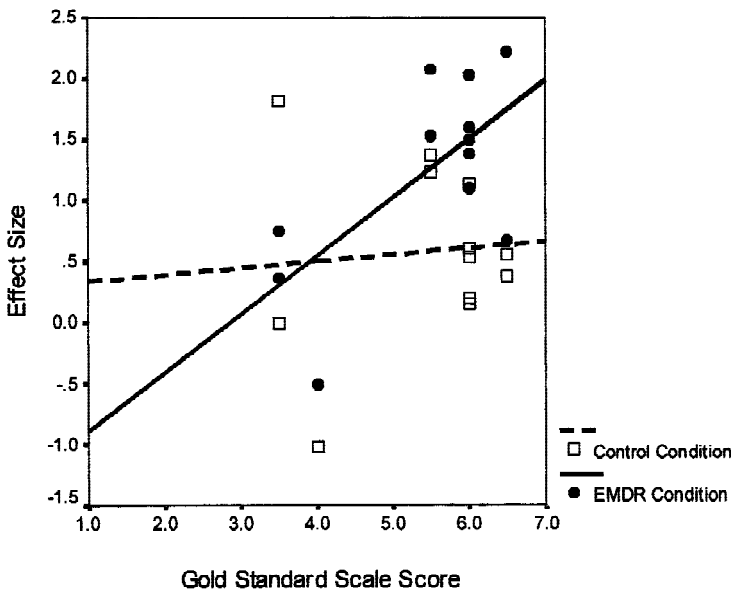


Figure 1. Gold standard scale score and effect size for EMDR and control conditions.

Table 3
*Bivariate Correlations of Gold Standard (GS) Scale
 with Pre-Post Effect Size for EMDR*

GS Scale Total Score		.69 ^a
GS #1	Diagnosis	-.20
GS #2	Measures	n/a
GS #3	Blind, independent assessor	.35
GS #4	Trained reliable assessor	.54 ^b
GS #5	Manualized treatment	n/a
GS #6	Random assignment	.21
GS #7	Treatment fidelity	.79 ^c

^asignificant at the 0.05 level (2-tailed); ^bsignificant at the 0.5 level (1-tailed); ^csignificant at the 0.01 level (1-tailed).

contacted to ensure accuracy, but this process prohibited the calculation of interrater reliability. Although the raters were highly concordant in their ratings and comprehensive attempts were made to minimize error, error cannot be ruled out. Differences between raters were resolved by consensus and/or by assigning the lower rating. Differences arose from varying interpretations or from overlooking a methodological aspect. For example, although one rater gave the S.A. Wilson et al. (1995) study a “1” for GS #4, indicating reliable assessment, the study received a “0.5” because, although the assessor was trained, there were no reliability checks. Similarly, one rater gave the Carlson et al. (1998) study a “1” for GS #3 because there was a blind independent assessor at 9-month follow-up; however, because the assessor was not independent nor blind at post-treatment, the study received a “0” for GS #3. The GS ratings for each study should be considered only preliminary and descriptive rather than qualitative.

Discussion

Relationship between Methodology and Outcome

The application of the GS Scale to the EMDR studies allowed a preliminary examination of the relationship between GSs and treatment outcome. The findings indicate a significant correlation between methodology and outcome. As methodology became more rigorous, the treatment effect became larger. The relationship between methodology and

Table 4
*Studies Grouped According to GS Scale Scores: Comparison of ES and Reported Decrease
 in PTSD Diagnosis*

	N	Mean Pre-Post Effect Size (SD)		Reported Decrease in PTSD Diagnosis After EMDR
		EMDR	Control	
Studies with scores above the mean of the GS Scale	9	1.57 (.50)	0.70 (.47)	61–90%
Studies with scores below the mean of the GS Scale	3	0.21 (.64)	0.27 (1.43)	36%

outcome is apparent when the studies are grouped according to methodological strength (see Table 4). The more rigorous methodological studies achieved large ESs and indicated that EMDR was efficacious, and more efficacious, than control conditions. This has not always been the trend in psychotherapy research, where increased rigor resulted in less robust ESs. It may be that methodological rigor removes noise and decreases error measurement, allowing for the more accurate detection of treatment effects. Perhaps attention to methodological detail also parallels attention to the quality of treatment delivery.

The GS Scale was significantly correlated with EMDR ES. Two items appeared to be important predictors of ES: assessment reliability and treatment fidelity. This suggests that the critical factor is the methodological rigor evidenced in those studies that used reliable assessment and provided adequate reliable treatment. Of interest is the fact that these two criteria are perhaps most reflective of quality treatment that is efficacious. Treatment effects are most apparent when there is an empirically supported treatment, which uses reliable and sensitive measures, allowing for the objectification of the method, with careful application of the treatment procedures. In a sense these are at the heart of the treatment.

The Gold Standard Scale

The Gold Standard (GS) Scale is a logical representation of reasonably understood standards of outcome research in psychotherapy. It was adapted from the seven GSs suggested by Foa and Meadows (1997). The quantification of methodological rigor using the GS Scale permits the assessment and comparison of methodological features, and allows clear differentiation among studies.

In this section we discuss the application of each standard, in terms of what this reveals about the methodology used in the EMDR studies, how each GS appears to relate to treatment effect, and the apparent relevance and usefulness of the standard.

GS #1: Clearly Defined Target Symptoms. GS #1 requires clearly defined target symptoms, with the specification of diagnosis and inclusion/exclusion criteria. As Foa and Meadows (1997) point out, there are two potential biases when using non-PTSD samples: Less severe participants may recover more readily (inflating efficacy), but their recovery may be in smaller increments and more difficult to measure (minimizing efficacy). The three EMDR studies with mixed samples had ESs in the mid-range (see Table 2). S.A. Wilson et al. (1995) and Scheck et al. (1998) used mixed samples, and reported that there were no differences between PTSD and non-PTSD response. At 15-month follow-up, S.A. Wilson et al. (1997) determined that those with and without PTSD diagnoses showed an equivalent amount of change, with an 84% reduction in PTSD diagnosis. However because those participants with PTSD diagnoses at intake had started with more severe symptoms, they were more symptomatic after three treatment sessions than those without PTSD.

GS #2: Reliable and Valid Measures. All the studies used measures with good psychometric properties. However, psychometric instruments also need to be accurate and sensitive enough to detect the amount of change that would be anticipated from the treatment. In research examining the effects of imaginal flooding for PTSD, Fairbank and Keane (1982) found that desensitization generalized from memory to memory only when features were similar, and that distress from nonsimilar traumatic memories was not reduced. Consequently, when treatment focuses on only one incident, change in symp-

toms can be specific to that event only. With survivors of single trauma, this may result in complete amelioration of symptoms. With survivors of multiple traumas, such as veterans, this may eliminate all distress related to the treated incident, but the individual may still suffer from severe PTSD related to other nontreated events. Therefore using a diagnostic tool as the sole outcome measure after two treatment sessions with combat veterans (e.g., Jensen, 1994) is inappropriate and a violation of GS #2. It is also possible that commonly used symptom measures may not detect minor amounts of symptom change. More precise measurement may result when assessment is specific to the treated memory (e.g., Rogers et al., 1999).

GS #3: Use of Blind Independent Evaluators. When assessors are not blind to treatment condition, their expectancies and bias may influence how they interact with and evaluate participants. When assessors are not independent, therapeutic alliance may influence the process. Some researchers maintain that this GS refers to the use of a “blind independent interviewer,” and that failure to use interview measures, results in failure to meet this GS. Regardless, it is the opinion of the authors that experimenter bias and type of measurement are two separate dimensions, which should be assessed and rated separately (see recommendations for GS #9). Accordingly, GS #3 was rated according to the blind independence of the assessor. This was one of the two least adequately performed methodological elements in the EMDR studies, but did not appear directly related to outcome. Three of the less rigorous and two of the more rigorous studies (Carlson et al., 1998; Lee et al., 2002) did not use blind independent evaluators to collect posttreatment measures. In the Marcus et al. (1998) study the assessor was independent but not blind due to participant response. In the Carlson et al. study, when measures were collected at nine-month follow-up by a blind independent assessor, earlier results were replicated, suggesting a minimal effect for GS #3.

GS #4: Assessor Reliability. Assessor training is essential to ensure that the administration of interview, observation, and self-report measures is standardized. Reliability checks should be done on a regular basis. Because most studies did not report on the training and qualifications of the assessors, information on this aspect was acquired by contacting the following researchers: Becker, Carlson, Devilly, Hyer, Lee, Marquis, Rothbaum, and Vaughan. All studies used assessors who were trained in administration of the instruments used. Those studies that provided reliability checks or close supervision of assessors tended to have larger effect sizes, and there was a significant correlation between ES and GS #4. Ensuring that measurement is reliable may decrease noise and increase the possibility of detecting change.

GS #5: Manualized Treatment. All studies were considered to have used a treatment manual because all claimed to have been conducted in accordance with the manuals received during training.

GS #6: Random Assignment. Random assignment is an integral part of all controlled research. Foa and Meadows (1997) state that each treatment must be delivered by more than one therapist, with random assignment to therapists within conditions, so that therapist and treatment effects can be separated. There are three studies that used only one (primary) therapist (Devilly et al., 1998; Devilly & Spence, 1999; Rothbaum, 1997). Random assignment to treatment conditions is also essential. While Rothbaum met this criterion, the other two studies used an incomplete randomization procedure (Devilly

et al., 1998) and nonrandom blocks (Deville & Spence, 1999). Both the studies with nonrandom assignment to treatment condition were among those with poor outcome.

GS #7: Treatment Fidelity. The importance of treatment fidelity is an issue that has been widely debated (Greenwald, 1994; Rosen, 1999). The findings of the current study shed some light upon this disagreement. The large significant correlation between GS #7 and ES indicates a strong relationship between outcome and fidelity: The studies with largest ESs tended to be those with assessed treatment fidelity. This finding is in accordance with the position of Elkin (1999) who argues that the role of the therapist is crucial and that therapist competency should be assessed in all treatment outcome studies. In three of the less rigorous studies with poor outcome (Deville et al. 1998; Devilly & Spence, 1999; Jensen, 1994) the EMDR procedure was not performed according to standard protocol. Although the therapists in the Marcus et al. (1997) study had pre-assessed treatment fidelity, pre-assessment does not prevent the possibility of therapist drift. Therefore the more conservative rating of 0.5 was assigned to this study. Other studies (Boudewyns & Hyer, 1996; Scheck et al., 1998; Vaughan et al., 1996) that did not assess treatment fidelity or that had variable fidelity had lower ESs.

Additional Methodological Standards

We recommend that a more complete scale be devised to include the three methodological items identified as not adequately assessed by the GS Scale. Although it could be argued that these issues are adequately addressed within the 7-item GS Scale, it is our opinion that they represent unique elements. This Revised Gold Standard (RGS) Scale would include the following additional items to permit accurate measurement of possible methodological shortcomings: GS #8, no concurrent treatment; GS #9, multimodal and interview measures; GS #10, adequate course of treatment length (see Table 5). Based on the observation that multiple-trauma survivors require more extensive treatment, GS #10

Table 5

Additional Items Added to GS Scale to Create the Revised Gold Standard (RGS) Scale

GS #8	No confounded conditions. 0: most subjects receiving concurrent psychotherapy .5: a few subjects receiving concurrent psychotherapy, or unspecified and no exclusion for concurrent treatment 1: no subjects receiving concurrent psychotherapy
GS #9	Use of multimodal measures: 0: self-report measures only .5: self-report, plus interview or physiological or behavioral measures 1: self-report plus two or more other types of measures
GS #10	Length of treatment for participants with single trauma (e.g., civilian) PTSD 0: 1–2 sessions .5: 3–4 sessions 1: 5+ sessions Length of treatment for participants with multiple trauma (e.g., combat) PTSD 0: 1–6 sessions .5: 7–10 sessions 1: 11+ sessions

is operationalized to evaluate the adequacy of treatment length, differentiating between multiple-trauma (e.g., combat veterans) and single-trauma (e.g., civilian) participants. This differentiates GS #10 from the actual number of sessions.

Although it is not appropriate to reanalyze the same data using a revised scale, a preliminary analysis was conducted to ascertain if the addition of these three items actually contributed to methodological evaluation. As expected, there was a large significant correlation between RGS scores and ES. The RGS Scale appeared to provide a more comprehensive evaluation of methodological strength and better differentiation among the studies, with seven studies above the mean RGS score and five below (see Figure 2). There was an unexpected interesting finding: a significant relationship between outcome and methodology for the control conditions, $R = .57$, $F(1, 10) = 4.9$, $p = .05$. This suggests that those studies with better methodology more accurately detected treatment effect even for the control conditions. Two of the additional items had significant one-tailed correlations with ES: GS #8 (no concurrent treatment) and GS #10 (adequate course of treatment). The correlation for GS #9 (interview/multimodal measures) was nonsignificant, even after controlling for GS #3 (blind independent assessment).

In the following section we discuss the application of these additional standards, in terms of what each reveals about the methodology used in the EMDR studies, how each appears to relate to treatment effect, and the apparent relevance and usefulness of the standard.

GS #8: No Concurrent Treatment. Confounded treatment conditions can obscure true effects by diminishing construct validity, and may increase the amount of “noise” and the likelihood of a Type II error. However, the provision of concurrent treatment could also indicate that the participants had more chronic PTSD, and/or more severe symptoms, suggesting that subjects were less responsive to treatment, or required a more extended course of treatment. Those studies whose participants received additional concurrent treatment tended to have poorer results. In two of the less rigorous studies (Deville

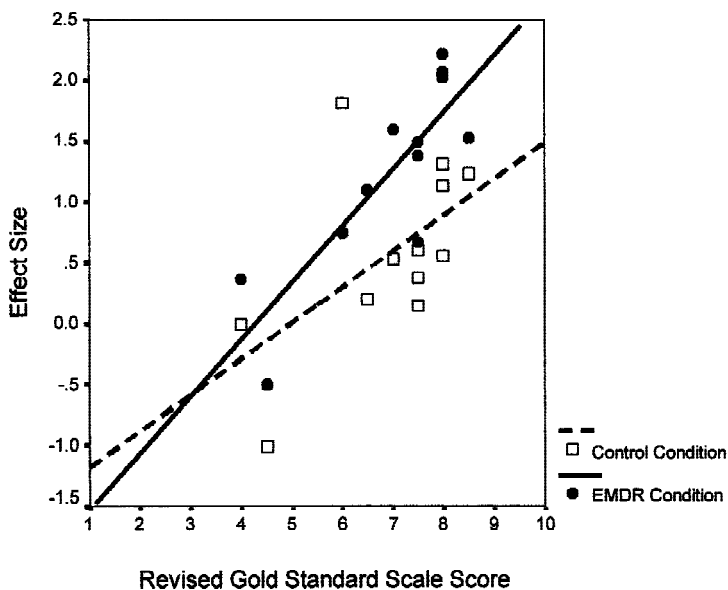


Figure 2. Revised gold standard scale score and effect size for EMDR and control conditions.

et al., 1998; Jensen, 1994), almost all subjects were receiving concurrent treatment. Only one of the more rigorous studies (Boudwyns & Hyer, 1996) had confounded treatment, with all subjects in group treatment, and this study had the poorest outcome of the more rigorous studies. It appears that GS #8 has a strong relationship with treatment outcome in the EMDR studies.

GS #9: Multimodal Assessment. When studies use multimodal measures, they assess a wide range of pathology and outcome with interview, behavioral, and physiological measures. Interviews are essential for diagnostic assessment, and it is assumed that multimodal measures provide a more accurate evaluation than self-report instruments. Although the use of such measures by researchers may indicate a commitment to methodological rigor, it could also indicate the availability of greater financial resources in the research program. There appeared to be a full range of outcomes among the EMDR studies using such instruments, and this GS did not have any direct relationship to effect size.

GS #10: Adequate Course of Treatment. It appears that persons with multiple traumas require multiple treatment sessions. Accordingly, GS #10 was operationalized to differentiate between single-trauma (civilian) and multiple-trauma (combat veteran) samples, with the assumption of a need for lengthier treatment in multiply traumatized populations (see Table 5). If an insufficient course of treatment is provided to participants, this may interfere with the assessment of treatment efficacy. GS #10 appeared to have a strong relationship with EMDR treatment outcome. Only one of the combat veteran studies provided a full course of treatment (Carlson et al., 1998) and it resulted in a large effect. The other veteran studies provided treatment of only one or two memories. With the exception of Rogers et al. (1999) who used measures specific to the treated incident, these studies (Boudewyns & Hyer, 1996; Devilly et al. 1998; Jensen, 1994) had smaller ESs. Only one of the civilian studies (Marcus et al., 1998) differentiated between single- and multiple-trauma participants and a more rapid treatment response was identified with single-trauma survivors. Although it is acknowledged that participants with complex PTSD require extended therapy, the optimal number of EMDR sessions for this population has yet to be empirically determined. It is expected that such persons generally require 24 or more sessions for adequate treatment response (Hyer & Sohnle, 2001).

The Revised Gold Standard Scale

We recommend that the three additional methodological standards be added to the GS Scale to create the Revised Gold Standard (RGS) Scale. The RGS Scale provides a way to quantify methodology and to examine its influence in treatment outcome studies. The reliability and validity of the scale require assessment. It is recommended that the RGS Scale be applied to examine the relationship between methodology and outcome for other disorders and treatments. It is expected that this relationship is not idiosyncratic to EMDR, and that similar results will be found for other effective treatments. In this investigation of the EMDR PTSD studies, we determined that four GSs were strongly related to treatment outcome. Future research can determine if these standards remain the largest predictors of outcome, or if these vary across treatments and/or disorders. The role and influence of specific standards can be assessed.

The RGS Scale may not be complete and additional items may be required to ensure that this is a comprehensive measure of methodology in treatment outcome studies. Other possible items include patient factors such as comorbidity, symptom severity, and multiple/single traumas.

Summary

This study of methodological features suggests that methodological rigor influences outcome, and that meticulous attention to detail can result in more clearly defined outcomes. The elimination of noise and decrease in error measurement appear to result in more precise evaluation of treatment outcome. In the EMDR PTSD treatment outcome studies, methodological standards were highly correlated with ES indicating that research with better methodology was more likely to reveal true effects. It should be noted that the association between methodology and outcome is purely correlational, and may actually be the effect of some unknown third variable. However, it can be argued that, when considering the aggregate evidence for the efficacy of EMDR, greater weight may be given to those studies with better methodology. Such studies found EMDR to be an efficacious treatment for PTSD.

Some aspects of the effectiveness and efficacy debate regarding psychotherapy outcomes hinge on the components of the standards outlined here. Perhaps future outcome studies can identify methodological features in a matrix where standards are rated according to the Revised GS Scale so that the reader can assess the value of the internal and external validation of the study. In time, given standards may even be assigned weights as estimates of their value in assessing outcome. The results of this study are at least suggestive of this possibility.

The current analyses had several limitations. There were only 12 studies in the sample, resulting in low power. However, this suggests that significant results represented consistent trends across the studies. Small samples are very sensitive to individual features and the findings may be sample specific and may not generalize. Nevertheless, these studies represent a wide range of participant populations, a wide range of researchers, numerous types of settings, and a variety of control conditions. Future research is needed to determine if the correlation between methodology and outcome is unique to this particular set of studies, or to EMDR or PTSD research, or if the correlation represents a robust relationship.

References

- Boudewyns, P.A., & Hyer, L.A. (1996). Eye movement desensitization and reprocessing (EMDR) as treatment for post-traumatic stress disorder (PTSD). *Clinical Psychology and Psychotherapy*, 3 (3), 185–195.
- Boudewyns, P.A., Stwertka, S.A., Hyer, L.A., Albrecht, J.W., & Sperr, E. V. (1993). Eye movement desensitization and reprocessing: A pilot study. *Behavior Therapist*, 16, 30–33.
- Carlson, J.G., Chemtob, C.M., Rusnak, K., Hedlund, N.L., & Muraoka, M.Y. (1998). Eye movement desensitization and reprocessing for combat-related posttraumatic stress disorder. *Journal of Traumatic Stress*, 11, 3–24.
- Devilley, G.J., & Spence, S.H. (1999). The relative efficacy and treatment distress of EMDR and a cognitive behavioral trauma treatment protocol in the amelioration of post traumatic stress disorder. *Journal of Anxiety Disorders*, 13 (1–2), 131–157.
- Devilley, G.J., Spence, S.H., & Rapee, R.M. (1998). Statistical and reliable change with eye movement desensitization and reprocessing: Treating trauma with a veteran population. *Behavior Therapy*, 29, 435–455.
- Elkin, I. (1999). A major dilemma in psychotherapy outcome research: Disentangling therapists from therapies. *Clinical Psychology: Science and Practice*, 6 (1), 10–32.
- Fairbank, J.A., & Keane, T.M. (1982). Flooding for combat-related stress disorders: Assessment of anxiety reduction across traumatic memories. *Behavior Therapy*, 13, 499–510.

- Fensterheim, H. (1996). Eye movement desensitization and reprocessing with complex personality patterns: An integrative therapy. *Journal of Psychotherapy Integration*, 6 (1), 27–38.
- Foa, E.B., & Meadows, E.A. (1997). Psychosocial treatments for posttraumatic stress disorder: A critical review. *Annual Review of Psychology*, 48, 449–480.
- Greenwald, R. (1994). Eye movement desensitization and reprocessing (EMDR): An overview. *Journal of Contemporary Psychotherapy*, 24 (1), 15–34.
- Hollon, S.D. (1999). Allegiance effects in treatment research. A commentary. *Clinical Psychology: Science & Practice*, 6, 107–112.
- Hyer, L., & Brandsma, J.M. (1997). EMDR minus eye movements equals good psychotherapy. *Journal of Traumatic Stress*, 10, 515–522.
- Hyer, L., & Sohnle, S. (2001). *Trauma among older people: Issues and treatment*. Philadelphia: Brunner/Routledge.
- Jensen, J.A. (1994). An investigation of eye movement desensitization and reprocessing (EMDR) as a treatment for posttraumatic stress disorder (PTSD) symptoms of Vietnam combat veterans. *Behavior Therapy*, 25, 311–326.
- Kazdin, A.E. (1994). Methodology, design, and evaluation in psychotherapy research. In A.E. Bergin & S.L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 19–71). New York: John Wiley & Sons.
- Kazdin, A.E., & Bass, D. (1989). Power to detect differences between alternative treatments in comparative psychotherapy outcome research. *Journal of Consulting and Clinical Psychology*, 57 (1), 138–147.
- Lee, C., & Gavriel, H. (1998). Treatment of post-traumatic stress disorder: A comparison of stress inoculation training with prolonged exposure and eye movement desensitization and reprocessing. *Proceedings of the World Congress of Behavioral and Cognitive Therapies, Acapulco*.
- Lee, C., Gavriel, H., Drummond, P., Richards, J., & Greenwald, R. (in press). Treatment of PTSD: Stress inoculation training with prolonged exposure compared to EMDR. *Journal of Clinical Psychology*.
- Marcus, S.V., Marquis, P., & Sakai, C. (1997). Controlled study of treatment of PTSD using EMDR in an HMO setting. *Psychotherapy*, 34, 307–315.
- Peniston, E.G. (1986). EMG biofeedback-assisted desensitization treatment for Vietnam combat veterans post-traumatic stress disorder. *Clinical Biofeedback Health*, 9, 35–41.
- Peniston, E.G., & Kulkosky, P.J. (1991). Alpha-theta brainwave neuro-feedback therapy for Vietnam veterans with combat-related post-traumatic stress disorder. *Medical Psychotherapy*, 4, 47–60.
- Pitman, R.K., Orr, S.P., Altman, B., Longpre, R.E., Poire, R.E., & Macklin, M.L. (1996). Emotional processing during eye movement desensitization and reprocessing therapy of Vietnam veterans with chronic posttraumatic stress disorder. *Comprehensive Psychiatry*, 37 (6), 419–429.
- Rogers, S., Silver, S., Goss, J., Obenchain, J., Willis, A., & Whitney, R. (1999). A single session, controlled group study of flooding and eye movement desensitization and reprocessing in treating posttraumatic stress disorder among Vietnam war veterans: Preliminary data. *Journal of Anxiety Disorders*, 13 (1–2), 119–130.
- Rosen, G.M. (1999). Treatment fidelity and research on eye movement desensitization and reprocessing (EMDR). *Journal of Anxiety Disorders*, 13, 173–184.
- Rothbaum, B.O. (1997). A controlled study of eye movement desensitization and reprocessing in the treatment of posttraumatic stress disorder sexual assault victims. *Bulletin of the Menninger Clinic*, 61, 317–334.
- Scheck, M.M., Schaeffer, J.A., & Gillette, C.S. (1998). Brief psychological intervention with traumatized young women: The efficacy of eye movement desensitization and reprocessing. *Journal of Traumatic Stress*, 11, 25–44.
- Shapiro F. (1989). Efficacy of the eye movement desensitization procedure: A new treatment for post-traumatic stress disorder. *Journal of Traumatic Stress*, 2 (2), 199–223.

- Shapiro, F. (1991). Stray thoughts. *EMDR Network Newsletter*, 1–3.
- Shapiro, F. (1994). Alternative stimuli in the use of EMD(R). *Journal of Behavior Therapy and Experimental Psychiatry*, 25, 89.
- Shapiro, F. (1995). *Eye movement desensitization and reprocessing: Basic principles, protocols and procedures*. New York: Guilford Press.
- Shapiro, F. (1999). EMDR and the anxiety disorders: Clinical and research implications of an integrated psychotherapy treatment. *Journal of Anxiety Disorders*, 13 (1–2), 35–67.
- Van Etten, M.L., & Taylor, S. (1998). Comparative efficacy of treatments for posttraumatic stress disorder: A meta-analysis. *Clinical Psychology & Psychotherapy*, 5, 125–144.
- Vaughan, K., Armstrong, M.S., Gold, R., O'Connor, N., Jenneke, W., & Tarrier, N. (1994). A trial of eye movement desensitization compared to image habituation training and applied muscle relaxation in post-traumatic stress disorder. *Journal of Behavior Therapy and Experimental Psychiatry*, 25 (4), 283–291.
- Vaughan, K., & Tarrier, N. (1992). The use of image habituation training with post-traumatic stress disorder. *British Journal of Psychiatry*, 164, 533–541.
- Wilson, D.L., Silver, S.M., Covi, W.G., & Foster, S. (1996). Eye movement desensitization and reprocessing: effectiveness and autonomic correlates. *Journal of Behavior Therapy and Experimental Psychiatry*, 27 (3), 219–229.
- Wilson, S.A., Becker, L.A., & Tinker, R. H. (1995). Eye movement desensitization and reprocessing (EMDR) treatment for psychologically traumatized individuals. *Journal of Consulting and Clinical Psychology*, 63 (6), 928–937.
- Wilson, S.A., Becker, L.A., & Tinker, R.H. (1997). 15-month follow-up of eye movement desensitization and reprocessing (EMDR) treatment for psychological trauma. *Journal of Consulting and Clinical Psychology*, 65 (6), 1047–1056.