The Effects of Eye Movement Desensitization and Reprocessing (EMDR) on Self-Reported Test Anxiety in College Students

John C. Hampel
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Disability and Equity in Education Commons, and the Educational Assessment, Evaluation, and Research Commons

Recommended Citation
https://scholarworks.wmich.edu/dissertations/1634
THE EFFECTS OF EYE MOVEMENT DESENSITIZATION AND REPROCESSING (EMDR) ON SELF-REPORTED TEST ANXIETY IN COLLEGE STUDENTS

by

John C. Hampel

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Psychology

Western Michigan University
Kalamazoo, Michigan
June 1997
Test anxiety is a common problem among students in western culture due to the importance of academic achievement and the consequences for failure. Many consider test anxiety to be primarily an issue of poor study habits and test readiness. However, some students who appear to possess excellent study habits also appear to experience severe anxiety during tests. A recent meta-analysis of test anxiety research substantiated these claims, finding that test anxiety appeared to be an emotionally-based as opposed to a cognitively-based problem. Despite these findings, the etiologies for test anxiety remain unknown. Similar to nearly all DSM-IV diagnostic categories, test anxiety is a syndrome with no known pathognomonic sign(s) which singularly diagnose the condition. Hence, treatments for test anxiety, as for nearly all other DSM-IV mental disorders are symptomatic as opposed to strategic. Unfortunately, there are few symptomatic treatments for test anxiety that are both efficient and effective.

Eye movement desensitization and reprocessing (EMDR), which was developed for the symptomatic treatment of Post-traumatic Stress Disorder (PTSD), was chosen to treat the symptoms of test anxiety for the following essential reasons: (a) the reported efficacy of EMDR with PTSD; (b) the similarities between test anxiety and PTSD that include intrusive thoughts, inability to concentrate, behavioral avoidance, and emotional symptomatology; and (c) the need for a brief, effective symptomatic treatment for test anxiety.
Using a waiting control group against which to compare the treatment group and subsequently replicate treatment effects, the results found that EMDR was highly effective for the symptomatic reduction of self-reported test anxiety as measured by all test anxiety scales. Moreover, these results also suggest that measures of study habits and attitudes are also sensitive to enhancement as a result of treatment with EMD/R. Although the current results did not suggest specific mechanism(s) by which EMDR was effective, the pattern of highly effective results across widely different types of test anxiety presentations suggests the actions of an active placebo treatment. It is suggested that future research contrast EMD/R with known active placebo protocols.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
ACKNOWLEDGMENTS

I would like to express my thanks to the following helpful individuals without whom this dissertation would never have made it to press. In no particular order they include: Ms. Mary Slominis, my research assistant, for her accuracy, consistency, and unwavering dedication; Ms. Martha Stout who has supported me and this work in so many ways throughout our close relationship. I think we made it! I owe an unimaginably large thank you to my parents, without whom this whole doctoral engagement would not have been possible ("You got that right, kid!"). To each and every one of my subjects, many of whom became very dear to me, I extend the hope that it worked and is still working as well as most of you self-reported. To my Committee, Drs. Spates, Robertson, and Kaye, my gratitude for brainstorming, encouraging, attending, reading, rereading, and trusting in my abilities and judgment. And finally, and especially, ever am I in the debt of my advisor, Dr. Galen Alessi, for his special wisdom (no placebos here!) without whom I would have never known the sublime complexity of psychology. I will miss our meetings and your exceptional supply of great stories.

John C. Hampel
# TABLE OF CONTENTS

**ACKNOWLEDGMENTS** ........................................................................................................ ii

**LIST OF TABLES** .................................................................................................................. vi

**LIST OF FIGURES** .................................................................................................................. ix

**CHAPTER**

I. INTRODUCTION ..................................................................................................................... 1
   Statement of the Problem ........................................................................................................ 1
   Review of the Pertinent Literature ....................................................................................... 3
   Eye Movement Desensitization and Reprocessing (EMDR) .................................................. 3
      Background, Development, and Procedure ...................................................................... 3
      Initial Enthusiasm and Subsequent Controversy ............................................................... 6
   Initial Controlled EMDR Research ...................................................................................... 7
   EMDR Case Studies .............................................................................................................. 11
   EMDR Controlled Experiments ........................................................................................... 23
   EMDR--Summary of Experimental Research ..................................................................... 50
   Test Anxiety: Theory, Research, and Treatment ................................................................. 59
   Syndromes Vs. Diseases; Symptoms Vs. Signs ................................................................. 78
   Confusing Effective Treatment With Identifying the Cause ............................................... 84
   Post-Traumatic Stress Disorder ........................................................................................... 91
   DSM-III-R Diagnostic Criteria: Signs and Symptoms ..................................................... 91
   Overlap Between Diagnostic Categories .......................................................................... 93
   Diagnosis by Client Self-report ......................................................................................... 96
   Statistical Vs. Clinical Significance ................................................................................... 99
   Purpose of the Current Study ............................................................................................. 104

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table of Contents--Continued

CHAPTER

Hypotheses Tested by the Current Study........................................ 107
II. METHOD ............................................................................................... 108

Subjects............................................................................................ 108
Design............................................................................................ 108
Statistics .......................................................................................... 112
Clinical Significance......................................................................... 114
Procedures.................................................................................... 117

Setting..................................................................................... 117
Apparatus ............................................................................... 117
Measures.................................................................................. 117

Independent Variables and Assessment Instruments.... 117
Dependent Variables and Assessment Instruments...... 119

Procedure................................................................................ 122

III. RESULTS............................................................................................... 124

Preliminary Analysis ....................................................................... 124
Reliability .......................................................................................... 124
Validity ................................................................................................. 126
Design Integrity: Group Equivalence ................................... 129
Delay to Treatment and Treatment Length ............................ 129
Primary Analysis ............................................................................... 139

Question 1. .................................................................................. 140
Question 2. .................................................................................. 158
Question 3. .................................................................................. 169
# Table of Contents—Continued

## CHAPTER

Secondary Analysis ................................................................. 173

IV. DISCUSSION .............................................................................. 185

Comparison With Previous Research ............................................. 185

   Consideration of the Main Hypotheses ........................................ 185

   Consideration of the Outcome Measures ......................................... 187

   Consideration of the Ways This Study Differs From Previous Research ........ 190

   Consideration of the Effective Components of EMDR ....................... 195

   Limitations of the Current Study ................................................ 202

Conclusion .................................................................................. 212

## APPENDICES

A. EMDR Treatment Protocol Used in the Current Study ......................... 214

B. Clinical Significance Formulas .................................................. 220

C. Protocol Clearance From Human Subjects Institutional Review Board ......................................................... 222

D. Consent Forms for Participation in the Current Study ...................... 225

E. Signed Consent From Irwin Sarason, Ph.D. to Use Reactions to Tests Assessment Device in the Current Study ........................................ 230

BIBLIOGRAPHY ............................................................................. 232
LIST OF TABLES

1. Shared Symptomatology Among Post-traumatic Stress Disorder (PTSD), Simple Phobias, Obsessive-Compulsive Disorder (OCD), Generalized Anxiety Disorder (GAD), Organic Mental Disorder, and Major Depression .................................................. 95

2. Dependent Variables and Their Acronyms .................................................................. 125

3. Correlation Analyses: A Comparison of Initial Screening and Post-Wait Correlation Coefficients in Group B With Published Reliability Coefficients ................................................................. 127

4. Correlation Analyses Among all Dependent Measures Pooled Across Groups A & B at Initial Screening .................................................................................. 128

5. Mean Differences Between Initial Screening Assessments for Groups A and B: Descriptive and Inferential Statistics ........................................................................................................... 130

6. Delay to Treatment and Length of Treatment for Subjects in Groups A and B .......................................................... 135

7. Groups A and B Delay to Treatment and Length of Treatment: Descriptive Statistics .......................................................... 136

8. EMDR Sessions Administered to Groups A and B: Descriptive Statistics .................. 137

9. Number of EMDR Sessions Administered to Subjects in Groups A & B ......................... 138

10. Analysis of Variance Summary Table for Two Factor Repeated Measures ANOVA on TAI-T ........................................................................ 142

11. Descriptive Data and Two-Sample t-tests on TAI-T, TAI-E, TAI-W, and RTT-T: Mean Differences Between Group A Post-Treatment and Group B Post-Wait ................................................... 144

12. Data Used for Determining Reliable Change Indices: Outcomes in Groups A and B on TAI-T .......................................................... 145

13. Clinical Significance (CS) Determination of TAI-T Outcomes: Group A Subjects (EMDR Treatment) and Group B Subjects (No-Treatment Waiting Control) ......................................................... 146

14. Analysis of Variance Summary Table for Two Factor Repeated Measures ANOVA on TAI-E .......................................................... 148
List of Tables—Continued

<table>
<thead>
<tr>
<th>Table No.</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>Data Used for Determining Reliable Change Indices: Outcomes in Groups A and B on TAI-E</td>
<td>149</td>
</tr>
<tr>
<td>16</td>
<td>Clinical Significance (CS) Determination of TAI-E Outcomes: Group A Subjects (EMDR Treatment) and Group B Subjects (No-Treatment Waiting Control)</td>
<td>150</td>
</tr>
<tr>
<td>17</td>
<td>Analysis of Variance Summary Table for Two Factor Repeated Measures ANOVA on TAI-W</td>
<td>153</td>
</tr>
<tr>
<td>18</td>
<td>Data Used for Determining Reliable Change Indices: Outcomes in Groups A and B on TAI-W</td>
<td>154</td>
</tr>
<tr>
<td>19</td>
<td>Clinical Significance (CS) Determination of TAI-W Outcomes: Group A Subjects (EMDR Treatment) and Group B Subjects (No-Treatment Waiting Control)</td>
<td>155</td>
</tr>
<tr>
<td>20</td>
<td>Analysis of Variance Summary Table for Two Factor Repeated Measures ANOVA on RTT-T</td>
<td>157</td>
</tr>
<tr>
<td>21</td>
<td>Clinical Significance (CS) Determination of RTT-T Outcomes: Group A Subjects (EMDR Treatment) and Group B Subjects (No-Treatment Waiting Control)</td>
<td>159</td>
</tr>
<tr>
<td>22</td>
<td>Analysis of Variance Summary Table for One Factor Repeated Measures ANOVA on TAI-T in Group B: Initial Screening, Post-Wait, Post-Treatment</td>
<td>160</td>
</tr>
<tr>
<td>23</td>
<td>Descriptive Data and Paired t-tests on TAI-T, TAI-E, TAI-W, and RTT-T: Mean Differences Between Group B Post-Wait and Post-Treatment</td>
<td>160</td>
</tr>
<tr>
<td>24</td>
<td>Descriptive Data and Two-Sample t-tests on TAI-T, TAI-E, TAI-W, and RTT-T: Mean Differences Between Group A Post-Treatment and Group B Post-Treatment</td>
<td>161</td>
</tr>
<tr>
<td>25</td>
<td>Clinical Significance (CS) Determination of TAI-T Outcomes: Group A Subjects (EMDR Treatment) and Group B Subjects (EMDR Treatment)</td>
<td>162</td>
</tr>
<tr>
<td>26</td>
<td>Analysis of Variance Summary Table for One Factor Repeated Measures ANOVA on TAI-E in Group B: Initial Screening, Post-Wait, Post-Treatment</td>
<td>164</td>
</tr>
<tr>
<td>27</td>
<td>Clinical Significance (CS) Determination of TAI-E Outcomes: Group A Subjects (EMDR Treatment) and Group B Subjects (EMDR Treatment)</td>
<td>165</td>
</tr>
</tbody>
</table>
List of Tables--Continued

28. Analysis of Variance Summary Table for One Factor Repeated Measures ANOVA on TAI-W in Group B: Initial Screening, Post-Wait, Post-Treatment ............................................. 166

29. Clinical Significance (CS) Determination of TAI-W Outcomes: Group A Subjects (EMDR Treatment) and Group B Subjects (EMDR Treatment) ........................................................ 168

30. Analysis of Variance Summary Table for One Factor Repeated Measures ANOVA on RTT-T in Group B: Initial Screening, Post-Wait, Post-Treatment ............................................. 169

31. Clinical Significance (CS) Determination of RTT-T Outcomes: Group A Subjects (EMDR Treatment) and Group B Subjects (EMDR Treatment) ........................................................ 170

32. Descriptive Data and Paired t-tests on TAI-T, TAI-E, TAI-W, and RTT-T: Mean Differences Between Group B Initial Screening and Post-Wait ........................................................ 171

33. Descriptive Data and Two-Sample t-tests on SSHA-DA, SSHA-WM, SSHA-SH, SSHA-TA, SSHA-EA, SSHA-SA, SSHA-SO, and STAI-T: Mean Differences Between Group A Post-Treatment and Group B Post-Wait ........................................................ 183

34. Summary of Statistical Significance (SS) and Clinical Significance (CS) Analyses ........................................................ 186

35. Comparison of Current Results With Previous TA Research Using the TAI ........................................................................ 189

36. Reduction in TAI-T Score and Corresponding Length of Treatment for Subjects in Groups A and B ..................................................... 208

37. Reduction in TAI-T Score and Corresponding Length of the No-Treatment Control Condition for Subjects in Group B ..................................................... 209
LIST OF FIGURES

1. Frequency Distribution of Group A Subjects on Days Following Initial Screening Assessment Prior to First Treatment Session .......... 131

2. Frequency Distribution of Group A Subjects on Total Days From First to Last Treatment Session ......................................................... 132

3. Frequency Distribution of Group B Subjects on Days Following Initial Screening Assessment Prior to First Treatment Session .......... 133

4. Frequency Distribution of Group B Subjects on Total Days From First to Last Treatment Session ......................................................... 134

5. A Comparison of the Effects of EMDR on the TAI-T in Group A With the Effects of Waiting on the TAI-T in Group B With the Effects of EMDR on the TAI-T in Group B ........................................ 141

6. A Comparison of the Effects of EMDR on the TAI-E in Group A With the Effects of Waiting on the TAI-E in Group B With the Effects of EMDR on the TAI-E in Group B ........................................ 147

7. A Comparison of the Effects of EMDR on the TAI-W in Group A With the Effects of Waiting on the TAI-W in Group B With the Effects of EMDR on the TAI-W in Group B ........................................ 151

8. A Comparison of the Effects of EMDR on the RTT-T in Group A With the Effects of Waiting on the RTT-T in Group B With the Effects of EMDR on the RTT-T in Group B ........................................ 156

9. A Comparison of the Effects of EMDR on the SSHA-DA in Group A With the Effects of Waiting on the SSHA-DA in Group B With the Effects of EMDR on the SSHA-DA in Group B ........................................ 175

10. A Comparison of the Effects of EMDR on the SSHA-WM in Group A With the Effects of Waiting on the SSHA-WM in Group B With the Effects of EMDR on the SSHA-WM in Group B ........................................ 176

11. A Comparison of the Effects of EMDR on the SSHA-SH in Group A With the Effects of Waiting on the SSHA-SH in Group B With the Effects of EMDR on the SSHA-SH in Group B ........................................ 177

12. A Comparison of the Effects of EMDR on the SSHA-TA in Group A With the Effects of Waiting on the SSHA-TA in Group B With the Effects of EMDR on the SSHA-TA in Group B ........................................ 178
List of Figures--Continued

13. A Comparison of the Effects of EMDR on the SSHA-EA in Group A With the Effects of Waiting on the SSHA-EA in Group B With the Effects of EMDR on the SSHA-EA in Group B ................................................ 179

14. A Comparison of the Effects of EMDR on the SSHA-SA in Group A With the Effects of Waiting on the SSHA-SA in Group B With the Effects of EMDR on the SSHA-SA in Group B ................................................ 180

15. A Comparison of the Effects of EMDR on the SSHA-SO in Group A With the Effects of Waiting on the SSHA-SO in Group B With the Effects of EMDR on the SSHA-SO in Group B ................................................ 181

16. A Comparison of the Effects of EMDR on the STAI-T in Group A With the Effects of Waiting on the STAI-T in Group B With the Effects of EMDR on the STAI-T in Group B ................................................ 184
CHAPTER I
INTRODUCTION

Statement of the Problem

Performance diminishing test anxiety (hereafter referred to as test anxiety or TA) is a common problem in this culture due to the emphasis that our education system places on performance evaluation through testing and the subsequent consequences based on test performance (Gonzalez, 1995; Spielberger & Vagg, 1995). Test anxiety is commonly believed to consist of at least two elements: (1) intrusive thoughts; and (2) autonomic arousal, although the exact nature of TA and how (and even if) it affects academic (or intellectual) performance has been debated through 40+ years of inconclusive research (Kirkland & Hollandsworth, 1979; Sarason & Sarason, 1990). Hembree (1988) appeared to have established, through an exhaustive meta-analysis, that TA does indeed diminish academic performance. Furthermore, contrary to the popular belief that intrusive cognitions are the essential problem in test anxiety, Hembree (1988) found that emotional or autonomic arousal is the element of test anxiety primarily responsible for academic performance decrements. However, subsequent findings reported by Spielberger & Vagg (1995) and their colleagues have contradicted Hembree's (1988) results, suggesting that cognitive factors are predominant in the TA condition and related performance decrements.

Generally, there is agreement among authors that diminishing the negative impact of TA could greatly enhance the overall quality of education for a substantial number of students (Allen, 1972; Gonzalez, 1995; Hembree, 1988; Spielberger & Vagg, 1995; Tryon, 1980). Hill (1984) estimated that there were over 10 million
students at precollege levels whose test performances suffered significantly as a result of excessive TA. Estimations of the prevalence of test anxiety in college students have been as high as 15% (Hill & Wigfield, 1984). TA often appears to be a dramatic force in the lives of those who experience it:

As the day approached when Ann was to take the final examination in geography, she began to get very nervous. She began to eat poorly and she even started smoking again, a habit which she was seriously trying to break. The night before the exam, Ann's sleep was disturbed many times and she found herself tossing and turning and talking to herself. "I can't remember a thing I studied. Now, what is the average rainfall in northern China? I don't even know all the countries in Africa. I just know I'm going to fail this final. Then my friends will think I'm stupid and I won't be able to become a teacher. Then I won't be able to repay my loans because I won't have a job. Then the bill collectors will pound on my door and they'll probably throw me in prison. Oh, what am I going to do? My life is doomed. My life is over." (Test Anxiety Reduction--Curriculum Guide, 1987, p.2)

TA has been noted by many to be unpleasant and to affect an individual's personal and professional growth and performance (Sarason & Sarason, 1990). For instance, if TA pervasively diminished an individual's academic test performance, that individual's abilities, aptitudes, academic progress, and IQ might be consistently underrated and misinterpreted (Hembree, 1988; Paris, Lawton, Turner & Roth, 1991). It is plausible that this could lead to damaging and inappropriate placements or diagnoses in school settings (e.g., emotionally impaired or learning disabled), the need for individual psychological treatment (Spielberger & Vagg, 1995; Topp, 1989), withdrawal from school, the failure to attain deserved levels of recognition and accomplishment (e.g., college graduates with high GPAs who score poorly on the GRE or GMAT), or to pursuing vocations that do not fully challenge the person's intellectual abilities. The drop in prevalence of TA between high school and college populations suggests that TA might contribute to the failure of many high school graduates to pursue higher levels of education and more satisfying and lucrative careers. Clearly, more conclusive research and consistently efficacious treatments are
needed for the prevention and reduction of TA in students at all levels (Hembree, 1988; Sarason & Sarason, 1990; Tyron, 1980). Toward this end, the current study examined the effectiveness of eye movement desensitization and reprocessing (EMDR), a brief psychosocial treatment that has shown promise in the treatment of PTSD symptomatology on self-reported test anxiety in a population of college students.

Review of the Pertinent Literature

Eye Movement Desensitization and Reprocessing (EMDR)

Background, Development, and Procedure

Eye movement desensitization and reprocessing (EMDR) is a novel, symptomatic, psychosocial treatment developed primarily for the purpose of abating clinical symptomatology related to traumatic memory events. Accordingly, the primary diagnostic category for which EMDR was originally intended was Post-traumatic Stress Disorder (PTSD). Initially called Eye Movement Desensitization (EMD), this treatment, fashioned in the tradition of effective, short-term, behavioral therapies (Herbert & Mueser, 1992), was conceived and developed by Francene Shapiro (1989). Shapiro (1995) has explained how she serendipitously observed the private events that lead her to develop EMD:

While walking through the park one day, I noticed that some disturbing thoughts I was having suddenly disappeared. I also noticed that when I brought these thoughts back to mind, they were not as upsetting or as valid as before. Previous experience had taught me that disturbing thoughts generally have a certain "loop" to them; that is, they tend to play themselves over and over until you consciously do something to stop or change them. What caught my attention that day was that my disturbing thoughts were disappearing and changing without any conscious effort.

Fascinated, I started paying very close attention to what was going on. I noticed that when disturbing thoughts came into my mind, my eyes spontaneously started moving very rapidly back and forth in an upward diagonal. Again the thoughts disappeared, and when I brought them back to
mind, their negative charge was greatly reduced. At that point I started making the eye movements deliberately while concentrating on a variety of disturbing thoughts and memories, and I found that these thoughts also disappeared and lost their charge. My excitement grew as I began to see the potential benefits of this effect. (p. 2)

Shapiro (1995) has reported that she developed and refined the initial EMD protocol through clinical trials with 70 individuals with varying diagnoses. Self-reported anxiety reduction was her primary goal. Hence she considered EMD primarily a behavioral treatment. In 1990 Shapiro (1995) renamed the EMD method Eye Movement Desensitization and Reprocessing (EMDR). She reportedly based this change on the belief that the processes activated by the eye movement procedure appeared to involve not only anxiety reduction but concurrent cognitive restructuring of relevant material. Patients not only seemed to feel more detached or less reactive toward previously upsetting memory events, they also spoke more rationally or adaptively regarding personal attributions vis-a-vis the original traumatic occurrences. Shapiro (1995) believed that the change from EMD to EMDR was highly significant, as she reports:

More than a name change, this was a shift in paradigm and perspective that would take EMDR far beyond its original clinical conceptualization as a treatment for PTSD. The continued refinement of the principles, protocols, and procedures that make up EMDR came to clearly define it as a methodology for a new approach to psychotherapy. (p. 10)

In a recent review of the literature Lohr et al. (1995) questioned the empirical evidence substantiating the addition of the word "Reprocessing" (the R of EMDR). However, these authors have justified the use of the term EMDR (as opposed to EMD) to refer to all eye movement treatments for four reasons (Lohr et al., 1995). These included: (1) EMDR is the moniker assigned to the treatment by its progenitor, F. Shapiro; (2) EMDR is the copyrighted name for Shapiro's organization; (3) EMDR is the name of the protocol which Shapiro describes and teaches; and (4) the term EMDR has become common and widely used. Hence, regardless of when a study or
discussion involving EMD or EMDR was published, henceforth in the current text the eye movement treatment will be referred to as EMDR.

The basics of the EMDR procedure will be described briefly here. For an authoritative, detailed account of the EMDR protocol refer to Shapiro's (1995) book, *Eye Movement Desensitization and Reprocessing: Basic Principles, Protocols, and Procedures*. Also, the basic steps of the EMDR protocol are described in Appendix A of the current text.

The essential component of EMDR involves holding “in mind” a target image, thought, or feeling associated with a disturbing memory event while simultaneously performing a series of large-magnitude, saccadic eye movements (Shapiro, 1991a, 1991b). The number and size of the saccades may vary, although Shapiro (1991a, 1991b) suggests starting with approximately 24 per set and maximizing the patient’s actual eye travel. Sets of approximately 24 saccades are repeated until two goals are achieved: (1) The patient reports that all uncomfortable or disturbing sensations associated with the target memory have been alleviated or maximally reduced; and, (2) the accompanying positive attributions describing the traumatic memories are maximally believed. To assess progress toward these two goals within session, two, self-reported measures are employed: A subjective units of distress scale (SUD), measured on a continuum from 0-10, which indicates the degree to which a memory event is disturbing; and a validity of cognition scale (VoC), measured on a continuum of 0-7, which indicates the extent to which healthier, more adaptive attributions and self-statements regarding the traumatizing event (e.g., “It’s over, it’s in the past, I’m safe now”) are subjectively believed. When all memories associated with the traumatizing event are desensitized thusly (i.e., when SUD = near 0 and VoC = near 7), EMDR treatment has been completed successfully.
Regarding eye movements themselves, there is no set pattern: They may be performed horizontally or diagonally. To achieve eye movements, generally, the therapist moves her outstretched arm and two fingers in a back and forth pattern in front of the patient's face. The patient tracks the therapist's fingers with his eyes thereby creating the saccadic eye movements. Other methods have been used to generate eye movements. These include looking alternately at two different points in space (e.g., two stationary lights or the therapist's, stationary outstretched arms) or following a moving rod or pointer manipulated by the therapist (Shapiro, 1991a, 1991b).

**Initial Enthusiasm and Subsequent Controversy**

Early published reports have suggested that EMDR is often effective for alleviating disturbing memories and their related symptomatology in a single, 90-minute session (Shapiro, 1989a; 1989b). Reports of this nature have generated substantial interest in and implementation of EMDR with a variety of populations (Boudewyns et al., 1993; Lohr et al., 1992, 1995; Lohr, Tolin, Kleinknecht, 1996). Perhaps as a result of such promising reports dissemination of the treatment has proceeded well in advance of controlled scientific studies substantiating its clinical efficacy (Acierno et al., 1994; Lohr et al., 1992; Herbert & Mueser, 1992).

The dyssynchrony between empirical validation and widespread implementation has generated substantial controversy, which does not appear to have been resolved (Baer et al., 1992; Lohr, Kleinknecht, Tolin, & Barrett, 1995; Montgomery, 1996; Shapiro, 1996). Many have cautioned practitioners regarding the use of EMDR, stating that dissemination of the procedure should await valid demonstrations of the treatment's empirical efficacy (Acierno et al., 1994; Herbert & Mueser, 1992; Lohr et al., 1992; Lohr, Kleinknecht, Tolin, & Barrett, 1995). These authors have urged practitioners to use more standard, proven treatments for PTSD and other anxiety disorders such as
exposure (flooding) and stress inoculation training (Cooper & Clum, 1989; Foa, Rothbaum, Riggs, & Murdock, 1991). Nevertheless, EMDR appears to have gained substantial popularity (Acierno et al., 1994; Lohr, Kleinknecht, Tolin, & Barrett, 1995), and the number of disorders for which it is reportedly effective has also increased (Lohr, Kleinknecht, Tolin, & Barrett, 1995). As Friedman (1996) has summarized:

Well-controlled empirical support for EMDR is lacking, the few completed controlled studies have been equivocal, and methodological questions have been raised...What’s remarkable, however, is that a number of seasoned PTSD clinicians are convinced that EMDR is the most effective available treatment for PTSD despite the fact that many others are highly skeptical of this approach. (p. 185)

Initial Controlled EMDR Research

Shapiro (1989a) conducted the first study using EMDR, which was also the first controlled investigation. Shapiro (1989a) randomly assigned 22 subjects to one of two groups, EMD treatment and a “placebo” control condition. The criteria for inclusion were self-reported traumatic memories and various symptoms of PTSD that had endured for a minimum of one year. Most of the subjects were concurrently undergoing some other form of therapy; the average time in other forms of treatment was 6 years. Most of the subjects had been diagnosed with PTSD; they included rape victims, molestation victims, and combat veterans. Most of the subjects had been referred by another therapist or counseling agency to the Shapiro (1989a) study. Five of the 22 were mental health counselors who expressed having chronic difficulty with specific traumatic memory events.

There were three primary differences between the placebo condition and EMDR treatment (Shapiro, 1989a). First, during EMDR treatment saccadic eye movements were administered during recollection of the traumatic events. By contrast, during the
placebo control condition subjects were instructed simply to focus on the traumatic memories. Second, during EMDR treatment eye movements were terminated contingent on reports of reduced SUD and increased VoC levels; compared with the placebo control condition wherein there were no contingencies between changes in self-reported SUD and VoC levels and termination of the instructions to continue imaginal exposure. And, third, heart rate was measured during the EMDR treatment condition and not during the placebo treatment condition. All other factors were designed to be identical. Both groups were administered their respective treatments for one session. The EMDR session lengths varied from 15 to 90 minutes (based on individual reports). The placebo session lengths were standardized at 50 minutes. Whereas the “placebo” treatment actually featured an active element (i.e., exposure), Shapiro (1989a) stated her belief that one, 50-minute exposure session was very likely insufficient to produce therapeutic effects.

Dependent measures included the SUD and VoC scales (Shapiro, 1989a). The results were as follows: The SUD level for the EMDR treatment group dropped from 7.45 (pre) to 0.13 (post). The SUD level for the placebo control group increased from 6.77 (pre) to 8.31 (post). The VoC for the EMDR treatment group rose from 3.95 (pre) to 6.75 (post). The VoC for the placebo control group dropped from 2.95 (pre) to 2.36 (post). Due to the apparent worsening of the placebo control group as a result of the placebo condition, EMDR treatment was administered to this group immediately following their participation in the placebo control condition. The resulting changes in SUD and VoC measures replicated closely those for the primary EMDR treatment group. SUD and VoC measures at 1- and 3-month follow ups for both the primary treatment and control groups (subsequently treated with EMDR) were virtually unchanged from the post treatment assessments (Shapiro, 1989a). In summary, the
Shapiro (1989a) study found that EMDR was effective for reducing self-reported symptoms related to traumatic memories, and the placebo control condition was not. The Shapiro (1989a) controlled study has been cited for a number of methodological limitations and procedural confounds (Acierno et al., 1994; Herbert & Mueser, 1992; Lohr et al., 1992). The initial diagnostic procedure did not verify or validate the PTSD diagnoses among all subjects. Rather, the PTSD condition was assumed or implicit among subjects. As Lohr et al. (1995) has indicated, without clear specifications of the disorder under study, any conclusions regarding the efficacy of treatment to remediate the symptoms of that disorder must be guarded. Conclusions regarding the clinical efficacy of EMDR are also circumspect because Shapiro (1989a) failed to employ standardized assessment instruments to evaluate subjects pre- and post-treatment (Herbert & Mueser, 1992; Lohr et al., 1992). The SUD and VoC ratings are highly subjective, phasic, self-report measures having no established relationships with the symptoms of any known mental disorder (Herbert & Mueser, 1992; Lohr et al., 1992). As Lohr et al. (1992) reported, "there are no empirical data to indicate that irrational beliefs are directly associated with PTSD per se." (p. 163). Moreover, Shapiro (1989a) did not assess SUD and VoC ratings objectively; the study was not double-blinded. F. Shapiro was both the primary therapist and the primary data collector. Hence, SUD and VoC ratings might reflect unintentional experimenter biases which might have been communicated either verbally or non-verbally (1989a).

Another concern with the Shapiro (1989a) investigation has been the potential sensitivity of SUD and VoC measures to demand variables (Acierno et al., 1994; Herbert & Mueser, 1992; Lohr et al., 1992). That is, SUD and VoC reports might have represented "impure" tacts, behavior partially under the control of environmental consequences and partially under the control of the subjects' perceived internal states (c.f., Skinner, 1957). One potentially influential environmental consequence was the
escape from or avoidance of eye movements, which, in Shapiro’s (1989a) EMDR treatment condition was made contingent on improved SUD and VoC reports (as previously described in this section). Hence, changes in subjects’ SUD and VoC reports might have been the result of a negative reinforcement process. Such environmental contamination of SUD and VoC reports cannot be ruled out especially given the temporal nature of SUD measures and the aversive nature of large magnitude, saccadic eye movements for some individuals (Acierno et al., 1994; Metter & Michelson, 1993). Lohr et al. (1992) referred to the escape/avoidance contingency built into Shapiro’s (1989a) EMDR protocol as a possible non-specific effect of EMDR and the “most compelling difference in the treatment procedures” (p. 164).

Other contaminating demand effects might have resulted from the same therapist-researcher (F. Shapiro— also EMDR’s progenitor) administering both EMDR and placebo treatment protocols (Acierno et al., 1994; Lohr et al., 1992; Herbert and Mueser, 1992). Using the same therapist to implement both procedures could have introduced unspecified performance demands favoring EMDR. Given her potential bias in favor of EMDR over placebo control, Shapiro (1989a) might have inadvertently confounded the relative levels of believability of both approaches, which was largely unassessed and uncontrolled (Lohr et al., 1992). Moreover, the two treatments’ relative believability might also have been confounded by the procedural assessment of heart rate in the EMDR condition and not in the placebo condition (Herbert and Mueser, 1992).

Another confound between the EMDR and placebo control groups might have resulted from the differential, non-specific effects of exposure. Although Shapiro (1989a) attempted to equate the number of presentations of traumatic material that the two groups received, there was no attempt to balance the resulting total length of exposure (Lohr et al., 1992).
Finally, as Lohr et al. (1992) have pointed out, the internal validity of the Shapiro (1989a) experimental design was compromised by the lack of a no-treatment, waiting control condition with which to compare the effects observed within the EMDR treatment group. Hence, the therapeutic effects observed cannot be attributed solely to EMDR as the effects of simple time passage or merely enrolling in a novel treatment protocol cannot be ruled out.

**EMDR Case Studies**

To date the majority of published EMDR investigations have involved poorly controlled, non-experimental case studies (Acierno et al., 1994; Herbert and Mueser, 1992; Lohr et al., 1992; Lohr et al., 1995; Spates & Burnette, 1995). Beyond the inherent limitations of the case study format, most of this work has suffered from many of the same methodological shortcomings that weakened the initial Shapiro (1989a) controlled study. As discussed previously in the Initial Controlled EMDR Research section of the current text, these shortcomings included poor diagnostic validity, the lack of standardized, objective measures assessed pre- and post-treatment, potentially confounding demand effects, and potentially confounding experimenter bias (Acierno et al., 1994; Herbert and Mueser, 1992; Lohr et al., 1992; Lohr et al., 1995). In addition to these difficulties, many of the pre-1993 EMDR case studies severely compromised the integrity of the EMDR protocol as they employed a variety of additional clinical procedures in separate sessions either concurrent with or proximal to the EMDR sessions (Acierno et al., 1994). These procedures included relaxation, cognitive restructuring, breathing training, and in vivo exposure. Consequently, most of the early EMDR studies achieved relatively low levels of internal or empirical validity (Acierno et al., 1994; Kazdin, 1981, 1992).
However, regardless of how well it is designed and conducted, the case study simply cannot function in the capacity of more tightly controlled, experimental research. As Lohr et al. (1995) have stated:

While case reports can provide guidance in the clinical application of EMDR to diverse clinical phenomena, they are inadequate in controlling for procedural confounds. Valid causal inferences can only be made in the context of experimental designs. (pp 287-288)

The current review will examine briefly the majority of EMDR case studies for purposes of explicating the range of clinical problems they have treated and the general results they have reported. In addition, each case report will be evaluated for its major methodological strengths and weaknesses.

Shapiro (1989b) presented the first EMDR case study. The subject was a 63-year-old woman who had reportedly experienced symptoms of PTSD subsequent to a rape 15-months prior to the study. However, no formal diagnostic assessment was conducted. Also, no standardized instruments or objective measures were employed. Dependent measures were SUD and VoC scales. Three memories associated with the rape were treated with EMDR in one, 50-minute session by the primary experimenter. The subject reported complete amelioration of anxiety related to the trauma immediately following treatment and at 3-month follow-up.

Wolpe and Abrams (1991) reported on the case of a 43-year-old woman with PTSD symptoms associated with a rape that occurred 9 years previous to EMDR treatment. This study lacked valid diagnostic procedures, standardized or objective dependent measures, and was confounded by extensive prior treatment immediately preceding EMDR treatment. Following six EMDR sessions, the subject reported complete symptomatic remission as assessed by the self-reported SUD scale.

Puk (1991) published a report of two, similar, EMDR case studies. One involved a 23-year-old woman who reported childhood sexual abuse. The other
involved a 33-year-old woman who reported experiencing traumatic memories of her dying sister. No formal diagnostic assessment was conducted for either case. The sole dependent measure was a 100-point SUD scale. In both cases a single session of EMDR ameliorated the self-reported symptoms. These results were maintained at 1 year follow-up (abuse) and 6 months follow-up (sister).

EMDR was applied to 100 patients with various clinical diagnoses presumed to be related to traumatic events (Marquis, 1991). In this study the primary experimenter was also the primary therapist. No formal diagnostic assessments were conducted and dependent measures consisted of nonobjective self-reports. It is unclear how many sessions of EMDR were administered to each subject. Many subjects were reportedly receiving some form of primary therapy concurrently with adjunctive EMDR treatment. Twenty-two of the original 100 subjects either refused treatment or dropped out completely. The 78 subjects who completed EMDR treatment obtained a mean improvement of 1.85 points (on a 3-point scale), which was described as significant clinical improvement (no statistical analysis was presented). As a group those diagnosed with PTSD obtained the greatest mean improvement.

Kleinknecht and Morgan (1992) reported a case of EMDR treatment with a 40-year old man presenting with an 8-year history of PTSD related to a nearly mortal gunshot wound sustained during a robbery. It is unclear who the primary therapist for this study was. No formal diagnostic assessment was completed. However, this study used a variety of standardized assessment measures in addition to the usual self-reported SUD scale. The assessment battery was administered at pre-treatment and at 4 and 8 months post-treatment. The presenting traumatic memory, as well as two other traumatic memories that “came up” during EMDR treatment were all desensitized within a single, 90-minute session. All standardized measures improved significantly at both follow-up evaluations.
McCann (1992) reported on the case of EMDR treatment of a 41-year-old man suffering from PTSD stemming from a serious, work related accident eight years prior that had left him severely physically handicapped and totally deaf. In this case EMDR was administered by the primary experimenter. Although there were no standardized assessment instruments employed, a diagnostic interview prior to treatment confirmed the DSM-III-R diagnosis of PTSD. Complete remission of PTSD symptomatology was achieved after one EMDR session as assessed by a self-report SUD scale. The self-reported gains were confirmed at various follow-ups as long as one year post-treatment.

Lipke and Botkin (1992) described EMDR treatment of five hospitalized Vietnam combat veterans with chronic PTSD. Four of the five had been diagnosed with PTSD at the onset of their current hospital admission. One had been diagnosed with PTSD two years earlier. No standardized assessments were used post-treatment. Subjects were evaluated using a self-reported SUD scale. Each subject received one session of EMDR administered by the primary investigator. Overall, four of the five subjects showed improvement, although the exact nature of this improvement was not clearly delineated. Lipke and Botkin (1991) reported:

Four of the five patients showed rapid marked decreases in distress associated with a severely troubling memory and two sustained this feeling of relief. Three of the five patients reported alteration or diminishing of the visual aspect of the memory, an experience they had not previously reported and the authors have not previously witnessed. Two patients appeared to demonstrate new insight, insight which was quite remarkable in the patient who revealed his need to retain uncomfortable memories (pp 593-594).

The authors also described their results as inferior to those obtained by Shapiro (1989a). Finally, no long-term follow-up was conducted in this case study.

EMDR has been administered to treat an adult female suffering from ongoing anxieties related to a traumatic, childhood, hospitalization experience and the subsequent scars with which it left her (Hassard, 1993). The author reports that the
initial SUD of 10 was reduced to zero in a single, 2-hour EMDR session. Anecdotal follow-ups at two weeks and six months suggested that the therapeutic gains had been substantial and permanent. The case study was marred by the lack of diagnostic validity and objective, physiological, or standardized assessment measures. In addition, the primary experimenter was the primary therapist which might have inadvertently biased the outcome.

One case study that involved eight subjects with PTSD or PTSD-like symptoms has found EMDR to produce mixed outcomes (Oswalt, Anderson, Hagstrom, & Berkowitz, 1993). Similar to many early case reports (this one was actually conducted in 1990) this design was poorly controlled across various factors. Diagnostic condition and history varied widely among subjects. Of the eight subjects, five had been hospitalized just prior to EMDR treatment wherein they had been diagnosed with PTSD, and three were college students who had been solicited via newspaper ads. The single diagnostic criterion for this study was a self-reported, intrusive traumatic memory. Evaluation methods included the SUD scale and anecdotal self-reports regarding the intrusive memories; no standardized or objective measures were obtained.

After a single session of EMDR, Oswalt, Anderson, Hagstrom, & Berkowitz (1993) reported generally positive outcomes for the three college students and generally negative outcomes for the five hospitalized patients. Treatment failures involved emotional discomfort with the procedure, extreme dizziness, confusion and disorientation, and a failure to replicate positive results in the second session following a productive first session. In addition to those confounds and shortcomings already discussed, the authors have delineated the following issues: (a) The number of saccads varied substantially among subjects, (b) all subjects were treated in a room wherein two observers were present (to record patient self-reports and insure treatment integrity),
and (c) no long-term follow-up was conducted. The authors encouraged continued outcome research to evaluate the utility of EMDR.

Spector and Huthwaite (1993) administered EMDR in a case involving a horrific car crash that left a 41-year-old woman suffering from a variety of anxieties related to driving. This case did not establish DSM-III-R diagnostic condition prior to treatment. This case used SUD and VoC ratings to assess therapeutic change. No objective or standardized instruments were employed. The primary investigator was also the primary therapist. The authors reported that one, 90-minute EMDR session diminished self-reported SUD ratings from 7 to 1. The accompanying VoC increased from 2 to 6. A single in vivo driving session revealed no overt expression of anxiety from the subject. Anecdotal follow-up data at 6-weeks and 6-months suggested that the treatment was a complete success clinically.

Pellicer (1993) reported the use of EMDR to treat chronic nightmares in a 10-year-old girl previously diagnosed with Mild Mental Retardation. At the outset of the study the girl met the DSM-III-R diagnostic criteria for Dream Anxiety Disorder (nightmare disorder). A 7-day, parentally reported baseline prior to EMDR treatment indicated that the girl emitted behavior indicative of nightmares on five of the seven days. Therapeutic assessment measures included the SUD scale and anecdotal reports from the girl’s parents. No standardized or objective measures were included. A single session of EMDR involving two sets of 20 saccadic eye movements resulted in a SUD score reduction of 10 to 0. The author reported that with each set of eye movements the girl reported that “the scene was becoming progressively unclear and losing its emotional impact” (Pellicer, 1993, p. 74). The girl’s parents reported that all behavioral symptoms (waking up screaming and clinging to mother) vanished immediately following treatment. The parents reported that these gains were maintained at 6-week and 6-month follow-ups.
Forbes, Creamer, and Rycroft (1994) conducted a “pilot study” investigating the efficacy of EMDR in the treatment of PTSD. This study featured better controls than most of the preceding case reports, however it falls short of a true experimental design. These authors established the validity of the PTSD diagnosis in all eight subjects at pre- and post-treatment. A variety of standardized, objective measures and one physiological measure (muscle tension measured via EMG) were employed both pre- and post-treatment. The primary investigators were also the primary therapists and SUD levels were used for within session assessments. The results of this study were mixed. Whereas significant differences were found in all three PTSD symptom categories and for depressive symptoms, four of the eight subjects still met the diagnostic criteria for PTSD at post-treatment and at 3-month follow-up. In accordance with the possible demand effects associated with assessing therapeutic progress during EMDR sessions using SUD self-reports, Forbes, Creamer, and Rycroft (1994) found that SUD levels had generally diminished substantially at post-treatment even when substantial pathology remained as measured by standardized, objective measures.

Muris and Merckelbach (1995) examined the effects of EMDR on two cases of spider phobia. These authors argued that previous EMDR research with phobias might have produced mixed results because EMD and not EMDR had been employed. These cases examined the effects of EMDR, then administered a single session of in vivo exposure. Both subjects met the DSM-IV criteria for specific phobia. Standardized and behavioral (directly observed physical approach) assessments were administered pre and post EMDR treatment as well as post in vivo exposure treatment. Also, SUD and VoC scales were used for assessing within session progress. As with most EMDR research, the primary experimenters were also the primary therapists. The results indicated that EMDR was effective for improving the self-reported SUD, VoC, and standardized phobia questionnaires (subjective indices of phobia). However, EMDR
was not nearly as effective for reducing actual avoidance behavior associated with phobia. In vivo exposure treatment resulted in major improvements in approach behaviors to spiders.

Silver, Brooks, and Obenchain (1995) reported a naturalistic or uncontrolled study wherein they assessed the additive benefits of EMDR, relaxation training, and biofeedback to a Veteran's hospital inpatient PTSD treatment program. These treatments and the experiment were started retroactively vis-a-vis the program. This raised some unique problems. Although an objective, standardized assessment instrument was employed, the original purpose for the assessment was evaluation of the hospital inpatient program (not the experiment), and completion rate of this assessment was erratic. Hence, the N for the experimental pre-treatment assessment was low. It was unclear if the subjects in this study were formally assessed for PTSD, however the authors stated that the pre-existing levels of PTSD varied greatly. Another confound was the number of treatments the subjects in each group received, which varied substantially. Given these weaknesses, these authors found that EMDR was significantly more beneficial (as assessed by the participants' self-reports) when added to the general program milieu than were either relaxation training or biofeedback.

Spates and Burnette (1995) reported three complex and quite different cases of EMDR treatment. Neither formalized diagnostic procedures nor objective, standardized assessment instruments were administered with two of the three cases. For the third case, a standardized, although non-diagnostic assessment was administered pre-and post-treatment. The primary authors functioned as therapists and SUD and VoC self-reports were assessed within sessions. The first case involved a priest with a history of multiple traumas one of which dated back 30 years prior to EMDR treatment. His symptomatology was suggestive of PTSD and panic disorder. He had received various treatments previous to this study and was receiving some form of outpatient therapy.
concurrently with this study. At post-treatment and at 30-day follow-up the subject reported having no significant anxiety. At 2-year follow-up, however, the subject reported some continuing symptoms of panic. The second case involved a 39-year-old woman with a history of childhood sexual abuse. Five sessions of EMDR resulted in resolution of the traumatic memories, which was maintained at 7-month follow-up. The third case involved a 38-year-old male police officer with PTSD-like symptomatology and occupational disability relating to a bullet wound sustained during the pursuit of a felon. One EMDR session resulted in a SUD of zero and significant reductions in the Symptom Checklist 90-R. At 15-month follow-up patient self-report indicated that his symptoms had not interrupted his work and that he had had no relapses.

By virtue of having addressed some of the previous shortcomings, a small number of post-1992 EMDR case studies appear to have achieved somewhat greater empirical or internal validity. In his second review of the EMDR literature (Lohr et al., 1995), Lohr, working with a somewhat different group of investigators, examined the 14 case studies that had been published since the writing of his first literature review (Lohr et al., 1992). Lohr et al. (1995) used guidelines established by Kazdin (1981) for evaluating the validity of case studies. These authors constructed what they termed a "validity profile" for the 14 EMDR case reports. Lohr et al. (1995) concluded that only three of the 14 studies employed the type of systematic assessment and measurement methodologies necessary to achieve "even the minimal procedures and data to support the empirical validity of the clinical intervention" (p. 287).

The first of these involved the case of a 21-year-old woman with long-standing fears of blood and injections (Kleinknecht, 1993). The diagnosis of simple phobia was confirmed in this case. The primary investigator was also the primary therapist. Standardized, physiological, and behavioral (directly observed) measures were
employed in addition to the standard SUD scale. Four, 30-minute, EMDR treatment sessions were administered which resulted in substantially diminished SUD ratings and physiologic measures. These gains were maintained at a 24-week follow-up. Moreover, these gains were observed during an injection (flu shot) and routine blood draw. Hence, unlike any previous case study, standardized, self-report, and behavioral measures cross-validated the positive results observed.

The next case study that Lohr et al. (1995) identified as using procedures sufficient to support the empirical validity of EMDR involved treatment of 10 clinical cases of PTSD (Vaughan, Weiss, Gold, & Tarrier, 1994). Unlike many previous case studies, for this study, the PTSD diagnoses were ascertained pre- and post-treatment. Moreover, objective, standardized measures were used that comprehensively described PTSD symptoms. Subjects in this study received an average of 3.3 EMDR sessions, which lasted from 45 to 60 minutes. Weaknesses of this study, as with most other EMDR case studies, included the primary investigator functioning as the primary therapist and SUD and VoC self-reports were used to assess within-session progress. Also, three subjects in this study were receiving psychotropic medications concurrent with EMDR treatment.

At post-treatment Vaughan, Weiss, Gold, & Tarrier (1994) found that 2 of the 10 cases showed some improvement and 8 of the 10 cases showed substantial improvement. Although they do not state whether or not the subjects still met the diagnostic criteria for PTSD at post-treatment, all three primary DSM-III-R categories of PTSD symptoms and the one depression measure were significantly reduced. At 8 to 12 week follow-ups, however, mean differences in one of the measures of PTSD and the measure of depression were no longer significant. Changes observed in the individual symptoms within the three categories were mixed at both post-treatment and at follow-up.
The third case study that Lohr et al. (1995) identified as employing procedures adequate to demonstrate the empirical validity of EMDR involved the treatment of seven subjects with panic disorder (Goldstein & Feske, 1994). All seven subjects in this study met the criteria for panic disorder according to the Structured Clinical Interview for DSM-III-R (SCID, outpatient version; Spitzer, Williams, & Gibbon, 1989). In addition, five of the subjects carried a premorbid diagnosis of simple phobia, five generalized anxiety disorder, and five agoraphobia. A number of standardized, self-report questionnaires and self-monitoring records were used to evaluate subjects pre- and post-treatment. These assessments were completed one week prior to the beginning of treatment and one week following the conclusion of treatment. No additional follow-up was conducted. Although the authors did not confirm the use of SUD and VoC values within sessions, it is assumed that these were employed to evaluate in vitro therapeutic progress as Goldstein & Feske (1994) employed the EMDR protocol as per Shapiro (1991). Five sessions of EMDR treatment were administered over a two-week period. The second author for the report was both the primary therapist and the diagnostic interviewer; this individual received EMDR training through Shapiro.

Data analysis found significant improvement on all standardized measures from pre- to post-treatment (Goldstein & Feske, 1994). Four of the subjects reported being panic-free at post-test. The others reported large reductions in frequency of panic episodes. Generally, outcomes for the two panic disorder without agoraphobia subjects realized were better than the five panic disorder with agoraphobia subjects. Nonetheless, the majority of subjects did not improve on measures of behavioral avoidance and fear of body sensations. Moreover, the clinical significance of the changes among subjects was somewhat less impressive than the statistical results. Clinical improvement was observed on only 17 of 35 or 49% of opportunities. Clinical
recovery was observed on only 13 of 35 or 37% of opportunities. The authors asserted that therapist engendered expectations for positive outcomes probably did not play a role in this study as, prior to this study, the primary therapist had no experience beyond training with EMDR, and she had expressed considerable skepticism regarding EMDR's utility. The authors also maintained that the generally positive outcomes they observed within the two-week EMDR treatment period were probably not due to the passage of time as a one-month, no-treatment, waiting control condition in a previous study produced virtually no changes in a similar group of subjects with panic disorder (Chambless, Goldstein, Gallagher, & Bright, 1986). The authors concluded that, although they did not consider any of their subjects to be "cured" as a result of five sessions of EMDR, the EMDR procedure appeared to be promising in the treatment of panic disorder.

A recently published case report by Carlson, Rusnak, Chemtob, and Hedlund (1996) also appears to merit special attention by virtue of its design methodology. In this study four Vietnam veterans were treated with EMDR. Unlike previous combat PTSD case studies that administered relatively limited EMDR treatment, this study administered 12, 60-75-minute treatment EMDR sessions. Also, diagnostic validity and therapeutic progress were ascertained via a comprehensive multimodal assessment battery consisting of objective, standardized instruments and various physiological measures. The entire battery was administered at pre-treatment, post-treatment, and at a 3-month follow-up. Plus, to insure the greatest possible efficacy, the integrity of the EMDR protocol was closely monitored and evaluated. Results at post-treatment and follow-up were very similar: The standardized self-report measures indicated that three of the four subjects showed "very substantial improvement and personally viewed the outcome of treatment very positively" (Carlson, Rusnak, Chemtob, and Hedlund,
However, none of the physiological measures showed changes suggestive of the self-reported improvements.

Collectively, the EMDR case reports and pilot studies have not achieved the magnitude of improvement realized in the initial Shapiro (1989a) controlled study (Vaughan et al., 1994). As a group they do suggest that EMDR is often effective for improving various self-report measures associated with a variety of different presenting conditions. However, in most cases, behavioral or physiological measures do not corroborate these results. Without the increased experimental power afforded by direct comparisons with alternative treatments, these results do not support EMDR as the treatment of choice in any given clinical circumstance. Moreover, without the inclusion of active placebo and, or, no-treatment control groups and double-blind methodology, or the use of single-subject experimental methodology, case and pilot studies cannot rule out possible non-specific or regression toward the mean effects. As Acierno et al. (1994) summarized, "Clearly, the utility of EMDR is not supported by these case studies" (p. 293). A number of authors have expressed similar viewpoints regarding selected portions of the EMDR literature (Herbert and Mueser, 1992; Lohr et al., 1992; Lohr et al., 1995). What the EMDR case studies clearly delineate is the need for definitive experimental investigations into the efficacy of the EMDR protocol.

**EMDR Controlled Experiments**

Following the Shapiro (1989a) study, the next experimental investigation using EMDR was published by Sanderson and Carpenter (1992). This study used a crossover design comparing EMDR with an alternative procedure named Image Confrontation (IC). The IC procedure, which was very similar to that used for Shapiro's (1989a) placebo control group, differed from EMDR only in that subjects were told to hold disturbing images in mind with their eyes closed and not moving.
There were 58 phobic subjects in this study. Thirty-one were arachnophobes. However, the authors presented no diagnostic details pertaining to pre- or post-treatment conditions. Subjects’ phobic conditions were rated pre-treatment on the subjective basis of an unspecified clinician. No standardized assessments were employed. Rather, a 100-point SUD scale was used for all evaluations of either feared objects (FO) or feared images (FI). The subjects were divided equally into two groups, one group which received seven sets of EMDR followed by seven sets of IC and a second group which received the same treatments in the reverse order. Following each of the seven treatment sets, a SUD level for the FI was obtained. To evaluate the relative effects of the treatments, SUD levels for a FO were obtained pre-treatment, between treatments, and post-treatment. Following this procedure all subjects were instructed to practice the IC procedure daily at home for one month. At a 1-month follow-up both FO and FI SUD levels were assessed.

The results indicated that both EMDR and IC treatment protocols were effective for significantly reducing SUD levels for the FI and FO (Sanderson and Carpenter, 1992). There was no significant difference between the treatments’ effectiveness. Treatment gains were maintained at the 1-month follow-up. An interesting post-hoc analysis contrasted those phobic subjects whose etiology appeared to be traumatic in nature (e.g., fear of elevators after a frightening escape), hence more similar to a PTSD condition. This seven member sub-group showed somewhat greater SUD reductions than the group as a whole. And though EMDR was more effective than IC for this sub-group, the difference was not significant. This study suffered from numerous methodological flaws including the lack of diagnostic specificity, the lack of objective, standardized dependent measures, and the lack of a behavioral measure of approach to the phobic stimulus. Nevertheless, based on their results Sanderson and Carpenter (1992) concluded: “it is clear from our results that the benefits of EMD, at least in
phobics, bear no relation to eye movements” (pp 274-275). In explaining their results vis-a-vis the Shapiro (1989a) study, they questioned the integrity of Shapiro’s (1989a) placebo control condition.

A pilot study that administered EMDR treatment to PTSD in 20 combat veterans has attempted to address many of the qualifying confounds that weakening the Shapiro (1989a) study (Boudewyns et al., 1993). In this study the patients were relatively homogeneous; moreover, PTSD was assessed via an interdisciplinary treatment team using DSM-III-R criteria. SUD reports, three standardized assessment instruments, and four psychophysiological measures were also administered pre- and post-treatment to evaluate outcomes for all three experimental conditions. No follow-up measures were obtained. Physiological measures were assessed in response to a tape-recorded, “original” description of the traumatic memory. These same recordings were used as the basis for pre- and post- treatment physiological assessments. The 20 subjects were randomly assigned to EMDR treatment (n=9), Exposure Control (n=6), or Control (n=5) conditions. EMDR was as per Shapiro (1991). The three therapists responsible for EMDR treatment were each trained by Shapiro. The Exposure Control condition involved a similar procedure as EMDR sans eye movements. Subjects in the Control condition received the milieu treatment on the Special Inpatient PTSD Unit (SIPU) at the Veterans Administration Hospital where this study was conducted. Two EMDR or Exposure Control sessions were conducted over a 14-day period.

These authors reported that EMDR produced significantly greater reductions in SUD scores compared with the Exposure Control and Control conditions (Boudewyns et al., 1993). However, SUD scores remained unchanged during the post-test assessment wherein the subjects listened to tape-recordings of their original memories. ANOVA analyses found no significant interactions between treatment conditions and
any of the self-reported, standardized assessments or any of the physiological measures.

Regarding their failure to replicate Shapiro's heart-rate reductions in the EMDR treatment group, Boudewyns et al. (1993) have pointed out that Shapiro (1989a) did not assess heart rates for both the EMDR and placebo treatment conditions. Hence, Shapiro (1989a) could not determine the degree to which EMDR improved physiological indices more than the placebo control group and any comparisons with these data might not be valid.

Boudewyns et al. (1993) have suggested that the difference in heart rate reductions observed in the Shapiro (1989a) study and those observed their study might have been due to procedural differences between the two research projects and the possible effects of EMDR. Shapiro (1989a) assessed heart rates within session and in response to private, silent memories that presumably represented the traumatic events. Moreover, if EMDR changes the individual's memory of the traumatic event and thereby also changes cognitive attributions toward the event, then the individual's physiological responses to the altered memory might themselves be altered from those relative to the original memory and cognitive attributions. If these occurrences were true, then Shapiro's (1989a) data are indicative of an EMDR treatment effect. This would also mean that at the pre-treatment assessment Shapiro (1989a) measured the untreated memory and at the post-treatment assessment she measured the EMDR treated memory. By contrast, at both pre- and post-treatment assessments the subjects in the Boudewyns et al. (1993) study were exposed to their personal tape-recorded message that reflected the original, untreated traumatic memory. These original memories might not have been indicative of the subjects' altered, more adaptive perspectives. That is, even after EMDR treatment the original, tape-recorded memories might have evoked strong physiological responses that the covert, EMDR treated memories no longer did.
or could. Moreover, Boudewyns et al. (1993) had reported that, prior to EMDR treatment, the original memory recordings had produced significant changes in all physiological measures.

Boudewyns et al. (1993) also have pointed out that powerful, hidden, monetary agendas might have confounded the results of their study. They stated that veterans seeking disability status have tended to overreport negative symptoms on self-report assessment devices. They posited that self-reported improvements might be threatening to this population for fear that, contingent on such improvements, disability support will be withdrawn, and subsequently they will relapse.

However, Boudewyns et al. (1993) have contended that the most "parsimonious" explanation for their failure to replicate the findings of Shapiro (1989a) and produce significant changes in standardized psychological and physiological measures might involve the actions of nonspecific, placebo variables. According to this explanation, unspecified placebo effects imparted greater, immediate, in-session efficacy to EMDR over the Exposure Control condition. These nonspecific effects were not enduring or potent enough to maintain therapeutic improvements until the post-treatment assessment, which was conducted several days following the second (and final) treatment session.

Jensen (1994) conducted an experiment with 25 combat veterans comparing the effects of EMDR and a no-treatment waiting condition on various self-reported, standardized measures associated with PTSD. In this study PTSD was assessed pre- and post-treatment. Also, in addition to SUD and VoC self-reports, personalized PTSD treatment goals were established, and a number of standardized assessment instruments were administered pre- and post-treatment to evaluate the effectiveness of EMDR treatment. No follow-up assessment was conducted. The therapists for this study were two psychology interns who had completed Shapiro's (1990) level 1
training workshop and practiced using EMDR with other interns. Of the 25 subjects, 13 were randomly assigned to EMDR treatment, and 12 were assigned to the no-treatment control condition. EMDR subjects received two treatment sessions. Post-treatment assessments were conducted with both groups approximately 17 days subsequent to the initial assessment.

As the two, controlled, group studies prior to the Jensen (1994) study had demonstrated (Boudewyns et al., 1993; Shapiro, 1989a) the results showed that EMDR was effective for reducing in-session anxiety as measured by the SUD scale. However, contrary to Shapiro’s (1989a) findings, EMDR did not significantly enhance VoC ratings. Similar to the Boudewyns et al. (1993) study, there were no significant differences between the EMDR group and the no-treatment control group on any standardized assessment device at post-treatment. Jensen (1994) concluded that “In general, converging data obtained in this study fail to support the effectiveness of EMDR with Vietnam combat veterans” (p. 321). One potential explanation for the lack of stronger results across self-reported measures might have been the relatively inexperienced therapists assigned to administer treatment.

Vaughan et al. (1994) contrasted the effects of three treatments in the treatment of PTSD and PTSD symptomatology with a no-treatment control condition. The treatments were image habituation training (IHT), applied muscle relaxation (AMR), and EMDR. IHT involved listening to a continuous loop of audio-taped descriptions of the original traumatic event while writing pertinent thoughts and recording anxiety levels for 60 minutes per day. AMR taught subjects how to relax physically, recognize incipient signals of their anxiety, and then apply the relaxation techniques contingent on these awarenesses. Whereas subjects in all three treatment groups received between three and five total sessions, both the IHT and AMR procedures involved daily homework assignments through the duration of the study and EMDR did not. Thirty-
six outpatients referred from various community sources served as subjects for this study. They were randomly assigned to treatment as follows: 12 in EMDR, 13 in imaginal exposure, and 11 in applied muscle relaxation. At the outset of the study, 17 subjects were assigned to the no-treatment waiting list for a period of two to three weeks prior to being assigned to one of the three primary treatment groups. Of the 36 subjects, 11 had received prior psychiatric treatment. Structured interviews were conducted by a blind rater. A number of objective, standardized instruments were employed in addition to SUD and VoC self-reports within EMDR sessions. Complete assessment batteries were administered at pre-treatment, following the no-treatment waiting control period, post-treatment, and at 3-month follow-up. Of the 36 subjects, 22% failed to qualify for a PTSD diagnosis. However, all subjects met the criteria for DSM-III-R Categories B (re-experiencing/intrusive) and D (hyperarousal). The authors indicated that this is a PTSD symptom pattern commonly observed in community samples. Comorbid personality disorders were diagnosed in 11 subjects, Generalized Anxiety disorder in 20 subjects, and Major Depression in six subjects.

Vaughan et al. (1994) found that all three treatment groups improved significantly compared with the no-treatment control group. Moreover, these gains were maintained at the 3-month follow-up. Whereas EMDR appeared more effective than the other treatments on some dependent measures, no significant differences emerged among the three treatment approaches. The authors identified small sample sizes as a possible reason that EMDR failed to show superiority over the other two treatments. The authors also cited possible placebo effects for the symmetry of their results.

Merckelbach, Hogervorst, and Kampman (1994) conducted an experiment wherein they compared the effects of EMDR and finger tapping to suppress the effects of an aversive visual stimulus in normal college students. These authors posited that if
EMDR is highly effective for reducing the potency of disturbing mental imagery (as some have claimed), then this effect might be observable following a relatively small number of saccadic eye movements. Forty students participated, 20 in each treatment group. The dependent measures in this study included heart rate and a 100-point visual analogue scale (VAS) which estimated the degree to which an imaginal picture was visualized. The aversive visual stimulus was a picture of a mutilated human hand. Subjects were told that they would see the picture, then answer some questions. There were four phases to this experiment. In phase 1 the students viewed the slide for 10 seconds. During phase 2 the students were instructed to hold a mental image of the picture they had just seen as vividly as possible while their heart rates were measured. Immediately following this the subjects’ VAS scores were obtained for the image. Then, in phase 3, one group visualized the image while receiving four sets of 24 eye movements while the other group visualized the image while engaging in four sets of 24 (right index) finger taps. In the fourth and final phase subjects were instructed to visualize the image again while heart rates were measured. After 10 seconds the VAS was obtained again.

Merckelbach, Hogervorst, and Kampman (1994) reported no significant differences between the EMDR and finger tapping groups on VAS scores or heart rates. These authors concluded that EMDR appears not to affect “emotional memory” more than an arbitrary, active placebo, control condition. Furthermore, Merckelbach, Hogervorst, and Kampman (1994) stated that “the present findings are not inconsistent with an interpretation of EMD effects in terms of placebo, demand, and/or expectancy mechanism” (p. 335).

Montgomery and Ayllon (1994a, 1994b) have conducted two, single-subject experiments involving EMDR. Both studies investigated the effects of EMDR in a multiple baseline design on some aspect of PTSD symptomatology. In both cases,
diagnostic validity was insured, and in both cases a variety of objective, standardized assessment devices were employed. However, the pre-post differences in these measures were not reported. Heart rate and systolic blood pressure were also assessed in both studies. The first study examined the effect of EMDR across two traumatic images as reported by the 42-year-old female subject. The baseline condition involved simple presentation of the traumatic images similar to systematic desensitization. A total of six EMDR sessions were administered. However, one traumatic memory was treated in the first half of the session and another was treated in the second half making for a total of 12 discrete, experimental sessions. The order in which the traumatic memories were treated was reversed for each full session. Between the first and second treatment sessions the authors required the subject to sit quietly for 15-minutes.

Montgomery and Ayllon (1994a) found no overlap between baseline and treatment data points for either traumatic memory for both SUD levels and heart rates. For image 1 there was one overlapping data point between baseline and treatment phases for systolic blood pressure. The degree and rate of improvement observed in SUD scores for both traumatic memories was very similar with no apparent interactivity between the two baselines (i.e., there was no generalization from one treatment to the other). Data patterns for heart rate and systolic blood pressure were not as uniform between images as for SUD scores. However, these two baselines appeared independent, and each decreased significantly with EMDR treatment. Follow-up data points showed some worsening for both images and all dependent measures. SUD levels and heart rate remained substantially lower than baseline levels. However, systolic blood pressure values began to approach baseline levels. Whereas a SUD of zero was not obtained for either memory, the six session trend during EMDR treatment suggested that somewhat lower SUD levels ultimately would have been obtained. The authors have pointed out that even though theirs was a single-subject, experimental
design, no reversal was attempted (predicated on the presumption that EMDR effects were permanent). Also, the degree of effects was not of the magnitude reported by Shapiro (1989a), nor was the efficiency (six sessions vs. one).

The second single-subject study conducted by Montgomery and Ayllon (1994b) examined the effects of EMDR in a phase-change, multiple baseline design across six subjects. In this study, EMDR was compared with the identical procedure sans saccadic eye movements. In total there were four conditions or phases which each subject experienced in the following order: baseline, EMDR sans eye movements, EMDR per Shapiro (1989a), and follow-up. Each subject experienced the same number of sessions within each phase; the EMDR phase involved six sessions.

The results indicated no interactivity (or generalization) among baselines across subjects (Montgomery and Ayllon, 1994b). For all six subjects the EMDR sans eye movement procedure (consisting of the cognitive restructuring and exposure components) was ineffective for improving significantly any self-report or physiological measure. Adding saccadic eye movements to the cognitive restructuring and exposure components produced clinically significant pre-post decreases in SUD reports for five of six subjects. However, contrasted with EMDR sans eye movements, EMDR with eye movements did not produce significant reductions in SUD levels. Only the overall reductions in SUD levels from baseline to follow-up were significant. The entire treatment sequence (from baseline to follow-up) also resulted in statistically significant reductions in the frequency of self-reported intrusive thoughts and disturbing dreams. Finally, pre- to post-treatment levels of self-reported depressive symptoms dropped significantly. However, total pre- to post-treatment changes in any dependent measure might be a function of repeated measurements or the passage of time (c.f., Lohr, Tolin, and Kleinknecht, 1996).
Unlike their initial single-subject EMDR study (Montgomery and Ayllon, 1994a), Montgomery and Ayllon (1994b) found no significant differences in any physiological measure for any subject. However, compared with EMDR sans eye movements, as it did on SUD levels, EMDR did produce systematic decreases in both blood pressure and heart rate for all six subjects. Regarding the subject for whom no changes in self-reported measures were observed, the authors reported that this individual, who was otherwise physically able and healthy, was involved in a legal proceeding wherein he was claiming total disability on the basis of his PTSD. Hence, this subject had a powerful secondary agenda that appeared antithetical to therapeutic improvement.

Montgomery and Ayllon, (1994b) concluded that their results supported the utility of the entire EMDR protocol in the treatment of PTSD. These authors found that EMDR as per Shapiro (1989a) was significantly more effective than EMDR sans eye movements (cognitive restructuring and exposure components alone). Although neither the magnitude of improvement nor the efficacy of the protocol were as remarkable as some previous studies (e.g., Shapiro, 1989a, 1989b; McCann, 1992), the authors stated that the treatment gains they observed were superior to traditional therapies for PTSD such as implosion and flooding. However, as these authors have implied, the sequential, additive nature of their experimental design does not support conclusive statements regarding the unique efficacy or contribution of eye movements to the EMDR protocol. Although the addition of the eye movement procedure to the cognitive restructuring and exposure components was sufficient to produce substantial reductions in self-reported SUD levels, intrusive thoughts, disturbing dreams, and depressive symptoms, this additive strategy does not support the necessity of eye movements to produce such effects. Without a suitable control group, the non-specific factors intrinsic to a credible, active placebo procedure cannot be ruled out.
Renfrey and Spates (1994) assigned 23 PTSD subjects to one of three conditions: (1) standard EMDR, (2) EMDR with eye movements generated by a light bar, and (3) EMDR with eye movements supplanted by fixed visual attention. PTSD was assessed via a standardized battery of assessment instruments which also functioned to evaluate the magnitude of therapeutic progress. At pre-screening, 21 of the 23 participants met the DSM-III-R criteria for PTSD. In addition, heart rate and SUD levels were monitored during pre, post, and follow-up assessments. The primary experimenter functioned as the primary therapist. EMDR treatment was provided up to a maximum of six sessions. The three treatment conditions involved identical procedures with the exception of the eye movement component, as described.

Renfrey and Spates (1994) reported that the magnitude and nature of EMDR effects on SUD and VoC scales replicated or exceeded those of Shapiro (1989a). However, all three treatment regimens produced nearly identical results across all dependent measures, from which Renfrey and Spates (1994) concluded that “EMD does appear to be an effective treatment for post-traumatic sequelae and...eye movements do not appear to play a necessary role” (p. 238). By way of explanation for these results, these authors discussed a number of hypothetical etiologies and constructs. However, it remained unclear why all three EMDR versions produced such similar results.

Bauman and Melnyk (1994) investigated the effects of EMDR and a placebo control condition (finger tapping) on self-reported levels of test anxiety on 15 pairs of college students. Subjects for this study were introductory statistics students who scored above the 50th percentile on the Test Anxiety Inventory (TAI; Spielberger, 1980). The TAI was administered one week prior to the first statistics exam (pre-treatment) and again one week prior to the second statistics exam (post-treatment). Also, scores from the first exam were analyzed for interaction with treatment effects.
The subjects were rank ordered in pairs by TAI score and then randomly assigned from the matched pairs to one of the two treatment groups. In order to match the number of sets of finger taps with eye movements, the subjects in the EMDR group were treated first. Sufficient sets of eye movements were administered to reduce the SUD levels to zero or until the maximum treatment time of 45 minutes had elapsed. The finger tapping group then received the same number of sets of finger taps as the number of eye movement sets that their paired counterpart received. For “ethical” reasons immediately following the conclusion of the finger tapping session, prior to the post-treatment assessment, all finger tapping subjects were offered EMDR treatment. Nine of the 15 finger tapping subjects chose this alternative. This was a major confound in this study that contaminates the post-treatment assessment which was subsequently administered one week prior to the subjects’ second statistics exam. Apparently, an unspecified number of additional finger tap subjects opted to receive EMDR following the post-treatment assessment. This confound “nullified follow-up comparisons between groups” (Bauman and Melnyk, 1994, p. 30). It is unclear why nine finger tap subjects were given EMDR treatment prior to the post-treatment assessment (perhaps the authors felt strongly biased toward EMDR). Moreover, it is unclear how this does not nullify the post-treatment results of this study any less than additional finger tap subjects receiving EMDR after the post-treatment assessment does nullify the pre-planned follow-up.

Nevertheless, Bauman and Melnyk (1994) have reported that both treatments produced fairly equivalent, significant reductions in SUD reports and the TAI emotion scale. However, only the (confounded) finger tapping procedure produced significant reductions in the TAI worry and total scales. Finally, regarding objective test performance, the finger tapping group had significantly higher first exam scores than the EMDR group.
Hekmat, Groth, and Rogers (1994) studied the effects of EMDR, a variant of EMDR wherein listening to music replaced the reprocessing portion of the protocol (hence, EMDM), and a control condition on management of acute pain induced by hand immersions in ice water. The subjects were 30 undergraduate psychology students. This study also considered whether subjects' hypnotizability was a factor in subsequent tests of pain endurance. Objective, standardized measures of hypnotic susceptibility, affect assessment, and pain threshold, tolerance, and endurance were employed. All subjects performed three hand immersion trials using a cold presser device. For both treatment conditions the first immersion was a baseline measure and the subsequent immersions involved treatment procedures. For the control condition all three hand immersion trials were done without any form of guidance, feedback, or special stimulation. The EMDR treatment "target" was structured around the first immersion trial. Individuals were instructed to formulate a positive self-statement (e.g., "I can handle the pain") regarding the worst sensations they experienced during the first trial and combine these two elements with eye movements (induced by following a moving light on a TV screen) during the second and third immersions. The EMDM procedure consisted of selecting preferred music subsequent to the baseline trial, then combining the preferred music with eye movements (induced following a moving light on a TV screen) during the second and third immersions.

In general, the results indicated that EMDR and EMDM both produced significant increases in subjects' ability to cope with acute pain induced by exposure to ice water (Hekmat, Groth, and Rogers, 1994). Moreover, both treatments reduced overall current anxiety levels as well. Subjects who appeared to be of greater hypnotic susceptibility did no better on the immersion trials than subjects of lesser hypnotic susceptibility. Hekmat, Groth, and Rogers (1994) concluded that those components loosely described as "reprocessing" are not essential to the efficacy of EMDR. These
authors have stated that Wolpe's theory of counterconditioning (Wolpe, 1990) might be a parsimonious explanation for the therapeutic effects of EMDR and EMDM observed in this study.

Foley and Spates (1995) compared the utility of EMDR with two EMDR variants and a no-treatment control condition in the treatment of public-speaking anxiety. Subjects were 40 college students assessed with excessive levels of public-speaking anxiety. In addition to SUD and VoC levels, public-speaking anxiety was evaluated at pre- and post-treatment using two, standardized, self-reported measures, one directly observed behavioral measure, and a physiological measure (heart rate). The standard EMDR treatment was as per Shapiro (1991a). The first variant substituted a bilateral, moving audio stimulus delivered via headphones. The second variant instructed subjects to focus their eyes on their hands resting on their laps. Subjects in the three treatment groups received a maximum of two treatment sessions. The post-treatment assessment was administered at least one week subsequent to the subject's last treatment session.

The results of this study found that all three EMDR variants had similar, significant effects on SUD and VoC levels as well as on one standardized measure of public speaking anxiety. All three treatments produced somewhat less improvement in the other standardized self-report measure and the standardized behavioral measure. None of the conditions affected heart rate. The control condition was generally ineffective for improving any of the dependent measures. Foley and Spates (1995) reported that "the results suggest that the eye movement component of the treatment was not essential to treatment outcome...This suggests a role for nonspecific factors in treatment outcome" (p. 327). The authors cite their small sample size and the failure to assess the self-reported SUD and VoC levels with the control group as limitations of this study.
Lohr, Tolin, and Kleinknecht (1995) conducted an investigation of EMDR treatment for medical phobias with two subjects. A within-series, phase-change design (Barlow, Hayes, & Nelson, 1984) was used with both subjects. For subject one three phases were administered which included: (1) presentation of the traumatic image similar to systematic desensitization (referred to as “a”), (2) EMDR sans eye movements (referred to as “b”), and (3) EMDR (referred to as “b+c”). For subject two only phases “b” (EMDR sans eye movements) and “b+c” (EMDR) were administered. For subject two the authors decided to dispense with the image presentation phase so that they could examine more thoroughly the results of adding eye movements to the cognitive restructuring and exposure components. For both subjects EMDR was administered within a single treatment session. Several standardized fear assessment instruments were administered at pre- and post-treatment assessments, primarily for screening purposes. However, the SUD scale and heart rate were the two primary, within-session dependent measures. No statistical analyses were performed; all data was analyzed visually. Pre-treatment measures of fear qualified both subjects as “phobic.”

For subject one Lohr, Tolin, and Kleinknecht (1995) reported that for two of the three traumatic images repeated presentation of these images within the “a” phase resulted in substantial reductions in self-reported SUD levels. However, for the third traumatic image, repeated presentations appeared to have less effect. At first, EMDR (“b+c”) produced substantial reductions in the SUD rating for this third memory. Then SUD levels returned to near baseline elevations before falling once more to levels below 10. No systematic changes in heart rate were observed across conditions. However, all standardized measures of fear diminished substantially from pre- to post-treatment assessments. Nonetheless, for subject one, anecdotal follow-up reports suggested limited generalization of treatment effects.
For subject two, the initial treatment with EMDR sans eye movements ("b") produced virtually no change in self-reported SUD levels. The contiguous EMDR ("b+c") treatment resulted in steadily falling SUD levels over eye movement sets. On a 100-point scale, the reduction was from approximately 70 to 15. SUD levels rose to near 50 at the beginning of the "b" phase that followed, and quickly fell to below 10 within that phase. A similar reduction in subject two's SUD levels was noted for the "b" and "b+c" phases of another traumatic memory image. As with subject one, no systematic changes in heart rate were observed across conditions, all standardized measures of fear diminished substantially from pre- to post-treatment assessments, and anecdotal post-treatment reports suggested limited generalization of treatment effects.

As is intrinsic with much single-subject research, conclusions based on this study were limited by the potential irreversibility of treatment effects; hence, replications were equivocal. For example, subject one's declining SUD scores when repeatedly exposed to target stimuli during the baseline "a" phase suggested that subsequent improvements during the EMDR treatment phase might have been falsely attributed to EMDR.

Wilson, Becker, and Tinker (1995) compared EMDR with delayed EMDR treatment on traumatic memories in 80 non-clinical subjects. Subjects were individuals reporting a history of at least one traumatic occurrence. They were recruited via various advertisements. Forty-six percent of participants met all DSM-IV criteria for PTSD. One hundred percent of participants met at least one criteria for PTSD. These diagnoses were validated by independent raters. Five therapists were used for this study. Four of them had received advanced EMDR training. In addition to the process measures of SUD and VoC, three standardized, self-report measures were evaluated. The 80 subjects were randomly assigned to either immediate or delayed EMDR treatment. For both groups EMDR was administered in three, 90-minute sessions.
The delayed treatment group began EMDR treatment at the point when the immediate treatment group terminated treatment. Complete assessments were conducted on five occasions: (1) pre-treatment for both groups, (2) post-treatment for the immediate group and a second pre-treatment for the delayed group, (3) post-treatment for the delayed group, (4) 90-day follow-up for the immediate group, and (5) 90-day follow-up for the delayed group.

Wilson, Becker, and Tinker (1995) found no significant differences between groups at pre-treatment. The immediate treatment group achieved significant improvements on all dependent measures. The delayed group showed virtually no change during the wait period wherein the immediate group received treatment. Upon completion of the three-session EMDR protocol, the delayed treatment group showed very similar effects as the immediate group. The authors concluded that the results of their study concurred with those presented by Shapiro (1989a). Further, Wilson, Becker, and Tinker (1995) reported that EMDR has “promise in the treatment traumatic memories, although the reasons for its effectiveness are yet to be understood” (p. 36). They stated that nonspecific factors cannot be ruled out.

Gosselin and Matthews (1995) conducted an investigation of eye movement and expectancy variables. These authors arranged an experiment using EMDR to treat test anxiety in 41 undergraduate college students. The TAI was used pre- and post-treatment and at a 1-month follow-up to assess test anxiety. SUD and VoC scales were used also assessed at these times and to rate intersession progress. Subjects were selected on the basis of high levels of self-reported test anxiety (above the eighty-fifth percentile). The primary investigator for this study also served as the primary EMDR therapist. The study used a 2 x 2 ANOVA design. The two manipulated factors were eye movement and expectancy. Hence, there were four conditions to which subjects were randomly assigned: (1) High-expectancy with EMDR, (2) low-expectancy with
EMDR, (3) high-expectancy with EMDR sans eye movements, and (4) low-expectancy with EMDR sans eye movements. All subjects received a 60-minute treatment session. The primary therapist was blind to the expectancy statement which all subjects heard just prior to the treatment session. EMDR followed the protocol described by Shapiro (1989a). The no eye movement condition was identical to the EMDR condition with the exception that subjects were instructed to look at the therapist’s unmoving fingers.

Gosselin and Matthews (1995) found that EMDR was significantly more effective for reducing SUD levels than EMDR sans eye movements. However, both EMDR with and without eye movements had similar effects on the standardized TAI. Neither treatment condition significantly improved VoC scores. These results held at the 1-month follow-up. Surprisingly, these authors found that the level of expectancy had no effect on outcomes. The authors contended that positive expectancy for change is likely a contributing factor to therapeutic outcomes. Further, they posited that the lack of an expectancy effect in their study might have been due to the low-expectancy condition not being low enough by comparison to the high-expectancy condition. These results support Shapiro’s (1989a) contention that EMDR is effective for reducing self-reported anxiety. However, the equivocal results between the eye movement and no eye movement conditions on the standardized TAI challenges the function and necessity of eye movements within the EMDR protocol.

Dunn, Schwartz, Hatfield, and Wiegele (1996) examined the effects of EMDR on self-reported anxiety in a “normal” college population. These authors drew 28 subjects from a pool of 603 potential subjects. The 28 were matched on sex, age, and severity and type of stressful or traumatic incident. One subject from each pair was then randomly assigned to receive either EMDR or EMDR sans eye movements (gazing at a spot on the wall). SUD scores were augmented by a number of physiological measures, all of which were obtained pre- and post-treatment. The EMDR procedure
was conducted with a stressful or traumatic memory until SUD levels reached zero or until SUD levels remained stable for 45 minutes. The EMDR subjects were treated first so that the paired subject's no eye movement treatment time could be yoked with their EMDR partner's treatment time.

Dunn, Schwartz, Hatfield, and Wiegele (1996) reported that EMDR and EMDR sans eye movements produced similar, significant reductions in self-reported SUD levels. Moreover, they reported that this was doubly remarkable because the EMDR subjects dictated the length of their treatment (within the design limit) and the non eye movement subjects had to accept the time limit with which they were yoked. The effects of EMDR and EMDR sans eye movements on the physiological measures were similarly mixed and generally corroborated the self-reported SUD levels. The authors posit that similar performance across the two groups might have been due to expectancy variables or to relatively equivalent non-specific or placebo effects.

Bates, McGlynn, Montgomery, and Mattke (1996) conducted an experiment examining the effects of a single EMDR session on fear of spiders in 14 non-clinical subjects. Subjects were solicited by various means and at pre-treatment met the DSM-III-R criteria for specific phobia. Fear of spiders was assessed by SUD and VoC reports, various standardized self-report measures, a manual self-report device, various physiological measures, and by in vivo approach to caged spiders. The assessment battery was administered pre- and post-treatment. At pre-treatment subjects were exposed to one in vivo trial that involved observing spiders at a distance of six feet for a period of four minutes. During and after this exposure all dependent measures were obtained. Then, seven subjects were randomly assigned to EMDR treatment while the remaining seven were assigned to a no-treatment control period. Approximately 10 to 18 days later a condition-blind assessor conducted post-treatment evaluations which involved a series of six trials similar to the pre-treatment trial. As before, the first four
trials brought the spiders to a distance of six feet from the subjects. To assess for behavioral avoidance, the final two trials gave the subjects manual control over their proximity to the spiders. EMDR was administered by a well trained assistant whose performance was closely monitored for treatment integrity through the entire experiment.

Bates, McGlynn, Montgomery, and Mattke (1996) found that, compared to the no-treatment condition, EMDR did not reduce any of the dependent measures from pre-to post-treatment nor did it reduce behavioral avoidance of spiders. The only effect these authors observed EMDR to have was the reduction of orally reported SUD scores across the first four post-treatment exposures. This result the authors attributed to an anxiolytic effect associated with repeated exposure. McGlynn, Montgomery, and Mattke (1996) concluded:

The failure of the EMDR-type experience to produce change among orthodox measures of fear is noteworthy because the experiment was relatively sound methodologically compared to much of the empirical work reported heretofore. (p. 567)

The authors have pointed out that their negative results should be viewed with caution in as much as the external validity of their experiment might have been compromised given the use of fearful college students (analogue subjects) as opposed to true clinical subjects. Nevertheless, they have also acknowledged the growing doubts among many researchers regarding EMDR as a viable behavioral therapy, especially for phobias for which viable standard approaches currently exist.

Wilson, Silver, Covi, and Foster (1996) conducted a study with 18 subjects that explored the effects of EMDR and two variant procedures on SUD and VoC levels and a range of autonomic measures. This study was a significant advancement over previous attempts to examine the autonomic correlates of EMDR in that heart rate (HR), blood pressure (BP), galvanic skin response (GSR), respiration (R), and fingertip skin
temperature (T) were all assessed before, during, and after all treatment conditions. Also, abreactions were assessed. These were operationally defined as sobbing and were reflected in irregular deep breathing, gross fluctuations in GSR, irregular HR, and decreased BP. The subjects for this study were mixed. They were recruited from three community agencies. All save one were already receiving some form of psychological treatment. The primary criterion for this study was the presence of at least one currently disturbing memory pertaining to a past event. Once referred the subjects were screened by the primary investigator using a history form, symptom review, and individual interview. Of the 18 subjects, 11 met criteria for PTSD, five met the criteria for Simple Phobia, one met the criteria for PTSD and Panic Disorder without Agoraphobia, and one met the criteria for Dream Anxiety Disorder without concurrent mental disorder. In the treatment groups, 12 subjects reportedly focused on traumatic memories, five focused on a phobia related event, and one on a recurrent nightmare.

The three treatments used in this study were EMDR, EMDR sans eye movements (TIC), and EMDR sans eye movements with thumb taps in time to a metronome—a competing motor activity (TAP) (Wilson, Silver, Covi, and Foster, 1996). Treatments were administered by two of the primary investigators both of whom were trained by Shapiro. All subjects received one treatment session which averaged 13.5 sets for the EMDR group and 10 sets for both control groups. Similar to the Bauman and Melnyk (1994) study, all subjects who completed the TIC and TAP procedures and reported less than total desensitization were immediately offered EMDR. This confounded post-treatment assessment comparisons, and also created two additional experimental groups: TIC-EMDR and TAP-EMDR. Moreover, this automatic default to EMDR treatment suggests that the primary therapists, who were also the principle investigators, might have been biased toward EMDR.
Wilson, Silver, Covi, and Foster (1996) reported that the results of this study were supportive of Shapiro's (1989a) original controlled study. At the end of treatment, EMDR had produced significantly greater reductions in SUD and VoC levels than had either TIC or TAP. Moreover, only 1 TIC or TAP subject reported complete desensitization whereas the remaining 17 (EMDR, TIC-EMDR, and TAP-EMDR) all reported complete desensitization. These gains were maintained at 3-, 6-, and 12-month follow-ups. During the administration of treatment, EMDR was the only treatment to improve significantly a number of autonomic variables. These included: GSR (commonly interpreted as a "relaxation response"), heart rate, and fingertip temperature (commonly interpreted as a sign of deep relaxation). EMDR also produced unique respiratory patterns and resulted in a much higher occurrence of "abreactions" than TIC or TAP. None of the treatments reduced blood pressure levels significantly. From these results Wilson, Silver, Covi, and Foster (1996) concluded that EMDR is effective within a single session, eye movements appear to contribute to EMDR in the form of a relaxation response, and the exposure component appears to be of lesser importance. These authors suggest that EMDR's effectiveness might be due in part to a reciprocal inhibition process (Wolpe, 1990).

Foley and Spates (1996) conducted a study extending and refining their earlier work with communication anxiety in college students. In addition to SUD and VoC scales, communication or speech anxiety was measured with a standardized self-report instrument, a behavioral instrument (administered by trained observers), and subject heart rate. Assessments were conducted pre- and post-treatment and at a 1-month follow-up. Working with 32 subjects who exhibited at least moderate communication anxiety, these authors employed a 2 x 2 repeated measures ANOVA design that manipulated a pre-treatment assessment speech and the imaginal exposure component of the standard EMDR protocol as factors. Hence, this study involved four treatment
groups: (1) EMDR with pre- and post-treatment assessment speeches, (2) EMDR with a post-treatment speech only, (3) EMDR sans the imaginal exposure component with pre- and post-treatment assessment speeches, and (4) EMDR sans the original exposure component with a post-treatment assessment speech only. EMDR sans the imaginal component was identical to EMDR except that subjects were not instructed to hold anything "in mind" during the eye movement procedure. All treatment was delivered by a therapist trained by the developer or by an experienced EMDR therapist. An observer viewed 30% of all sessions from behind a one-way mirror to observe the treatment procedures and verify that subjects were treated according to their group designation. Subjects received one or two treatment sessions contingent on their response to treatment.

Foley and Spates (1996) found that all treatment conditions produced significant reductions from pre- to post-treatment on several measures. No significant differences were found between groups on any measure. The authors have stated that their results were inconclusive regarding either the necessity of the imaginal exposure component to the EMDR protocol or the treatment utility of the pre-treatment assessment speech. However, the authors also have reported that small group sizes might have contributed to the lack of significant differences on the pre-treatment assessment factor.

Largo-Marsh (1996) examined the effects of EMDR and structuring writing sessions on PTSD. She also examined the influences of hypnotizability and expectancies as factors affecting EMDR treatment outcomes. For this study, 24 subjects of a clinical nature were recruited via print media and community agencies. Diagnostic validity was established using a standardized diagnostic interview and various self-report instruments. Hypnotizability was assessed via a standardized, self-report device. Expectancies for treatment outcomes were evaluated with a subjective, non-standardized instrument that the author reported had obvious face validity.
Assessments were conducted at pre- and post-treatment and at 1-month follow-up. Inclusion criteria for the study included: (a) having experienced a traumatic event as defined by Category A of the DSM-IV diagnostic criteria for PTSD, (b) reexperiencing the event (i.e., intrusive thoughts, dreams, flashbacks), (c) persistent avoidance of stimuli associated with the trauma, (d) and hyperarousal (i.e., sleep disturbances, hypervigilance, increased physiologic reactivity to associated stimuli met the DSM-IV. Subjects received a maximum of three, 60-minute treatment sessions. EMDR was presented as per Shapiro (1991). The structured writing sessions involved many of the same exposure and reprocessing components of the EMDR procedure. The structured writing technique involved the subject sitting at a table alone in a room with paper and pen. They were instructed to visualize the event, label it, evaluate their level of distress vis-a-vis the event, and identify a negative and positive cognition for the event. These subjects were interrupted every 15 minutes to assess SUD levels and at the beginning and end of the sessions to assess VoC levels. No manipulation such as eye movements was performed.

Largo-Marsh (1996) reported that a repeated measures ANOVA analysis found both EMDR and structured writing were equally effective for reducing PTSD symptoms. Neither measures of hypnotic susceptibility nor subject expectations were found to be significant factors contributing to either treatment’s efficacy. Similar results were observed at 1-month follow-up. Based on these results, Largo-Marsh (1996) concluded that writing about a traumatic event can be therapeutic. Moreover, similar to other sensory alternatives, writing, as it was used herein, might be successfully substituted for eye movements and still maintain the effectiveness of the EMDR protocol. This study joins a growing body of research that suggests that eye movements are not an essential ingredient within the EMDR clinical recipe.
Lohr, Tolin, and Kleinknecht (1996) have presented two cases of EMDR treatment of simple phobia (claustrophobia) using a within-series, phase-change, experimental design. This design represented a substantial improvement over previous single-subject designs and case studies. Both subjects met the DSM-IV criteria for simple phobia. A standardized, self-report instrument was administered pre-treatment and at two-month follow-up. SUD, VoC, and heart rate values were obtained following each imaginal exposure which lasted 30 seconds. Three phases were administered that included: Image presentation (A; similar to systematic desensitization), analogue EMDR (B; all components of EMDR excluding the eye movements), and EMDR (B+C). A minimum of four imaginal exposures were administered per experimental phase. The sequence of the phases and the number of traumatic memories treated differed for each of the two subjects. Only one of subject one's three traumatic memories were treated within the experimental design. Subject one experienced the following sequence of phases once: A-A-A/B-B-B/B+C-A-A-A. Due to time constraints and the desire to examine in greater depth the significance of the eye movements themselves, subject two was not exposed to phase one. Rather, subject two's treatment began with phase two and consisted of alternations between phases two and three before and after each of the four traumatic memories that were desensitized. Specifically, untreated memories were probed within the analogue EMDR or phase two condition before the first EMDR or phase three condition, untreated and treated memories were probed within the analogue EMDR or phase two condition before and after the second and third EMDR or phase three conditions, and treated memories were probed within the EMDR analogue or phase two condition after the fourth and final EMDR or phase three condition. This arrangement facilitated an analysis of generalization and permanence of treatment effects among all four traumatic memories.
for subject two. For each of the four traumatic memories, subject two experienced the experimental phases in the following sequence: B-B+C-B-B.

For subject one Lohr, Tolin, and Kleinknecht (1996) found no significant change in the standardized, self-report measure. The authors considered statistical analyses for the time-series data (SUD, VoC, and heart rate), however, due to limitations of the data and, or, the design these efforts were abandoned for both subjects. Hence, all analyses of time-series data involved visual inspection only. For both subjects EMDR appeared to have no effect on heart rate levels; these were relatively stable across experimental conditions. For subject one, the addition of eye movements to the EMDR analogue condition appeared to result in substantial reductions in self-reported SUD levels related to the treated traumatic memory (which subject one described as the “etiological event” for the disorder). Moreover, this treatment effect appeared to generalize to the two other traumatic memories. However, the image presentation condition (phase one) produced marked decreases in SUD levels for all three traumatic memories which suggested that further presentations of the images with no additional intervention might have produced results similar to those of EMDR.

For subject two, Lohr, Tolin, and Kleinknecht (1996) found no significant change in the standardized, self-report measure. The EMDR analogue condition (administered prior to EMDR) produced no change or slight increases in SUD levels on all but one occasion when it resulted in a substantial drop in SUD levels. The addition of eye movements to the EMDR analogue condition produced substantial reductions in self-reported SUD levels for all four traumatic memories. However, unlike subject one, subject two did not receive the image presentation condition. Hence, it is possible that simple exposure to the traumatic images might have reduced self-reported SUD levels as it did for subject one. Also, for all traumatic memories, the effects of EMDR for subject two did not appear to be permanent. SUD levels (indicative of fear) rose
substantially on one of four analogue probes of traumatic memories administered immediately following EMDR treatment for those memories. Likewise, the SUD level for the follow-up analogue probe of memory one, which was conducted one month after the conclusion of EMDR treatment for that memory, immediately following the conclusion of EMDR treatment for memory four, found a substantial increase in that SUD level (immediate post treatment SUD for memory one was below 10; at one-month follow-up SUD had rebounded to over 80).

Lohr, Tolin, and Kleinknecht (1996) indicated that a spontaneous, anecdotal, follow-up report from subject two suggested that EMDR treatment had lead to substantial improvements in numerous aspects of that individual’s life. Despite what appeared like a successful outcome, these authors have counseled that their results be interpreted cautiously. They cite the potentially confounding effects of concomitant variables such as dosed exposure, within session demand contingencies for improved SUD reports, and expectations for positive outcomes all of which need to be ruled out before the true efficacy of EMDR can be known. These authors have stated that, whereas single-subject research is appropriate for strategic explorations of EMDR’s applications, discovering the nature and extent of concomitant variables such as exposure or those owing to non-specific functions will require manipulation within group experimental designs.

EMDR—Summary of Experimental Research

In terms of the types of subjects and diagnostic categories treated, the duration and nature of the independent variables manipulated, and the range of dependent variables measured and analyzed, the controlled investigations of EMDR are a diverse group of projects. Moreover, the types of experimental designs employed among controlled EMDR investigations has varied widely. As a result of this variance within
the EMDR research the empirical evidence supporting or refuting EMDR's clinical utility is arguably incomplete and guarded. Nevertheless, a number of trends have emerged from the literature extant, and these suggest both the functional characteristics of EMDR as well as the avenues of research needed to gain greater analytic clarity.

One of the more salient details that surfaces from an analysis of the experimental EMDR research literature is that subsequent to the initial Shapiro (1989a) study no investigation has been able to replicate the rapid, powerful effects demonstrated by Shapiro (1989a). A number of case studies have claimed to realize fantastic treatment effects within a single 90-minute session (Kleinknecht and Morgan, 1992; McCann, 1992; Puk, 1991; Pellicer, 1993; Shapiro, 1989b; Spector and Huthwaite, 1993). However, to date, only one controlled EMDR group study (Wilson, Silver, Covi, & Foster, 1996) has approached the type of pre- to post-treatment changes in SUD levels (relative to a placebo control condition) that Shapiro (1989a) attained in one or two sessions.

Another striking aspect within the experimental EMDR research literature is that the SUD scale is the only dependent measure upon which EMDR has had a reliable impact (Lohr, Tolin, & Kleinknecht, 1996). Without exception, every controlled research study reviewed in the current text that has incorporated a SUD measure has reported that EMDR reduced SUD levels significantly, or, in the cases of the single-subject investigations, appeared to reduce SUD levels compared to control conditions. In addition, each of the four group studies within the current review that compared EMDR treatment to a no-treatment waiting condition found that EMDR was significantly more effective for reducing SUD levels than a no-treatment control condition (Bates, McGlynn, Montgomery, & Mattke, 1996; Boudewyns et al., 1993; Jensen, 1994; Wilson, Becker, & Tinker, 1995).
However, as Lohr, Tolin, and Kleinknecht (1996) have pointed out, the treatment utility or validity of the SUD measure has yet to be established. Moreover, most studies have found that the SUD scale has been equally sensitive to EMDR treatment and a wide range of placebo control procedures and, or, alternative treatments (Bauman & Melnyk, 1994; Dunn, Schwartz, Hatfield, & Wiegele, 1996; Foley & Spates, 1995, 1996; Largo-Marsh, 1996; Renfrey & Spates, 1994; Sanderson & Carpenter, 1992; Vaughan et al., 1994). Of the 17 controlled EMDR group studies being reviewed currently, only two of them reported that EMDR produced significantly greater reductions in SUD levels than a control condition (Gosselin & Matthews, 1995; Wilson, Silver, Covi, & Foster, 1996). What’s more, Gosselin and Matthews (1995) found that EMDR and the control condition (EMDR sans eye movements) produced nearly identical results on the VoC and on the one standardized, self-report instrument. Unfortunately, in the Wilson, Silver, Covi, and Foster (1996) study, outcome comparisons among EMDR and the two control conditions on all standardized, self-report measures were not viable due to a procedural confound (as previously discussed).

Nevertheless, of the 10 experimental EMDR group studies in the current review that compared the effects of EMDR with one or more placebo control or alternative treatment conditions on one or more standardized measures (Bauman & Melnyk, 1994; Boudewyns et al., 1993; Dunn, Schwartz, Hatfield, & Wiegele, 1996; Foley & Spates, 1995, 1996; Gosselin & Matthews, 1995; Largo-Marsh, 1996; Merckelbach, Hogervorst and Kampman, 1994; Renfrey & Spates, 1994; Vaughan et al., 1994), none of them found EMDR to be superior to their respective control conditions for improving standardized, self-report scores. Of the three experimental EMDR group studies in the current review that compared the effects of EMDR with a no-treatment waiting condition on one or more standardized measures (Bates, McGlynn,
Montgomery, & Mattke, 1996; Jensen, 1994; Vaughan et al., 1994), only one found EMDR to improve standardized, self-report scores significantly more than the respective no-treatment condition (Vaughan et al., 1994).

Unlike its effects on SUD reports, the effects of EMDR on physiological measures has been inconsistent. Of the five experimental studies comparing the sensitivity of various physiological measures to EMDR and a range of placebo control conditions and alternative treatments (Boudewyns et al., 1993; Foley & Spates, 1995, 1996; Renfrey & Spates, 1994; Wilson, Silver, Covi, and Foster, 1996), only one has demonstrated EMDR's superiority to a control condition for improving various physiological processes (Wilson, Silver, Covi, and Foster, 1996). However, it should be noted that the Wilson, Silver, Covi, and Foster (1996) investigation was the only one in this group to replicate the methodology (and the results) of the original Shapiro (1989a) study and assess physiological parameters within session. The others evaluated physiological measures at pre- and post-treatment assessments. It could well be that EMDR does impact a range of bodily processes. However, the precise therapeutic value of these effects has yet to be established.

EMDR's ability to enhance directly observed, behavioral measures of performance significantly more than a placebo control condition or alternative treatment has yet to be demonstrated. Of the four EMDR group studies within the current analysis that employed an objective, directly observed dependent measure (Bates, McGlynn, Montgomery, & Mattke, 1996; Foley & Spates, 1995, 1996; Hekmat, Groth, & Rogers, 1994), none of them reported differential results on their respective objective measure among the interventions employed.

Although they have involved only a handful of experimental subjects (nine subjects within four studies), the results of single-subject EMDR research have been qualitatively similar to the experimental group efforts. Of the four single-subject
designs currently reviewed (Lohr, Tolin, & Kleinknecht, 1995, 1996; Montgomery & Ayllon, 1994a, 1994b), all of them found SUD measures sensitive to EMDR treatment, and all of them found that EMDR produced greater improvements in SUD scores and standardized, self-report measures than EMDR sans eye movements. However, in the two studies that evaluated statistically differences in SUD scores and standardized variables (Montgomery & Ayllon, 1994a, 1994b), no significant differences were found on these variables between EMDR and EMDR sans eye movements. Mean differences in these variables reached significance only when the effects of the entire pre- to post-treatment sequence was considered.

As with most controlled group research, the impact of EMDR on physiological variables within the single-subject literature has been inconsistent. Montgomery and Ayllon (1994a) found that EMDR reduced heart rate and blood pressure levels significantly more than repeated image presentation, but not more than EMDR sans eye movements. The other single-subject investigations currently being reviewed were unable to replicate these results (Lohr, Tolin, & Kleinknecht, 1995, 1996; Montgomery & Ayllon, 1994b). These authors reported that, overall, physiological measures were fairly stable across and consistent within experimental conditions.

Similar to most group designs, objective measures of change in single-subject studies have been lacking. Moreover, follow-up reports largely have been anecdotal; the real world experiences that subjects have reported post-treatment have been mixed, ranging from spectacular (Montgomery & Ayllon, 1994a) to poor (Lohr, Tolin, & Kleinknecht, 1995).

The single-subject EMDR research projects have extended the applications of EMDR beyond the treatment of PTSD and traumatic memories to include medical phobias and claustrophobia (Lohr, Tolin, & Kleinknecht, 1995, 1996). However, given the additive nature of treatment conditions within single-subject methodology,
and the potential non-specific and demand effects inherent with the EMDR procedure, single-subject methodology cannot provide unequivocal evidence of the utility of EMDR's individual components. As Lohr, Tolin, and Kleinknecht (1996) have pointed out, adequately controlled group investigations are required for addressing these issues. Toward this end the accumulated group-based evidence has suggested that saccadic eye movements, the presumed primary component of the EMDR protocol, is not uniquely essential to EMDR's efficacy. Of the 13 controlled group studies currently under scrutiny that have compared EMDR with a placebo control condition or significantly different, alternative treatment, 12 have reported equivocal results among treatment conditions on one or more self-reported dependent variables (Bauman & Melnyk, 1994; Boudewyns et al., 1993; Dunn, Schwartz, Hatfield, & Wiegele, 1996; Foley & Spates, 1995, 1996; Gosselin & Matthews, 1995; Hekmat, Groth, & Rogers, 1994; Largo-Marsh, 1996; Merckelbach, Hogervorst and Kampman, 1994; Renfrey & Spates, 1994; Sanderson & Carpenter, 1992; Vaughan et al., 1994). Of these 13 studies, six compared EMDR and EMDR sans eye movements, seven compared EMDR and EMDR with eye movements replaced by an alternative behavior, and one compared EMDR with two alternative treatments. The one investigation whose results generally have supported the utility of eye movements was conducted by Wilson, Silver, Covi, and Foster (1996). These authors found EMDR to be superior to EMDR sans eye movements for improving both SUD reports and a number of physiological measures within session. However, as previously discussed, a procedural confound prevented the analysis of pre- and post-treatment standardized measures.

On the one experimental occasion when EMDR was compared with an alternative treatment that was substantially different than EMDR, EMDR demonstrated non-significant, differential efficacy contrasted with two bonafide alternatives (Vaughan et al., 1994). These authors blamed small sample sizes for the failure of between group
differences to reach significance. These results suggest that, the bulk of the experimental evidence notwithstanding, some aspect of EMDR does have a true therapeutic effect. Moreover, the failure of numerous efforts to replicate Shapiro’s (1989a) initial findings might be explained by the confluence of influential flaws and confounds within the Shapiro (1989a) study which, acting cumulatively, might have greatly exaggerated and confused EMDR’s true clinical efficacy.

What appears to be at least as likely, however, is that EMDR’s power is largely a function of non-specific, expectancy, or placebo effects (Frank & Frank, 1991; Ilardi & Craighead, 1994). Non-specific effects are those that do not derive their therapeutic efficacy from a specific, theory-based, therapeutic mechanism or protocol. The exact mechanisms underlying non-specific, expectancy, or placebo effects have yet to be identified (Fisher & Greenberg, 1989). However, J. D. Frank and J. B. Frank (1991) have specified four essential criteria that identify a (psychotherapy) placebo treatment. These criteria are: (1) that treatment be delivered in a culturally sanctioned setting by a culturally sanctioned helper, (2) that the client and therapist have established a therapeutic relationship, (3) that there is a protocol (or ritual) for treatment, and (4) that there is a believable rationale for the treatment. These non-specific effects of psychotherapy, which are germane to most psychotherapies, suggest that no active therapeutic mechanism, beyond what is defined as psychotherapeutic placebo, is required to produce therapeutic changes. In other words, any believable treatment delivered by a seemingly competent professional should lead to clinically relevant results (Kirsch, 1990). As this passage from Lewis Carroll’s (1865/1946) Alice in Wonderland illustrates, it doesn’t matter how you run the (therapy) race—all participants win.

First it marked out a race course, in a sort of circle ("the exact shape doesn’t matter," it said), and then all the party were placed along the course here and there. There was no "One, two, three, and away," but they began running
when they liked, and left off when they liked, so that it was not easy to know when the race was over. However, when they had been running half an hour or so, and were quite dry again, the Dodo suddenly called out, "The race is over!" and they all crowded round it, panting, and asking, "But who has won?"

This question the Dodo could not answer without a great deal of thought, and it sat for a long time with one finger pressed upon its forehead (the position in which you usually see Shakespeare, in the pictures of him), while the rest waited in silence. At last the Dodo said, "Everybody has won, and all must have prizes." (pp 25-26)

Arguably, "non-specific" factors are the most frequently cited explanation for equivocal experimental results among EMDR researchers. Kirsch (1990) states that the key factor for obtaining positive outcomes in psychotherapy (based on placebo or expectancy effects) is the degree to which the client believes the treatment will work. Treatments that seem unreasonable or lead clients to doubt the efficacy of the procedure will probably not be as successful as more believable placebos, and vice-versa. For this reason, "active" placebos (e.g., EMDR) are preferred. An active placebo produces an effect or sensation that induces the client to believe that the treatment is working (e.g., atropine as an active placebo for anti-depressant research). Furthermore, if treatment success is primarily based on placebo effects, then any number of large or small alterations in the placebo protocol will have little effect on treatment success, provided that these alterations do not tarnish the believability of the treatment.

Kirsch (1990) reports that placebo and expectancy effects might be central to the success of many psychotherapies. In short, most therapies might be successful because clients expect them to be. This is not to suggest that therapies whose successes rest primarily (or solely) on placebo effects should be discounted or removed from the armamentarium of available therapies. Clearly, some placebos are more effective, cost-efficient, and humane than others. And placebo effects that can enhance the more specific effects of certain therapy protocols should be added to those treatments.
The fact that EMDR with eye movements is usually found to be no more effective than EMDR with a placebo replacement for eye movements (such as finger or thumb tapping, listening to music, listening to an auditory stimulus that moves from ear to ear, writing thoughts and feelings, or staring at a point on the wall) might simply mean that all of these efforts engender approximately equivalent non-specific power. Moreover, this shouldn't be surprising as, contextually, these treatments are administered similarly and each eye movement replacement would appear to qualify as an "active" placebo. Conversely, because the alternative treatments with which EMDR has been compared (Vaughan et al., 1994) are dramatically different than EMDR, it should not be surprising that there would be outcome differences among them as they probably engender differing non-specific efficacies.

However, at this time, the underlying mechanism(s) responsible for EMDR's effectiveness have not been verified. This is true in part because the causes for PTSD and almost all other psychological syndromes are themselves unknown. What has been determined (based on the research published through 1993 when the current investigation was conceptualized) is that EMDR might be effective for alleviating or managing many of the self-reported symptoms (as well as potential signs) of PTSD and any number of related disorders. Shapiro's training workshops (1991a, 1991b) have promoted the use of EMDR as an "information processing enhancer" that reduces self-reported emotional reactivity to traumatic memories, thereby allowing the adoption of more adaptive perspectives toward the traumatic material, regardless of etiology. In effect, Shapiro (1991a, 1991b, 1995) makes no practical distinction within EMDR technique or theory between distressing thoughts or images that are symptomatic of PTSD and similarly distressing thoughts and images that result from more common human experiences. According to these guidelines, disturbing levels of emotional
arousal within a continuum of presenting conditions can be treated and desensitized using EMDR.

Therefore, it is both reasonable and conservative to extend EMDR applications to those conditions that present with symptom pictures similar to PTSD and which EMDR is designed to address (Lohr, Tolin, and Kleinknecht, 1996; Maxmen & Ward, 1995). Even though test anxiety (TA) is not a DSM-III-R or DSM-IV diagnostic category, TA appears to be one such psychological condition. The primary diagnostic symptoms of both TA and PTSD are intrusive cognitions, high levels of autonomic arousal, and emission of escape or avoidance behaviors (DSM-III-R, 1987; Brown & Nelson, 1983; Geen, 1985; Hembree, 1988; Tryon, 1980). In addition, both TA and PTSD are characterized by symptoms of worry and apprehension (Barlow, 1990; Sarason, 1978, 1984). When using EMDR, TA and PTSD (as well as similar anxiety based disorders) are conceptualized and treated in similar fashion. Moreover, the utility of EMDR in the treatment of TA hinges neither on identifying the pathognomonic sign for TA nor knowing the functional mechanism(s) of EMDR. Rather, EMDR's efficacy as a symptomatic treatment for TA is presently an empirical issue to be established irrespective of the etiologic or functional issues.

Test Anxiety: Theory, Research, and Treatment

Over the last four decades the term TA has been used to describe a collection of behaviors presumably responsible for decrements in academic performance (Hembree, 1988; Kirkland & Hollandsworth, 1980; Sarason, 1978). During this period there has been widespread disagreement regarding the specific behaviors and, or, circumstances that give rise to or influence the TA condition (Hembree, 1988; Sarason, 1984). Most investigators have contended that emotions are primary (Hembree, 1988). Others have proposed that both emotions and cognitions are involved but that cognitions play the
more central role in performance decrements (Sarason & Sarason, 1990; Tryon, 1980; Wine, 1970). Still others have described TA as the product of both cognitive processing and study skills deficits (Benjamin, McKeachie, Lin, & Holinger, 1981; McKeachie, 1984). Some have even contended that, in and of itself, TA does not diminish academic performance; rather, poor study habits and, or, poor test-taking skills do (Kirkland & Hollandsworth, 1979, 1980). Arguably, the diversity among TA theories has resulted, in part, from the many contradictory and inconsistent findings that have been reported within the vast TA research literature (Brown & Nelson, 1983; Hembree, 1988; Sarason, 1980; Spielberger & Vagg, 1995; Tryon, 1980).

Although there have been numerous explanations for the phenomena associated with the construct of TA, most theories appear to fit within two, broad conceptual models (Hembree, 1988). These include: Interference models and deficit models. Interference model theories were the first theories proposed for TA (Hembree, 1988). In one manner or another, these theories have posited that TA hinders the retrieval of previously learned information, thereby attenuating intellectual performance. The first interference model theory of TA was delineated by Mandler and Sarason (1952) who pioneered early TA studies at Yale University with elementary school children. These authors classified students with high or low levels of TA (using the Test Anxiety Questionnaire—TAQ, Sarason & Mandler, 1952) then administered a standard intelligence test. On the timed subtests of the intelligence test they found that students with low levels of self-reported TA scored higher and more consistently than students with high levels of self-reported TA. Based on these findings, they postulated that test-taking situations evoked two kinds of learned, inversely correlated, psychological drives: Task-directed drives and anxiety drives. Task-directed drives encouraged behaviors that led to the completion of the test-taking assignment and a subsequent reduction in the drive. Anxiety drives evoked two competing sets of behaviors: task-
relevant behaviors, which led to finishing the test and a reduction in anxiety, and task-irrelevant behaviors, which consisted of, "feelings of inadequacy, helplessness, heightened somatic reactions, anticipations of punishment or loss of status and esteem, and implicit attempts to leave the testing situation" (Mandler & Sarason, 1952, p. 166). Task-irrelevant behaviors led to reductions in anxiety via escape of the test situation. The TAQ produced a single scale value, that for debilitating anxiety. Because of the inverse, unidimensional relationship that Mandler and Sarason (1952) posited between the two anxieties, levels of facilitating anxiety were inferred from levels of debilitating anxiety.

The cumulative work of Mandler and Sarason (1952) found that TA usually emerged in second grade, was more prevalent in boys than girls, and tended to worsen from one school year to the next and become increasingly related to measures of academic performance (Hill, 1972). The cumulative work of Mandler and Sarason (1952) and their colleagues also demonstrated that evaluative testing conditions, wherein achievement was emphasized, and negative cognitions, or “failure feedback,” tended to exacerbate TA, while “game-like,” non-demanding test conditions tended to minimize exam related anxiety (Doris & Sarason, 1955; I. G. Sarason, 1958, 1960, 1961; S. B. Sarason et al., 1952, 1960).

Whereas Mander and Sarason (1952) and their colleagues focused on identifying the environmental and academic conditions that provoked or exacerbated test anxiety, McKeachie (1951) was one of the first to experimentally investigate treatments for anxiety related to classroom examinations. McKeachie (1951) reported that test anxious students performed better on multiple-choice examinations when they had the opportunity to write notes about the questions during the examination. McKeachie (1951) believed that this procedure reduced anxiety because it enabled the student to "dispel or channelize some of the tensions built up by tests" (p. 156).
The second major interference model (chronologically) was the bidimensional theory of Alpert and Haber (1960). Unlike Mandler and Sarason (1952), who believed that facilitating and debilitating anxieties were highly correlated in an inverse, unidimensional relationship, Alpert and Haber (1960) hypothesized that facilitating and debilitating anxieties were uncorrelated, independent drives. An individual could possess any degree of facilitating anxiety, which led to task-directed behavior, regardless of their level of debilitating anxiety, which led to task-irrelevant behavior. Because the two anxieties or drives could coexist in any proportion, test related anxieties were seen as a bidimensional condition. Alpert and Haber developed the Anxiety Achievement Test (AAT) which included both facilitating (AAT+) and debilitating (AAT-) subscales.

The work of Liebert and Morris (1967) began a period wherein cognitive rather than emotional influences were seen as the primary causes for diminished test performance. Based on a factor analysis of the TAQ, Liebert and Morris (1967) posited that the debilitating anxiety associated with test conditions actually constituted two separate components: worry and emotionality. These authors described worry as "any cognitive concern about one's own performance" (p. 975). They described emotionality as autonomic or physiological responses such as rapid heartbeat, dry mouth, nausea, dizziness, and perspiration. Liebert and Morris (1967) modified the TAQ to include worry and emotion subscales. The results of their investigations using this instrument suggested that worry diminished academic performance, whereas emotionality appeared to influence performance only in cases where the worry subscale was low.

Wine (1971) continued the emphasis on cognitive influences and suggested an attention theory to explain performance deficits associated with TA. Wine (1971) posited that individuals with high levels of TA spent less time attending to task-relevant
stimuli and more time attending to self-criticisms, worries, and physiological sensations. Hence, with elevated levels of attention spent on test-irrelevant tasks, valuable test-taking time is wasted and performance suffers.

Spielberger (1972) has investigated the nature of anxiety more generally. His research has suggested that there are two basic types of anxiety: A-State and A-Trait anxieties. Spielberger (1972) has described A-State anxiety as relatively brief emotional reactions (e.g., tension, nervousness, and various somatic symptoms) to specific environmental conditions. Conversely, he characterized A-Trait anxiety as a long-standing tendency to respond in an A-State fashion across various environments. Based on the state-trait distinction, Spielberger, Anton, & Bedell (1976) have described TA as a type of A-Trait anxiety wherein academic testing conditions typically elicit strong A-State responses. These emotional responses then give rise to negative cognitions (e.g., worry, fear of negative evaluations, or self-criticisms) which result in diminished academic performance. Therefore, within this conceptualization, cognitive and emotional factors both contribute to levels of self-reported TA and to lowered academic performance. This formulation of TA has become one of the more widely held and enduring of interference model theories (Sarason, 1978; Sarason & Stoops, 1978; Wine, 1982). Sarason (1984) has reported that although a direct relationship between academic performance deficits and emotional arousal has not been established, it cannot be ruled out. Nevertheless, Sarason's (1984) findings have suggested that performance decrements are primarily a function of cognitive influences.

Cognitive theories of TA have evolved, in part, because a functional link between objective performance measures (e.g., quiz scores or course grades) and emotional factors has not been identified. Denny (1974) has estimated that 67% of those TA studies investigating the effects of systematic desensitization failed to find differences between treatment and control conditions on academic performance.
measures. Moreover, numerous studies have been unable to differentially identify those who report high levels of TA from those who report low levels of TA on the basis of various physiological measures assessed during test-like conditions (Deffenbacher & Hazaleus, 1985; Hollandsworth, Glazeski, Kirkland, Jones, & Van Norman, 1979). Cognitive explanations also have gained validity from experiments which demonstrated that academic performance suffered as a result of high levels of task-irrelevant thoughts or cognitive interference during analogue, test-like conditions (Paulman & Kennelly, 1984; Sarason, 1978; Sarason, 1984; Sarason & Stoops, 1978; Tryon, 1980). However, similar to arousal based approaches such as systematic desensitization, treatments spawned by cognitive explanations for TA, such as coping-skills training, have not consistently lowered TA and increased performance (Allen, Elias, & Zlotlow, 1980; Denny, 1980; Hembree, 1988; Kirkland & Hollandsworth, 1980; Tryon, 1980).

The general ineffectiveness of treatment approaches designed to offset the presumed underlying causes of TA (emotional and, or, cognitive) for enhancing academic or cognitive performance (the ultimate dependent criterion) has led some to assert that TA is the result of academic deficits rather than the cause of them. In contrast to interference model theories, deficit model theories have purported that TA might result from inabilities or weaknesses in a number of academically related skill areas (Hembree, 1988). These include: acquisition deficits (i.e., inabilities to learn the material being tested) (Bruch, Juster, & Kaflowitz, 1983), organization/rehearsal or study skill deficits (Bruch, Juster, & Kaflowitz, 1983; Culler & Holahan, 1980; Spielberger, Anton, & Bedell, 1976; Tobias, 1985; Wittmaier, 1972), and deficits in test-taking strategies (Bruch, Juster, & Kaflowitz, 1983; Kirkland & Hollandsworth, 1979, 1980; Tobias, 1985). Some investigators have demonstrated that simply
enhancing test-taking skills can improve academic performance and relieve high TA levels (Kirkland & Hollandsworth, 1980).

However, the effects of study skills counseling alone on both self-reported TA and real-world, academic performance have been inconsistent (Allen, 1971; Cornish & Dilley, 1973; Hembree, 1988; Mitchell & Ng, 1972; Paulman & Kennelly, 1984; Tryon, 1980). Based on these results, and those findings suggesting that coping-skills training has been generally ineffective for improving academic performance, some have posited that TA might constitute dual deficits: concurrent cognitive or information processing and academic skill deficits (Benjamin, McKeachie, Lin, & Holinger, 1981; Lin & McKeachie, 1970; McKeachie, 1984; Meichenbaum & Butler, 1980; Mueller, 1992; Spielberger, Anton, & Bedell, 1976; Tobias, 1978, 1985). Dual deficit theories maintain that neither weak study skills nor excessive TA alone hold the key to successful remediation of poor academic performance traditionally associated with high TA. Rather, as many authors have reported, both study skills enhancement and anxiety-reduction procedures are generally necessary to reduce TA and improve academic performance (Algaze, 1995; Allen, Elias, & Zlotlow, 1980; Benjamin, McKeachie, Lin, & Holinger, 1981; Denny, 1980; Fiore & Pescar, 1987; Vagg & Spielberger, 1995).

The lack of theoretical specificity regarding the etiologies, correlates, and effects of TA might have contributed to the abundance of treatments employed and investigated over the past 40 years of TA research. Moreover, no single theory has yet demonstrated consistent treatment utility. B. Algaze (1995) summarized the TA outcome literature thusly:

On the basis of theoretical expectations, cognitive and emotionality-focused treatments should have different effects on the worry and emotionality components of test anxiety, but these differences have never been demonstrated. (p. 150).
In fact, most investigations have found that nearly any credible, traditional or placebo treatment has achieved greater reductions in self-reported TA than a no-treatment control condition (Hembree, 1988; Tryon, 1980). A non-exhaustive list of treatments that have been found effective for reducing self-reported TA includes: systematic desensitization (Donner & Guerney, 1969), implosive therapy (Cornish & Dilley, 1973), anxiety-management training (Richardson & Suinn, 1974), massed desensitization (Richardson & Suinn, 1974), relaxation (Bedell, 1976), cue-controlled relaxation (Chang-Liang & Denny, 1976), covert reinforcement (Finger & Galassi, 1977), study counseling (Allen, 1971, 1973), cognitive or cognitive-behavioral modification (Meichenbaum, 1972), regular aerobic exercise (Tryon, 1980), vicarious desensitization (Denny, 1974), rational emotive therapy (Fletcher & Spielberger, 1995), biofeedback (Parker, Vagg, & Papsdorf, 1995), computer-assisted treatment (Buglione et al., 1990), attention-skills placebo (Allen, 1971), and pseudo-therapy/group meditation (Holroyd, 1976).

Equivocal outcomes across traditional and placebo treatments suggests the shared actions of non-specific factors such as the therapeutic environment and expectancies for improvement (Frank & Frank, 1991; Ilardi & Craighead, 1994; Kirsch, 1990). Hence, the inability of any singular TA treatment to improve consistently measures of actual performance suggests that, despite numerous claims to the contrary, none of the treatments investigated have included components whose therapeutic efficacy actually derive from a therapeutic mechanism, accepted theory, or treatment protocol. Alternatively, various treatments for TA might possess specific active components. Attempting to enhance poor academic performance by way of alleviating high levels of self-reported TA might be applying sound treatments to the wrong problem. The inconsistent connection between TA reduction and academic
performance suggests that the construct of TA does not possess strong utility for the remediation of poor academic performance (Kirkland & Hollandsworth, 1980).

However, an exhaustive meta-analysis of 562 TA outcome studies by Hembree (1988) suggests that the inability for TA treatments to lower both TA and raise academic performance consistently might be an interpretational artifact of the extant research. Most studies that have investigated the effects of various treatments for TA on self-reported levels of TA and academic performance have featured small treatment and control group sizes with high within-group variability. Hence, it is plausible that most studies have not had the statistical power to detect between-group mean differences of a given magnitude (in this case, approximately 0.5 standard deviations). Hembree (1988) reported:

Improved test performance and GPA consistently accompany TA reduction. This finding differs with the conclusions of prior reviews (e.g., Allen, Elias, & Zlotlow, 1980). Their findings issued from the dearth of studies that showed significant performance differences between treated and untreated subjects. Significant is the critical word; the treatment studies need to be probed for their ability to detect performance differences usually found between TA levels...an improvement of about 6 points (on a 100 point scale) should be expected as a result of TA treatment. For alpha = 0.05 and a pooled standard deviation of 12, a 6-point difference requires experimental and control group sample sizes in the neighborhood of 30 before significance can be observed. Of 137 treatment studies, only 16% had samples larger than 20; one third used samples of 10, 9, or 8. Apparently, these small samples could not detect the significance of performance differences near 6 points. Of 120 performance comparisons, 107 showed higher test scores for the treated subjects. In the aggregate, these effects possess a highly significant mean. (p. 73)

Hembree’s (1988) conclusions stand in sharp contrast to those drawn by many TA researchers over the past 15 to 20 years (cf., Sarason & Sarason, 1990; Tryon, 1980). Hembree (1988) concluded that purely behavioral treatments, designed to remediate (only) the emotional component of TA, were significantly more effective for reducing both the emotional (autonomic) and cognitive (worry) components of TA than purely cognitive treatments, which were designed to remediate (only) the cognitive component of TA. This finding implies that emotionality triggers worry and that, at
least in part, emotionality diminishes performance, as opposed to the converse position held by deficit theorists. This finding also suggests that TA is a unidimensional, behavioral construct. That is, worry might be a byproduct of emotional arousal in the same sense that diminished performance appears to be (Hembree, 1988).

Considered at face value, Hembree's (1988) meta-analysis cuts through some of the confusion extant in the TA literature research. After four decades and 562 TA research studies we can now say with some confidence that TA does, in fact, impair intellectual performance. We can also say that, despite dozens of assertions to the contrary, of the two primary components of TA, emotional arousal and worry, emotional arousal appears to be the primary, operative component.

However, what do we know about the possible contributions to the emotional component of TA? The literature has suggested a number of variables that might determine, correlate with, or result from an individual's TA. A non-exhaustive, non-mutually exclusive list of these variables includes: emotionality and physiological symptoms (Sarason, 1984), intrusive thoughts (Sarason, 1984), worry (Sarason, 1984), tension (Sarason, 1984), the type of TA (whether facilitative or debilitative) (Alpert & Haber, 1960), fear of negative evaluation (Brown & Nelson, 1983), self-deprecatory thoughts (Wine, 1971), poor attentional skills (Hunsley, 1985), excessive familial pressures to succeed (Test Anxiety Reduction--Curriculum Guide, 1987), unrealistically high standards (Brown & Nelson, 1983), escape and/or avoidance of tests (Geen, 1987), inadequate study and test taking skills (Hunsley, 1985; Paulman & Kennelly, 1984), unrealistically high study standards (i.e., have to "know it all" in order to feel prepared) (Brown & Nelson, 1983), excessive procrastination (Kirkland & Hollandsworth, 1980), lack of academic preparation (Klinger, 1984), history of poor test-taking performance (Klinger, 1984), gender differences (Best & Stanford, 1983), self-efficacy for good performance (Klinger, 1984), self-efficacy for controlling...
negative thoughts and anxiety (Brown & Nelson, 1983), the perceived difficulty and importance of the test (Sarason & Sarason, 1990), and individual coping skills for anxiety (Kirkland & Hollandsworth, 1980).

The fact that most studies have grouped subjects together based solely on their self-reported levels of TA, that is, irrespective of a wide range of potential etiologies and histories, might have masked significant population differences (Paulman and Kennelly, 1984). Perhaps the syndrome of TA is comprised of many disorders or subgroups (Meichenbaum & Butler, 1980). This might partially account for the lack of effective standardized treatments. Symptomatic treatments for TA might have differential effectiveness resulting from differing etiologies. As Paulman and Kennelly (1984) state: "Knowledge of how skill levels combine with TA in mediating cognitive performance would hold direct implications for remedial intervention" (p. 280).

Moreover, it is certainly not the case that all students who experience TA perform poorly (Brown & Nelson, 1983). Some students who report high TA levels perform quite well. Conversely, it is certainly not the case that all students who perform poorly experience TA. Some students appear not to be bothered by their relatively weak academic performance.

Despite the collective power of the Hembree (1988) meta-analysis, the validity of its conclusions rests on the soundness of the methodology of each contributing study. There are a number of reasons to suspect patterns of error within the existing research. These error patterns might be due to: (a) the emphasis on correlational research designs that might (falsely) lead to cause and effect interpretations; (b) the emphasis on group designs that might obscure important individual differences and introduce extraneous variables; (c) the nature of the constructs investigated and the imprecision of many assessment devices; (d) the complexity and types of treatments investigated for TA that reduce construct validity; (e) the lack of one standardized TA
assessment that prevents direct comparisons between many studies and suggests that results might depend on which device you employ; and (f) demand characteristics of the experimental environment (other than those related to TA) that might influence subjects' responses to TA assessment devices. Each of these error patterns will now be considered in turn.

The first plausible reason for error across TA studies is the emphasis on statistically analyzed, correlational research designs wherein cause and effect relationships (whether true or false) might be readily implicated. In this fashion correlates might be turned into causes. If, for example, 100 studies find similar correlation values between some aspect of TA and a measure of academic performance (using similar research designs), does this imply a causal relationship between the variables under examination (cf., Stanovich, 1992)? Although the answer is obvious, in practice this detail is often overlooked. The theories that sprout from the results of correlational research designs sometimes make headlines, sway public opinion, and often determine the direction for future research. In a sense, they might function as self-fulfilling prophecies. They might also take on greater significance when one renowned expert quotes another (cf., Gilovich, 1991).

Another plausible error pattern within the existing research is the emphasis on group designs that compare outcome measures for treatment and control groups that might obscure the character and magnitude of treatment effects at the individual level (cf., Kazdin, 1982). Hembree (1988) reported that, when considered individually, most outcome studies found decreases in TA did not lead to statistically significant improvements in academic performance. However, when the results of those same 120 outcome studies were pooled and evaluated via meta-analysis, the aggregate mean difference in objective measures of academic performance between treated and untreated groups was highly significant. The average amount of improvement was 6%. Despite
the importance this differential performance increment might have (e.g., for institutional service planning and large-scale program development), averaging performance across hundreds of subjects fails to acknowledge that some subjects' performances increased substantially more than 6% when treated for symptoms of TA, whereas others' performances might have either remained constant or deteriorated. What the Hembree (1988) meta-analysis explains is that most previous research studies have used group sizes too small (groups of at least 30 were required) to detect average group academic performance improvements (that presumably resulted from reductions in TA) as statistically significant. What the Hembree (1988) analysis and extant research have failed to describe are the differences between those clients for whom treatment has lowered TA and improved academic performance and those clients for whom treatment lowered TA and did not improve academic performance. In other words, individual successes (and the variables or treatment processes that contribute to these successes) have not been discriminated from individual failures (cf., Gendlin, 1986). Moreover, Hembree (1988) does not specify whether the individual studies controlled for possible ceiling effects. That is, how many studies have treated individuals who exhibited both high levels of TA and near-perfect, pre-treatment academic performance over which very little or no improvement was possible?

The lack of clarity and specificity regarding individual responding to treatments for TA is not surprising considering that most research studies have selected subjects solely on the basis of high levels of self-reported TA, ignoring other, sometimes obvious factors that might contribute to reported levels of TA (Brown & Nelson, 1983; Klinger, 1984; Paulman & Kennelly, 1984). This type of unidimensional selection procedure automatically includes subjects that represent a wide variety of differing TA etiologies and levels of academic functioning. Ignoring these possible confounding influences (i.e., variables that could differentially influence academic performance
and/or co-vary with any other variable or measure—refer to the list on p. 21 of this report) when selecting subjects could partially explain the inconsistency of academic improvement associated with reductions in TA reported in the literature over the last 40+ years. It could also help explain the large levels of variance generally found between individuals within treatment conditions, and the necessity of using treatment groups of at least n=30 in order to find the average differences between groups statistically significant.

In order to offset the influence of confounding variables (that gross variances in performance appear to represent), most studies have selected subjects randomly from a pool of individuals assessed (via some instrument) as being high in TA, then randomly placed these individuals in one of the treatment (or no treatment) groups (Klinger, 1984). In this fashion, no one confounding variable should have been over-represented in any one of the experimental conditions. In order to demonstrate the relative symmetry among treatment groups, experimenters have frequently performed a pre-study analysis of variance, and, typically, no significant differences have been found. However, symmetry among groups can only be validated for a specific variable when an experimenter assesses for that variable. Therefore, experimenters might have failed to notice or misinterpreted the effects of many intervening variables. Also, with smaller treatment groups, as were typically used in most studies, random assignment to groups is substantially less effective for offsetting the effects of any one confounding variable. In summary, treating all subjects in a group in like fashion might gloss over many individual differences that are critical to the ultimate understanding and treatment of TA at the individual level. In their summary of the outcome literature, Brown and Nelson (1983) state: “From a practical standpoint, these findings provide the counselor with few guidelines for tailoring treatments to individual client needs” (p. 367).
Patterns of error in the existing research also might be caused by the nature of the constructs examined and how they are assessed. Many studies have examined the relationship between self-reports of TA and other, hard to assess "cognitive" constructs such as cognitive interference. Strong correlations between measures of two such constructs might indicate the presence of a cause and effect relationship (i.e., either variable might have caused the other, or some third variable might have influenced the two measured variables). However, a strong correlation might be the result of significant survey item overlap. That is, strong, positive correlations between measures generated by different assessment instruments might be due to those instruments measuring the same construct (Tryon, 1980).

Another plausible source of error across TA research studies arises from the complexity of many of the treatments for TA. Many outcome studies have compared the effectiveness of two or more complex treatments and invariably failed to identify the specific components responsible for observed outcomes. Tryon (1980) states that these types of studies lack construct validity. Furthermore, the treatments compared in these studies are sometimes described as essentially behavioral or essentially cognitive (Hembree, 1988; Tryon, 1980). The distinction between those treatment elements that are behavioral and those that are cognitive, although useful for heuristic purposes, is arbitrary. That is, the distinctions that these descriptors yield might bear no relationship to the differing backgrounds and current needs of individuals experiencing TA.

Error patterns in the extant TA studies might be due to the lack of a standardized TA assessment instrument. The lack of an accepted, standard instrument makes interpretations or conclusions across studies problematic. There are at least eight different major TA assessment devices (Sarason, 1984; Tryon, 1980). Results measured by one might differ dramatically from those measured by any other. Tryon (1980) has reported:
Even though correlations among the scales are high, they are rarely high enough to account for most of the variance between scales. For example, the Suinn Test Anxiety Behavior Scale and The Test Anxiety Scale correlate .60, which accounts for only 36 percent of the variance. (p. 346)

To confuse matters further, there are also multiple versions of some of the instruments (Tryon, 1980). Sarason and his colleagues (1978) developed and used three versions of the Test Anxiety Scale. However, for many of their studies, they did not stipulate which version was used (Spielberger, Anton, & Bedell, 1976). At least three different versions of the Anxiety Achievement Test have been used for research. Using the three AAT versions to examine levels of self-reported TA within a single experimental group, Huck and Jacko (1974) reported statistically significant mean differences between the three scales. Clearly, the lack of a standardized assessment instrument constrains us to view meta-conclusions regarding the relationship between the construct of TA and academic performance with caution.

Patterns of error in the existing research might be the result of subjects' responses to self-report TA instruments falling under the control of variables other than those pertaining to TA. This problem is not unique to the assessment of TA. Subjects' responses to any self-report device might be affected by the demand characteristics of the experiment. Allen (1970) demonstrated how subjects' self-reported TA changed with different instructions. For this experiment, subjects' responses to the TAS and AAT were compared under standard and role-play conditions. Under role-play conditions subjects were instructed to respond to the items of the TAS and AAT as if they were a student about to take an important exam on which they did not anticipate doing well. Under standard conditions the subjects were told to respond to the two scales based on their own TA. Allen (1970) reported that under the role-play condition, TA scores were significantly higher than under the standard condition. Based on these results it appears that subjects' responses to TA assessment instruments might vary as a
function of the demands put upon them by the social, clinical, and/or experimental environment. In as much as previous studies have not assessed or controlled for the effects of demand variables, it is plausible that they have differentially affected the outcomes of TA research.

Subsequent to both the Hembree (1988) meta-analysis and the formulation of the current study, Spielberger and Vagg (1995) have offered a "Transactional Process Model" for TA. This model has incorporated and interwoven the perspectives of many previous theories. Similar to state-trait theory, the Transactional Process Model holds that TA is a situation-specific anxiety trait that predisposes an individual to high levels of S-Anxiety under evaluative conditions such as tests. Similar to various cognitive and deficit theories it considers the full range of antecedent and consequent variables such as study- and test-taking skills as well as intrusive, test-irrelevant thoughts.

Spielberger and Vagg (1995) described the Model's structure and function:

The Transactional Process Model...is intended as a heuristic framework for representing the antecedent conditions and dispositions that influence students' reactions to tests, the mediating emotional and cognitive processes involved in responding to evaluative situations, and the correlates and consequences of TA. This model provides a cross-sectional analysis of TA phenomena as a situation-specific dynamic process in which examinations and other evaluative situations evoke mediating affective states and task-irrelevant cognitions that have important behavioral consequences. (p. 11)

Spielberger and Vagg (1995) have designed the Transactional Process Model to accommodate flexibly the complex, intrapersonal, idiopathic nature of the TA condition. Using the Transactional Process Model's framework, all measurable components and correlates of TA can be analyzed and evaluated, regardless of their theoretical or treatment orientation. Moreover, the Model provides a basis for comparing the treatment utility of differing theories: The impact of multiple treatment approaches on any of the Model's elements can be compared along a cognitive-emotionality continuum.
Spielberger and Vagg (1995) have employed the Transactional Process Model to evaluate the outcomes of five controlled TA investigations (Algaze, 1995; Gonzalez, 1995; Parker, Vagg, & Papsdorf, 1995; Spielberger & Fletcher, 1995; Vagg & Papsdorf, 1995). Cumulatively, their analysis has suggested that cognitively-oriented treatments are more effective than emotionality-oriented treatments for reducing test anxiety. These results stand in clear contrast with those reached by Hembree (1988). Conversely, they are in keeping with the majority of TA literature. However, Spielberger & Vagg (1995) consistently found that the most effective treatments among these five studies were those that employed both cognitive- and emotionality-focused components and also provided opportunities for students to discuss and practice their newly developed coping mechanisms.

In summary, extensive individual research efforts investigating the nature and treatment of TA have been unable to delineate clear and consistent relationships between the apparent components of TA (emotionality and worry) and measures of academic or cognitive performance. An exhaustive meta-analysis (Hembree, 1988) of the research extant has suggested that depressed academic or cognitive performance is largely a function of emotional arousal. However, various error patterns and inconsistencies within the body of TA research render conclusions based on the Hembree (1988) meta-analysis tentative at best. In addition, more recent research has challenged Hembree’s (1988) findings and suggested that cognitive factors are more salient and influential in the TA condition than emotional factors (Spielberger & Vagg, 1995).

In conclusion, TA remains a poorly understood syndrome (or group of syndromes) for which there is neither one or more pathognomonic markers nor a DSM-IV diagnostic category. Using the DSM-IV (1994) diagnostic decision tree for anxiety disorders (pp. 698-699), the diagnostic categories that seems to fit or describe TA most accurately are Social Phobia (Social Anxiety Disorder) and Specific Phobia. The DSM-
IV describes Social Phobia as "fear of humiliation or embarrassment in social or performance situations" (p. 698). The DSM-IV describes Specific Phobia as "fear cued by object or situation" (p. 698). It is plausible that a response pattern described as TA could evolve from a severe traumatic event such as in PTSD. For example, a parent who physically abuses a child contingent on unacceptable academic performance. Regardless of etiological influences, however, the symptoms of TA appear to overlap with a number of DSM-IV anxiety disorders including Social Phobia, Specific Phobia, Panic Disorder, and Post-traumatic Stress Disorder. Given this lack of diagnostic precision, it is not surprising that there is no standard "best practice" assessment procedure or treatment regimen for TA by virtue of consistent demonstrations of effectiveness, generality, and economy.

The findings of the cumulative TA research literature reviewed for the current study yield a number of implications for future TA treatment and research. These include: (a) In order to optimize the quality of care provided to individuals who suffer from excessive TA, maximally efficient and effective symptomatic treatment approaches for TA need to be designed and evaluated; (b) in order to reduce self-reported levels of TA and concurrently improve academic performance treatments for TA should address both cognitive and emotional issues; and (c) in order to understand better the various etiologies responsible for TA that have previously been collapsed into a single category the relationship between academic variables such as study habits and attitudes and TA needs to be investigated.

EMDR appears to be a treatment protocol that could efficiently alleviate the symptoms of TA. EMDR comprises both emotional and cognitive components. Similar to systematic desensitization, the most common procedure used for treating the emotional component of TA, (Hembree, 1988; Spielberger et al., 1976; Spielberger & Vagg, 1987), EMDR is designed to systematically diminish emotional reactivity to
feared or anxiety inducing stimuli. Similar to various cognitive therapies that have been employed in the treatment of TA, EMDR would attempt to construct and enhance more adaptive attributions, beliefs, and expectations regarding evaluative, academic situations. Similar to another cognitive treatment approach for TA, anxiety-coping training, EMDR employs imaginal rehearsal, which has been shown to diminish TA levels and enhance the efficacy of desensitization procedures (Algaze, 1995; Allen, 1971; Spielberger & Vagg, 1995). Moreover, any behavior or environment associated with the TA condition, whether long past or current, could be addressed and treated via the protocol's desensitization, cognitive restructuring, and imaginal rehearsal components.

The exact nature of what EMDR contributes to the treatment of TA over and above the current standard approaches is an empirical issue yet to be addressed. However, as discussed previously in the current text, the mechanism(s) through which EMDR achieves positive, symptomatic effects are not understood at this time. Just as EMDR does not treat the cause(s) of PTSD, it would not treat the cause(s) of TA. Rather, EMDR would provide symptom management. Moreover, the primary outcome measures used to infer EMDR's effectiveness for treating PTSD have been self-reported symptoms. Reports of symptoms are verbal behavior, and verbal behavior is usually multiply determined. Given the commonality between the symptoms of PTSD and TA, the EMDR protocol might change the nature of social contingencies controlling the verbal responses describing TA symptomatology as it appears to change the social contingencies controlling verbal responses describing PTSD symptomatology.

Syndromes Vs. Diseases; Symptoms Vs. Signs

Now that the EMDR and TA literatures have been reviewed and the basis for the current study's rationale has been established, the focus shifts to examine a number of
issues pertinent to nearly all treatment outcome studies. In the following sections a number of issues will be covered that are typically assumed or ignored in most studies. However, as shall be discussed, understanding many of these issues might be vital toward the development of maximally effective, empirically valid clinical interventions.

The terms syndrome and disease refer to two conceptually distinct constructs. Likewise, the terms symptom and sign refer to two conceptually distinct constructs. Although clinicians often blur these pairs of terms in common practice, researchers need to maintain the clarity provided by these distinctions.

A syndrome is defined as: "(in med.) a concurrence of symptoms, fr. syn, together, + dromos, a running. The aggregate of signs and symptoms associated with any morbid process, and constituting together the picture of the disease" (Stedman, 1982, p. 1382). The DSM-IV (1994, p. 771) defines a syndrome as: "A grouping of signs and symptoms, based on their frequent co-occurrence, that may suggest a common underlying pathogenesis, course, familial pattern, or treatment selection." By contrast, a disease is defined as an: "...entity characterized usually by at least two of these criteria: a recognized etiologic agent (or agents), an identifiable group of signs and symptoms, or consistent anatomical alterations" (Stedman, 1982, p. 403). In brief, then, a syndrome is a collection of symptoms and signs with no known etiology or cause, whereas a disease is a collection of symptoms and signs with a known etiology or cause.

Webster's Second International Unabridged Dictionary (1944) has defined sign (in reference to medicine) as:

An objective evidence of disease; that is, one appreciable by some one other than the patient. The terms symptom and sign are often used synonymously, but properly they differ in that the former is perceived only by the patient. The term sign is often further restricted to the purely local evidences of disease afforded by direct examination of the organs involved, as distinguished from evidences of general disturbance afforded by the temperature, pulse, etc., and is then often called physical sign. (p. 2334)
In brief, then, a symptom is a subjective verbal report by a client, whereas a sign is an objective indicator confirmed by a therapist (or other examiner).

Historically, scientists have identified syndromes as the first step toward identifying specific diseases (Magner, 1992). Being able to diagnose reliably a collection of symptoms and signs as marking a definable and stable syndrome has allowed researchers to identify homogeneous groups of subjects for research purposes (Hayes & Follette, 1994; Kiple, 1993). Further research has often focused on identifying the (often multiple) etiologic factors that caused the appearance of such symptoms and signs as was the case prior to the discovery of typhus and typhoid fever as separate disease entities (Wilson, 1993). A syndrome, thus, usually comprises more than one disease process. Typically, the various causes for the different disease processes in a syndrome are identified one factor (or disease) at a time. As Kiple (1993) has stated: "The twentieth-century war on contagious disease has seen the identification of microbe after microbe, the subsequent control through vaccination of disease after disease in the developed world..." (p. 375).

For example, in the mid nineteenth-century anyone presenting with certain signs and symptoms of congestive respiratory distress and a progressively deteriorating (wasting) condition might have been diagnosed with "consumption" (Hafner, 1970, p. 122). The disease process for this syndrome was isolated by German physician Robert Koch (1843-1910) who identified the tubercle bacillus bacteria as the cause of what would subsequently be diagnosed as tuberculosis (Magner, 1992). Other specific disease processes that have emerged from the consumption diagnosis include fungus of the lung, arthritis of the lung, and cystic fibrosis of the lung (Wilson, 1993). Today one no longer diagnoses the syndrome of "consumption," because the ability to diagnose the several specific disease processes that comprise consumption has made the syndrome obsolete.
Likewise, "febrile syndrome" was the diagnosis given to anyone in the mid-nineteenth century who presented with signs of fever (Pellini, 1993; Wilson, 1993). Today, the febrile syndrome has yielded (through research) to a large number of specific disease processes, some bacterial, some viral, as well as other causes such as heat prostration, thyroid dysfunction, and reactions to medications all of which can cause a fever.

In modern medical practice, diseases are commonly diagnosed when a pathognomonic sign for that disease is identified. Pathognomonic is defined as "Pathognostic; characteristic or indicative of a disease; denoting especially one or more typical symptoms" (Stedman, 1982, p. 1040). In other words, a pathognomonic sign is that sign decisively characteristic of a specific, underlying etiologic or disease process.

Identifying the cause of a disease can lead to specific and strategic treatment (i.e., a "cure") if a way can be determined to eliminate the cause found (e.g., streptomycin kills tubercle bacillus bacteria) (Magner, 1992). However, not having an effective strategic treatment for a disease of known cause (e.g., measles), or not knowing the cause of a syndrome (e.g., Bipolar Disorder) does not mean that no treatment can provide relief for patient distress. Most syndromes, as well as all diseases for which there are no available strategic treatments (e.g., tuberculosis prior to streptomycin), are treated "symptomatically." For example, people with colds take medications to alleviate the symptoms of their colds (e.g., antihistamine for "runny" nose; aspirin for aches and fever reduction). Their immune systems do the work of fighting the causes of their distress--viral infections.

The causes for physical diseases almost always involve biologic factors. However, most psychopathology, the manifestations of mental or behavior disorders, is multiply determined by a combination of biologic and environmental influences.
(Maxmen & Ward, 1995). This is sometimes referred to as the "stress-diathesis" model of psychological disorders. Individually, neither a biological predisposition (e.g., a gene) nor a precipitating environmental stressor are necessary and sufficient to cause most mental disorders. Rather, predispositions and environmental events interact to produce the conditions necessary and sufficient for mental disorders. Maxmen and Ward (1995) have described the interaction between environmental stressors and biologic predispositions:

Having a gene, and having it produce visible effects, are two different things. The gene itself is a genotype; whether the gene is expressed—that is, if it becomes manifest—is a phenotype. If there is a gene (or genotype) for an ulcer, whether it develops into an ulcer (the phenotype) depends on environmental factors such as diet, stress, and smoking. Because of this interrelationship, to ask whether a mental disorder is caused by genes or environment is naive. When a person inherits the genotype of a mental disorder, how the environment affects that gene determines the degree, form, and existence of the disorder. (p. 68)

Whereas the biological causes for many infectious diseases are known, the unique admixture of biologic and environmental causes for most psychological disorders are not known; they are idiopathic. The difficulty in identifying pathognomonic causes or etiologies for mental disorders is due, in part, to the complexity of environmental and predispositional influences. As Maxmen and Ward (1995) have stated:

To speak of "the cause" of a mental disorder is naive; it is akin to patients who enter treatment and say, "I want to get to the root of my problem." This "root myth" assumes that a single factor has caused the problem and that unearthing it will result in cure. Unfortunately, there is no such thing as the root to a problem; at most, there are roots, with many sprouting, variegated phenomena. Formulations, therefore, should consider diverse etiological influences...(p. 58)

Maxmen & Ward (1995) suggest that etiology should refer not only to the origins of a disorder, but also to its pathogenesis—all factors involved in the disorder’s production. They delineate four dimensions or levels of causation which clinicians should consider when determining causation. These include: predisposition, initiation, perpetuation,
and exacerbation. For some predisposed individuals what initiates a disorder (loss of job leads to depression) might be very different from what perpetuates it (low rates of rewarding social activity due to lack of income) or exacerbates it (the holiday season with no disposable income and few friends).

Maxmen and Ward (1995) define two general classes of causal predispositions, or latent conditions, that increase an individual’s susceptibility to certain disorders or diseases. The first class of predispositions is “disorder-specific” and involve conditions such as a gene or physical disease that results in a vulnerability to a certain disorder. The second class of predispositions is “disorder-nonspecific” and involve conditions such as having an alcoholic parent or being assaulted which can result in vulnerability to any number of disorders.

Arguably, the etiologies of most mental disorders are complex and indeterminable. In Appendix C of the DSM-III-R (1987), syndrome has been defined as follows:

A group of symptoms that occur together and that constitute a recognizable condition. “Syndrome” is less specific than “disorder” or “disease.” The term disease generally implies a specific etiology or pathophysiologic process. In DSM-III-R most of the disorders are, in fact, syndromes. (p. 405)

Hence, (biologic) pathognomonic signs and, or, environmental pathogeneses cannot be identified for most DSM-III-R and DSM-IV disorders. Moreover, for most of these disorders there are no strategic or curative treatments. The exceptions are few and mostly involve etiologies that are obvious and inherent to the conditions they create such as alcohol or amphetamines in alcohol or amphetamine intoxication (Maxmen & Ward, 1995).

Nevertheless, effective symptomatic treatments have been validated for many of the syndromes in the DSM-III-R and DSM-IV. The signs and symptoms of the syndrome Major Depression are often successfully ameliorated with antidepressant
medications (Hyman, 1995) and/or cognitive-behavioral therapy (Beck, Rush, Shaw, & Emery, 1979). The positive signs and symptoms of the syndrome Schizophrenia (and related psychotic syndromes) are often minimized or eliminated using antipsychotic drugs such as Halperidol, Thorazine, Prolinxin, and Respirdal (Hyman, 1995). The signs and symptoms of the syndrome Bipolar Disorder are virtually eliminated in the majority of cases when treated with a salt of lithium (e.g., lithium carbonate) (Baastrup & Shou, 1967; Hyman, 1995; Kline, 1969).

As it pertains to the current study, TA is a syndrome for which there is no known pathogenesis or pathognomonic sign and for which there is no cure (Spielberger & Vagg, 1995). There are, however, numerous symptomatic treatments that have been somewhat effective for reducing the self-reported symptoms of TA. A few of the more popular treatments have included systematic desensitization (Hembree, 1988), anxiety-management training (Richardson & Suinn, 1974), cognitive therapy (Fletcher & Spielberger, 1995), and study skills counseling (Holroyd, 1976).

Confusing Effective Treatment with Identifying the Cause

By definition symptomatic treatments are effective for reasons other than offsetting the mechanism(s) responsible for the disorder. However, based on the therapeutic effects of medical or psychological treatments, many clinicians have inferred unobservable, psychological or biological etiologies (Stuart, 1970). Such interpretations of medical or psychological treatment effects are grossly misleading and empirically indefensible. Nevertheless, it has not been uncommon for mental health researchers to espouse causal theories for mental disorders inferred or deduced from the active mechanisms of effective medications. Two prominent examples include: The dopamine theory for schizophrenia that was inferred from the successful treatment of schizophrenic symptoms by neuroleptic medications such as chlorpromazine that
blocked dopamine D2 receptors in the brain (Spiegel & Aebi, 1983). And, the seratonin theory for depression that was inferred from the successful treatment of depressive symptoms by medications such as Prozac and Zoloft (SSRIs) that blocked seratonin system re-uptake in the brain (Spiegel & Aebi, 1983).

The DSM’s atheoretical, descriptive structure also might have contributed to the confusion between medical treatment effects and causal inferences. All DSM’s have de-emphasized environmental etiologies and influences to such an extent that psychopathology appears to reside within the individual (Kratochwill & McGivern, 1996). The DSM approach has emphasized the form of the psychopathology and correspondingly de-emphasized its environmental impact and function. Hence, individuals appear to have mental disorders in much the same way that individuals have diseases as described by the medical model. The disorder, like the disease, is conceptualized in a disembodied, contextless fashion, in isolation of environmental causes or functions. The DSM nosology suggests that there is some final common organic pathway—some ultimate disease process to which any number of functional, environmental roads might lead and for which effective treatment requires medicine.

Medications are, after all, the only effective treatment modality that the DSM-III, DSM-III-R, or DSM-IV specify differentially and reliably. However, as Maxmen and Ward (1995, p.6) point out, “DSM-IV classifies mental disorders; it does not classify individuals with mental disorders.” A mental disorder is only one facet of the person who has it. Identifying the topography of an individual’s disorder neither describes the whole person nor explains the function that the disorder serves in the individual’s environment. Nor, for that matter, do the medications which the DSM differentially prescribes usually resolve the patient’s underlying problems or issues (Kirsch, 1990). Maxmen and Ward (1995) have observed:
In general, biological therapies eliminate or alleviate symptoms, such as a depressive's insomnia, a schizophrenic's delusions, or a manic's spending sprees: psychosocial therapies usually address issues, such as coping with job losses, a failed marriage, medical condition or psychiatric disorder. When used properly, biological and psychosocial therapies do not impede, but facilitate, one another...Correctly medicated patients gain more from psychotherapy...Conversely, psychotherapies can accomplish what drugs cannot. (pp. 82-83)

Explicating symptoms allows the clinician to explain a disorder, topographically, and prescribe appropriate medicine to control the symptoms. Only by explicating the pertinent issues, the meanings that underlie the symptoms, can clinicians understand a disorder, empathize with the person who has it, and assist them to resolve their issues and ultimately the problems they present. Maxmen and Ward (1995) have indicated:

Issues contain ideas; symptoms do not. However, just as mail transmits ideas, symptoms can express issues. The ideas patients communicate when they manifest symptoms reflect the issues concerning them. For instance, two men with the severe mental disorder of major depression may have the symptom of overwhelming hopelessness. The first, an elderly, devout Catholic, insists he is hopeless “because I sinned 20 years ago by cheating on my wife.” The second, a young up-and-coming actor, feels equally hopeless because “I’m not a star.” Both patients have psychopathology—that is, hopelessness—yet, since their issues are so different, each expresses hopelessness quite differently. (p. 7)

Whereas the appropriate medications that the Catholic and the actor might receive easily could be the same (e.g., Prozac), very likely each would require different psychotherapeutic strategies to address effectively their unique sets of issues.

This analysis of the relative contributions of chemo- and psychosocial therapies in the treatment of mental disorders suggests that neither treatment modality is effective for remediating all the signs and symptoms of most disorders. Hence, in most all cases, neither modality represents a strategic treatment or “cure.” For many disorders medications and psychosocial therapies team up to form the most effective symptomatic treatment (Maxmen & Ward, 1995).
The tendency to misconstrue treatment effectiveness as offsetting or remediating the cause of a disorder (and thereby identifying the cause of the disorder) might be exacerbated by the general tendency in Western culture to confuse syndromes and diseases, symptomatic treatment with strategic, curative treatment. Moreover, this might be fostered by the belief that doctors “cure” people when they give them medicines; quite often signs and symptoms do diminish following a course of symptomatic treatment. Antihistamines, however, do not cure the common cold. Individuals also might draw premature conclusions from such pairings of treatment administration and diminished signs and symptoms because of our culture’s general tendency to identify and assign causes when we observe temporal relationships between events (Gilovich, 1992). Unfortunately, without scientifically controlled experimental conditions, often this tendency leads to spurious assumptions regarding the relationships among variables. We observe what amounts to a correlation between one event (treatment) and another (improved health) and erroneously deduce not only that the former controlled the latter, but the manner by which it did so (offsetting the cause).

In a similar manner, a therapist’s assumption that therapeutic changes in their client’s mental health is the result of the prescribed treatment regimen might be specious (Stuart, 1970). Improvements perceived following psychotherapeutic treatment might be due to a host of variables, none of which involve the specific mechanisms or actions of the treatment. As Myerson (1939) poignantly observed:

The neuroses are ‘cured’ by Christian Science, osteopathy, chiropractic, nux vomica and bromides, benzedrine sulfate, change of scene, a blow on the head, and psychoanalysis, which probably means that none of these has yet established its real worth in the matter...Moreover, since many neuroses are self-limited, anyone who spends two years with a patient gets credit for the operation of nature (Myerson, 1939, p. 641)
Frank (1961) has observed that the majority of "neurotic" patients improve immediately from psychotherapy regardless of the type or nature of therapy. Stuart (1970) has offered three general explanations for these equivocal results. First, seemingly heterogeneous treatments might all be effective for reasons that have nothing to do with their claimed powers. Rather, their effectiveness might involve basic characteristics that they share. As previously discussed in the *EMDR--Summary of Experimental Research* section of the current text, these shared characteristics involve aspects of the therapeutic process and environment that are nearly universal among psychotherapies and produce what are usually described as non-specific, expectancy, or placebo effects (J. D. Frank & J. B. Frank, 1991; Ilardi & Craighead, 1994; Kirsch, 1990).

The second explanation that Stuart (1970) offers for how seemingly heterogeneous treatments might all be equally effective is: the simple passage of time and the non-therapy events that are, nonetheless, therapeutic. Some common factors unrelated to treatment that nonetheless might be responsible for therapeutic improvement could include (although are certainly not limited to) the following: natural healing processes, enhanced coping abilities, the cyclic nature of the disease or disorder (such that milder periods naturally follow more severe periods, hence, regression toward mean functioning levels), expectations for positive outcomes, improvements in the individual's environment such as better housing, improved diet, improved sleep (new bed!), obtaining a better job, establishing more desirable close relationships, buying a better automobile, going on vacation, or terminating some type of stressful relationship or condition.

Regardless of the specific reasons why, the severity or acuity of many medical and mental disorders naturally appears to abate over time (Myerson, 1939). The evidence for such abatement is usually inferred from changes in self-reported scores on
repeated administrations of normative assessment instruments; the individual’s follow-up or post-treatment score is somewhat improved over their initial or pre-treatment score. However, scores are just scores, not directly observed measures of actual change. Presuming no active involvement in psychotherapy, regression toward mean values might occur because of therapeutic, non-therapy events (as described above—real events that have real impact), or might reflect the unreliability or instability of the assessment device itself (Alessi, 1982). In the latter case, scores change because the individual has guessed or estimated differently on certain assessment items even though real world conditions have remained substantively the same. Nunnally (1978) has described the regression principle generically as follows:

Obtained scores are biased estimates of true scores. Scores above the mean are biased upward, and scores below the mean are biased downward. The farther scores are in either direction from the mean of obtained scores, the more, in an absolute sense, scores are biased. As a group, people with high obtained scores have a preponderance of positive errors of measurement, and the opposite is true for people who have low obtained scores. (p. 217)

The extent to which scores can be expected to regress toward their assessment instrument’s mean value between administrations depends upon the reliability coefficient for the instrument and the distance that the scores fall from the mean (Salvia & Ysseldyke, 1978). The lower the reliability coefficient of the instrument, the higher the error component in the obtained score, the greater the variability in scores between administrations. The more distal an individual’s obtained score is from an assessment instrument’s normative mean, the greater the impact that the device’s reliability coefficient has on that score (a greater percentage of the obtained score is error, or achieved by random variables). Hence, when reliability coefficients are low and scores are distal from the mean, more regression toward mean values on subsequent administrations should be expected. Conversely, when reliability coefficients are high and scores are proximal to mean values, very little regression toward mean values on
subsequent administrations should be expected. Regression affects are most potent in research groups selected by cutoff scores well below the mean, using pre-post test designs rather than a control group design. Failure to control for statistically predictable regression effects is probably a major reason that so many educational studies conducted in the 1950’s, 1960’s, and 1970’s, using only pre- and post-test designs, produced statistically significant treatment effects that subsequently could not be replicated in real-world clinical practice (Travers, 1982). Many of these studies concluded that their treatments were effective, and often what they were measuring were the inherent effects of regression. Adequate control group experimental designs would have controlled for such regression effects.

The third explanation that Stuart (1970) offers for how seemingly heterogeneous treatments might all be equally effective is: The subjective nature of outcome assessments and the lack of objective, consensually valid, change criteria. What constitutes meaningful, clinical improvement? What changes in self-report, third-party report, in-session behavior, norm-referenced scores, rating scales, or any other assessment modality are indicative of substantive, real-world differences? For that matter, what degree of what measure constitutes a clinical problem and by whose standards? Given the lack of accepted standards, frequently clinical researchers apply their own, subjective values of change when evaluating experimental subjects. As a result, all manner of treatments might appear effective, and the risks of measuring the researcher’s personal or theoretical biases, rather than actual changes in the patient, appear substantial. Moreover, without clear-cut, objective measures of change and with the motivation of demonstrating a treatment’s effectiveness, misinterpreting the true effects of causes and overlooking the real effects of regression also might be more probable.
This analysis raises serious doubts about the actual, therapeutic effectiveness of many traditional psychotherapies as well as the potential treatment utility of the DSM diagnostic system. If a number of disorders improve with just about any treatment, then what specific therapeutic utility do those treatments have? Further, given that a variety of mental disorders (e.g., all types of depressions, anxieties, and personality disorders) have been treated effectively with a limited number of psychosocial treatments (i.e., various cognitive and cognitive-behavioral therapies), then of what treatment utility can the DSM nosology be for differentially identifying effective treatments (Hayes & Follette, 1992)?

Post-Traumatic Stress Disorder

DSM-III-R Diagnostic Criteria: Signs and Symptoms

Given that one of the current study’s rationales for administering EMDR to the problem of TA were the similarities between TA and PTSD, it is then reasonable to examine the PTSD syndrome more closely. An analysis, similar to that which follows, would have been presented for TA as well. However, TA is not a diagnostic category within either the DSM-III-R or DSM-IV.

All standardized, diagnostic nomenclature used in the current study was taken from the DSM-III-R. During the course of the current study the DSM-IV was published. To establish how the current study’s rationale, organization, and conclusions vis-a-vis the DSM-III-R might be affected by changes introduced in the DSM-IV, all pertinent, substantive differences between the language of the DSM-III-R and that of DSM-IV are noted when a reference is made to the DSM-III-R (portions of the DSM-IV that differ from the DSM-III-R are printed in italics).
Post Traumatic Stress Disorder (PTSD) is a syndrome, diagnosed when the following presenting criteria are met (DSM-III-R, 1987; DSM-IV, 1994):

**DSM-III-R**

A. The person has experienced an event that is outside the range of usual human experience and that would be markedly distressing to almost anyone.

**DSM-IV**

A. The person has been exposed to a traumatic event in which both of the following were present:
   1. the person experienced, witnessed, or was confronted with an event or events that involved actual or threatened death or serious injury, or a threat to the physical integrity of self or others.
   2. the person's response involved intense fear, helplessness, or horror.

**DSM-III-R**

B. The traumatic event is persistently re-experienced in at least one of the following ways:
   1. recurrent and intrusive distressing recollections of the event.
   2. recurrent distressing dreams of the event.
   3. sudden acting or feeling as if the traumatic event were recurring (includes a sense of reliving the experience, illusions, hallucinations, and dissociative flashback episodes, including those that occur on awakening or when intoxicated).
   4. intense psychological distress at exposure to events that symbolize or resemble an aspect of the traumatic event.

**DSM-IV**

B. The traumatic event is persistently re-experienced in at least one of the following ways:
   1. recurrent and intrusive distressing recollections of the event, including images, thoughts, or perceptions.
   4. intense psychological distress at exposure to internal or external cues that symbolize or resemble an aspect of the traumatic event.
   5. physiological reactivity on exposure to internal or external cues that symbolize or resemble an aspect of the traumatic event.

**DSM-III-R**

C. Persistent avoidance of stimuli associated with the trauma or numbing of general responsiveness (not present before the trauma), as indicated by at least three of the following:
   1. efforts to avoid thoughts or feelings associated with the trauma.
   2. efforts to avoid activities or situations that arouse recollections of the trauma.
   3. inability to recall an important aspect of the trauma (psychogenic amnesia).
   4. markedly diminished interest in significant activities.
   5. feeling of detachment or estrangement from others.
   6. restricted range of affect, e.g., unable to have loving feelings.
(7) sense of a foreshortened future, e.g., does not expect to have a career, marriage, or children, or a long life.

DSM-IV-R
C. Persistent avoidance of stimuli associated with the trauma or numbing of general responsiveness (not present before the trauma), as indicated by at least three of the following:
(1) efforts to avoid thoughts, feelings, or conversations associated with the trauma.
(2) efforts to avoid activities, places, or people that arouse recollections of the trauma.
(3) markedly diminished interest or participation in significant activities.

DSM-III-R
D. Persistent symptoms of increased arousal (not present before the trauma), as indicated by at least two of the following:
(1) difficulty falling or staying asleep.
(2) irritability or outbursts of anger.
(3) difficulty concentrating.
(4) hypervigilance.
(5) exaggerated startle response.
(6) physiologic reactivity upon exposure to events that symbolize or resemble an aspect of the traumatic event. (p. 250)

DSM-IV
D. Identical to DSM-III-R except that point 6 is omitted.

DSM-III-R
E. Duration of the disturbance (symptoms in B, C, and D) of at least one month.

DSM-IV
E. Identical to DSM-III-R.

DSM-III-R
F. None (pp 250-251)

DSM-IV
F. The disturbance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning. (pp 427-429)

Overlap Between Diagnostic Categories

As can be seen from the above, the DSM-III-R and DSM-IV criteria for PTSD consist almost entirely of (self-reported) symptoms. There are potential signs (objective criteria) for the disorder. However, none of these qualify as definitive,
pathognomonic signs that indicate the etiological processes underlying PTSD. For a diagnostic sign to be a pathognomonic indicator for a syndrome or disease, its presence must indicate the existence of one and only one diagnostic category (Stuart, 1970). Table 1 lists the (potential) signs and symptoms that, based on DSM-III-R diagnostic criteria, might overlap or coexist among PTSD, Simple Phobias, Obsessive-Compulsive Disorder (OCD), Generalized Anxiety Disorder (GAD), Organic Mental Disorder, and Major Depression. As Table 1 clearly displays, there is significant overlap among the (potential) signs and symptoms of PTSD and other DSM-III-R diagnostic categories. Each (potential) sign and symptom reported for PTSD appears, or is implicated, in at least one other DSM-III-R syndrome and is therefore predictive of more than one syndrome.

The DSM-III-R (1987) lists the primary diagnostic criterion for PTSD as "an event that is outside the range of usual human experience that would be markedly distressing to almost anyone" (p. 250). In describing the "predisposing factors" for Multiple Personality Disorder (MPD; now relabeled Dissociative Identity Disorder in the DSM-IV), the DSM-III-R (1987) states: "Several studies indicate that in nearly all cases, the disorder has been preceded by reported abuse (often sexual) or another form of severe emotional trauma in childhood" (p. 271). Braun (1993) conceptualized that MPD is a special instance of PTSD. He believes that the unifying concept between the two syndromes is dissociation as the central response reaction to trauma. Braun (1993) states: "The understanding of PTSD and MPD begins with an awareness that severe psychological trauma is the underlying etiology, and that maladaptive use of dissociation is the underlying pathology" (p. 5).

If PTSD and MPD are different names for similar conditions, perhaps they share similar pathognomonic signs. The biophysical response of most humans to trauma appears to be quite similar. Van der Kolk and Saporta (1993) have reported:
Table 1
Shared Symptomatology Among Post-traumatic Stress Disorder (PTSD), Simple Phobias, Obsessive-Compulsive Disorder (OCD), Generalized Anxiety Disorder (GAD), Organic Mental Disorder, and Major Depression

<table>
<thead>
<tr>
<th>Syndromes</th>
<th>Avoidance</th>
<th>Irritability</th>
<th>Difficulty Concentrating</th>
<th>Restricted Range of Affect</th>
<th>Loss of Interest in Activities</th>
<th>Sleep Difficulties</th>
<th>Memory Dysfunction</th>
<th>Hyper-vigilance</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTSD</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Simple Phobias</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OCD</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAD</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Organic Mental Disorder</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Major Depression</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
The human response to trauma is so constant across traumatic stimuli that it is safe to say that the central nervous system (CNS) seems to react to any overwhelming, threatening, and uncontrollable experience in quite a consistent pattern. Regardless of the circumstances, traumatized people are prone to have intrusive memories of elements of the trauma, to have a poor tolerance for arousal, to respond to stress in an all-or-nothing way, and to feel emotionally numb. (p. 26)

However, how can trauma be an essential aspect of etiology if trauma appears to have differential effects on behavior among individuals? For instance, most combat veterans do not develop PTSD, whereas some non-combat veterans do develop PTSD (Foa, 1993). Hence, it would seem that, by definition, combat is not "traumatic."

**Diagnosis by Client Self-report**

The reliable differential diagnosis of PTSD (or any other DSM-III-R diagnostic category) is made more difficult by the fact that most of the potential signs, such as "markedly diminished interest in significant activities," or "exaggerated startle response" are observed infrequently by a clinician in a treatment session (Foa, 1993). Rather, these events or conditions are reported to the clinician by the client and therefore would be categorized as (self-reported) symptoms.

Self-reports are verbal behavior, and verbal behavior is usually multiply controlled and highly "plastic" within different contexts (cf., Skinner, 1957, Chapter 9). It is important for clinicians and researchers to be aware of the kind of behavior that they are assessing. Most all of the (dependent) behavior that psychologists evaluate and measure, either in direct contact treatment sessions, or in clinical research studies, is verbal behavior. And verbal reports describing behavior and behavior itself often are controlled by substantially different environmental conditions. Across many contexts there is little empirical support that suggests essential agreement between behavior and verbal reports of the same behavior (Azar, 1997; Frank & Frank, 1991; Loftus, 1994; Patterson, 1982). Yet, as parents, friends, teachers, and professionals it appears that
humans routinely expect such agreement. As Alessi (1988, pp. 15-16) has pointed out, it is a rational model of human behavior that predicts behavior and verbal reports describing the same behavior will agree. Alternatively, a behavior analytic model of human behavior (Skinner, 1957) would anticipate differing content between behavior and verbal reports of behavior on the basis of the inherently evanescent controlling conditions between these events. Such variables as conditional discriminative stimuli, motive and emotional states, reinforcement and punishment histories, and current response contingencies impact behavior pragmatically, lawfully, and differentially across settings and occasions (Michael, 1993).

The anxiety related connotations of topographically similar self-reports can vary widely from person to person (Marks, 1987). For example, three individuals report experiencing fear when crossing bridges; should they each be diagnosed with Agoraphobia? The first person reports that she becomes afraid when she thinks about how far away from land she is, and that she is trapped on the bridge. The second person reports feeling afraid when she remembers a car crash that she witnessed on a bridge and the horror of the carnage. In fact, she reports that the slightest deviation in the traffic flow around her car while she is driving on a bridge now evokes intense fear. The third person reports becoming intensely afraid while driving across a bridge when she remembers a picture of a collapsed bridge in a newspaper. Now, she reports that just the sight of a newspaper lying in the car sometimes causes feelings of fear.

How can these diverse and complex responses be understood, analyzed, and diagnosed? Unlike most animals for which the rules of stimulus generalization are relatively well established, humans have the ability to create abstract, symbolic representations of any (overt or covert) event that might then function to elicit self-reported symptoms of anxiety (Marks, 1987). This tendency can make evaluations of self-reported events highly imprecise and unreliable. Marks (1987) has stated:
Undoubtedly mediational or symbolic cues can become linked to anxiety. A serious problem is that anything can in theory be a symbol for anything else. Given inspiration we could juxtapose almost any two things in our minds and find a link between them, although some links will seem more plausible than others. Strict criteria are therefore needed to exclude the chance that the symbolism or attribution ascribed to a given phobic or obsessive-compulsive worry is not merely a Rorschach test of the investigator. Such criteria for distinguishing causal association from post-hoc rationalization are very hard to establish. (p. 256).

In summary, given the lack of specificity among the pertinent DSM-III-R diagnostic categories and the plasticity, complexity, and imprecision of verbal behavior in humans, it does not appear possible to differentially diagnose PTSD with a high degree of reliability. If this is true, how can a diagnosis of PTSD predict that a treatment appropriate for the individual will be administered? Moreover, by what mechanism(s) are treatments for PTSD such as exposure effective? (Cooper & Clum, 1989). As has already been discussed in the current paper, treatments for syndromes such as PTSD are symptomatic treatments. Therefore, it is not necessary to identify precisely an individual's psychopathology and, or, the mechanisms underlying the psychopathology to prescribe effective treatment. A treatment merely must be suitable for the client it is intended to serve and the symptoms it is intended to offset (Kirsch, 1990; Maxmen & Ward, 1995). Likewise, treatment utility does not necessitate the identification of underlying therapeutic mechanisms. Given the probability that most psychosocial treatments incorporate sufficient non-specific elements that yield significant, broad-based therapeutic power, no strategic treatment component is required to achieve statistical and clinically significant improvements (Frank & Frank, 1992; Ilardi & Craighead, 1994; Kirsch, 1990).
Statistical vs. Clinical Significance

Psychological treatments often change the self-reports of the individuals receiving treatment. When a group of individuals participating in a psychological study report improvements of a certain average magnitude between pre- and post-treatment assessments, those pre-post mean differences achieve statistical significance. That is, the probability of attaining results of that certain magnitude by chance is sufficiently low that an active influence is suggested, presumably the experimental intervention.

However, just because obtained changes are minimally probable bears no relationship with their real-world importance, their clinical significance (Jacobson & Truax, 1991). As the issue is sometimes stated, “Does the difference make a difference?” Do the differences that the treatment has (presumably) facilitated suggest a return to normal functioning? Have the issues that initiated therapy been resolved? All too often statistically significant results on standardized self-report measures across groups of individuals have not led to clinically significant behavior changes at the individual level (Goldfried, Greenberg, & Marmar, 1990; Jacobson, Wilson, & Tupper, 1988). As Jacobson & Truax (1991) have stated:

Conventional statistical comparisons between groups tell us very little about the efficacy of psychotherapy...Consumers, clinicians, and researchers all expect psychotherapy to accomplish particular goals, and it is the extent to which psychotherapy succeeds in accomplishing these goals that determines whether or not it is effective or beneficial. (p. 12)

In order to estimate the extent to which psychotherapy accomplishes the meaningful goals described and desired by the individuals receiving and, or, impacted by the treatment, Jacobson and Truax (1991) have developed objective, standardized methodologies for evaluating clinical significance. In establishing these methodologies they have attempted to accomplish three essential goals: (1) to define parameters of clinical significance that were universally applicable across settings, persons, and
disorders, (2) to define parameters of clinical significance that were consistent with both professional and lay understandings and expectations of change resulting from psychotherapy, and (3) to establish specific criteria by which individual clients could be judged as clinically changed or unchanged.

In essence, Jacobson and Truax (1991) have defined clinical significance as the treatment-induced shift of dependent values from a dysfunctional to a functional distribution. They have explicated three different methods for operationalizing this definition. Each method is suggested for specific experimental conditions. The three methods included (Jacobson & Truax, 1991):

(a) The level of functioning subsequent to therapy should fall outside the range of the dysfunctional population, where range is defined as extending to two standard deviations beyond (in the direction of functionality) the mean for that population.

(b) The level of functioning subsequent to therapy should fall within the range of the functional or normal population, where range is defined as within two standard deviations of the mean of that population.

(c) The level of functioning subsequent to therapy places that client closer to the mean of the functional population than it does to the mean of the dysfunctional population (p. 13)

Jacobson and Truax (1991) have recommended method a for establishing the clinical significance critical value when no normative data are available for the dependent measures. When normative data are available, they have recommended adopting the critical values described by either methods b or c. When normative data are available and functional and dysfunctional populations do not overlap they have recommended method b. When normative data are available and functional and dysfunctional populations do overlap, they have prescribed using method c.

All three methods evaluate the relative position of a client's score vis-a-vis functional and dysfunctional populations. All three methods estimate what Ankuta and Abeles (1993) have described as "meaningful change" (p. 72) in pre- to post-treatment
101

scores. That is, a post-treatment score has satisfied the condition of meaningful change when, by virtue of its proximity to the functional population, the score is more likely to have been sampled from the functional rather than the dysfunctional distribution.

Although these three methods estimate a score’s relative pre- and post-treatment positions, none of them measure how much a score has changed between pre and post assessments (Jacobson & Truax, 1991). This limitation has little impact on methods a or b. When using either of these methods for establishing clinical significance, knowing the magnitude of an individual’s improvement is generally not crucial. In both of these cases, clinical significance inherently reflects relatively substantial pre to post movement. For a score to reach clinical significance using method a the post-treatment score must be at least two standard deviations distal from the pre-treatment dysfunctional group mean. Using method b the post-treatment score must move from within the original dysfunctional distribution to within two standard deviations of the non-overlapping functional distribution (Jacobson and Truax, 1991). Again, this generally involves substantial change.

However, where functional and dysfunctional populations overlap, sole reliance on the method c position criterion could identify clinically significant changes that are too small to achieve statistical significance. To guard against this possibility, Jacobson and Truax (1991) have suggested using an additional procedure to verify that the magnitude of improvement exceeds what would be expected on the basis of chance alone (i.e., regression toward the mean effects). Jacobson and Truax (1991) have referred to this procedure as the “reliable change index” (RC, p. 14). This procedure calculates a standard error of difference between the pre and post assessment scores and compares this with actual pre to post changes. When the pre to post difference exceeds what would be expected on the basis of chance (p < .05), the magnitude of change is statistically reliable; thus, it is unlikely that the measured change is not valid. The
accepted critical RC value, which the obtained pre to post difference must exceed in order to attain statistical significance at the .05 level, has typically been 1.96. In summary, when using method c, both the relative position (or "meaningful change") and reliable change (RC) conditions must be satisfied in order to achieve clinical significance.

Because clinically significant outcome criteria are more stringent than their statistical counterparts (Ankuta & Abeles, 1993), comparisons between clinically and statistically significant outcomes typically have made the psychotherapy under scrutiny appear less effective (Ankuta & Abeles, 1993; Jacobson & Truax, 1991; Jacobson, Wilson, & Tupper, 1988). Some number of subjects in these studies have reported positive changes which have not achieved clinical significance. However, this loss of apparent effectiveness has been offset by the identification of some subjects for whom the treatment protocol has appeared to be unequivocally effective. Ankuta and Abeles (1993) have reported that clinically effective outcomes are differentially associated with increased levels of client satisfaction. Moreover, these authors have reported that practitioners find clinically significant data more useful for purposes of effective treatment planning. When clinical significance is assessed "the variability in the data is made clearly visible rather than camouflaged in group effects" (Ankuta & Abeles, 1993, p. 73). Hence, for individual treatment applications, and for comprehending both treatment successes and failures, clinically significant results might provide more pertinent, specific guidance than statistically averaged group data (Ankuta & Abeles, 1993; Jacobson & Truax, 1991; Jacobson, Wilson, & Tupper, 1988).

There are, however, at least four unsolved problems associated with clinical significance methodologies (Jacobson & Truax, 1991). First, it is not known whether functional and dysfunctional populations are distributed normally and how violations of distribution normality will impact clinical significance calculations. All formulas
presented by these authors assume that the two populations are normally distributed. Second, operationalizing clinical significance as the return to "normal" functioning might be inappropriate for the psychotherapeutic treatment of some mental disorders. It is patently unrealistic to expect the symptoms of schizophrenia, bipolar disorder, or dementia (among other severe mental disorders) to be completely alleviated by any known form of psychotherapy. Third, the paucity of normative psychometric values for many commonly used outcome measures deters the use and dissemination of clinical significance assessments. Without normative dependent data, standardized cutoff points cannot be determined, rendering the procedures inoperative.

The fourth, and perhaps most important problem with clinical significance procedures, is that they are unvalidated statistical manipulations or abstractions. Clinical significance statistics that describe an individual’s progress in treatment might be useful tools in guiding the treatment protocol and determining when treatment should be terminated. However, similar to the SUD measure used in EMDR, clinical significance itself cannot predict when and, or, to what extent the problem or behavior of concern has been alleviated in vivo. The critical cutoff points currently used to determine clinical significance have yet to be validated against more reliable measures of treatment outcome, recidivism rates, or quality of life measures. Are outcomes that attain clinical significance indicative of "real-world" success? The evidence for the answer to this and related questions has yet to be documented.

Despite these limitations, there has been strong interest in and support for the development and use of clinically significant criteria (Goldfried, Greenberg, & Marmar, 1990; Jacobson, 1988). Jacobson and Truax (1991) have contended that the procedures are sufficiently developed that it is not unreasonable to expect most researchers to report on the clinical significance of their data. Moreover, widespread use of clinical significance analyses might help to standardize applications and
methodologies. For instance, Jacobson and Truax (1991) have suggested that norms should be established based on aggregated samples among a large number of similar studies as this would help to eliminate inter-study cutoff point differences. Nevertheless, these authors also have advised that our current understanding of clinical significance analyses vis-a-vis specific treatments, dependent variables, and, or mental disorders is inadequately comprehensive to prescribe the exact method or strategy by which such analyses should be conducted.

Purpose of the Current Study

Now that we have established an understanding of the current problem's nature (TA is a syndrome as are most mental disorders), the nature of the proposed treatment (EMDR is a symptomatic approach as are all others for PTSD or TA), the potential, shared, non-specific effects of all psychotherapies (that obviate the necessity of therapeutically active mechanisms), the nature of most dependent measures (self-report), the limitations of self-reports (they often differ from the behavior they describe), the imprecision and uncertainty of most diagnostic categories (including TA), and the potentially misleading nature of between-group statistical analyses (masking crucial individual variability), the purpose of the current study will be explicated. By way of review: EMDR was selected for the treatment of TA because of its brief protocol, flexibility across presenting issues, appropriateness for treatment of the symptoms of TA (owing to the similarity between the symptoms of TA and PTSD), ability to address the major components of TA—maladaptive emotionality and cognitions—that many previous TA treatment strategies have targeted, emphasis on improved future performance (using imaginal rehearsal), and potentially swift remedial action. From a generality standpoint, it was hoped that the current study would extend
the demonstrated effectiveness of the EMDR protocol for clinical conditions other than PTSD.

Many previous EMDR investigations have employed case study methodologies or designs with inadequate experimental controls. The current study has employed a no-treatment waiting control group which provided a sound baseline with which to contrast the effects of EMDR treatment. Moreover, following the initial no-treatment waiting control period, EMDR was administered to the control group, thus replicating the primary experimental objective. Unlike many previous EMDR studies, diagnostic validity of the to be treated condition was assessed at both pre- and post-treatment evaluations. Moreover, standardized diagnostic measures were employed, as opposed to SUD or VoC measures.

The current investigation made no attempt to “dismantle” the EMDR procedure or evaluate its individual components. Rather, the current project represented an efficacy study using the entire EMDR protocol. As explained previously, the current study demonstrated the efficacy of EMDR via the comparison of concurrent treatment and no-treatment control conditions. To estimate the “real world” validity of the results, both statistical and clinical significances were calculated and presented. The author believed that the current study is the first TA study and one of the few EMDR studies to examine the clinical significance of its treatment effects.

Much of the previous EMDR research has been flawed or limited by the use of artificial treatment guidelines (e.g., administering only seven sets of eye movements or a single session of EMDR regardless of individual needs). This study has redressed this issue by applying iterative eye movements to the extent that they are needed to help individuals reprocess distressing material. The number of saccades used for each set of eye movements was not determined by a preset or fixed number, rather depended on
individual client needs and the nature of the imaginal task. The number of sessions was also flexible although it was capped at a maximum of 10.

Some have argued that TA is just another name for poor study habits, and/or inadequate study and test-taking skills (Kirkland & Hollandsworth, 1979, 1980; Paulman & Kennelly, 1984). As previously discussed in this text, the contention is that an individual might very well become anxious while taking a test if he or she is unprepared and does not comprehend the materials being tested, and or waits until the last minute to begin studying. This TA condition however, might be quite different from one in which an individual possesses at least average study and test-taking skills, studies diligently, and still experiences high levels of debilitating TA. Moreover, failing to control for these academic variables and subsequently treating all subjects in a group in like fashion (as if their symptoms of TA had identical etiologies) might gloss over individual differences that are critical to the ultimate understanding and treatment of TA at the individual level. In consideration of these potential academic confounds, inclusion criteria for the current study involved both high levels of TA and at least average (self-reported) study habits and attitudes. Whereas academic factors influencing TA could not be ruled out as insignificant, for the subjects in the current study they were expected to be minimal.

In the current study, TA, the primary dependent variable, was measured using two instruments:

1. The three scales of the Test Anxiety Inventory (TAI; Spielberger, 1978): The Test Anxiety Total Scale (TAI-T), the Test Anxiety Emotion Scale (TAI-E), and the Test Anxiety Worry Scale (TAI-W).

2. A cross-validation measure was obtained from the Reactions to Tests Total Scale (RTT-T; Sarason, 1984).
Hypotheses Tested by the Current Study

In summary, this study focused on three primary hypotheses or experimental questions. They were as follows:

1. Would EMDR be more effective for treating test anxiety than spontaneous recovery that occurs during a waiting control condition as measured by the TAI-T, TAI-E, TAI-W, or RTT-T in a population of college students whose TA is not the result of obvious academic factors?

2. Is the effectiveness of EMDR, as observed in the primary treatment group, replicated when applied to the waiting control group subsequent to a waiting control period as measured by the TAI-T, TAI-E, TAI-W, or RTT-T?

3. Is there spontaneous remission of test anxiety during a waiting control period as measured by the TAI-T, TAI-E, TAI-W, or RTT-T?
CHAPTER II

METHOD

Subjects

Subjects were male and female college students selected from a large midwestern university and a large midwestern junior college. Thirty-two subjects, 11 males and 21 females, were identified in pairs matched on the basis of date of initial qualifying assessment. The members of each pair were randomly assigned to two groups. There were 16 subjects in each group, the EMDR treatment group (Group A) and the no-treatment waiting control group (Group B). There were six males in Group A and five males in Group B. Subjects were recruited via bulletin board postings, ads in the college newspaper, and screenings at a university academic assistance center.

All subjects selected for the current study granted their informed consent to participate (see Appendix B). All subjects reported excessive TA according to their scores on any of the three TAI scales. All subjects also reported at least mean levels of either study habits or study attitudes as measured by the Survey of Study Habits and Attitudes (SSHA; Brown & Holtzman, 1967). In excess of 200 students were screened in order to identify 32 subjects who met the investigation’s criteria. Four individuals qualified for the study and ultimately declined to participate due to personal reasons.

Design

The experimental design employed an EMDR treatment condition (Group A) and concurrent no-treatment waiting control condition (Group B) with subsequent
replication of the EMDR treatment condition in Group B. Thirty-two subjects were selected in matched pairs based on their initial assessment date on a clinical trials basis. When two individuals who completed the assessment questionnaires within approximately one week of each other met the design criteria, and both agreed to participate in the study, they were randomly assigned to either Group A or Group B. Hence, the time between initial screening and mutual acceptance to participate in the study ranged from approximately 24 hours to one week. For most yoked subject pairs mutual agreement to participate was obtained within 48 hours of initial screening.

The 16 Group A individuals were scheduled for immediate EMDR treatment sessions. The 16 Group B individuals were informed that the treatment schedule was currently full and that they would be contacted within a few weeks to schedule their first treatment sessions. When the Group A subject completed treatment, the Group B subject was contacted to arrange for post-wait re-assessment and their first treatment session. On occasion, when treatment for the Group A individual was lengthy, the Group B individual was contacted by phone and reassured that their treatment would start soon.

Yoking the onset and offset of measurement between each member in the treatment group and the corresponding paired member in the no-treatment waiting control group was vital in that it minimized the confounding influences involved with time passage, regression toward the mean, and other effects unrelated to the EMDR protocol. In the current study, regression toward the mean effects were especially likely due to the restricted range of the subjects’ TA scores (at initial screening all subjects were above the eighty-fifth percentile on the TAI-T). Therefore, it was considered essential that the experimental design permit the influence of such confounds to be estimated and their effects on outcome data to be minimized.
Yoking the assessments within each subject pair also permitted the estimation of effects unrelated to the EMDR protocol over the course of the waiting control period. The waiting period, however, was not standard for all subjects. Rather, the waiting control condition for each subject in Group B was equivalent to the EMDR treatment period for the yoked subject in Group A. Because this was an efficacy study wherein EMDR would be applied flexibly as determined by individual goals, needs, and schedules, treatment periods were expected to vary substantially across subjects. Thus, the current design did not (could not) specify a standard fixed length for the waiting control period and subsequently conclusions regarding the effects of a given wait period on a group of individuals could not be drawn.

The use of a no-treatment waiting control condition against which to contrast the effects of the EMDR treatment condition served the purpose of controlling for various confounding influences as follows: In Group A, all therapeutic effects unrelated to the EMDR protocol were active within the treatment period and could not be partialed out or measured. Hence, all therapeutic effects, both treatment specific and unrelated to treatment, were included within Group A’s mean differences. By contrast, in Group B, regression toward the mean and all other therapeutic effects unrelated to the EMDR protocol, such as the passage of time, were observed in isolation of treatment effects. Hence, only the effects unrelated to the EMDR protocol were included within Group B’s initial screening to post-wait mean differences. Assuming the equivalence of the two groups at initial screening, which was likely as both Groups were assessed at approximately the same time and subjects were randomly assigned, any changes in Group B’s dependent scores during the no-treatment waiting control condition would estimate the confounding effects unrelated to the EMDR protocol subsumed within Group A’s initial screening to post-treatment mean differences.
Therefore, the extent to which the effects observed in Group A (during EMDR treatment) exceeded the effects observed in Group B (during the equivalent waiting interval) yielded an estimate of EMDR effects alone (Question 1). For each dependent measure of TA (TAI-T, TAI-E, TAI-W, RTT-T), the magnitude of this difference was analyzed statistically using a two-step process. First, a two factor, repeated measures analysis of variance (ANOVA; 2 groups X 2 assessment occasions) determined the presence of significant contrasts among the four combinations of the two factors (Group A/initial screening; Group A/post-treatment; Group B/initial screening; Group B/post-wait). Second, contingent on a significant F statistic, a post-hoc, two-sample t-test was performed that compared Group A’s post-treatment mean with Group B’s post-wait mean. These analyses, which derived from and addressed Question 1, required a minimum of four, two factor, repeated measures ANOVAs and as many as four, post-hoc, two sample t-tests.

Evaluating the re-administration of EMDR treatment (Question 2) in Group B was accomplished in three steps for each of the four TA dependent measures (TAI-T, TAI-E, TAI-W, RTT-T). First, a one factor ANOVA (1 group X 3 assessment occasions) determined the presence of significant contrasts among Group B’s initial screening, post-wait, and post-treatment assessments. Next, contingent on a significant F statistic, a post-hoc, paired t-test that contrasted Group B’s post-wait and post-treatment means was conducted. The results of this analysis provided a relatively unconfounded or unbiased estimate of EMDR’s efficacy as it automatically excluded, and hence controlled for, the effects unrelated to EMDR treatment observed during the initial, no-treatment waiting control condition. Finally, contingent on a significant F statistic, the equivalence of EMDR treatment efficacy across groups was analyzed by comparing the post-treatment means for Groups A and B via a post-hoc, two-sample t-test. These analyses, which derived from and addressed Question 2, required a
minimum of four, one factor, repeated measures ANOVAs and as many as eight post-hoc t-tests: four paired t-tests and four, two-sample t-tests.

The degree to which each dependent measure of TA (TAI-T, TAI-E, TAI-W, RTT-T) spontaneously remitted or was effected by variables unrelated to the EMDR protocol (Question 3) was analyzed in two steps. First, a one factor ANOVA (1 group X 3 assessment occasions) determined the presence of significant contrasts among Group B’s initial screening, post-wait, and post-treatment assessments. Then, contingent on a significant F statistic, a post-hoc, paired t-test that contrasted Group B’s initial screening and post-wait means was conducted. These analyses, which derived from and addressed Question 3, required four, one factor, repeated measures ANOVAs (the same as used for Question 2) and as many as four, post-hoc, paired t-tests.

In summary, the current experimental design provided the mechanisms by which to address all three primary hypotheses. The incorporation of treatment and no-treatment waiting control conditions, random assignment to those conditions, and synchronized assessment occasions delimited the influence of effects unrelated to the EMDR protocol, and thus yielded a more accurate picture of EMDR’s efficacy with the current problem and population. Moreover, this design enhanced the internal validity (Kazdin, 1992) of the current study’s two essential experimental conditions: EMDR in Group A contrasted with waiting in Group B, and the re-application of EMDR in Group B.

Statistics

For all occasions wherein mean comparisons were analyzed using t-statistics, simultaneous sets of t-tests (cf., Bonferoni t procedure) were conducted on the mean differences of all four dependent variables for that occasion. As explained above, in the
current study there were four such occasions that involved a total of 16 t-tests. Organizing the statistical analyses into sets of simultaneous t-tests served the primary purpose of establishing critical p-values against which to compare the obtained p-values of the individual t-statistics within the set. Moreover, this approach was considered to be the most conservative and, or, stringent of the alternative analytic methods examined which thereby minimized the chances of type I errors. Comparisonwise p-values were found by dividing the critical p-value for the simultaneous set, which functioned as the experimentwise p-value, by the number of t-tests within the set (four in all cases). The critical p-value for each simultaneous set, which determined the probability of a type I error within the entire set of t-tests, was set at p<.05 based on two-tailed t-tests. Therefore, the comparisonwise p-value for a simultaneous set with four t-tests was .05/4 = .0125. That is, to reach statistical significance, the probability of obtaining results of a given magnitude by chance, if the specific null hypothesis was true, had to be less than .0025.

All 16 subjects in Group A (EMDR treatment) completed both the initial screening and post-treatment assessments. All 16 subjects in Group B (no-treatment control) completed both the initial screening and post-wait assessments. Unfortunately, due to unacceptably long initial wait periods (that followed initial screening), only 12 subjects in Group B followed through with treatment and completed the post-treatment assessment. Therefore, allowing for one degree of freedom (n-1), the obtained t-statistics for mean comparisons between Group A’s post-treatment TA scores and Group B’s post-wait TA scores were based on 15 subjects (16-1). Likewise, the obtained t-statistics for mean comparisons between Group B’s initial screening TA scores and Group B’s post-wait TA scores were based on 15 subjects (16-1). However, the obtained t-statistics for mean comparisons between Group A’s post-treatment TA scores and Group B’s post-treatment TA scores were based on 11 and 15
subjects (12-1; 16-1). Finally, the obtained t-statistics for mean comparisons between Group B's post-wait TA scores and Group B's post-treatment TA scores were based on 11 subjects (12-1).

Clinical Significance

For the individual subjects in Groups A and B the clinical significance of TAI-T, TAI-E, and TAI-W outcomes was estimated using method c and the reliable change index (RC) concurrently (Jacobson & Truax, 1991; discussed in the Statistical vs. Clinical Significance section of the current text; refer to Appendix B for formulas). Ankuta and Abeles (1993) have described the conditions that these procedures evaluate as “meaningful change” and “reliable change” respectively (p. 72). Both conditions must be satisfied for outcomes to achieve clinical significance. Ankuta and Abeles (1993) have employed method c and the RC procedure concurrently when using a different assessment device with psychometric properties similar to those of the TAI.

Method c for calculating clinical significance stipulated the following: “The level of functioning subsequent to therapy places that client closer to the mean of the functional population than it does to the mean of the dysfunctional population” (Jacobson & Truax, 1991, p. 13). Method c was selected for assessing the clinical significance of the TAI and its subscales because normative data were available for these measures (hence, relative post-treatment positions could be assessed) and because the functional and dysfunctional TAI distributions appeared to overlap. In order for the two distributions to overlap, it had to be possible for obtained scores to fall within two standard deviations of either distribution (Jacobson & Truax, 1991). Such overlap readily occurred as the TAI inclusion criterion for the current study, which defined the dysfunctional distribution, was a TAI score above the seventy-fifth percentile, a mere three-quarters of a standard deviation above the normative or functional mean. Using
these criteria, nearly all of the obtained, pre-treatment, dysfunctional TAI scores fell within two standard deviations of the normative, functional distribution.

Using method c, a condition of "meaningful change" was verified in the current study when the post-treatment score was "more likely to be drawn from the functional than from the dysfunctional distribution" (Ankuta & Abeles, 1993, p. 72). Method c established this probability by determining a cutoff value against which post-treatment scores were compared. When an individual’s post-treatment score fell below the cutoff point, that outcome was considered to be closer to the functional population mean rather than the dysfunctional population mean. Hence, that outcome was considered to have satisfied a condition of meaningful change.

The purpose and method for the RC index procedure used in the current study were identical to those described in Jacobson and Truax (1991) and employed in the Ankuta and Abeles (1993) study. The purpose of the RC index was to ensure that relatively small changes that resulted in scores falling closer to the functional rather than the dysfunctional mean, which can happen frequently when distributions overlap, also exceeded differences anticipated on the basis of chance alone (i.e., as a function of regression toward the mean effects). That is, the RC index verified that the small changes that satisfied procedure c were also statistically reliable. In order to accomplish this, the standard error of difference for the pre-treatment assessment scores was calculated across Groups A and B for both males and females for each of the three TAI variables. Each of these computations involved the test-retest coefficient for the TAI scale under consideration. For the current study, the test-retest coefficients between the Group B initial screening and post-wait assessments were used rather than the published test-retest coefficients. Assuming the equivalence of Groups A and B at initial screening, the Group B correlation coefficients represented the best estimates of
the reliability of the TAI-T, TAI-E, and TAI-W with this population (Alessi & Kaye, 1983).

Once the gender-specific, standard errors of difference were calculated, each individual's pre to post change was expressed in standard errors of difference. This result was compared against the critical value of 1.96 (p < .05). When the pre to post raw score difference exceeded 1.96 standard errors of difference, the magnitude of change was judged to be statistically reliable and the measured change was considered valid. In summary, in the current study, an individual's post-treatment score reached clinical significance when both the proximity condition (meaningful change), as described by method c, and the magnitude of change condition (reliable change), as described by the RC index, were satisfied.

For both Groups A and B the clinical significance of the RTT-T was estimated using what Jacobson and Truax (1991) have described as method a (discussed in the Statistical vs. Clinical Significance section of the current text). Method a for calculating clinical significance stipulated the following (Jacobson & Truax, 1991):

The level of functioning subsequent to therapy should fall outside the range of the dysfunctional population, where range is defined as extending to two standard deviations beyond (in the direction of functionality) the mean for that population. (p. 13)

Method a was selected because the RTT does not supply normative mean data. Without normative population data, method a is the only method for calculating clinical significance. In the current study the administration of this procedure was straightforward. The mean raw score and standard deviation for the RTT-T were calculated. Then, individual post-treatment scores were expressed in standard deviations. Those scores which improved more than two standard deviations achieved clinical significance.
Procedures

Setting

Most assessment sessions were conducted in the primary investigator's, university, four-room laboratory. A moderate number of initial screening assessments (perhaps 60 of approximately 200 total) were administered at the university student assistance center. Nearly all of the treatment sessions were conducted in the primary investigator's, university, four-room laboratory. A few treatment sessions (perhaps five or six of approximately 160 total sessions) were conducted in a treatment room in the university clinic. All treatment sessions took place with only the therapist and subject present.

Apparatus

The only items or devices required by the current study were the office furniture and psychometric instruments.

Measures

Independent Variables and Assessment Instruments

There was one experimentally manipulated independent variable in the current study. It was Eye Movement Desensitization and Reprocessing (EMDR). There were two diagnostic and selection independent variables in this experiment. They included the following:

1. Test anxiety was measured by the Test Anxiety Inventory (TAI; Spielberger, 1978).
The 20-item TAI was created by Spielberger (1978) through factor analysis of items mostly from the TAS and features two, 10-item subscales. The two subscales of the TAI include the emotionality scale (TAI-E) and the worry scale (TAI-W). Spielberger (1978) described the two major goals of the TAI as:

1. to construct a brief, objective, self-report scale that was highly correlated with other widely used measures of test anxiety and
2. to employ factor analysis to identify subscales measuring worry and emotionality as major components of test anxiety (p. 2)

In order to mask the intentions of the instrument, that is, to avoid cueing respondents that anxiety is being assessed, questionnaire booklets are labeled Test Anxiety Inventory. However, as Galassi (1985) has remarked, upon reading, the content of the TAI becomes quickly apparent.

Means, standard deviations, and percentile ranks are available for undergraduate college students, first year college students, community college students, and high school students on the TAI-T, TAI-E, and TAI-W. In addition, standardized T scores are available for the TAI-T scale. The norm groups for the TAI were small according to some authors (Knapp, 1985). The norm groups included: 1,449 undergraduates at the University of South Florida, 1,129 incoming freshman at that same institution, 320 students at Hillsborough Community College, and 1,118 high school students in St. Petersburg and Clearwater, Florida. Spielberger (1978) has reported relatively high internal consistency coefficients for all three scales: TAI-T, .92-.96; TAI-E, .85-.91; TAI-W, .83-.91. Spielberger (1978) has reported test-retest reliability coefficients of .80 for college students over a three week time lapse. Others have reported test-retest reliability coefficients for the TAI between .62 and .81 during a span of 2 weeks to 6 months (Schaer & Isom, 1988).

For the current study the inclusion criterion was a TAI total score (TAI-T) above the seventy-fifth percentile. RTT total scores (RTT-T) could not be used for
inclusion purposes because there were no published normative data that specified raw scores and equivalent percentile ranks as there was for the TAI. It was expected that selecting subjects with TAI scores above the seventy-fifth percentile would result in a restricted range of scores among subjects, thus opening up possibilities of regression toward the mean effects on post test scores (cf., Nunnally, 1978, p. 217). It was also expected that the restricted range in scores would result in lower test-retest and inter-test correlations than were the sample unrestricted in range.

2. Study Habits and attitudes were assessed by the college form of the Survey of Study Habits and Attitudes (SSHA; Brown & Holtzman, 1967).

The current, extensively used, version of the SSHA features 100 items that probe students for how frequently they engage in specific study behaviors and how they feel and think regarding professors and the educational process. The SSHA yields one overall measure, the Study Orientation (SSHA-SO) scale. The SO is comprised of two 50-item subscales, the Study Habits (SSHA-SH) scale and the Study Attitudes (SSHA-SA) scale. Each of these in turn is comprised of two 25-item subscales. The SH scale consists of the Delay Avoidance (SSHA-DA) scale and the Work Methods (SSHA-WM) scale. The SSHA-SA scale consists of the Teacher Approval (SSHA-TA) scale and the Educational Acceptance (SSHA-EA) scale. Adequate test-retest reliability (r = .90+) and criterion-related validity have been established for the SSHA (Brown & Holtzman, 1967). Inclusion criterion for the current study was an SSHA-SO, SSHA-SA, or SSHA-SH scale score above the 50th percentile rank.

**Dependent Variables and Assessment Instruments**

There were three primary dependent variables analyzed in the current study. They included the following:
1. Test anxiety as measured by two separate self-report instruments, the TAI (described in the Independent Variables and Assessment Instruments section of the current text) and the Reactions to Tests Total scale (RTT; Sarason, 1984).

Sarason (1984) devised the 40-item RTT empirically through factor analysis on a pool of 91 items, some of which were extracted from the Test Anxiety Scale (TAS; Sarason, 1978). The RTT features four, 10-item subscales that evaluate four discrete aspects or dimensions of TA. However, for the current study only the RTT Total scale (RTT-T) will be considered. Sarason (1984) has reported excellent criterion and construct validity for the RTT. He has reported that the TAS and RTT-T were highly correlated, although he didn’t report actual data.

There were three reasons for using both the TAI and RTT TA assessment instruments in the current study. First, although correlations between most of the available rating scales are high, they generally do not account for the majority of the variance between scales. Consequently, conclusions may be partially determined by the TA instrument being used. Using more than one instrument may enhance the construct validity and treatment utility of the current study’s outcomes. The second reason for using two TA assessment devices was that it allowed cross-validation of the devices. This was considered especially important for the RTT, which appears to be a useful, well-conceived instrument that has received little attention or empirical validation. Finally, using two dependent measures for TA allowed greater comparisons between the current study and other studies. Tryon (1980) has encouraged the adoption of a single, universal TA scale. However, she indicates that many researchers have employed more than one TA assessment methodology.

The TAI was selected for use in this study because it assessed both worry and emotionality, the two components commonly thought to constitute the TA condition. It is also (arguably) the most widely cited and well researched TA device available over
the past 15 years. As mentioned above, the RTT was selected as an overall cross-validation measure. In order to establish the RTT's overall commonality with and validity vis-a-vis the TAI, only the RTT-T scale was evaluated in the current analysis. It should be noted, however, that the RTT's multifactor design can provide a breakdown of the TA condition that is more detailed than the TAI. In addition to the overall scale, the RTT evaluates tension, worry, test-irrelevant thinking, and bodily reactions. This greater refinement and detailing of responses may lead to a more precise and comprehensive understanding of the TA condition, and ultimately to more effective treatment.

2. Study habits and attitudes were assessed by the SSHA (Refer to the Independent Variable/Instruments Section for references and descriptions).

The SSHA was selected as a dependent measure to determine whether reductions in anxiety produced changes in study habits and, or, attitudes. Some have suggested that poor study skills and being poorly prepared for examinations may be a major cause of TA. However, as we have discussed, remediating poor study skills does not alleviate all test-related anxiety. Because the two tests correlated at .00, it might appear that the anxiety a student experiences with taking tests is unrelated to study habits or attitudes. However, it is conceivable that anxiety in response to academic tasks could eventually lead to avoidance of or escape from academic tasks generally. If this was true, then it would follow that reductions in self-reported TA could result in enhanced self-reported study habits and, or, attitudes.

The results of the SSHA were not evaluated within the primary analysis of the current study. Rather, pre-selected comparisons among the SSHA scales were examined within the secondary analysis.

3. A self-reported level of trait anxiety was measured by the Trait form of the State-Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, & Lushene, 1970). This

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
instrument was used to assess for generalization of TA treatment effects to general levels of anxiety (Smith & Nye, 1989).

As with the SSHA, the results of the STAI were not evaluated within the primary analysis of the current study. Rather, pre-selected comparisons of this dependent measure were examined within the secondary analysis.

Procedure

All subjects in Groups A and B completed the following four psychometric assessment instruments at the initial screening and after the termination of treatment: (a) the RTT and TAI to assess test anxiety, (b) the SSHA to assess study habits and attitudes, and (c) the STAI-T to assess trait anxiety. In addition, subjects in Group B completed the four psychometric assessment instruments (described above) at the conclusion of the wait period (post-wait), just prior to the beginning of treatment.

The EMDR protocol administered closely followed that described by Shapiro (1991a, 1991b, 1995) and discussed in the Review of the Current Literature section of the current text (also, see Appendix A). EMDR treatment was administered by the principal investigator who, prior to the current study, had attended both beginning (level 1) and intermediate (level 2) EMDR workshops and treated 14 individuals clinically using EMDR. It may be argued that the principal investigator should not also have been the primary therapist. It was deemed acceptable for these experiments given that the principal investigator was insulated from the subjects during all collections of dependent variable data and had no awareness of ongoing changes in these measures. Initial screening, post-wait, and post-treatment assessment batteries involved paper and pencil questionnaires and no interaction with the primary therapist. Moreover, at least half of all assessment batteries were administered by someone other than the primary investigator. Insulation from outcome data was accomplished by postponing scoring.
and analysis until the conclusion of the study. It was believed that using student therapists trained by the principal investigator or others might have threatened the integrity of the independent variable by using less experienced (hence, less effective) therapists, and by introducing variability across therapists.

As recommended by Shapiro (1991a, 1991b), each EMDR treatment session was approximately 90 minutes in length. Across Groups A and B, subjects received an average of approximately 4.4, 90-minute EMDR sessions. At the outset of the study it was estimated that between one and four sessions would be sufficient to attain maximally beneficial results. As has been suggested by Shapiro (1991a, 1991b), EMDR treatment was highly individualized among all subjects. Based on the therapist’s perceptions of the subjects’ differing preferences and needs, the number of sessions, the number of eye movement sets within each session, the speed, magnitude, and number of saccads, the degree of “cognitive interweaving” (Shapiro, 1991b), and the extent of therapist encouragement and, or, involvement with the evolution of treatment were all varied substantially among subjects.

Treatment was terminated after a maximum of 10 sessions or when the subject indicated a SUD of 0 (or as near 0 as the subject believed he or she could attain) on all pertinent target images. Post-treatment assessments were conducted between one and two weeks after the final treatment sessions.
CHAPTER III

RESULTS

Preliminary Analysis

The results of the current study are presented in the order of the three experimental questions (refer to page 107 of the current text). Summary data for these analyses are presented in 37 tables and 16 figures that accompany and support the text. For clarification of the acronyms used to signify the dependent variables discussed in the current text, refer to Table 2.

Reliability

The test-retest component required by the current experimental design for Group B (the waiting control condition) provided an estimate of the reliability coefficient for each of the dependent variables, with this population. That is, the estimated reliability coefficients are equivalent to the correlations between the initial Group B and the post-wait Group B values for each overall or total dependent variable and their subscales. These correlation coefficients represented the best estimates of the reliability of these instruments with this population (Alessi & Kaye, 1983). These correlation coefficients, which were calculated using raw scores, were as follows:

- initial--post-wait TAI Total scale (TAI-T), $r = .78$
- initial--post-wait TAI Emotion scale (TAI-E), $r = .70$
- initial--post-wait TAI Worry scale (TAI-W), $r = .85$
- initial--post-wait RTT Total scale (RTT-T), $r = .84$
- initial--post-wait RTT Tension scale (RTT-TE), $r = .80$
- initial--post-wait RTT Worry scale (RTT-W), $r = .57$
- initial--post-wait RTT Test-irrelevant Thinking scale (RTT-I), $r = .85$

124
### Table 2
Dependent Variables and Their Acronyms

<table>
<thead>
<tr>
<th>Acronyms</th>
<th>Full Names of Dependent Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAI-T</td>
<td>Test Anxiety Inventory—Total Score</td>
</tr>
<tr>
<td>TAI-E</td>
<td>Test Anxiety Inventory—Emotion Scale</td>
</tr>
<tr>
<td>TAI-W</td>
<td>Test Anxiety Inventory—Worry Scale</td>
</tr>
<tr>
<td>RTT-T</td>
<td>Reactions to Tests—Total Score</td>
</tr>
<tr>
<td>RTT-TE</td>
<td>Reactions to Tests—Tension Scale</td>
</tr>
<tr>
<td>RTT-W</td>
<td>Reactions to Tests—Worry Scale</td>
</tr>
<tr>
<td>RTT-I</td>
<td>Reactions to Tests—Test Irrelevant Thinking Scale</td>
</tr>
<tr>
<td>RTT-B</td>
<td>Reactions to Tests—Bodily Reactions Scale</td>
</tr>
<tr>
<td>SSHA-DA</td>
<td>Survey of Study Habits and Attitudes—Delay Avoidance Scale</td>
</tr>
<tr>
<td>SSHA-WM</td>
<td>Survey of Study Habits and Attitudes—Work Methods Scale</td>
</tr>
<tr>
<td>SSHA-SH</td>
<td>Survey of Study Habits and Attitudes—Study Habits Scale</td>
</tr>
<tr>
<td>SSHA-TA</td>
<td>Survey of Study Habits and Attitudes—Teacher Approval Scale</td>
</tr>
<tr>
<td>SSHA-EA</td>
<td>Survey of Study Habits and Attitudes—Educational Acceptance Scale</td>
</tr>
<tr>
<td>SSHA-SA</td>
<td>Survey of Study Habits and Attitudes—Study Attitudes Scale</td>
</tr>
<tr>
<td>SSHA-SO</td>
<td>Survey of Study Habits and Attitudes—Total Study Orientation Scale</td>
</tr>
<tr>
<td>STAI-T</td>
<td>Strait-Trait Anxiety Inventory—Trait Scale</td>
</tr>
</tbody>
</table>

scale (RTT-B), $r = .83$; initial--post-wait SSHA Delay Avoidance scale (SSHA-DA), $r = .86$; initial--post-wait SSHA Work Methods scale (SSHA-WM), $r = .82$; initial--post-wait SSHA Study Habits scale (SSHA-SH), $r = .83$; initial--post-wait SSHA Teacher Acceptance scale (SSHA-TA), $r = .74$; initial--post-wait SSHA Educational

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Acceptance scale (SSHA-EA), $r = .55$; initial--post-wait SSHA Study Attitudes scale (SSHA-SA), $r = .66$; initial--post-wait SSHA Study Orientation scale (SSHA-SO), $r = .65$; initial--post-wait STAI Trait Anxiety scale (STAI-T), $r = .81$. These correlation coefficients between Group B initial and Group B post-wait dependent variable data are presented also in Table 3.

Validity

The best estimates of the assessment instruments' validity, as they were used in the current study, were the correlations among all of the primary dependent variables and their subscales as measured at initial screening for Groups A and B. The construct validity of each assessment instrument would appear to be enhanced if certain correlations among the instruments were strong and others weak. That is, if the TAI and RTT instruments are both measuring the TA condition (or construct), then the TAI-T and RTT-T scales should be strongly correlated with each other. Moreover, if the TAI and RTT are not measuring the construct of study habits and attitudes or trait anxiety, then the TAI and RTT should correlate weakly with the SSHA and STAI. Finally, if the SSHA and STAI are measuring the constructs of study habits/attitudes and trait anxiety respectively, then the SSHA-SO and STAI-T should be weakly correlated with one another. The correlation coefficients among all dependent instruments used in the current study are presented in Table 4. The correlation coefficients among the primary assessment devices, presented in rank order from strongest to weakest degree of correlation (ignoring the directionality or sign of the correlation), were as follows: TAI-T--RTT-T, $r = .80$; RTT-T--STAI-T, $r = .39$; SSHA-SO--STAI-T, $r = -.34$; RTT-T--SSHA-SO, $r = -.17$; TAI-T--STAI-T, $r = .14$; TAI-T--SSHA-SO, $r = .00$.
Table 3
Correlation Analyses: A Comparison of Initial Screening and Post-Wait Correlation Coefficients in Group B With Published Reliability Coefficients

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Obtained Correlation Coefficients</th>
<th>Published Reliability Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAI-T</td>
<td>.78</td>
<td>.80</td>
</tr>
<tr>
<td>TAI-E</td>
<td>.70</td>
<td>*</td>
</tr>
<tr>
<td>TAI-W</td>
<td>.85</td>
<td>*</td>
</tr>
<tr>
<td>RTT-T</td>
<td>.84</td>
<td>*</td>
</tr>
<tr>
<td>RTT-TE</td>
<td>.80</td>
<td>*</td>
</tr>
<tr>
<td>RTT-W</td>
<td>.57</td>
<td>*</td>
</tr>
<tr>
<td>RTT-I</td>
<td>.85</td>
<td>*</td>
</tr>
<tr>
<td>RTT-B</td>
<td>.83</td>
<td>*</td>
</tr>
<tr>
<td>SSHA-DA</td>
<td>.87</td>
<td>*</td>
</tr>
<tr>
<td>SSHA-WM</td>
<td>.82</td>
<td>*</td>
</tr>
<tr>
<td>SSHA-SH</td>
<td>.83</td>
<td>*</td>
</tr>
<tr>
<td>SSHA-TA</td>
<td>.74</td>
<td>*</td>
</tr>
<tr>
<td>SSHA-EA</td>
<td>.55</td>
<td>*</td>
</tr>
<tr>
<td>SSHA-SA</td>
<td>.66</td>
<td>*</td>
</tr>
<tr>
<td>SSHA-SO</td>
<td>.65</td>
<td>.90</td>
</tr>
<tr>
<td>STAI-T</td>
<td>.81</td>
<td>.84</td>
</tr>
</tbody>
</table>

* Reliability Coefficient Unavailable

These results support the predictions that the TAI and RTT are measuring similar variables or constructs, and, by contrast, the SSHA and STAI are each
Table 4
Correlation Analyses Among all Dependent Measures Pooled
Across Groups A & B at Initial Screening

<table>
<thead>
<tr>
<th>Dependent Variables</th>
<th>TAI-T</th>
<th>TAI-E</th>
<th>TAI-W</th>
<th>RTT-T</th>
<th>RTT-TE</th>
<th>RTT-W</th>
<th>RTT-I</th>
<th>RTT-B</th>
<th>SSHA-DA</th>
<th>SSHA-WM</th>
<th>SSHA-SH</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAI-T</td>
<td>1.0000</td>
<td>.7071</td>
<td>.7326</td>
<td>.7963</td>
<td>.5492</td>
<td>.2766</td>
<td>.6713</td>
<td>-.1228</td>
<td>.0401</td>
<td>-.087</td>
<td></td>
</tr>
<tr>
<td>TAI-E</td>
<td>.7071</td>
<td>1.0000</td>
<td>.1477</td>
<td>.4223</td>
<td>.7575</td>
<td>.0787</td>
<td>-.1266</td>
<td>.6276</td>
<td>.1251</td>
<td>.3250</td>
<td>.197</td>
</tr>
<tr>
<td>TAI-W</td>
<td>.7326</td>
<td>.1477</td>
<td>1.0000</td>
<td>.7693</td>
<td>.4384</td>
<td>.7109</td>
<td>.5752</td>
<td>.3330</td>
<td>-.2143</td>
<td>-.1720</td>
<td>-.228</td>
</tr>
<tr>
<td>RTT-T</td>
<td>.7963</td>
<td>.4223</td>
<td>.7693</td>
<td>1.0000</td>
<td>.7168</td>
<td>.7152</td>
<td>.6820</td>
<td>.5957</td>
<td>-.3085</td>
<td>-.0971</td>
<td>-.261</td>
</tr>
<tr>
<td>RTT-TE</td>
<td>.7676</td>
<td>.7575</td>
<td>.4384</td>
<td>.7168</td>
<td>1.0000</td>
<td>.3050</td>
<td>.1686</td>
<td>.6608</td>
<td>-.0210</td>
<td>.0876</td>
<td>.001</td>
</tr>
<tr>
<td>RTT-W</td>
<td>.5492</td>
<td>.0787</td>
<td>.7109</td>
<td>.7152</td>
<td>.3050</td>
<td>1.0000</td>
<td>.5004</td>
<td>.0102</td>
<td>-.4078</td>
<td>-.1710</td>
<td>-.356</td>
</tr>
<tr>
<td>RTT-I</td>
<td>.2766</td>
<td>-.1266</td>
<td>.5752</td>
<td>.6820</td>
<td>.1686</td>
<td>.5004</td>
<td>1.0000</td>
<td>-.0176</td>
<td>-.4292</td>
<td>-.3068</td>
<td>-.421</td>
</tr>
<tr>
<td>RTT-B</td>
<td>.6713</td>
<td>.6276</td>
<td>.3330</td>
<td>.5957</td>
<td>.6608</td>
<td>.1020</td>
<td>.0176</td>
<td>1.0000</td>
<td>.1039</td>
<td>.1970</td>
<td>.137</td>
</tr>
<tr>
<td>SSHA-DA</td>
<td>-.1228</td>
<td>.1251</td>
<td>-.2143</td>
<td>-.3085</td>
<td>-.0210</td>
<td>-.4078</td>
<td>-.4292</td>
<td>1.0000</td>
<td>.5684</td>
<td>.921</td>
<td></td>
</tr>
<tr>
<td>SSHA-WM</td>
<td>.0401</td>
<td>.3250</td>
<td>-.1720</td>
<td>-.0971</td>
<td>.0876</td>
<td>-.1710</td>
<td>-.3068</td>
<td>.1970</td>
<td>.5684</td>
<td>1.0000</td>
<td>.831</td>
</tr>
<tr>
<td>SSHA-SH</td>
<td>-.0873</td>
<td>.1972</td>
<td>-.2284</td>
<td>-.2649</td>
<td>.0015</td>
<td>-.3507</td>
<td>-.4260</td>
<td>.1379</td>
<td>.9269</td>
<td>.8306</td>
<td>.1000</td>
</tr>
<tr>
<td>SSHA-TA</td>
<td>.2121</td>
<td>.2818</td>
<td>.1104</td>
<td>.1915</td>
<td>.2452</td>
<td>.0507</td>
<td>.0488</td>
<td>.2284</td>
<td>-.1978</td>
<td>.1275</td>
<td>.081</td>
</tr>
<tr>
<td>SSHA-EA</td>
<td>.0346</td>
<td>.1653</td>
<td>-.0162</td>
<td>-.0528</td>
<td>.0051</td>
<td>-.1484</td>
<td>-.2390</td>
<td>.2746</td>
<td>.3661</td>
<td>.4861</td>
<td>.46</td>
</tr>
<tr>
<td>SSHA-SA</td>
<td>.0995</td>
<td>.2445</td>
<td>.0027</td>
<td>.0418</td>
<td>.1207</td>
<td>-.0927</td>
<td>-.1188</td>
<td>.2528</td>
<td>.0593</td>
<td>.3479</td>
<td>.19</td>
</tr>
<tr>
<td>SSHA-SO</td>
<td>.0036</td>
<td>.2850</td>
<td>-.1584</td>
<td>-.1719</td>
<td>.0614</td>
<td>-.2926</td>
<td>-.3829</td>
<td>.2344</td>
<td>.7193</td>
<td>.8038</td>
<td>.84</td>
</tr>
<tr>
<td>STAI-T</td>
<td>.1364</td>
<td>-.0610</td>
<td>.1902</td>
<td>.3862</td>
<td>.1321</td>
<td>.3829</td>
<td>.4441</td>
<td>.0176</td>
<td>-.2477</td>
<td>-.3893</td>
<td>-.32</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th></th>
<th>RTT-I</th>
<th>RTT-B</th>
<th>SSHA-DA</th>
<th>SSHA-WM</th>
<th>SSHA-SH</th>
<th>SSHA-TA</th>
<th>SSHA-EA</th>
<th>SSHA-SA</th>
<th>SSHA-SO</th>
<th>STAI-T</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>.2766</td>
<td></td>
<td>.6713</td>
<td>-.1228</td>
<td>.0401</td>
<td>.2121</td>
<td>.0346</td>
<td>.0995</td>
<td>.0036</td>
</tr>
<tr>
<td></td>
<td>-.1266</td>
<td>.6276</td>
<td></td>
<td>.1251</td>
<td>.3250</td>
<td>.1972</td>
<td>.2818</td>
<td>.1653</td>
<td>.2445</td>
<td>.2850</td>
</tr>
<tr>
<td></td>
<td>.5752</td>
<td>.3330</td>
<td></td>
<td>-.2143</td>
<td>-.1720</td>
<td>-.2284</td>
<td>.1104</td>
<td>-.0162</td>
<td>.0027</td>
<td>-.1584</td>
</tr>
<tr>
<td></td>
<td>.6820</td>
<td>.5957</td>
<td></td>
<td>-.3085</td>
<td>-.0971</td>
<td>-.2649</td>
<td>.1915</td>
<td>-.0528</td>
<td>.0418</td>
<td>-.1719</td>
</tr>
<tr>
<td></td>
<td>.1686</td>
<td>.6608</td>
<td></td>
<td>-.0210</td>
<td>.0876</td>
<td>.0015</td>
<td>.2452</td>
<td>.0051</td>
<td>.1207</td>
<td>.0614</td>
</tr>
<tr>
<td></td>
<td>.5004</td>
<td>.1020</td>
<td></td>
<td>-.4078</td>
<td>-.1710</td>
<td>-.3501</td>
<td>.0507</td>
<td>-.1484</td>
<td>-.0927</td>
<td>-.2926</td>
</tr>
<tr>
<td></td>
<td>1.0000</td>
<td>-.0176</td>
<td></td>
<td>.4292</td>
<td>-.3068</td>
<td>-.4260</td>
<td>.0489</td>
<td>-.2390</td>
<td>-.1188</td>
<td>-.3829</td>
</tr>
<tr>
<td></td>
<td>.0176</td>
<td>1.0000</td>
<td></td>
<td>.1039</td>
<td>.1970</td>
<td>.1379</td>
<td>.2284</td>
<td>.2746</td>
<td>.2528</td>
<td>.2344</td>
</tr>
<tr>
<td></td>
<td>-.4292</td>
<td>.1039</td>
<td></td>
<td>1.0000</td>
<td>.5684</td>
<td>.9269</td>
<td>-.1978</td>
<td>.3661</td>
<td>.0593</td>
<td>.7193</td>
</tr>
<tr>
<td></td>
<td>-.3068</td>
<td>.1970</td>
<td></td>
<td>.5684</td>
<td>1.0000</td>
<td>.8306</td>
<td>-.1275</td>
<td>.4861</td>
<td>.3479</td>
<td>.8038</td>
</tr>
<tr>
<td></td>
<td>-.4260</td>
<td>.1379</td>
<td></td>
<td>.9269</td>
<td>.8306</td>
<td>1.0000</td>
<td>-.0802</td>
<td>.4641</td>
<td>.1940</td>
<td>.8455</td>
</tr>
<tr>
<td></td>
<td>.0488</td>
<td>.2284</td>
<td></td>
<td>-.1978</td>
<td>.1275</td>
<td>-.0802</td>
<td>1.0000</td>
<td>.4516</td>
<td>.8763</td>
<td>.4014</td>
</tr>
<tr>
<td></td>
<td>-.2390</td>
<td>.2746</td>
<td></td>
<td>.3661</td>
<td>.4861</td>
<td>.4641</td>
<td>.4516</td>
<td>1.0000</td>
<td>.8142</td>
<td>.7992</td>
</tr>
<tr>
<td></td>
<td>-.1188</td>
<td>.2528</td>
<td></td>
<td>.0593</td>
<td>.3479</td>
<td>.1940</td>
<td>.8763</td>
<td>.8142</td>
<td>1.0000</td>
<td>.6795</td>
</tr>
<tr>
<td></td>
<td>-.3829</td>
<td>.2344</td>
<td></td>
<td>.7193</td>
<td>.8038</td>
<td>.8455</td>
<td>.4014</td>
<td>.7992</td>
<td>.6795</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>.4441</td>
<td>.0176</td>
<td></td>
<td>-.2477</td>
<td>-.3893</td>
<td>-.3238</td>
<td>-.0980</td>
<td>-.2451</td>
<td>-.1993</td>
<td>-.3407</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
measuring different variables or constructs. These results therefore enhance the construct validity of each assessment device used in the current study.

**Design Integrity: Group Equivalence**

The 32 subjects were assigned randomly to Group A or B. However, random assignment cannot guarantee that the groups were functionally equivalent, or eliminate the possibility that some initial, pre-existing differences between the groups influenced substantially the outcome of treatment. To establish that any initial between-group differences were minimal, and thereby enhance the confidence in and validity of the current study's overall treatment results, t-tests were used to evaluate the groups' initial mean differences across all 16 assessment instrument scales and subscales. The results of these t-tests, which are displayed in Table 5, found no significant differences between Groups A and B among any of the dependent measures. These results indicate that Groups A and B were functionally equivalent relative to the assessment devices used in the current study.

**Delay to Treatment and Treatment Length**

Once they met the criteria for the study and agreed to participate, subjects in Group A commenced treatment as expeditiously as possible and subjects in Group B were assigned to the no-treatment wait condition. However, due to differing work and class schedules, treatment sessions could not be scheduled in an equally expeditious or regular manner for all subjects. As can be seen in Figures 1 through 4 (frequency distributions), this resulted in substantial variance in both the delay to treatment and the overall length of treatment for both Groups A and B (Table 6).

For Group A subjects the mean number of days between the initial screening assessment and the first treatment session was 19.00 (s = 25.39) (Table 7). The mean
Table 5
Mean Difference Between Initial Screening Assessments for Groups A and B: Descriptive and Inferential Statistics

16 Simultaneous t-tests; 30 degrees of freedom

<table>
<thead>
<tr>
<th>Dependent Variables</th>
<th>Group A Initial Mean</th>
<th>Group B Initial Mean</th>
<th>Obtained t statistic</th>
<th>Obtained p-value</th>
<th>Critical t statistic</th>
<th>Critical p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAI-T</td>
<td>86.81</td>
<td>93.50</td>
<td>1.108</td>
<td>.2768</td>
<td>3.214</td>
<td>.0031</td>
</tr>
<tr>
<td>TAI-E</td>
<td>91.69</td>
<td>89.19</td>
<td>.5834</td>
<td>.5650</td>
<td>3.214</td>
<td>.0031</td>
</tr>
<tr>
<td>TAI-W</td>
<td>85.94</td>
<td>91.91</td>
<td>1.312</td>
<td>.1996</td>
<td>3.214</td>
<td>.0031</td>
</tr>
<tr>
<td>RTT-T</td>
<td>* 101.38</td>
<td>* 108.06</td>
<td>1.152</td>
<td>.2596</td>
<td>3.214</td>
<td>.0031</td>
</tr>
<tr>
<td>RTT-TE</td>
<td>* 34.19</td>
<td>* 35.19</td>
<td>0.613</td>
<td>.5446</td>
<td>3.214</td>
<td>.0031</td>
</tr>
<tr>
<td>RTT-W</td>
<td>* 26.81</td>
<td>* 30.50</td>
<td>1.828</td>
<td>.0792</td>
<td>3.214</td>
<td>.0031</td>
</tr>
<tr>
<td>RTT-I</td>
<td>* 18.31</td>
<td>* 20.69</td>
<td>0.868</td>
<td>.3924</td>
<td>3.214</td>
<td>.0031</td>
</tr>
<tr>
<td>RTT-B</td>
<td>* 22.06</td>
<td>* 21.56</td>
<td>0.217</td>
<td>.8296</td>
<td>3.214</td>
<td>.0031</td>
</tr>
<tr>
<td>SSHA-DA</td>
<td>57.81</td>
<td>62.00</td>
<td>0.429</td>
<td>.6707</td>
<td>3.214</td>
<td>.0031</td>
</tr>
<tr>
<td>SSHA-WM</td>
<td>36.56</td>
<td>50.31</td>
<td>1.655</td>
<td>.1083</td>
<td>3.214</td>
<td>.0031</td>
</tr>
<tr>
<td>SSHA-SH</td>
<td>47.81</td>
<td>56.25</td>
<td>0.930</td>
<td>.3600</td>
<td>3.214</td>
<td>.0031</td>
</tr>
<tr>
<td>SSHA-TA</td>
<td>67.81</td>
<td>61.88</td>
<td>0.660</td>
<td>.5161</td>
<td>3.214</td>
<td>.0031</td>
</tr>
<tr>
<td>SSHA-EA</td>
<td>67.31</td>
<td>68.69</td>
<td>0.171</td>
<td>.8651</td>
<td>3.214</td>
<td>.0031</td>
</tr>
<tr>
<td>SSHA-SA</td>
<td>68.44</td>
<td>64.81</td>
<td>0.458</td>
<td>.6515</td>
<td>3.214</td>
<td>.0031</td>
</tr>
<tr>
<td>SSHA-SO</td>
<td>57.50</td>
<td>61.50</td>
<td>0.528</td>
<td>.6013</td>
<td>3.214</td>
<td>.0031</td>
</tr>
<tr>
<td>STAI-T</td>
<td>73.20</td>
<td>75.38</td>
<td>0.266</td>
<td>.7921</td>
<td>3.214</td>
<td>.0031</td>
</tr>
</tbody>
</table>

* Raw Scores--all other means are percentile scores.
Figure 1. Frequency Distribution of Group A Subjects on Days Following Initial Screening Assessment Prior to First Treatment Session.

Legend: ■ Group A Subjects
Figure 2. Frequency Distribution of Group A Subjects on Total Days From First to Last Treatment Session.

Legend: ■ Group A Subjects
Figure 3. Frequency Distribution of Group B Subjects on Days Following Initial Screening Assessment Prior to First Treatment Session.
Figure 4. Frequency Distribution of Group B Subjects on Total Days From First to Last Treatment Session.
Table 6

Delay to Treatment and Length of Treatment for Subjects in Groups A and B

<table>
<thead>
<tr>
<th>Group A Subjects</th>
<th>Group B Subjects</th>
<th>Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Delay to Treatment</td>
<td>Length of Treatment</td>
</tr>
<tr>
<td>--------</td>
<td>------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Female</td>
<td>6</td>
<td>30</td>
</tr>
<tr>
<td>Female</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>Male</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>Female</td>
<td>30</td>
<td>33</td>
</tr>
<tr>
<td>Male</td>
<td>4</td>
<td>114</td>
</tr>
<tr>
<td>Female</td>
<td>92</td>
<td>47</td>
</tr>
<tr>
<td>Female</td>
<td>5</td>
<td>122</td>
</tr>
<tr>
<td>Male</td>
<td>17</td>
<td>102</td>
</tr>
<tr>
<td>Female</td>
<td>66</td>
<td>44</td>
</tr>
<tr>
<td>Male</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>Female</td>
<td>6</td>
<td>21</td>
</tr>
<tr>
<td>Female</td>
<td>10</td>
<td>91</td>
</tr>
<tr>
<td>Female</td>
<td>6</td>
<td>63</td>
</tr>
<tr>
<td>Male</td>
<td>31</td>
<td>28</td>
</tr>
<tr>
<td>Female</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>Male</td>
<td>5</td>
<td>22</td>
</tr>
</tbody>
</table>

--- Did not Complete Tx

--- Did not Complete Tx

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 7

<table>
<thead>
<tr>
<th></th>
<th>Group A Delay to Treatment</th>
<th>Group A Length of Treatment</th>
<th>Group B Delay to Treatment</th>
<th>Group B Length of Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>2</td>
<td>8</td>
<td>20</td>
<td>13</td>
</tr>
<tr>
<td>Maximum</td>
<td>92</td>
<td>122</td>
<td>123</td>
<td>79</td>
</tr>
<tr>
<td>Range</td>
<td>90</td>
<td>114</td>
<td>103</td>
<td>66</td>
</tr>
<tr>
<td>Mean</td>
<td>19.00</td>
<td>47.31</td>
<td>76.75</td>
<td>45.42</td>
</tr>
<tr>
<td>Median</td>
<td>7.00</td>
<td>30.50</td>
<td>90.00</td>
<td>32.50</td>
</tr>
<tr>
<td>Mode</td>
<td>6</td>
<td>10</td>
<td>110</td>
<td>None</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>25.39</td>
<td>39.13</td>
<td>38.77</td>
<td>25.01</td>
</tr>
</tbody>
</table>

For Group B subjects the mean number of days between the initial screening assessment and the first treatment session (the no-treatment waiting control period) was 76.75 ($s = 38.77$) (Table 7). The mean number of days that Group B subjects were in treatment was 45.42 ($s = 25.01$). The mean number of EMDR sessions administered to Group B during this period was 3.58 ($s = 1.00$) (Table 8), the median number of
EMDR sessions was 3, and the mode number of sessions was also 3. Across Groups A and B, the mean number of EMDR sessions administered per subject was 4.36 ($s = 2.61$), the median number of sessions was 3, and the mode was 3.

Whereas the median and mode number of EMDR sessions administered to Groups A and B were identical (3), the mean number of sessions was somewhat different (Group A = 4.94; Group B = 3.58). This difference is largely due to the four Group A members who received the maximum of 10 sessions. By contrast, the highest number of sessions received by a Group B member was 5 (Table 9).
<table>
<thead>
<tr>
<th>Group A Subjects Gender</th>
<th>EMDR Sessions</th>
<th>Group B Subjects Gender</th>
<th>EMDR Sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>3</td>
<td>Female</td>
<td>---</td>
</tr>
<tr>
<td>Female</td>
<td>2</td>
<td>Male</td>
<td>3</td>
</tr>
<tr>
<td>Male</td>
<td>3</td>
<td>Male</td>
<td>3</td>
</tr>
<tr>
<td>Female</td>
<td>10</td>
<td>Female</td>
<td>3</td>
</tr>
<tr>
<td>Male</td>
<td>10</td>
<td>Female</td>
<td>5</td>
</tr>
<tr>
<td>Female</td>
<td>2</td>
<td>Female</td>
<td>5</td>
</tr>
<tr>
<td>Female</td>
<td>10</td>
<td>Female</td>
<td>3</td>
</tr>
<tr>
<td>Male</td>
<td>7</td>
<td>Male</td>
<td>2</td>
</tr>
<tr>
<td>Female</td>
<td>3</td>
<td>Female</td>
<td>---</td>
</tr>
<tr>
<td>Male</td>
<td>2</td>
<td>Male</td>
<td>3</td>
</tr>
<tr>
<td>Female</td>
<td>3</td>
<td>Male</td>
<td>5</td>
</tr>
<tr>
<td>Female</td>
<td>10</td>
<td>Female</td>
<td>4</td>
</tr>
<tr>
<td>Female</td>
<td>5</td>
<td>Female</td>
<td>---</td>
</tr>
<tr>
<td>Male</td>
<td>3</td>
<td>Female</td>
<td>---</td>
</tr>
<tr>
<td>Female</td>
<td>2</td>
<td>Female</td>
<td>4</td>
</tr>
<tr>
<td>Male</td>
<td>4</td>
<td>Female</td>
<td>3</td>
</tr>
</tbody>
</table>

--- Did not Complete Tx
Primary Analysis

Dependent data will be presented in order of the experimental questions. For each experimental question and accompanying data set, the current text will proceed in the following order: First, descriptive data will be described and presented graphically in figures for visual inspection. Next, statistical analyses will be discussed and presented in table form. Finally, the clinical significance of individual outcomes will be described and presented in table form.

The analysis of Question 1 involved four, two factor repeated measures ANOVAs, one ANOVA for each of the four dependent measures of TA (TAI-T, TAI-E, TAI-W, RTT-T). These four analyses contrasted the initial screening and post-treatment scores in Group A with the initial screening and post-wait scores in Group B, which represented the effects of EMDR treatment controlled for effects unrelated to the treatment protocol (e.g., regression toward the mean effects). Four, two-sample, simultaneous t-tests functioned as post-hoc, pair-wise analyses to verify significant differences between Group A’s post-treatment means and Group B’s post-wait means. One t-test was run for each of the four dependent measures of TA. Finally, for each subject in Group A, an estimate of clinical significance was calculated for the initial screening to post-treatment differences on all TA measures.

The analyses of experimental Questions 2 and 3 involved a set of four, one factor repeated measures ANOVAs, one ANOVA for each of the four dependent measures of TA (TAI-T, TAI-E, TAI-W, RTT-T). These four analyses contrasted the initial screening and post-wait scores as well as the post-wait and post-treatment scores in Group B which represented the effects of the no-treatment waiting control condition and EMDR treatment respectively. Paired, simultaneous t-tests functioned as post-hoc, pair-wise analyses to verify significant contrasts. For experimental Question 2, four
paired t-tests were conducted to compare the mean differences between Group B's post-wait and post-treatment scores. One paired t-test was conducted for each dependent measure of TA. In addition, in order to evaluate the relative magnitude of treatment effects between the groups, four, two-sample t-tests comparing Group A's and Group B's post-treatment means were conducted. One, two-sample t-test was conducted for each of the four dependent measures of TA. For experimental Question 3, four paired t-tests were conducted to compare the mean differences between Group B's initial screening and post-wait scores. One paired t-test was conducted for each TA measure. Finally, for Questions 2 and 3, for each subject in Group B, estimates of clinical significance were calculated. Question 2 evaluated the clinical significance of all individual changes in TA from post-wait to post-treatment assessments. Question 3 evaluated the clinical significance of all individual changes in TA from initial screening to post-wait assessments.

**Question 1**

Would EMDR be more effective for treating test anxiety than spontaneous recovery that occurs during a waiting control condition as measured by the TAI-T, TAI-E, TAI-W, or RTT-T in a population of college students whose TA is not the result of obvious academic factors? Referring to Figure 5, visual inspection suggests that the mean reduction of 54.00 (s = 27.98) percentile points in Group A's TAI-T score from initial screening to post-treatment (over the course of EMDR treatment) was substantially larger than the mean reduction of 6.00 (s = 9.72) percentile points in Group B's TAI-T score from initial screening to post-wait (over the course of waiting for and receiving no treatment). A two factor ANOVA on TAI-T scores (2 groups X 2 assessment occasions) showed a significant main effect for the groups (p < .0001), a
Figure 5. A Comparison of the Effects of EMDR on the TAI-T in Group A With the Effects of Waiting on the TAI-T in Group B With the Effects of EMDR on the TAI-T in Group B.
significant main effect for the repeated assessment ($p < .0001$), and a significant treatment by assessment interaction ($p < .0001$) (Table 10).

Table 10
Analysis of Variance Summary Table for Two Factor Repeated Measures ANOVA on TAI-T

<table>
<thead>
<tr>
<th>Source</th>
<th>dF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>1</td>
<td>10557.56</td>
<td>10557.56</td>
<td>52.06</td>
<td>.0001</td>
</tr>
<tr>
<td>Error (Treatment)</td>
<td>30</td>
<td>6083.87</td>
<td>202.80</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Assessment</td>
<td>1</td>
<td>14,400.00</td>
<td>14,400.00</td>
<td>65.64</td>
<td>.0001</td>
</tr>
<tr>
<td>Treatment X Assessment Interaction</td>
<td>1</td>
<td>9216.00</td>
<td>9216.00</td>
<td>42.01</td>
<td>.0001</td>
</tr>
<tr>
<td>Error (Treatment X Assessment Interaction)</td>
<td>30</td>
<td>6581.00</td>
<td>219.37</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

These results indicated that, on the TAI-T, the main effect of the treatment factor did not consistently reach significance across the main effect of the assessment factor. Rather, the significance of the treatment factor was contingent on the level of the assessment factor. In consideration of the visual and descriptive analyses, the results of the ANOVA suggested that Group A's post-treatment mean TAI-T score was significantly lower than Group B's post-wait mean TAI-T score. A two-sample t-test comparing the Group A (EMDR) TAI-T mean value at post-treatment (37.81) with the
Group B (no-treatment control) TAI-T mean value at post-wait (87.50) verified that this intergroup mean difference reached significance (p < .0001) (Table 11).

The clinical significance (CS) of TAI-T reductions was analyzed for each member of Group A via the two-stage process suggested by Jacobson and Truax (1991). First, method c was used to determine the gender-specific cutoff points below which post-treatment scores were more proximal to the functional rather than the dysfunctional mean. Second, RC indices were calculated to ensure that the magnitudes of the pre to post changes were statistically reliable. Table 12 lists the data used in determining the RC indices for Group A on the TAI-T. Table 13 displays the gender-specific cutoff points, RC indices, and the results of the two-stage process for determining clinical significance.

The results of the method c analyses indicated that the post-treatment TAI-T scores for all 16 Group A subjects fell below the gender-specific cutoff points. Hence, at post-treatment, all 16 subject's TAI-T scores were more proximal to the functional rather than the dysfunctional population means. However, it should be noted that for three of these subjects the TAI-T score fell below the gender-specific cutoff point at the initial screening assessment.

The results of the RC analyses found that for one female subject the RC index failed to exceed the 1.96 critical value. The magnitude of her change did not reach statistical significance, thus did not satisfy the reliable change requirement for clinical significance. Overall, TAI-T outcomes for 15 of the 16 Group A subjects attained clinical significance.

Referring to Figure 6, visual inspection suggests that the mean reduction of 55.56 (s = 25.87) percentile points in Group A's TAI-E score from initial screening to post-treatment was substantially larger than the mean reduction of 5.19 (s = 14.97) percentile points in Group B's TAI-E score from initial screening to post-wait. A two
Table 11

Descriptive Data and Two-Sample t-tests on TAI-T, TAI-E, TAI-W, and RTT-T:
Mean Differences Between Group A Post-Treatment and Group B Post-Wait

4 simultaneous t-tests; 15 degrees of freedom

<table>
<thead>
<tr>
<th>Measure</th>
<th>Group A Post-Treatment Mean</th>
<th>Group A Standard Deviation</th>
<th>Group B Post-Wait Mean</th>
<th>Group B Standard Deviation</th>
<th>Obtained p-value</th>
<th>Critical p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAI-T</td>
<td>37.81</td>
<td>25.56</td>
<td>87.50</td>
<td>11.10</td>
<td>.0001</td>
<td>.0125</td>
</tr>
<tr>
<td>TAI-E</td>
<td>36.13</td>
<td>24.55</td>
<td>84.00</td>
<td>17.95</td>
<td>.0001</td>
<td>.0125</td>
</tr>
<tr>
<td>TAI-W</td>
<td>44.88</td>
<td>23.23</td>
<td>83.94</td>
<td>19.26</td>
<td>.0001</td>
<td>.0125</td>
</tr>
<tr>
<td>RTT-T</td>
<td>* 64.94</td>
<td>12.81</td>
<td>* 102.81</td>
<td>12.84</td>
<td>.0001</td>
<td>.0125</td>
</tr>
</tbody>
</table>

* Raw scores—all other mean differences are percentiles

factor ANOVA on TAI-E scores (2 groups X 2 assessment occasions) showed a significant main effect for the groups (p < .0001), a significant main effect for the repeated assessment (p < .0001), and a significant treatment by assessment interaction (p < .0001) (Table 14).

These results indicated that, on the TAI-E, the main effect of the treatment factor did not consistently reach significance across the main effect of the assessment factor. Rather, the significance of the treatment factor was contingent on the level of the assessment factor. In consideration of the visual and descriptive analyses, the results of the ANOVA suggested that Group A's post-treatment mean TAI-E score was significantly lower than Group B's post-wait mean TAI-E score. A two-sample t-test comparing the Group A (EMDR) TAI-E mean value at post-treatment (36.13) with the
### Table 12
Data Used for Determining Reliable Change Indices:
Outcomes in Groups A and B on TAI-T

<table>
<thead>
<tr>
<th>Data</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normative Mean for the Functional Female Population on the TAI-T</td>
<td>42.79</td>
</tr>
<tr>
<td>Normative Mean for the Functional Male Population on the TAI-T</td>
<td>38.48</td>
</tr>
<tr>
<td>Initial Screening Mean for Groups A &amp; B Females Combined on the TAI-T</td>
<td>64.86</td>
</tr>
<tr>
<td>Initial Screening Mean for Groups A &amp; B Males Combined on the TAI-T</td>
<td>62.91</td>
</tr>
<tr>
<td>Standard Deviation for Groups A &amp; B Females Combined on the TAI-T</td>
<td>7.41</td>
</tr>
<tr>
<td>Standard Deviation for Groups A &amp; B Males Combined on the TAI-T</td>
<td>7.01</td>
</tr>
<tr>
<td>Standard Deviation for the Functional Female Population on the TAI-T</td>
<td>13.70</td>
</tr>
<tr>
<td>Standard Deviation for the Functional Male Population on the TAI-T</td>
<td>12.43</td>
</tr>
<tr>
<td>Test-Retest Reliability of the TAI-T (Based on Group B’s Initial Screening to Post-Wait Correlation)</td>
<td>0.78</td>
</tr>
<tr>
<td>Standard Error of Measurement for Females on the TAI-T</td>
<td>3.48</td>
</tr>
<tr>
<td>Standard Error of Measurement for Males on the TAI-T</td>
<td>3.29</td>
</tr>
<tr>
<td>Standard Error of Difference for Females on the TAI-T</td>
<td>4.92</td>
</tr>
<tr>
<td>Standard Error of Difference for Males on the TAI-T</td>
<td>4.65</td>
</tr>
</tbody>
</table>

* Raw Scores

Group B (no-treatment control) TAI-E mean value at post-wait (84.00) verified that this intergroup mean difference reached significance ($p < .0001$) (Table 11).

The clinical significance of TAI-E reductions for each member of Group A was analyzed in the same fashion as described for TAI-T reductions. Table 15 lists the data used in determining the RC indices for Groups A and B on the TAI-E. Table 16
Table 13

Clinical Significance (CS) Determination of TAI-T Outcomes: Group A Subjects (EMDR Treatment) and Group B Subjects (No-Treatment Waiting Control)

<table>
<thead>
<tr>
<th>Group A Subjects Gender</th>
<th>TAI-T Initial Screening &amp; Post-Tx Raw Scores</th>
<th>Obtained RC Index CS?</th>
<th>Group B Subjects Gender</th>
<th>TAI-T Initial Screening &amp; Post-Wait Raw Scores</th>
<th>Obtained RC Index CS?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>59, 43</td>
<td>3.25 Yes</td>
<td>Female</td>
<td>61, 47</td>
<td>2.85 Yes</td>
</tr>
<tr>
<td>Female</td>
<td>55, 49</td>
<td>1.22 No</td>
<td>Male</td>
<td>67, 65</td>
<td>0.43 No</td>
</tr>
<tr>
<td>Male</td>
<td>60, 31</td>
<td>6.24 Yes</td>
<td>Male</td>
<td>70, 69</td>
<td>0.22 No</td>
</tr>
<tr>
<td>Female</td>
<td>53, 38</td>
<td>3.05 Yes</td>
<td>Female</td>
<td>58, 56</td>
<td>0.41 No</td>
</tr>
<tr>
<td>Male</td>
<td>70, 24</td>
<td>9.89 Yes</td>
<td>Female</td>
<td>63, 59</td>
<td>0.81 No</td>
</tr>
<tr>
<td>Female</td>
<td>59, 33</td>
<td>5.28 Yes</td>
<td>Female</td>
<td>67, 67</td>
<td>0.00 No</td>
</tr>
<tr>
<td>Female</td>
<td>77, 51</td>
<td>5.28 Yes</td>
<td>Female</td>
<td>65, 49</td>
<td>3.25 Yes</td>
</tr>
<tr>
<td>Male</td>
<td>58, 48</td>
<td>2.15 Yes</td>
<td>Male</td>
<td>63, 45</td>
<td>3.87 Yes</td>
</tr>
<tr>
<td>Female</td>
<td>70, 46</td>
<td>4.88 Yes</td>
<td>Female</td>
<td>77, 78</td>
<td>-0.20 No</td>
</tr>
<tr>
<td>Male</td>
<td>71, 25</td>
<td>9.89 Yes</td>
<td>Male</td>
<td>59, 60</td>
<td>-0.22 No</td>
</tr>
<tr>
<td>Female</td>
<td>70, 31</td>
<td>7.93 Yes</td>
<td>Male</td>
<td>49, 46</td>
<td>0.65 No</td>
</tr>
<tr>
<td>Female</td>
<td>61, 29</td>
<td>6.50 Yes</td>
<td>Female</td>
<td>76, 75</td>
<td>0.20 No</td>
</tr>
<tr>
<td>Female</td>
<td>56, 29</td>
<td>5.49 Yes</td>
<td>Female</td>
<td>65, 66</td>
<td>-0.20 No</td>
</tr>
<tr>
<td>Male</td>
<td>57, 38</td>
<td>4.09 Yes</td>
<td>Female</td>
<td>60, 64</td>
<td>-0.81 No</td>
</tr>
<tr>
<td>Female</td>
<td>73, 23</td>
<td>10.16 Yes</td>
<td>Female</td>
<td>64, 54</td>
<td>2.03 Yes</td>
</tr>
<tr>
<td>Male</td>
<td>67, 22</td>
<td>9.68 Yes</td>
<td>Female</td>
<td>73, 71</td>
<td>0.41 No</td>
</tr>
</tbody>
</table>

Method c Cutoff Scores: Females = 57.11; Males = 54.10.
Figure 6. A Comparison of the Effects of EMDR on the TAI-E in Group A With the Effects of Waiting on the TAI-E in Group B With the Effects of EMDR on the TAI-E in Group B.
Table 14

Analysis of Variance Summary Table for Two Factor Repeated Measures ANOVA on TAI-E

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>1</td>
<td>8235.56</td>
<td>8235.56</td>
<td>21.33</td>
<td>.0001</td>
</tr>
<tr>
<td>Error (Treatment)</td>
<td>30</td>
<td>11,581.44</td>
<td>386.05</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Assessment</td>
<td>1</td>
<td>14,762.25</td>
<td>14,762.25</td>
<td>66.10</td>
<td>.0001</td>
</tr>
<tr>
<td>Treatment X Assessment Interaction</td>
<td>1</td>
<td>10,150.56</td>
<td>10,150.56</td>
<td>45.45</td>
<td>.0001</td>
</tr>
<tr>
<td>Error (Treatment X Assessment)</td>
<td>30</td>
<td>6700.19</td>
<td>223.34</td>
<td>----</td>
<td>----</td>
</tr>
</tbody>
</table>

displays the gender-specific cutoff points, RC indices, and the results of the two-stage process for determining clinical significance.

The results of the method c analyses indicated that the post-treatment TAI-E scores for all 16 Group A subjects fell below the gender-specific cutoff points. Hence, at post-treatment, all 16 subject's TAI-E scores were more proximal to the functional rather than the dysfunctional population means. However, it should be noted that for one of these subjects the TAI-E score fell below the gender-specific cutoff point at the initial screening assessment.

The results of the RC analyses found that for one male subject the RC index failed to exceed the 1.96 critical value. The magnitude of his change did not reach statistical significance, thus did not satisfy the reliable change requirement for clinical
Table 15
Data Used for Determining Reliable Change Indices:
Outcomes in Groups A and B on TAI-E

<table>
<thead>
<tr>
<th>Data (Value)</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normative Mean for the Functional Female Population on the TAI-E</td>
<td>* 18.94</td>
</tr>
<tr>
<td>Normative Mean for the Functional Male Population on the TAI-E</td>
<td>* 16.85</td>
</tr>
<tr>
<td>Initial Screening Mean for Groups A &amp; B Females Combined on the TAI-E</td>
<td>* 28.29</td>
</tr>
<tr>
<td>Initial Screening Mean for Groups A &amp; B Males Combined on the TAI-E</td>
<td>* 25.45</td>
</tr>
<tr>
<td>Standard Deviation for Groups A &amp; B Females Combined on the TAI-E</td>
<td>3.00</td>
</tr>
<tr>
<td>Standard Deviation for Groups A &amp; B Males Combined on the TAI-E</td>
<td>4.32</td>
</tr>
<tr>
<td>Standard Deviation for the Functional Female Population on the TAI-E</td>
<td>6.31</td>
</tr>
<tr>
<td>Standard Deviation for the Functional Male Population on the TAI-E</td>
<td>5.64</td>
</tr>
<tr>
<td>Test-Retest Reliability of the TAI-E (Based on Group B's Initial Screening to Post-Wait Correlation)</td>
<td>0.70</td>
</tr>
<tr>
<td>Standard Error of Measurement for Females on the TAI-E</td>
<td>1.64</td>
</tr>
<tr>
<td>Standard Error of Measurement for Males on the TAI-E</td>
<td>2.37</td>
</tr>
<tr>
<td>Standard Error of Difference for Females on the TAI-E</td>
<td>2.32</td>
</tr>
<tr>
<td>Standard Error of Difference for Males on the TAI-E</td>
<td>3.35</td>
</tr>
</tbody>
</table>

* Raw Scores

significance. Overall, TAI-E outcomes for 15 of the 16 Group A subjects attained clinical significance.

Referring to Figure 7, visual inspection suggests that the mean reduction of 41.06 (s = 26.41) percentile points in Group A's TAI-W score from initial screening to post-treatment was substantially larger than the mean reduction of 8.00 (s = 14.07)
Table 16

Clinical Significance (CS) Determination of TAI-E Outcomes: Group A Subjects (EMDR Treatment) and Group B Subjects (No-Treatment Waiting Control)

Critical RC Index = 1.96, p < .05

<table>
<thead>
<tr>
<th>Gender</th>
<th>Initial Screening &amp; Post-Tx Raw Scores</th>
<th>Obtained RC Index</th>
<th>CS?</th>
<th>Group B Subjects Gender</th>
<th>Initial Screening &amp; Post-Wait Raw Scores</th>
<th>Obtained RC Index</th>
<th>CS?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female 26, 17</td>
<td>3.88</td>
<td>Yes</td>
<td>Female 26, 17</td>
<td>3.88</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female 26, 21</td>
<td>2.16</td>
<td>Yes</td>
<td>Male 27, 28</td>
<td>-0.30</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male 24, 10</td>
<td>4.18</td>
<td>Yes</td>
<td>Male 28, 29</td>
<td>-0.30</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female 27, 18</td>
<td>3.88</td>
<td>Yes</td>
<td>Female 27, 27</td>
<td>0.00</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male 29, 9</td>
<td>5.97</td>
<td>Yes</td>
<td>Female 27, 24</td>
<td>1.29</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female 26, 13</td>
<td>5.60</td>
<td>Yes</td>
<td>Female 31, 32</td>
<td>-0.43</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female 32, 20</td>
<td>5.17</td>
<td>Yes</td>
<td>Female 27, 18</td>
<td>3.88</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male 24, 20</td>
<td>1.19</td>
<td>No</td>
<td>Male 28, 21</td>
<td>2.10</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female 31, 19</td>
<td>5.17</td>
<td>Yes</td>
<td>Female 32, 31</td>
<td>0.43</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male 25, 9</td>
<td>7.16</td>
<td>Yes</td>
<td>Male 24, 24</td>
<td>0.00</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female 29, 13</td>
<td>6.90</td>
<td>Yes</td>
<td>Male 14, 16</td>
<td>-0.60</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female 30, 12</td>
<td>7.76</td>
<td>Yes</td>
<td>Female 32, 32</td>
<td>0.00</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female 21, 12</td>
<td>3.88</td>
<td>Yes</td>
<td>Female 27, 29</td>
<td>-0.86</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male 28, 17</td>
<td>3.28</td>
<td>Yes</td>
<td>Female 25, 27</td>
<td>-0.86</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female 32, 10</td>
<td>5.17</td>
<td>Yes</td>
<td>Female 32, 30</td>
<td>0.86</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male 30, 9</td>
<td>6.27</td>
<td>Yes</td>
<td>Female 28, 28</td>
<td>0.00</td>
<td>No</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Method C Cutoff Scores: Females = 25.28; Males = 21.72
Figure 7. A Comparison of the Effects of EMDR on the TAI-W in Group A With the Effects of Waiting on the TAI-W in Group B With the Effects of EMDR on the TAI-W in Group B.
percentile points in Group B's TAI-W score from initial screening to post-wait. A two
factor ANOVA on TAI-W scores (2 groups X 2 assessment occasions) showed a
significant main effect for the groups (p < .0001), a significant main effect for the
repeated assessment (p < .0001), and a significant treatment by assessment interaction
(p < .0001) (Table 17).

These results indicated that, on the TAI-W, the main effect of the treatment
factor did not consistently reach significance across the main effect of the assessment
factor. Rather, the significance of the treatment factor was contingent on the level of
the assessment factor. In consideration of the visual and descriptive analyses, the
results of the ANOVA suggested that Group A's post-treatment mean TAI-W score was
significantly lower than Group B's post-wait mean TAI-W score. A two-sample t-test
comparing the Group A (EMDR) TAI-W mean value at post-treatment (44.88) with the
Group B (no-treatment control) TAI-W mean value at post-wait (83.94) verified that
this intergroup mean difference reached significance (p < .0001) (Table 11).

The clinical significance of TAI-W reductions was analyzed for each member of
Group A in the same fashion as for TAI-T and TAI-E reductions. Table 18 lists the
data used in determining the RC indices for Groups A and B on the TAI-W. Table 19
displays the gender-specific cutoff points, RC indices, and the results of the two-stage
process for determining clinical significance.

The results of the method c analyses indicated that the post-treatment TAI-W
scores for all 16 Group A subjects fell below the gender-specific cutoff points. Hence,
at post-treatment, all 16 subject's TAI-E scores were more proximal to the functional
rather than the dysfunctional population means. However, it should be noted that for
five of these subjects the TAI-W score fell below the gender-specific cutoff point at the
initial screening assessment.
Table 17
Analysis of Variance Summary Table for Two Factor Repeated Measures ANOVA on TAI-W

<table>
<thead>
<tr>
<th>Source</th>
<th>dF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>1</td>
<td>8122.52</td>
<td>8122.52</td>
<td>20.36</td>
<td>.0001</td>
</tr>
<tr>
<td>Error (Treatment)</td>
<td>30</td>
<td>11,968.09</td>
<td>398.94</td>
<td></td>
<td>---</td>
</tr>
<tr>
<td>Assessment</td>
<td>1</td>
<td>9628.52</td>
<td>9628.52</td>
<td>43.02</td>
<td>.0001</td>
</tr>
<tr>
<td>Treatment X Assessment</td>
<td>1</td>
<td>4372.52</td>
<td>4372.52</td>
<td>19.54</td>
<td>.0001</td>
</tr>
<tr>
<td>Error (Treatment X Assessment)</td>
<td>30</td>
<td>6714.47</td>
<td>223.82</td>
<td></td>
<td>---</td>
</tr>
</tbody>
</table>

The results of the RC analyses found that for five subjects (three females and two males) the RC index failed to exceed the 1.96 critical value. The magnitude of these changes did not reach statistical significance, thus did not satisfy the reliable change requirement for clinical significance. Overall, TAI-W outcomes for 11 of the 16 Group A subjects attained clinical significance.

Referring to Figure 8, visual inspection suggests that the mean reduction of 36.44 (s = 20.85) raw score points in Group A’s RTT-T score from initial screening to post-treatment was substantially larger than the mean reduction of 5.25 (s = 7.21) raw score points in Group B’s RTT-T score from initial screening to post-wait. A two factor ANOVA on these RTT-T scores (2 groups X 2 assessment occasions) showed a significant main effect for the groups (p < .0001), a significant main effect for the
Table 18
Data Used for Determining Reliable Change Indices:
Outcomes in Groups A and B on TAI-W

<table>
<thead>
<tr>
<th>Data</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normative Mean for the Functional Female Population on the TAI-W</td>
<td>* 14.90</td>
</tr>
<tr>
<td>Normative Mean for the Functional Male Population on the TAI-W</td>
<td>* 13.61</td>
</tr>
<tr>
<td>Initial Screening Mean for Groups A &amp; B Females Combined on the TAI-W</td>
<td>* 23.00</td>
</tr>
<tr>
<td>Initial Screening Mean for Groups A &amp; B Males Combined on the TAI-W</td>
<td>* 24.27</td>
</tr>
<tr>
<td>Standard Deviation for Groups A &amp; B Females Combined on the TAI-W</td>
<td>5.24</td>
</tr>
<tr>
<td>Standard Deviation for Groups A &amp; B Males Combined on the TAI-W</td>
<td>5.04</td>
</tr>
<tr>
<td>Standard Deviation for the Functional Female Population on the TAI-W</td>
<td>5.51</td>
</tr>
<tr>
<td>Standard Deviation for the Functional Male Population on the TAI-W</td>
<td>4.98</td>
</tr>
<tr>
<td>Test-Retest Reliability of the TAI-W (Based on Group B’s Initial Screening to Post-Wait Correlation)</td>
<td>.85</td>
</tr>
<tr>
<td>Standard Error of Measurement for Females on the TAI-W</td>
<td>2.03</td>
</tr>
<tr>
<td>Standard Error of Measurement for Males on the TAI-W</td>
<td>1.95</td>
</tr>
<tr>
<td>Standard Error of Difference for Females on the TAI-W</td>
<td>2.87</td>
</tr>
<tr>
<td>Standard Error of Difference for Males on the TAI-W</td>
<td>2.76</td>
</tr>
</tbody>
</table>

* Raw Scores

repeated assessment (p < .0001), and a significant treatment by assessment interaction (p < .0001) (Table 20).

These results indicated that, on the RTT-T, the main effect of the treatment factor did not consistently reach significance across the main effect of the assessment factor. Rather, the significance of the treatment factor was contingent on the level of
Table 19
Clinical Significance (CS) Determination of TAI-W Outcomes: Group A Subjects (EMDR Treatment) and Group B Subjects (No-Treatment Waiting Control)

Critical RC Index = 1.96, p < .05

<table>
<thead>
<tr>
<th>Group A Subjects Gender</th>
<th>TAI-W Initial Screening &amp; Post-Tx Raw Scores</th>
<th>Obtained RC Index</th>
<th>CS?</th>
<th>Group B Subjects Gender</th>
<th>TAI-W Initial Screening &amp; Post-Wait Raw Scores</th>
<th>Obtained RC Index</th>
<th>CS?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>22, 16</td>
<td>2.09</td>
<td>Yes</td>
<td>Female</td>
<td>21, 19</td>
<td>0.70</td>
<td>No</td>
</tr>
<tr>
<td>Female</td>
<td>16, 18</td>
<td>-0.35</td>
<td>No</td>
<td>Male</td>
<td>27, 29</td>
<td>-0.72</td>
<td>No</td>
</tr>
<tr>
<td>Male</td>
<td>25, 12</td>
<td>4.71</td>
<td>Yes</td>
<td>Male</td>
<td>28, 26</td>
<td>0.72</td>
<td>No</td>
</tr>
<tr>
<td>Female</td>
<td>13, 9</td>
<td>1.39</td>
<td>No</td>
<td>Female</td>
<td>17, 15</td>
<td>0.70</td>
<td>No</td>
</tr>
<tr>
<td>Male</td>
<td>30, 10</td>
<td>7.25</td>
<td>Yes</td>
<td>Female</td>
<td>26, 23</td>
<td>1.05</td>
<td>No</td>
</tr>
<tr>
<td>Female</td>
<td>20, 13</td>
<td>5.92</td>
<td>Yes</td>
<td>Female</td>
<td>22, 23</td>
<td>-0.35</td>
<td>No</td>
</tr>
<tr>
<td>Female</td>
<td>30, 19</td>
<td>3.83</td>
<td>Yes</td>
<td>Female</td>
<td>24, 23</td>
<td>0.35</td>
<td>No</td>
</tr>
<tr>
<td>Male</td>
<td>17, 16</td>
<td>0.36</td>
<td>No</td>
<td>Male</td>
<td>19, 13</td>
<td>2.17</td>
<td>Yes</td>
</tr>
<tr>
<td>Female</td>
<td>23, 15</td>
<td>2.79</td>
<td>Yes</td>
<td>Female</td>
<td>30, 27</td>
<td>1.05</td>
<td>No</td>
</tr>
<tr>
<td>Male</td>
<td>31, 11</td>
<td>7.25</td>
<td>Yes</td>
<td>Male</td>
<td>23, 23</td>
<td>0.00</td>
<td>No</td>
</tr>
<tr>
<td>Female</td>
<td>29, 12</td>
<td>5.92</td>
<td>Yes</td>
<td>Male</td>
<td>26, 21</td>
<td>1.81</td>
<td>No</td>
</tr>
<tr>
<td>Female</td>
<td>15, 10</td>
<td>1.74</td>
<td>No</td>
<td>Female</td>
<td>29, 27</td>
<td>0.70</td>
<td>No</td>
</tr>
<tr>
<td>Female</td>
<td>25, 11</td>
<td>4.88</td>
<td>Yes</td>
<td>Female</td>
<td>25, 22</td>
<td>1.05</td>
<td>No</td>
</tr>
<tr>
<td>Male</td>
<td>16, 12</td>
<td>1.45</td>
<td>No</td>
<td>Female</td>
<td>20, 20</td>
<td>0.00</td>
<td>No</td>
</tr>
<tr>
<td>Female</td>
<td>26, 8</td>
<td>6.27</td>
<td>Yes</td>
<td>Female</td>
<td>19, 11</td>
<td>2.79</td>
<td>Yes</td>
</tr>
<tr>
<td>Male</td>
<td>25, 9</td>
<td>5.80</td>
<td>Yes</td>
<td>Female</td>
<td>31, 28</td>
<td>1.05</td>
<td>No</td>
</tr>
</tbody>
</table>

Method c Cutoff Scores: Females = 19.05; Males = 18.91

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 8. A Comparison of the Effects of EMDR on the RTT-T in Group A With the Effects of Waiting on the RTT-T in Group B With the Effects of EMDR on the RTT-T in Group B.
Table 20
Analysis of Variance Summary Table for Two Factor Repeated Measures ANOVA on RTT-T

<table>
<thead>
<tr>
<th>Source</th>
<th>dF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>1</td>
<td>735.77</td>
<td>735.77</td>
<td>19.97</td>
<td>.0001</td>
</tr>
<tr>
<td>Error (Treatment)</td>
<td>30</td>
<td>1105.47</td>
<td>36.85</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Assessment</td>
<td>1</td>
<td>1048.14</td>
<td>1048.14</td>
<td>67.68</td>
<td>.0001</td>
</tr>
<tr>
<td>Treatment X Assessment Interaction</td>
<td>1</td>
<td>534.77</td>
<td>534.77</td>
<td>34.53</td>
<td>.0001</td>
</tr>
<tr>
<td>Error (Treatment X Assessment)</td>
<td>30</td>
<td>464.59</td>
<td>15.49</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

the assessment factor. In consideration of the visual and descriptive analyses, the results of the ANOVA suggested that Group A’s post-treatment mean RTT-T score was significantly lower than Group B’s post-wait mean RTT-T score. A two-sample t-test comparing the Group A (EMDR) RTT-T mean value at post-treatment (64.94) with the Group B (no-treatment control) RTT-T mean value at post-wait (102.81) verified that this intergroup mean difference reached significance (p < .0001) (Table 11).

Clinical significance for the individual RTT-T outcomes was analyzed. The methodology was substantially less complex than that used for the TAI scales. Because the RTT-T did not provide normative data, only method a could be used. Using this method clinical significance was achieved when post-treatment scores fell at least two standard deviations below the initial screening mean. Based on the RTT-T mean of
104.72 and standard deviation of 16.50 (pooled across Groups A and B at initial screening), the critical cutoff value, below which individual scores achieved clinical significance, was 71.72.

The results of the method analysis found that post-treatment RTT-T scores for 11 of the 16 subjects fell below the critical cutoff value (Table 21). Hence, these 11 individual outcomes achieved clinical significance.

**Question 2**

Is the effectiveness of EMDR, as observed in the primary treatment group, replicated when applied to the waiting control group subsequent to the waiting control period as measured by the TAI-T, TAI-E, TAI-W, or RTT-T? Referring to Figure 5, visual inspection suggests that the mean reduction of 46.42 (s = 26.86) percentile points in Group B's TAI-T score from post-wait to post-treatment assessments was substantial. Moreover, this mean TAI-T reduction in Group B appears similar to Group A's (statistically significant) mean TAI-T reduction of 54.00 (s = 27.98) percentile points.

Statistical analyses supported these visual and descriptive conclusions. A one factor repeated measures ANOVA (1 group X 3 assessment occasions) on Group B's initial screening, post-wait, and post-treatment TAI-T scores indicated the presence of at least one significant difference among the three pairwise contrasts (p < .0001) (Table 22). A post-hoc, paired t-test contrasting the mean difference between Group B's post-wait (87.42) and post-treatment (41.00) TAI-T scores verified that this intragroup mean difference reached significance (p < .0001) (Table 23).

The equivalence of treatment outcomes between groups was evaluated using a post-hoc, two-sample t-test that compared Group B's TAI-T post-treatment mean

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 21

Clinical Significance (CS) Determination of RTT-T Outcomes: Group A Subjects (EMDR Treatment) and Group B Subjects (No-Treatment Waiting Control)

<table>
<thead>
<tr>
<th>Group A Subjects Gender</th>
<th>RTT-T Initial Screening &amp; Post-Tx Raw Scores</th>
<th>Standard Deviations Below Mean CS?</th>
<th>Group B Subjects Gender</th>
<th>RTT-T Initial Screening &amp; Post-Wait Raw Scores</th>
<th>Standard Deviations Below Mean CS?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>98, 75</td>
<td>1.80</td>
<td>No</td>
<td>Female</td>
<td>98, 86</td>
</tr>
<tr>
<td>Female</td>
<td>85, 73</td>
<td>1.92</td>
<td>No</td>
<td>Male</td>
<td>114, 110</td>
</tr>
<tr>
<td>Male</td>
<td>93, 65</td>
<td>2.41</td>
<td>Yes</td>
<td>Male</td>
<td>128, 124</td>
</tr>
<tr>
<td>Female</td>
<td>73, 62</td>
<td>2.59</td>
<td>Yes</td>
<td>Female</td>
<td>84, 79</td>
</tr>
<tr>
<td>Male</td>
<td>135, 53</td>
<td>3.13</td>
<td>Yes</td>
<td>Female</td>
<td>109, 111</td>
</tr>
<tr>
<td>Female</td>
<td>96, 65</td>
<td>2.41</td>
<td>Yes</td>
<td>Female</td>
<td>118, 111</td>
</tr>
<tr>
<td>Female</td>
<td>116, 77</td>
<td>1.68</td>
<td>No</td>
<td>Female</td>
<td>116, 105</td>
</tr>
<tr>
<td>Male</td>
<td>82, 71</td>
<td>2.04</td>
<td>Yes</td>
<td>Male</td>
<td>97, 98</td>
</tr>
<tr>
<td>Female</td>
<td>114, 72</td>
<td>1.98</td>
<td>No</td>
<td>Female</td>
<td>126, 112</td>
</tr>
<tr>
<td>Male</td>
<td>110, 58</td>
<td>2.83</td>
<td>Yes</td>
<td>Male</td>
<td>114, 107</td>
</tr>
<tr>
<td>Female</td>
<td>129, 97</td>
<td>0.47</td>
<td>No</td>
<td>Male</td>
<td>101, 87</td>
</tr>
<tr>
<td>Female</td>
<td>88, 52</td>
<td>3.20</td>
<td>Yes</td>
<td>Female</td>
<td>115, 115</td>
</tr>
<tr>
<td>Female</td>
<td>73, 51</td>
<td>3.26</td>
<td>Yes</td>
<td>Female</td>
<td>99, 107</td>
</tr>
<tr>
<td>Male</td>
<td>101, 68</td>
<td>2.23</td>
<td>Yes</td>
<td>Female</td>
<td>91, 98</td>
</tr>
<tr>
<td>Female</td>
<td>127, 53</td>
<td>3.13</td>
<td>Yes</td>
<td>Female</td>
<td>97, 84</td>
</tr>
<tr>
<td>Male</td>
<td>102, 47</td>
<td>3.50</td>
<td>Yes</td>
<td>Female</td>
<td>122, 111</td>
</tr>
</tbody>
</table>

RTT-T: Mean = 104.72, Standard Deviation = 16.50; Critical Cutoff Value = 71.72
Table 22
Analysis of Variance Summary Table for One Factor Repeated Measures ANOVA on TAI-T in Group B: Initial Screening, Post-Wait, Post-Treatment

<table>
<thead>
<tr>
<th>Source</th>
<th>dF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment</td>
<td>2</td>
<td>19,948.17</td>
<td>9974.08</td>
<td>44.11</td>
<td>.0001</td>
</tr>
<tr>
<td>Error (Assessment)</td>
<td>22</td>
<td>4975.17</td>
<td>226.14</td>
<td>----</td>
<td>----</td>
</tr>
</tbody>
</table>

Table 23
Descriptive Data and Paired t-tests on TAI-T, TAI-E, TAI-W, and RTT-T: Mean Differences Between Group B Post-Wait and Post-Treatment

<table>
<thead>
<tr>
<th>Measure</th>
<th>Group B Post-Wait Mean</th>
<th>Standard Deviation</th>
<th>Group B Post-Treatment Mean</th>
<th>Standard Deviation</th>
<th>Obtained p-value</th>
<th>Critical p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAI-T</td>
<td>87.42</td>
<td>10.10</td>
<td>41.00</td>
<td>22.79</td>
<td>.0001</td>
<td>.0125</td>
</tr>
<tr>
<td>TAI-E</td>
<td>84.92</td>
<td>17.03</td>
<td>38.67</td>
<td>25.26</td>
<td>.0003</td>
<td>.0125</td>
</tr>
<tr>
<td>TAI-W</td>
<td>82.58</td>
<td>22.05</td>
<td>42.92</td>
<td>28.00</td>
<td>.0001</td>
<td>.0125</td>
</tr>
<tr>
<td>RTT-T</td>
<td>* 103.50</td>
<td>13.69</td>
<td>* 67.83</td>
<td>13.06</td>
<td>.0001</td>
<td>.0125</td>
</tr>
</tbody>
</table>

* Raw scores—all other mean values are percentile scores

(41.00) with Group A’s TAI-T post-treatment mean (37.81). This intergroup mean difference did not reach significance (p < .7311) (Table 24).

The clinical significance (CS) of TAI-T reductions was analyzed for each member of Group B using the same methodology described for all TAI reductions in
Table 24

Descriptive Data and Two-Sample t-tests on TAI-T, TAI-E, TAI-W, and RTT-T: Mean Differences Between Group A Post-Treatment and Group B Post-Treatment

<table>
<thead>
<tr>
<th>Measure</th>
<th>Group A Post-Treatment Mean</th>
<th>Group A Standard Deviation</th>
<th>Group B Post-Treatment Mean</th>
<th>Group B Standard Deviation</th>
<th>Obtained p-value</th>
<th>Critical p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAI-T</td>
<td>37.81</td>
<td>25.56</td>
<td>41.00</td>
<td>22.79</td>
<td>.7311</td>
<td>.0125</td>
</tr>
<tr>
<td>TAI-E</td>
<td>36.13</td>
<td>24.55</td>
<td>38.67</td>
<td>25.26</td>
<td>.7920</td>
<td>.0125</td>
</tr>
<tr>
<td>TAI-W</td>
<td>44.88</td>
<td>23.23</td>
<td>42.92</td>
<td>28.00</td>
<td>.8459</td>
<td>.0125</td>
</tr>
<tr>
<td>RTT-T</td>
<td>* 64.94</td>
<td>12.81</td>
<td>* 67.83</td>
<td>13.05</td>
<td>.5638</td>
<td>.0125</td>
</tr>
</tbody>
</table>

* Raw scores--all other mean differences are percentile scores

Group A. The entire initial screening to post-treatment period (which included the no-treatment waiting condition) was evaluated because the Jacobson and Truax (1991) RC index compensates for variables that produce regression toward the mean. Therefore, removing the changes that occurred during the waiting control period for the RC computation would have been redundant and led to inaccurate conclusions regarding the clinical significance of EMDR effects in Group B. Table 12 lists the data used in determining the RC indices for Group B on the TAI-T. Table 25 displays the gender-specific cutoff points, RC indices, and the results of the two-stage process for determining clinical significance.

The results of the method c analyses indicated that the post-treatment TAI-T scores for all 12 of the Group B subjects who completed EMDR treatment fell below the gender-specific cutoff points. Hence, at post-treatment, all 12 of the Group B
Table 25

Clinical Significance (CS) Determination of TAI-T Outcomes: Group A Subjects (EMDR Treatment) and Group B Subjects (EMDR Treatment)

<table>
<thead>
<tr>
<th>Group A Subjects Gender</th>
<th>TAI-T Initial Screening &amp; Post-Tx Raw Scores</th>
<th>Obtained RC Index</th>
<th>CS?</th>
<th>Group B Subjects Gender</th>
<th>TAI-T Post-Wait &amp; Post-Tx Raw Scores</th>
<th>Obtained RC Index</th>
<th>CS?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>59, 43</td>
<td>3.25</td>
<td>Yes</td>
<td>Female</td>
<td>61, ---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Female</td>
<td>55, 49</td>
<td>1.22</td>
<td>No</td>
<td>Male</td>
<td>67, 32</td>
<td>7.53</td>
<td>Yes</td>
</tr>
<tr>
<td>Male</td>
<td>60, 31</td>
<td>6.24</td>
<td>Yes</td>
<td>Male</td>
<td>70, 26</td>
<td>9.46</td>
<td>Yes</td>
</tr>
<tr>
<td>Female</td>
<td>53, 38</td>
<td>3.05</td>
<td>Yes</td>
<td>Female</td>
<td>58, 34</td>
<td>4.88</td>
<td>Yes</td>
</tr>
<tr>
<td>Male</td>
<td>70, 24</td>
<td>9.89</td>
<td>Yes</td>
<td>Female</td>
<td>63, 27</td>
<td>7.32</td>
<td>Yes</td>
</tr>
<tr>
<td>Female</td>
<td>59, 33</td>
<td>5.28</td>
<td>Yes</td>
<td>Female</td>
<td>67, 30</td>
<td>7.52</td>
<td>Yes</td>
</tr>
<tr>
<td>Female</td>
<td>77, 51</td>
<td>5.28</td>
<td>Yes</td>
<td>Female</td>
<td>65, 47</td>
<td>3.66</td>
<td>Yes</td>
</tr>
<tr>
<td>Male</td>
<td>58, 48</td>
<td>2.15</td>
<td>Yes</td>
<td>Male</td>
<td>63, 31</td>
<td>6.88</td>
<td>Yes</td>
</tr>
<tr>
<td>Female</td>
<td>70, 46</td>
<td>4.88</td>
<td>Yes</td>
<td>Female</td>
<td>77, ---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Male</td>
<td>71, 25</td>
<td>9.89</td>
<td>Yes</td>
<td>Male</td>
<td>59, 49</td>
<td>2.15</td>
<td>Yes</td>
</tr>
<tr>
<td>Female</td>
<td>70, 31</td>
<td>7.93</td>
<td>Yes</td>
<td>Male</td>
<td>49, 32</td>
<td>3.66</td>
<td>Yes</td>
</tr>
<tr>
<td>Female</td>
<td>61, 29</td>
<td>6.50</td>
<td>Yes</td>
<td>Female</td>
<td>76, 41</td>
<td>7.11</td>
<td>Yes</td>
</tr>
<tr>
<td>Female</td>
<td>56, 29</td>
<td>5.49</td>
<td>Yes</td>
<td>Female</td>
<td>65, ---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Male</td>
<td>57, 38</td>
<td>4.09</td>
<td>Yes</td>
<td>Female</td>
<td>60, ---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Female</td>
<td>73, 23</td>
<td>10.16</td>
<td>Yes</td>
<td>Female</td>
<td>64, 50</td>
<td>2.85</td>
<td>Yes</td>
</tr>
<tr>
<td>Male</td>
<td>67, 22</td>
<td>9.68</td>
<td>Yes</td>
<td>Female</td>
<td>73, 36</td>
<td>7.52</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Method c Cutoff Scores: Females = 57.11; Males = 54.10; --- Did not Complete Tx

Critical RC Index = 1.96, p < .05

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
subject's TAI-T scores were more proximal to the functional rather than the dysfunctional population means. However, it should be noted that for one of the 12 subjects the TAI-T score fell below the gender-specific cutoff point at the initial screening assessment (prior to treatment).

The results of the RC analyses found that the TAI-T reductions for all 12 Group B subjects who completed treatment exceeded the 1.96 critical value. Thus, all 12 reached statistical significance. Considered together with the results of the method c analysis, the RC indices indicated that all 12 Group B subjects who completed EMDR treatment attained clinical significance on the TAI-T.

Referring to Figure 6, visual inspection suggests that the mean reduction of 46.25 (s = 30.26) percentile points in Group B's TAI-E score from post-wait to post-treatment assessments was substantial. Moreover, this mean TAI-E reduction in Group B appears similar to Group A's (statistically significant) mean TAI-E reduction of 55.56 (s = 25.87) percentile points.

Statistical analyses supported these visual and descriptive conclusions. A one factor repeated measures ANOVA (1 group X 3 assessment occasions) on Group B's initial screening, post-wait, and post-treatment TAI-E scores indicated the presence of at least one significant difference among the three pairwise contrasts (p < .0001) (Table 26). A post-hoc, paired t-test contrasting the mean difference between Group B's post-wait (84.92) and post-treatment (38.67) TAI-E scores verified that this intragroup mean difference reached significance (p < .0003) (Table 23).

The equivalence of treatment effects between groups was evaluated using a post-hoc, two-sample t-test that compared Group B's TAI-E post-treatment mean (38.67) with Group A's TAI-E post-treatment mean (36.13). This intragroup mean difference did not reach significance (p < .7920) (Table 24).
Table 26

Analysis of Variance Summary Table for One Factor Repeated Measures ANOVA on TAI-E in Group B: Initial Screening, Post-Wait, Post-Treatment

<table>
<thead>
<tr>
<th>Source</th>
<th>dF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment</td>
<td>2</td>
<td>16,363.39</td>
<td>8181.69</td>
<td>34.85</td>
<td>.0001</td>
</tr>
<tr>
<td>Error (Assessment)</td>
<td>22</td>
<td>5164.61</td>
<td>234.76</td>
<td>----</td>
<td>----</td>
</tr>
</tbody>
</table>

The clinical significance of TAI-E reductions was analyzed for each member of Group B using the same methodology as for TAI-T reductions in Group B. Table 15 lists the data used in determining the RC indices for Group B on the TAI-E. Table 27 displays the gender-specific cutoff points, RC indices, and the results of the two-stage process for determining clinical significance.

The results of the method c analyses indicated that the post-treatment TAI-E scores for 11 of the 12 Group B subjects who completed EMDR treatment (after the initial waiting period) fell below the gender-specific cutoff points. Hence, at post-treatment, 11 of the 12 treated Group B TAI-E scores were more proximal to the functional rather than the dysfunctional population means. However, it should be noted that for two of the 12 subjects the TAI-E score fell below the gender-specific cutoff point at the initial screening assessment.

The results of the RC analyses found that for one female and two male subjects the RC index failed to exceed the 1.96 critical value. The magnitude of these individual changes did not reach statistical significance, thus did not satisfy the reliable change requirement for clinical significance. Overall, TAI-T outcomes for nine of the 12 Group B subjects who completed EMDR treatment reached clinical significance.
Table 27
Clinical Significance (CS) Determination of TAI-E Outcomes: Group A Subjects (EMDR Treatment) and Group B Subjects (EMDR Treatment)

<table>
<thead>
<tr>
<th>Group A Subjects</th>
<th>TAI-E Initial Screening &amp; Post-Tx Raw Scores</th>
<th>Obtained RC Index CS?</th>
<th>Group B Subjects</th>
<th>TAI-E Initial Screening &amp; Post-Tx Raw Scores</th>
<th>Obtained RC Index CS?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female 26, 17</td>
<td>3.88</td>
<td>Yes</td>
<td>Female 26, ---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Female 26, 21</td>
<td>2.16</td>
<td>Yes</td>
<td>Male 27, 13</td>
<td>4.18</td>
<td>Yes</td>
</tr>
<tr>
<td>Male 24, 10</td>
<td>4.18</td>
<td>Yes</td>
<td>Male 28, 10</td>
<td>5.37</td>
<td>Yes</td>
</tr>
<tr>
<td>Female 27, 18</td>
<td>3.88</td>
<td>Yes</td>
<td>Female 27, 12</td>
<td>6.47</td>
<td>Yes</td>
</tr>
<tr>
<td>Male 29, 9</td>
<td>5.97</td>
<td>Yes</td>
<td>Female 27, 19</td>
<td>3.45</td>
<td>Yes</td>
</tr>
<tr>
<td>Female 26, 13</td>
<td>5.60</td>
<td>Yes</td>
<td>Female 31, 13</td>
<td>7.76</td>
<td>Yes</td>
</tr>
<tr>
<td>Female 32, 20</td>
<td>5.17</td>
<td>Yes</td>
<td>Female 27, 19</td>
<td>3.45</td>
<td>Yes</td>
</tr>
<tr>
<td>Male 24, 20</td>
<td>1.19</td>
<td>No</td>
<td>Male 28, 15</td>
<td>3.88</td>
<td>Yes</td>
</tr>
<tr>
<td>Female 31, 19</td>
<td>5.17</td>
<td>Yes</td>
<td>Female 32, ---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Male 25, 9</td>
<td>7.16</td>
<td>Yes</td>
<td>Male 24, 19</td>
<td>1.49</td>
<td>No</td>
</tr>
<tr>
<td>Female 29, 13</td>
<td>6.90</td>
<td>Yes</td>
<td>Male 14, 11</td>
<td>0.90</td>
<td>No</td>
</tr>
<tr>
<td>Female 30, 12</td>
<td>7.76</td>
<td>Yes</td>
<td>Female 32, 16</td>
<td>6.90</td>
<td>Yes</td>
</tr>
<tr>
<td>Female 21, 12</td>
<td>3.88</td>
<td>Yes</td>
<td>Female 27, ---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Male 28, 17</td>
<td>3.28</td>
<td>Yes</td>
<td>Female 25, ---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Female 32, 10</td>
<td>5.17</td>
<td>Yes</td>
<td>Female 32, 28</td>
<td>1.72</td>
<td>No</td>
</tr>
<tr>
<td>Male 30, 9</td>
<td>6.27</td>
<td>Yes</td>
<td>Female 28, 16</td>
<td>5.17</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Method c Cutoff Scores: Females = 25.28; Males = 21.72. --- Didn’t Complete Tx

Critical RC Index = 1.96, p < .05

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Referring to Figure 7, visual inspection suggests that the mean reduction of 39.67 (s = 23.75) percentile points in Group B's TAI-W score from post-wait to post-treatment assessments was substantial. Moreover, this mean TAI-W reduction in Group B appears similar to Group A's (statistically significant) mean TAI-W reduction of 41.06 (s = 26.41) percentile points.

Statistical analyses supported these visual and descriptive conclusions. A one factor repeated measures ANOVA (1 group X 3 assessment occasions) on Group B's initial screening, post-wait, and post-treatment TAI-W scores indicated the presence of at least one significant difference among the three pairwise contrasts (p < .0001) (Table 28). A post-hoc, paired t-test contrasting the mean difference between Group B's post-

<table>
<thead>
<tr>
<th>Source</th>
<th>dF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment</td>
<td>2</td>
<td>18,866.06</td>
<td>9433.03</td>
<td>32.11</td>
<td>.0001</td>
</tr>
<tr>
<td>Error (Assessment)</td>
<td>22</td>
<td>6463.94</td>
<td>292.82</td>
<td>------</td>
<td>------</td>
</tr>
</tbody>
</table>

wait (82.58) and post-treatment (42.92) TAI-W scores verified that this intragroup mean difference reached significance (p < .0001) (Table 23).

The equivalence of treatment effects between groups was evaluated using a post-hoc, two-sample t-test that compared Group B's TAI-W post-treatment mean (42.92) with Group A's TAI-W post-treatment mean (44.88). This intragroup mean difference did not reach significance (p < .8459) (Table 24).
The clinical significance of the individual TAI-W reductions was analyzed for each member of Group B using the same methodology as for TAI-T reductions in Group B. Table 18 lists the data used in determining the RC indices for Group B on the TAI-W. Table 29 displays the gender-specific cutoff points, RC indices, and the results of the two-stage process for determining clinical significance.

The results of the method c analyses indicated that the post-treatment TAI-W scores for 11 of the 12 Group B subjects who completed EMDR treatment (after the initial waiting period) fell below the gender-specific cutoff points. Hence, at post-treatment, 11 of the 12 treated Group B TAI-W scores were more proximal to the functional rather than the dysfunctional population means. However, it should be noted that for one of the 12 subjects the TAI-W score fell below the gender-specific cutoff point at the initial screening assessment (prior to treatment).

The results of the RC analyses found that the TAI-W reductions for 11 of the 12 treated Group B subjects exceeded the 1.96 critical value and reached statistical significance. Considered together with the results of the method c analysis, the RC indices indicated that 11 of the 12 Group B subjects who completed EMDR treatment attained clinical significance on the TAI-W.

Referring to Figure 8, visual inspection suggests that the mean reduction of 35.67 (s = 19.61) raw score points in Group B’s RTT-T score from post-wait to post-treatment assessments was substantial. Moreover, this mean RTT-T reduction in Group B appears similar to Group A’s (statistically significant) mean RTT-T reduction of 36.44 (s = 20.85) percentile points.

Statistical analyses supported these visual and descriptive conclusions. A one factor repeated measures ANOVA (1 group X 3 assessment occasions) on Group B’s initial screening, post-wait, and post-treatment RTT-T scores indicated the presence of at least one significant difference among the three pairwise contrasts (p < .0001)
Table 29
Clinical Significance (CS) Determination of TAI-W Outcomes: Group A Subjects (EMDR Treatment) and Group B Subjects (EMDR Treatment)

Standard Error of Difference: Females = 2.87; Males = 2.76

<table>
<thead>
<tr>
<th>Group A Subjects Gender</th>
<th>TAI-W Initial Raw Scores &amp; Post-Tx Obtained RC Index CS?</th>
<th>Group B Subjects Gender</th>
<th>TAI-W Initial Raw Scores &amp; Post-Tx Obtained RC Index CS?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female 22, 16</td>
<td>2.09 Yes Female</td>
<td>21, --</td>
<td>-- --</td>
</tr>
<tr>
<td>Female 16, 18</td>
<td>-0.35 No Male</td>
<td>27, 12</td>
<td>5.43 Yes</td>
</tr>
<tr>
<td>Male 25, 12</td>
<td>4.71 Yes Male</td>
<td>28, 11</td>
<td>6.16 Yes</td>
</tr>
<tr>
<td>Female 13, 9</td>
<td>1.39 No Female</td>
<td>17, 9</td>
<td>2.79 Yes</td>
</tr>
<tr>
<td>Male 30, 10</td>
<td>7.25 Yes Female</td>
<td>26, 8</td>
<td>6.27 Yes</td>
</tr>
<tr>
<td>Female 20, 13</td>
<td>5.92 Yes Female</td>
<td>22, 10</td>
<td>4.18 Yes</td>
</tr>
<tr>
<td>Female 30, 19</td>
<td>3.83 Yes Female</td>
<td>24, 18</td>
<td>2.09 Yes</td>
</tr>
<tr>
<td>Male 17, 16</td>
<td>0.36 No Male</td>
<td>19, 9</td>
<td>3.62 Yes</td>
</tr>
<tr>
<td>Female 23, 15</td>
<td>2.79 Yes Female</td>
<td>30, --</td>
<td>-- --</td>
</tr>
<tr>
<td>Male 31, 11</td>
<td>7.25 Yes Male</td>
<td>23, 20</td>
<td>1.09 No</td>
</tr>
<tr>
<td>Female 29, 12</td>
<td>5.92 Yes Male</td>
<td>26, 14</td>
<td>4.35 Yes</td>
</tr>
<tr>
<td>Female 15, 10</td>
<td>1.74 No Female</td>
<td>29, 16</td>
<td>4.53 Yes</td>
</tr>
<tr>
<td>Female 25, 11</td>
<td>4.88 Yes Female</td>
<td>25, --</td>
<td>-- --</td>
</tr>
<tr>
<td>Male 16, 12</td>
<td>1.45 No Female</td>
<td>20, --</td>
<td>-- --</td>
</tr>
<tr>
<td>Female 26, 8</td>
<td>6.27 Yes Female</td>
<td>19, 9</td>
<td>3.48 Yes</td>
</tr>
<tr>
<td>Male 25, 9</td>
<td>5.80 Yes Female</td>
<td>31, 14</td>
<td>5.92 Yes</td>
</tr>
</tbody>
</table>

Method c Cutoffs: Females = 19.05; Males = 18.91. Critical RC Index = 1.96, p < .05

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
(Table 30). A post-hoc, paired t-test contrasting the mean difference between Group B’s post-wait (103.50) and post-treatment (67.83) RTT-T scores verified that this intragroup mean difference reached significance (p < .0001) (Table 23).

Table 30
Analysis of Variance Summary Table for One Factor Repeated Measures ANOVA on RTT-T in Group B: Initial Screening, Post-Wait, Post-Treatment

<table>
<thead>
<tr>
<th>Source</th>
<th>dF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment</td>
<td>2</td>
<td>1542.72</td>
<td>771.36</td>
<td>63.97</td>
<td>.0001</td>
</tr>
<tr>
<td>Error (Assessment)</td>
<td>22</td>
<td>265.28</td>
<td>12.06</td>
<td>----</td>
<td>----</td>
</tr>
</tbody>
</table>

The equivalence of treatment effects between groups was evaluated using a post-hoc, two-sample t-test that compared Group B’s RTT-T post-treatment mean (67.83) with Group A’s RTT-T post-treatment mean (64.94). This analysis did not reach significance (p < .5638) (Table 24).

The clinical significance of RTT-T reductions was analyzed for each member of Group B in the same fashion as for Group A subjects. The results of the method analysis found that post-treatment RTT-T scores for nine of the 12 subjects who completed EMDR treatment were two or more standard deviations lower than the RTT-T mean (Table 31). Hence, these nine individual outcomes achieved clinical significance.

Question 3

Is there spontaneous remission of test anxiety over the waiting control period as measured by the TAI-T, TAI-E, TAI-W, or RTT-T? Referring to Figure 5, visual

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
## Table 31
Clinical Significance (CS) Determination of RTT-T Outcomes: Group A Subjects (EMDR Treatment) and Group B Subjects (EMDR Treatment)

<table>
<thead>
<tr>
<th>Group A Subjects Gender</th>
<th>RTT-T Initial Screening &amp; Post-Tx Raw Scores</th>
<th>Change in Standard Deviations CS?</th>
<th>Group B Subjects Gender</th>
<th>RTT-T Initial Screening &amp; Post-Tx Raw Scores</th>
<th>Change in Standard Deviations CS?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>98, 75</td>
<td>1.39 No</td>
<td>Female</td>
<td>98, ---</td>
<td>--- ---</td>
</tr>
<tr>
<td>Female</td>
<td>85, 73</td>
<td>0.73 No</td>
<td>Male</td>
<td>114, 63</td>
<td>3.09 Yes</td>
</tr>
<tr>
<td>Male</td>
<td>93, 65</td>
<td>1.70 No</td>
<td>Male</td>
<td>128, 53</td>
<td>4.55 Yes</td>
</tr>
<tr>
<td>Female</td>
<td>73, 62</td>
<td>0.67 No</td>
<td>Female</td>
<td>84, 65</td>
<td>1.15 No</td>
</tr>
<tr>
<td>Male</td>
<td>135, 53</td>
<td>4.97 Yes</td>
<td>Female</td>
<td>109, 54</td>
<td>3.33 Yes</td>
</tr>
<tr>
<td>Female</td>
<td>96, 65</td>
<td>1.88 No</td>
<td>Female</td>
<td>118, 66</td>
<td>3.15 Yes</td>
</tr>
<tr>
<td>Female</td>
<td>116, 77</td>
<td>2.36 Yes</td>
<td>Female</td>
<td>116, 83</td>
<td>2.00 Yes</td>
</tr>
<tr>
<td>Male</td>
<td>82, 71</td>
<td>0.67 No</td>
<td>Male</td>
<td>97, 57</td>
<td>2.42 Yes</td>
</tr>
<tr>
<td>Female</td>
<td>114, 72</td>
<td>2.55 Yes</td>
<td>Female</td>
<td>126, ---</td>
<td>--- ---</td>
</tr>
<tr>
<td>Male</td>
<td>110, 58</td>
<td>3.15 Yes</td>
<td>Male</td>
<td>114, 99</td>
<td>0.91 No</td>
</tr>
<tr>
<td>Female</td>
<td>129, 97</td>
<td>1.94 No</td>
<td>Male</td>
<td>101, 60</td>
<td>2.48 Yes</td>
</tr>
<tr>
<td>Female</td>
<td>88, 52</td>
<td>2.18 Yes</td>
<td>Female</td>
<td>115, 69</td>
<td>2.79 Yes</td>
</tr>
<tr>
<td>Female</td>
<td>73, 51</td>
<td>1.33 No</td>
<td>Female</td>
<td>99, ---</td>
<td>--- ---</td>
</tr>
<tr>
<td>Male</td>
<td>101, 68</td>
<td>2.00 Yes</td>
<td>Female</td>
<td>91, ---</td>
<td>--- ---</td>
</tr>
<tr>
<td>Female</td>
<td>127, 53</td>
<td>4.48 Yes</td>
<td>Female</td>
<td>97, 74</td>
<td>1.39 No</td>
</tr>
<tr>
<td>Male</td>
<td>102, 47</td>
<td>3.33 Yes</td>
<td>Female</td>
<td>122, 71</td>
<td>3.09 Yes</td>
</tr>
</tbody>
</table>

RTT-T Initial Screening Across Groups: Mean = 104.72; Standard Deviation = 16.50

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
inspection suggests that from initial screening to post-wait assessments there was minimal reduction in Group B’s mean TAI-T percentile score. Statistical analyses supported this conclusion. A one factor repeated measures ANOVA (1 group X 3 assessment occasions) on Group B’s initial screening, post-wait, and post-treatment TAI-T scores indicated the presence of at least one significant difference among the three pairwise contrasts (p < .0001) (Table 22). A post-hoc, paired t-test contrasting Group B’s initial screening (93.50) and post-wait (87.50) TAI-T scores verified that this intragroup mean difference did not reach significance (p < .0261) (Table 32).

Table 32
Descriptive Data and Paired t-tests on TAI-T, TAI-E, TAI-W, and RTT-T: Mean Differences Between Group B Initial Screening and Post-Wait

<table>
<thead>
<tr>
<th>Measure</th>
<th>Group B Initial Screening Mean</th>
<th>Standard Deviation</th>
<th>Group B Post-Wait Mean</th>
<th>Standard Deviation</th>
<th>Obtained p-value</th>
<th>Critical p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAI-T</td>
<td>93.50</td>
<td>9.72</td>
<td>87.50</td>
<td>11.10</td>
<td>.0261</td>
<td>.0125</td>
</tr>
<tr>
<td>TAI-E</td>
<td>89.19</td>
<td>14.97</td>
<td>84.00</td>
<td>17.95</td>
<td>.1860</td>
<td>.0125</td>
</tr>
<tr>
<td>TAI-W</td>
<td>91.94</td>
<td>14.07</td>
<td>83.94</td>
<td>19.26</td>
<td>.0381</td>
<td>.0125</td>
</tr>
<tr>
<td>RTT-T</td>
<td>* 106.06</td>
<td>7.21</td>
<td>* 102.81</td>
<td>12.84</td>
<td>.0107</td>
<td>.0125</td>
</tr>
</tbody>
</table>

* Raw scores—all other mean differences are percentile scores

The clinical significance of TAI-T reductions between initial screening and post-wait assessments was calculated for each member of Group B using the same methods and procedures as for Group A members on the TAI-T in Question 1 and Group B.
members on the TAI-T in Question 2. Table 12 lists the data used in determining the
RC indices for Group B on the TAI-T. Table 13 displays the gender-specific cutoff
points, RC indices, and the results of the two-stage process for determining clinical
significance. Over the course of the no-treatment waiting control period, four of the 16
Group B subjects' TAI-T reductions reached clinical significance.

Referring to Figure 6, visual inspection suggests that from initial screening to
post-wait assessments there was minimal reduction in Group B's mean TAI-E
percentile score. Statistical analyses supported this conclusion. A one factor repeated
measures ANOVA (1 group X 3 assessment occasions) on Group B's initial screening,
post-wait, and post-treatment TAI-E scores indicated the presence of at least one
significant difference among the three pairwise contrasts (p < .0001) (Table 26). A
post-hoc, paired t-test contrasting Group B's initial screening (89.12) and post-wait
(84.00) TAI-E scores verified that this intragroup mean difference did not reach
significance (p < .1860) (Table 32).

The clinical significance of TAI-E reductions between initial screening and post-
wait assessments was calculated for each member of Group B using the same methods
and procedures as for Group A members on the TAI-E in Question 1 and Group B
members on the TAI-E in Question 2. Table 15 lists the data used in determining the
RC indices for Group B on the TAI-E. Table 16 displays the gender-specific cutoff
points, RC indices, and the results of the two-stage process for determining clinical
significance. Over the course of the no-treatment waiting control period, three of the 16
Group B subjects' TAI-E reductions reached clinical significance.

Referring to Figure 7, visual inspection suggests that from initial screening to
post-wait assessments there was minimal reduction in Group B's mean TAI-W
percentile score. Statistical analyses supported this conclusion. A one factor repeated
measures ANOVA (1 group X 3 assessment occasions) on Group B's initial screening,
post-wait, and post-treatment TAI-W scores indicated the presence of at least one
significant difference among the three pairwise contrasts (p < .0001) (Table 28). A
post-hoc, paired t-test contrasting Group B’s initial screening (91.94) and post-wait
(83.94) TAI-W scores verified that this intragroup mean difference did not reach
significance (p < .0381) (Table 32).

The clinical significance of TAI-W reductions between initial screening and
post-wait assessments was calculated for each member of Group B using the same
methods and procedures as for members of Group A on the TAI-W in Question 1 and
members of Group B on the TAI-W in Question 2. Table 18 lists the data used in
determining the RC indices for Group B on the TAI-W. Table 19 displays the gender-
specific cutoff points, RC indices, and the results of the two-stage process for
determining clinical significance. Over the course of the no-treatment waiting control
period, two of the 16 Group B subjects’ TAI-W reductions reached clinical
significance.

Referring to Figure 8, visual inspection suggests that Group B’s initial
screening and post-wait mean scores were somewhat different. Statistical analyses
supported this conclusion. A one factor repeated measures ANOVA (1 group X 3
assessment occasions) on Group B’s initial screening, post-wait, and post-treatment
RTT-T scores indicated the presence of at least one significant difference among the
three pairwise contrasts (p < .0001) (Table 30). A post-hoc, paired t-test contrasting
Group B’s initial screening (108.26) and post-wait (102.81) RTT-T scores verified that
this intragroup mean difference reached significance (p < .0107) (Table 32).

Secondary Analysis

In this section the effects of EMDR treatment on a number of additional
dependent measures has been reviewed. Specifically, treatment effects on the seven
scales of the SSHA and the STAI-T were examined. These data have been presented as it was believed, a priori, that EMDR would have little impact on most of these measures. It is reasonable to suggest that the less impact that EMDR has on peripheral measures such as the SSHA and STAI-T, the greater its specific treatment utility when that is concomitantly demonstrated.

Within the Secondary Analysis only the effects of EMDR treatment on Group A have been reviewed. Figures have included Group B's EMDR treatment outcomes. However, all descriptive and statistical analyses have been limited to differences between Group A's post-treatment scores and Group B's post-wait scores. Statistical analyses were limited to two-sample t-tests.

The presentation of data for the Secondary Analysis section was similar to that described for the Primary Analysis. However, the descriptive and statistical information presented has been less detailed and more general. Moreover, clinical significance calculations have not been performed.

Figures 9 through 15 display the changes in means for Group A from initial assessment to post-treatment, for Group B from initial assessment to post-wait, and for Group B from post-wait to post-treatment for all seven SSHA scales. Visual inspection suggests that Group A's post-treatment and Group B's post-wait scores differed most in the cases of the SSHA-DA, SSHA-WM, and SSHA-SH scales. Conversely, they appear to have differed least in the cases of the SSHA-TA, SSHA-EA, and SSHA-SA scales. In terms of the magnitude of difference between Groups A and B, the SSHA-SO scale appears to be midway between the first and second groups of SSHA scales. This is reasonable as the SSHA-SO is the overall composite scale that, in effect, aggregates the first group with the second.

Despite clear visual differences between the groups on the SSHA-DA, SSHA-WM, and SSHA-SH scales, none of the differences between Group A's post-treatment
Figure 9. A Comparison of the Effects of EMDR on the SSHA-DA in Group A With the Effects of Waiting on the SSHA-DA in Group B With the Effects of EMDR on the SSHA-DA in Group B.
Figure 10. A Comparison of the Effects of EMDR on the SSHA-WM in Group A With the Effects of Waiting on the SSHA-WM in Group B With the Effects of EMDR on the SSHA-WM in Group B.
Figure 11. A Comparison of the Effects of EMDR on the SSHA-SH in Group A With the Effects of Waiting on the SSHA-SH in Group B With the Effects of EMDR on the SSHA-SH in Group B.
Figure 12. A Comparison of the Effects of EMDR on the SSHA-TA in Group A With the Effects of Waiting on the SSHA-TA in Group B With the Effects of EMDR on the SSHA-TA in Group B.
Figure 13. A Comparison of the Effects of EMDR on the SSHA-EA in Group A With the Effects of Waiting on the SSHA-EA in Group B With the Effects of EMDR on the SSHA-EA in Group B.
Figure 14. A Comparison of the Effects of EMDR on the SSHA-SA in Group A With the Effects of Waiting on the SSHA-SA in Group B With the Effects of EMDR on the SSHA-SA in Group B.
Figure 15. A Comparison of the Effects of EMDR on the SSHA-SO in Group A With the Effects of Waiting on the SSHA-SO in Group B With the Effects of EMDR on the SSHA-SO in Group B.
scores and Group B’s post-wait scores on these SSHA scales reached statistical
significance (Table 33). It should be noted, however, that paired t-tests examining
mean differences between initial screening and post-treatment assessments in Group A
(critical p < .0063) reached statistical significant on the SSHA-WM scale (p < .0008),
SSHA-SH scale (p < .0018), and the SSHA-SO scale (p < .0039). In addition, the
difference on the SSHA-DA scale (p < .0069) approached a significant magnitude.

Figure 16 displays the changes in STAI-T mean scores for Group A from initial
assessment to post-treatment and for Group B from initial assessment to post-wait.
Visual inspection suggests that the groups differed, although minimally. At the
beginning of the study they appear to have been nearly identical. At the post-treatment,
post-wait comparison point, Group A’s STAI-T score was somewhat below that of
Group B. However, a two-sample t-test found that the mean difference between Group
A’s post-treatment score and Group B’s post-wait score did not reach statistical
significance (p < .2862) (Table 33).
Table 33

Descriptive Data and Two-Sample t-tests on SSHA-DA, SSHA-WM, SSHA-SH, SSHA-TA, SSHA-EA, SSHA-SA, SSHA-SO, and STAI-T: Mean Differences Between Group A Post-Treatment and Group B Post-Wait

8 simultaneous t-tests; 15 degrees of freedom

<table>
<thead>
<tr>
<th>Measure</th>
<th>Group A Post-Treatment Mean</th>
<th>Group A Post-Treatment Standard Deviation</th>
<th>Group B Post-Wait Mean</th>
<th>Group B Post-Wait Standard Deviation</th>
<th>Obtained p-value</th>
<th>Critical p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSHA-DA</td>
<td>75.00</td>
<td>26.37</td>
<td>60.75</td>
<td>32.29</td>
<td>.1820</td>
<td>.0063</td>
</tr>
<tr>
<td>SSHA-WM</td>
<td>65.06</td>
<td>25.10</td>
<td>60.00</td>
<td>25.23</td>
<td>.5737</td>
<td>.0063</td>
</tr>
<tr>
<td>SSHA-SH</td>
<td>71.94</td>
<td>24.61</td>
<td>60.63</td>
<td>29.83</td>
<td>.2515</td>
<td>.0063</td>
</tr>
<tr>
<td>SSHA-TA</td>
<td>79.06</td>
<td>21.04</td>
<td>77.06</td>
<td>29.34</td>
<td>.8263</td>
<td>.0063</td>
</tr>
<tr>
<td>SSHA-EA</td>
<td>77.50</td>
<td>21.80</td>
<td>74.31</td>
<td>30.05</td>
<td>.7339</td>
<td>.0063</td>
</tr>
<tr>
<td>SSHA-SA</td>
<td>79.75</td>
<td>20.47</td>
<td>74.88</td>
<td>29.99</td>
<td>.5957</td>
<td>.0063</td>
</tr>
<tr>
<td>SSHA-SO</td>
<td>76.75</td>
<td>20.89</td>
<td>68.81</td>
<td>26.54</td>
<td>.3551</td>
<td>.0063</td>
</tr>
<tr>
<td>STAI-T</td>
<td>50.25</td>
<td>30.59</td>
<td>61.94</td>
<td>30.82</td>
<td>.2862</td>
<td>.0063</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 16. A Comparison of the Effects of EMDR on the STAI-T in Group A With the Effects of Waiting on the STAI-T in Group B With the Effects of EMDR on the STAI-T in Group B.
CHAPTER IV

DISCUSSION

Comparison With Previous Research

Consideration of the Main Hypotheses

The results of the current investigation support the primary hypothesis that EMDR is more effective for the alleviation of excessive, self-reported TA in college students, whose TA is not due to academic factors, than waiting (for a period equivalent to that of treatment) and receiving no treatment. For both statistical and clinical significance analyses, for all four dependent measures, there was a clear differential between Group A's (treatment) and Group B's (no treatment) performances (Table 34). On the repeated measures ANOVAs and post-hoc, two-sample t-tests that compared the effects of EMDR treatment with those for the no-treatment waiting condition, the obtained p-values for all four TA measures fell below .0001. By contrast, the mean changes in Group B's dependent measures during the waiting period were modest, arguably no greater than expected on the basis of regression toward the mean effects.

Following EMDR treatment, on the TAI-T and TAI-E, 15 of the 16 Group A subjects attained clinically significant improvements. On the TAI-W and RTT-T, 11 of the 16 Group A subjects' change scores reached clinical significance. By contrast, following the no-treatment waiting control period, four of the 16 Group B subjects attained clinically significant improvements on the TAI-T, three attained clinically significant improvements on the TAI-E, two attained clinically significant

185
Table 34

Summary of Statistical Significance (SS) and Clinical Significance (CS) Analyses

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>TAI-T</td>
<td>Yes 15 of 16</td>
<td>Yes 12 of 12</td>
<td>No 4 of 16</td>
</tr>
<tr>
<td>TAI-E</td>
<td>Yes 15 of 16</td>
<td>Yes 9 of 12</td>
<td>No 3 of 16</td>
</tr>
<tr>
<td>TAI-W</td>
<td>Yes 11 of 16</td>
<td>Yes 11 of 12</td>
<td>No 2 of 16</td>
</tr>
<tr>
<td>RTT-T</td>
<td>Yes 11 of 16</td>
<td>Yes 9 of 12</td>
<td>Yes 0 of 16</td>
</tr>
</tbody>
</table>

improvements on the TAI-W, and none attained clinically significant improvements on the RTT-T.

The current results support the hypothesis that EMDR was equally effective for alleviating excessive TA in a population of college students (as described above) whether administered immediately following diagnostic assessment or subsequent to a yoked, no-treatment waiting control period of the approximate length required for treatment (although, as predicted a priori, the length of the waiting period varied substantially). Visually, the results observed in group B were very similar to those observed in Group A. For all one factor repeated measures ANOVAs and paired t-tests evaluating Group B’s post-wait to post-treatment mean differences, p values of less than .0001 were obtained. Moreover, Groups A and B were statistically equivalent at initial screening and post-treatment. Clinical significance analyses on EMDR outcomes were also similar across groups. Following the no-treatment waiting control period, on the TAI-T, 4 of 12 Group B subjects attained clinically significant improvements. On

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
the TAI-E, 3 of 12 Group B subjects' change scores reached clinical significance. On the TAI-E, 2 of 12 Group B subjects' change scores reached clinical significance. And on the RTT-T, none of the Group B subjects attained clinically significant improvements.

There were small quantitative differences between Group A's and Group B's mean treatment effects. However, it is not surprising that these differences are virtually eliminated when the effects unrelated to the EMDR protocol are controlled for in Group A (as the experimental design naturally did in Group B), thereby equating the influences that the groups' results represent. That is, when the initial screening to post-wait mean differences observed in Group B are removed from the initial screening to post-treatment mean differences observed in Group A, the resulting mean treatment effects in Group A closely resemble those in Group B.

The current results do not support the hypothesis that TA symptoms remit spontaneously over a no-treatment waiting control period. The various, unspecified variables unrelated to the EMDR protocol that influenced Group B's performance during the waiting period did not do so substantially. Only one of the four TA mean reductions reached significance, and that by a minimal amount (p < .0107). Following EMDR treatment, on the TAI-T, 12 of 12 Group B subjects attained clinically significant improvements. On the TAI-W, 11 of the 12 Group B subjects' change scores reached clinical significance. On the TAI-E and RTT-T, 9 of 12 Group B subjects attained clinically significant improvements.

Consideration of the Outcome Measures

With respect to the TA outcomes generally, the current findings are consistent with those from a large number of previous investigations. In a review of 38 TA treatment studies that compared a treatment of interest with a no-treatment control
condition, Allen, Elias, and Zlotlow (1980) have found the treatment condition to be superior to the no-treatment condition in 22 (58%) of those studies. These same authors have found that systematic desensitization was superior to a no-treatment control condition for diminishing self-reported TA in 20 of 26 studies (77%). Other reviewers have reported similar success rates for systematic desensitization across large numbers of studies (Hembree 1988; Tyron, 1980).

With respect to the TAI, the current findings are consistent with those from numerous studies (Algaze, 1995; Deffenbacher, Michaels, Daley, & Michaels, 1980; Fletcher & Spielberger, 1995; Gonzalez, 1995; Himle, Thyler, Papsdorf, & Caldwell, 1984; Parker, Vagg, & Papsdorf, 1995; Register, May, Beckman, & Gustafson, 1991; Vagg, & Papsdorf, 1995; Van Der Ploeg & Ploeg-stapert, 1986) (Table 35). These projects employed an array of interventions including systematic desensitization, stress inoculation bibliotherapy, group therapy, relaxation training, cognitive-behavioral coping, cognitive coping, anxiety management training, rational emotive therapy, cognitive therapy, biofeedback, study skills counseling, test-taking practice, paradoxical directives, or various combinations of the treatment protocols above.

Although SUD scores were not analyzed or presented in the current study, similar to most EMDR research, they appeared to diminish substantially across all subjects. Nevertheless, SUD reports appear to have little treatment utility or predictive validity for other, more objective measures of performance (Lohr, Tolin, & Kleinknecht, 1996). With regard to the outcome measures generally, the current findings are consistent with one of the three other controlled group studies that have found EMDR superior to a no-treatment control condition for enhancing self-reported standardized measures (Vaughan et al., 1994). It should be noted, however, that in both studies wherein EMDR was found to be no more effective than a no-treatment control condition (Bates, McGlynn, Montgomery, & Mattke, 1996; Jensen, 1994),
Table 35
Comparison of Current Results With Previous TA Research Using the TAI

<table>
<thead>
<tr>
<th>Authors</th>
<th>TA Treatment</th>
<th>Change in TAI-T Raw Score</th>
<th>Mean Reduction</th>
<th>Number of Sessions</th>
<th>Length of Sessions (Minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hampel (1997)</td>
<td>EMDR</td>
<td>64 to 35</td>
<td>29</td>
<td>4.4</td>
<td>90</td>
</tr>
<tr>
<td>Algaze (1995)</td>
<td>CT, SD, &amp; SC</td>
<td>64 to 37</td>
<td>27</td>
<td>8</td>
<td>120</td>
</tr>
<tr>
<td>Algaze (1995)</td>
<td>CT &amp; SD</td>
<td>59 to 33</td>
<td>26</td>
<td>8</td>
<td>120</td>
</tr>
<tr>
<td>Algaze (1995)</td>
<td>CT &amp; SC</td>
<td>59 to 32</td>
<td>27</td>
<td>8</td>
<td>120</td>
</tr>
<tr>
<td>Fletcher &amp; Spielberger (1995)</td>
<td>CT</td>
<td>56 to 35</td>
<td>21</td>
<td>7</td>
<td>90</td>
</tr>
<tr>
<td>Gonzalez (1995)</td>
<td>SD &amp; AC</td>
<td>64 to 46</td>
<td>18</td>
<td>8</td>
<td>75</td>
</tr>
<tr>
<td>Himle, Thyler, Papsdorf, &amp; Caldwell (1984)</td>
<td>RET &amp; Package</td>
<td>61 to 43</td>
<td>18</td>
<td>16</td>
<td>90</td>
</tr>
<tr>
<td>Parker, Vagg, &amp; Papsdorf (1995)</td>
<td>Biofeedback &amp; CC</td>
<td>55 to 41</td>
<td>14</td>
<td>7</td>
<td>90</td>
</tr>
<tr>
<td>Van der Ploeg &amp; Van der Ploeg-Stapert (1986)</td>
<td>Group Therapy</td>
<td>54 to 45</td>
<td>9</td>
<td>8</td>
<td>90</td>
</tr>
<tr>
<td>Deffenbacher, Michaels, Daley, &amp; Michaels (1980)</td>
<td>AM</td>
<td>44 to 39</td>
<td>5</td>
<td>6</td>
<td>50</td>
</tr>
</tbody>
</table>

SD = Systematic Desensitization, AM = Anxiety Management, CT = Cognitive Therapy, CC = Cognitive Control, SC = Study Counseling, AC = Anxiety Control

only one or two sessions of EMDR treatment were administered as contrasted with the average 4.4 sessions in the current study.
Consideration of the Ways This Study Differs From Previous Research

The current study differed from previous research in two essential ways. First, unlike much of the controlled EMDR research, this study adopted a treatment efficacy approach. Most experimental EMDR studies have adopted a more theoretical approach and examined the extent of treatment effects subsequent to a certain limited number of sessions or eye movement sets (Bates, McGlynn, Montgomery, & Mattke, 1996; Bauman & Melnyk, 1994; Boudewyns, Stwertka, Hyer, Albrecht, & Sperr, 1993; Dunn, Schwartz, Hatfield, & Wiegele, 1996; Foley & Spates, 1995, 1996; Gosselin & Matthews, 1995; Hekmat, Groth, & Rogers, 1994; Jensen, 1994; Largo-Marsh, 1996; Merckelbach, Hogervorst, & Kampman, 1994; Sanderson & Carpenter, 1992; Wilson, Silver, Covi, & Foster, 1996). By contrast to these more theoretical approaches, the essential goal in the current project was alleviation of the presenting problem. Approaching the occurrence of excessive TA with the intention of alleviating the condition more closely resembles clinical treatment modalities in natural applied settings. It is believed that this strategy has enhanced the external validity or treatment generality of the current results (Kazdin, 1992).

The second essential way in which the current study differed from previous research was the differential identification of students who reported what appeared to be adequate levels of study habits or attitudes as an independent selection variable. The primary goal of the current study was to demonstrate EMDR's effectiveness for reducing or eliminating excessive levels of TA in a population of college students whose TA was not due solely to obvious, academic factors. Potentially influential academic factors were loosely defined as poor study habits and, or, attitudes and were measured using the three primary scales of the SSHA. Prior to the beginning of the study it was determined arbitrarily that an assessed level of study habits and attitudes
below the 50th percentile would constitute an increased risk of academically-influenced TA, and conversely, an assessed level of study habits and, or, attitudes above the 50th percentile would constitute a decreased risk of academically-influenced TA.

Most all previous TA studies have treated subjects reporting high levels of TA in a similar fashion. Hence, most TA research outcomes may have been confounded by variables related to study habits or attitudes (Algaze, 1995). Some students may have reported excessive TA because of their poor study habits or attitudes. Others may have reported problems taking tests for wholly "emotional" or "psychological" reasons relating to families, teachers, or test related experiences. Using any number of cognitive or behavioral treatments to treat the former group high in TA may have produced less than satisfactory results for many of the same reasons that using cognitive therapy or systematic desensitization are generally ineffective for enhancing poor study skills (Fletcher & Spielberger, 1995). In contrast, as this study has demonstrated, for those students whose TA largely appears to be more a product of various psychological or emotional rather than academic factors, a suitable psychological treatment tends to be highly effective for alleviating TA symptoms.

Many authors and clinicians have suggested that TA may be just an expression of poor test taking skills, poor preparation for tests, or some other aspect involving study habits and, or, attitudes (Kirkland & Hollandsworth, 1979, 1980; Paulman & Kennelly, 1984). There can be no doubt that poor academic preparation gives rise to what is referred to as TA in many cases. However, the results of the current suggest that there are individuals who experience extreme levels of TA (at or above the ninety-fifth percentile) who also report very high, nearly unsurpassable levels of study habits and attitudes (at or above the 90th percentile). Whereas it is no doubt true that every student could improve the quality of their study habits and attitudes to at least some small degree and that poor preparation leads to TA for some students, these results
suggest that there is a subset of college students for whom very good to exemplary
description preparation appears to be ineffective for reducing or preventing excessive levels of TA.
Moreover, the current results suggest that diminishing excessive self-reported TA does
not require enhancing study habits or attitudes directly. Those who consider TA to be
strictly an issue involving poor study habits and attitudes may find these results surprising.

In regards to EMDR treatment of TA, two previous experimental studies have
examined this application (Bauman & Melnyk, 1994; Gosselin & Matthews, 1995),
and the current study differs from both of these substantially. In both of the previous
studies, EMDR was found to be effective for lowering excessive TA levels. However,
in neither case was EMDR found to be more effective than an alternative treatment.
Moreover, unlike the current study, neither the Bauman & Melnyk (1994) nor Gosselin
& Matthews (1995) study employed a no-treatment control condition. Moreover,
neither of these studies adjusted prescores to estimate true prescores. Hence, neither
controlled for potential regression toward the mean effects, which were probable given
their pre-treatment mean TAI percentile ranks (92—Bauman & Melnyk, 1994; 93—
Gosselin & Matthews, 1995).

The current study also reported substantially greater treatment effects than either
the Bauman & Melnyk (1994) or Gosselin & Matthews (1995) studies. The current
study found a mean TAI-T raw score reduction (across groups) of 29 points (64 to 35).
By comparison, Bauman & Melnyk (1994) reported a mean TAI-T raw score reduction
of seven points (58 to 51), and the Gosselin & Matthews (1995) study reported a mean
TAI-T raw score reduction of six points (58 to 52). These differences are not
surprising given the differential number of EMDR treatment sessions administered
between the three studies. The current study administered an average of 4.4, 90-minute
EMDR sessions; the Bauman & Melnyk (1994) study administered a single 45-minute
session; and the Gosselin & Matthews (1995) study administered a single 60-minute session.

Finally, the treatment goals for the current study were different than those for either previous EMDR/TA study. As discussed previously in this section, the current study attempted to alleviate self-reported TA symptoms. In contrast, the Bauman & Melnyk (1994) and Gosselin & Matthews (1995) studies were attempts to demonstrate a treatment effect within a single EMDR session and thereby replicate the efficacy reported by Shapiro (1989a).

In regard to other dependent measures, the current study appears to be one of the few TA studies to employ the RTT assessment device. Many studies have employed more than one standardized measure of TA (Algaze, 1995; Fletcher & Spielberger, 1995; Gonzalez, 1995; Parker, Vagg, & Papsdorf, 1995). However, the RTT has been virtually ignored by researchers. This is surprising in as much as the RTT offers a finer breakdown of TA influences and responses than any other existing self-report instrument. Sarason (1984) has discussed the rationale for the multifactor approach used in the formulation of the RTT as follows:

Worry and emotionality, like anxiety, are concepts. They may or may not be unitary. Wine (1982) has argued that a concept as complex as anxiety may obscure important processes, have too much excess meaning, and, therefore, be misleading. An approach that would reduce these problems is one that deals more explicitly with the scope of phenomena that may pertain to traditional definitions of test anxiety. As indicated by work related to the worry-emotionality distinction, an additional useful step would be to define more reliably the reactions people have when placed in evaluational situations. (p. 931)

The most probable reason that the RTT has not been used more frequently is its lack of normative data. It is hoped that in the future TA researchers will consider using the RTT. If sufficient data were generated using the instrument in this fashion, standardization could become feasible.
One of the great surprises of the current study was the effect that both EMDR treatment and the no-treatment waiting control condition had on self-reported levels of various study habits and attitudes. These results were not expected for a number of reasons. Whereas it seemed reasonable that EMDR treatment might have a therapeutic impact on study habits and attitudes (especially “Delay Avoidance”), the waiting control condition, as experienced by Group B, was not expected to enhance these measures. Rather, as is typical between testing occasions these subjects were expected to exhibit some regression toward mean values. However, with the exception of the SSHA-DA scale, all SSHA mean values increased between initial screening and post-wait assessments.

It is conceivable that anticipating a powerful treatment might have this type of effect. Yet, how anticipating a treatment for TA should affect self-reported measures of study habits and attitudes is not apparent. It is plausible that subjects may have derived some therapeutic value from the initial screening session wherein they were informed that they qualified for the experimental study because they appeared to be a student whose TA was not due to academic causes (those causes were often discussed briefly). Just hearing this explanation and the accompanying causes may have prompted some subjects to engage in certain study related behaviors more frequently. Likewise, just reading the SSHA the first time may have performed a similar function. Hence, subjects’ scores improved on the second administration because they either changed their real-world behavior as a result of the first administration, or they may have simply changed how they reported their behavior. Some may have reported more how they think they should be, as opposed to how they really are. It seems unlikely, although it is possible, that some number of subjects gained substantial study skills or information thereon during the control waiting condition.
Also surprising is the pattern of generally higher scores and mean differences that reached significance among the SSHA scales as a result of EMDR treatment in both Groups A and B. Whereas these results might be understood on the basis of reduced anxiety about academic activity in general, which might lead to less avoidance responding, the nature of the increases does not support this theory. Given this theory, one would expect that the greatest changes would be seen on the SSHA-DA (Delay Avoidance) scale. However, in neither group did the mean value of this scale increase significantly. Rather, it was the SSHA-WM (work methods) scale that increased significantly in Group A and nearly so for group B. Did subjects engage in more adaptive study habits because they were now less aversive? Were the EMDR sessions that focused on imaginal rehearsal of studying helpful in some way? Or, did they alter their verbal report of these events for some other reason that may or may not have reflected real-world differences? Whereas it is possible that subjects learned new study habits and attitudes from the EMDR therapist, the limited number of sessions and limited time spent on these topics in session reduces the plausibility of this supposition.

Consideration of the Effective Components of EMDR

The current results are indicative of EMDR’s effectiveness as it would be employed in most clinical settings. The current study attempted no theoretical dismantling of the EMDR protocol: that is, EMDR was implemented as a seamless package. Moreover, with the goal of demonstrating the effectiveness of the procedure, the sole primary therapist implemented the most complete, advanced application of EMDR of which he was capable. Therefore, the current results suggest that EMDR was effective as a refined, integrated protocol as taught by EMDR’s progenitor (Shapiro, 1991a, 1991b).
To date, virtually no evidence has emerged to suggest that eye movements, in and of themselves, are uniquely essential to EMDR's effectiveness. In fact, the cumulative evidence suggests just the opposite (Lohr, Tolin, & Kleinknecht, 1995). The mechanism(s) responsible for EMDR's functionality (if any) have defied precise systematic identification. As EMDR's progenitor, Francene Shapiro (1995), has stated, "Theories as to why EMDR works are currently only speculations—and will probably remain so for many years" (p. 310).

Nevertheless, Shapiro's (1995) explication of what EMDR achieves, a process she has described in the "Accelerated Information Processing model," suggests that EMDR's functions are known, and that only the specific mechanisms behind the functions remain to be discovered. Shapiro (1995) has stated:

In its simplest form, the model incorporates the physiological notions of network activation, counterconditioning, and assimilation of emotionally corrective, adaptive information. During EMDR treatment, the neuro network containing the target information is activated, the negative effects are mitigated, and the resultant information is weighted positively and then functionally stored in memory. By whatever means, EMDR causes this delayed learning to take place and to do so at an accelerated rate. (p. 310)

Shapiro (1995) has suggested and explored various neurological mechanisms that may be involved with EMDR. These include: (a) EMDR distorts or disrupts the "stereotypic" physiological state associated with the traumatic memory which allows adaptive processing to occur; (b) EMDR distracts the individual from focusing on the traumatic images which prevents these thoughts from being reinforced by previously anticipated anxiety; (c) EMDR creates a suggestibility state similar to hypnosis; (d) EMDR produces synaptic changes which may shift the potential of the traumatic material until it has reached an adaptive resolution; (e) EMDR functions similar to REM sleep; (f) EMDR produces a relaxation response which results in counterconditioning or functions to diminish avoidance of the traumatic memories; and (g) EMDR results in altered cortical functions that diminishes previous anxiety-induced blockages—this may
entail alternating hemispheric activity. This last supposition evolved, in part, from Pavlov’s (1927) theory involving “excitatory-inhibitory” neurological states of balance and the accompanying notion that all successful psychotherapeutic treatments involve restoring correct neurological balance.

Empirical evidence supporting any of these theories is scanty, inferential, and provisional at best. This is not surprising in that, by and large, these theories are genotypic in nature (i.e., they purport unobservable, dysfunctional, dispositional qualities that result in psychopathology; the dysfunctional qualities need to be offset to remediate the psychopathology). Establishing cause and effect among hypothetical physiological or psychophysiological mechanisms (that are often not observable directly) and psychopathology represents a daunting, highly complex endeavor. For what does one look when there is no precedent upon which to base the investigation? Where will one find evidence that counterconditioning, distraction, or changing synaptic potentials are occurring? Moreover, even if one knew where to look, by what means would one look? The technology does not exist to establish whether EMDR functions similarly to REM sleep, or, for that matter, whether REM sleep in and of itself is therapeutic for specific types of memory events. With what equipment would one determine if anxiety was neurologically blocking healthy processing? Where does one hook up what machine to measure healthy processing or altered cortical functioning?

Moreover, the failure to establish empirically that eye movements are an essential component of EMDR is not surprising given what appears to be the circular nature of the primary theory upon which EMDR’s procedures are based and, or, justified. On the one hand the language of Shapiro’s (1995) “Accelerated Information Processing model,” which subsumes many of the elements from the group of hypothetical explanations discussed above, suggests that EMDR functions strategically,
as a cure, by “mitigating” the neurological imbalances that the original distressing event(s) produced. Yet, on the other hand, Shapiro (1995) states, “It was the observations of EMDR treatment effects that led to the Accelerated Information Processing model” (p. 310). In other words, the model or theory was derived from or substantiated by EMDR’s observed effects; the model, in turn, has functioned to explain the very treatment effects from which it was derived. The model has also functioned to hypothesize etiology, guide the treatment process, and refine treatment procedures. This type of reasoning is circular and empirically unsupported. Similar to the psychoanalytic model upon which the DSM-I and DSM-II were based, Shapiro (1995) and others appear to have confused EMDR’s effectiveness as a symptomatic treatment with curing presumed, dysfunctional, underlying mechanisms.

As tempting as it may be to infer substantive, therapeutic mechanisms, given robust results as obtained in the current study, scientific parsimony demands caution and empirical substantiation. One of the primary axioms of science is the adoption of theories that explain the greatest amount of data in the simplest, least complicated manner (Catania, 1992; Johnston & Pennypacker, 1994). This type of parsimony also is one of the guiding strategies for the differential diagnostic process described in the DSM-IV. By contrast, the many theories underpinning EMDR, and the primary explanatory model which subsumes their content to some unspecified degree, are anything but simple or uncomplicated. Their genotypic nature, which, due to the lack of known neurological processes and, or, the lack of the necessary technologies with which to measure them, renders most of EMDR’s most popular theories non-researchable. Moreover, existing research results suggest that eye movements, the very heart of EMDR, while sufficient for therapeutic change, are not essential. Hand taps, finger taps, listening to music, writing, and eye fixation all appear equally effective for achieving positive treatment outcomes (Bauman & Melnyk, 1994; Boudewyns,
Stwertka, Hyer, Albrecht, & Sperr, 1993; Dunn, Schwartz, Hatfield, & Wiegele, 1996; Hekmat, Groth, & Rogers, 1994; Largo-Marsh, 1996; Merckelbach, Hogervorst, & Kampman, 1994; Renfrey & Spates, 1994; Sanderson & Carpenter, 1992). Do all of these seemingly disparate techniques induce similar neurological mechanisms? Specifically, how can eye fixation and eye movement involve similar neurological activity? It seems more likely that the explanation for equivocal effects across disparate treatment rituals is less esoteric and more fundamental than shared neurological impact.

It does not make good scientific sense to suspect or investigate the virtually unknowable, albeit potentially influential, complex effects of a treatment if it hasn’t yet been established that the treatment possesses any clinical facility in excess of what would be expected on the basis of a therapeutic placebo. Most research studies have examined EMDR’s efficacy against either standard treatment approaches, such as exposure, or against modified versions of the EMDR protocol (e.g., substituting eye fixation for eye movement). Whereas such comparative, “dismantling” analyses have merit in that they suggest that eye movements are not singularly therapeutic (Renfrey & Spates, 1994), no study to date comparing EMDR with another treatment has controlled the comparison procedure for the essential, non-specific, therapeutic features sufficient and necessary for a psychosocial treatment to achieve positive outcomes. That is, no comparisons have been conducted with EMDR wherein the alternative approach did not involve elements derived from or germane to a specific, theory based, therapeutic mechanism or protocol. Conversely, no comparisons have been conducted using an alternative treatment comprised entirely of elements non-specific to any theory based therapeutic mechanism or treatment protocol. Most alternative treatment conditions have consisted of treatments similar to EMDR such as exposure or an EMDR variation wherein some “active,” specifically therapeutic component may still have been present.
Research is needed, therefore, wherein EMDR is compared with procedures believed to contain no components germane to known therapeutic mechanisms, theories, or protocols; procedures whose therapeutic efficacy is due solely to non-specific elements (J. D. Frank & J. B. Frank, 1991). However, if this research were to suggest that EMDR were significantly more effective than therapeutically non-specific placebo treatments, non-specific elements could not be ruled out.

EMDR incorporates at least two features that may function to enhance one or more of the four basic, non-specific, therapeutic qualities (J. D. Frank and J. B. Frank, 1991). First, at the outset of EMDR treatment, the clinician establishes expectations for swift, powerful treatment effects via information deduced from the “Accelerated Information Processing model.” Even though the mechanisms described by this model may be illusory and unverifiable, paradoxically, the model itself may be beneficial to EMDR’s effectiveness as a symptomatic treatment. The certitude with which the Accelerated Information Processing model prescribes EMDR’s swift and powerful effects is exactly the type of language that should enhance the therapist’s credibility as an expert care provider, augment trust in or acceptance of the EMDR regimen, and establish an expectation for salient treatment effects when the eye movement procedure takes place (Kirsch, 1990).

The second feature of EMDR that may enhance the basic elements necessary for a psychosocial therapy to function as a therapeutic placebo is the stimulation produced by eye movements. Regardless of any psychological effects they may entail or facilitate, large magnitude, saccadic eye movements, as prescribed by the EMDR protocol, produce distinct physical sensations (a brief trial will prove the point). Eye movements may be instrumental in EMDR’s effectiveness because they produce unique proprioceptive stimulation which the individual being treated has been prepared to expect and subsequently interpret as a sign of the treatment’s swift power. The
physical sensations produced by EMDR simply may confirm to the patient the procedure's predicted effectiveness (Kirsch, 1990). Clients are told that EMDR's effects will be "powerful"; then, when they can "feel" it working, they may "believe" that it works and report this perceived consequence. As a result of these feelings and how they are interpreted, EMDR may function especially well as an "active" placebo. Therefore, research is needed to compare EMDR with "active" therapeutic placebos other than variations of EMDR or exposure as has been done.

One operant explanation for EMDR's effectiveness is notable. Numerous authors have suggested that EMDR may function to alter in-session verbal behavior within an escape/avoidance paradigm (Acierno et al., 1994; Lohr et al., 1992; Herbert & Mueser, 1992). According to this analysis repetitive eye movements are considered to be somewhat aversive. Therefore, their cessation is mildly reinforcing and any behavior that produces cessation is reinforced. Behaviors that stop or limit the sets of eye movements during EMDR treatment are generally those implicating improved subject functioning, that is, lower SUDS levels and more positive thoughts and feelings. It is conceivable that a subject would express feeling or thinking "better" as a function of escaping or avoiding additional eye movements, if the eye movements were aversive.

However, the current results do not support this hypothesis. There are two sets of pertinent evidence. First, were such an escape/avoidance pattern occurring, then one would expect to see the SUDS levels, which had fallen within a session, rise again by the beginning of the next session. This occurred rarely if at all. When a SUDS level did increase in regard to a specific thought or situation between sessions it was almost always because of some substantive environmental event (e.g., parents having declared that the student must achieve a certain GPA during that semester or lose financial support). Second, if such an escape/avoidance paradigm was in effect, then one would
anticipate that the post-treatment scores would rise from the level anticipated by lowered SUDS scores or maintain at or near initial screening levels once the treatment period was terminated and there were no escape/avoidance contingencies for reporting reduced symptom intensity. However, post-treatment assessments were generally reduced dramatically from initial screening levels. Moreover, nearly all subjects completed the final, post-treatment evaluation some number of days after they officially completed treatment. If lowered symptoms had been feigned during the sessions, there was adequate time between the last session and the post-treatment assessment during which symptom levels could have migrated back to more honest levels.

Limitations of the Current Study

One of the major limitations of the current study is the lack of a follow-up assessment. The course of continued therapeutic changes and, or, the durability of a therapeutic intervention can vary dramatically subsequent to the conclusion of treatment (Kazdin, 1992, p. 276). Hence, follow-up assessments often alter conclusions regarding the effects of treatment. In the current study most of the post-treatment evaluations were conducted between one and two weeks after the termination of treatment. Ideally, an additional assessment should have been conducted at least one month subsequent to termination.

A second major limitation of this study is the failure to perform integrity checks on the independent variable, the EMDR treatment protocol. Two integrity procedures were considered: (1) unscheduled observations using a checklist to record the essential steps in the EMDR protocol as they occurred in actual treatment sessions, and (2) creating a videotaped demonstration wherein correct and incorrect EMDR vignettes would be demonstrated with therapy confederates. The videotape would be evaluated and scored by an independent expert such as EMDR’s progenitor. Accurately depicting
correct and incorrect applications of EMDR would serve as a skills test for the primary therapist. A strong showing on the videotaped test would presumably enhance the believability that the primary therapist had applied the EMDR protocol faithfully and consistently.

The validity of the checklist integrity method was questioned on the grounds that the essential features of the EMDR protocol as applied in this study were unclear and therefore unmeasurable. Certainly the basic steps in the protocol could have been observed independently and verified. However, given the highly idiosyncratic presentations of TA and the corresponding flexibility of EMDR application across a number of parameters (necessary for and appropriate to a clinical efficacy study such as this), ultimately, it was not clear how to quantify or objectify the current EMDR protocol in a meaningful way. How would one rate the thoroughness or accuracy of the background assessment, the utility of the EMDR “target” memory, the exact speed, number, and direction of the saccades, the appropriateness of various imaginal strategies, or the timing, need, and implementation of “cognitive interweave” techniques? Certainly, developing reliable measures for these and other illusive, ill-defined treatment components remains one of the essential challenges for psychotherapy outcome research.

The validity of the vignettes approach to assessing treatment integrity was questioned on the grounds that demonstrating simple correct and incorrect usage of EMDR with a therapy confederate is far removed from the dynamic interaction characteristic of most psychotherapy. The conditions of and contingencies for making a demonstration tape and performing EMDR treatment are drastically different. Demonstrating competent EMDR performance with confederates in front of a camera might have predictive validity for competent implementation of the protocol with real
clients (and real problems). However, this validity would be based on logic or reasoning, not empirical findings.

Another limitation of the current study is inherent with self-report data such as the TAI and SSHA. Self-report is a form of verbal behavior that is readily influenced in part by the events which they describe (the pertinent antecedents) and in part by variables unrelated to those events such as social consequences or social histories. For example, subjects may have reported experiencing less TA because, in fact, they did; because they wanted to please the therapist; because they wanted to escape the therapy; because they adopted various conventions from the therapist that made them sound improved; or for any number of other reasons unrelated to the TA condition itself. Hence, verbal behavior, especially verbal behavior that describes the self, is notably inaccurate or "plastic." (cf., Skinner, 1957).

The verbal nature of the primary dependent measures limits the generality and external validity of the current results (Kazdin, 1992). These outcomes say virtually nothing about whether EMDR actually alleviated excessive worrying and emotionality while taking tests as the self-reported results of the TAI suggested. The current results only describe indirectly how EMDR may have changed the subjects' behavior in their day-to-day lives. The current results do not indicate whether EMDR actually changed time spent studying, the promptness of studying vis-a-vis assignments, test-taking ability, and, or, objective performance on examinations. Ultimately, these are the crucial issues which self-report measures fail to evaluate accurately and which future research needs to address. More generally, establishing and implementing more reliable and objective outcome criteria for psychological research remains a major unmet challenge (Azar, 1997).

Regarding the evaluation of the EMDR protocol, one limitation of the current study was the lack of a comparative or placebo treatment condition. Unfortunately, the
inclusion of both no-treatment control and placebo control conditions in addition to the EMDR treatment condition was beyond the scope and resources of the current investigation. As discussed in the Consideration of the Effective Components of EMDR section of the current text, systematic examinations of EMDR with active placebos (that include no active elements specific to a therapeutic mechanism or component) would help to establish the probable existence of therapeutic effects above and beyond those anticipated on the basis of various non-specific mechanisms.

Another limitation of the current study involves the variance in delay to treatment in Group A subjects. Due to various influences (e.g., competing work, academic, and family obligations), the number of days between the subjects’ initial screenings and first treatment sessions varied substantially. Whereas some subjects in Group A began treatment within a few days of the initial screening others did not begin for a number of weeks. Because Group A subjects were not systematically reassessed following their respective (non-programmed) wait periods prior to treatment administration (as were Group B subjects), it is plausible that some of the improvement in TA that some Group A subjects achieved may have been due to time passage as opposed to EMDR treatment. However, the significance of treatment effects in Group A were contrasted against changes in the waiting control condition, which should have controlled for confounding effects unrelated to the EMDR protocol. Furthermore, the effects of the no-treatment waiting control condition in Group B and the subsequent effects of EMDR in Group B, that very nearly matched those measured in Group A, suggest that the palliative effects of time prior to treatment on self-reported TA in Group A were minimal.

As expected, the clinical trials nature and efficacy goals of the current study produced variance in the EMDR treatment period. As they presented themselves, each Group A subject was treated in a flexible fashion according to their idiosyncratic needs,
goals, and schedules. Treatment was not administered within a fixed time or defined by a limited number of sessions. Rather, the course and length of treatment were driven by the alleviation of excessive, self-reported TA. Hence, treatment required varying numbers of sessions and varying lengths of time in which to administer those sessions. Some Group A subjects required three or four sessions that were completed within two or three weeks. Others needed as many as 10 sessions that required as long as 122 days to complete. Naturally occurring events, such as the Christmas holidays, which occurred during the middle of the current study, extended treatment for some subjects by a number of weeks. As a consequence of the temporal inconsistencies among treatment periods for Group A subjects, conclusions regarding the effects of EMDR over a fixed time are limited. However, the interaction of treatment effects and length of treatment can be examined at the individual level. Table 36 displays reductions in TAI-T scores for subjects in Groups A and B and each subject’s corresponding time in treatment. A cursory examination of these results suggests that there was no correlation between the magnitude of improvement on the TAI-T and the time in treatment. In fact, there were many striking comparisons in both groups. For example, one Group A subject (male) achieved an 8-point reduction in TAI-T percentile score over a period of 102 days. Another Group A subject (male) achieved a 91-point reduction in TAI-T percentile score over a period of eight days.

Also, as expected, the yoked control between each Group A subject’s post-treatment assessment and the corresponding Group B subject’s post-wait assessment resulted in substantial variance in the no-treatment waiting period among Group B subjects. Hence, the results of the current study describe the mean effects of varying waiting periods on self-reported TA; the information that these results provide regarding the effects of a fixed length no-treatment waiting condition is limited.
Nonetheless, the interaction of waiting condition effects and the length of the waiting condition can be examined at the individual level. Table 37 displays reductions in TAI-T scores for Group B subjects and each subject's corresponding time in the no-treatment waiting control condition. A cursory examination of these relationships suggests that there is no correlation between the length of time waiting in a no-treatment condition and the palliative effects experienced. Most Group B subjects reported little or no change in their TAI-T levels. However, four subjects reported substantial reductions in the TAI-T from initial screening to post-wait assessments (each of these changes reached clinical significance). The controlling variables for these spontaneous reductions were uncertain. Generally, identifying the controlling variables for spontaneous remissions of self-reported psychological symptoms remains a challenge for psychological research.

Subsequent to EMDR treatment, for most subjects in both groups, reductions on all dependent measures of TA attained clinical significance. As discussed previously in this text, clinical significance might be a useful tool for treatment planning and assessment. However, in the treatment of TA (and nearly all other clinical conditions) clinical significance remains a derivative statistic which has yet to be validated against measures of treatment permanence, relapse rates, quality of life, and other variables pertinent to "real-world" improvements. In regard to the limitations of clinical significance measures, the current individual treatment outcomes, most of which attained clinical significance, should be interpreted with caution.

An issue related to non-specific variables in EMDR research that limits the current study involves the subjects' belief in, acceptance of, and, or, expectations for the treatment protocol and how these self-reported parameters may be influenced by experimental procedures and subsequently influence outcomes at either the individual or group level. These are variables that the current study did not assess, analyze, or
Table 36
Reduction in TAI-T Score and Corresponding Length of Treatment for Subjects in Groups A and B

<table>
<thead>
<tr>
<th>Group A Subjects Gender</th>
<th>Group A Reduction in TAI-T</th>
<th>Group A Length of Treatment</th>
<th>Group B Subjects Gender</th>
<th>Group B Reduction in TAI-T</th>
<th>Group B Length of Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>32</td>
<td>30</td>
<td>Female</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Female</td>
<td>13</td>
<td>12</td>
<td>Male</td>
<td>66</td>
<td>76</td>
</tr>
<tr>
<td>Male</td>
<td>56</td>
<td>10</td>
<td>Male</td>
<td>80</td>
<td>79</td>
</tr>
<tr>
<td>Female</td>
<td>38</td>
<td>33</td>
<td>Female</td>
<td>49</td>
<td>77</td>
</tr>
<tr>
<td>Male</td>
<td>86</td>
<td>114</td>
<td>Female</td>
<td>80</td>
<td>58</td>
</tr>
<tr>
<td>Female</td>
<td>56</td>
<td>47</td>
<td>Female</td>
<td>73</td>
<td>66</td>
</tr>
<tr>
<td>Female</td>
<td>27</td>
<td>122</td>
<td>Female</td>
<td>4</td>
<td>30</td>
</tr>
<tr>
<td>Male</td>
<td>8</td>
<td>102</td>
<td>Male</td>
<td>46</td>
<td>13</td>
</tr>
<tr>
<td>Female</td>
<td>33</td>
<td>44</td>
<td>Female</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Male</td>
<td>91</td>
<td>8</td>
<td>Male</td>
<td>12</td>
<td>47</td>
</tr>
<tr>
<td>Female</td>
<td>71</td>
<td>21</td>
<td>Male</td>
<td>36</td>
<td>22</td>
</tr>
<tr>
<td>Female</td>
<td>71</td>
<td>91</td>
<td>Female</td>
<td>46</td>
<td>28</td>
</tr>
<tr>
<td>Female</td>
<td>64</td>
<td>63</td>
<td>Female</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Male</td>
<td>37</td>
<td>28</td>
<td>Female</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Female</td>
<td>92</td>
<td>10</td>
<td>Female</td>
<td>9</td>
<td>35</td>
</tr>
<tr>
<td>Male</td>
<td>89</td>
<td>22</td>
<td>Female</td>
<td>56</td>
<td>14</td>
</tr>
</tbody>
</table>

--- Did not Complete Tx

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 37

Reduction in TAI-T Score and Corresponding Length of the No-Treatment Control Condition for Subjects in Group B

<table>
<thead>
<tr>
<th>Group B Subjects Gender</th>
<th>Reduction in TAI-T (Percentile Rank)</th>
<th>No-Treatment Waiting Control Condition (Days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>25</td>
<td>48</td>
</tr>
<tr>
<td>Male</td>
<td>2</td>
<td>37</td>
</tr>
<tr>
<td>Male</td>
<td>1</td>
<td>35</td>
</tr>
<tr>
<td>Female</td>
<td>2</td>
<td>110</td>
</tr>
<tr>
<td>Female</td>
<td>4</td>
<td>110</td>
</tr>
<tr>
<td>Female</td>
<td>0</td>
<td>122</td>
</tr>
<tr>
<td>Female</td>
<td>24</td>
<td>111</td>
</tr>
<tr>
<td>Male</td>
<td>23</td>
<td>119</td>
</tr>
<tr>
<td>Female</td>
<td>-1</td>
<td>123</td>
</tr>
<tr>
<td>Male</td>
<td>-1</td>
<td>34</td>
</tr>
<tr>
<td>Male</td>
<td>7</td>
<td>33</td>
</tr>
<tr>
<td>Female</td>
<td>1</td>
<td>102</td>
</tr>
<tr>
<td>Female</td>
<td>-1</td>
<td>87</td>
</tr>
<tr>
<td>Female</td>
<td>-4</td>
<td>93</td>
</tr>
<tr>
<td>Female</td>
<td>13</td>
<td>20</td>
</tr>
<tr>
<td>Female</td>
<td>1</td>
<td>44</td>
</tr>
</tbody>
</table>

Control for and which could have contributed to its successful results. It is conceivable that the inherent nature of efficacy and theoretical studies may affect belief, acceptance, expectancy, and other non-specific variables differentially. These two types of
Experimental arrangements would deliver very different messages involving the extent of help the study intends to provide. For instance, a theoretical study that administers one or two EMDR sessions cannot possibly offer its subjects the same levels of reassurance and hope as an efficacy study wherein subjects may receive as many as 10 sessions.

Future studies that offer a greater number of sessions could attempt to determine if believability, acceptance, and expectancy levels would be impacted by the pre-treatment knowledge of the total number of available sessions. Would subjects who understood a priori that as many as 10 sessions were possible express stronger belief in, acceptance of, or expectancies for EMDR therapy than subjects who understood a prior that a total of two sessions would be administered, although in reality 10 were available? How would the subjects in the latter group respond when, upon completing their second and presumably final session, they were informed that some number of additional sessions remained optional? Subsequent to these manipulations, the respective impact that differences on belief, acceptance, expectancies, and, or, other non-specific variables exert on outcome measures could be assessed.

Whereas this study demonstrated the efficacy of EMDR to treat self-reported TA in a college population, providing unlimited treatment up to 10 sessions may have limited conclusions about the lower limits of EMDR's efficacy for this application. The current study has demonstrated that an average of 4.4 sessions of EMDR is effective for alleviating the self-reported symptoms of excessive TA. However, it cannot be determined that 4.4 sessions were sufficient and necessary to attain these results. It is possible that, for various personal reasons, some subjects in the current study may have continued with treatment beyond the point where their TA was alleviated. During the study, a number of subjects reported having stressful, distressing relationships with friends and family. Although resolving academic issues often involved resolving
family issues, occasionally, the connection between the reported, distressed relationship and TA seemed tenuous. When this occurred subjects were probed to ascertain the pertinence of the presented material vis-a-vis TA. Sometimes subjects reported no connection between the stressful relationship and testing. When this occurred, the primary therapist redirected the session's focus to an academically related issue and, ostensibly, very little time was lost. Nevertheless, the existence of these occasions suggests that some subjects may have gained something of value from the EMDR treatment protocol (e.g., support and empathy) other than the relief of their TA. Hence, it is conceivable that fewer sessions than were administered may actually have been required to resolve successfully the self-reported symptoms of TA. Future EMDR efficacy research may wish to consider how variables of this nature distort observed treatment length and strategize ways to circumvent it. One possible solution might be to track or reassess subjects at regular intervals during treatment to determine their progress and need for additional therapy. Accordingly, single-subject research methodology might be suitable to address this concern.

Finally, the results of the current study were also limited to a very small percentage of the total college population who report high levels of TA. Although the current research did not record exact data, it is estimated that in excess of 200 individuals were screened in order to identify the 32 who ultimately participated. Of the dozens of students who did not qualify for this study due to poor study habits and, or, attitudes, all but seven or eight reported TA levels above the 75th percentile. The fact that the current study systematically excluded students low in study habits and, or, attitudes in no way reflects EMDR's potential efficacy with this population. Algaze (1995) found that students assessed as either good or poor in study habits responded equally well across four different TA treatment conditions, two of which involved study skills counseling and two of which did not. If, as Algaze (1995) reported,
cognitive therapy is effective for alleviating the self-reported symptoms of TA in students with poor study habits, then it seems reasonable to posit that EMDR would be as well.

Conclusion

For purposes of alleviating self-reported TA in college students, whose TA does not appear to be determined by obvious academic variables, the current results suggest that EMDR is more effective and requires less time than nearly any other TA treatment strategy. These results do neither verify nor disconfirm the presence of one or more active, EMDR-specific, therapeutic effects. Rather, when considered together with the (now) extensive published EMDR literature (as reviewed in the current text) and what is known regarding the non-specific effects of treatment (J. D. Frank and J. B. Frank, 1991; Kirsch, 1990), the broad-based success of the current findings suggests a more parsimonious explanation for EMDR's efficacy.

The demonstrated ability of EMDR to impact therapeutically a variety of self-reported problems and dependent measures suggests that the nature of its efficacy is non-specific to a theory-based, therapeutic mechanism or protocol. In the current study EMDR worked for nearly everyone and improved (at least to some extent) nearly everything. Sweeping results such as these are not suggestive of a complex, neurological palliative as EMDR has been described by various theories. Rather, it appears reasonable and conservative to conclude that EMDR is (no less than) a unique combination of non-specific elements that yield a highly effective active placebo treatment for a variety of anxiety conditions including what has been described herein as TA. Moreover, before elaborate, non-objectifiable, and seemingly improbable theories are entertained, it is merely basic science to rule out those etiological factors.
which appear most probable: the variables non-specific to any theory-based, therapeutic mechanism or protocol.

However, the non-strategic nature of EMDR should not be interpreted as disparaging its clinical value. Efficient and effective symptomatic treatments are valuable tools within the clinician’s armamentarium. Nevertheless, it is vital for clinicians to understand that, regardless of its active constituent components, whether EMDR is or is not a placebo, EMDR most definitely is (no more than) a symptomatic treatment. Given that the mechanisms underlying the disorders for which EMDR is applied remain unknown, EMDR cannot, by definition, be a strategic or curative treatment. EMDR developed neither from an understanding of the pathogenesis of PTSD (or any other disorder for which it has been used including TA), nor from known, functional, empirically-determined principles of behavior. EMDR manages or alleviates symptoms, it does not cure their causes.

Falsely believing that a symptomatic treatment is curative could lead to the belief that the cause of the disorder being treated is known when it is not. Such beliefs could mislead, confuse, and disillusion clients and therapists. Moreover, clients who believe they are “cured” by EMDR could ultimately reject psychosocial interventions generally when and if they should experience a symptomatic relapse. Falsely believing that a mental disorder’s cure exists could suppress efforts at discovering the actual mechanisms for the disorder as well as curtail the development of more effective symptomatic treatments. Unfortunately, neither Shapiro (1995), nor any other published EMDR researcher, has advanced the understanding unequivocally that EMDR is a symptomatic treatment strategy.
Appendix A

EMDR Treatment Protocol Used in the Current Study
The EMDR Protocol Used in the Current Study

Although the EMDR protocol is ostensibly simple and straightforward, it should not be viewed as a rigid routine, that, upon administration, will result automatically in clinical improvements. "EMDR is not a cookie cutter. Each client is approached interactively...No two clients are exactly alike—and no two sessions are exactly alike" (Shapiro, 1991). Individual tailoring of the protocol aside, a typical EMDR session (based on level 1 EMDR training) would include the following steps. The first step would include preliminaries such as establishing a therapeutic relationship, explaining treatment rationale and method, and establishing client expectancies for quick, positive outcomes. The second step would include an assessment of the client's problems, targets for treatment and any contra-indications for the use of EMDR (e.g., eye problems, client's current psychological stability, and severe dissociative disorders). The third step would include having clients rate their level of subjective autonomic arousal (in the here and now) to the traumatic material using a Subjective Units of Distress (SUD, 0-10) scale (SUDS; Wolpe & Lazarus, 1966). The therapist would then ask the client to subjectively pinpoint the location in their bodies where they feel the distress and to name the feeling that they are experiencing. It should be noted that regardless of the material presented, the therapist asks for subjective verbal reports of any physiological manifestations of clients' distress and uses changes in such verbal reports as the yardstick of therapeutic progress. In this third step the therapist also asks clients for reports of negative thoughts or beliefs they may hold about themselves and their role relative to the reported historical distressing event. This subjective verbal report is referred to as the negative cognition. The therapist then asks the client what they would rather believe about themselves relative to the event. This subjective verbal report is referred to as the
positive cognition. The subjective degree to which the client believes the positive cognition ("in their gut") is estimated using an eight point (0-7) Validity of Cognition scale (VoC, Shapiro, 1991a).

The fourth step in the EMDR protocol involves the therapist directing clients to subjectively focus on ("hold in mind") some salient aspect of the reported distressing memory (e.g., a thought, bodily sensation, or image) while following a moving object (typically the therapist's fingers or a pointer) with their eyes, to produce large-magnitude saccadic eye movements at the rate of about one complete back and forth cycle per second. Following a moderate number (20 to 30) of complete saccades (the number might be much higher in the case of client abreactions or lengthier imaginal sequences), the therapist stops the eye movement cycles, instructs the client to subjectively "blank it out," then asks the client to report on what they are currently subjectively experiencing (e.g., "What comes up for you now?"). SUD levels are assessed as the treatment progresses, although this is not necessary following each set of eye movement cycles.

In the fifth step the therapist would include strategies for reinitiating therapeutic progress when a client's subjective reports (SUDS) stabilize at a level higher than zero for two or more sets. In these cases the therapist asks the client to report which thought, image, or feeling is preventing further reduction in SUDS feelings. The response directs the therapist in choosing the target for further desensitization and the eye movement process continues.

When the client reports a consistently low SUDS level (usually near "0"), the therapist instructs the client to re-imagine the original, distressing material. If the client reports a higher SUDS level, further reprocessing is done until therapist directives to return to the original covert material evokes no verbal reports of subjectively felt, physical distress (i.e., the SUDS level does not change from its low value). Once the
target image, thought, or kinesthetic feeling is reported to be completely desensitized, the client is asked to subjectively rate (using the 0-7 VoC scale) how thoroughly they believe their own positively reframed cognition. A high value, near 7, indicates that they believe it absolutely. A lower value requires that the therapist instruct the client to hold the positively reframed statement in mind while eye movements cycles are resumed.

In the final step, once the subjective VoC value for the positive cognition reaches a value near 7, the therapist has the client mentally juxtapose the positive cognition with thoughts of the original distressing material while another set of eye movement cycles is conducted. Treatment cessation (for a particular problem or target) occurs when the client: (a) reports no subjective tension remaining in the body (the SUD=0), (b) reports that the positive cognition is thoroughly believed (the VoC=7), and (c) reports a sense of closure in regard to the disturbing thoughts.

Within the conceptualization of EMDR used in the current study, eye movements were done to the extent that they were needed to help reprocess the dissociated material via the pairing of memories and current alertness. The number of saccades used for each set of eye movements depends on the nature of individual client reports and the nature of the imaginal task.

More specifically, EMDR treatment for TA differed from EMDR treatment for PTSD in the material targeted for desensitization and the imaginal strategies employed. In the case of PTSD, the treatment targets include images and thoughts germane to the one-time traumatic event. For treating TA, the EMDR therapist would initially target and sequentially desensitize at least three memories in the following order: (1) the client's first recollection of the problem; (2) the most disturbing recollection of the event or time relating to the problem; and (3) the most recent recollection of the problem (Shapiro, 1991b). The rationale for targeting memories in sequential chronological
order is that early events are thought to establish a pattern into which later events become associated. In order to attain the best results treating test anxious individuals with EMDR, the therapist might need to target and desensitize any number of test-relevant scenes or thoughts, such as studying the night before a big exam, imagining having only five minutes left in the exam period, or thinking, "I'm going to fail this exam, I just know it, and everybody is going to think I'm stupid and a failure."

To desensitize and reprocess an irrational or dysfunctional belief or thought (such as "I'm a failure"), the same strategy (as is used for anxiety-provoking material that has no clear or dramatic onset) of identifying three chronologically-sequenced memory targets is employed (Shapiro, 1991b). Targeting a belief as opposed to a memory of an event is considered to be another way of gaining access to the important memory information.

Once the EMDR desensitization process is completed with each of these three past memories (with either specific memory images or beliefs), a future imaginal scenario is identified and desensitized. Specifically, the client would be instructed to imagine themselves engaged in the (previously) anxiety-provoking activity, functioning calmly and effectively while reminding themselves (sometimes out loud) of their positive cognitions. This imaginal rehearsal component affords the client the opportunity to engage the previously dreaded circumstances in vitro and desensitize/reprocess any troublesome thoughts or images experienced. However, the value of this component to the EMDR ritual seems questionable. Behavior therapy research has indicated that in vitro or imaginal desensitization is ineffective for serious phobias, etc.

An imaginal strategy that might be effective for assessing the thoroughness of desensitization is the "process approach" to anxiety treatment, wherein the EMDR memory involves not just a single image or thought, but an entire sequence of events...
(Shapiro, 1991b). This would be employed following the successful desensitization of all other reported memories, and would function to identify any additional anxiety provoking circumstances. Clients would be instructed to begin imagining themselves at the moment in time just before they typically experience anxiety symptoms. Then, during eye movement cycles, they would imagine the entire sequence of events that follows "like a movie." For a test anxious person the imagined sequence might begin with walking into the classroom to take a test, and conclude when the person hands in the completed test form. Any and all anxiety provoking images or thoughts that the client encounters would be desensitized before proceeding with the sequence. The client would progress imaginally through the sequence of events until they could run the whole "test-taking movie" from the beginning without experiencing any increase in somatic arousal (Shapiro, 1991b).
Appendix B

Clinical Significance Formulas
Dysfunctional Functional

\[ c = \frac{S_0M_1 + S_1M_0}{S_0 + S_1} = \frac{.31(1.39) + .601(.31)}{.31 + .601} = .677 \]

\[ RC = (X_{post} - X_{pre})/S_{Diff} \]

\[ S_E = S_1\sqrt{1 - r_{xx'}} = .601\sqrt{1 - .838} = .242 \]

\[ S_{Diff} = \sqrt{2(S_E)^2} = \sqrt{2(.24)^2} = .35 \]
Appendix C

Protocol Clearance From Human Subjects Institutional Review Board
Date: December 2, 1993

To: John Hampel

From: M. Michele Burnette, Chair

Re: HSIRB Project Number 93-11-15

This letter will serve as confirmation that your research project entitled "A survey of test anxiety, trait anxiety, and study habits in two kinds of college courses" has been approved under the exempt category of review by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the application.

You must seek reapproval for any changes in this design. You must also seek reapproval if the project extends beyond the termination date.

The Board wishes you success in the pursuit of your research goals.

Approval Termination: December 2, 1994

xc: Alessi, Psych.
Date: February 16, 1994
To: John Hampel
From: M. Michele Burnette, Chair
Re: HSIRB Project Number 93-11-16

This letter will serve as confirmation that the requested revisions to your research protocol, "The effects of Eye Movement Desensitization and Reprocessing (EMD/R) on self-reported test anxiety and state anxiety in two kinds of college courses" have been approved by the Human Subjects Institutional Review Board.

Please submit verification from Kalamazoo Community College of permission to advertise on their campus.

xc: Alessi, Psych
Appendix D

Consent Forms for Participation in the Current Study
I have been invited to participate in a research project entitled "The effects of eye movement desensitization and reprocessing (EMDR) on self-reported test anxiety in college students." I understand that this research is experimental in nature. I understand that this research is intended to study how well EMDR lowers test anxiety. I further understand that this project is John Hampel’s dissertation project.

My consent to participate in this project indicates that I will be asked to complete a series of questionnaires. During the first meeting with John Hampel I will be asked to complete four questionnaires: the Reactions to Tests, the Test Attitude Inventory, the Survey of Study Habits and Attitudes, and the Self-evaluation Questionnaire. I will also provide general information about myself such as my telephone number, level of education, and career goals. My consent to participate in this project also indicates that I will be asked by John Hampel to complete the same series of questionnaires at the end of the study that I am asked to complete at the beginning of this study.

Should I have a problem associated with participation in this study, I understand that John Hampel is prepared to talk with me at a convenient time. As in all research, there may be unforeseen risks to the participant. If an accidental injury occurs, appropriate emergency measures will be taken; however, no compensation or treatment will be made available to the subject except as otherwise stated in this consent form.

One way in which I may benefit from this activity is learning more about test anxiety and whether it is a problem for me. Another way in which I may benefit from this study is learning about my study habits and attitudes. I understand that, if I request it, John Hampel will provide a summary of my performance on all of the survey forms that I complete. If I am chosen to participate in the treatment part of the experimental study I may benefit from reductions in the amount of anxiety that I experience when taking tests. I also understand that others who experience test anxiety may benefit from the knowledge that is gained from this research.

I understand that all the information collected from me is confidential. That means that my name will not appear on any papers on which this information is recorded. The forms will be coded, and John Hampel will keep a separate master list with the names of participants and the corresponding code numbers. Once data are collected and analyzed, the master list will be destroyed. All other forms will be retained for three years in a locked file in the principal investigator’s laboratory and then destroyed (shredded).

I understand that I may refuse to participate or quit at any time during the study without prejudice or penalty. If I have any questions or concerns about this study, I may contact either John Hampel at 343-9466 or 387-4456, or Dr. Galen Alessi at 387-4470. I may also contact the Chair of Human Subjects Institutional Review Board at 387-8293 or the Vice President for Research with any concerns that I have. My signature
below indicates that I understand the purpose and requirements of the study and that I agree to participate.

_________________________  ____________________________
Print name                                    Signature

_________________________  ____________________________
Phone number                                  Date

QUESTIONS:

1) In what year of study are you at WMU? (Circle one)  1  2  3  4  GRAD

2) Briefly describe your professional career goals.

_________________________________________________________________
_________________________________________________________________
_________________________________________________________________
_________________________________________________________________
_________________________________________________________________
_________________________________________________________________
_________________________________________________________________
_________________________________________________________________

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
I have been invited to participate in a research project entitled "The effects of eye movement desensitization and reprocessing (EMDR) on self-reported test anxiety in college students." I understand that this research is experimental in nature. I understand that this research is intended to study how well EMDR lowers test anxiety and improves classroom test performance. I understand that this research is also intended to study how well EMDR lowers test anxiety for students who seem to become nervous when taking tests even though they have studied well. I further understand that this project is John Hampel's dissertation project.

My consent to participate in this project indicates that I will be asked to attend one, one hour pre-treatment assessment session with John Hampel. I will also be asked to attend between one and three 90-minute EMDR treatment sessions with John Hampel to be scheduled within a three week period. I understand that the first of my three treatment sessions may be scheduled at any time within the next six weeks. I understand that if, after three, 90 minute EMDR sessions I need further counseling about my test anxiety, John Hampel is prepared to make a referral for appropriate psychological services. However, I will be responsible for the cost of therapy if I choose to pursue it.

During the first one hour session I will be asked to provide some specific information about myself including: (a) any history of eye injury, disease, or weakness, and (b) any current or expected (formal) training in study skills that I expect to receive during this study. If the results of this interview disclose that either of these items are (or will, during the course of this study, be) true for me, then I will be excluded from the remainder of the study and will be provided with any and all appropriate referrals. If neither of these items mentioned pertain to me, and I agree to participate in this study, then the balance of the first one hour meeting will be spent discussing my test anxiety related problems.

During this first session I understand that I will be asked to complete (or review) three questionnaires: the Reactions to Tests, the Test Attitude Inventory, the Self-evaluation Questionnaire. I will also be asked two questions about my expectations for EMDR treatment. Further, I understand that I will be asked to complete these same three questionnaires at the conclusion of my treatment.

I understand that treatment for test anxiety using Eye Movement Desensitization and Reprocessing will begin in the second session with John Hampel. During this 90-minute session I will be asked to concentrate on or imagine various aspects of taking tests or studying while following John Hampel's moving fingers with my eyes. During this session I will also be asked to report on the nature and intensity of the anxious thoughts and feelings that I experience. I will be asked to report how much discomfort I feel. I will also be asked to report how much I believe in certain thoughts.

My consent to participate in this research project also indicates that I will be asked to complete the Self-evaluation Questionnaire and answer six brief questions prior to each test that I take in this course. I will also be asked to complete the Self-evaluation
Questionnaire and the Cognitive Interference Questionnaire, and answer two questions following each test that I take in this course.

As in all research, there may be unforeseen risks to the participant. If an accidental injury occurs, appropriate emergency measures will be taken; however, no compensation or treatment will be made available to the subject except as otherwise stated in this consent form.

I understand that one potential risk of my participation in this study is the loss of study time due to time spent in treatment sessions. I understand that John Hampel will be prompting me to plan ahead so that I have sufficient time to study for all of my classes.

I understand that other forms of treatment that have been shown to be effective for reducing test anxiety (and that I may wish to consider) include systematic desensitization, various forms of relaxation training, stress management or stress inoculation training, and various forms of cognitive-behavioral therapy.

One way in which I may benefit from this activity is reducing my levels of anxiety that I feel when I take tests. This may help me improve my academic test performance and my overall academic functioning. This may also help me attain a more academically challenging career goal. I also understand that others who experience test anxiety may benefit from the knowledge that is gained from this research.

I understand that all the information collected from me is confidential and will be used only in a scientific, scholarly fashion. That means that the information I provide during this study will be written and/or published in such a way that I cannot be identified. It also means that my name will not appear on any papers on which this information is recorded. The forms will be coded, and John Hampel will keep a separate master list with the names of participants and the corresponding code numbers. Once data are collected and analyzed, the master list will be destroyed. All other forms will be retained for three years in a locked file in the principal investigator's laboratory.

If I have any questions or concerns about this study, I may contact either John Hampel at 343-9466 or 387-4456, or Dr. Galen Alessi at 387-4470. I may also contact the Chair of Human Subjects Institutional Review Board at 387-8293 or the Vice President for Research with any concerns that I have.

Answering and initialing the questions below and signing this form below indicates that I understand the purpose and requirements of the study and that I agree to participate. I understand that I may refuse to participate or quit at any time during the study without prejudice or penalty.

1. Do you have a history of eye injury, disease, or weakness?
   YES______ NO______ Initials:______

2. Will you receive formal study skills training during the period of this study?
   YES______ NO______ Initials:______

__________________________________________________________________________  ____________
Signature                                      Date
Appendix E

Signed Consent From Irwin Sarason, Ph.D. to Use Reactions to Tests Assessment Device in the Current Study
September 15, 1993

The Graduate College
Western Michigan University
Kalamazoo, MI. 49008-5121

To whom it may concern:

I, Irwin G. Sarason, do hereby give my permission to John C. Hampel to use the Reactions to Tests (RTT) scale, the Cognitive Interference Questionnaire (CIQ), and the Thought Occurrence Questionnaire (TOQ), three psychometric assessment instruments that I have created, for purposes of the assessment and treatment evaluation of experimental subjects in John C. Hampel's dissertation. I do further give my permission to John C. Hampel, Western Michigan University and UMI to publish references to these instruments, their psychometric characteristics, and the levels of individual and group performance as measured by these instruments, in John C. Hampel's dissertation and any ensuing document describing this research that John C. Hampel may publish. I give my permission so that UMI may supply copies on demand.

Cordially,

Dr. Irwin G. Sarason, Ph.D
Department of Psychology
NI-25
University of Washington
Seattle, Washington 98195
BIBLIOGRAPHY


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


