

COUNSELLING COUPLES:  
A META-ANALYTIC REVIEW OF THE LITERATURE

by

Carolyn Brenda Woszczyzna

B.A., York University, 1989

B.A.(Hons.), York University, 1991

THESIS SUBMITTED IN PARTIAL FULFILMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF  
MASTER OF ARTS  
in the Faculty  
of  
Education

© Carolyn B. Woszczyzna 1995

SIMON FRASER UNIVERSITY

July 1995

All rights reserved. This work may not be reproduced in whole or in part, by photocopy or other means, without permission of the author.

## APPROVAL

**NAME** Carolyn Brenda Woszczyzna  
**DEGREE** Master of Arts  
**TITLE** Counselling Couples: A Meta-analytic  
Review of the Literature

**EXAMINING COMMITTEE:**

**Chair**

  
Lucy LeMare

---

Adam Horvath  
Senior Supervisor

---

/  
Janny Thompson  
Assistant Professor  
Member

---

Dr. John Friesen, Professor  
Counselling Psychology Department  
Faculty of Education  
University of British Columbia  
External Examiner

**Date:**

Aug 1, 1995

## PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay

**Counselling Couples: A Meta-analytic Review of the Literature**

---

---

---

---

Author:

\_\_\_\_\_/\_\_\_\_\_  
(Signature)

\_\_\_\_\_  
(Name)  
August 1/95  
\_\_\_\_\_  
(Date)

## Abstract

Meta-analysis was employed to explore the efficacy of and examine the impact of moderator and mediator variables on the outcome of couples counselling. Sixty-two pieces of research, involving more than three-thousand participants, were reviewed in this study. Results showed an overall effect size (ES) of 0.69. This ES was weighted by sample size and methodological quality of the research.

Studies were contrasted based on research design and treatment factors. Research design factors included the study design, whether the research used random assignment, whether participants were self-referred and whether the study was published. Treatment factors that were considered were the age of the participants, the length of the relationship of the couples, the therapist's experience/education, the length of treatment and theoretical orientation. Results yielded no significant differences among these variables. A discussion of the findings and implications for future research are considered.

## Acknowledgements

To my esteemed committee: Drs. Adam Horvath, Janny Thompson, John Friesen. Thank you so much for all of your help and guidance. Thank you for the opportunity to work on this project Adam, I am grateful.

To, Reo Audette, a huge thanks for all your help with those never-ending hours of ugly statistics.

For the nearly two years of support from my "thesis group": Brent, Mat, Sheila, Sophia, Steve. The five of you have truly made this a reality.

To my extended family in B.C.: Gill, our credo: How would I have made it through without you?! From last minute administrative assistance to hours of fun, THANKS! It's a good thing you're in crawling distance. Heather and Ron, where would I be without breakfasts, "fam dinners", and endless hours of talk? I'll never think about this crazy grad school thing without undying gratitude to the two of you. Sandy, I can't tell you how those hours of gabbing and squash made a difference this year! Kirk, thank you for being such an incredible M.A./Ph.D. inspiration. You've been a kind and wonderful friend and mentor.

To my long distance family in T.O.: Jan, not only do I need to thank you for all your brilliant insight into this thesis process, but especially for your help with

Chapter Two when I was most muddled. Ro, what can I say? Your undying support has been remarkable. Don't worry, one down, one to go. I'll almost home.

Most importantly, to my real family: Hen, for those fabulous letters when I needed most to hear about what was going on at home. Paul (ever-lovin') and Jenn, I don't know what I'd've done without all your love. Your phone calls over the miles have kept me going through all my turbulence. Finally, to Dad, the real author in the family. Your love and support have kept me full, kept a roof over my head, kept me laughing, and kept me proud of myself when I'd forget to be. I truly appreciate all your help. I love you with all my heart.

Love and miss you mom.

## Table of Contents

Approval . . . . .	ii
Abstract . . . . .	iii
Acknowledgements . . . . .	iv
List of Tables . . . . .	xiv
List of Figures . . . . .	xv
Introduction . . . . .	1
The Development of Couples Therapy . . . . .	1
The influence of the social work tradition . . . . .	1
The influence of the child guidance movement . . . . .	2
The influence of psychoanalysis on couples therapy . . . . .	4
The influence of behavioural theory on couples therapy . . . . .	5
The influence of systems theory on couples therapy . . . . .	6
Efficacy in Therapy . . . . .	8
Controversy within psychotherapy outcome research . . . . .	8
Approaches to Summarizing Research . . . . .	9

Early qualitative reviews . . . . .	9
Quantitative reviews . . . . .	10
The Literature Summarizing Efficacy of Individual Therapy . . . . .	11
The Literature Summarizing Efficacy of Couples Therapy . . . . .	11
Meta-analyses conducted on research with couples . . . . .	13
Research Questions . . . . .	14
Definition of Terms . . . . .	15
Ratings of the research quality . . . . .	19
Limitations of the Study . . . . .	20
Organization of this Thesis . . . . .	20
 Literature Review . . . . .	 21
Summary of the Research on Couples Therapy Literature . . . . .	 21
Issues in Qualitative Review Methodology . . . . .	22
The qualitative reviews . . . . .	22
Efficacy as Summarized in the Qualitative Reviews	25
Efficacy of couples counselling at post-treatment . . . . .	 25
Efficacy of couples counselling at follow-up	25
Moderators and Mediators in Literary Reviews . . .	26
Moderators and Mediators of Research Design . . .	27
Study design . . . . .	27



Theoretical orientation . . . . .	31
Length of treatment and number of therapists	32
Therapist characteristics . . . . .	32
Participant characteristics . . . . .	33
Length of the couple's relationship . . . . .	33
Issues in Quantitative Review Methodology . . . . .	34
The art and science of meta-analysis . . . . .	34
Psychotherapy outcome and meta-analysis	
controversy . . . . .	35
Meta-Analytic Reviews of Couples Therapy . . . . .	37
Hahlweg and Markman (1988) . . . . .	37
Plattor (1990) . . . . .	38
Shadish, Montgomery, Wilson, Wilson, Bright and	
Okwumabua (1993) . . . . .	38
Summary of the Meta-Analytic Reviews . . . . .	39
Efficacy at post-treatment . . . . .	39
Efficacy at follow-up . . . . .	41
Research Design Moderators and Mediators . . . . .	42
Study design . . . . .	42
Use of outcome measures. . . . .	43
Quality rating . . . . .	44
Experimenter allegiance . . . . .	45
Self-referred vs. recruited participants . . . . .	46
Published vs. unpublished data . . . . .	46
Treatment Moderators and Mediators . . . . .	48
Theoretical orientation . . . . .	48

Experience/education of therapists . . . . .	50
Length of treatment . . . . .	50
Age and length of relationship . . . . .	51
Comparison of Meta-Analyses with Qualitative Reviews	
. . . . .	51
Critical issues in meta-analysis . . . . .	52
Method . . . . .	55
Data Gathering Procedures . . . . .	55
Bibliographies of outcome studies and reviews	
. . . . .	55
Current journals . . . . .	56
Criteria for inclusion . . . . .	56
Analogue and single-case studies . . . . .	58
Missing data . . . . .	58
Data Extraction . . . . .	59
Effect Size Estimation . . . . .	59
Procedures for Data Analysis . . . . .	60
Conversion of raw data to ESs . . . . .	62
Combining and contrasting ESs . . . . .	63
Procedures for Quality Rating . . . . .	66
Results . . . . .	69
Demographics . . . . .	69
Characteristics of studies . . . . .	69

Treatment. . . . .	70
Efficacy . . . . .	71
Research Design and Treatment Moderators and Mediators . . . . .	72
Diffuse comparisons . . . . .	72
Study design . . . . .	72
Self-referred versus recruited participants .	73
Published versus unpublished research . . . .	74
Treatment Moderators and Mediators . . . . .	74
Theoretical orientations . . . . .	74
Participant's age . . . . .	75
Length of the couples' relationship . . . . .	75
Length of treatment . . . . .	75
Influence of the experience/education of the therapist . . . . .	76
Quality Rating . . . . .	76
Intra-class correlations . . . . .	76
Overall quality weighting . . . . .	76
The questions tested with quality weighted Ess . . . . .	77
Diffuse tests with quality weighted Ess . . .	77
Focused tests with quality weighted ESSs . . .	77
Discussion & Conclusions . . . . .	79
Summary of Findings . . . . .	79

Efficacy at posttreatment . . . . .	79
Efficacy at follow-up . . . . .	79
Research design moderators and mediators . . .	79
Treatment moderators and mediators . . . . .	80
Comparison of These Findings with the Findings of	
Other Research . . . . .	80
Efficacy at posttreatment . . . . .	80
Efficacy at follow-up . . . . .	80
Research Design Moderators and Mediators . . . . .	81
Study design . . . . .	81
Published vs. unpublished research . . . . .	81
Self-referred vs. recruited participants . . .	82
Treatment Moderators and Mediators . . . . .	82
Theoretical orientation . . . . .	82
Experience/education of therapists . . . . .	82
Interpretation of Findings . . . . .	83
Efficacy . . . . .	83
Research design and treatment . . . . .	83
Limitations of This Study . . . . .	84
Clinical significance . . . . .	84
Outcome measures. . . . .	85
Recommendations for Future Research . . . . .	85
Efficacy . . . . .	86
Clinical significance . . . . .	86
Separation/divorce . . . . .	88
Individual change . . . . .	88

Follow-up . . . . .	89
Prevention . . . . .	89
Research Design and Treatment Moderators and Mediators . . . . .	90
Study design . . . . .	90
Qualitative research . . . . .	91
Theoretical orientation . . . . .	91
Participant and therapist information . . . . .	92
Conclusions . . . . .	93
References . . . . .	95
Appendix A . . . . .	105
Appendix B . . . . .	107
Appendix C . . . . .	108
Appendix D . . . . .	109
Appendix E . . . . .	114
Appendix F . . . . .	115
Table 1 . . . . .	126

Table 2 . . . . . 131

Table 3 . . . . . 133

Figure 1 . . . . . 134

Figure 2 . . . . . 135

## List of Tables

Table 1	
Study Characteristics . . . . .	130
Table 2	
Tests of Heterogeneity . . . . .	131
Table 3	
Tests of Heterogeneity weighted by Quality .	132

## List of Figures

### Figure 1

Weighted ES by study design . . . . 132

### Figure 2

Weighted Ess by orientation . . . . 133



## Chapter I

### Introduction

According to Statistics Canada (1994), 80,998 Canadians were divorced in 1989, 78,463 in 1990, and 77,020 in 1991. For many people, marital discord is and will be a disturbing problem. In fact, next to the death of a close family member, divorce and marital separation are the most stressful events in an adult's life (Holmes & Rahe, 1967).

These numbers illustrate the amount of marital discord that occurs among people who are married. However, couples counselling is not only utilized by people who are married. A couple can be any two people, married or unmarried, heterosexual, gay, or lesbian. Couples counselling is sought for a multitude of problems that occur in a relationship. Although it is impossible to categorize the various factors that motivate couples to seek counselling, we do know that couples have been treated formally for more than seventy years.

This chapter includes a description of the development of psychotherapy with couples. In addition, this chapter presents the rationale for conducting this study and outlines the research questions that will be considered.

#### The Development of Couples Therapy

The influence of the social work tradition. One group of professionals who work with couples are social workers. Social work began in the late nineteenth century and grew

out of the charity movements in the US and Britain (Nichols & Schwartz, 1991). The role of the social worker was to improve conditions for the underprivileged and poor. Some social workers assessed the needs of families by visiting them in their homes. This social worker was known as the *friendly visitor* (Nichols & Schwartz, 1991). She or he interviewed parents conjointly to assess the family's problems. This group of professionals was one of the first to consider the family as a unit. These interviews, conducted by social workers, were the first formal marital assessments.

Soon after, the first marriage counselling centres were established in the United States. Three centres were opened in the 1920s. Two of these centres were opened by Professor Ernest Groves in 1928 (Reevy, 1967). He was likely the first university professor to teach a marriage and family course in the United States. He also encouraged other sociologists to teach and counsel couples (Reevy, 1967). The third centre was opened in 1929, by two physicians, Abraham and Hannah Stone (Reevy, 1967). The Stones opened a consultation centre for those seeking premarital and marital guidance.

The influence of the child guidance movement. Another group of professionals who influenced the development of couples therapy were those involved in the child guidance movement. At the end of the last century, changes in the

legal status of children created mandatory education of children and restrictions on child labour. During this period, Freud postulated that problems in adulthood stemmed from the emotional disorders that began in childhood. One of Freud's most distinguished followers, Alfred Adler, believed that to prevent adult neuroses, therapy should be conducted with children. In Vienna, Adler organized child guidance clinics for not only children, but also for parents and teachers.

In the 1920s, child guidance clinics were opened across the United States. Treatment was conducted by psychiatrist-psychologist-social worker teams, that worked mostly on the child's environment (Nichols & Schwartz, 1991). At first, there was no consideration of the interpersonal dynamics that occurred within the family. However, within a few decades, these professionals concluded that the real problem lay in the tension wrought in the family. Mothers and children were treated separately. Most often the child was seen by a psychiatrist, while the mother was seen by social workers. The social workers helped the mother cope with her anxiety and stress about dealing with the child. Usually fathers were ignored.

In the 1940s and 1950s, American child guidance workers focused a great deal on the role of parental "psychopathology." It is clear that the emphasis shifted

from blaming the problems of the child on the parents, to seeing pathology as inherent in the relationships that developed between children and parents. The new goal became understanding relationships between parents and children, and teaching parents ways to support their children. Some believe that at this time, professionals practiced couples counselling without a theory that was particular to the treatment of couples (Gurman, 1979).

The influence of psychoanalysis on couples therapy.

The American Association of Marriage Counselors (AAMC) was the first organization dedicated to work with couples. In 1945, small groups of physicians conducting marriage counselling organized themselves into this single professional organization (Plattor, 1990).

At the time of its inception, the members of the AAMC did not yet practice marital counselling with the couple directly. Rather, in the 1940s, the partners in a relationship were usually seen separately by a counsellor (Plattor, 1990). This was a result of the powerful influence of psychoanalysis; therapists viewed marital difficulties as a consequence of intrapersonal impasses. Psychoanalysis continued to be thought of as the superior form of therapy by most clinicians into the 1960's (Garfield & Bergin, 1986).

In 1961, the "Joint Commission on Mental Illness and Health" report included a description of the clientele of

psychoanalysis as middle and upper middle class. The investigators on the committee expressed dissatisfaction with psychoanalysis because it was not available to all people in the United States (Garfield & Bergin, 1986). This report illuminated the needs of underprivileged people and the need for crisis intervention, as well as briefer therapies. Psychoanalysis could not meet all these needs.

The influence of behavioural theory on couples therapy. By the middle of the 1960s, new theoretical orientations were being introduced in the field of marital counselling. The principles of learning and behaviour theory, were creating a stir because they opposed the psychoanalyst's belief that problems were due to unconscious conflicts (Garfield & Bergin, 1986). Behaviour therapists posited that all behaviours were learned. Consequently, behavioural therapists concluded that problem behaviours could also be "unlearned".

For some clinicians, the 1960s brought the influence of behavioural theory to their treatment of couples. Therapists posited that the reinforcers partners provided for one another emphasized the quid pro quo exchange in relationships. They put forth the "exchange model" which stated that troubled marriages were reflections of low rates of positive reinforcement (Jacobson & Martin, 1976). Articles that described the outcomes of behaviourally-oriented couples therapy were first published in the late

1960s (Whisman, Jacobson, Fruzzetti & Waltz, 1989). Today, behaviour therapy is one of the most researched therapies in the area of couples counselling (Gurman, Kniskern, & Pinsof, 1986).

The influence of systems theory on couples therapy.

In 1946 to 1947, systems theory began to emerge. Systems were being examined by professionals from such contexts as biology, anthropology and engineering. Scientists worked to reduce sequences of reality into small analysable components (known as reductionism). In other words, they attempted to discover the laws according to which the world operated. As a result, individuals were seen as reacting to reality rather than creating reality (Becvar & Becvar, 1993). This type of linear cause-and-effect thinking dominated the world of therapy until theories about systems created a shift.

Theories of systems focus on patterns of interactions rather than on the specific content of those interactions. From these varying theories of systems came a rudimentary family therapy position (Becvar & Becvar, 1993). That is, the theory included new assumptions about reality, which included describing it in terms of feedback mechanisms, patterns and recursion, reciprocal causality, subjective/perceptions, a here-and now focus, relativism, context and wholism (Becvar & Becvar, 1993).

In the 1960s and 1970s, different schools of systems

aimed at working with the family and couples emerged. Clinicians created therapeutic interventions that were consistent with this new way of thinking. For example, the representative intervention associated with the "strategic" school of systems theory and therapy was the use of the paradoxical intervention. One form of this intervention includes "prescribing the symptom." For example, clients who present with anxiety that is interfering with their enjoyment of sexual contact with their partner might be told to make themselves feel as anxious as possible during their next sexual encounter. The client may then resist this suggestion and feel less anxiety. Alternately, the client may follow the suggestion, feel more anxious, and therefore learn that he or she has some control over the anxiety. Either way, the client may experience relief from the anxious symptoms.

This paradigmatic shift changed the way clinicians conceptualized problems among couples and in the family. The way they made sense of problems had changed, and with it came a change in the practice of therapy. "Family systems" therapy became defined as any psychotherapeutic endeavour that had as its focus changing the interactions between (or among) family members and improving the performance of the family unit, its subsystems, and/or the functioning of individual family members (Gurman, Kniskern, and Pinsof, 1986).

It is important to understand the distinction between family systems counselling and couples counselling. Although it is true that systems theory has influenced many mental health professionals in their research and practice of couples counselling, "systems theory" is just one of the theories that guides the practice of couples therapy. Unfortunately, the terms "family" and "marital/couples counselling" are often used incorrectly to mean the practice of counselling which is based on systems theory. As described in the previous paragraphs, the practice of family and couples counselling may be guided not only on systems theory, but on other theories as well.

#### Efficacy in Therapy

Controversy within psychotherapy outcome research. In 1952, Hans Eysenck stated that dynamic psychotherapy was no more effective for 75% of "neurotics" than no treatment at all. Eysenck's remarks resulted in a controversy over efficacy in psychotherapy. "The outpouring of praise and invective, and of claims and counterclaims [e.g., Bergin, 1971; Rachman, 1971, 1973; Shapiro & Shapiro, 1977], has been an extraordinary phenomenon..." (Bergin & Lambert, 1978, p.140).

Eysenck based his claim on the results of 24 outcome studies. When other scholars reviewed these same 24 studies, they discovered much variation in their results. In fact, the reviewers found different percentages of



improvement depending on their method and criteria of analysis. In addition, one reviewer retabulated the results of the same set of studies to show that researchers with different allegiances can arrive at different percentages of improvement (Garfield and Lambert, 1978).

Perhaps the most significant consequence of Eysenck's article was the recognition that there were several contradictory results in psychotherapy outcome. These differences were difficult at best, and impossible at worst, to resolve. As a result, a host of researchers in the 1960s and 1970s worked to improve the methodological assessment of the effects of psychotherapy. For example, in one chapter of The Handbook of Psychotherapy and Behavior Change (2nd Ed.), Garfield and Lambert (1978, pp. 171-179) outlined six recommendations for improving the quality of outcome research. These and other recommendations created a new standard of quality in psychotherapy research. With new and better ways of assessing the efficacy of psychotherapy, behavioural scientists began to work on finding improved methods for combining and comparing the results of psychotherapy research.

#### Approaches to Summarizing Research

Early qualitative reviews. Perhaps the earliest and most common approach used to review outcome literature was the qualitative method. Reviewers using this method chose

exemplary articles from the area being evaluated and commented on the strengths of these articles. A second approach to reviewing outcome literature was the "box count" method. This method is considered "qualitative" since each study is placed in a category depending on whether a statistically significant relationship was found. Each category is totalled and compared.

A shortcoming of this procedure is that it does not examine variables that contribute to change. Smith and Glass (1977) stated that this traditional method is biased in favour of large sample studies. Rosenthal (1987) adds that this method has the possibility of having a higher incidence of Type II errors (failing to reject null hypotheses that are false) when compared to a quantitative method.

Quantitative reviews. Meta-analysis is a quantitative procedure that converts the results of diverse studies into a common metric. This enables the researcher to compare and combine results. The strength of this method is that it allows not only a relatively unbiased aggregation of outcome statistics, but also a comparison of different variables between studies. For example, a question commonly researched is "what contributions are made by theoretical orientation in therapy?" In a qualitative review in 1975, Luborsky, Singer, and Luborsky claimed that all therapeutic orientations were equally effective in

individual counselling; i.e., as the Dodo bird in Alice in Wonderland proclaimed: "Everyone has won and all must have prizes." In a meta-analysis conducted by Smith and Glass (1977), no differences between theoretical orientations in individual psychotherapy were found. The claim put forth then, by the qualitative review by Luborsky et al. (1975) was supported.

#### The Literature Summarizing Efficacy of Individual Therapy

Recently, authors (e.g., Whiston & Sexton, 1993; Lambert, 1991) have described the results of individual psychotherapy outcome research in a way that is so compelling that the question of whether psychotherapy is effective, is no longer the main concern of researchers. Instead, the more important questions now involve discovering which aspects of therapy moderate psychotherapeutic changes in clients.

In other words, the debate has moved from the territory of efficacy, to the less charted domain of the variables that contribute to change. The cumulation of outcome research has made it possible to begin summarizing the variables that contribute to change.

#### The Literature Summarizing Efficacy of Couples Therapy

Although a great deal of research is available on psychotherapy outcome with individuals, comparatively little research focused on psychotherapy with couples. This is surprising, given that by the 1970's, marital

therapy had become one of the most important psychotherapeutic techniques employed in the mental health field (Gurman, 1979).

Nevertheless, some reviews have been published on couples psychotherapy outcome research (e.g., Gurman, Kniskern, & Pinsof, 1986; Beck, 1975; Gurman, 1979). To date, therapies from a number of differing theoretical orientations have claimed evidence of efficacy (Jacobson & Addis, 1993; Beck, 1975; Gurman, Kniskern & Pinsof, 1986). For example, behavioural couples therapy (BCT) has shown efficacy over control groups in over two dozen published studies (Hahlweg & Markman, 1988). A second form of couples therapy that has been shown to be effective is emotion-focused couple therapy (EFT). EFT has been evaluated in three trials (Goldman, 1987; James, 1991; Johnson & Greenberg, 1985) two of which were replications of the initial trial.

Nevertheless, it may be shortsighted to attribute differences in outcome of psychotherapy to theoretical orientation of treatment alone. To date, the reviews of therapy with couples include an examination of the efficacy of different theoretical perspectives, but little research has considered the role of research design and treatment variables that may contribute to changes in therapy (Shadish & Sweeney, 1991).

There are still several uncertainties regarding

variables that influence efficacy. The present review includes not only an aggregated estimate of the efficacy of psychotherapy with couples, but also an investigation of some of the variables that influence psychotherapy outcome. I have asked questions such as, are there differences between published or unpublished studies? How do the age and length of relationship of the subjects effect outcome? Is there a difference in outcome if clients are self-referred or solicited by the investigator?

Meta-analyses conducted on research with couples. To date, three meta-analyses on couples therapy have been published. The first of these (Hahlweg & Markman, 1988) examined the effects of one theoretical orientation, Behavioural Couples Therapy (BCT). A second study (Plattor, 1990) explored the effects of three types of BCT and insight oriented couples therapy. The third was a meta-analytic review of couples and family counselling of differing theoretical orientations (Shadish, Montgomery, Wilson, Wilson, Bright & Okwumabua, 1993). In this study, only behavioural measures of change were used to aggregate and compare research design and treatment variables. All these meta-analyses included studies that used experimental designs only (i.e., treatment vs. control). These three reviews will be examined in more detail in Chapter Two.

The present research differs from the aforementioned reviews in five ways. First, only articles that included

conjoint therapy are included; i.e., couples who were seen together by one or more therapists. All three of the previously mentioned articles included both individual and group therapy for couples. Second, this research includes a review of over 60 articles, chapters and dissertations from 1970 to 1994. Therefore, this review includes more studies than ever considered previously, as well as more recent publications. Third, this overview of the outcome research includes studies from diverse theoretical orientations. Fourth, quasi-experimental designs were included in this analysis. Fifth, this research includes the results from client, therapist, and observer measurements from each study based on the information available.

In sum, the purpose of this study was to employ meta-analysis to evaluate the efficacy of couples counselling. This thesis also included an investigation of the contributions of theoretical orientation and therapist and client variables to couples psychotherapy outcome.

### Research Questions

Each of the following eleven research questions was considered in this research.

#### Questions Related to Efficacy

- 1) Is couples therapy effective?
- 2) Is couples therapy effective at follow-up?

#### Research Design Moderators and Mediators That May Influence

ES

- 3) Are there any differences between studies when the subjects are self-referred compared to when they are solicited by the investigator(s)?
- 4) Does efficacy differ significantly when experimental studies are compared to quasi-experimental studies?
- 5) Does efficacy differ significantly between treatment versus treatment studies, treatment versus control studies and pre-post studies?
- 6) Is there a significant difference in efficacy between studies published and studies that were not published?

Treatment Moderators and Mediators That May Influence ES

- 7) Are there differences in efficacy between theoretical models?
- 8) Does the age of the participants in the study influence outcome?
- 9) Does the length of the couples' relationship influence outcome?
- 10) How does outcome differ according to the experience/education of the therapist?
- 11) Does length of treatment influence the effect of therapy?

#### Definition of Terms

Meta-analysis. A quantitative method used to aggregate and compare effects of research studies.

Effect size (ES). The effectiveness of therapy is

expressed in effect sizes. These effect sizes were computed based on client reports, therapist reports, and observer reports of change in therapy.

Couples therapy. Two people in a significant heterosexual relationship, either married or living together, who are engaged in therapy together with one or more therapists.

Behavioural Couples Therapy (BCT). Includes both behavioural and cognitive interventions which focus on directly changing the presenting symptoms. The therapy is based on both social learning theory and behavioural exchange theory. Specifically, this therapy typically includes interventions such as communication skill training, problem-solving skill training, and cognitive restructuring (Hahlweg & Markman, 1988)

Generic/pragmatic therapy. "Generic" counselling refers to therapy that was not clearly defined by the author; i.e., the treatment orientation was only described as "marital/couples counselling." The term "pragmatic" was used when the therapists in the study each employ their own approach with the couples in therapy, or the interventions used by the practitioners originated from different theoretical approaches to therapy.

Psychodynamic therapy. Any therapy that investigates the individual's intrapsychic dynamics (which are partly unconscious) and the contributions of the individual's



personality, motivations, etc., to the interactions of the couple. Such therapies include insight-oriented, psychoanalytic, or psychodynamic.

Eclectic therapy. The combination of three or more distinguishable orientations of therapy. For example, Relationship Enhancement (or Conjugal) therapy, is a weave of psychodynamic, humanistic, and behavioural interventions into one treatment.

Systemic therapy. All therapies which have as their foundation the theory that the members of the couple are part of a stable system governed by patterns and the thoughts/behaviours of the individuals in the system. Examples of systemic orientations include: strategic, structural, emotionally focused, and Milan models.

Moderating variable. A third variable that influences the strength of the relationship of the independent and dependent variable (Shadish & Sweeney, 1991). For example, a moderating variable may be the length of the treatment, the therapist's experience, or the presenting problem.

Mediating variable. A third variable that is caused by the independent variable that in turn effects outcome (Shadish & Sweeney, 1991). For example, a psychodynamic orientation leads the therapist to examine the subject's childhood that in turn leads to a new insight that enables the subject to become "unstuck" in his/her relationship. Such mediating variables are also called process variables.

### Overview of Data Gathering Procedures

The data used in this thesis came from four sources. The first sources were databases (PsychInfo, PsychLit, Medline, Dissertation Abstracts, and ERIC). Next, a cross-check of all bibliographies of the articles collected was conducted to ensure that all relevant articles were assembled. Other reviews and meta-analyses were reviewed to ensure that all outcome studies conducted on couples were collected. Finally, a manual search of all relevant journals was carried out.

Articles were not included in this review if there were fewer than five couples involved in the study, or if there were no usable statistical analyses reported by the author(s).

### Overview of Procedures for Data Analysis

Once the articles for the analysis were gathered, all of the potentially relevant data were extracted from them and placed in a database. Next, Hedges's  $g$  (Hedges, 1982) effect sizes (ESs) were calculated for each outcome measure employed in the 62 pieces of research.

Prior to combining and comparing, groups of ESs were tested for heterogeneity and adjusted for bias. A bias occurs because as more and more ES samples are taken from the population, the distribution of the ESs becomes skewed (Rosenthal, 1994). In addition to correcting this bias, each of the overall ESs were weighted by sample size before

combining them. This weighting procedure ensured that each of the data sets received a weight that was in proportion to the variance. Finally, the groups of ESSs were compared in order to answer the research questions addressed in this thesis. The meta-analytic procedures were based on the methods outlined by Rosenthal (1990), Cooper and Hedges (1994), and Shadish (1992).

Ratings of the research quality. Each of the 62 pieces of research were reviewed by three graduate students. These three students independently rated the research on issues of methodological quality and design. The first quality rating scale consisted of fourteen questions designed to evaluate the internal and external validity of the studies (see Appendix A). This scale is called the Methodological Validity Scale (MVS; Woszczyzna, 1995).

The second quality rating scale was the Smith, Glass and Miller (1980) Reactivity Scale (see Appendix B). This scale includes five questions that aid the meta-analyst to rate the amount of reactivity that may occur as a result of the goals of the therapist or experimenter. That is, "these questions discern whether or not outcomes which were influenced by the therapist, who had an acknowledged interest in achieving predetermined goals; or which are subject to the client's need and ability to alter his [sic] scores to show more or less change than what actually took

place" (Smith et al., 1980, pp. 66-67).

### Limitations of the Study

Some might argue (e.g., Greenwald & Russell, 1991) that a limitation of this research is that there were no exclusion criteria based on the quality of the studies. For example, articles were included if the primary investigator was also the therapist, if there was only one therapist employed by the investigator, if there was a lack of equivalency between therapists, if there was clear experimenter bias, or if there was no follow-up conducted. On the other hand, all these questions and several others were considered when first, each article included in the research was rated by the quality and second, when we considered how much impact the quality of each article had on its ES.

### Organization of this Thesis

Chapter Two is a review of the literature on qualitative and quantitative research on couples therapy. Chapter Three is a detailed description of the method used for this study. In Chapter Four, a report and discussion of the findings can be found. Chapter Five includes conclusions and implications of the findings of this thesis.

## Chapter II

### Literature Review

#### Summary of the Research on Couples Therapy Literature

This chapter is a review of the research literature on couples therapy. As described in the previous chapter, the processes used to summarize research became a controversial issue as a result of the Eysenck invectives in 1952. Since 1952, researchers in psychotherapy outcome have worked hard to strive for quality and rigour.

This chapter is broken into two parts: Issues in Qualitative Review Methodology and Issues in Quantitative Review Methodology. Within each of these parts, three major areas will be considered: Efficacy; Moderators and Mediators of Research Design; and Moderators and Mediators of Treatment. Finally, qualitative and quantitative methodologies will be compared.

Perhaps it is obvious that the main and original purpose for not only conducting quality psychotherapy research, but also for reviewing it, was to answer the question "does therapy work?". In addition to the question of efficacy, researchers tried to discover answers about the comparative effectiveness of treatments based on different theories and interventions (e.g., behavioural and non-behavioural), how they are applied to different presenting problems, whether there are harmful effects on

clients, what qualifications and skills of the therapist effect outcome, and many other relevant matters. Research synthesis provides the researcher and practitioner with a framework for discovering what is known about therapy with couples.

Olson, Russell, and Sprenkle (1980), and Gurman (1979), described the growth of couples therapy in developmental terms; in the 1950s family therapy was described as an infant; in the 1960s, marital and family therapy reached childhood; in the 1970s this field reached young adulthood. Typical of young adulthood, couples therapy became self-involved and struggled with its independence from family therapy (Gurman, 1979).

Part of this adolescent struggle was the beginning of self-reflective reviews of couples therapy literature. The following paragraphs examine traditional qualitative assessments of couples therapy from the first reviews of the 1970s to the present. The two fundamental aspects of couples therapy research examined the effectiveness and mediating and moderating variables in research design and treatment.

#### Issues in Qualitative Review Methodology

The qualitative reviews. One of the first reviews of marital therapy was conducted by Gurman in 1973. Like other researchers at this time, Gurman was most interested

in the efficacy of psychotherapy with couples. He considered 26 studies in his review. An overall improvement rate of 66% was found across a group of heterogeneous studies of couples therapy. That is, 15 of the 26 studies resulted in improvements for couples.

In 1975, Beck also found positive rates of change when she summarized three sets of studies. The first of these sets was the 1970 census conducted by the Family Service Association of America (FSAA). This census documented the results of 1,919 cases, of which 1,257 reported a marital problem as the primary issue. About two-thirds of these couples showed improvement. Eight percent of the couples reported a deterioration in the relationship, and only 5% of therapists reported this deterioration.

In addition to the FSAA study, Beck examined the results of several smaller studies which she divided into two categories. The first set included studies that did not utilize a control group. The results indicated that the majority of these cases improved. Next, Beck examined the results of controlled studies which included individual, group, and conjoint therapy for couples. The results indicated that there were always greater improvements for the treatment than no treatment group, regardless of the mode of treatment.

In 1978, Gurman and Kniskern reviewed the couples

counselling literature. They found an improvement rate of 66% for couples in non-behavioural marital therapy, and 64% for couples in behavioural marital therapy. In addition, they found that conjoint couples therapy was more effective for problems with couples than individual marital therapy.

In 1980, Olson, Russell, and Sprenkle reviewed the outcome research on family and marital therapy from the 1970s. Olson et al. agreed with Gurman and Kniskern (1978), that couples therapy was proving the best option for couples experiencing discord. In particular, they stated that conjoint marital therapy proved to be better than individual therapy for marital problems. They also concluded that couples therapy was not only efficacious, but that in the 1970s it "emerged as a viable treatment approach for most mental health problems" (p. 973).

In 1985, Beach and O'Leary reviewed the research on couples therapy conducted using psychodynamic and BCT approaches. Specifically, they reviewed three articles on insight-oriented therapy and two articles on contracting with communication training. Of the three articles on insight-oriented therapy, only one showed significant changes for the couples. The BCT therapies reviewed were based on contracting with communication training. Both of these therapies yielded positive significant changes for the couples.



Most recently, in 1993, Jacobson and Addis reviewed the current state of couples counselling. They concluded that all treatments appear to be equally effective. However, they did note that effects of new therapies may be exaggerated when first researched. Nevertheless, with replication, they do tend to show effects that are consistent with other couples therapies.

#### Efficacy as Summarized in the Qualitative Reviews

Efficacy of couples counselling at post-treatment. It is important to begin by noting that qualitative reviews reported a positive change for the majority of couples who received counselling. Although the merits and contributions of these studies are appreciated by many, others believe that the results found in these reviews were yet to be corroborated through replication, and by quantitative methods. In fact, Garfield and Bergin (1986) stated that until quantitative methods confirmed these early results, the outcome of qualitative reviews should be considered probationary.

Efficacy of couples counselling at follow-up. Most research studies did not include follow-up measurements. When follow-up measurements were done, they were often measured within a short period after treatment. In their article "Methodological issues in marital therapy," Whisman, Jacobson, Fruzzetti and Waltz (1989) note that

follow-ups should be conducted after 12 or 24 months post-therapy. The reason for choosing 12 to 24 months is that Whisman et al. found that nearly one-third of couples who showed improvement at post-test had returned to baseline at the 12 and 24 month follow-ups.

### Moderators and Mediators in Literary Reviews

Some authors suggest that because the efficacy of therapy with couples is established, it is important to change the focus of research from efficacy to other considerations (Jacobson & Addis, 1993; Lambert, 1991; Raffa, Sypek, & Vogel, 1990). As a result, it is important to research questions such as how does study design effect outcome? How does the length of the couples relationship effect outcome? How do such variables as theoretical orientation or therapist experience effect outcome?

Moderators and mediators, as described in the previous chapter, refer to the variables which influence the dependent variable. The influence on the dependent variable is due to either a third variable (moderator), or a process variable (mediator) which is under the influence of the independent variable. Specifically, moderators are any variable that influence the strength of the relationship of the independent and dependent variable (Shadish & Sweeney, 1991). Examples of moderators are therapist experience/education and the referral source.

Mediators are any variable that is caused by the independent variable which in turn effects outcome (Shadish & Sweeney, 1991). Examples of mediators are the standardization of the treatment (did they use a manual?) and the theoretical orientation used in treatment.

#### Moderators and Mediators of Research Design

Study design. In 1989, Whisman et al. stated that one great "strength of marital therapy research methodology is its legacy of elegant sophisticated experimental designs" (p.177). It will become clear from the succeeding paragraphs that what Whisman et al. considered "elegant sophisticated experimental designs" is one of the controversial methodological considerations in couples' research.

For example, Jacobson and Addis (1993), argued that the success of experimental couples versus controls may be less significant than first believed. The reason for this is that the improvement of the control group is so minimal that it does not require great improvements in the experimental group (when compared to the control group) in order to show statistically significant differences. They also stated that it is not uncommon for couples who are not receiving treatment to experience further deterioration in their relationship. Therefore, it is important to consider why and if the significance found in experimental designs

are indeed the most "elegant [and] sophisticated" compared to other research designs.

It has not yet been determined how the design of studies influences outcome or what variables may influence the effectiveness of therapy (Greenwald & Russell, 1991; Rosenthal, 1991b). Nevertheless, as far back as 1973, Gurman stated that the studies included in his review were of "questionable" methodological standard. Specifically, he did not believe that quasi-experimental designs were methodologically sound. Gurman's rationale for this statement was that quasi-experiments do not provide a no-treatment comparison, or placebo, against which improvement could be compared. Consequently, he believed that it was not possible to determine what changes the couples made as a result of therapy. Despite his belief, Gurman included in his review 8 experimental and 18 quasi-experimental design studies.

Similarly, in Beck's 1975 FSAA project, many of the articles used in her study did not include a control group. Like Gurman, Beck did not believe that this study design was adequate. She argued that the results of the controlled studies were the only informative ones because they were compared to a no-treatment group; i.e., results were due to "real gains", and not just "a wish to please

the counselor, mutual good feeling, or a need of either party to feel that his [sic] substantial investment in treatment was worthwhile" (p.161).

In Beach and O'Leary's review (1985), treatment versus control studies were included but quasi-experimental studies were not included. They included treatment versus treatment studies whenever a control group was included. In these cases they considered only the treatment versus control results. These authors maintained that treatment versus treatment and quasi-experimental treatment studies are not relevant to research in such a new field until experimental research has "proven" the effectiveness of the treatment being considered. When that time comes, they argued, quasi-experimental research will be useful to "provide a basis for treatment decision" (p. 1039). However, Gurman, Kniskern, and Pinsof (1986) and Jacobson and Addis (1993) found that the only time a difference was found in treatment versus treatment studies was when the investigator had an allegiance to a particular treatment.

The use of manuals in therapy. Beach and O'Leary (1985) argued that the use of manuals in psychotherapy research helps to reduce investigator bias. They believe that not only is it helpful to include details of the therapy employed in a study to evaluate the validity of the treatment, but it also allows replication of the study.

Similarly, Whisman et al. (1989) maintained that the use of manuals "promote the generality of findings to practice" (p.176).

However, it has been argued that manuals restrict the latitude that practitioners have in therapy, and also reduces the generalizability of the results of experimental research. Whisman et al. (1989), considered the question of treatment flexibility. The authors acknowledged the danger of employing a protocol which was "too" structured. That is, if the conventions for conducting therapy were highly structured then there was a danger that the therapy would reduce clinical flexibility. That is, if the treatment is too structured because it is serving the needs of research, then it may reduce the integrity of the treatment. Normally, clinicians alter their treatment based on the needs and characteristics of the clients (Jacobson & Addis, 1993). It is clear that with regard to the use of manuals there may be a gap between researcher and clinical practice.

Investigator bias can be reduced by ensuring that the primary investigator is not a therapist in the study (Jacobson & Addis, 1993). A difficulty with the studies Gurman (1973) reviewed, was that in almost all of the studies, the investigator was also the therapist in the study; i.e., of the 19 articles which acknowledged who the

therapist(s) was/were, 16 were the author. The therapist/investigator may have had the opportunity to influence the results of this research.

There are other important moderators and mediators which were not considered in the qualitative reviews. This is likely a result of the fact that these variables were not described in the original research. For example, factors such as whether the participants in the studies were self-referred or solicited or whether the study was published or not published are not considered in the early qualitative reviews.

#### Treatment Moderators and Mediators

Theoretical orientation. Jacobson and Addis (1993) concluded that BCT is "the closest thing that couple therapy has to an established treatment. It can also be said that the demonstration of an effect, having been replicated so many times, is unequivocal" (p. 85). However, Jacobson et al. (1984) have also stated that "the absence of conventions for designating a couple as improved has led to an inflated estimate of [BCT's] success rate" (p.503).

In addition to BCT, Jacobson and Addis (1993) note that both emotion-focused couple therapy (EFT), and insight-oriented couple therapy (IOCT), are "providing the beginnings of an empirical basis for recommending these two

widely practiced treatments" (p. 85).

Length of treatment and number of therapists. Apart from theoretical orientation, other moderating variables of treatment examined by Gurman (1973) were length of treatment and the effectiveness of one versus two therapists. Neither of these moderating variables yielded significant differences. However, there was a trend toward more positive outcome for the couples who received shorter treatment. In 1978, Gurman and Kniskern considered length of treatment in their review. They found that only one of the seven studies that examined length of treatment, showed that time-limited couples therapy had better outcome than open-ended therapy. Another study that Gurman and Kniskern reviewed found that briefer therapy yielded better results for husbands.

Therapist characteristics. It was rare for authors of the qualitative reviews to consider therapist characteristics. Gurman and Kniskern (1978) wrote one of the two reviews that considered the experience of the therapist. Gurman and Kniskern stated that more positive outcomes do seem to occur with more experienced therapists. In 1993, Jacobson and Addis also investigated the relationship between therapist experience and outcome. They concluded that there can not be an assumed relationship between therapist experience and therapist



competence. However, they did not review differences in outcome based on therapist experience because they stated that in order to evaluate therapist competence, the study had to include manualization of the therapy and have the therapists judged by "disinterested" evaluators. They concluded that in order to discover which characteristics may be associated with positive outcome, better quality therapist information must be collected.

Participant characteristics. Problems with the outcome studies published were not only that little demographic information was collected from the therapists, but also that little information was collected from the participants in the studies. Only two reviews included a discussion of the relationship between the age of subjects and outcome. First, Gurman and Kniskern (1978) found mixed results. Only two primary studies were reviewed. In one of the studies outcome with couples who had a mean age of 29 showed the best results. The worst outcome were with couples whose mean age was 42. In a second study no differences were found. Second, Jacobson and Addis (1993) found that subjects who received BCT found an inverse relationship between age and outcome. That is, younger couples were more likely to have positive outcomes.

Length of the couple's relationship. Length of relationship of the couple was examined only by Gurman and

Kniskern (1978). They found one study showed the most favourable outcome with couples married 13 to 16 years and another study showed no differences for length of relationship. Clearly, more information needs to be gathered from participants, both as a couple and individually in order to understand how and if the length of a couples relationship influences outcome.

### Issues in Quantitative Review Methodology

The art and science of meta-analysis. Meta-analysis is a statistical technique used to aggregate and compare the data gleaned by different investigations. That is, meta-analysis "converts" the results of many different types of outcome measures into a common metric. An effect size (ES) is the unit used to express the relationship between variables. The meta-analyst may choose from a number of different methods for estimating ESs.

There are two "families" of ES, the  $r$  family and the  $d$  family. The  $r$  family of ES employs Pearson produced moment correlations. The  $d$  family of ES is used to index the magnitude of a linear relationship between two variables by means of a comparison between the means of two conditions (Rosenthal, 1994). However, there are different formulas for calculating ES within the  $d$  family. With the exception of Hedges's  $g$ , the other ES estimators in this family employ the standard deviation of the distribution (or other

sampling units) from the control group (Rosenthal, 1991a). Hedges's  $g$  utilizes a pooled standard deviation, computed from both control and experimental groups, which "tends to provide a better estimate in the long run of the population standard deviation" (Rosenthal, 1991a, p.16).

#### Psychotherapy outcome and meta-analysis controversy.

Meta-analysis was first used in the realm of psychotherapy outcome research by Smith and Glass (1977). Published in an issue of American Psychologist, their study included 375 pieces of research and found an overall ES of 0.68. That is, an effect of .68 standard deviations implies that couples who received treatment moved to the 75th percentile as compared to the average untreated couples who remained at the 50th percentile.

The controversy over the question of efficacy and differential effectiveness began because Smith and Glass' meta-analysis showed less conservative results than qualitative reviews and found all therapies equally effective (within one-half of a standard deviation; Cook & Leviton, 1980).

These results sparked criticism, not surprisingly, by Eysenck. Recall that in 1952 Eysenck argued that dynamic psychotherapy did not work for 75% of "neurotics". Eysenck (1978) replied to the Smith and Glass (1977) publication

with an article entitled "An exercise in megasilliness", also published in the American Psychologist. Eysenck, joined by Rachman and others (e.g., Rachman, 1971; Wilson & Rachman, 1983), maintained that behaviour therapy was superior to non-behavioural therapy and that psychoanalysis and other "verbal" therapies were inferior. Eysenck (1978) and Rachman (Rachman & Wilson, 1980; Wilson & Rachman, 1983) argued that there were several problems with the Smith and Glass meta-analysis. They stated that Smith and Glass omitted too many behavioural articles (in particular, single-case studies); that their attempt to generalize their findings was inappropriate; that as a result of faults these results were erroneous; and that statistical techniques did not compensate for the inclusion of poor quality primary research in the database.

Garfield (1983) criticized Eysenck's arguments by stating that he believed that Eysenck was letting personal feelings colour his opinions of the Smith and Glass results. Despite these criticisms, in 1980, Smith, Glass and Miller expanded their first meta-analysis by compiling an even more comprehensive database of psychotherapy outcomes with 475 studies. Their results showed an overall ES of 0.85, indicating that treated couples moved to the 80th percentile compared to the average couples that were not treated who remained at the 50th percentile.

In order to address some of the criticisms of Smith and her colleagues, Shapiro and Shapiro (1982) conducted a meta-analysis with the intent of taking into consideration Rachman and Wilson's (1980) recommendations. Their results showed that there was a small advantage to the cognitive and behavioural therapies when compared with the psychodynamic and humanistic methods. Nevertheless, the Shapiros' meta-analysis was also criticized by Rachman and Wilson in 1983.

As Garfield (1983) cautions, "we should not be too surprised if individuals representing different value orientations evaluate research studies differently, and this is by no means limited to the area of psychotherapy" (p.39). Thus, meta-analysis, like qualitative reviews, will continue to yield varying results which will leave some more satisfied than others.

#### Meta-Analytic Reviews of Couples Therapy

To date, three meta-analyses have been published on the couples psychotherapy literature.

Hahlweg and Markman (1988). Hahlweg and Markman (1988) were the first to carry out a couples therapy meta-analysis. This cross-cultural comparison of couples therapy research from Europe and the United States showed no differences in outcome based on where the data were collected. They compared the effectiveness of BCT with

that of cognitive-behavioural pre-marital intervention programs.

Plattor (1990). Plattor reviewed 25 studies conducted with distressed couples who sought marital psychotherapy. His review excluded the marital enrichment literature. This study examined four approaches: communication training, contracting, contracting with communication training, and insight-oriented couples therapy. The first three of the four approaches listed are types of BCT and the fourth, insight-oriented couples therapy, is a psychodynamic approach. Plattor included research which only used random assignment and no-treatment/wait-list control groups. He did not include studies which employed a quasi-experimental design, or a placebo control group. In addition, Plattor used only self-report dependent measures "with a relationship change focus" (p.46). Thus, any results from observational/behavioural outcome measures from the primary research were not included in his statistical analysis.

Shadish, Montgomery, Wilson, Wilson, Bright and Okwumabua (1993). Shadish et al. did an extensive review of family and marital psychotherapies. Sixty-two of the 163 studies he reviewed were on couples therapy. He included only articles that used random assignment and either a treatment versus treatment or a treatment versus

control group design. He included research from all theoretical orientations, only if the subjects were distressed. In contrast with Plattor, Shadish used observational/behavioural outcome measurements only.

#### Summary of the Meta-Analytic Reviews

Efficacy at post-treatment. Hahlweg and Markman (1988) found that 17 behavioural couples therapy articles yielded an ES of 0.95. That is, the average experimental participants moved from the 50th percentile to the 83rd percentile (i.e., improvement on outcome measures); "the chance of getting better increases from about 30% (the estimated improvement rate for the control group) to about 70%" (p.445). To be even more conservative, Hahlweg and Markman compared the experimental group with the most difficult control group--participants who received a placebo. They found that BCT resulted in a 26% increase in improvement over the placebo control subjects. Hahlweg and Markman concluded that their findings illustrated the power of the nonspecific factors that operate in therapy, and also that consideration of the control group employed is critical.

Similar to Hahlweg and Markman, Plattor (1990) found a large overall ES of 0.85. This was identical to the ES found by Smith, Glass and Miller, in their 1980 psychotherapy meta-analysis. Broken down by treatment

approach, the ESs were 0.81 for contracting with communication training (in 17 studies); 0.75 for communication training alone (based on 14 studies); 0.90 for contracting alone (based on 4 studies); and 1.22 for insight-oriented couples therapy (based on 6 studies). No significant differences were found among approaches.

Shadish et al. (1993) found an overall ES of 0.60 for marital therapy. There were no significant differences among approaches. ESs by theoretical orientation for between-studies treatment versus control comparisons were: 0.74 for behavioural; 0.62 for systemic; 0.12 for humanistic; 0.63 for eclectic; and 0.62 for unclassified. An ES for psychodynamic couples therapy was not calculated because there was only one study that met the criteria for this category. Within-study treatment-treatment comparisons also revealed no significant differences between theoretical orientations.

Shadish et al. also tested their BCT results against Hahlweg and Markman's results. Recall that Hahlweg and Markman found an overall ES of 0.95 for BCT. Shadish and his group replicated the procedures used by Hahlweg and Markman and found that when they excluded dissertations from their analysis, they found the identical 0.95 ES. When the the dissertations were added, the ES for BCT dropped to 0.74.



Efficacy at follow-up. Hahlweg and Markman found that of the 17 BCT articles, 13 had a follow-up measurement. Five of the studies included follow-ups between 9 to 12 months and eight between 3 to 6 months. The results were ESs of 1.16, and 1.17, respectively. They concluded that there is some stability in the effects from 3 months to 1 year after therapy. No significant differences were found between post and follow-up measurements.

Plattor found that follow-up was conducted in less than half of the articles he analyzed (12 of 25). The range for post-measurement was from 3 to 12 months, with a mean of 6.6 months. Results of ESs for contracting with communication training (based on 10 of 17 studies) was 0.76, communication training alone (based on 5 of 14 studies) was 0.61, contracting (based on 4 of 4 studies included) was 0.68, and finally, insight-oriented couples therapy (based on 2 of 6) was 0.96. The overall follow-up ES was 0.73, which was not significantly different from post-therapy results.

Like Hahlweg and Markman, and Plattor (1990), Shadish et al. found that there were no significant differences between post-therapy and follow-up ESs. Only one study had a follow-up period longer than 1 year, and the median follow-up period was 5 months, none of which exceeded 9

months. Shadish et al. did not consider within-group deterioration at follow-up and stated that future studies are warranted before drawing any conclusions about the long-term effects of therapy.

### Research Design Moderators and Mediators

In 1975, Luborksy, Singer and Luborsky considered the differences between orientations in psychotherapy outcome. They claimed that "Everyone has won and all must have prizes" (p.995) because no differences were found between the orientations. Shadish and Sweeney (1991), attempted to untangle the relationship between orientation and outcome. Specifically, they concluded (as did Luborsky et al. in their discussion) that Dodo birds are not very bright, and therefore did not recognize the large number of moderators and mediators which influence the direction or strength of the dependent, or criterion variables. The following paragraphs examine some of the moderators and mediators that have (or have not) been considered in the three meta-analyses of couples psychotherapy outcome.

Study design. All of the meta-analysts chose their articles for review based on the type of research design of the studies. As described earlier in this chapter, at one end of the argument is the choice to include all studies, and on the other, to choose studies with particular research designs of the "highest" quality. In their

article "Assessing rationales for inclusiveness in meta-analytic studies", Greenwald and Russell (1991) assert that what is important is to have a group of carefully evaluated studies that "fit" with the rest of the studies so that generalizations can be made about treatment effectiveness.

Plattor included studies only if they employed random assignment and a no-treatment/wait-list control group. Hahlweg and Markman (1988) included studies with either no treatment or placebo control subjects. They found that the placebo group yielded lower ESs compared to BCT treatment. That is, compared to the no-treatment group (which yielded higher ESs for BCT (ES = 1.02), the placebo control group compared to BCT only yielded an ES of 0.55. Shadish et al. (1993) found no differences based on type of control group used. They also looked at differences between random assignment and "haphazard" assignment and found that there were no differences between them. In addition, behavioural versus nonbehavioural presenting problem, matching, use of patient exclusion criteria, and treatment standardization (e.g., manualized or not) also made no differences to outcome.

Use of outcome measures. Plattor used only self-report measures with a relationship change focus. Hahlweg and Markman compared self-report with observational measures. Ten of the BCT studies they reviewed used

observational measures and all 17 used self-report measures. They found ESs of 0.99 and 0.97, respectively. Shadish et al. replicated Hahlweg and Markman's finding. They compared all self-reports (ES = 0.77) to ratings of behaviours (ES = 0.70), and also found a nonsignificant difference.

Shadish et al. found that ratings by others versus ratings by self produced higher ESs, so self-reports were considered the more conservative measure. Curiously, Shadish et al. only used behavioural measures, despite their findings that behavioural measures rather than nonbehavioural measures were correlated with higher ESs.

Quality rating. Wilson and Rachman (1983) criticized meta-analysis, stating that it does not eliminate subjectivity or bias. They believe that it is the responsibility of the researcher to include only studies of greatest quality and omit studies of lesser quality. Greenwald and Russell (1991), Smith et al., (1980), Strube & Hartmann (1983), and Shapiro and Shapiro (1982), have suggested weighting studies according to their quality. Meta-analysts have the additional opportunity of using statistical procedures to create a differential weighting system which is chosen a priori and used logically to reduce bias. Using this method, Smith and her colleagues (1980) showed that there was no relationship between

quality of the study and its ES.

Nonetheless, only one of the three meta-analyses examined here included a quality rating. The quality rating used by Shadish et al. was the Smith, Glass and Miller Reactivity Scale (1980). Shadish et al. found that the more reactive the measures were in the primary research, the higher the ES. They also found that the more tailored the measures were to the treatment, the higher the ES. They also found that when total attrition was low, the ES was higher. Although Shadish et al.'s. results are important to acknowledge, recall that they are the result of a combination of the literature on both couple and family therapies, not couples therapy literature alone.

Experimenter allegiance. Plattor states in his research that experimenter allegiance is an inherent part of every study. As a result, he postulates that because this effect is evenly distributed across classifications that it may have a "cancelling out" effect contributing to the finding of no significant differences between approaches.

Plattor's belief is not consistent with the findings of Shadish et al. They found a significant difference between ESs of various predictor variables "when researchers could have influenced subjects' treatment

during the study" (p.996).

Self-referred vs. recruited participants. Plattor (1990) recorded, but did not calculate the differences in outcome based on whether subjects were solicited or self-referred. Interestingly, he did note that 17 of the 25 studies reported the source of referral of their subjects. Of these 17, 71% of the subjects were recruited through various media such as newspapers, radio, and television. Shadish et al. were the only reviewers who examined the effects of referral on psychotherapy outcome. They found that referral source did not correlate with therapy outcome.

Published vs. unpublished data. None of the qualitative reviews considered differences between published and unpublished research. However, it is important in both qualitative and quantitative reviews to consider what Rosenthal (1979) coined the "file drawer problem." One possible methodological problem is that articles are more likely to be published if they report significant findings than those that are equally well-designed but report nonsignificant findings. Some (e.g., Shadish, Doherty & Montgomery, 1989) believe that the "file drawer problem" results in overestimating the significance of effects found in published articles.

In their article "How many studies are in the file

drawer? An estimate from the family/marital psychotherapy literature", Shadish, Doherty and Montgomery (1989), attempted to discover how many unpublished articles could be found. Of the 519 randomly sampled members of five relevant organizations, three articles that would qualify for inclusion in their meta-analysis were recovered. Still, these authors maintain that ESs should only be considered approximately 70% to 90% as large as those found in published studies. The reason for this may be that individual studies that are published may only include selective results (i.e., those that are significant and in the expected direction), or may not report other dependent variables that were examined.

Greenwald and Russell (1991) tested the file drawer problem in their meta-analysis of child psychotherapy. They found that for the twenty-nine treatment/control comparisons they used in their research, they would have to find "390 studies averaging null results before the combined  $p$  level of the average effect size would be less than .05" (p.23).

In the couples meta-analysis literature, Plattor (1990) did not report how many of the articles he reviewed were published or unpublished. However, three dissertations are listed in his reference section. Hahlweg

and Markman also did not report that they included four unpublished manuscripts of their twenty-four studies. From their references we see that two of these four were dissertations and two were papers delivered at conferences. It is interesting that in their meta-analysis, Shadish et al. (1993) criticized Hahlweg and Markman (1988) for not including any unpublished research in their review. As previously described, Shadish et al. attempted to replicate Hahlweg and Markman's results. To do this, Shadish et al. recalculated their own BCT results, excluding dissertations. Recall that without dissertations, they found the same ES as Hahlweg and Markman (0.95), but when they added unpublished dissertations, their ES went down (0.74). In addition, Shadish et al. found that combined couples and family research results showed that published data correlated significantly with higher ESs.

#### Treatment Moderators and Mediators

Theoretical orientation. Hahlweg and Markman (1988) and Shadish et al. (1993) concluded that it is no longer necessary to conduct more research comparing BCT with a control group because the efficacy of BCT is clearly established. Plattor's results confirm this conclusion. Shadish et al. added that systemic and eclectic orientations are not far behind BCT in establishing their efficacy. However, there are some other theoretical



orientations that have yet to "prove" their effects with the couples population. For example, there is such a small number of research articles on psychodynamic therapy with couples that Shadish et al. could not draw any conclusions about this orientation.

There is also a consensus in the results of these meta-analyses that there are no differences in efficacy between theoretical orientations. In four of the studies that Hahlweg and Markman reviewed, BCT was compared to other approaches. The ESs calculated from post-therapy measures yielded an ES of 0.88 for BCT, and 0.83 for the other therapies. No significant differences were found. Shadish et al. found no differences between theoretical orientations in between-studies treatment-control comparisons or within-study treatment-treatment comparisons. However, they did discover a consistent failure of the humanistic approach to create change in couples. That is, the ES that this approach yielded, did not differ significantly from zero.

Gurman, Kniskern and Pinsof stated in 1986 that finding significant differences between therapies in psychotherapy outcome studies "are so rare that they should be considered highly suspect" (p.569). The results of these meta-analyses support this statement.

Experience/education of therapists. In their article,

"Meta-analysis of therapist effects in psychotherapy outcome studies," Crits-Christoph et al. (1991) examined four variables that may influence therapist efficacy. One of these variables was therapist experience. The dependent variable of their study was outcome variability due to differences between therapists. They found that more experienced therapists using a treatment manual showed less variability between therapist effects. Conversely, inexperienced therapists not using a manual showed larger therapist effects. Crits-Christoph et al. concluded that the use of manuals by experienced therapists would minimize differences in psychotherapy outcome caused by therapist variability.

How therapist experience affected outcome was not examined by Hahlweg and Markman (1988) or Plattor (1990). Plattor stated that there was not enough information about the therapists in the primary research for him to do an analysis of therapist factors. However, Shadish et al. (1993) did show that in their couples and family meta-analysis, therapist degree/experience did not correlate with outcome.

Length of treatment. Crits-Christoph et al. (1991), considered the relationship between length of treatment and therapist effect. They found that length of treatment did not predict therapist effects.

Of the three meta-analyses considered here, only Plattor considered the effect of length of treatment on outcome. He found that the longest treatment had the lowest ES (however, this information was based on only one study). Plattor also concluded that couples therapy which lasted for 10 sessions over a 10 week period, seemed to provide positive results in the couples relationship.

Age and length of relationship. These variables were not considered in any of the three meta-analyses on couples therapy.

#### Comparison of Meta-Analyses with Qualitative Reviews

There are a number of reasons why some researchers believe that meta-analysis is a more reliable and superior method of synthesizing research. Qualitative methods such as the box-count method, have been accused of imprecision, omitting important information, and including biased samples of data. However, some researchers believe that meta-analysis has the same shortcomings as traditional qualitative reviews plus an additional set of statistical problems to overcome (cf., Cook & Leviton, 1980; Strube & Hartmann, 1983). The following paragraphs include a consideration of these important issues: Is meta-analysis superior to qualitative reviews because statistical interactions can be uncovered? Does meta-analysis include relevant information that is often not included in

traditional qualitative research? Is the inclusion of articles in a meta-analysis less biased than those reviewed by qualitative reviews?

Critical issues in meta-analysis. Although proponents of meta-analysis claim to examine the statistical impact of independent variables on dependent variables, it is probably more accurate to state that meta-analysis describes the impact of predictor on criterion variables of outcome (Shadish & Sweeney, 1991; Strube & Hartmann, 1983). That is, when we speak of the results of meta-analysis it often sounds as though we are describing causal inferences. However, meta-analysis is not experimental. The meta-analyst does not randomly assign original research to different levels of independent variables, nor do the articles have control comparisons.

This point is especially important with regard to using meta-analysis as a causal model rather than a mediational model. As described earlier, it would be naive to attribute the magnitude of ES to a single variable such as theoretical orientation. Nonetheless, early qualitative reviews often considered theoretical orientation the primary reason for differences in outcome. Clearly, there are numerous moderator variables that influence ESs in meta-analysis. Nonetheless, it is difficult enough to identify moderators in primary research; in a mediational

model, a great number of moderator variables are considered, and as a result, the problems in primary research are compounded for a host of reasons. The most obvious reason is that meta-analysis, due to the consideration of a large number of variables, is unlikely to reach the same standard of quality that is found in primary research (Shadish & Sweeney, 1991). Also, poor quality in primary research affects meta-analytic results in the same way that it affects the results of qualitative reviews. Nevertheless, meta-analysis does afford researchers the opportunity to precisely consider a great many variables when a mediational model is employed.

The fact that meta-analysis and qualitative reviews can only provide accurate results if the primary research is accurate has been the basis of an important question in research synthesis: Which pieces of primary research should be included in the reviews? Some (e.g., Luborsky et al., 1975; Smith & Glass, 1977) have argued that there is no relationship between quality of primary research and results. Therefore, on one side of the argument is the belief that all studies should be included in the reviews. On the other side, is the belief that only very selective data should be included in meta-analyses (e.g., Rachman & Wilson, 1983).

Meta-analysis was once considered less biased than

qualitative reviews because it often included all the research available in a particular area. Qualitative reviews, on the other hand, were considered "subjective" because reviewers chose the articles for review. This claim that meta-analysis is less biased is debatable. Meta-analytic reviews normally include only a subset of studies from any given research area. There are many criteria on which to choose the articles to be included in a meta-analysis. Behavioural scientists recognize that these decisions are influenced by the meta-analyst in the same way selection decisions are influenced by qualitative reviewers. Like qualitative reviewers, meta-analysts should choose articles based on conceptual, methodological, and statistical grounds in order to reduce bias.

## Chapter III

### Method

This chapter describes the method used to conduct this research. It is divided into five sections: data gathering procedures, inclusion/exclusion criteria, data extraction, statistical procedures, and quality rating procedures.

#### Data Gathering Procedures

The first step of this meta-analysis was to collect the literature on couples psychotherapy outcome. The data were extracted from three areas: data bases, bibliographies of relevant articles and reviews of couples therapy literature, and current journals.

Data bases. For this meta-analysis PsychInfo, PsychLit, Medline, Dissertation Abstracts, and Educational Resources Information Center were searched. These data bases were searched from 1950 to the present. Although the data bases were searched from 1950, it should be noted that no articles published in the 1950s and 1960s met the inclusion criteria for this study.

Keywords searched were: marriage/marital therapy, marriage/marital counselling, psychotherapy effectiveness/impact, couples therapy/counselling, and psychotherapy outcome.

Bibliographies of outcome studies and reviews. First,

the bibliography of each article/dissertation/chapter collected was checked manually to find relevant literature. Next, bibliographies of relevant literature were cross-checked with material already collected to discover any missing research. This method was used each time a piece of relevant data was found.

Second, bibliographies of reviews and meta-analyses of couples therapy were checked to ensure that all outcome studies were assembled. Once again, articles were added, and the manual check of bibliographies commenced with the new articles.

Current journals. A manual, page by page search of all relevant journals was conducted monthly from August 1994 to March 1995. The journals included: Journal of Consulting and Clinical Psychology, Journal of Counselling and Development, Journal of Counseling Psychology, Journal of Marriage and the Family, Family Process, Psychotherapy Research, and American Journal of Family Therapy.

#### Inclusion/Exclusion Criteria

When all of the potentially relevant articles on couples therapy outcome were assembled, each article was reviewed to determine if it would meet the inclusion criteria of this research. This section describes how the data were selected for inclusion.

Criteria for inclusion. For the purpose of this



thesis, couples therapy was defined as conjoint therapy. This meant that there were always two persons in the room with one or more therapists. We chose conjoint therapy as criteria for inclusion because it involves both partners in the therapy process. As a result, researchers measure change that occurs for both partners. This information was essential to conducting a meta-analysis of therapy with couples.

A consequence of including only conjoint counselling, was the exclusion of group counselling with couples or family counselling. Although both group and family counselling may involve both partners in the process of therapy, I felt that couples who seek therapy alone may be different from couples who go to therapy with their family or with other couples. In addition, I felt that the treatment received by families or couples in groups may be different from couples therapy. As a result, I decided to examine the changes that occur in therapy when the couple is alone with the therapist.

Another type of counselling which was not included was individual counselling for couples. That is, any therapy done with individuals presenting a marital issue, was not included. Furthermore, if both the partners were not actively involved in the treatment process they were not included in this research.

Articles were included regardless of the theoretical orientation of the therapists conducting treatment. However, articles were not included when the treatment the couples received was defined explicitly as sex therapy (e.g., Masters and Johnson program). In addition, articles were not included if the focus of the investigation was preventive or premarital counselling/marriage preparation counselling.

Analogue and single-case studies. Neither analogue nor single-case studies were included in this research. Analogue studies were excluded because they are quite distinct from clinical research. Most important, the motivation and expectancies of analogue participants probably vary from participants who actively seek treatment. Analogue study participants are often solicited and given incentives for participation. As a result, it would be inappropriate to generalize the results of this type of research to the clinical setting (Kazdin, 1986). Single-case studies were also excluded. No articles were included if less than five couples were used in the study.

Both types of studies belong to a different design paradigm. That is, the results of these types of studies could not be fairly summarized under the same headings as clinical research.

Missing data. In cases where statistical data were

not included, then every effort was made to contact the author. If the author(s) could not be contacted, or did not comply with the request for the results of their study, those articles were not included in this review. It should be noted that only in very rare situations were data totally irretrievable.

### Data Extraction

A data base was created to compile the wealth of information afforded by the pieces of research gathered. Once the ESs were calculated for each article they were entered into the database along with the variables needed to answer the research questions. From this data base, ESs were copied and used in the aggregation and comparisons of ESs. The data base also included such categories as title of the article, author's name(s), year, and the purpose of each article gathered. A copy of this data base record sheet can be found in Appendix C.

### Effect Size Estimation

ESs are the statistic used to convert the results of diverse outcome studies into a common matrix. The preferred effect size estimate chosen for this research was Hedges's  $g$  (Hedges, 1981, 1982), a member of the  $d$  family of ES estimating statistics. The  $d$  family is used when the primary data consists mostly of means and standard deviations.

The ES estimators in the  $d$  family employ the standard deviation from the control group to calculate the denominator of the ES (Rosenthal, 1991). In contrast to the other members of this family, Hedges's  $g$  was chosen because it uses a pooled standard deviation. That is, the denominator used to calculate  $g$  is computed from both control and experimental groups. Therefore,  $g$  takes advantage of all of the available data.

As previously mentioned, the procedures used to calculate  $g$  depended on the original statistics reported in each article. Sometimes it was not possible to use  $g$ . For example, one of the three designs of the studies used in this research was the pre-post design which does not include a control group. As a result, it is not possible to pool the standard deviation of the experimental and control groups. In this case, we used Cohen's  $d$  (Cohen, 1988). Cohen's  $d$  was then easily converted to Hedges's  $g$ . The results reported in Chapter 4 refer to  $g$ s, unless otherwise stated. In addition, if there were no statistical data reported in an article but the results were described as non-significant, an ES of "0" was used.

#### Procedures for Data Analysis

Once an ES was calculated for each of the dependent measurements used in the research, all of these ESs were pooled to yield one overall ES. However, some studies

included follow-up data, and in these cases, the data were summarized in three ways: The first of these ESs was the average of the ES calculated for post-treatment change. A second ES for follow-up only, was calculated. Third, an ES was calculated by averaging both the post and follow-up ESs. The ESs reported in the results section refer to the post-measurement data unless otherwise specified.

In the event that research conducted on one data set was reported more than once, these ESs were combined to form a single ES. That is, when the same data were used by an investigator in more than one article or report, these data were weighted by sample size, averaged, and combined into one ES for the data set. For example, after the publication of a primary article (Jacobson, 1984), 1 year (Jacobson, Follette, Follette, Holtzworth-Munroe, Katt & Schmaling, 1985), and 2 year (Jacobson, Schmaling, & Holtzworth-Munroe, 1987) follow-up articles were published. The data from these two follow-up articles were averaged and included in the analysis as one data set.

Data were not averaged when both a dissertation and a publication reported the same findings. When this occurred, the relevant information was gathered from the dissertation because of the richness and detail of the material. All of the estimation methods can be found in Appendix D.

Conversion of raw data to ESs. To calculate Hedges's  $g$  (Rosenthal, 1994, p.232), the mean of the control group was subtracted from the mean of the experimental group, and divided by the pooled standard deviation of both experimental and control groups.

When the research presented did not include a control group, and only means and standard deviations were reported, the procedure employed for calculating  $g$  was subtracting the mean pre-score of the measurement from the mean post-score, and dividing by the pooled standard deviation (Rosenthal, 1994, p.232).

When calculating Hedges's  $g$  using between-groups  $t$ -tests (Rosenthal, 1994, p.239) assuming that  $n_1$  equals  $n_2$ , the  $t$ -value is converted to  $g$ . A different procedure was employed for calculating  $g$  from  $t$ -values (Rosenthal, 1994, p.238), when  $n_1$  was not equal to  $n_2$ .

In order to obtain  $g$  from research which used two-group ( $df = 1$ ), between-groups analysis of variance (ANOVA),  $F$ -statistics were converted according to Rosenthal's (1991a, p.238) recommendations.

For a two-group or more analysis of covariance (ANCOVA)  $F$ -statistics between groups, I first converted the  $F$ -statistics into Cohen's  $d$  (Rosenthal, 1994, p.240), and then converted the Cohen's  $d$  ES into Hedges's  $g$ , as

outlined by Shadish (1993, p.7).

When Pearson correlations ( $r$ ) were reported in the research, the conversion to  $g$  was computed using Rosenthal's procedure (1991a, p.66).

If only exact probability levels were reported, a corresponding  $t$ -statistic was found in a table, considering  $N - 2$  degrees of freedom. Next, this  $t$ -value was converted into  $g$  using the procedure described previously in this section.

If only inexact probabilities were reported, the procedure was identical to the one in the previous paragraph. That is, the level of significance for a two-tailed test, with  $N - 2$  degrees of freedom was found in a standardized table. Using this  $t$ -value,  $g$  was calculated.

Combining and contrasting ESs. As previously described, each piece of research yielded one overall ES. Prior to contrasting the ESs, each one was weighted first by sample size. This weighting produced an ES which was in proportion to the number of participants in the study. That is, because studies with a large number of participants have a smaller variance and more accurately estimate the true population mean, more weight is given to them. Conversely, a study which used a small number of

participants would have a larger variance and less accurately estimate the true population mean. To obtain the weighted  $g$ , the sum of each of the ESs were added and divided by the reciprocal of the estimate variance of  $g$ . This procedure was outlined by Rosenthal in 1991.

The next statistical procedure entailed using the overall ES from each piece of research to conduct the Hedges's adjustment for bias (Hedges, 1981, 1982). This procedure is performed because as more and more  $g$ s are sampled from the original population, the distribution of  $g$  becomes skewed (Rosenthal, 1994). Hedges's unbiased estimator  $g$  transforms the effect size estimator so that it is distributed nearly normally. All of the following computations were carried out with this unbiased estimate  $g$ .

Combining ESs (as recommended by Rosenthal, 1991, p.88), was done in order to yield an estimate of the overall efficacy of couples therapy. This procedure was conducted using weighted and adjusted  $g$ s. By giving more weight to the data sets with larger sample sizes (which are likely to have smaller error variance) meta-analysts have a better chance of accurately estimating the true population mean.

Prior to testing between-group differences (i.e., focused tests), a diffuse test, or test of heterogeneity



was conducted. This test determined the within-group variability of ES estimates. To the extent that the within group ES estimates are not significantly heterogeneous, the meta-analyst has greater confidence that the ESs are accurate estimates of the true population. The diffuse comparison uses chi square statistics to determine whether there are significant differences within groups of ES estimators. However, like the omnibus F-test, results indicate only whether there are differences, not where the differences lie.

Finally, focused tests, or "contrasts" of ES were conducted (Rosenthal, 1991, p.82). This procedure compared groups of studies by testing specific questions. For example, different theoretical orientations can be compared to see whether the results of therapy of one orientation show a greater amount of effect than another. The means of the two groups are compared to see whether the results differ (with 95% confidence) significantly. Each ES is assigned a lambda weight. The lambda weights always sum to zero. This value (lambda) is distributed approximately as Z. Next, the Z score and the associated probability level are found. If the null hypothesis is true, then the Z value will be approximately 0.

### Procedures for Quality Rating

In order to discover any differences in ES due to some aspects of research design, each article was evaluated by three graduate students. Each of the students was trained to use the rating scales by the researcher. The first part of the training consisted of a group meeting which lasted for two hours. During this meeting, the raters were taught to complete the rating scales and were given three articles to rate. These articles were not included in this research, and each student was given the same articles. One week later, the group reconvened to confirm their understanding of the rating scales. This meeting lasted for three hours.

The first of the two scales completed was the Methodological Validity Scale (MVS; Woszczyzna, 1995). It included 14 questions and was created by the researcher (see Appendix A). The second rating scale was the Smith, Glass and Miller Reactivity Scale (1980). This scale consisted of five questions (see Appendix B). For the purpose of this research, only 4 of the 5 questions were used. The question that was omitted was not relevant to this research. Each of the 18 questions was answered with either a "yes" or "no"; i.e., the raters scored each question with either a "1" to indicate "yes", or a "0" to indicate "no". When the question was unanswerable, the

rater scored the question .5.

When the ratings were completed for each article, the scores were compiled in a data base. Next, the researcher computed the intraclass correlations between the judges ratings (see Shrout and Fleiss, 1979) on each of the 18 questions. Intraclass correlations describe the reliability of two or more judges' scores on particular questions. Those questions which yielded a pre-designated reliability score on the intraclass correlation would be used for weighting the ESs.

The pre-designated reliability rating that was chosen was an intra-class of 0.41. Landis and Koch (1977), outlined acceptable ratings for intraclass correlations. They suggested that any intraclass correlation less than zero should be considered poor; 0 to .20 should be considered slight; .21 to .41 should be considered fair; .41 to .61 should be considered moderate; .61 to .81 should be considered substantial; and .81 and higher should be considered almost perfect. Consequently, if the ratings of the three judges yielded an intraclass correlation of .41 or greater, the intra-class would be considered moderately significant and used to weight the articles.

Next, the quality rating scores of each of the articles were summed. However, because one of the quality rating questions pertained only to data sets which included

a follow-up, the articles were split on this criterion. The articles which did not include a follow-up were summed separately without adding the question that pertained to the data sets with follow-ups.

The score for each data set in the post and follow-up groups were then converted to a standardized score with a mean of one and a standard deviation of one. This new standardized score was used for weighting each of the ESs by the methodological quality of the data set. Comparisons and contrasts were carried out using this new weighted  $g$ .

## Chapter IV

### Results

This chapter is organized in three sections. First, an overview of the demographic characteristics of the studies and participants are presented, followed by the results pertaining to each research question. Finally, the results of the quality rating will be examined.

#### Demographics

Characteristics of studies. Five hundred and seventy-three individual articles, books, theses, etc., were reviewed for this research. Of these articles, books, theses, etc., over 100 were retrieved. Of these approximately 100 pieces of research, 62 were chosen for inclusion in this study. Eleven of the 62 articles used non-independent data. Therefore, the results of these pieces of research were combined. Consequently, the original 62 pieces of research made up 51 data sets. The study characteristics can be found in Table 1.

Characteristics of participants. These 51 data sets included 3346 participants, (1,673 couples). The number of couples in the studies ranged from 20 to 318, with an average of 33 couples per study. Of the 51 data sets, 34 reported the age of the participants. The age of participants ranged from 26 to 46, with the average participant being 35 years old. In only 26 of the 51 data

sets was the length of the couples' relationship reported. The range of the relationship length was 5.9 to 17.2 years. The mean length of all of these relationships was 10 years.

Treatment. Within the 51 data sets, there were 56 "tests" of efficacy based on theoretical orientation: 30 of BCT, 8 of generic/pragmatic, 7 of systemic, 2 of psychodynamic and 9 of eclectic. Within the 51 data sets, on five occasions more than one theoretical orientation was tested in the same study.

Two hundred and seventy-three therapists (9 studies did not report the number of therapists) took part in the studies. There was a range from 1 to 29 therapists used in the research, with a mean of 6.5 therapists per study. In 40 of the 51 data sets, the therapists' level of clinical experience/education was reported. Based on the information provided in the primary research, the therapists were classified in one of five categories: a) undergraduate student (n = 2), b) Master's level student and/or 1 year of clinical experience (n = 5), c) M.A. or M.S.W. and/or 2 years of experience (n = 12), d) Doctoral candidate and/or 3-4 years of experience (n = 12), e) PhD or M.D. and/or 5+ years of experience (n = 9). Length of treatment, reported in 44 of the data sets, varied from 3 weeks to 21 weeks with a mean of 11 weeks.

## Efficacy

The 51 data sets yielded 234 self-report and behavioural ESs. There was an average of 4 ESs per study which reveals an attempt to assess the outcome of therapy in a multi-dimensional way by the researchers. As described in the previous chapter, all of the results were converted to the ES estimate  $g$ , with the exception of the few statistics that were converted to  $d$ . These ESs were averaged within studies and then combined. Next, the ESs were adjusted for bias (Hedges, 1982), and then weighted by sample size (Rosenthal, 1991) in order to give each of the studies a proportionate "vote".

The mean of all ESs at posttest was 0.73. The weighted overall ES was 0.69. The range of weighted ESs was from -0.52 to 2.0. Only three ESs fell in the negative range, showing the likelihood of an overall positive effect. In fact, an overall ES of 0.69 means that the average couple that received treatment moved from the 50th percentile to the 76th percentile posttreatment on outcome measures (Shadish, 1992). This interpretation comes from the idea that ESs can be described as standard scores that estimate the overlap between the distribution of the experimental group and the distribution of the control group. Reference to the normal distribution table shows that an effect of .69 standard deviations, implies that couples who received treatment moved from the 76th

percentile as compared to the average couples in the control group who remained at the 50th percentile.

In two of the three types of studies included in this thesis, couples were in an experimental condition being compared to either couples receiving another kind of treatment (treatment versus treatment studies), or a control group (treatment versus control studies). In the third group, couples were not compared with other couples (pre-post studies) but were compared against their own pre-treatment scores.

Twenty-five of the 51 data sets included data about a follow-up. The average ES at follow-up was 0.74, and the weighted follow-up ES was 0.68. There were no significant differences found between overall posttest and follow-up ESs ( $z = -0.01$ ,  $p = .50$ ).

#### Research Design and Treatment Moderators and Mediators

Diffuse comparisons. As previously described, groups of ESs were tested for heterogeneity within groups before ESs were compared between different groups. On all moderators and mediators of research design and treatment, groups of ESs were found to be significantly heterogeneous (see Table 2).

Study design. No significant differences were found ( $z = 0.006$ ,  $p = .50$ ) when a focused test was conducted on research using random versus nonrandom assignment. Average ESs were  $M = 0.74$  [ $n = 37$ ] for studies using random



assignment and for studies using non-random assignment,  $M = 0.70$  [ $n = 14$ ]. Their weighted individual ESs were 0.75 and 0.57, respectively.

Thirty-three data sets included a control group, 14 were pre-post designs and 4 were treatment versus treatment studies (see Figure 1). The experimental versus control group studies yielded an average effect size of 0.72 and a weighted ES of 0.66. The 14 pre-post data sets yielded an average ES of 0.80, and a weighted ES of 0.78. Finally, the 4 treatment versus treatment data sets yielded an average ES of 0.84, and a weighted ES of 0.90. When these study designs were compared to one another no significant differences were found (studies with control groups compared to treatment versus treatment designs,  $z = 0.005$ ,  $p = .50$ ; treatment versus treatment designs compared with pre-post design,  $z = 0.04$ ,  $p = .48$ ; studies with control groups compared with pre-post design studies,  $z = 0.016$ ,  $p = .49$ ).

Self-referred versus recruited participants. For the 30 data sets with self-referred participants, the average ES was 0.74, and the weighted ES was 0.70. Of the 18 data sets which used participants that were solicited, the average ES was 0.78 and the weighted ES was 0.74. When these two groups were compared, differences were nonsignificant ( $z = -0.01$ ,  $p = .50$ ).

Published versus unpublished research. Forty-one of

the data sets came from published articles or books. Their average ES was 0.75, and their weighted ES was 0.69. Ten of the data sets were from unpublished dissertations. They yielded an average ES of 0.61, and a weighted ES of 0.65. When these groups were contrasted, there was a non-significant finding of  $z = -0.03$ ,  $p = .49$ .

### Treatment Moderators and Mediators

Theoretical orientations. All of the different data sets based on theoretical orientation yielded heterogeneous ESs, with the exception of the psychodynamic data sets (see Table 2). However, the psychodynamic test of heterogeneity of ES was based on the results of only two studies. Average/non-weighted ESs by orientation were eclectic ( $n = 9$ ), 0.67; BCT ( $n = 30$ ), 0.78; systemic ( $n = 7$ ), 0.90; generic/pragmatic ( $n = 8$ ), 0.52; psychodynamic ( $n = 2$ ), 0.50. Once adjusted and weighted, ES by orientations were eclectic, 0.70; BCT, 0.75; systemic, 0.88; generic/pragmatic, 0.43; psychodynamic, 0.61 (see Figure 2). When the theoretical orientations were contrasted to one another no significant differences were found: BCT with eclectic,  $z = -0.023$ ,  $p = .49$ ; generic/pragmatic with systemic,  $z = 0.079$ ,  $p = .47$ ; systemic with psychodynamic,  $z = 0.033$ ,  $p = .49$ ; systemic with BCT,  $z = 0.006$ ,  $p = .50$ . In addition, there were no significant differences between posttest and follow-up ESs between theoretical orientations (systemic,  $z = 0.08$ ,  $p = .47$ ; generic/pragmatic,  $z = -0.17$ ,

$p = .43$ ; psychodynamic,  $z = 0.25$ ,  $p = .40$ ; eclectic,  $z = 0.03$ ,  $p = .49$ ; BCT,  $z = 0.04$ ,  $p = .48$ ).

Participant's age. As previously described, the mean ages of participants were described in 34 of the 51 data sets. The ages had a mean ES of 0.79. A median split was carried out (median = 35) and the two groups were compared. The results indicated no significant differences ( $z = -0.05$ ,  $p = .48$ ). In order to further test whether there may be differences based on age, the group was split into three separate groups. The participants from the two outlying groups ( $n = 22$ ) were compared. The results were not significant ( $z = -0.12$ ,  $p = .45$ ).

Length of the couples' relationship. The length of the relationship of the couple was divided and compared in the same manner as the age of the participants. The length of relationship of the couples was reported in 26 data sets and had a mean ES of 0.88. The median length was 10 years. Results showed that the two groups were not significantly different ( $z = 0.007$ ,  $p = .50$ ). When the groups were split into three ( $n = 18$ ), nonsignificant differences were found once again ( $z = -0.06$ ,  $p = .48$ ).

Length of treatment. Length of treatment was explored using the same method as age of participants, and length of relationship. Length of treatment was reported in 44 of the data sets, with a mean ES of 0.75. A median (10 weeks) split was conducted and results were found to be

nonsignificant ( $z = -0.02$ ,  $p = .49$ ). When the data were split into three and the two outlying groups were compared ( $n = 28$ ), nonsignificant differences were found ( $z = -0.02$ ,  $p = .49$ ).

Influence of the experience/education of the therapist. As previously described, the experience/education of therapist was divided into five sections. A comparison of the therapists in the least experienced group ( $M = 0.9$ ,  $n = 2$ ) with the most experienced group ( $M = 0.5$ ,  $n = 9$ ) yielded nonsignificant differences ( $z = 0.46$ ,  $p = .48$ ). Due to the small number of data sets in the least experienced group, a second comparison of groups 1 and 2 ( $M = 0.61$ ,  $n = 7$ ) with groups 4 and 5 ( $M = .60$ ,  $n = 21$ ) was conducted. This contrast also yielded nonsignificant differences ( $z = -0.07$ ,  $p = .47$ ).

#### Quality Rating

Intra-class correlations. Of the 18 questions (see Appendixes A and B) that were rated by the judges, 11 met the reliability cut-off of 0.41 (see Appendix E). The sums of these 11 questions were combined as outlined in Chapter 3 to weight the ESs by methodological quality.

Overall quality weighting. First, all of the scores were converted to a standard score with a mean of 1 and a standard deviation of 1. These scores were then used to weight the ESs of each data set. The overall quality

weighted ES was 0.69 ( $M = 0.72, n = 51$ ).

The questions tested with quality weighted Ess. The comparative tests conducted using the quality weighted ESs were: a) differences between published and unpublished research and b) differences between theoretical orientations. It should be noted that the other comparative questions (i.e., questions of methodological validity) asked in this research were not tested again with quality weighted ESs. The other questions were not tested because some elements of those research questions were embedded in the questions that made up the quality rating.

Diffuse tests with quality weighted Ess. Tests of heterogeneity of ESs were conducted prior to conducting focused tests. All tests of ESs proved to be nonsignificantly heterogeneous (see Table 3). It seems that their estimates of the size of the true population mean were consistent with one another, as they had very small standard deviations.

Focused tests with quality weighted ESs. When the theoretical orientations were contrasted to one another no significant differences were found: BCT with eclectic ( $M = 0.02, n = 30; M = 0.02, n = 9$ )  $z = 0.00, p = .50$ ; generic/pragmatic with systemic ( $M = 0.02, n = 8; M = 0.02, n = 7$ ),  $z = 0.00, p = .50$ ; systemic with psychodynamic ( $M = 0.02, n = 7; M = 0.01, n = 2$ ),  $z = 0.00, p = .50$ ; systemic with BCT ( $M = 0.02, n = 7; M = 0.02, n = 30$ ),  $z =$

0.00,  $p = .50$ .

The published versus unpublished contrast ( $M = 0.01$ ,  $n = 41$ ;  $M = 0.01$ ,  $n = 10$ ) also yielded nonsignificant differences ( $z = 0.00$ ,  $p = .50$ ).

## Chapter V

### Discussion & Conclusions

#### Summary of Findings

Efficacy at posttreatment. The purpose of this study was to discover whether or not couples therapy is effective and which variables influence the effectiveness of therapy. An overall ES of 0.69 was found. This ES found for couples therapy at posttreatment reveals that the average couple who seeks therapy improves from the 50th percentile (of all of the couples seeking therapy) to the 76th percentile.

Efficacy at follow-up. Twenty-five of the 51 data sets included a follow-up (M = 11 months). Their overall ES was 0.68. Therefore, couples who were tested at follow-up appear to move from the average 50th percentile to the 75th percentile compared to those couples who did not receive treatment. Consequently, we can say that couples in therapy appear to maintain posttest gains.

Research design moderators and mediators. In this study, no differences were found between experimental and quasi-experimental studies. In addition, no differences were found between studies that used a control group, studies which were pre-post test designs, or studies that used a treatment versus treatment comparison design. Furthermore, no differences were found between the outcomes of self-referred compared to recruited participants or between published and unpublished research.

Treatment moderators and mediators. In this meta-analysis, there were no significant differences in outcome found between theoretical orientations, age of participants, length of participants' relationships, length of treatment, or experience/education of therapists.

#### Comparison of These Findings with the Findings of Other Research

Efficacy at posttreatment. The overall efficacy found in this research is very similar to the results of the three meta-analyses already conducted on couples therapy. In 1988, Hahlweg and Markman conducted a cross-cultural comparison of couples therapy. They looked at the differences between BCT and pre-marital intervention programs from Europe and the United States. They found that the efficacy of the BCT articles yielded an ES of 0.95. Plattor (1990) found an ES of 0.85 when he looked at the effects of BCT and Insight-Oriented Couples Therapy (IOCT). Finally, Shadish et al. (1993) found an overall ES of 0.60 for BCT, systemic, humanistic, eclectic and "unclassified" therapies.

Efficacy at follow-up. At follow-up, the couples therapy research included in this study yielded an ES of 0.68. This number is more conservative, but comparable to the follow-up results found by Hahlweg and Markman (1988) and Plattor (1990). Hahlweg and Markman found an ES of 1.16, and Plattor found an ES of 0.73. As suggested by



Jacobson and Addis (1993), it is likely that with replication, this ES will become even more conservative.

Shadish et al. (1993) did not report a follow-up ES for couples therapy alone. However, like Hahlweg and Markman and Plattor (1990), Shadish et al. stated that there were no significant differences found between post and follow-up ESs.

### Research Design Moderators and Mediators

Study design. This was the first meta-analysis of couples therapy that considered primary research with quasi-experimental designs (i.e., without random assignment). Plattor (1990) used experimental groups versus no treatment/wait-list control groups, Hahlweg and Markman (1988) used studies with control groups only, and Shadish et al. (1993) included treatment versus control groups or treatment versus treatment groups with random assignment.

Published vs. unpublished research. Published research was compared with unpublished research in only one of the other meta-analyses conducted on couples therapy. Like the results of this study, Shadish et al. (1993) found no statistically significant differences between published and unpublished research. In addition, Shadish et al. claimed that combined couples and family therapy research that was published correlated with higher ESs than couples and family therapy research that was not published.

Self-referred vs. recruited participants. Shadish et al. (1993) tested for differences between the outcome of therapy with solicited versus self-referred participants. Comparable to the results of this study, no differences were found.

#### Treatment Moderators and Mediators

Theoretical orientation. The results of this study are the same as those found by Plattor (1990) and Shadish et al. (1993). In this study no differences were found between BCT, systemic and eclectic therapies. Plattor found no differences between BCT and psychodynamic therapy. Shadish et al. found no differences between BCT, humanistic, systemic, eclectic and unclassified therapies. Shadish and his colleagues also found that all of the therapies yielded ESs that differed significantly from zero except for the humanistic group of studies. No humanistic therapy research pieces met the inclusion requirements of this thesis and therefore cannot be compared.

Experience/education of therapists. Shadish et al. (1993) also considered the contributions of the therapist's experience/education to the outcome of therapy. Like the results of this study, they found no significant differences in outcome between the levels of experience of the therapists. However, Shadish et al. did suggest that there may be a correlation between greater experience/education of the therapist with higher ESs.

## Interpretation of Findings

Efficacy. From this and other reviews, it appears that couples are changing as a result of couples psychotherapy. Participants seem to be experiencing positive change both at the end of therapy and at follow-up. Later in this chapter, efficacy will be readdressed.

Research design and treatment. It is difficult to say why there were no significant differences between moderators and mediators of research design and treatment. I do not presume to have the definitive answer to this question. However, alone or in combination, the following possible explanations are worth considering.

First, it has been suggested (e.g., Smith et al., 1981) that no differences may be found due to experimenter allegiance/therapist bias. That is, therapists may influence the results of therapy in the direction of their hypotheses. Plattor (1990) suggested that as a result of experimenter allegiance, all results may "even out."

Second, the fact that no differences were found between theoretical orientations has been considered by Shadish et al. (1993). These authors have pointed out that it is hard to differentiate between theoretical orientations in the age of eclecticism and the integration of couples/family therapy approaches.

Third, non-specific psychotherapeutic processes may be the pith of each therapeutic experience. For example, the

therapeutic alliance has proven to be equally effective among all theoretical schools (Horvath & Symonds, 1991). It is possible that there may be something inherent in the working relationship and/or the process of therapy that dilutes the impact of moderating and mediating variables. Shadish et al. (1993) stated that "If all treatments were equally well designed, implemented, measured and reported, significant univariate differences among orientations may not be found" (p.999).

#### Limitations of This Study

Clinical significance. Clinical significance refers to a technique created by Jacobson, Follette, Revenstorf, Baucom, Hahlweg and Margolin (1984). Baucom and Hoffman(1986) adapted Jacobson et al's. technique for meta-analysis. This technique is used for comparing post-test results on inventories to normative data. For example, a commonly used inventory created to assess change in the couple's relationship is the Dyadic Adjustment Scale (Spanier, 1976). By examining pre-test and post-test means from the experimental and control condition groups, it is possible to see if couples moved from distressed to nondistressed levels on the inventory. If the participants who received therapy moved from the severely distressed level to the distressed level, then compared to distressed couples who did not receive treatment, they appear to be doing much better at the end of therapy (and follow-up),

however, they may remain distressed.

There appears to be controversy over whether or not clinical significance is different from efficacy as reported by ES strength. Hahlweg and Markman (1988) tested clinical significance (in nine studies) and found that many of the BCT couples were dissatisfied or in distress after therapy. However, Shadish et al. (1993), found no differences between clinical significance and ES. It seems that the question of whether clinical significance is the same as efficacy (as expressed by ES) is still unresolved. In this study, clinical significance was not tested. This issue will be readdressed in the next section.

Outcome measures. In this study, outcome measures were not divided into self-report and behavioural measures. Nor was there an examination of the differences between self-report and "other" report measures. There is controversy in the current literature over which measures yield more conservative results. Shadish et al. (1993), found that self-report measures yielded lower ES ratings than observer ratings. However, Hahlweg and Markman (1988) found that there were no differences between self-report and observational measures.

#### Recommendations for Future Research

In order to assess what needs to be done in future research, I believe we must reflect on the work that has been done to date. Although the results of psychotherapy

synthesis research appear to be reliable, we must ask if it is also valid. By this I mean that I believe it is essential that we concern ourselves with whether our indices of psychotherapeutic "success" or "failure" of couples therapy reflect the changes made in the couple's relationship. The next section includes some personal and "other" suggestions for answering the question, How can we do a better job of finding how much, and how, participants are changing? The next section is broken down into sections on efficacy and moderators and mediators of research design and treatment.

### Efficacy

Clinical significance. It has been argued that compared to couples not in therapy, distressed couples may still end up in a relatively distressed state. For example, Jacobson and Addis (1993) report that only about 50% of couples are happily married at the end of treatment. In other words, just because participants show change, does not mean couples are happy, but rather, it may mean that couples are only less unhappy. If this poor prognosis is the result of what is happening post-therapy, then we must ask if couples are experiencing a meaningful change, and if therapists are doing their job effectively.

I believe that it may prove useful to examine the change experienced in couples who receive treatment on a couple to couple basis. Perhaps the clinical significance

technique would yield meaningful information if it were conducted for each couple in primary research as well as for each meta-analysis conducted.

Clinical significance is different from other outcome measurements such as ES. It is important to ask what ES and/or clinical significance truly mean. Garfield (1983, p.41) eloquently states "I find it hard to translate a .85 sigma effect size into a meaningful appraisal of change in terms of clinical criteria. For example, does this represent 1 point on a 5-point rating scale of change? ...Does it represent a significant change in behavior? Or what?" Perhaps using the clinical significance technique teamed with a definition of the couple's level of distress should be considered.

Many researchers (e.g., Plattor, 1990; Jacobson & Addis, 1993) have suggested creating a type of classification like those inherent in the "medical model." Raffa, Sypek and Vogel (1990) argue that without a system of diagnosis like that offered by the Diagnostic and Statistical Manual of the American Psychiatric Association (APA, 1987), it is impossible to discuss efficacy of couples therapy. They state that in order to create meaningful definitions of change for the couple, specific definitions of distress must be created. Once this type of classification is created, they believe that it will be easier to start matching interventions/approaches as well

as appropriate control groups with the specific problems that couples face.

Separation/divorce. To date, separation and/or divorce has been considered a liability of couples counselling. I would like to suggest that for some participants, separation/divorce should not be seen as a failure of therapy, but rather a success of therapy. It is likely that there are couples seeking therapy in order to find mediation for their separation/divorce. If this is indeed true, then there must be a way to consider separation/divorce in a way that does not inappropriately deflate the results of the study's efficacy.

Perhaps researchers can test ways to measure whether positive change occurs for individuals at post-treatment/follow-up regardless of marital status.

Individual change. Occasionally, the results of quantitative studies list only individual partner's results, without a combined result for "relationship" differences. More often, the results of quantitative studies on couples are reported in terms of a single index of the partners' combined scores. It would be helpful to find a way to look at the changes for the individuals. Jacobson and Addis (1993) suggest that outcome studies report changes in both the couple and individuals. They believe that this is important because they have seen women enter therapy with the agenda of finding positive change.



However, they believe that men often enter therapy with the agenda of maintaining the status quo of the relationship. As a result, it is important to remember that positive change for one partner may not be positive change for the other.

Follow-up. It has been suggested by Jacobson, Schmaling and Holtzworth-Munroe (1987) as well as Grady and Fletcher (1991) that a significant amount of deterioration in the couples' relationship may occur by one year post-therapy. There is very little research about how long the effects of therapy last. The longest follow-up to date is only 4 to 5 years long, and only a very small fraction of studies include follow-ups. It seems that if researchers and clinicians are going to claim that therapy is efficacious, it would increase the soundness of this claim if they could also say for how long the effectiveness of therapy may last. However, this and other research supports that approximately 1 year posttreatment (the mean of the follow-up of this study was 11 months) the efficacy of therapy is not diminished.

Prevention. One method of couples therapy which is proving to be efficacious is preventive and pre-marital counselling. In fact, Jacobson and Addis (1993) suggest that it may be easier to prevent relationship problems than to treat problems once they materialise. For example, relationship satisfaction commonly declines when the female

partner first becomes pregnant. Jacobson and Addis (1993) talk about doing work with "pregnant couples" previous to the birth of the child. Perhaps identifying other times when couples are likely to experience dissatisfaction could prevent or aid in the future of the couple's relationship.

#### Research Design and Treatment Moderators and Mediators

Study design. Unlike the other couples therapy meta-analyses, this study found that when quasi-experimental research was included, there were no differences in ESs. Some methodologists (e.g., Hedges, 1983; Rosenthal, 1991) maintain that the inclusion of quasi-experimental research means that inferences cannot be made because participants were not randomly assigned to treatment and control conditions. Still, others (e.g., Greenwald & Russell, 1991; Raffa, Sypek and Vogel, 1990) disagree; they maintain that any inclusion/exclusion criteria does not necessarily indicate biases in the research. Clearly, this is a controversy that is yet to be resolved through continued research.

Jacobson and Addis (1993) recommend using an intra-treatment model. One example of this model is, instead of conducting treatment versus treatment research, they suggest comparing and contrasting different treatments which come from the same theoretical orientation. As a result, they believe that researchers will be able to learn what components, or active ingredients influence treatment.

Another intramodel comparison method they recommend is a "constructive treatment" design. This method allows the researcher to selectively add new components to a treatment in order to discover whether or not the new component enhances the original treatment. It is likely that the use of these methods will enhance not only our understanding of efficacy in therapy, but also the processes that contribute to outcome.

Qualitative research. It seems that qualitative research would be an excellent direction to take in order to discover more about how couples experience change. Jacobson and Addis (1993) state that process variables can best be discovered by looking at the couple interaction processes. They also purport that in BCT, client effort in therapy and out-of-session correlate most with success in therapy.

It is possible that process variables could be uncovered if qualitative interviewing was done both with clients who "succeed" and with clients who "fail." It is important to find out why participants do not succeed. We should ask questions such as with whom, when, and why does therapy fail? Conversely, it would be helpful to discover which treatments do succeed, why, and with which presenting issues. Perhaps more naturalistic rather than experimental studies would provide interesting information.

Theoretical orientation. There appear to be two

areas of research that need to be considered when theoretical orientation is examined. First, there is a need for more replication of studies from the psychodynamic and humanistic schools of therapy. Shadish et al. (1993) found that studies of humanistic couples therapy did not significantly differ from zero. There were no studies that used the humanistic approach to couples therapy that met the criteria for this study, and to date, no one else has considered the humanistic approach in their couples meta-analysis.

Second, Gurman, Kniskern and Pinsof (1986) claim that there is failure on the part of authors to describe the techniques used in therapy in detail. Gurman, Kniskern and Pinsof further claim that therapists may also use techniques which are not associated with the theoretical framework that they represent.

Participant and therapist information. If researchers collected thorough data on the therapist(s) and the couple, it will make the job of the meta-analyst much easier. For example, by looking at changes in men and women separately, it will be possible to see if changes occur differently for men and women as a result of such factors as specific interventions, the sex of the therapist, characteristics of the therapist, etc.

There is also little information about counselling gay and lesbian couples. In fact, only one article (Olson,

Russell, & Sprenkle, 1980) has addressed therapy with gay and lesbian couples.

Other variables such as length of relationship, length of distress in the relationship, age of the couples, severity of the distress and cultural issues/differences in the relationship, often go unreported. Once this information is collected, not only will the meta-analyst be able to discover more about efficacy, but she or he will also have the added information necessary to learn more about moderators and mediators of treatment.

### Conclusions

One aim of this study was to discover whether couples therapy works. Results indicated that couples therapy is, on the whole, effective. The other purpose of this study was to investigate which aspects of couples therapy influence outcome. In other words, I sought to investigate which variables made a difference. Given the results of this study, it seems that a more fruitful avenue of investigation would be to look for similarities instead of differences.

How does couples therapy create change? What universal/common factors are the threads that weave psychotherapy into an effective and meaningful tapestry? The following points are only a few of many possible reasons why therapy produces effective results.

Therapy allows people/couples to be heard--people have

the opportunity to self-reflect, often without judgment, in a place where they may feel validated, and/or see their experiences as "normal". Together with the therapist, the clients create new meanings through discoveries ("uh-huhs") about their beliefs, experiences, feelings. Clients may try on/express/experience behaviours, emotions, and cognitions. The couple has the opportunity to learn new ways to make sense of their relationship.

Couples may experience positive change in therapy because of the commitment they bring with them into therapy. Perhaps the commitment is a result of the monetary investment they intend to make. It is also possible that their commitment to therapy comes at a time when the couple has exhausted all other resources. Couples may initiate counselling as a last resort. It is possible that couples may be committed to make changes in their relationship because they have children and fear the break-up of the family.

Whatever the reason, it seems likely that couples who enter counselling are motivated to make changes in their relationship. As a result, it seems that regardless of who administers therapy, what theoretical paradigm governs the treatment, what method or design or treatment therapy takes, couples experience positive change.

## References

- Baucom, D. H., & Hoffman, J. A. (1986). The effectiveness of marital therapy: Current status and application to the clinical setting. In N. Jacobson & A. Gurman (Eds.). Clinical Handbook of Marital Therapy (pp. 597-620). New York: Guilford Press.
- Beach, S. R. H., & O'Leary, K. D. (1985). Current status of outcome research in marital therapy. In L. L'Abate (Ed). The Handbook of Family Psychology and Therapy, Volume II (pp. 1035-1072). Homewood, Ill: The Dorsey Press.
- Beck, D.F. (1975). Research findings on the outcomes of marital counseling. Social Casework, 153-181.
- Becvar, D.S. & Becvar, R.J. (1993). Family Therapy: A Systemic Integration (2nd Ed). Boston: Allyn and Bacon.
- Bergin, A. E. (1971). The evaluation of therapeutic outcomes. In A. E. Bergin & S. L. Garfield (Eds.) Handbook of Psychology and Behaviour Change (pp. 217-270). New York: Wiley.
- Bergin, A. E. & Lambert, M. J. (1978). The evaluation of therapeutic outcomes. (pp. 139-189). In S. L. Garfield & A. E. Bergin (Eds.). Handbook of Psychotherapy and Behavior Change: An Empirical Analysis (2nd ed.). New York: John Wiley & Sons.
- Carr, A. (1991). Milan systemic family therapy: A review

- of ten empirical investigations. Journal of Family Therapy, 13, 237-263.
- Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). New York: Academic Press.
- Cook, T. D., & Leviton, L. C. (1980). Reviewing the literature: A comparison of traditional methods with meta-analysis. Journal of Personality, 48(4), 449-472.
- Cooper, H. & Hedges, L. V. (Eds.) (1994). The Handbook of Research Synthesis. New York: Russell Sage Foundation.
- Crits-Christoph, P., Baranackie, K., Kurcias, J. S., Beck, A. T., Carroll, K., Perry, K., Luborsky, L., McLellan, A. T., Woody, G. E., Thompson, L., Gallagher, D., & Zitrin, C. (1991). Meta-analysis of therapist effects in psychotherapy outcome studies. Psychotherapy Research, 1(2), 81-91.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297-334.
- Devine, & Cook (1983).
- Eysenck, H.J. (1952). The effects of psychotherapy: An evaluation. Journal of Consulting Psychology, 16, 319-324.
- Eysenck, H. J. (1978). An exercise in megasilliness. American Psychologist, 33, 517.



- Foley, V. D. (1989). Family therapy. In R. J. Corsini & D. Wedding (Eds.). Current Psychotherapies (4th Ed.). (pp. 455-500). Itasca, Ill: F.E. Peacock Publishing, Inc.
- Garfield, S. L. & Bergin, A. E. (Eds) (1986) (3rd Edition) Handbook of Psychology and Behaviour Change. (pp. 3-22). New York: John Wiley and Sons.
- Garfield, S. L., & Lambert, M. J. (1978). In S. L. Garfield & A. E. Bergin (Eds.) (2nd Edition). Handbook of Psychotherapy and Behavior Change: An Empirical Analysis (pp. 171-179). New York: John Wiley & Sons.
- Garfield, S. L. (1983). Effectiveness of psychotherapy: The perennial controversy. Professional Psychology: Research and Practice, 14(1), 35-43.
- Garfield, S.L. & Bergin, A.E. (1978) (Eds.). Handbook of Psychotherapy and Behavior Change: An Empirical Analysis (2nd Ed.). New York: John Wiley & Sons.
- Greenwald, S., & Russell, R.L. (1991). Assessing rationales for inclusiveness in meta-analytic samples. Psychotherapy Research, 1(1), 17-24.
- Gurman, A. S. (1973). The effects and effectiveness of marital therapy: A review of outcome research. Family Process, 145-170.
- Gurman, A. S., & Kriskern, D. P. (1978). Research on marital and family therapy: Progress, perspective and prospect. In S. L. Garfield & A. E. Bergin (Eds.),

- Handbook of Psychological Change: An Empirical Analysis (2nd Edition) (pp. 817-901). New York: Wiley.
- Gurman, A.S. (1979). Dimensions of marital therapy: A comparative analysis. Journal of Marital and Family Therapy, 5-15.
- Gurman, A. S., Kniskern, D. P. & Pinsof, W. M. (1986). Research on the process and outcome of marital and family therapy, (pp.565-624.) In S. L. Garfield & A. E. Bergin (Eds.). Handbook of Psychotherapy and Behavior Change (3rd ed.). New York: John Wiley & Sons.
- Hahlweg, K., & Markman, H. J. (1988). Effectiveness of behavioral marital therapy: Empirical status of behavioral techniques in preventing and alleviating marital distress. Journal of Consulting and Clinical Psychology, 56(3), 440-447.
- Hazelrigg, M. D., Cooper, H. M., & Borduin, C. M. (1987). Evaluating the effectiveness of family therapies: An integrative review and analysis. Psychological Bulletin, 101(3), 428-442.
- Hedges, L., & Olkin, I. (1985). Statistical methods for meta-analysis. Orlando, FL: Academic Press.
- Holmes, T. H. & Rahe, R. H. (1967). The social readjustment rating scale. Journal of Psychosomatic Research, 11, 123-128.
- Holtzworth-Munroe, A., Jacobson, N. S., DeKlyen, M., &

- Whisman, M. A. (1989). Relationship between behavioral marital therapy outcome and process variables. Journal of Consulting and Clinical Psychology, 57(5), 658-662.
- Horvath, A. O., & Symonds, E. D. (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. Journal of Consulting and Clinical Psychology, 38, 139-149.
- Jacobson, N. S., & Addis, M. E. (1993). Research on couples and couple therapy: What do we know? Where are we going? Journal of Consulting and Clinical Psychology, 61(1), 85-93.
- Jacobson, N. S., Follette, W. C., Revenstorff, D., Baucom, D. H., Hahlweg, K., & Margolin, G. (1984). Variability in outcome and clinical significance of behavioral marital therapy: A reanalysis of outcome data. Journal of Consulting and Clinical Psychology, 52, 497-504.
- Jacobson, N. S. & Martin, B. (1976). Behavioral marriage therapy: Current status. Psychological Bulletin, 83(4), 540-556.
- Kazdin, A. E. (1986) In S. L. Garfield, and A. E. Bergin (eds) (3rd edition). Handbook of Psychotherapy and Behavior Change. (pp.23-68). New York, NY: John Wiley & Sons.
- Lambert, M. J. (1991). Introduction to psychotherapy

- research. In L. E. Beutler and M. Crago (Eds.).  
Psychotherapy Research: An International Review of  
Programmatic Studies (pp. 1-23). Washington, DC:  
American Psychological Association.
- Landis, J. R. & Koch, G. G. (1977). The measurement of  
observer agreement for categorical data. Biometrics,  
33, 1599-1674.
- Luborsky, L., Singer, B., & Luborsky, L. (1975).  
Comparative studies of psychotherapies: Is it true  
that "everyone has won and all must have prizes"?  
Archives of General Psychiatry, 32, 995-1008.
- Nicols, M. P. & Schwartz, R. C. (Eds) 2nd Edition (1991).  
Family Therapy: Concepts and Methods. Needham  
Heights, MA: Simon & Schuster.
- O'Leary, K. D. & Turkewitz, H. (1978). Methodological  
errors in marital and child treatment research.  
Journal of Consulting and Clinical Psychology, 46,  
747-758.
- Olson, D. H., Russell, C. S., & Sprenkle, D. H. (1980).  
Marital and family therapy: A decade review. Journal  
of Marriage and the Family, 42, 973-993.
- Plattor, J. R. (1990). Evaluating the Effectiveness of  
Marital Therapies: A Meta-Analytic Investigation.  
Dissertation Abstracts.
- Rachman, S. (1971). The Effects of Psychotherapy. Oxford:  
Pergamon Press.

- Rachman, S. J., & Wilson, G. T. (1980). The Effects of Psychological Therapy (Second Enlarged Edition). New York: Pergamon.
- Raffa, H., Sypek, J., & Vogel, W. (1990). Commentary on reviews of "outcome" studies of family and marital psychotherapy. Contemporary Family Therapy, 12(1), 65-73.
- Reevy, W. R. (1967) In H. L. Silverman (Ed.). Marital Counseling: Psychology, Ideology, Science (pp. 5-21). Springfield, ILL: Charles C Thomas Publisher.
- Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. Psychological Bulletin, 85, 185-193.
- Rosenthal, R. (1987). Judgment Studies: Design, Analysis and Meta-Analysis. Cambridge, England: Cambridge University Press.
- Rosenthal, R. (1991a). Meta-analytic Procedures for Social Research (Revised Edition). Applied Social Research Methods Series, Volume 6. Newbury Park, CA: SAGE Publishing, Inc.
- Rosenthal, R. (1991b). Quality-weighting of studies in meta-analytic research. Psychotherapy Research, 1(1), 25-28.
- Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper and L. V. Hedges (Eds.). The Handbook of Research Synthesis (pp.231-244). New York: Russell

Sage Foundation.

- Shadish, W. R. Jr. (1992). Do family and marital psychotherapies change what people do? In T. D. Cook, H. Cooper, D. S. Cordray, H. Hartmann, L. V. Hedges, R. J. Light, T. A. Louis, & F. Mosteller (Eds.). Meta-Analysis for Explanation: A Casebook (pp. 129-208). New York: Russell Sage Foundation.
- Shadish, W. R., Jr., Doherty, M., & Montgomery, L. M. (1989). How many studies are in the file drawer? An estimate from the family/marital psychotherapy literature. Clinical Psychology Review, 2, 589-603.
- Shadish, W. R. Jr., Montgomery, L. M., Wilson, P., Wilson, M. R., Bright, I., & Okwumabua, T. (1993). Effects of family and marital psychotherapies: A meta-analysis. Journal of Consulting and Clinical Psychology, 61(6), 992-1002.
- Shadish, W. R. Jr., & Sweeney, R. B. (1991). Mediators and moderators in meta-analysis: There's a reason we don't let Dodo Birds tell us which psychotherapies should have prizes. Journal of Consulting and Clinical Psychology, 59(6), 883-893.
- Shapiro, D. A., & Shapiro, D. (1977). The 'double standard' in evaluation of psychotherapies. Bulletin of the British Psychological Society, 30, 209-210.
- Shapiro, D. A. & Shapiro, D. (1982). Meta-analysis of comparative therapy outcome research: A critical

- appraisal. Behaviour Psychotherapy, 10, 4-25.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 86(2), 420-428.
- Smith, M.L., & Glass, G.V. (1977). Meta-analysis of psychotherapy outcome studies. American Psychologist, 32, 752-760.
- Smith, M. L., Glass, G. V., & Miller, T. I. (1980). The Benefits of Psychotherapy. Baltimore: John Hopkins University Press.
- Spanier, G. B. (1976). Measuring dyadic adjustment: New scales for assessing the quality of marriage and similar dyads. Journal of Marriage and the Family, 38, 15-28.
- Statistics Canada (June, 1994). Ministry of Industry, Science and Technology, Cat 84-213 Annual.
- Strube, M. J., & Hartmann, D. P. (1983). Meta-analysis: Techniques, applications, and functions. Journal of Consulting and Clinical Psychology, 51(1), 14-27.
- Whisman, M. A., Jacobson, N. S., Fruzzetti, A. E., & Waltz, J. A. (1989). Methodological issues in marital therapy. Advances in Behavioral Research and Therapy, 11, 175-189.
- Whiston, S. C. & Sexton, T. L. (1993). An overview of psychotherapy outcome research: Implications for practice. Professional Psychology: Research and

Practice, 24(1), 43-51.

Wilson, G. T., & Rachman, S. J. (1983). Meta-analysis and the evaluation of psychotherapy outcome limitations and liabilities. Journal of Consulting and Clinical Psychology, 51(1), 54-64.

Woszczyzna, C. B. (1995). The Methodological Validity Scale (MVS). Unpublished rating scale.



## Appendix A

## The Methodological Validity Scale (Woszczyzna, 1995)

1. The study used a comparison group.
2. The principle investigator did not conduct the treatment.
3. There was diagnostic homogeneity amongst the subjects (i.e., tested BEFORE admittance to study).
4. There was a pre-treatment assessment.
5. If there is a follow-up measurement, there is evidence that there was no further treatment after therapy ended.
6. Appropriate statistical analyses were used.
7. There is evidence that the therapists are competent with this method of therapy (e.g., were trained for this study, have M.A. level of education, or 1 year or more practice with this method).
8. Treatments were equally valued by therapists.
9. Treatment was controlled for quality/consistency (e.g., manualized).
10. Subjects were randomly assigned/or there was an attempt to match subjects/or a statistical demonstration indicated that there was no differences between groups.
11. Therapists were blind to the hypotheses of the study.
12. 50%+ or all clients were self-referred/other referred (not solicited).
13. 50%+ or all clients were not from a University setting.

14. Give "1" if randomization occurred, and overall attrition was less than 15%, and the difference in attrition between conditions was less than 10%, OR, an appropriately strong within-subjects design was used (either an interrupted time series with at least 30 pre- and post-intervention data points for both a treatment and a control group, or an aggregation of results from several identical ABAB type designs that used multiple baselines and reversals).

## Appendix B

## Smith, Glass &amp; Miller Reactivity Scale (1980)

1. The study included blinded ratings and decisions -- blind projective test ratings, blind ratings of symptoms, blind discharge from hospital.
2. The study used standardized measures of traits having minimal connection with treatment or therapist (MMPI, Rotter I-C, DAS, Locke-Wallace MAS,...).
3. The study did not include experimenter-constructed inventories (nonblind), ratings of symptoms (nonblind), any client self-report to experimenter, blind administration of client self-report to experimenter, blind administration of Behavioural Approach Tests.
4. The study did not include therapist rating of improvement or symptoms, projective tests (nonblind), behaviour in the presence of therapist or nonblind evaluator (e.g., behaviour approach test), instruments that have a direct and obvious relationship with treatment (e.g., where desensitization hierarchy items were taken directly from measuring instrument).

Appendix C  
Data Base Record Sheet

<b>Set#</b>	<b>Type</b>	<b>Related Sets</b>	
<b>Author</b>			
<b>Title</b>			
<b>Publisher</b>			
<b>Date</b>	<b>Vol#</b>	<b>Page#</b>	<b>Source</b> <input type="text"/>
<b>Purpose</b>			
<b>Client type</b>	<input type="text"/>		
<b>Treatment</b>			
<b>Treatment Defn's</b>			
<b>Problems</b>			
<b>Control</b>	<b># of Couples</b>	<b>Group assignment</b>	<input type="text"/>
<b>Treatment fidelity</b>	<b>Therapist (M/F)</b>	<b>No.</b>	<b>Age</b>
<b>Experience</b>			
<b>Student (Y/N)</b>	<b>Training</b>		
<input type="text"/>			
<b>Length of treatment</b>	<b># of Sessions</b>		
<b>Design Type</b>	<b>Outcome</b>	<input type="text"/>	<b>Process</b> <input type="text"/>
<b>Features</b>			
<b>Outcome Measures</b>			
<b>Process measures</b>			

## Appendix D

## Statistical Procedures

To calculate Hedges's  $g$  (1982) from means and standard deviations, the mean of the control group was subtracted from the mean of the experimental group, and divided by the pooled standard deviation of both experimental and control groups (Rosenthal, 1994, p.232):

$$\frac{M_e - M_c}{S_p}$$

When the research presented did not include a control group, and only means and standard deviations were reported, the procedure employed for calculating  $g$  was subtracting the mean pre-score of the measurement from the mean post-score, and dividing by the pooled standard deviation (Rosenthal, 1994, p.232):

$$\frac{M_2 - M_1}{S_p}$$

When calculating Hedges's  $g$  using between-groups  $t$ -tests assuming that  $n_1$  equals  $n_2$ , the  $t$ -value is converted to  $g$  using the following procedure, as outlined by Rosenthal (1994, p.239):

$$\frac{2t}{(N)}$$

When calculating  $g$  from  $t$ -values (Rosenthal, 1994, p.238), when  $n_1$  is not equal to  $n_2$ , the procedure used was:

In order to obtain the  $g$  from research which used two-

group (when  $df = 1$ ), between-groups analysis of variance (ANOVA),  $F$ -statistics were converted using the following formula (Rosenthal, 1991a, p.238):

$$\sqrt{F(1/n_1 + 1/n_2)}$$

For a two-group or more analysis of covariance (ANCOVA)  $F$ -statistics between groups, this formula first required converting the  $F$ -statistics into the ES Cohen's  $d$ , and then converting the Cohen's  $d$  ES into Hedges's  $g$ , as outlined by Shadish (1992, p.7):

For ANCOVA to Cohen's  $d$ :

$$\sqrt{F(1/n_E + 1/n_C)(1-r_{xy}^2)}$$

To convert Cohen's  $d$  to Hedges's  $g$  (Rosenthal, 1994, p.240):

$$\frac{d}{\sqrt{\frac{n_1 + n_2}{df}}}$$

When correlations were reported in the research, the conversion to  $g$  was conducted using the following procedure (Rosenthal, 1991a, p.66):

$$\frac{r}{\sqrt{1-r^2}} \times \sqrt{\frac{df(n_1 + n_2)}{n_1 n_2}}$$

If only exact probability levels were reported in the research, a corresponding  $t$ -statistic was found in a standardized table, considering  $N - 2$  degrees of freedom. Next, this  $t$ -value was converted into  $g$  using the procedure described previously in this section.

If only inexact probabilities were reported in the

research, the procedure was identical as the one in the previous paragraph. That is, the level of significance for a two-tailed test, with  $N - 2$  degrees of freedom was found in a standardized table. Using this  $t$ -value,  $g$  was calculated.

As previously described, each piece of research yielded one overall ES. This ES was the average of all of the post-measurement effects.

The next statistical procedure entailed using the overall ES from each piece of research to conduct the Hedges's adjustment for bias. The Hedges's adjustment was conducted using the following procedure (Rosenthal, 1994, p.240):

$$(g) \left( 1 - \frac{3}{4(df) - 1} \right)$$

To combine ESs, each data point was weighted first by sample size. To obtain the weighted  $g$ , the sum of each of the ESs were added and divided by the reciprocal of the estimate variance of  $g$ . This procedure was outlined by Rosenthal in 1991, pp.75-88:

$$\frac{E(w)(g)}{Ew}$$

The next test was conducted to see if the ESs were significantly heterogeneous. The first step of the procedure is to compute the estimated variance of  $g$  ( $1/w$ ), as described by Rosenthal (1991). For  $w$  (Rosenthal, 1991a,

p.75), assuming that  $n_1$  and  $n_2$  are equal,  $t$  was obtained first:

$$\frac{(g)\sqrt{N}}{2}$$

to obtain  $w$ :

$$\frac{N(N - 2)}{2(t^2 + 2N - 4)}$$

to obtain the weighted mean  $g$  (as above):

$$\frac{E(w)(g)}{Ew}$$

to compute  $\chi^2$ :

$$Ew_j(g_j - g)^2$$

Once the chi square is computed, then the df ( $K - 1$ ), were calculated, and the associated probability is found in the chi square standard table. If the probability level is significant, then the ESs are significantly heterogeneous.

The final statistical procedure carried out was focused tests, or "contrasts" of ES. The first step of this procedure was to calculate the adjusted  $g$ , known as the estimated variance of  $g$  ( $1/w$ ) for each of the studies (see above). The second step was to use the  $w$ 's to test the significance of each of the contrasts, or predictions we made. The letter " $\lambda$ " represents the lambda symbol in the following equation as recommended by Rosenthal, 1991, p.82:



$$\frac{E l_j a_j}{E w_j}$$

## Appendix E

## Intraclass Correlation Scores

Scale Correlation	Question	Intraclass
Methodological Validity Scale (Woszczyzna, 1995)	1	<b>0.92*</b>
	2	<b>0.65*</b>
	3	0.31
	4	0.40
	5	<b>0.50*</b>
	6	0.40
	7	0.29
	8	0.36
	9	<b>0.44*</b>
	10	<b>0.48*</b>
	11	<b>0.49*</b>
	12	<b>0.57*</b>
	13	0.29
	14	<b>0.56*</b>
Smith, Glass & Miller Reactivity Scale (1980)	1	0.17
	2	<b>0.81*</b>
	3	<b>0.51*</b>
	4	0.35

\*These 11 questions were included in the quality rating.

## Appendix F

## References of Articles Used In This Study

- Azrin, N. H., Naster, B. J., & Jones, R. (1979).  
Reciprocity counseling: A rapid learning-based  
procedure for marital counseling. Behavior Research  
and Therapy, 11, 365-382.
- Azrin, N. H., Besalel, V. A., Bechtel, R., Michalicek, A.,  
Mancera, M., Carroll, D., Shuford, D., & Cox, J.  
(1980). Comparison of reciprocity and  
discussion-type counseling for marital problems. The  
American Journal of Family Therapy, 8(4), 1-28.
- Baucom, D. H., Sayers, S. L., & Sher, T. G. (1990).  
Supplementing behavior marital therapy with cognitive  
restructuring and emotional expressiveness training:  
An outcome investigation. Journal of Consulting and  
Clinical Psychology, 58(5), 636-645.
- Baucom, D., & Lester, G. W. (1986). The usefulness of  
cognitive restructuring as an adjunct to behavioral  
marital therapy. Behavior Therapy, 17, 385-403.
- Baucom, D. (1982). A comparison of behavioral contracting  
and problem-solving/communications training in  
behavioral marital therapy. Behavior Therapy, 13,  
162-174.
- Bennun, I. (1985). Prediction and responsiveness in  
behavioural marital therapy. Behavioural

- Psychotherapy, 13, 186-201.
- Beutler, L. E. (1971). Attitude similarity in marital therapy. Journal of Consulting and Clinical Psychology, 37(2), 298-301.
- Boelens, W., Emmelkamp, P., MacGillavry, D., & Markvoort, M. (1980). A clinical evaluation of marital treatment: Reciprocity counseling vs. system-theoretic counseling. European Journal of Behavioural Analysis and Modification, 4(2), 85-96.
- Cline, V. B., Jackson, S. L., Klein, N., Mejia, J., & Turner, C. (1987). Marital therapy outcome measured by therapist, client, and behavior change. Family Process, 26, 255-267.
- Cookerly, J. (1980). Does marital therapy do any lasting good? Journal of Marital and Family Therapy, 6, 393-397.
- Crowe, M. J. (1978). Conjoint marital therapy: A controlled outcome study. Psychological Medicine, 8, 623-636.
- Dadds, M. R., Schwartz, S., & Sanders, M. R. (1987). Marital discord and treatment outcome in behavioral treatment of child conduct disorders. Journal of Consulting and Clinical Psychology, 55(3), 396-403.
- Everaerd, W., & Dekker, J. (1981). A comparison of sex therapy and communication therapy: Couples

- complaining of orgasmic dysfunction. Journal of Sex and Marital Therapy, 7(4), 278-289.
- Friedman, A. S. (1975). Interaction of drug therapy with marital therapy in depressive patients. Archives of General Psychiatry, 32, 619-637.
- Goldman, A., & Greenberg, L. (1992). Comparison of integrated systemic and emotionally focused approaches to couples therapy. Journal of Consulting and Clinical Psychology, 60(6), 962-969.
- Hahlweg, K., Revenstorf, D., & Schindler, L. (1984). Effects of behavioral marital therapy on couples communication and problem-solving skills. Journal of Consulting and Clinical Psychology, 52(4), 553-566.
- Hahlweg, K., Revenstorf, D., & Schindler, L. (1982). Treatment of marital distress: Comparing formats and modalities. Advances in Behaviour Research and Therapy, 4, 57-74.
- Halford, W. K., Sanders, M. R., & Behrens, B. C. (1993). A comparison of the generalization of behavioral marital therapy and enhanced behavioral marital therapy. Journal of Consulting and Clinical Psychology, 61(1), 51-60.
- Hefner, C. W., & Prochaska, J. O. (1984). Concurrent vs. conjoint marital therapy. Social Work, 29(3), 287-291.

- Huber, C. H., Milstein, B. (1985). Cognitive restructuring and a collaborative set in couples work. The American Journal of Family Therapy, 13(2), 17-27.
- Iverson, A., & Baucom, D. H. (1990). Behavior marital therapy outcomes: Alternative interpretations of the data. Behavior Therapy, 21, 129-138.
- Jacobson, N. (1978). Specific and nonspecific factors in the effectiveness of a behavioral approach to the treatment of marital discord. Journal of Consulting and Clinical Psychology, 46(3), 442-452.
- Jacobson, N. S., Schmaling, K. B., Holtzworth-Munroe, A., Katt, J. L., Wood, L. F., & Follette, V. M. (1989). Research-structured vs. clinically flexible versions of social learning-based marital therapy. Behavior Research and Therapy, 27(2), 173-181.
- Jacobson, N., Schmaling, K., & Holtzworth-Munroe, A. (1987). Component analysis of behavioral marital therapy: Two-year follow-up and prediction of relapse. Journal of Marital and Family Therapy, 13(2), 187-195.
- Jacobson, N. S., Dobson, K., Fruzzetti, A. E., Schmaling, K. B., & Salusky, S. (1991). Marital therapy as a treatment for depression. Journal of Consulting and Clinical Psychology, 59(4), 547-557.
- Jacobson, N. S., & Follette, W. C. (1985). Clinical

significance of improvement resulting from two behavioral marital therapy components. Behavior Therapy, 16, 249-262.

Jacobson, N. S., Follette, V. M., Follette, W. C., Holtzworth-Munroe, A., Katt, J. L., & Schmaling, K. B. (1985). A component analysis of behavioral marital therapy: One-year follow-up. Behavior Research and Therapy, 23(5), 549-555.

Jacobson, N. (1977). Problem solving and contingency contracting in the treatment of marital discord. Journal of Consulting and Clinical Psychology, 45(1), 92-100.

Jacobson, N. (1984). A component analysis of behavioral marital therapy: The relative effectiveness of behavior exchange and communication/problem-solving training. Journal of Consulting and Clinical Psychology, 52(2), 295-305.

James, P. S. (1991). Effects of a communication training component added to an emotionally focused couples therapy. Journal of Marital and Family Therapy, 17(3), 263-275.

Johnson, S. M. (1984). A comparative treatment study of experiential and behavioral approaches to marital therapy. Unpublished doctoral dissertation, University of British Columbia, Canada.

- Johnson, S. M., & Greenberg, L. S. (1985). Emotionally focused couples therapy: An outcome study. Journal of Marital and Family Therapy, 11(3), 313-317.
- Johnson, S. M., Greenberg, L. S. (1985). Differential effects of experiential and problem solving in resolving marital conflict. Journal of Consulting & Clinical Psychology, 53(2), 175-184.
- Kaplan Mehlman, S., Baucom, D. H., & Anderson, D. (1983). Effectiveness of co-therapists versus single therapists and immediate versus delayed treatment in BMT. Journal of Counselling and Clinical Psychology, 51(2), 258-266.
- Lefebvre, N. R. (1981). Conjoint couple-perceived therapist empathy as a correlate of outcome in short-term conjoint therapy. Unpublished Doctoral Dissertation, Boston University, United States.
- Margolin, G., & Weiss, R. (1978). Comparative evaluation of therapeutic components associated with behavioral marital treatments. Journal of Consulting and Clinical Psychology, 46(6), 1476-1486.
- McLean, P. D., Ogston, K., & Grauer, L. (1973). A behavioral approach to the treatment of depression. Journal of Behavior Therapy and Experimental Psychiatry, 4, 323-330.
- Mejia, J. A. (1981). Contributions of therapist's



- relationship and structuring skills to marital therapy: Outcome variance in a Hispanic clinical couples population. Unpublished Doctoral Dissertation, University of Utah, United States.
- O'Leary, K. D., & Beach, S. R. H. (1990). Marital therapy: A viable treatment for depression and marital discord. American Journal of Psychiatry, 147(2), 183-186.
- O'Leary, K. D., & Turkewitz, H. (1981). A comparative outcome study of behavioral marital therapy and communication therapy. Journal of Marital and Family Therapy, 7 (April), 159-169.
- O'Leary, K., Arias, I. (1983). The influence of marital therapy on sexual satisfaction. Journal of Sex and Marital Therapy, 9(3), 171-181.
- Padgett, V. R. (1983). Videotape replay in marital therapy. Psychotherapy: Theory, Research and Practice, 20(2), 232-242.
- Revenstorff, D., Schindler, L., & Hahlweg, K. (1983). Behavioral marital therapy applied in a conjoint-group modality: Short and long-term effectiveness. Behavior Therapy, 14, 614-625.
- Ross, E. R., Baker, S. B., & Guerney, B. G. (1985). Effectiveness of relationship enhancement therapy vs. therapist preferred therapy. American

Journal of Family Therapy, 13(1), 11-21.

Sayers, S. L., Baucom, D. H., Sher, T. G., Weiss, R. L., & Heyman, R. E. (1991). Constructive engagement, behavioral marital therapy, and changes in marital satisfaction. Behavioral Assessment, 13, 25-49.

Schindler, L., Hahlweg, K., & Revenstorff, D. (1983). Short and long-term effectiveness of two communication training modalities with distressed couples.

American Journal of Family Therapy, 11(3), 54-64.

Schmitz, K. L. (1988). Effects of the Guided Communication Model in marital therapy on marital satisfaction, marital communication, intimacy, and self-esteem.

Unpublished Doctoral Dissertation, University of Northern Colorado, United States.

Sher, T., Baucom, D., & Larus, J. (1990). Communication patterns and response to treatment among depressed and nondepressed maritally distressed couples.

Journal of Family Psychology, 4(1), 63-79.

Snyder, D. K., & Wills, R. M. (1989). Behavioral versus insight-oriented marital therapy: Effects on individual and interspousal functioning. Journal of Consulting and Clinical Psychology, 57(1), 39-46.

Snyder, D. K., Wills, R. M., & Grady-Fletcher, A. (1991). Long-term effectiveness of behavioral versus insight-oriented marital therapy: A four-year

- follow-up study. Journal of Consulting and Clinical Psychology, 59(1), 138-141.
- Sokolof, P. A. (1990). Perception of change in communication attitudes by couples with varying levels of marital dysfunction. Unpublished Doctoral Dissertation, University of Northern Colorado, United States.
- Swan, R. W. (1972). Differential counseling approaches to conflict resolution in the marital dyad. Unpublished Doctoral Dissertation, Arizona State University, United States.
- Thompson, R. A. (1980). Effects of videotape playback on causal attribution in distressed couples. Unpublished Doctoral Dissertation, North Texas State University, United States.
- Tsoi-Hoshmand, L. (1976). Marital therapy: An integrative behavioral-learning model. Journal of Marriage and Family Counseling, 2, 179-191.
- Waring, E. M., Carver, C., Stalker, C. A., Fry, R. & Schaefer, B. (1990). A randomized clinical trial of cognitive marital therapy. Journal of Sex and Marital Therapy, 16(3), 165-180.
- Waring, E. M. , & Stalker, C. A. (1991). Waiting list controlled trial of cognitive marital therapy in severe marital discord. Journal of Marital and

Family Therapy, 17(3), 243-256.

Wells, R. A., Figurel, J. A., & McNamee, P. (1977).

Communication training vs. conjoint marital therapy.

Social Work Research and Abstracts, 13, 31-39.

Welsh, E. C. (1976). Counseling and self-help: The

effect of three treatment modes on interpersonal

perception and marital adjustment. Unpublished

Doctoral Dissertation, University of Missouri, United States.

Wheeler, C. M. (1986). Enhancing behavioral marital

therapy with cognitive restructuring and emotional

expressiveness training. Unpublished Doctoral

Dissertation, University of North Carolina, United States.

Whisman, M. A. (1990). The use of booster maintenance

sessions in behavioral marital therapy. Unpublished

Doctoral Dissertation, University of Washington, United States.

Wilson, G. L., Bornstein, P. H., & Wilson, L. J. (1988).

Treatment of relationship disfunction: An empirical evaluation of group and conjoint marital therapy.

Journal of Consulting and Clinical Psychology, 56(6),

929-931.

Zimmer, D. (1987). Does marital therapy enhance the

effectiveness of treatment for sexual dysfunction?

Journal of Sex & Marital Therapy, 13(3), 193-209.

Table 1

## Study Characteristics

Data Set	Study	# of Ss	Mean Age of Ss	M length of Relationship	Treatment	Design	Random Assignment	Follow-Up	M # Sessions	M Length of Treatment	ES
1	Beutler (1971)	20	35.7	n/a	generic	pre-post	n	none	12	12	1.06
2	Swan (1972)	60	n/a	n/a	eclectic	control	n	none	n/a	n/a	-.04
3	Azrin, Naster & Jones (1973)	24	31	9	bct	control	n	none	14	7	1.15
4	McLean, Ogston, & Grauer (1973)	40	n/a	n/a	bct	control	y	3	8	8	1.1
5	Friedman (1975)	80	36.5	n/a	systemic	control	y	none	12	12	0.36
6	Tsoi-Hoshmand (1976)	40	26.2	4.8	bct	control	n	n/a	8.3	10	0.67
7	Welsh (1976)	40	38.5	14.3	generic	control	y	none	6	6	0.49
8	Wells, Figurel, & McNamee (1977)	42	32.9	10.9	systemic	control	y	4.5	10	10	0.06
9	Crowe (1978)	42	37	n/a	psycho-dynamic/ bct	control	y	18	10	10	0.30
10	Jacobson (1977)	64	30.9	5.2	bct	control	y	12	8	8	3.57
10	Jacobson (1978)	24	n/a	n/a	bct	control	y	none	8	8	1.61
11	Margolin & Weiss (1978)	54	46	n/a	bct	control	y	none	4	4	0.64
12	Azrin et al. (1980)	110	29.6	7.3	bct	control	y	24	4	4	0.82

13	Boelens, Emmelkamp, MacGillavry & Markvoort (1980)	42	32.4	8.1	bct/ systemic	control	y	6	10	8	1.19
14	Cookerly (1980)	318	n/a	n/a	generic	pre-post	n	60	n/a	n/a	0.19
15	Thompson (1980)	30	30.82	7	bct	control	y	none	3	3	-0.14
16	Everaerd & Dekker (1981)	44	n/a	n/a	bct	pre-post	n/a	6	16	n/a	0.22
17	Lefebvre (1981)	56	n/a	5.9	generic	pre-post	n	none	15.4	n/a	0.34
18	Mejia (1981)	116	31	n/a	generic	control	n	none	n/a	n/a	1.03
19	O'Leary & Turkewitz (1981)	60	35.4	12.4	bct	control	y	none	10	10	0.66
20	Baucom (1982)	144	32	n/a	bct	control	y	3	10	10	0.73
21	Kaplan Mahlman, Baucom & Anderson (1983)	60	35	n/a	bct	control	y	3	10	10	0.62
22	O'Leary & Arias (1983)	88	n/a	n/a	bct	pre-post	n	none	14	14	0.75
23	Padgett (1983)	48	n/a	n/a	eclectic	treat't versus treat't	y	none	4	4	0.70
24	Hahlweg, Revenstorf & Schindler (1982)	170	n/a	n/a	bct	control	y	12	15	15	0.76
24	Revenstorf, Schindler & Hahlweg (1983)	100	32.9	7.7	bct	control	y	12	15	15	0.78

25	Schindler, Hahlweg & Revenstorf (1983)	104	33.7	8.1	eclectic	control	y	12	15	15	0.51
25	Hahlweg, Revenstorf & Schindler (1984)	86	33.5	8.1	bct	control	y	none	15	15	1.03
26	Hefner & Prochaska (1984)	54	37.6	13.4	generic	pre-post	y	none	6	12	0.03
27	Johnson (1984)	90	35.3	8.6	bct/ systemic	control	y	2	8	8	1.03
27	Johnson & Greenberg (1985)	90	35.3	8.6	systemic	pre-post	y	2	8	8	1.03
28	Bennun (1985)	90	35	7.2	bct	treat't versus treat't	y	6	10	10	1.75
29	Huber & Milstein (1985)	17	n/a	n/a	bct	control	y	none	6	6	0.39
30	Johnson (1985)	28	33	6.9	systemic	alone	n	2	8	8	0.93
31	Ross, Baker & Guerney (1985)	48	41	17.2	generic	pre-post	y	none	10	10	1.08
32	Baucom & Lester (1986)	48	32	n/a	bct	control	n	6	12	12	0.63
33	Wheeler (1986)	72	n/a	n/a	bct	control	y	none	12	12	0.43
34	Cline, Jackson, Klein, Mejia & Turner (1987)	176	n/a	n/a	eclectic	pre-post	n	none	8.2	n/a	1.15



35	Dadds, Schwartz,& Sanders (1987)	48	n/a	n/a	generic	control	y	none	6	n/a	0.63
36	Jacobson, Schmaling & Holtzworth- Munroe (1987)	34	37.4	10.1	bct	control	y	24	12	12	0.51
36	Jacobson (1984)	72	37.4	10.1	bct	control	y	none	12	12	1.17
36	Jacobson et al. (1985)	120	37.4	9.7	bct	control	y	12	12	12	0.49
36	Jacobson (1985)	60	n/a	n/a	bct	control	y	none	12	12	0.49
37	Zimmer (1987)	56	29	n/a	bct	control	y	3	12	12	1.1
38	Schmitz (1988)	42	n/a	n/a	eclectic	control	y	none	8	8	0.25
39	Wilson, Bornstein & Wilson (1988)	30	33	9.6	bct	control	y	6	4	4	1.17
40	Jacobson et al. (1989)	60	41	11.4	bct	pre-post	y	6	21	21	1.22
41	Snyder & Wills (1989)	79	38.6	12.6	psycho- dynamic/ bct	control	y	6	19	20	0.72
41	Baucom, Sayers & Sher (1990)	60	34	8.1	bct	control	y	6	12	12	0.48
41	Sher, Baucom & Larus (1990)	94	34	n/a	bct	pre-post	n/a	none	12	n/a	0.60
41	Sayers, Baucom, Sher & Heyman (1991)	60	34	8.1	bct	control	y	none	12	12	0.37

41	Snyder, Wills & Grady- Fletcher (1991)	55	n/a	n/a	psycho- dynamic/ bct	pre-post	y	none	20	20	0.00
42	Iverson & Baucom (1990)	96	32.7	n/a	bct	pre-post	n	none	12	12	0.14
43	O'Leary & Beach (1990)	36	n/a	n/a	bct	control	y	none	16	16	2.00
44	Sokolof (1990)	20	35.5	9.4	eclectic	pre-post	n	none	8	8	1.6
45	Waring, Carber, Stalker, Fry & Schaefer (1990)	66	n/a	n/a	bct	control	y	none	10	10	0.16
46	Whisman (1990)	54	39.7	10.4	bct	pre-post	y	12	21.5	20	1.11
47	Jacobson, Dobson, Fruzzetti, Schmaling & Salusky (1991)	60	39.5	11.3	bct	control	y	none	20	20	-0.52
48	James (1991)	42	n/a	9.6	systemic	control	y	4	12	12	0.78
49	Waring & Stalker (1991)	82	n/a	n/a	bct	control	y	none	10	10	0.16
50	Goldman & Greenberg (1992)	66	38.5	11.3	systemic	control	y	4	10	10	1.53
51	Halford, Sanders & Behrens (1993)	26	41.7	15.7	bct	pre-post	y	3	13.5	15	1.09

Table 2

Test of Heterogeneity	N	Chi Square	Weighted g	Prob.
Treatment vs. Treatment	4	35.65	0.78	P<.005
Pre-Post	14	32.57	0.75	P<.005
Treatment vs. Control	33	118.00	0.62	P<.005
Random Assignment	37	128.91	0.68	P<.005
Non-Random Assignment	14	49.71	0.55	P<.005
Self-Referred Clients	30	119.74	0.66	P<.005
Recruited Clients	18	44.44	0.71	P<.005
Published Research	41	148.00	0.65	P<.005
Non-Published Research	10	28.12	0.62	P<.005
Psychodynamic	2	2.28	0.59	P<.14
Systemic	7	20.20	0.83	P<.005

BCT	30	95.40	0.70	$P < .005$
Eclectic	9	34.09	0.65	$P < .005$
Generic	8	17.18	0.41	$P < .02$

Table 3

Test of Heterogeneity	N	Chi Square	Quality	Probability
			Weighted g	
Published	41	0.09	0.013	p=1.0
Unpublished	10	0.008	0.012	p=1.0
Systemic	7	0.01	0.018	p=1.0
Generic	8	0.01	0.007	p=1.0
BCT	30	0.06	0.015	p=1.0
Eclectic	9	0.01	0.012	p=1.0
Psychodynamic	2	0.00	0.01	P=1.0

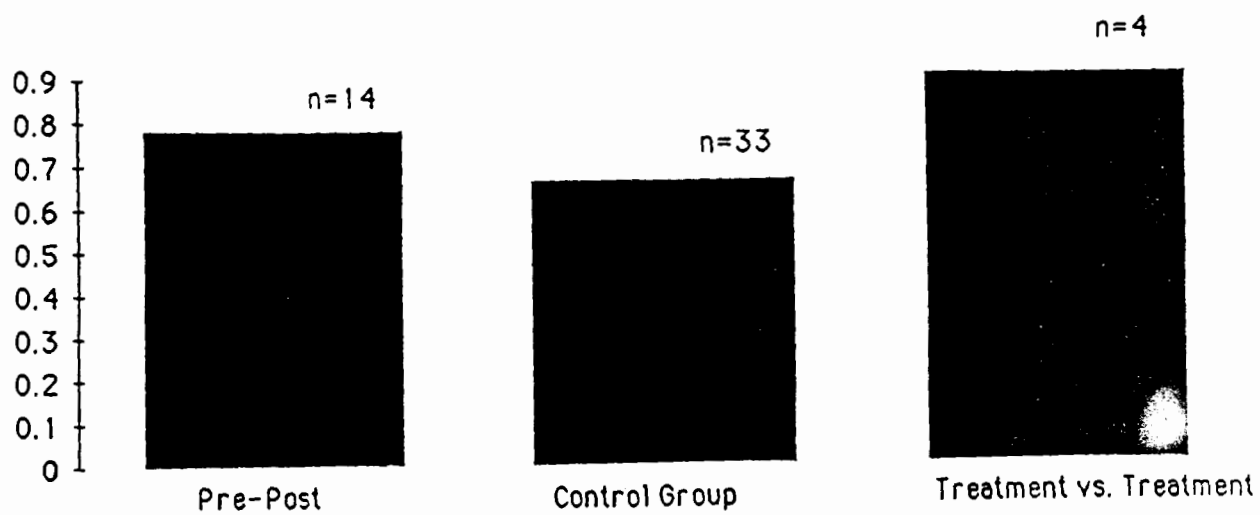


Figure 1. Weighted ESs by study design.

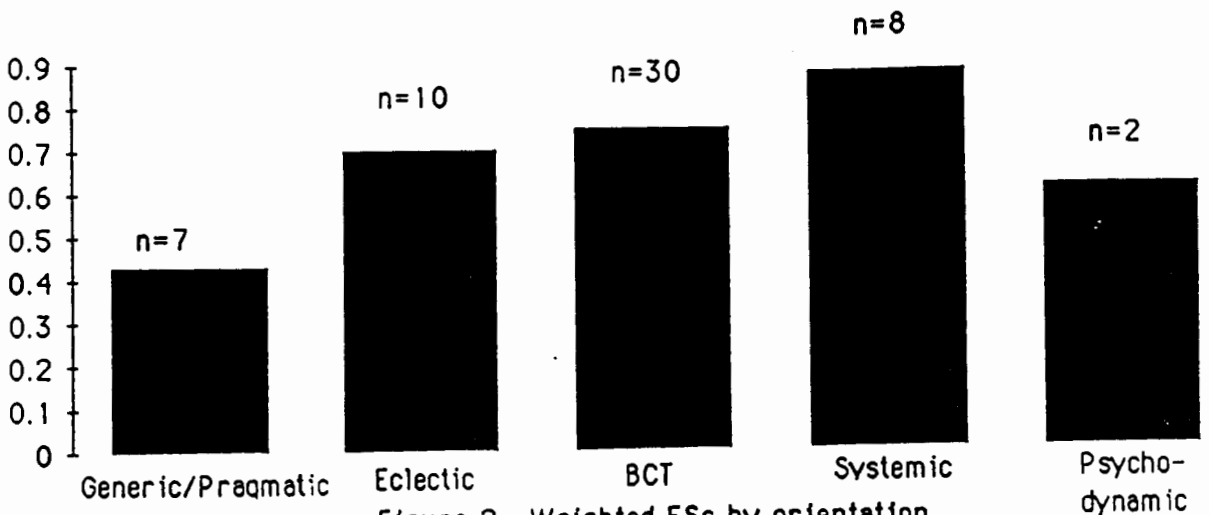


Figure 2. Weighted ESs by orientation.