

A META-ANALYSIS OF EMPATHY TRAINING PROGRAMS  
FOR CLIENT POPULATIONS

by

Robert Paul Butters

A dissertation submitted to the faculty of  
The University of Utah  
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

College of Social Work

The University of Utah

May 2010

Copyright © Robert Paul Butters 2010

All Rights Reserved

## STATEMENT OF DISSERTATION APPROVAL

The dissertation of Robert Paul Butters  
has been approved by the following supervisory committee members:

<u>Brad W. Lundahl</u>	, Chair	<u>03/14/10</u>
<u>David Derezotes</u>	, Member	<u>3/14/10</u>
<u>[REDACTED] A. Pett</u>	, Member	<u>3/14/10</u>
<u>[REDACTED] L. Smith</u>	, Member	<u>3/14/10</u>
<u>Joanne Yaffe</u>	, Member	<u>3/14/10</u>

and by [REDACTED]  
the Department of Social Work

and by Charles A. Wight, Dean of The Graduate School.

## ABSTRACT

Empathy is a core component of human relationships and a cornerstone of effective interpersonal skills. Low levels of empathy are associated with negative outcomes while high empathy contributes to prosocial behaviors. A large body of literature exists on empathy but as yet there is no comprehensive review or meta-analysis of empathy training programs with client populations. This study is a meta-analysis that includes 24 studies of empathy training programs for client populations. The findings suggest that empathy training programs are effective in increasing empathy with an overall large effect size of  $g = 0.911$ . Moderator analyses were conducted on study design variables, program characteristics, and client variables. Moderator analyses suggest that empathy training is effective across a variety of client populations and training modalities. One significant moderator effect related to the way empathy is measured; studies using self-report outcomes show moderate gains in empathy ( $g = 0.386$ ) while studies using observer-report outcomes show large gains ( $g = 1.488$ ). Because of this large difference, the studies were divided by measurement type. Once the studies were divided by outcome type, none of the moderators continued to be significant across both measurement types. So while it appears that empathy training has a moderate to large effect on increasing empathy, the way empathy is measured is important and moderates the strength of the effect. These findings lead to a discussion on the measurement of empathy, the components of empathy, and the very construct of

empathy. The study concludes with recommendations for further research on empathy training programs.

## TABLE OF CONTENTS

	Page
ABSTRACT.....	iii
LIST OF TABLES.....	vii
LIST OF FIGURES.....	vii
ACKNOWLEDGMENTS.....	ix
CHAPTERS	
1. INTRODUCTION.....	1
Research Questions.....	3
Review of Empathy Training Programs.....	4
Research Goals.....	6
2. LITERATURE REVIEW.....	8
Definition of Empathy.....	9
Theoretical Foundations .....	15
Development and Mechanism of Empathy.....	19
Importance and Function of Empathy.....	25
Review of Empathy Training Programs.....	29
Review of Possible Moderators to Empathy.....	34
Empathy Scales (Dependent Variable and Moderator).....	43
Summary and Limitations of Literature.....	48
3. METHODOLOGY.....	51
Overview of Meta-analysis.....	53
Formulate Specific Research Questions.....	54
Literature Search.....	54
Study Selection.....	57
Inclusion Criteria.....	57
Statistical Procedures.....	64

Publication Bias.....	70
4. RESULTS.....	73
Sorting and Screening Studies.....	73
Meta-analysis of Data.....	77
Partitioned Data Analysis.....	96
Publication Bias.....	106
Conclusion.....	107
5. DISCUSSION.....	109
Summary of Results.....	109
What Specific Interventions Promote Empathy?.....	113
Does Training Increase Cognitive or Affective Empathy, or Both?.....	114
Does the Location or Setting of the Training Make a Difference? .....	114
Is Empathy Training More Effective in Groups?.....	115
Are Results Durable Over Time?.....	115
Does the Dosage of the Training Influence Outcomes?.....	116
Does the Number of Training Ingredients Influence Outcomes?.....	117
Does the Age of the Client Influence Outcomes?.....	117
Does the Gender of the Client Influence Outcomes?.....	118
Is Empathy Training More Effective at Certain Levels of Prevention?.....	119
Does Empathy Training Work Across Client Populations?.....	119
Does Voluntary Status Influence Outcomes?.....	120
Do Certain Ethnic Populations Respond Better to Empathy Training?.....	121
Contribution to the Literature on Empathy.....	121
Implications of the Study.....	123
Limitations of Study.....	126
Future Directions.....	129
Conclusion.....	130
 APPENDICES	
A. STUDY SEARCH AND SORTING STRATEGY.....	132
B. PROTOCOL FOR SCREENING, SORTING, AND CODING ARTICLES FOR INCLUSION .....	134
C. CODE SHEET AND CODE BOOK .....	136
D. STUDY QUALITY RATING FORM-EMPATHY TRAINING PROGRAMS .....	142
REFERENCES.....	144

## LIST OF TABLES

<u>Table</u>	<u>Page</u>
1. Excluded Studies.....	76
2. Included Studies.....	78
3. Windsorizing Data.....	80
4. Effect Sizes and Forest Plots by Study Quality.....	82
5. Empathy Training Moderators: Study Variables.....	85
6. Empathy Training Moderators: Program Variables.....	89
7. Empathy Training Moderators: Program Ingredients.....	91
8. Empathy Training Moderators: Participant Variables.....	93
9. Self-report Outcomes: Study Variables.....	98
10. Observer-report Outcomes: Study Variables.....	98
11. Self-report Outcomes: Program Variables.....	99
12. Observer-report Outcomes: Program Variables.....	100
13. Self-report Outcomes: Program Ingredients.....	101
14. Observer-report Outcomes: Program Ingredients.....	103
15. Self-report Outcomes: Participant Variables.....	104
16. Observer-report Outcomes: Participant Variables.....	105

## LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1. PRISMA-type Flow Chart of Sampling Procedure.....	74
2. Meta-regression of Study Quality of Hedges' $g$ .....	87
3. Meta-regression of Year Published on Hedges' $g$ .....	87
4. Partitioning Data by Outcome Type.....	95
5. Funnel Plot to Assess Publication Bias.....	108

## ACKNOWLEDGMENTS

My entire committee, Dr. Brad Lundahl, Dr. David Derezotes, Dr. Marge Pett, Dr. Larry Smith, and Dr. Joanne Yaffe, deserve the highest praise and thanks for sharing their experience, wisdom, and insight throughout this project. A special thanks to my chairperson, Dr. Lundahl, for his generosity, fortitude, attention to detail, and narrative. The selfless hours you spent teaching, editing, discussing methods and results, and maybe most importantly, being patient and positive, facilitated the completion of this project.

Thank you to the graduate research assistants who assisted with various aspects of this project, Marcella Hurtado, Christian Sarver, and Sean Baenzinger. You all surely have bright futures as social workers and researchers.

Thank you to the entire faculty and staff at the University of Utah's College of Social Work. Special thanks to Ph.D. Director Hank Liese and Dean Jannah Mather for unwavering support throughout my education. Also, a special thanks to Professor Russ Van Vleet, for guidance, wisdom, and inspiration over the past decade. I have been fortunate to work alongside you and to learn to ask the hard questions and fight for true justice.

I couldn't have done any of this without the support of my friends and family. My partner Cindy Butters loves me, inspires me, and reminds me to maintain balance in all things. My children Madison, Samantha, Robert, and Jack have sacrificed a lot of daddy time for this project; thank you for understanding and for being a constant

reminder of what's really important. My extended family, friends, and co-workers at LifeMatters have always been there to help and to support me, thank you.

## CHAPTER 1

### INTRODUCTION

Interpersonal difficulties lead a number of people to seek social work related services; interventions that target interpersonal functioning consistently describe the cultivation of empathy as a central component of treatment across a wide variety of clients and settings (Hepworth, Rooney, Rooney, Strom-Gottfried, & Larsen, 2010). Empathy is a basic human quality that dissolves feelings of alienation and allows people to connect with others (Rogers, 1959). The development of empathy starts in early childhood attachment experiences and provides the necessary foundation for moral development (Kagan, 1984). Deficits in empathy can manifest in a variety of social problems and many social work interventions seek to increase client empathy in an attempt to decrease dysfunctional behaviors and improve social functioning. In the published literature there are numerous programs that use various modalities to increase empathy. However, the overall effectiveness of these programs is not clear and the heterogeneity seen in empathy training programs makes it difficult to determine which variables contribute to effective programs. To date, there is no published review of empathy training programs for client populations or an analysis of the factors that contribute to effective empathy training. The present study seeks to fill this void.

Empathy is a critical component in prosocial behaviors and moderates aggressive behavior (Eisenberg & Miller, 1988) and the absence of empathy is a hallmark of autism

and psychopathy (Blair, 2008). Empathy helps individuals establish and maintain friendships (Del Barrio, Aluja, & Garcia, 2004; Hay, 1994), enhances satisfaction in intimate relationships (Long, Angera, & Hakoyama, 2008), improves the quality of family relationships (Guernsey, 1988), and is positively associated with family cohesion, parental support, and communicative responsiveness (Henry, Sager, & Plunkett, 1996). Empathy has been positively correlated with helping behaviors and negatively correlated with aggression (Jolliffe & Farrington, 2006a; Mehrabian & Epstein, 1972). Studies indicate that increases in empathy can reduce sexual assaults and the incidence of domestic violence (Berg, Lonsway, & Fitzgerald, 1999; Foubert, 2006; O'Donohue, 2003). Further support for the salutary effects of empathy comes from studies of parent training programs, where increasing parental empathy is believed to reduce children's aggressive and acting-out behaviors (Eyberg, Nelson, & Boggs, 2008; McMahon & Washburn, 2003). Further, some research has demonstrated that improving empathy through family and couples counseling leads to improved relationships and decreased family problems (Long, 1999).

Empathy is a complex construct. In an attempt to understand empathy, it has often been understood to have two main components: cognitive and affective empathy. Cognitive empathy is the ability to accurately perceive and appropriately respond to the thoughts and feeling of another person, while affective empathy is the actual emotional sharing of another's emotional state (Carkhuff & Truax, 1965; Hodges & Meyers, 2007). While affective empathy is considered to be a more genuine empathic response, cognitive empathy is frequently the target of training programs. In empathy training programs people are taught to recognize emotions and respond appropriately as a social skill. The

central assumptions of most empathy training programs are (a) empathizing with another person creates a shared emotional experience and this common ground is the basis for building trust and affiliation, and (b) empathy for another being is likely to deter or reduce hurtful behavior toward them because the other's pain becomes shared. Restated, if I share your emotional experience, your interests become my interests and I will avoid hurting you because I will feel your pain.

Both cognitive and affective empathy are fundamental parts of meaningful human relations and are important to everyday social interchanges as well the problems often targeted by social workers. Given the important role empathy plays in social relations, clinicians and researchers have developed programs designed to teach empathy. The assumption is that if empathy can be taught, clients and the people around them will enjoy the benefits typically associated with higher levels of empathy. Developing such programs does not, of course, guarantee that they will be effective. Clinicians and researchers have, therefore, investigated the effectiveness of such programs independently. Of particular interest to this dissertation are questions regarding the effectiveness of empathy training programs and factors influencing such effectiveness.

### Research Questions

The research questions for this study are as follows:

1. Do empathy training programs increase empathy in client populations?
2. Is there variation in empathy training program outcomes?
3. What moderating variables contribute to empathy training programs and what is the relative strength of these variables?

To investigate these questions, meta-analysis will be used to provide a quantitative analysis of empathy training programs.

### Review of Empathy Training Programs

This section highlights studies on empathy training programs, studies from diverse client groups demonstrating the ubiquitous nature of empathy, and the importance of this study in determining the salient qualities of effective programs.

In a study of incarcerated child molesters (Marshall, O'Sullivan, & Fernandez, 1996), participants showed significant increases in victim empathy following a group empathy training program. The increase in empathy was negatively correlated with future sexual offending.

In another study targeting violence prevention with African American youth (McMahon & Washburn, 2003), a school based program showed significant increases in empathy following participation in the program. The increases in empathy were associated with a reduction in aggressive behaviors.

In an empathy training program targeting couples in romantic relationships (Long, Angera, Carter, Nakamoto, & Kelso, 1999), a 10-hour empathy training group demonstrated increased empathy following the program. The increase in empathy was positively correlated with relationship satisfaction at a 6-month follow-up.

In a study of mothers receiving a two-week training program (Grskovic, 2008) that included a strong empathy component, the participants demonstrated increases in acceptance, empathy, and positive attention towards their children following the training. This intervention was correlated with a decrease in child behavior problems.

These programs demonstrated that empathy training programs can be effective and lead to other positive outcomes. Yet, little is known about what types of empathy programs are most effective, or if certain populations respond better than others. A cursory review of these studies reveals considerable variation with regard to target populations and program delivery. This variation makes it difficult for practitioners to make informed decisions when selecting and implementing effective programs.

Gordon Paul (1967) stated the classic question: "What treatment, by whom, is most effective for this individual, with that specific problem, and under which set of circumstances?" (p. 111). Further information about empathy training programs is needed to provide clearer and more confident direction about their use. This dissertation explores questions about the effectiveness of empathy training programs based on clients' presenting problem, target populations, modalities of training, and other contextual factors. To date, there is no known research that broadly synthesizes the existing literature on empathy training programs and provides an overall estimate of the effectiveness of empathy training programs or examines the variables that may moderate program effectiveness.

It cannot be assumed that all programs are effective overall or that all training components are connected to successful outcomes. For example, in a study of rape prevention programs on college campuses (Berg et al., 1999), college fraternity men were randomly assigned to one of two treatment groups or a control group. One treatment group listened to an audiotape of a woman describing her sexual assault and the other treatment group listened to a male describing his sexual assault. The group that listened to the woman describing the assault actually increased the endorsement of rape

supportive behaviors, like using alcohol or coercion, following the intervention; neither group showed any significant increases in empathy as a result of the intervention. This study illuminates the importance of empirically assessing the effectiveness of empathy training programs, especially considering that even programs with good face validity may have unintended outcomes. Said differently, there is no guarantee that professional intervention will help clients and it is possible that it may actually have a negative effect on client outcomes.

The use of a meta-analytical research strategy to investigate empathy training programs is well suited to answer the research questions proposed in this study. This methodology is a form of survey research that uses research reports rather than subjects and yields an effect size measure that will allow for comparison of programs using a common metric (Lipsey & Wilson, 2001). The study synthesizes the published literature on empathy training with client populations to assess overall effectiveness, investigate heterogeneity of effect, and explore potential outcome moderators. The systematic process and procedures used in this methodology will yield research results valuable and accessible to practitioners and researchers working in child abuse, sexual violence, criminal offending, and interpersonal relationships.

### Research Goals

Using this meta-analytical methods, the goals of this study are:

1. Investigate the overall effectiveness of empathy training programs with client populations through meta-analytical methodology.
2. Assess potential heterogeneity of effect.

3. Examine potential moderators to the overall effectiveness of empathy training programs with client populations through meta-analysis.
4. Make recommendations for effective empathy training programs.

The following chapters provide the literature review, methodology, results, and the discussion of results for the study.

## CHAPTER 2

### LITERATURE REVIEW

Empathy is a core component in human relationships and a cornerstone of effective interpersonal skills (Carkhuff & Truax, 1965). Methods for training clients to understand, feel, and demonstrate empathy are firmly entrenched in intervention programs and theoretical models to improve client functioning (Hepworth et al., 2010; Miller & Rollnick, 2002). Although there appears to be universal acceptance of empathy as an important component of social interventions, as well as a body of literature describing various empathy training programs, as yet there is no comprehensive review and analysis of empathy training programs with client populations to determine effectiveness and guide program development. This study will provide an analysis of important program and client moderators that will be accessible to practitioners and researchers seeking to design quality empathy training programs or improve existing programs. The goals of this research project are:

1. Investigate the overall effectiveness of empathy training programs with client populations through meta-analytical methodology.
2. Assess potential heterogeneity.
3. Examine potential moderators to the overall effectiveness of empathy training programs with client populations through meta-analysis.
4. Make recommendations for effective empathy training programs.

This literature review will provide an overview of the extant literature on empathy training programs. The review will be structured according to the following: (a) define empathy; (b) examine the theoretical foundations of empathy to inform an understanding of the construct of empathy and provide a foundation for this analysis of empathy training programs; (c) review the development of empathy and the mechanisms by which empathy emerges in human relations; (d) discuss the importance of empathy and its function in human interactions, including the impact of having high or low empathy; (e) review studies of empathy training programs that will identify variation and gaps in the literature on empathy training; (f) explore factors that emerge in the literature related to moderators of empathy training programs; and (g) summarize the limitations of the existing research on empathy training programs, providing a rationale and justification for this study.

### Definition of Empathy

One of the tensions in studying empathy relates to the way empathy is defined. Currently, no consensus exists on how to define empathy, how it functions and manifests in interpersonal relationships, or how to train or measure it. This conceptual fragmentation, which informs this inquiry, confounds a comprehensive definition of empathy and contributes to inadequate methodology and measurement of outcomes of training programs (Varker, Devilly, Ward, & Beech 2008). Prior to investigating the effectiveness of empathy training programs, I need to start by defining and operationalizing empathy as a construct. By first discussing what empathy is, we better understand the challenges faced by programs working to teach empathy and measure outcomes.

Carl Rogers, the progenitor of client-centered therapies, proposes that empathy is a core competency for therapists, along with unconditional positive regard, warmth, and genuineness. Rogers offers a classical definition of empathy for helping professionals:

The state of empathy or being empathic is to perceive the internal frame of reference of another with accuracy and with the emotional components and meanings which pertain thereto as if one were the person, but without ever losing the “as if” condition. (1959, pp. 210-211).

One limitation of this definition is the specifier “as if” because it denotes a professional distance or boundary. This prescribed barrier serves to protect against over-identification with the client, which may result in reduced objectivity and professional efficacy. In non-professional interactions, the “as if” condition is not always present; rather, it may be a function of our ego that allows us to experience the emotions of another while maintaining a differentiated self. This as if condition or boundary can vary based on the setting, nature, and strength of relationship. The emotional state of a close loved one generally engenders a stronger emotional response than that of an acquaintance. Our empathic responses may vary depending on the context of the interaction (Zaki, 2008), individual temperament (Strayer & Roberts, 2004), age (Hoffman, 1987), gender (Baron-Cohen, 2003), and personal history (Eisenberg & Fabes, 1998).

A relational-feminist approach to empathy redefines it as “an affective attunement...that reflects the ability and willingness to feel the presence of each other and the impact each has made on the other” (Freedberg, 2007, p. 254). The relational-feminist theory is drawn from relational-cultural theory developed at the Stone Center of Wellesley College. This theory emphasizes feminist theory and an interpersonal framework of mutuality and equality. The founders of the Stone Center define empathy

as “the capacity to feel and think something similar to the feelings and thoughts of another person that exists in all people” (Miller & Stiver, 1997, p. 27). This broader definition, which is inclusive of the relational-feminist theoretical framework, recognizes the complex and interactive nature of empathy that is an essential component of empathy training programs. This complexity informs this study that examines various client and program moderators for effective programs. Clarifying the definition of empathy will allow for better evaluation of empathy training programs and ultimately improve interventions that seek to increase empathy.

One of the major tensions in defining empathy concerns the true nature of empathy as a construct. As with any complex human quality, there are a multitude of emotional, cognitive, and behavioral components to empathy. Because we can only reliably assess something that can be seen and measured, with many social and psychological constructs we must use indicators and proxies to approximate true values. For this reason, cognitive or behavioral manifestations of empathy are more often the target of empathy training programs. Evaluations of these programs rely mainly on observer or self-reported outcomes that may be insufficient because such measures fail to assess the more internal, emotional components of empathy. Most researchers agree that empathy has a strong affective component that is considered by many to constitute a true empathic response (Thwaites, 2007). These complexities in defining empathy have led many leading researchers to operationalize empathy into very specific definitions that may be assessed by their proprietary instruments.

Hogan (1969), who created the Hogan Empathy Scale, described empathy as "the intellectual or imaginative apprehension of another's condition or state of mind without

actually experiencing that person's feelings" (p. 308). Clearly, this definition reflects a purely cognitive process. Empathy has also been defined as emotional arousal or sympathy in response to the feelings or experiences of others. For example, Mehrabian and Epstein (1972), who created the Questionnaire Measure of Emotional Empathy, defined empathy as "the heightened responsiveness to another's emotional experience" (p. 526). Finally, there is an integrative approach, proposed by Davis, the creator of the Interpersonal Reactivity Index, which employs both cognitive and emotional approaches to the study of empathy (Davis, 1983; Thornton & Thornton, 1995).

Empathy has been conceptualized broadly to capture the full range of empathic responses and feelings. Davis conceptualizes general empathy as "a set of constructs having to do with the response of one individual to the experiences of another" (1994, p.12). Hoffman (2001) proposes that empathy is "an emotional response that is focused more on another person's situation or emotion than on one's own" (p. 62). Although broad definitions of empathy may capture a larger portion of a true empathic response, they lack the specificity to inform interventions and development of outcome instruments that can measure treatment progress or effectiveness. As a result, many researchers have developed multifaceted and sequential models for empathy.

Davis (1983) suggested that empathy is a multi-component response, involving the following four stages: (a) perspective-taking, which is the ability to adopt the viewpoint of another person; (b) fantasy, which is the ability to transpose oneself into the feelings of a fictitious character; (c) empathic concern, which are feelings of concern for another person; and (d) personal distress, which are self-oriented feelings of distress. This model of empathy more clearly delineates specific skills and stages that could be targets

of an intervention, but it is not without critics. Marshall, Hudson, Jones, and Fernandez (1995) suggest that this multifactor definition of empathy simply repackages constructs that already exist in the literature, arguing that perspective-taking is a cognitive component of empathy, personal distress is an affective component of empathy, and empathic concern is more like sympathy than empathy.

A leader in the field of treating sexual offenders, Marshall conceptualized empathy with the following invariant sequence of stages: (a) emotional recognition, which is the ability to discriminate the emotional state of another person; (b) perspective-taking, which is the ability to see situations from another's perspective; (c) emotion replication, which is replication of the observed emotion; and (d) response decision, which is decision making that is based on the feelings experienced. Unlike other conceptualizations, Marshall's view of the process involves both affective and cognitive empathic responses, requiring self-awareness and an ability to separate one's own experiences from that of the other person.

Hodges and Meyer (2007) propose three main components of emotional empathy. The first is feeling the same emotion as another person (sometimes attributed to emotional contagion, e.g., unconsciously "catching" someone else's tears and feeling sad oneself). The second component, personal distress, refers to one's own feelings of distress in response to perceiving another's plight, which may or may not mirror the emotion that the other person is actually feeling. For example, one may feel distress, but not specifically depression, when another person says he or she is so depressed he or suicidal; similarly, one feels distress, but not actual pain, when one sees someone fall. The third emotional component, feeling compassion for another person, is the one most frequently

associated with the study of empathy in psychology. It is often called *empathic concern* and sometimes *sympathy*. Empathic concern is thought to emerge later developmentally and to require more self-control than either emotional contagion or personal distress, although these earlier components (along with the ability to imitate) probably lay the groundwork for later, more sophisticated forms of empathy.

Cognitive empathy, which requires sensitivity and knowledge about emotions and expressivity, is not devoid of emotion; however, it does not generally include caring about the other person. Cognitive empathy can be described as a social skill that may be developed and improved (Henry et al., 1996). There is evidence that individuals who lack the perceptual or neurobiological capacities for an automatic empathic response can be taught to reflect and respond empathically. Autistic children are being taught, through conditioning techniques, to respond to facial expressions in prosocial and seemingly empathic ways (Baron-Cohen, 2003). Empathy skills are commonly taught in professional schools of social work, psychology, medicine, and nursing as a means of building trust and affiliation with clients. In recognition of the power of empathic responsiveness, the concept and practice of empathy is often seen in the fields of social psychology and client-centered psychotherapy. In professional training, much has been made of the distinction between empathy and sympathy, but the two terms are often used interchangeably. When a distinction is made, empathy is often defined as understanding another person's experience by imagining oneself in that other person's situation. Sympathy may involve the experience of being emotionally moved by, or responding in tune with, another person. Sympathy is sometimes referred to as a more emotional response, while empathy is considered by some to be a more cognitive response. Schafer

(1959) describes generative empathy as the inner experience of sharing in and comprehending the momentary psychological state of another person.

Hobart and Fahlberg (1965) conceptualize empathy by comparing it with its opposite: projection. If empathy is an accurate perception of another's thoughts, feelings, and behaviors, then projection is the opposite: seeing one's self in others and imagining them to share your feelings and behaviors. This distinction is particularly important as we attempt to measure empathy and empathic responding, because some instruments fail to discriminate between empathic responding and projective identification. Many questions on empathy scales ask about how others feel, reflecting empathic responding, projection, or both (Jolliffe & Farrington, 2006b).

The psychodynamic perspective described above, that includes projective identification, illuminates some of the complexities in defining and conceptualizing empathy. Either in the form of empathy or projective identification, it is clear how important such relatedness is in social situations. A person with strong empathic abilities has a clear advantage in nearly all circumstances over a person with impaired empathy. The cumulative impact of this subtle advantage is magnified over time in our interpersonal, educational, and vocational relationships. A friend, spouse, manager, employee, or social worker with strong empathic abilities will nearly always reap benefits from this connectedness, while an emotionally unavailable, socially awkward individual may struggle with day-to-day tasks.

### Theoretical Foundations

The construct of empathy is firmly entrenched in philosophical and psychological theory and has been discussed for hundreds of years as a central component of human

relatedness and morality by philosophers like Aristotle (1976), David Hume (2004), and Immanuel Kant (as cited in Kennett, 2002). In the 20<sup>th</sup> century, empathy has been discussed by most major theorists in social work and allied professions. Mary Richmond recognized the importance of sympathetic responding in casework with clients (Richmond, 1922). Hodges and Myers (2007) traced the history of empathy and found that while the concepts of empathy and sympathy are very old, we find specific references to empathy appearing at the beginning of the 20th century, often in discussions of art. The origin of the word empathy may be traced to the German word *Einfühlung*, which translates literally as "feeling into," as in projecting oneself into something else.

The development of empathy contributes to a less egocentric view that allows for others to be valued as important. This developmental process is linked to moral development (Piaget, 1932) and judgment (Kohlberg, 1981). The ability to respond to another's experience helps us to more effectively and meaningfully interact in social situations (Kohut, 1959). Our ability to feel with another allows us to relate on a deeper emotional level and serves as a deterrent to harmful behaviors. These kind of social and moral rules are embedded in the structure of our society; individuals who miss the subtle cues or misunderstand messages based on deficits in empathic perception and communication are at risk for a myriad of social ills and of becoming dangerous to others.

Christine Olden discusses empathy as a complex and dichotomous construct. In *Notes on the Development of Empathy* (1958), she argues that empathy is the sensibility of one person towards another that allows for the maintenance of one's unique separateness. She stresses the importance of being able to understand another's thoughts

and feelings and create emotional intimacy, while also maintaining a sense of the independence and uniqueness of the person with whom one is empathizing. This idea of joining while remaining objective and separate is also endorsed by Giuseppe Di Chiara (1982), who describes the importance of working to develop intimacy and closeness while remaining separate.

Carkhuff and Truax developed one of the first empathy training programs as well as the first widely used empathy measurement scale (Truax and Carkhuff, 1967). They conceptualized empathy as consisting of three main qualities: accurate empathy, non-possessive warmth, and genuineness. They theorize that a combination of these qualities in a helping professional contribute to positive outcomes for patients because the patient is able to engage in self-exploration. The Carkhuff and Truax training scales are used extensively in professional schools of social work, psychology, and in the health professions in addition to client populations, to evaluate empathy.

Feminist-relational theory provides a conceptual framework that emphasizes growth, development, and the fostering of relational activities (Freedberg, 2007). This theory describes empathy more broadly as “the capacity to feel and think something similar to the feelings and thoughts of another person that exists in all people” (p. 254). This relational view describes an alternate model for healthy development of the self. According to this theory, maturation of the self or ego occurs through the process of relationship-differentiation, rather than through the more traditional models of separation-individuation (Erickson, 1963; Mahler, 1975). Using this model, with relationship at its core, we see empathic responding and relating as a central process in development, socializations, and healthy functioning.

Empathy is generally discussed as relating to another's emotional state in a way that is responsive to a person's unfortunate or disadvantaged circumstance. If we expand this traditional definition to include a broader range of emotional connectedness and expression, empathy may be conceptualized as any response to another's emotional state. Clearly, we have the capacity to respond empathically not just to a negative emotional state but to another's positive emotions as well. The positive emotional response that accompanies the connection to and sharing of a loved one's positive emotional states is an important component of empathy that deserves further investigation. Empathy training programs that focus on more holistic emotional expression and connectedness have been largely neglected in the empathy training literature. This holistic definition of empathy, which may enhance our understanding of empathy training programs and interventions, needs further investigation.

Positive empathic responding may not be discussed as frequently because many empathy interventions are targeted to remediate deficiencies in social and behavioral areas. Varker and associates (2008) and Marshall (1995) suggest that appropriate response to positive emotions may not deter harmful behaviors; as a result, programs that focus on positive emotional response may have been neglected in the literature and in training programs. A more inclusive definition of empathy that includes the full range of emotional sensitivity and responding is surely more accurate and reflective of human experience.

Although there remains considerable disagreement in the professional literature about an inclusive definition of empathy (Davis, 1983; Jolliffe & Farrington, 2006b; Thornton & Thornton, 1995; Thwaites, 2007) most clinicians and researchers agree that

empathy includes a *cognitive component*, or an ability to perceive and decode another's emotional state; an *affective component*, or an emotional connection to another's emotional state; and a *behavioral component*, or an action taken to demonstrate empathy. In addition, more recent definitions of empathy also include an *interactive component* (Freedberg, 2007; Zaki, 2008) that moderates the perception and expression of empathy between individuals.

### Development and Mechanism of Empathy

How do humans develop empathy and how is empathy manifested? Object relations theory and ego psychology speak directly to these questions (Fairbairn, 1952; Goldstein, 2001; Hartman, 1958; Mahler, 1968). The foundations of empathy and social interactions begin in our early infancy and are firmly entrenched in core psychological theories. Empathic responding is generally taught or modeled by caregivers and others in the social milieu throughout our development. The quality and timing of these social interactions offer an explanation for the variability we see in empathic attunement and responding in individuals. Attachment theory contends that empathy is learned first through experiencing caregiver empathy and reflective responding (Bowlby, 1998). Empathic responding may be modeled by caregivers and others in a child's environment. In healthy development, the child's attunement to empathic cues is positively reinforced, thus reinforcing a pattern of social relatedness. If the child does not receive reciprocation or reinforcement for this emotional attunement, this skill develops to a lesser degree if it develops at all. Attachment theorists describe a window in early childhood in which the child either attaches to a caregiver or loses the capacity to attach. Much of the attachment process --mirroring, attunement, responsiveness, and eye contact--can be

conceptualized as empathic responding. Most attachment clinicians theorize that if this attachment relationship is not internalized at an early age it is difficult to replicate in later life and the client is then vulnerable to a myriad of devastating relationships and interpersonal problems including reactive attachment disorder, anxiety, depression, and borderline personality disorder (Ainsworth & Bowlby, 1991).

Development of cognitive abilities contributes to mentalizing or the theory of mind (Apperly, Samson, & Humphreys, 2009). This developmental milestone allows us to predict the thoughts and internal state of another. This process, which is part of healthy separation and individuation, requires an understanding that others have different thoughts and internal states from one's self (Mahler, 1968). The ability to conceive of other people as separate individuals and to understand their thoughts is linked to the ability for abstract thought and the development of executive functioning. This skill refers to the abstract ability to understand another person's mind and thoughts (Frith & Frith, 2005). This is strictly a cognitive process and most often linked to the cognitive, rather than the affective, components of empathy.

In addition to our early childhood experiences with caregivers, peer relationships appear to be an important medium through which individuals develop empathy (Eisenberg & Fabes, 1998). The supportive parenting that is a central feature of secure parent-child relationships creates an emotional climate that can foster the development of empathy and reciprocity within peer relations (Garber, Robinson, & Valentiner, 1997).

Social interactions can influence not only the development of empathy but also the social reinforcement of expressions of empathy. The cultural, developmental, and familial influences on empathy vary according to whether or not empathy is modeled,

taught, or reinforced (Webster-Stratton & Reid, 2004). Static factors, such as gender and intelligence, may also play an important role in the development of empathy. Gender is often linked to empathic responding and females are generally thought to have higher levels of empathy based on their attunement to relationship factors. This correlation is supported by Baron-Cohen (2003), whose neuroimaging studies suggest clear gender differences in emotional responsivity and empathic ability. In addition to gender, empathy may be correlated to intellectual abilities, specifically the developmental process of abstract thought.

Intelligence has been discussed as a correlate to empathy as a higher capacity to learn, analyze and discriminate between subtle differences is more likely in individuals with higher intellectual capacities. Because social interactions are complex, having higher levels of analytical and discriminatory abilities may provide for an increased capacity to perceive and respond empathically. The concept of emotional intelligence is central to empathy and higher emotional intelligence is largely conceptualized in the components of empathy. The concept of emotional intelligence gained popular notoriety in Daniel Goleman's 1995 book *Emotional Intelligence*. Emotional intelligence is described as an alternate competency that explains the variance in interpersonal functioning that is not accounted for by intelligence or other constructs. The author explores the importance of emotional attunement and social skills as crucial human qualities and skills. Goleman describes empathy as the foundation of emotional intelligence. Such attention to the individual differences in empathic responding, the consequences of high and low empathy, and the differential responses to empathy training programs inform the current study.

According to Marshall and associates (1995), empathy training programs should target cognitive recognition skills like reading nonverbal signals, subtleties of body language, and facial expressions, to decipher the emotional state of another. Marshall's research supports empathy and social skills training programs for criminal and sexual offenders similar to the training of autistic children that emphasize recognizing emotions and learning prosocial responses to social cues.

Marshall conceptualizes the origins of empathy deficit as the initial lack of recognition of the emotions of another, while Graziano, Habashi, Sheese, and Tobi (2007) describe an alternate pathway. In this sequential model of empathy, an observer deciphers and feels the initial distress of another person automatically. Following this feeling of distress, those individuals low in empathy will seek to escape the situation to alleviate their distress. Individuals high in empathy will use emotional regulation skills to deal with their own distress, and quickly shift their energies to formulate a plan, manifest as some kind of helping behavior or empathic response, to alleviate the other person's distress. The authors report that the initial emotional response is automatic and does not require higher cognitive processing or recognition skills. An empathic response requires the ability to tolerate distress and make use of this negative event. According to this empathic pathway, increasing emotional regulation skills would increase empathic responding. This theory is supported by the neuroimaging and physiological research demonstrating the automatic responsiveness of humans and primates to the emotional state of another. Empathy has been found to be an essential part of human social responsiveness, with a strong psychological and social basis as well as a neurobiological component.

Studies on the function of different areas of the brain suggest that empathic responses are linked to neural activity in the amygdala (Blair, 2008), the medial prefrontal cortex, the temporal-parietal junction, and the temporal poles (Frith & Frith, 2006). Schulte-Ruther (2008) found that humans have a human mirror neuron system (hMNS) that promotes autonomic mirroring responses, possibly providing clues to the origins of empathic responding. Dapretto and associates (2006) found that the amount of frontal lobe activity in our brains is positively correlated with empathic ability; individuals with autistic spectrum disorders have decreased activity in their frontal lobes that is negatively correlated with symptom severity. This human mirroring system is believed to be the neural basis of the observer's recognition of, and resonance with, the emotional state of another individual (Gallese, 2003). Neuroimaging evidence also suggests that a mirror neuron system is linked to the ability to infer the intentions and emotions of other people (Iacoboni & Lenzi, 2002).

The amount of activation in the inferior frontal cortex during the empathy-related attribution of emotions is correlated with individual empathic ability. Deficits in empathic behavior and other forms of socioemotional processing may be related to a deficient hMNS. For example, Dapretto and associates (2006) demonstrate that individuals with autism (who often have difficulties in the expression and understanding of emotions) have reduced activation of the inferior frontal cortex during the observation and imitation of facial expressions. This idea is also supported by research on primates, especially monkeys (Bock & Goode, 2007). The capacity for empathy seems to be both hard wired into our brains and influenced by our social environment.

Empathy, the capacity for projective identification with another, usually means that one projects one's self into another to understand what the other is feeling. Empathy may also mean any ability to project oneself into something. The creation of art has been described as the projection of the artist into their work, creating a representation of their internal state outside of themselves (Root-Bernstein, 1987). Viewing is described by the authors as an emotional event: we connect with the artist's outward representation and this evokes a feeling in us that can be described as an empathic response. The ability to imagine oneself as part of one's creative product is described in the literature on creativity

Empathy is not necessarily limited to a feeling toward another person, but may also refer to feelings toward nonhuman objects. Scientists conducting research describe identifying emotionally with their subject matter (Lovecky, 1993). For example, Keller (1983), a genetic researcher, describes an ability to become one with the plants she was researching. By knowing each plant thoroughly, she was able to understand the relationship between what she saw in the field and what she would eventually see under the microscope. This deeper emotional connection between researchers and research subjects is seen strongly when conducting qualitative research, in which qualities like reflexivity, authenticity, and empowerment are held as primary research values and the researcher is required to engage on an emotional level with the research to ensure accuracy (Denzin & Lincoln, 2000).

In many ways empathy bridges the "right brain-left brain" dichotomy, requiring both a cognitive and emotional response for accuracy. Many intellectually and artistically gifted children show exceptional empathic abilities and sensitivity to others that are well beyond normal developmental expectations. For example, a 10-year-old girl

with an IQ in the 170s was gifted in producing creative and complex origami. She explained that her designs began with a feeling she had inside herself that translated into a visual/kinesthetic form that was a part of her. From this highly personal identification and visualization, she could create nearly any form because she “was the form” (Lovecky, 1992, p. 23). In another example, a 9-year-old boy, with an IQ over 200, described “putting himself on paper” as he drew a continuous comic strip in which he was the central superhero (Silverman, 1993, p. 55). He described the process as reciprocal, and reports both feeling and acting as the hero and that the hero knew him. Another boy understood nuclear physics as a process of merging with the subatomic particles and feeling his relationship with space, energy, and other particles (Piechowski, 1991). In a final example of the interconnectedness of intelligence, art, and empathy, a 3-year-old child described “falling into the music” after coming home from hearing classical music. The child had memorized the entire pieces of music after a single exposure (Feldman, 1986).

### Importance and Function of Empathy

Based on social, psychological, and biological studies of empathy, it appears that humans have both a predisposition to, and natural affinity for, the development of empathy. Yet, we see significant variation in how empathy is expressed or felt. Some individuals seem to have naturally high levels of trait empathy while others may be sorely lacking in empathic understanding and expressiveness. Lack of empathy has been linked to disrupted attachments, trauma, neglect, or other life experiences that preclude the development or expression of empathy (Garber, 1997). Empathy helps individuals establish and maintain friendships (Del Barrio et al., 2004; Hay, 1994); enhances

satisfaction in intimate relationships (Davis & Oathout, 1987); improves the quality of family relationships (Guernsey, 1988); and is positively associated with family cohesion, parental support, and communicative responsiveness (Henry et al., 1996). Additionally, empathy is related to social intelligence and can serve as a moderator for all forms of aggression in adolescence (Bandura, 1999; Burke, 2001; Feshbach, 1987; Jolliffe & Farrington, 2006a; Miller & Eisenberg, 1988). Many of the most promising parenting and child abuse prevention programs emphasize empathic responding as a core component to reduce externalizing behaviors and improving attachment (Wiehe, 1997).

A lack of empathy, conversely, implies the inability to view the world from other individuals' perspectives or to feel sympathy toward their suffering (Davis, 1994) and a predisposition toward prejudice (McFarland, 1998). Impairment in empathy is a primary diagnostic characteristic of pervasive developmental disorders, like autism, and psychotic disorders, like schizophrenia, and has been strongly linked to aggression (Wiehe, 1997), psychopathy (Blair 2008), criminal behaviors, and sexual offending (Varker et al., 2008). Low levels of empathy in adolescents are also associated with bullying others (Endresen & Olweus, 2001; Jolliffe & Farrington, 2006a). In contrast, high levels of empathy in adolescents are positively associated with helping behaviors toward victimized schoolmates (Gini et al., 2007) and, more generally, with prosocial and helping behaviors (Davis, 1994; Eisenberg et al., 2006; Eisenberg & Miller, 1987; Hoffman, 2000).

The importance of empathy can be seen in people with empathy deficits, like sexual offenders and criminals. If a person initially fails to recognize that a behavior causes emotional distress in another person, they are missing a main preventative safeguard that would prevent further harmful behavior towards that person. If empathy

deficits result in the failure to recognize distress, this could facilitate harmful behavior. Many practitioners who work with antisocial and sociopathic clients propose that psychotherapy (with all its empathic responding and emotional education components) is not only ineffective but can actually be dangerous (Samenow, 2002). In therapy, true sociopaths will increase their cognitive empathy and hence be more socially savvy and more able to manipulate others. This creates the possibility for a kind of Machiavellian cognitive empathy (Andrew et al., 2008) that can be used to harm and manipulate others. While this concept runs counter to most conventional uses of the term empathy, this more inclusive definition of empathy informs our study of empathy training programs with several populations of criminals and individuals who have a history of hurting others. Providing an intervention that is intended to increase empathy but is not guided by research could potentially create more socially adept, and hence more effective and dangerous, criminals.

Varker and associates (2008) describe the relationship between empathy and sexual offending using meta-analytical techniques. The development of empathy is commonly recognized as a central treatment goal for adolescent sexual offenders, with a national U.S. survey finding that 94% of programs treating male sexual offenders included a component of empathy training (Freeman-Longo, Bird, Stevenson, & Fiske, 1995). There is debate in the literature as to whether offenders possess general (i.e., nonperson-specific) empathy deficits, victim empathy deficits (i.e., deficits towards potential victims such as women and children) or victim-specific empathy deficits (Fisher, Beech, & Browne, 1999; Marshall et al., 1995). Without a clear conception of the nature of offender empathy deficits, it is extremely difficult to develop effective, useful

treatment programs. Although many current treatment programs teach sexual offenders victim empathy, it has been argued that sufficient empirical support does not currently exist to support victim-specific empathy training as a specific treatment process (Burke, 2001; Pithers & Gray, 1996). Victim empathy deficits in sexual offenders are considered to be empathy deficits for specific classes of potential victims (e.g., women or children). For example, Finkelhor and Lewis (1988), suggest that an inability to empathize with children in general allows child molesters to be able to sexually abuse their victims. Similarly, Barbaree, Marshall, and Lanthier (1979) asserted a failure to recognize and feel compassion for a woman's distress reduces rapists' inhibitions and allows rapists to become sexually aroused during an attack.

While victim-specific empathy deficits are generally considered to be empathy deficits in sexual offenders, Fernandez, Marshall, Lightbody, and O'Sullivan (1999) suggest that because victim-specific empathy deficits often occur in the absence of other general empathy deficits, we can consider them more as cognitive distortions. By using this distortion, the offender protects himself from the negative judgments of himself and others, allowing him to continue offending without being restrained by sympathy for his victim. The role of cognition in suppressing empathy provides additional rationale for combined cognitive and affective empathy education. Cognitive distortions are the automatic thoughts that lead to offense-supportive beliefs, such as *all children benefit from sex*. An offender's tendency to interpret children's behavior in a sexual way, using selective logic and a narcissistic framework, makes a victim's pain or distress easy to dismiss (Ward, 2000).

### Review of Empathy Training Programs

A large volume of literature addresses debates on the true and theoretical definitions of empathy as well as correlational studies linking empathy to criminal behavior, aggression, social functioning, and a variety of social and psychological ills. A smaller literature focuses on evaluating empathy training programs for their effectiveness. Only a handful of studies have systematically evaluated empathy training programs to determine which specific parts of a program contributed to overall effectiveness. One of the barriers, as discussed in previous sections, has to do with the true nature of empathy. Some researchers have proposed that empathy is organic in origin and part of our neuroanatomy (Frith & Frith, 2006). Others discuss empathy as a social skill that is learned and reinforced in early childhood (Bowlby, 1998). Still others, working in the field of criminal behaviors, describe empathy as a cognitive distortion or *thinking error* (Jolliffe, 2004). Many of these theorists and researchers propose that our capacity for empathy is largely static and difficult to change, exemplifying the classic *either you have it or you don't* philosophy. The conceptualization of empathy as a static factor has been generally disproven through research on empathy training programs, as described below.

Social learning theory provides a theoretical basis for empathy training programs intending to teach empathy to individuals with social or behavioral problems who are thought to have empathy deficits. Social learning theory posits that people learn from one another, through observational learning, imitation, and modeling (Bandura, 1969). Learning theory may account for the variance in empathy training programs, as many target different theoretical constructs of empathy, employ various empathy intervention

techniques, and assess different interventions. Research has shown improvement in individuals previously believed to have little or no capacity for empathy, like persons with autism (Baron-Cohen, 2003) and even hardened criminal offenders (Varker et al., 2008).

A narrative review of empathy training programs was conducted to inform the development of better child abuse prevention programs (Wiehe, 1997). This review relates specifically to parenting, and highlights the relationship between empathy and child abuse. In this study, abusive parents showed less empathy than nonabusive parents and empathy was a moderator in preventing child abuse. In concluding, Wiehe recommends that empathy training should be a part of all parenting programs and suggests a general framework for providing empathy training for parents. In addition, the author recognizes the need for better and ongoing evaluation of empathy training programs. Since this review, there have been several new reports published on empathy training programs and the next logical step to understand empathy training is to conduct an aggregate statistical analysis of program effectiveness that will build upon the results of Wiehe's review (1997).

A systematic review and meta-analysis of empathy training for medical students (Stepien & Baernstein, 2006) detailed several empathy training programs that target medical students and interns. The review revealed a positive effect in most empathy training programs. Communication skills programs that address the behavioral, or cognitive, component of empathy yielded the strongest effects. Unfortunately, the validity of the review is questionable because most studies included in the analysis have a small sample size, there is little consensus in defining empathy, and there is inconsistency

in outcome variables. These methodological challenges are consistent with the challenges in the present study.

There are three meta-analyses on empathy training with psychotherapists and helping professionals (Baker, 1990; Baker & Daniels, 1989; Hill & Lent, 2006). All three found that empathy training has moderate to large effect sizes. All three studies report limitation in the ability to draw conclusions based on significant heterogeneity of effect and poor study quality.

In the most recent of these studies, Hill and Lent (2006) included 14 studies and 526 participants, the majority of whom were undergraduate or graduate students in the social sciences. The study found that multi-method interventions were more effective than single methods, and that modeling empathy was the most effective training method.

Empathy enhancement programs have been used extensively in rape prevention programs on college campuses. One study of a rape prevention program (Berg, 1999) found no significant difference between the untreated control group and the treatment groups in a post-intervention empathy scale. Other programs utilizing empathy interventions with similar populations (Foubert, 2006; Lee, 1987; O'Donohue, 2003), however, found positive results for such interventions as a means to increase empathy and decrease future sexual assaults. These conflicting findings in the literature further illuminate the need for the present study.

The variance in empathy training programs, populations, and results of empathy training programs raises the question: *Can empathy be trained?* Monroe (2006) explored this question in a qualitative pilot study, conducted in a college course about individual differences, prejudice and discrimination, in which “students learned to combat prejudice

by thinking themselves into another's position" (p. 61). The author found that students who were exposed to information about prejudice and discrimination, who were intimately engaged with others different from them and who had an opportunity to explore their personal history, increased their observed and self-reported empathy. Monroe (2006) used a highly intellectual training program that included reading in philosophy and psychology and reviewing current human rights violations. This study illuminates two of the complexities in training for empathy and the need for further study of these programs: (a) although this program provided students with what appears to be an effective program for increasing empathy, the highly intellectual interventions may not have the same impact on many of the common targets of empathy training programs like juvenile offenders, adult sexual offenders, or domestic violence perpetrators; and (b) the study methodology was inadequate to determine the real impact of the empathy interventions. The author did not collect data or utilize an experimental method, but relied upon student comments and written assignments to qualitatively assess changes in empathy.

Victim Impact Training programs (VIT) are used often in offender rehabilitation programs to induce intense emotional impact, shame, guilt, and empathy in the offender in order to reduce future offending (Jackson, 2009). The VIT programs often include direct victim confrontation, victim panels, or victim mediation programs supplemented with psycho education about victimization. As with most empathy training programs, the VIT programs struggle to conceptualize the programs, measure reliable change, and produce outcomes. The extant research on criminal offenders reliably demonstrates that offenders have lower levels of empathy toward victims; however, the research on the

effectiveness of empathy training programs, and the relationship to recidivism, remains inconclusive.

Zucker (1985) describes a time-ordered process for increasing empathic responding. In this research, a short-term empathy training group was used with the goal of increasing overall empathy. In the study, empathy scores declined initially as the participants increased their listening skills. This attention and attunement to others supports prior research about the development of empathic responses and the multifaceted nature of empathy. It also supports findings from other studies that describe a strong time effect in empathy training programs, with longer programs being more effective (Jackson, 2009).

Steibe (1979) describes an empathy training program for 62 Roman Catholic nuns, in which the major effects predicting increases in empathy were the trait levels of empathy (pretest scores), the age of the participants, and the participants' perception of the trainer's empathy. In this study, the participants with high initial empathy, young age, and those who felt their trainer was empathic, had greater increases in empathy as a result of the training. In this study, participants were randomly assigned to either a 6- or a 12-hour empathy training program and analysis revealed no significant time effect between the two groups. Based on the findings about an age effect, the authors speculated that younger participants may have incorporated new skills and knowledge more easily. Baseline, or trait, empathy is positively correlated to increases in empathy scores, leading to the conclusion that some individuals are naturally empathic rather than that empathy training programs are effective. In this study, the time spent in the training

program was not significant in contrast to other studies that have found that longer interventions are more effective.

Many psychosocial interventions target empathy development, but as yet there are no guidelines for the provision of effective training. Most programs have been developed based on theoretical models and few have been systematically evaluated (Wiehe, 1997). Many programs demonstrate promising results and seem to have developed training models that show direct improvement on an empathy scale as well as decreases in aggression, bullying, criminal offending, and rape-supportive attitudes. Other empathy training program evaluations show little or no effect or have inferior evaluation strategies, leaving readers to evaluate program effectiveness based on personal experiences and assumptions about empathy. An analysis of the moderating variables that influence outcomes for empathy training programs suggests that statistical analysis is the best way to inform program development and understand the factors that contribute to effective programs.

#### Review of Possible Moderators to Empathy

In addition to assessing the effectiveness of various empathy training programs, the present study will analyze training program and client variables to explain heterogeneity in effect sizes. In a meta-analysis these variables are usually referred to as moderators, as they account for, or moderate, the overall effect size of the intervention. The following section will discuss possible moderators of empathy training programs, as described in the literature, and provide a rationale for their inclusion in this study. The three main categories of moderators in this study are participant/client characteristics, program characteristics, and outcome variables used to assess change. In addition, the

year of publication will be coded and assessed as a moderator to identify trends in empathy training programs over time to describe the prevalence of published studies on empathy training. These moderators are more fully described in the following sections.

### Participant and Client Moderators

Client variables include usual demographic variables (e.g., gender, age, ethnicity), in addition to important client differences like presenting problem, level of dangerousness, voluntary status, treatment setting, presence of mental health diagnosis, and cognitive impairment. These variables are described in more detail in this section.

Age. The age of the participant is important because the ability to empathize with others is an abstract construct that may vary with age, level of cognitive development, and life experience. Empathy is thought by some authors to reach its highest developmental stage during late adolescence (Hoffman, 1987). Research on empathy as a construct highlights the role empathy plays in the acquisition of social competence during adolescence. As adolescents negotiate expanding social roles they are acutely attuned to the perceptions and emotional states of others. This heightened sensitivity has been conceptualized as high levels of empathy. Steibe (1979) found that younger women demonstrated an increased responsiveness to an empathy training program. The author speculates that younger participants were more likely to acquire new empathic skills more easily. The study found that age was a significant moderator in the effectiveness of the empathy training program and that younger women increased their post-intervention scores more than older women.

Gender. Empathic responding is often linked to compassion, sensitivity, and other characteristics typically attributed to females. The empathy literature contains

studies linking base levels of empathic responding to gender, but the findings are inconsistent. For example, females, compared to males, have higher levels of empathy and may be more likely to pick up on emotional cues (Baron-Cohen, 2003). The higher levels of baseline empathy may contribute to the differential effects of empathy training programs. In addition, males tend to commit more violent and sexual crimes (Varker et al., 2008), the most serious of the behaviors targeted by empathy training programs. Although females are believed to be more empathic, research indicates that this depends on how empathy is measured. Generally, research indicates that if you ask for self-reported empathic responses, females show higher levels of empathy (Norris, 2006). However, when physiological stress responses are measured or behaviors are observed unobtrusively, no gender-related differences are found.

Bohart and Greenberg (1997) report that the literature reveals only modest gender differences between girls and boys, with girls having higher levels of empathy. They report several functional differences between male and female empathy, including that female empathy emerges from a socialization history that is punctuated by positive parenting rather than a family constellation fraught by distress, conflict, and discord. It may be that empathy in boys may emerge from different socialization experiences and therefore manifest in less overt expressions of empathy. This hypothesis is supported by literature on risk and protective factor differences between boys and girls (Bohart & Greenberg, 1997). Girls are more sensitive to family violence and discord than males, presumably because girls rely more heavily on their socialization experiences at home while boys have a more external focus. This same research reports that low empathy

levels are not indicative of maladjustment or emotional problems in men but are strong indicators of such problems in women.

Owen-Anderson, Jenkins, Bradley, and Zucker (2008) report significant sex differences in the display of empathy, with females consistently demonstrating higher levels of empathy from early childhood through adulthood. Females show significantly more empathy than males on parent, teacher, and self-reports of empathic and sympathetic behaviors (Eisenberg, 1996; Hastings, 2000). A study of toddlers found that girls displayed significantly more prosocial responses to simulated distress events than did boys (Hill et al., 1976). Aggregating observer, teacher, and parent reports in a study of 5-year-olds, Strayer and Roberts (2004) found that girls were more empathic than boys and that gender accounted for 16% of the variance. In another study, Roberts and Strayer (1996) measured facial and verbal responses to emotionally evocative videotapes in children and found that girls often had higher empathy scores than boys on both measures (a method that typically minimizes gender differences). The results from empathy studies contrasting males and females are fairly consistent, demonstrating that females are more likely than males to respond empathically and to endorse an empathic response. Several questions emerge from this difference.

The first is a discussion of the social construction of gender and the social benefits as compared to the social costs of responding according to traditional gender roles. The second concerns the modes of measurement and the validity of the instruments used to assess empathy. Most empathy self-report measures assess cognitive empathy and may be missing the more valid mixed construct of empathy (cognitive and affective). In addition, observer-reported assessments are subject to the same observer gender role bias

that may contribute to girls displaying more empathic responses. Observer reports rely heavily on a behavioral and verbal display of empathy, which is also subject to gender role bias, as males tend to be less emotionally expressive.

In addition to persuasive biological evidence that females generally have higher capacities for empathy, observations of behaviors support females having higher levels of empathy than males throughout the lifespan. In a study that compared boys and girls between 4 and 8 years of age, researchers found that girls scored higher on both self-reported parent-reported empathy scores than any of the other groups. The study compared girls with three groups of boys: boys with Gender Identity Disorder, boys with externalizing behaviors, and a control group. The study reported that the girls consistently scored higher in empathy, followed closely by the boys with Gender Identity Disorder. The boys with externalizing behaviors scored significantly lower than any of the other groups, supporting the theory that empathy both contributes to better social skills and inhibits aggressive behaviors (Owen-Anderson, 2006).

Ethnicity. It is important to account for cultural and ethnic differences in any research study, especially considering the goals of this study, which are to inform practice and generalize results. There is little research on the relationship between culture, ethnicity, and empathy; none on an interaction of ethnicity and empathy training. It is possible that interventions may work for one ethnicity due to their grounding in a cultural norm or phenomena, while others may not.

Cognitive functioning. There is some evidence that empathy may be a function of intelligence and that a baseline level of intellectual functioning is required to attend to, translate, and respond empathically. It is important to investigate how the effectiveness

of empathy training programs may vary according to a client's level intellectual functioning. Broom (2002) studied a small group of adjudicated adolescents to determine the relationship between empathy, age, and intelligence. She discovered that overall, age and intelligence had no significant correlations with empathy. While the correlations were insignificant overall, there were some surprising within-group findings: the youth with borderline mental retardation (IQs ranging from 55 to 70) demonstrated higher perspective-taking empathy scores.

Diagnostic presentation. Many clients receiving empathy training present with impairments in empathy as well as mental health problems, either in addition to or related to a mental health disorder. Recall that Baron-Cohen (2003) postulates that females have a more empathic brain while males have a more systemizing brain. The systemizing brain responds to the behavior of others by analyzing the input-operation-output of the relation and inferring rules about the system to respond. Females respond by inferring mental states and responding with appropriate emotion. Baron-Cohen argues that autistic spectrum disorders are the manifestation of an extreme male brain that has impaired abilities to empathize, but is a strong systemizer. Autistic spectrum disorders are much more common in males, lending support to the idea of a hyper-masculinized systemizing brain (Baron-Cohen, 2006). This theory lends support to the commonly held belief that women are more attuned to emotional content and relationship factors while men have stronger analysis and manipulative abilities.

Presenting problem. Most empathy training programs target potential future or previous hurtful behavior towards another person. Theoretically, the empathy training program seeks to increase empathic response and thereby decrease negative behaviors

and increase positive and prosocial behaviors. Understanding how interventions affect different target behaviors is central to this study.

Level of prevention. Good treatment programs are targeted for specific populations. A tertiary prevention program that is highly effective in reducing harmful behaviors may be ineffective and even harmful if implemented as a primary prevention program. Assessing where an empathy training program falls in the prevention continuum is an important variable.

Voluntary status. The voluntary status of participants is a variable that may affect motivation to attend, participate, and internalize an empathy training program. One of the myths about interventions is that mandated treatment does not work, but we have good evidence that treatment works regardless of the mandated status of participants (Farabee et al., 1998). It is therefore important to assess voluntary status as a moderator.

### Program Moderators

The program components are another main category of variables that are important to study. The theoretical foundation of the program, the modality of the training, the dosage of treatment, the specific program ingredients and the number of different ingredients, the degree to which the program addresses the interactive nature of empathy, and whether the program stands alone or is part of a larger treatment program are all variables of interest in this study.

Program's theoretical foundation. This moderator assesses the degree to which the program is grounded in the three major theoretical models of empathy: cognitive empathy, affective empathy, or victim impact training. Nearly all empathy training programs fall into one or more of these categories, which constitute the major ways that

we think about empathy. Cognitive and affective empathy have been described extensively in previous sections and victim impact training is a special category of empathy training that targets criminal offenders and may use both cognitive and affective techniques.

Interactive empathy. Zaki (2008) describes empathy as a mutually interactive process that cannot exist alone. Zaki's research demonstrates the importance of conceptualizing empathy as an interactive process that can be measured only after considering the empathic availability, receptiveness, and expressivity of both parties. This framework provides an important possible moderator of an empathy training program that has been largely neglected in the literature to date. The level of interactive empathy involved in a program could be an important factor in training effectiveness and is also an important framework for comparing the ways that empathy is measured as an outcome variable. Most assessments of empathy rely on the measurement of one individual. This theoretical framework, however, reinforces the concept that empathy is a developmental process that must be learned through interaction with another person, either the trainer or another participant. This contention is supported by Steibe (1979), who found that the trainee's perception of the trainer's level of empathy was an important predictor in the outcome of the empathy training program. Much of the variability in empathic responding we see in individuals may be accounted for by the quality of early childhood interactions with caregivers (Bowlby, 1998). This model provides support for a reciprocal corrective and healing process that may occur in a safe, therapeutic relationship and may be replicated in effective empathy training programs.

Modality of treatment. The treatment modality, whether individual, couples, family, or group, counseling, is an area where considerable variation exists in empathy training programs. Understanding how modality contributes to outcomes is important for developing more effective, cost efficient programs. For instance, if group interventions are as effective as individual interventions, this could represent a significant resource savings to programs.

Program dosage. The dosage of an intervention contributes to the outcomes in many areas of practice. In the case of empathy training programs there is a large variation in the overall program dosage, including weeks in treatment, total number of sessions, and total number of contact hours. It is important to understand the lowest dosage threshold necessary to see a significant change in empathy and also when saturation occurs, wherein continued treatment no longer realizes significant gains in outcomes. Additionally, it is important to understand how frequency and number of treatment sessions contribute to outcomes.

Setting for treatment. Empathy training programs are conducted in various community and institutional settings. Little is known about whether empathy training programs delivered in institutional settings are more or less effective than those conducted in outpatient settings.

Role of intervention. Empathy training programs are often offered as stand-alone programs that are specifically targeted at remediating empathy deficits. Though this is not always the case as programs can be part of larger rehabilitation efforts for criminal offenders, violence prevention programs, or parenting or social skills training programs.

Understanding how empathy training interfaces with these larger programs is an important variable to consider.

### Program Ingredients

The impact of specific program ingredients, and how they contribute to overall effect, is one of the most important questions in this research study. Considerable variability in program ingredients exists and identifying those that are more strongly connected to desirable outcomes could guide practice decisions. Role plays, victims guest speakers, watching videos, visualization, listening to victim statements, psycho-education, independent reading materials, positive empathy, and culturally or spiritually grounded training components are hypothesized to contribute to the heterogeneity we see in program effectiveness. This study will examine if one or more ingredient contributes more strongly to desirable outcomes.

### Empathy Scales (Dependent Variable and Moderator)

The ways in which outcomes are measured contributes to the perceived success or failure of the intervention. In the extant literature on empathy training, there are several commonly used instruments for assessing empathy. An analysis of the outcome variable as a moderating variable is included in the present study as a means to understand how different scales may capture different components of empathy and contribute to the overall effect of a given study. In addition, many treatment programs target empathy as a proxy for other outcome variables like criminal recidivism, social skills, rape-supportive attitudes, and relationship satisfaction. This study will also code for data regarding the durability of outcomes or the longitudinal effects of the empathy training program.

While most studies report data on changes postintervention, follow-up assessments of the outcome variables are of interest to understand the long-term impact of empathy training programs. The five most commonly used instruments for measuring empathy are reviewed below.

Several reviews of instruments to measure empathy are provided in the literature (Albiero, 2009; Jolliffe & Farrington, 2006b; Mehrabian, 1997). As noted, the heterogeneity in operationally defining empathy, the components of empathy, and the sequence of events leading to an empathic response create problems in reliably measuring empathy. Several scales have been developed independently and subjected to psychometric analyses. Most empathy scales assess empathy according to cognitive, affective, or combined empathy components. How empathy is measured will be a moderator.

One of the most commonly used empathy scales is the Carkhuff and Truax Empathy Scales (Truax & Carkhuff, 1967). This was the first empathy scale to reliably operationalize empathy into specific behaviors and statements and was specifically developed for use in teaching empathy to students in the helping professions. This scale is completed by an observer or team of observers as they watch participants respond to a stimulus person or in a role-play setting. The scale originally developed as a nine-point scale but was later revised (Carkhuff, 1969) as a five-point empathy scale with a rating of one representing low levels of empathic responding and five as high levels of empathy. The Carkhuff and Truax scales have been used and validated in research (Truax & Carkhuff, 1967) and continue to be widely used both in social work education (Hepworth et al., 2010; Larsen & Hepworth, 1978) to teach and train in empathy. The Carkhuff and

Truax scales are no longer the instruments of choice outside of professional training programs.

The most frequently used self-report empathy scales are Hogan Empathy Scale (Hogan, 1969), the Questionnaire Measure of Emotional Empathy (Mehrabian & Epstein, 1972), and its more recent version, the Balanced Empathy Emotional Scale (Mehrabian, 1996), the Interpersonal Reaction Index (Davis, 1980), and the Basic Empathy Scale (Jolliffe & Farrington (2006b)). The most popular and widely used of these is the Interpersonal Reactivity Index (IRI). The IRI consists of four subscales measuring four different dimensions of interpersonal reactivity (Davis, 1980). Davis' model views empathy as a set of distinct, but related constructs, two of which are cognitive dimensions and two of which are emotional: (a) fantasy, which is the tendency to project one's self into the feelings and actions of fictitious characters portrayed in movies, book, etc; (b) perspective taking, which is the tendency to spontaneously adopt another person's psychological viewpoint; (c) empathic concern, which consists of other-oriented feelings of sympathy and concern for unfortunate others and; (d) personal distress, which consists of self-oriented feelings of personal anxiety and unease in charged interpersonal settings (Davis, 1983).

Although the IRI is used widely, it has clear limitations common to most assessments of empathy: the underlying theoretical model may be questionable and it may not adequately capture the complexities involved in the empathy process. Several factor analysis studies have been conducted to examine the IRI and few have confirmed the validity of Davis' four-factor model (Albiero et al., 2006; Litvack-Miller, McDougall,

& Romney, 1997). Most analyses find that the original model does not provide an adequate data fit (Cliffordson, 2002; Poulos, Elison, & Lennon, 2004).

Jolliffe and Farrington (2006b) also highlight two serious limits of the IRI. The first, which can also be observed with the Questionnaire Measure of Emotional Empathy (QMEE; Mehrabian & Epstein, 1972), is that the construct of empathy is easily confounded with sympathy--two distinct, albeit closely linked, constructs (Hoffman, 2001). In addition, the IRI's assessment of the cognitive components of empathy mostly focuses on the ability to take another person's perspective, rather than the ability to recognize the emotions felt by another person. This is a significant shortcoming, given that the emotional aspect of empathy has consistently been considered a fundamental component of empathy since the earliest theorizations (Feshbach & Roe, 1968). Moreover, both the IRI and the QMEE were developed on a population of university undergraduates; which casts doubt on their reliability when used with a more diverse population.

A further problem, frequently accompanying the use of the IRI, is linked to the method proposed for calculating a total score, which is obtained by summing individual subscale scores and is considered an index of high or low empathy. Yet, as D'Orazio (2004) discusses, this use of a total empathy score is inappropriate because the IRI's subscales are not positively correlated (Albiero et al., 2006; Davis, 1983; Eisenberg & Fabes, 1990). While the IRI has limitations, with proper use it is a useful instrument to assess empathy.

Other frequently used scales, like the Hogan Empathy Scale (HES), which assesses primarily cognitive empathy, and the QMEE, which measures the emotional

component of empathy, fail to provide a more inclusive measurement of empathy. This results in a discrepancy between the goals of empathy training programs (to improve overall empathic responding) and the outcome instruments that gauge effectiveness. The psychometric assessment instruments used currently do not correspond to more recent theoretical models underscoring the multidimensionality of the construct; nor do they meet the theoretical and methodological imperative to account for the ways in which affect and cognition interact to give rise to empathic responses (Davis, 1994; Hoffman, 2000).

Jolliffe and Farrington (2006b) constructed and validated the Basic Empathy Scale (BES), a new scale for measuring empathic responsiveness in adolescence. Starting with a more inclusive and multidimensional definition of empathy, the scale assesses both emotional congruence (affective empathy) and the understanding of another person's emotion (cognitive empathy). An indication of the comprehensive nature of this definition of empathy is reflected in the fact that the BES measures empathic responsiveness to positive emotions, whereas previous questionnaires had mainly detected responsiveness to negative emotions. Responsiveness to other people's positive emotions can represent a crucial aspect of empathy assessment during this developmental period. Specifically, the opportunity for adolescents to experience and share positive emotions (i.e. happiness, joy) with their peers plays a fundamental role in enhancing both their physical health and sense of well-being (Mahon et al., 2005; Natvig et al., 2003) and in creating further positive experiences and a strong desire to contribute to society (Magen, 2004). This holistic view, that captures the full range of emotional responsiveness, has been identified as a significant gap in the literature and study of

empathy. While this instrument has been validated for use only with adolescents, it holds promise for use with other populations.

Given the crucial role of empathy in mediating interpersonal relations, social and antisocial behavior, and psycho-physical well-being, it is essential for professionals working with clients in a variety of settings to have valid and reliable procedures available for measuring empathic responsiveness for both intervention and research purposes. Although there has been an upsurge in the number of studies investigating the construct of empathy and though these studies are increasingly sophisticated as far as theory and methodology, the basic approach for measuring empathy in the social sciences has changed little since the 1980s.

#### Summary and Limitations of Literature

The preceding literature review describes the numerous programs, populations, and variables that must be considered as we consider the central questions for the present research: are empathy training programs effective and what variables influence their effectiveness. Based on this review we have an indication of the variation in literature regarding these programs, which is confusing for a practitioner or researcher who is interested in empathy. More important, this significant but unfocused knowledge base has the potential to hurt clients.

Consider clients who have invested time and energy into an empathy program who are left with poor outcomes, not because of their own failure or lack of investment, but because of the failure of professionals to provide effective programs. It is not adequate to rely on intuitive responses or personal experiences to develop and implement programs to reduce egregious violations like sexual assault, domestic violence, and child

abuse. The perpetrators of these intensely personal and damaging behaviors are often the subjects of empathy training programs and it is important that widely used interventions are derived from scientific investigation and research. Otherwise, there is a risk of wasting resources and time and causing harm to clients and the larger community.

The responsibility to provide effective interventions for clients is a professional and ethical responsibility as prescribed in the National Association of Social Workers (NASW) Code of Ethics. One of the core values of competency requires social workers to provide effective interventions, to engage in research and evaluation, and to protect clients from harm (NASW, 1999, para. 1.04). In addition, social workers are required to critically evaluate research and current interventions to ensure that the treatment we provide is effective and based on empirical knowledge (NASW, 1999, para. 4.01). Considering empathy training programs through this lens, it is important to evaluate the numerous programs that target empathy with the purpose of preventing future harm. Among these programs, we have multiple target populations, ages, treatment modalities and outcome measurement instruments. Empathy training programs have a large capacity to improve interpersonal functioning and prevent future harmful behaviors. Conversely, if empathy training programs fail to increase empathy, especially with dangerous offenders, we thereby increase the risk of future harm.

A study of a college-based rape prevention program illustrates how a well-intentioned program can produce an untoward effect (Berg, 1999). Researchers used audiotapes of both male and female victims describing a sexual assault as part of a prevention program with college-aged males. The analysis revealed that, while neither audiotape increased empathic responding nor rape-supportive attitudes, the subjects who

listened to the female victim showed an increase in rape-supportive behaviors (i.e., encouraging females to consume more alcohol with the intention of having sex). This study illuminates a harmful effect of a well-intended but ineffective empathy training program: listening to a female recount a sexually assaultive event may increase the risk for college males to commit a rape. As a result of this study, most sexual assault and domestic violence prevention programs now use gender-matched audio and videotapes.

This finding has resulted in the development of more effective rape prevention programs. This finding is useful for this specific population; however, empathy training programs are used in a wide variety of settings and with numerous client populations. The variation in published reports of empathy training programs also creates confusion for practitioners searching for scientifically grounded interventions to increase empathy. Although there is some literature that describes specific program outcomes, these results are highly specific and this limits the applicability of these findings for most practitioners. An analysis of empathy training programs across client populations and treatments is necessary to determine if empathy can be increased through training and which factors contribute to the success of an intervention. Such analysis will provide a foundation for developing more effective programs. This research project seeks to identify which factors are most important in empathy training programs and to disseminate this information to allow practitioners to develop new programs and modify existing programs to maximize success. Successful empathy training programs will not only benefit the clients in the program but will also have an impact on others in the social milieu.

## CHAPTER 3

### METHODOLOGY

Providing effective and evidence-based psychotherapeutic interventions requires utilizing research findings to inform practice. Meta-analysis is one method to assess the effectiveness of an intervention and to understand the moderators that impact its effectiveness. Social work has typically lagged behind other professions in the use and publication of meta-analyses (Lundahl & Yaffe, 2007). As the pressure to provide efficient and evidence-based interventions increases, it becomes increasingly important to investigate empathy training programs using this methodology. This dissertation utilizes a meta-analytic strategy to examine the effectiveness of empathy training programs in increasing empathy and the moderating variables that contribute to their effectiveness.

A meta-analysis statistically combines the results from multiple studies to yield a single metric, or overall effect size, which can be used to compare different studies. In the case of intervention research, a meta-analysis may examine the outcomes from multiple interventions across studies and determine their relative effectiveness. The computation of effect size for various outcomes within the studies makes it possible to compare the characteristics of both the interventions and the clients. These moderators are then analyzed to determine how each of them contributes to the overall effect size. This section will synthesize the existing literature on empathy training and clarify the key moderators essential to effective programs.

This dissertation is a treatment effectiveness meta-analysis that summarizes the effects of a specific kind of treatment and the relationship of the treatment to other variables (Durlak & Lipsey, 1991). The search, coding, and statistical analysis is systematic in nature and governed by standards of scientific rigor appropriate for a meta-analysis, as described in this section. There are numerous research articles on empathy training programs that report findings from independent evaluations. These reports vary in their intervention methods, target populations, and evaluation methods, making it difficult, when reviewing the reports separately, to decipher the salient variables that contribute to an effective program. The strength of meta-analysis is its ability to combine the results of various studies and statistically analyze various program components. This meta-analysis yields answers to the main research questions and attempts to accomplish research goals as follows:

The research questions for this study are:

1. Do empathy training programs increase empathy for client populations?
2. Is there variation in empathy training program outcomes?
3. What moderating variables contribute to empathy training programs and what is the relative strength of these variables?

The goals of this research project are to:

1. Investigate the overall effectiveness of empathy training programs for client populations through meta-analytical methodology.
2. Assess potential heterogeneity.
3. Examine potential moderators to the overall effectiveness of empathy training programs for client populations through meta-analysis.

4. Make recommendations for effective empathy training programs.

### Overview of Meta-analysis

Prior to reviewing the specific methodology used in this study, it is important to briefly review the core terminology and goals of a meta-analysis (see Borenstein, Hedges, Higgins, & Rothstein, 2009; Cooper & Hedges, 1994; Lipsey & Wilson, 2001). Meta-analysis is a method for combining and summarizing the quantitative results from independent primary studies that share a similar focus. Independent empirical studies provide information about information about specific populations, interventions tested, and outcome tools. A meta-analysis combines treatment effects into a common metric that is standardized: an effect size. An effect size refers to the magnitude of the effect or the strength of the intervention. In effect size nomenclature, an effect size of  $d = \pm 1.00$  would suggest a group difference score equal to one full standard deviation between clients in the treatment condition relative to the comparison group, or the difference between pre-post scores. An effect size of  $d = -0.50$  would suggest a group difference of one half of a standard deviation in the negative, direction. By calculating the effect size as a function of the standard deviation for each outcome reported, we standardize the scale and can make comparisons between studies even though they may use different outcome instruments with varying scales. Conventionally, an effect size in the range of  $d = 0.20$  is considered small, while effect sizes in the range of  $d = 0.50$  and  $d = 0.80$  are considered moderate and large, respectively (Cohen, 1988).

Ensuring scientific integrity, transparency, and replicability is central to any scientific inquiry (Kerlinger & Lee, 2000). This study follows the gold standard procedures, as described by Durlak and Lipsey (1991), in conducting the analysis. The steps are: (1)

formulate specific research questions, (2) search the literature systematically and sort the articles for inclusion, (3) code the studies, (4) calculate the index of effect sizes in the studies (5) select the appropriate statistical test and conduct the analysis, and (6) report conclusions and findings of study. Using the aforementioned steps as a guide, the following sections provide additional detail regarding the steps required to complete this study.

### Formulate Specific Research Questions

It is vital to any scientific inquiry to formulate research questions a priori to ensure the study has adequate specificity parameters. The research questions of this study, which have been stated in previous sections, are repeated here for clarity. In this exploratory study the research questions are:

1. Are empathy training programs effective in increasing empathy for client populations?
2. Is there variation in empathy training program outcomes?
3. If variance exists, what moderating variables contribute to effective empathy training programs and what is the relative strength of these variables?

### Literature Search

A systematic, multistep process was used to identify potential studies for inclusion in the analysis. To identify potential articles, a literature search was conducted of major electronic databases using the search terms *empathy and (train\* or interven\* or treat\* or educ\* or program\*)*. The asterisk is a Boolean *truncation* symbol that captures various versions of the root word i.e. training, trainer, trained, etc. The data bases were

searched using online portals: EBSCO was used to search *CINAHL*, *ERIC*, *MEDLINE*, and *PSYCHINFO*; CSA Illumina for *Social Service Abstracts*, *Sociological Abstracts*, *Criminal Justice Abstracts*; and OVID for *Social Work Abstracts*.

The initial search strategy was broad in scope, attempting to capture the full range of the professional literature on empathy. Following this broad search, more specific limiters are applied to the search strategy and only those studies that used some form of comparison and yield adequate statistical information are included in the study.

Many of the studies captured through the initial literature search are used to identify the theoretical foundations of empathy as well as the terms of the debate about the definition and the measurement of empathy. This broad search also includes other meta-analyses that identify key moderators and narrative reviews that provide overviews of the literature on empathy.

The initial search included all relevant databases that commonly publish research findings related to psychosocial interventions as described in subsequent sections. By first conducting a broad search and then systematically narrowing the inclusion criteria for the study, a highly representative sample of published literature on empathy training programs is generated.

One of the first major methodological decisions in any systematic review is to determine the breadth of the search strategy. Many systematic reviews attempt to capture all available literature by including all published and unpublished literature, including both peer-reviewed and non-peer-reviewed articles. This study restricts the sampling frame to published literature contained in electronic databases. In a meta-analysis, the exclusion of unpublished reports has the potential to introduce a *file drawer* bias and

artificially inflate the overall effect size (Rothstein, Sutton, & Borenstein, 2005). This is caused by the fact that studies yielding a positive result are more likely to be published in peer-reviewed journals while those yielding less significant results are more likely to end up in the researcher's file drawer (Epstein, 2004). One problem with attempting to gather unpublished reports by contacting authors, searching for research presented at conferences, or using other methods, is that there are problems with replication of the study, which could violate a key condition of the scientific method. Additionally, conducting an unscientific and largely unsystematic search for unpublished reports in an attempt to address the file drawer and publication bias may contribute to a false sense of security that all studies have been considered. Fortunately, we have sophisticated statistical tests and methods to account for missing studies and may even adjust our reported findings based on these results. The potential effect of unpublished research on the study will be assessed using statistical methods (Lipsey & Wilson, 2001) including fail-safe N, forest plots, and funnel plots. The results of these tests are included in the final report and if necessary, reported data will be adjusted accordingly (Borenstein et al., 2009).

Identified studies were evaluated for inclusion in the analysis by the principal investigator and trained graduate research assistant in conjunction with the supervisory committee. The first step was to read the abstracts of all articles captured in the initial search and eliminate any article that is strictly theoretical or describes training professionals (e.g., counselors or medical students). The remaining articles were then assessed for the presence of quantifiable results. Further details of the sorting, screening, and vetting process are detailed in Appendix A. This appendix includes the specific

search strategy, search terms, databases searched, and screening and inclusion rules, used at each step. An overview of these steps is provided below.

### Study Selection

In searching the literature, several rules were developed to identify studies that fit the goals of this study. There are no universally prescribed rules for which studies should be included in a meta-analysis (Borenstein et al., 2009). As with any research, there are generally accepted guidelines for research methods, but it is ultimately up to the individual researcher to determine specific selection criteria appropriate for the study and research questions. In a meta-analysis, the study selection and inclusion criteria are a large part of the research and deserve special attention (Petticrew & Roberts, 2006). As with the sampling strategies in primary research, the sampling frame and the decisions whether to include or exclude subjects can bias and skew the results. In a meta-analysis it becomes even more important to exercise caution when selecting articles for inclusion and to clearly describe the consequences of this selection strategy. The strategy for weighing the pros and cons of sampling frames and making decisions about inclusion criteria was guided by consultation with the supervising committee and grounded in the best literature on meta-analytical methodology (Borenstein et al., 2009; Cooper & Hedges, 1994; Durlak & Lipsey, 1991; Lipsey & Wilson, 2001; Littell, Corcoran, & Pillai, 2008; Petticrew & Roberts, 2006).

### Inclusion Criteria

Understandably, not all identified primary studies were included in the final meta-analysis. For example, some identified studies have evaluated empathy training

programs, others will be purely theoretical, and some do not provide sufficient information to compute effect sizes. To guide decision-making in this area Lipsey and Wilson (2001, pp. 16-17) propose seven eligibility criteria for the inclusion of a research article in a meta-analysis. The criteria are: (a) distinguishing features of a qualifying study, (b) research respondents, (c) key variables, (d) research designs, (e) cultural and linguistic range, (f) time frame, and (g) publication type.

Specific eligibility criteria were developed for study inclusion as follows: 1) study design must be pre-post, comparison group, or treatment vs. control; 2) must be empathy intervention; 3) qualitative studies excluded-must report quantitative data; 4) must be a client population; 5) must have at least 5 subjects in each treatment group; 6) must report an empathy scale or subscale as outcome variable.

The primary investigator screened all articles for inclusion. To promote consistency in the screening process, a minimum of 20% of the studies were double-screened by a trained graduate research assistant. Reliability statistics were calculated to assess bias in the screening process and any discrepancies were resolved through discussion. The full training and screening protocol is included in Appendix B and additional details on eligibility criteria are described in the following sections.

### Distinguishing Features

The studies that are eligible for inclusion in this study use an empathy training intervention as their primary treatment target, or as an integral part of the intervention. An empathy training program was defined as an intervention that targets an individual or group with the purpose of increasing empathic responding (affective empathy), empathic understanding (cognitive empathy).

### Research Respondents

The characteristics of research participants for included studies will be any client or clinical population with the exclusion of professionals or professional training programs (e.g., doctors, nurses, or mental health professionals). Professional training programs that seek to increase empathic responding for the purpose of improving bedside manner or building rapport are qualitatively different from programs that target violence prevention or clients with identified empathy deficits. In addition, professionals-in-training are demographically different than most client populations.

Studies with small populations and single-case studies will be excluded from the meta-analysis. Studies with fewer than 5 participants in any treatment condition will be excluded from the study.

### Key Variables

The key outcome variable of interest in this study is a quantifiable empathy (outcome) score that reflects either a group difference or treatment response score. The forms of this outcome variable may be one of several empathy scales or subscales. Several studies include a client self-report, professional rating scale, or other form of observation. There is significant diversity in empathy training programs (independent variables) and this analysis will include all methods such as role-plays, psycho-education, mirroring exercises, videos, videotaping, encounter sessions, and victim-offender programs. There were no other exclusionary criteria other than the respondents must be client populations. Key moderators included in the study fall into three main categories, program variables, client variables, and study quality.

The year of publication is also an important moderator that may help to identify trends in empathy training programs over time. The year of publication will also be a descriptive indicator of the prevalence of published studies on empathy training.

Program variables. Program variables are central to this inquiry, as they provide information about how different modalities and components of empathy training programs affect clients. The literature review section reveals wide variability in how empathy training programs are implemented and the specific training ingredients used to increase empathy. Some of the major moderators in this area are: type of empathy targeted (cognitive or affective empathy), modality of intervention (individual, group, couples counseling), dosage and length of treatment (weekly, total hours of training, whether the empathy training program is independent or part of a larger psychosocial or social skills program, and the specific intervention techniques (role play, video, psycho-education, exposure of victim, etc.) employed.

Client variables. Client variables such as age, cognitive functioning, presenting problem, setting in which intervention is provided, gender, and ethnicity are important to study because one or more may influence the relative effectiveness of a given empathy intervention. By calculating the relative strength of client and program variables, we can determine which factors contribute to effective programs, which will provide insight into the effective implementation of empathy training programs.

Study quality. The relative quality of the included studies is a key variable in meta-analysis. Meta-analysis uses primary studies as the unit of analysis to answer research questions. As such, the results of the analysis rely heavily on the quality of the research being conducted in each study. If the primary studies were conducted poorly or

have significant methodological flaws, then the findings of the primary research and any subsequent analysis of the research is suspect.

Determination of quality of an individual study is an important variable in understanding and interpreting the relative influence of a study on the overall effect size. The quality of a study can be correlated to the effect size and may help to explain the distribution and accuracy of the weighting strategy used in the analysis. With typical weighting strategies in meta-analysis, larger studies are given more weight than smaller studies. Generally, larger studies tend to have more rigorous methodology and lower overall calculated effect sizes (Weisburd, Lum & Petrosino, 2001). Smaller studies can yield much larger effect sizes, but because of small sample sizes they do not influence the overall effect as much as a larger study. Assessing the rigor of the study will provide additional information about individual and overall effect sizes for empathy training programs. Studies included in this analysis were evaluated for quality and rigor using a rating scale developed for this study, the Empathy Training Program Study Quality Rating Form. This form was developed using the Cochrane guidelines regarding the five major types of bias in intervention studies, selection bias, performance bias, attrition bias, detection bias, and reporting bias (Higgins & Altman, 2008). The study rating form is provided in Appendix B. This rating form yields both an ordinal rating of low, moderate, and high and a continuous rating of 0-20. Both categorical and ordinal values are used to assess study quality and are included in the meta-analysis as a moderator variable.

The first rating scale is the sum of all the study quality variables that yields a continuous variable between 0-20 points. Although this sum taken alone does not pertain

to any real value regarding the overall value of the study, it does provide a relative evaluation of the studies and an approximation of the study quality.

Overall study quality was also coded as a nominal variable as low, moderate, or high quality. This ordinal rating was coded after rating and summing the total study quality points. This coding of this variable considered the overall points awarded to the study; high quality = 12-20 points, moderate quality = 7-11 points, or low quality < 7 points, in addition to considering if the study had any major research design flaws. If the study had a significant design flaw, like a large attrition rate or the use of a convenience sample, it was downgraded to low quality rating regardless of the number of points awarded.

### Research Methods

Studies included in the statistical analysis, must utilize an experimental design, quasi-experimental design, or pre-post design. Studies included in the analysis are likely to be control or comparison group designs or pre-post designs that yield mean difference statistics. Correlational, theoretical, qualitative, or other experimental designs that do not report quantitative data are excluded from the analysis. The study must report sufficient statistical and descriptive data for inclusion in the analysis.

### Cultural and Linguistic Range

Only studies reported in English are included in the current analysis. Although this exclusionary criteria could result in some bias and limit the external validity of the findings, this is a pragmatic decision as the time and resources needed to translate studies pose a significant difficulty. This study includes studies completed in all parts of the

world and with respondents of all cultures and races, as long as the study can be obtained in English.

#### Time Frame

All available studies were evaluated for inclusion in this study without any pre-set limit on date range. The study inclusion criteria stipulate that only reports included in electronic data bases were eligible for inclusion. The earliest publication date for inclusion was dependent on the start date of the data base. In other words, the search included all studies contained in the electronic database from the earliest studies through the end date. The end date for inclusion was July 16, 2009, as that is when the initial search of databases was conducted. The initial search was conducted in the summer of 2009, the sorting and coding in the fall 2009, and the analysis and final dissemination in winter 2010.

#### Publication Type

All studies cited and available through a systematic search of online academic databases and abstracts are eligible for inclusion in this study. Although peer-reviewed studies are more likely to be included in the final analysis, other studies of appropriate design and rigor may be included as well. Dissertations were also included in the analysis. Although there is potential for an upward bias when utilizing only published reports (Epstein, 2004), this decision was made for pragmatic reasons and to increase the replicability of the study procedures and protocols. The decision not to include unpublished reports is based on the fact that replicability was determined to be more important than potential upward bias. In addition, publication bias testing is conducted on the included studies to determine the potential publication bias present in the analysis.

Findings are also presented with the assumption that we have not included all studies on empathy training programs, as is standard practice in many meta-analytical studies. In the end, the decision to conduct a more thorough, but less systematic search for unpublished studies could be a possible limitation of this study.

### Coding Studies

All articles selected for inclusion were coded and rated by the primary investigator and a graduate research assistant. The outcome of the coding was compared and any discrepancies resolved through discussion. A code sheet was developed for this study to capture statistical data in addition to study characteristics, program characteristics, program ingredients, and participant variables. The study code sheet is attached in Appendix C and the study quality code sheet is attached in Appendix D.

### Statistical Procedures

This section describes the statistical procedures employed in the study including the index of effect that was used to calculate the effect size, the statistical analysis including tests of significance, the frame for the analysis, and the model used for the analysis.

### Effect Size Statistic

Hedges's  $g$  (Hedges & Olkin, 1985) is used to calculate a weighted effect size in this analysis. The calculation for Hedges's  $g$  is derived from Cohen  $d$  but accounts for small sample size. The simplest way to calculate  $g$  is to use the correction factor  $J$  to convert from  $d$  to  $g$ . The equation for  $d$  is presented below in (1) followed by the

equation for  $J$ , which leads to the calculation of Hedges's  $g$  (Borenstein, 2009 pp. 26-27)

(1)

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}}$$

$$J = 1 - \frac{3}{4df - 1}$$

$$\text{Hedges' } g = J \times d$$

An effect size  $g$  is considered significant if the 95% confidence interval does not include 0.000. This statistic is a more accurate measure of effect than a standardized mean difference  $d$  because it weights studies according to the standard error of the effect, also referred to as inverse variance weighting (Lipsey & Wilson, 2001). A simple way to interpret an effect size is that the value represents a ratio of the standard deviation. For example if  $g = 1.000$ , this represents a treatment effect of one standard deviation in the positive direction from the mean of the sample. In normally distributed data, this can also be expressed as the mean difference moving from the 50<sup>th</sup> percentile to the 84<sup>th</sup> percentile. This weighted, standardized effect size is used with the random effects model for calculating the effect sizes throughout the analysis.

Many studies in the analysis report more than one empathy outcome variable or study multiple groups in the same study. It was possible to code these subgroups separately allowing for more precise determination of effect and further moderator

analyses. Within the included studies, multiple unique effect sizes were calculated. These unique effect sizes emerged from the reported data in one of three ways; 1) from multiple comparison groups within the study, that is a study with two treatment groups and a control group could have three different comparisons; Treatment Group A to Control, Treatment Group B to Control, and Treatment Group A to Group B; 2) when a study used more than one outcome measure to assess the effect of empathy training, this creates additional unique effect sizes within the study, that is, a study reporting data from both Interpersonal Reactivity Index and the Hogan Empathy Scale would have two effect size calculations; 3) the study reports more than one post-training measure, that is the study assesses the outcome immediately after the training and at a follow-up time point. The calculation of separate effect sizes within each subgroup of a study provides information for further moderator analysis, but is not reflective of the overall effect of the empathy training program.

The overall effect of empathy training programs was calculated by combining the weighted means of the included studies. The effect size for each of the included studies is calculated by averaging the sub-groups within each study. These study-level effects are reported and used in conducting regression analyses of continuous variables. The unique effect sizes were then used to compare the difference in moderator variables only. To use all unique effects to compute overall effects would violate the assumption of independence, overestimate the precision of the studies, and underestimate the error of the overall effect (Borenstein et al., 2009).

## Statistical Analysis

This section describes meta-analysis as a research strategy and provides the framework for the statistical analysis including justification and explanation of the statistical models and tests used in this meta-analysis.

In addition to describing the basic characteristics of the empirical studies of empathy training programs, this study attempted to address three questions that are commonly explored via a meta-analysis (Johnson, Mullen, & Salas, 1995). First, meta-analysis investigates the central tendency of the combined effect sizes. In this way, the overall treatment effect of empathy training programs is assessed. Second, meta-analysis seeks to understand the variance of the overall effect size. If variance is low, then the overall effect size may be a good estimate of the effect of this intervention across the included studies. If the variance of the effect size is high, as indicated by a significant Cochran  $Q$  score, then the overall effect size may not be a good estimate of overall effect. The  $Q$ -within ( $Q_w$ ) tests the heterogeneity within a group of effect sizes. A significant  $Q_w$  statistic that reaches the ( $p < 0.05$ ) suggests a probability that the effect sizes are not homogeneous and the overall effect cannot be interpreted as representative of all empathy training programs. The  $I^2$  statistic is also useful in assessing heterogeneity. The  $I^2$  is derived from the  $Q_w$  but represents only the real, not spurious, dispersion of effects, while the  $Q_w$  is a calculation of the true heterogeneity plus the random error (Higgins, Thompson, Deeks, & Altman, 2003).

This leads to the third common question in meta-analysis: what moderators contribute to or predict the variability? To predict or understand high variability and understand the effects of the moderators, two types of analyses can be conducted: (a) an

analog to the ANOVA of  $Q$ -between ( $Q_b$ ), wherein effect size differences are examined based on categorical variables within studies (e.g., treatment format, type of comparison group used), and (b) a weighted multiple regression, which uses continuous variables (e.g., treatment length) as potential predictors of the mean effect size (Bornstein et al., 2009). By first assessing the overall effect, then analyzing the variability in effect sizes, and then investigating the moderators, this study will answer the proposed research questions.

### Model for Meta-analysis

An important decision to make prior to undertaking a meta-analysis is to decide between fixed or random effects models for analysis. A fixed effects model assumes that the effect size for all the studies is homogenous and that any variance seen in effect sizes is due to sampling error. The fixed effect model is inappropriate for this study because we cannot assume that empathy training programs have consistent effects across studies. The random effects model, also called the mixed effect model, assumes that the effect sizes for each study are heterogeneous and the effect sizes are clustered around the mean and reflect a true difference or effect size rather than sampling error. In addition, the random effects model is more balanced in assigning weights to studies and this model allows for the analysis of more diverse studies and outcomes. (Borenstein et al., 2009) Based on these assumptions of the models for analysis, the random effects model was used throughout the analysis.

### Frame for Analysis

The analysis of aggregate data requires making several methodological and statistical decisions prior to the analysis. These decisions about how to treat the data fall

into two categories: 1) decisions guided by the main research questions and literature review that determine how data are grouped and coded, and 2) the statistical analysis of the data that considers the underlying assumptions and limitations of the various statistical tests, and selects the most appropriate test considering the data. The strategy for searching the literature, the coding, and the grouping of variables reviewed in earlier in this chapter. A description of the frame for statistical analysis follows.

The analysis of the data was conducted using two statistical software packages, SPSS Version 17.0 ([www.spss.com](http://www.spss.com)) and Comprehensive Meta-Analysis Version 2.2.027 (Borenstein et al. 2005; [www.Meta-Analysis.com](http://www.Meta-Analysis.com)). The data entry was conducted by the primary researcher under the guidance and oversight of the supervisory committee chair.

All studies meeting all inclusion criteria are included in the meta-analysis. The study level data was first entered into SPSS for ease of data entry, to calculate descriptive statistics, and to explore and determine the distribution of the data. The data were then copied into Comprehensive Meta-Analysis. This software is specifically designed to analyze data, test for heterogeneity, and conduct moderator and meta-regression analyses in a meta-analysis. The data entered included sample size, means, standard deviations, group difference scores, correlations, statistical test values ( $t$ ,  $F$ ), and  $p$  values.

Studies that reported nonsignificant findings without reporting descriptive statistics (i.e. means and standard deviations), can be a troublesome area in a meta-analysis. Although the study may report nonsignificant findings, is it unlikely that the findings are equivalent to an effect of exactly zero. Including nonsignificant studies, and using zero as the effect index of the study, has the potential to underestimate true effect size while not including nonsignificant studies has the potential to overestimate the index

of effect (Durlak & Lipsey, 1991). This study includes findings that report non-significant findings and reports the effect size as zero.

In summary, the statistical meta-analysis was conducted using *Hedges' g* as the index of effect. The statistical testing was conducted using *Q* tests with categorical data and meta-regression with continuous data. The study used a random effects model and conducted analysis using specialized statistical and meta-analytic software.

The final section of this chapter describes how the included studies were assessed for publication bias.

### Publication Bias

One of the largest threats to the validity of a meta-analysis is publication bias. Publication bias has the potential to skew the results of the analysis due to potential sampling bias. Because published literature is more easily accessible to researchers. Most meta-analyses rely heavily on published literature. The problem inherent in sampling from published literature is that research that gets published is more likely to have found a positive results than research that report nonsignificant results (Epstein, 2004). This bias has been described as a file-drawer bias because many research reports with nonsignificant findings are not published and sit instead in the researcher's file drawer. This threat has the potential to have an upward bias, as more positive studies are included in the analysis and fewer nonsignificant or small effect studies are excluded, the analysis may report an inflated overall effect. An additional threat to the validity of a meta-analysis is the likelihood that no matter how exhaustive the literature search, that some studies will not be found and therefore excluded from the analysis. Fortunately,

there are several tests that may be conducted to statistically assess for publication bias. Publication bias was assessed using four statistical tests as described below.

Traditionally, a Rosenthal Fail-safe  $N$  (Rosenthal, 1979) is conducted that offers an estimate of the number of additional studies not included in the analysis, that would be needed to make the p-value insignificant. A small value would reflect that only a few studies are need to nullify the effect, while a very large number would increase our confidence in the findings.

More recently Rosenthal's  $N$  has been supplanted by Orwin's Fail-safe  $N$  as a more appropriate test for publication bias because it allows for more flexibility in assigning substantive values to missing studies and assigning an effect value, other than zero, to the missing studies (Borenstein et al., 2009).

Another assessment of publication bias is conducted by plotting effect size by the standard error of the studies on a funnel plot. This visual representation of the studies should look like an inverted funnel, with studies clustered around the mean symmetrically. If the studies are evenly distributed in an inverted funnel shape, we can be confident that there is not a systemic publication bias. To address any unbalance in the distribution of effects, Duval and Tweedie's Trim and Fill (Duval & Tweedie, 2000) is conducted that further assesses for bias by removing, or trimming, the most extreme smaller studies. This process then adds, or imputes, a mirror-image of the studies to make an even distribution on both sides of the mean. By balancing the distribution, we can assess the impact missing studies and determine if the overall effect changes significantly.

This concludes the description of the methodology used to generate data for this meta-analysis of empathy training programs. The following chapter will present the results of search strategy, the quantitative results of the meta-analysis and moderator analysis, the exploration of the data, and will conclude with the assessment of publication bias.

## CHAPTER 4

### RESULTS

This chapter describes the results of the meta-analysis of empathy training programs. A discussion of the study searching and screening process is presented first followed by descriptive data on included and excluded studies. Next, the research questions guide the presentation of the results of the meta-analysis. Overall results are presented first followed by tests of heterogeneity. Next, moderator analyses are presented organized by study variables, empathy program variables, empathy program ingredients, and participant variables. The analysis will show that one moderator, the measurement of empathy, is significant. This leads to partitioning the data into two main groups. Further moderator analyses are presented on this partitioned data and compared to the overall effects. Lastly, potential publication bias is explored and presented.

#### Sorting and Screening Studies

A total of 24 studies on empathy training met inclusion criteria. The process of screening and sorting the studies is discussed below and presented in a PRISMA-type flow chart in Figure 1 (Moher, Liberati, Tetzlaff, & Altman, 2009).

Using the search methods described in the method section, the initial search identified 2433 potential studies on empathy training. These study abstracts were reviewed by the primary author and screened according to the following criteria: 1)

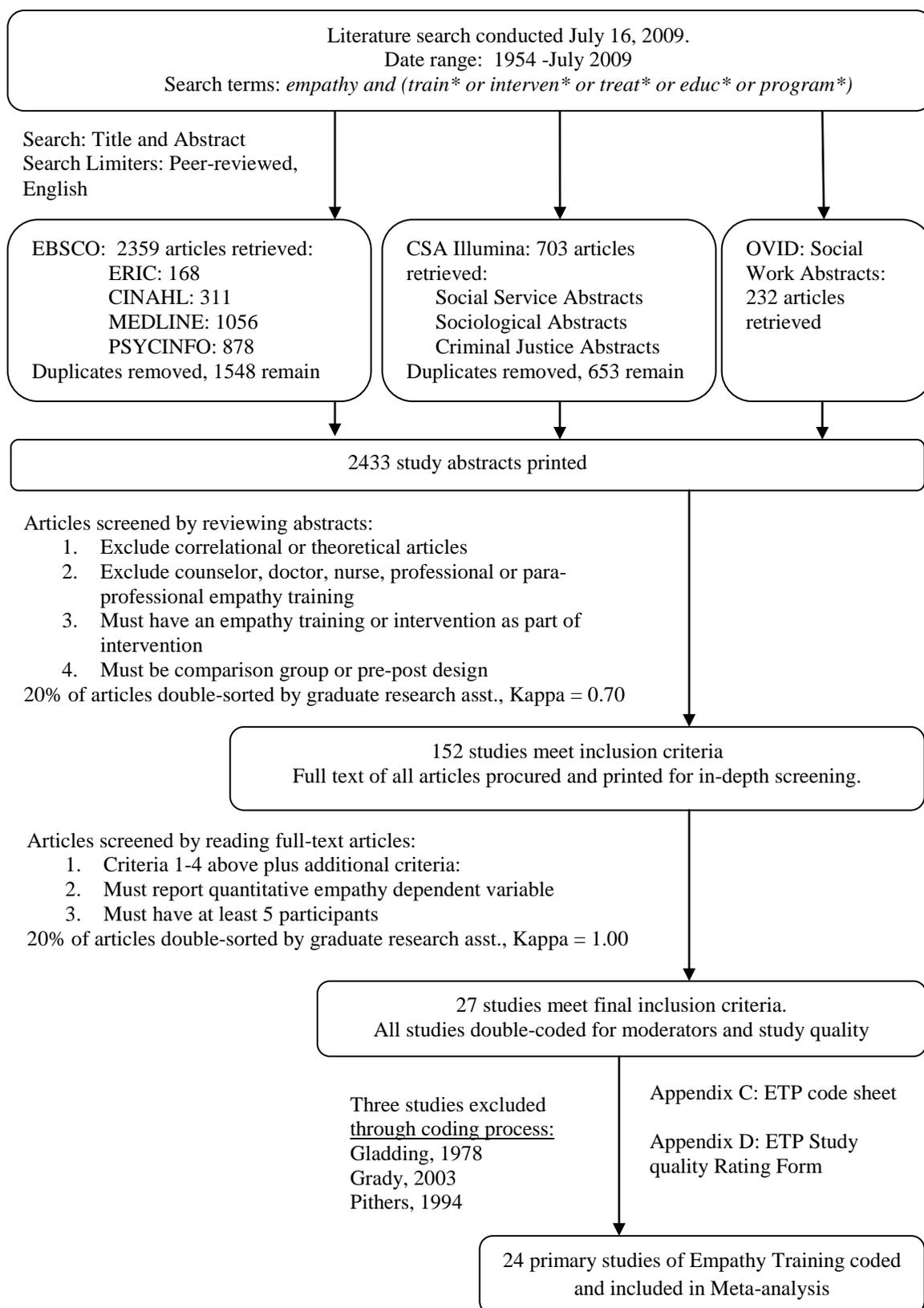


Figure 1. PRISMA-type Flow Chart of Sampling Procedure

exclude correlational or theoretical articles, 2) exclude counselor, doctor, nurse, professional or para-professional empathy training, 3) must have an empathy training or intervention as part of intervention, and 4) must be comparison group or pre-post design. This initial screening excluded 2281 studies, leaving 152 potential studies. To ensure the integrity of the screening process, 20% of the studies were double-screened by a trained doctoral level graduate research assistant and the primary author. Agreement in decision making was calculated at  $Kappa = 0.70$  ( $n = 600$ ,  $p < 0.001$ ), which indicates substantial agreement (Landis & Koch, 1977). The screening instructions, training protocol, and inclusion criteria are included in Appendix B.

The remaining 152 studies were screened again for inclusion by obtaining the full text articles and reviewing them for inclusion using the aforementioned criteria as well as two additional conditions; 1) Must report a quantitative empathy outcome or dependent variable; 2) Must have at least five study participants. The primary investigator screened all 152 articles and the graduate assistant double-screened 20% of these articles with  $Kappa = 1.00$  ( $n = 30$ ,  $p < 0.001$ ) indicating perfect agreement. This screening process produced 27 articles that met all inclusion criteria.

All 27 studies were double coded by the primary investigator and graduate research assistant using the Empathy Training Program Code Sheet (Appendix C) and Empathy Training Program Study Quality Rating Form (Appendix D). Disagreement in coding was resolved through discussion. Through the process of coding the studies, three additional studies (Gladding, 1978; Grady, 2003; Pithers, 1994) were identified that did not meet the inclusion criteria and were excluded from the analysis. The excluded studies and the reasons for their exclusion are listed in Table 1.

Table 1

## Excluded Studies

Author	Date of Publication	Population	Reason for Exclusion
Gladding	1978	College students	No empathy intervention, stimulus only (infant cry)
Grady (Diss.)	2003	Sex Offenders	CBT intervention, not empathy training
Pithers	1994	Sex Offenders	Reports same data as Pithers (1999) study (included)

Descriptive data for the included studies are listed in Table 2. The publication date range for the studies is 1976 through 2008 and includes 21 journal articles and three dissertations. The total number of participants for all studies is  $N = 1319$ . Eighteen of the studies used control or comparison groups and six used a pre-post design with a single experimental group. Only five studies included a follow-up assessment of empathy to assess the durability of the intervention effect, ranging from 2 months to 15 months posttraining. Seven of the studies used college students, four targeted incarcerated sexual offenders, and six studies were conducted on incarcerated individuals.

The following section describes the analysis of the data, organized by the original research questions.

#### Meta-analysis of Data

##### R.Q. # 1: Do Empathy Training Programs Increase

##### Empathy in Client Populations?

This question was explored by initially calculating an effect size for the 24 studies. The overall effect size is  $d = 1.048$  ( $sd\ 1.620$ ),  $k = 24$  which reflects a large effect (Cohen, 1988). However this value may not be representative of all empathy training programs because there may be heterogeneity within the data. Significant heterogeneity can undermine confidence in the overall effect and is why moderator analyses were conducted on the data and reported in the subsequent sections.

In an attempt to understand potential sources of heterogeneity in the data, a visual inspection was conducted. Scanning the effect sizes from each study revealed at least

Table 2

## Included Studies

Author	Publ. Date	Population	N in each group		Group Assignment Protocol	Comparison Groups	Outcome Scale(s)	Empathy Outcome
			Treatment	Control				
Angera & Long	2006	Couples	42		None	P	I,O	self-report
Berg et al.	1999	College males	18,19	17	Randomized	T,C	I,R	self-report
Dorfman (Dissertation)	1993	Adol. sex offend.	9		None	P, F	H,Q	self-report
Foubert & Newberry	2006	College males	93,73	93	Randomized	T,C	R	self-report
Guzzetta	1976	Parents	10,8,8	8	Matched groups	P,T	C	observer-report
Haynes & Avery	1979	Adolescents	25	23	Matched groups	C	O	observer-report
Heran (Dissertation)	2005	Adol. sex offend.	31	9	None	C	I,M,C,O	self-report
Isquick	1981	Aged population	9,9	10	Randomized	T,C	C	observer-report
Janoka & Scheckenbach	1978	Prisoners	7		Randomized	P	C	observer-report
Kalliopucka & Tiitinen	1991	Children	14,17	32	None	T,C	F	observer-report
Kalliopuska & Ruokonen	1993	Children	16	16	None	C,F	F	observer-report
Kimberlin & Friesen	1977	College students	20,20,20,20	40	Matched groups	T,C	C,O	observer-report
Kipper & Ben-Ely	1979	H.S. students	16,16,16	16	Matched groups	T,C	T	observer-report
Kremer & Dietzen	1991	College students	23,21	16	Matched groups	T,C,F	O	observer-report
Lomis & Baker	1985	Psychiatric	8	8	Randomization	C	C,H	both
Long et al.	1999	Couples	96		Randomization	C,F	I,O	self-report
Long et al.	2008	Couples	32		None	P	O	self-report
Marshall et al.	1996	Sex offender	29		None	P	M	self-report
Ogle (Dissertation)	2008	College students	26	26	Matched groups	T,C	I,B	self-report
Pecukonis	1990	Adol.females	24		None	P	H,B	self-report
Pithers	1999	Sex offenders	20		None	P	I	self-report
Schewe & O'Donohue	1993	College students	14,14	27	Randomized	T,C	R	self-report
Therrien	1979	Parents	30	17	Tested for equivalence	C,F	T	observer-report
Uhlemann & Koehn	1989	College students	16,16,16	16	Randomized	T,C	C	observer-report
			Total N = 1319					

**Group Assignment:** The method by which the participants were assigned to groups.

**Comparison groups:** P= pre-post comparison, C= treatment to control, T=treatment to treatment, F=Follow-up

**Empathy Scale:** B=Balanced Emotional Empathy Scale (BEES), C=Carkhuff, F= Feshbach & Roe, H= Hogan Empathy Scale (HES), I=Interpersonal Reactivity Index (IRI), M= Child Molester Empathy Measure (CMEM), O=Other (proprietary) empathy measure, Q= Questionnaire Measure of Emotional Empathy (QMEE), R= Rape Empathy Scale (RES), T=Truax.

two obvious outliers with effect sizes of  $d = 10.063$  (Haynes, 1979) and  $d = 8.775$  (Guzetta, 1976). Because the purpose of a meta-analysis is to calculate a reasonable summary or average of quantitative findings (Lipsey & Wilson, 2001), the data, including these two outliers and one other (Kremer, 1991;  $d = 4.812$ ), were Winsorized. This process maintains the contribution of the study effects, but reduces the impact of the extremes scores. In this study, effect sizes from three studies (Haynes, 1979; Guzetta, 1976; Kremer 1991) fell outside the 95% C.I. range and were, therefore, Winsorized to a value at two standard deviations from the mean, or 4.29. This value was calculated using the following equation 2.

$$d \pm 2sd = 95\% \text{ C.I.} \quad (2)$$

$$1.048 \pm 2(1.620) = 4.29, -1.192$$

Both the pre- and post-Winsorized values are reported in Table 3. This process reduced both the overall calculated effect size and standard deviation to  $d = 0.928$  (1.106)  $k = 24$ . The Winsorized data and a weighted effect size ( $g$ ) are used for the remainder of the analysis.

Having Winsorized the data, the overall standardized and weighted effect (Hedges'  $g$ ) are calculated for the 24 empathy training programs, using the random effects model, to answer the first research question: do empathy training program increase empathy? The result  $g = 0.911$ ,  $p < 0.000$ ,  $k = 24$  suggests that empathy training

Table 3

## Windsorizing Data

	<i>k</i>	Minimum value	Maximum value	Effect size <i>d</i>	<i>sd</i>
Pre-Windsorized	24	-1.03	10.06	1.0477	1.621
Post-Windsorized	24	-1.03	4.29	0.9280	1.106

programs are associated with increases in empathy. The 24 studies are sorted by study quality with Forest plots are visually presented in Table 4. This overall large effect (Cohen, 1988) suggests that empathy training programs hold promise, although this overall statistic may not be reflective of all empathy programs. The variation in study quality is another reason to be cautious in interpreting the results. Further analyses of the data are necessary, leading to the second research question.

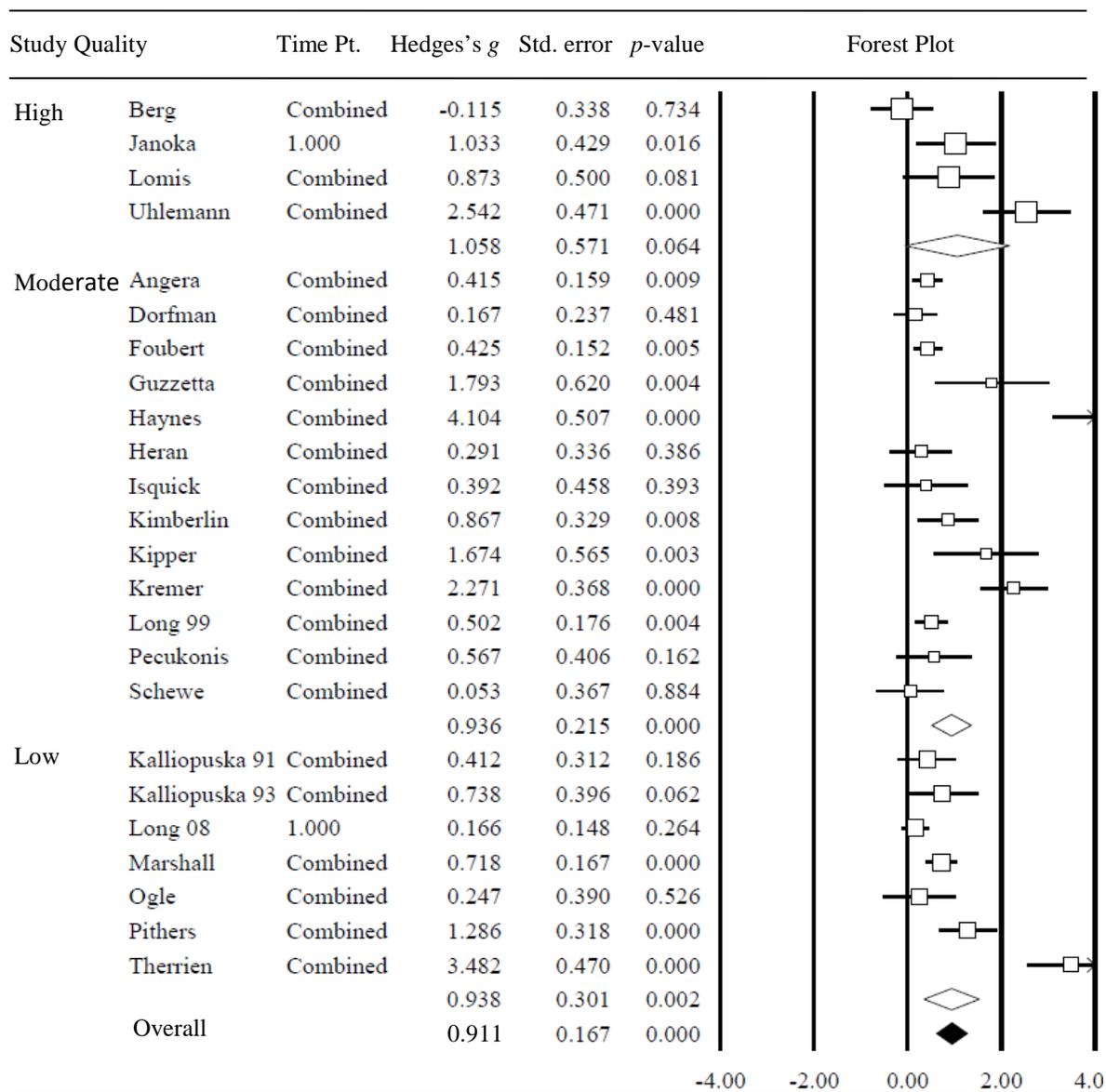
#### R.Q. # 2: Is There Variation in Empathy Training Program Outcomes?

The next step in the meta-analysis is to determine if the overall effect size is consistent across studies or if significant heterogeneity exists. If significant heterogeneity exists this leads to further moderator analyses, if not, then the overall effect is accepted as representative of the included studies.

The analysis reveals significant heterogeneity in the effect sizes,  $Q_w (23, k = 24) = 358.407, p = 0.001$ , of the included studies. In addition to the  $Q$  statistic, another statistic, the  $I^2$  may be used to assess heterogeneity. The included studies have an  $I^2 = 93.583$ . The large value for  $I^2$  also reflects significant heterogeneity. A value of this statistic is that it may also be expressed as a percentage for ease of interpretation. In this case, 93.583% of the observed variation is likely to be true heterogeneity and not due to error. An  $I^2$  value greater than 75 indicates high levels of heterogeneity (Higgins et al., 2003) and confirms that too much heterogeneity exists to simply accept the overall effect and that it should be interpreted with caution. Further moderator analysis is needed to account for this variation.

Table 4

## Effect Sizes and Forest Plots by Study Quality



### R.Q. # 3: What Moderators Contribute to Empathy Training Programs?

Given that significant heterogeneity was found, moderator analyses were conducted. Moderator analyses are conducted to determine if grouping effect sizes by certain variables eliminates the heterogeneity allowing for confidence in the effect size based the variable being used as a moderator. Further, moderator analyses allow for between-group tests. That is, subsets within a variable class can be directly compared which allows for both exploratory testing and a priori hypothesis testing. As discussed in chapter 3, potential moderators were organized in the four groups: 1) study variables (including study quality), 2) program variables, 3) program ingredients, and 4) participant variables.

The moderator analysis and tests of heterogeneity are conducted and reported using either study level data ( $k = 24$ ) or using subgroup data ( $k = 90$ ). The use of study level data will report one effect size for each study, and will be reported when the moderating variable is consistent across the subgroups in the study. If the moderator varies across the subgroups of the study, subgroup level data will be reported.

Study variables. Meta-analysis relies on the results of primary studies to calculate effects. The methodological quality of the included studies is, therefore, important to explore through moderator analysis. Confidence in the results of a meta-analysis is weakened if the primary studies were conducted poorly or have significant methodological flaws.

One indicator of study quality is the way participant were assigned to groups within the study. Group assignment and the use of randomization are important to protect against bias and threats to validity (Rubin, 2008). Among the 24 studies included

in the analysis, eight used randomization, seven used matched groups, one tested for group equivalence in existing groups, and eight used no randomization (Table 5). These four conditions of group assignment were tested for heterogeneity and the findings were significant,  $Q_b(3, k = 24) = 78.031, p < 0.001$ , suggesting group assignment may moderate effects. To further explore the source of the group differences, the one study that tested for group equivalence was partitioned out of the analysis because it appears to be an outlier ( $g = 3.390$ ) and may be considered too unstable. Tests for heterogeneity were conducted again with the remaining studies. The result,  $Q_b(2, k = 23) = 7.666, p = 0.022$ , still shows heterogeneity between the groups. Subsequent group to group comparisons reveals that matched case groups show the strongest training effects, while non-randomized groups and randomized groups show similar effects,  $Q_b(1, k = 16) = 0.583, p = 0.445$ .

Another important consideration in study design is the use of comparison groups. This moderator analysis explores if empathy training programs have differential absolute (treatment vs. control) or relative effects (treatment vs. treatment) as discussed by Littell and associates (2008). In this meta-analysis includes three types of comparisons; 1) pre-post designs with a single group, 2) pre-post designs comparing one empathy treatment to another, and 3) designs comparing empathy training to a no-treatment control group. Outcomes differed significantly based on the type of comparison group  $Q_b(2, k = 90) = 6.833, p = 0.033$  and the results of the test are shown in Table 5. Treatment vs. control group designs have a significantly higher effect ( $g = 1.349, k = 32$ ) than either treatment to treatment designs ( $g = 0.638, k = 41$ ) or pre-post only designs ( $g = 0.507, k = 17$ ). This

Table 5  
Empathy Training Moderators: Study Variables

Group	k	Hedge's g	Std. Error	95% conf. inter.	Q <sub>w</sub> p-value	Q <sub>b</sub>	df	Q <sub>b</sub> p-value
Group Assignment	24					78.031	3	0.000*
Random	8	0.670	0.223	0.223	1.107	0.000		
Matched Case	7	1.397	0.321	0.767	2.027	0.000		
Tested for equivalence	1	3.390	0.327	2.749	4.032	0.000		
No randomization	8	0.487	0.088	0.314	0.660	0.000		
Comparison groups	90					18.796	2	0.000*
Treatment to Control	32	1.429	0.186	1.065	1.793	0.000		
Treatment to Treatment	41	0.580	0.108	0.368	0.792	0.000		
None: pre-post	17	0.546	0.103	0.343	0.748	0.000		
Study Quality (ordinal)	24					0.066	2	0.967
High	4	1.059	0.653	-0.221	2.339	0.105		
Moderate	13	0.887	0.154	0.584	1.189	0.000		
Low	7	0.902	0.251	0.410	1.394	0.000		
Study Quality (meta-regression)	24	<i>b</i> = 0.000				0.002	1	0.964

Note: *k* = number of studies or unique effects contributing to an effect size. Variables that are consistent across effects within for studies will have a total *k* = 24. Variables reporting different levels with studies due to subgroups or outcome variables will have a total *k* = 90.

suggests that empathy training programs have a significantly larger absolute effect than relative effect.

To investigate the degree to which overall study quality moderated effects, two separate but related ratings of study quality are used. Specifically, a continuous variable of total quality and an ordinal rating for low, moderate or high quality are examined. The first is a simple a sum of scores on the study quality rating form (Appendix D). The mean value for all studies is  $M = 9.30$  ( $sd\ 2.2$ ),  $k = 24$ . Meta-regression is used to assess whether overall quality is related to moderated outcomes. No effect was found ( $b = 0.000$ ,  $p = 0.965$ ). Figure 2 depicts the regression.

Overall study quality is also coded as an ordinal variable with three levels: low, moderate, or high quality. There are a total of 4 high quality studies, 13 studies of moderate quality, and 7 low quality studies. Consistent with the effects of the meta-regression, the between group analysis reveals no significant differences based on study quality,  $Q_b(2, k = 24) = 0.066$ ,  $p = 0.967$  (Table 5).

The association between year of publication and strength of effect is also analyzed. This analysis could identify trends or patterns of empathy training programs over time. The meta-regression of year published by effect size reveals a significant negative trend over time  $b=-0.050$ ,  $p < 0.001$ ,  $k = 24$ . The results are shown in Figure 3. The negative slope of the regression line shows that empathy training programs have become significantly less effective from 1976 to 2008. Further analysis of the data will show that this trend may not be entirely a time effect, but an effect of the outcome instruments used to measure empathy that have evolved over the last 30 years and have more conservative outcomes.

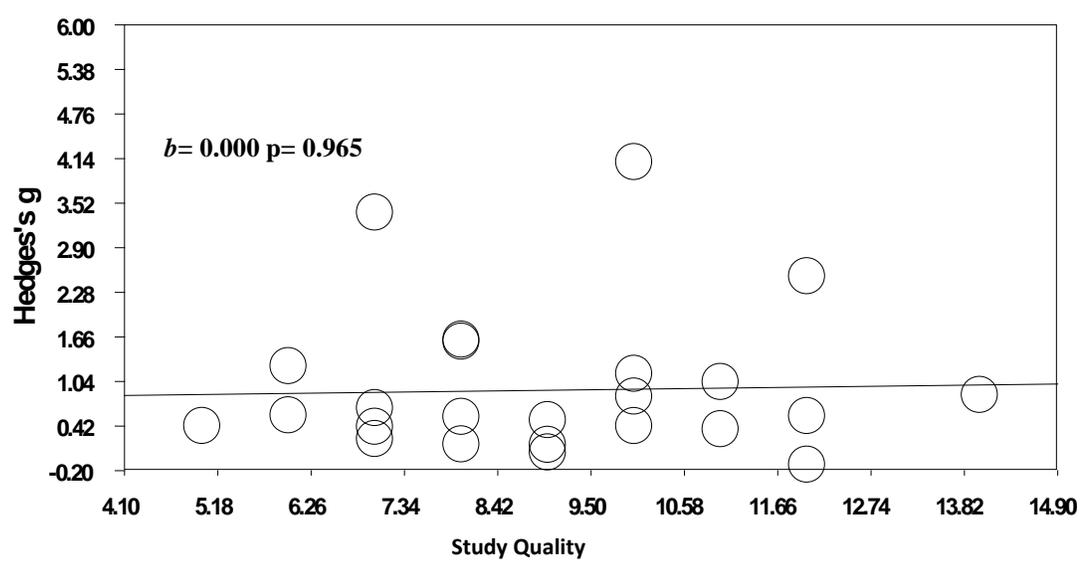


Figure 2. Meta-regression of Study Quality on Hedges's g.

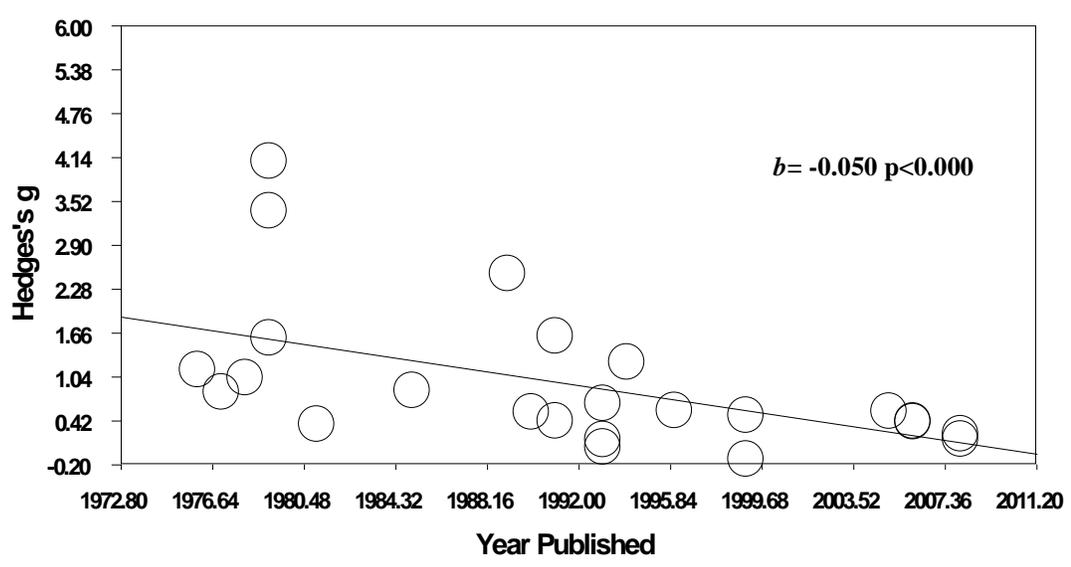


Figure 3. Meta-regression of Year Published on Hedges's g

Program variables. Program variables were explored for heterogeneity to identify characteristics of empathy programs contribute to more effective empathy training. This section provides the results of the moderator analysis of the program variables.

The measurement of empathy emerged as a significant source of debate and disagreement in the literature on empathy (Albiero, 2009; Jolliffe & Farrington, 2006; Mehrabian, 1997). The way empathy is conceptualized and subsequently measured may contribute to variation in effectiveness of empathy training programs. In fact, this is what was found. Outcomes vary as a function of the instrument used to measure empathy,  $Q_b(11, k = 90) = 65.399, p = 0.000$ . Table 6 shows the results.

Visual inspection of effects grouped by type of empathy outcome variable reveals that many of the observer ratings scales, like the Carkhuff and Truax scales, have a much larger overall effect than the other outcome measures. Based on this observed difference in the effect by outcome scale, the outcome scales were grouped into two main groups, self-report scales and observer-report.

All 12 empathy scales were re-coded into a dichotomous variable, empathy outcome type, and tested for heterogeneity. This comparison reveals a significant difference in effect sizes  $Q_b(1, k = 90) = 36.998, p < 0.001$ , shown in Table 6. Specifically, when empathy is assessed through observation  $g = 1.488, p < 0.001, k = 43$  and when a self-report measure is used  $g = 0.386, p < 0.001, k = 47$ . Grouping the data in this way reveals that observer reported empathy scales show a mean effect nearly four times larger than that of self-report measures. This large difference suggests that the way empathy is measured is a powerful moderator in the effect of empathy training programs. This measurement effect leads to further moderator analysis and exploration of the

Table 6

## Empathy Training Moderators: Program Variables

Group	k	Hedge's Std.		95%		Q <sub>w</sub> p-value	Q <sub>b</sub>	df	Q <sub>b</sub> p-value
		g	Error	conf.	interval				
Empathy Outcome Scale	90						65.399	11	0.000*
BEES	3	0.530	0.230	0.078	0.982	0.021			
Carkhuff	22	1.206	0.192	0.830	1.582	0.000			
CMEM	12	0.473	0.143	0.192	0.753	0.001			
Empathy-other	12	1.601	0.290	1.033	2.169	0.000			
Feshbach & Roe	4	0.543	0.257	0.040	1.046	0.034			
HES	4	0.142	0.146	-0.143	0.427	0.329			
IRI	11	0.511	0.136	0.245	0.777	0.000			
ODPT	2	0.369	0.116	0.141	0.597	0.001			
QMEE	2	0.248	0.190	-0.125	0.622	0.193			
RES	8	-0.037	0.177	-0.385	0.310	0.834			
S(O)DPT	2	0.533	0.120	0.298	0.768	0.000			
Truax	8	2.136	0.375	1.401	2.870	0.000			
Empathy Outcome Type	90						38.998	1	0.000*
Observer	43	1.488	0.165	1.164	1.812	0.000			
Self-report	47	0.386	0.062	0.266	0.507	0.000			
Type of empathy measured	90						3.320	2	0.190
Cognitive	74	0.884	0.091	0.706	1.062	0.000			
Affective	11	0.547	0.162	0.229	0.865	0.001			
Both	5	0.756	0.290	0.187	1.325	0.000			
Dosage of training (meta-regression)									
Total Hours	90						0.717	1	0.397
Number of sessions	90						0.537	1	0.463
Number of weeks	90						0.507	1	0.476
Setting for training	24						7.517	4	0.111
College	7	0.773	0.258	0.267	1.280	0.003			
Group Home/Residential	3	0.404	0.158	0.094	0.714	0.011			
Jail/prison/secure facility	4	0.879	0.203	0.482	1.277	0.000			
Outpatient	5	0.900	0.307	0.298	1.503	0.003			
School preK-12	5	1.560	0.520	0.541	2.580	0.003			
Positive Empathy	90						19.849	1	0.000*
Yes	37	0.448	0.074	0.303	0.592	0.000			
No	53	1.077	0.120	0.841	1.313	0.000			
Trainer models empathy	90						0.235	1	0.628
Yes	18	0.744	0.174	0.403	1.085	0.000			
No	72	0.838	0.086	0.670	1.007	0.000			
Modality of training	24						12.186	2	0.002*
Couples	3	0.396	0.083	0.223	0.559	0.000			
Individual	2	1.553	0.998	-0.402	3.508	0.120			
Groups	19	1.000	0.161	0.685	1.316	0.026			
Independent Empathy Training	24						0.170	1	0.680
Empathy training only	19	0.888	0.151	0.592	1.183	0.000			
Empathy plus other	5	0.642	0.126	0.512	1.508	0.000			

Note:  $k$  = number of studies or unique effects contributing to an effect size. Variables that are consistent across effects within for studies will have a total  $k = 24$ . Variables reporting different levels with studies due to subgroups or outcome variables will have a total  $k = 90$ .

partitioned data within each of these subgroups later in this chapter.

Below, results from further overall moderator analyses are presented even though little confidence can be placed in them as they may be confounded by measurement type. These results will be presented organized by program variables (Table 6), program ingredients (Table 7), and participant variables (Table 8).

Seven other program variables are tested for group difference with results shown in Table 6. Two of the moderators show a group difference; 1) modality of the training, 2) and the inclusion of a positive empathy component in the training. There is a significant heterogeneity between the three training modalities; group, individual, and couples.  $Q_b(2, k = 24) = 12.186, p = 0.001$ . Grouping by this modality indicates that groups are the most effective modality ( $g = 0.895$ ), followed by individual ( $g = 0.798$ ), and couples ( $g = 0.402$ ), although further analysis reveals that group and individual modalities are statistically equivalent. Programs that include a positive empathy component  $Q_b(1, k = 90) = 19.849, p = 0.001$  were significantly less effective ( $g = 0.448$ ) than programs that did not ( $g = 1.077$ ).

The five program variables that show no group differences are: 1) the type of empathy targeted in training, 2) the dosage of the empathy training programs, 3) the setting of the empathy training, 4) if the empathy trainer models empathy, and 5) if stand-alone empathy training or empathy training plus other psychosocial interventions are associated with changes in empathy.

Program ingredients. Program ingredients are the specific interventions or activities within an empathy training program that are provided to participants. The results of the moderator analyses on these variables are shown in Table 7. Only four of

Table 7

## Empathy Training Moderators: Program Ingredients

Group	k	Hedges' g	Std. Error	95% conf. interval	Q <sub>w</sub> p-value	Q <sub>b</sub>	df	Q <sub>b</sub> p-value
Audio tape	90					1.187	1	0.276
Yes	15	1.139	0.346	0.462	1.817	0.001		
No	75	0.754	0.074	0.610	0.899	0.000		
Video tape	90					5.192	1	0.023*
Yes	41	0.631	0.088	0.457	0.804	0.000		
No	49	0.983	0.127	0.734	1.231	0.000		
Role-play	90					0.008	1	0.928
Yes	62	0.826	0.089	0.652	1.000	0.000		
No	28	0.810	0.022	0.518	1.103	0.000		
Exposure to victim	90					3.512	1	0.061
Yes	15	0.586	0.134	0.323	0.849	0.000		
No	75	0.890	0.092	0.711	1.070	0.000		
Visualization	90					1.135	1	0.287
Yes	4	1.270	0.430	0.427	2.113	0.003		
No	86	0.804	0.078	0.651	0.957	0.000		
Homework	90					7.153	1	0.007*
Yes	13	0.534	0.097	0.334	0.724	0.000		
No	77	0.907	0.100	0.711	1.103	0.000		
Written assignments	90					0.239	1	0.625
Yes	20	0.742	0.189	0.371	1.112	0.000		
No	70	0.843	0.085	0.676	1.010	0.000		
Reading Materials	90					0.086	1	0.770
Yes	13	0.882	0.223	0.445	1.319	0.000		
No	77	0.813	0.083	0.650	0.975	0.000		
Psycho-education	90					1.921	1	0.166
Yes	55	0.742	0.098	0.549	0.934	0.000		
No	35	0.964	0.127	0.715	1.212	0.000		
Psychodrama	90					23.527	1	0.000*
Yes	8	0.257	0.098	0.065	0.449	0.009		
No	82	0.881	0.083	0.718	1.044	0.000		
Total Training ingredients								
Meta-regression:	90	<i>b</i> =-0.145	0.044	-0.231	-0.058	0.001	10.761	1 0.001*

Note: *k* = number of studies or unique effects contributing to an effect size. Variables that are consistent across effects within for studies will have a total *k* = 24. Variables reporting different levels with studies due to subgroups or outcome variables will have a total *k* = 90.

the eleven training ingredients show a differential empathy training effect: 1) watching a video, 2) homework, 3) psychodrama, and 4) total number of training ingredients. These results are shown in Table 7. Programs that watched a video as part of the program ( $g = 0.631, k = 41$ ) were less effective than programs that did not watch a video ( $g = 0.983, k = 49$ ). Programs that used homework ( $g = 0.534, k = 13$ ) were less effective than program that did not ( $g = 0.907, k = 77$ ). Programs that engaged in psychodrama ( $g = 0.257, k = 8$ ) were less effective than programs that did not ( $g = 0.881, k = 82$ ). The total number of empathy training interventions was explored through meta-regression ( $b = -0.145, p = 0.001, k = 90$ ), revealing a significant negative relationship between effect size and number of training ingredients.

Participant moderators. Participant moderators are the characteristics of the client or participant in the empathy training program and are shown in Table 8. Three of seven client variables are found to be significant; gender, level of prevention, and client presenting problem.

The gender make-up of the groups is a significant moderator  $Q_b (3, k = 90) = 9.160, p = 0.002$ . Female-only groups ( $g = 1.363$ ) and mixed gender groups ( $g = 1.298$ ) are statistically equivalent,  $Q_b (1, k = 40) = 0.034, p = 0.853$ , and show a much larger effect than male-only groups ( $g = 0.403$ ). Group gender differences are also explored through meta-regression by percentage of males, this also shows a significant gender effect ( $b = -0.004, p = 0.000$ ), with larger numbers of males in the program associated with lower empathy gains.

Primary prevention programs ( $g = 1.235, k = 14$ ) have outcomes significantly higher than either secondary ( $g = 0.134, k = 3$ ) or tertiary prevention programs ( $g =$

Table 8

## Empathy Training Moderators: Participant Variables

Group	k	Hedge's Std.		95%		Q <sub>w</sub>	Q <sub>b</sub>	df	Q <sub>b</sub>
		g	Error	conf. interval		p-value			p-value
Gender	90						29.318	3	0.000*
All Female	11	1.363	0.306	0.764	1.963	0.000			
All Male	37	0.403	0.085	0.237	0.569	0.000			
Both genders	29	1.298	0.181	0.944	1.651	0.000			
Not reported	13	0.860	0.130	0.605	1.114	0.000			
Percent Males	77	<i>b</i> = -0.004	0.001	-0.006	-0.002	0.000	15.560	1	0.000*
Level of Prevention	24						14.469	2	0.001*
Primary	14	1.235	0.210	0.823	1.646	0.000			
Secondary	3	0.134	0.200	-0.257	0.525	0.502			
Tertiary	7	0.625	0.114	0.403	0.848	0.000			
Presenting Problem	24						17.779	5	0.003*
No presenting problem	9	1.345	0.301	0.754	1.936	0.000			
Couples	3	0.388	0.085	0.222	0.553	0.000			
Prim/secondary prevention	3	0.134	0.200	-0.257	0.525	0.502			
Parenting skills	2	2.256	1.122	0.057	4.455	0.044			
Delinquency/criminal	3	0.762	0.184	0.402	1.121	0.000			
Sexual offender	4	0.582	0.139	0.309	0.855	0.000			
Dangerousness (meta-regression)	24	<i>b</i> = -0.047	0.041	-0.126	-0.033	0.249	1.324	1	0.249
Voluntary Status	24						3.601	1	0.058
Involuntary	7	0.625	0.114	0.403	0.848	0.000			
Voluntary	17	1.028	0.179	0.667	1.379	0.000			
Age	24						6.351	4	0.174
0-12 years	2	0.517	0.173	0.179	0.856	0.003			
13-18 years	5	1.335	0.424	0.503	2.166	0.002			
19-30 years	8	0.782	0.237	0.316	1.247	0.001			
31-60 years	8	0.964	0.187	0.597	1.331	0.000			
60+ years	1	0.376	0.324	-0.259	1.011	0.246			
Age (meta-regression)	24	<i>b</i> = -0.027	0.126	-0.275	0.220	0.829	0.046	1	0.829
Ethnicity Reported	24						23.395	1	0.000*
Yes	6	0.252	0.101	0.054	0.450	0.013			
No	18	1.177	0.163	0.859	1.496	0.000			
Percent Caucasian (meta-regression)	6	<i>b</i> = -3.651	2.919	-9.373	2.069	0.211	1.565	1	0.211

Note: *k* = number of studies or unique effects contributing to an effect size. Variables that are consistent across effects within for studies will have a total *k* = 24. Variables reporting different levels with studies due to subgroups or outcome variables will have a total *k* = 90.

0.625,  $k = 7$ ),  $Q_b(2, k = 24) = 14.469$ ,  $p = 0.001$ .

Training outcomes vary by the presenting problem of the client  $Q_b(5, k = 24) = 17.799$ ,  $p = 0.003$ , with parenting programs ( $g = 2.256$ ,  $k = 2$ ) and no presenting problem populations ( $g = 1.345$ ,  $k = 9$ ) showing the largest effects. This categorical variable was re-coded as a continuous variable for client dangerousness and was explored through meta-regression ( $b = -0.047$ ,  $p = 0.249$ ). This result shows significant variability in the outcome of training programs by client presenting problem, but do not show a trend in the relationship between the level of dangerousness of the client and training outcome.

Both moderator analyses on prevention levels and presenting problems indicate that clients with no presenting problems and at a primary level of prevention show larger increases in empathy than programs targeting other problems or levels of prevention.

Non-significant participant variables include voluntary/involuntary status, age, and an ethnicity variable, percent Caucasian. The age of the participant was analyzed through a moderator analysis for group differences  $Q_b(4, k = 24) = 6.351$ ,  $p = 0.174$ , and through meta-regression  $b = -0.027$ ,  $p = 0.829$ . Neither analysis demonstrates a relationship between age and effect.

This concludes the presentation of the overall meta-analytical data. The type of empathy outcome remains the most significant moderating variable. The results reported above should only be interpreted with caution due to this confound which affects the overall analysis. The following section begins the analysis of the data partitioned by the moderator variable empathy outcome type. A visual representation of the partitioning process is provided in Figure 4.

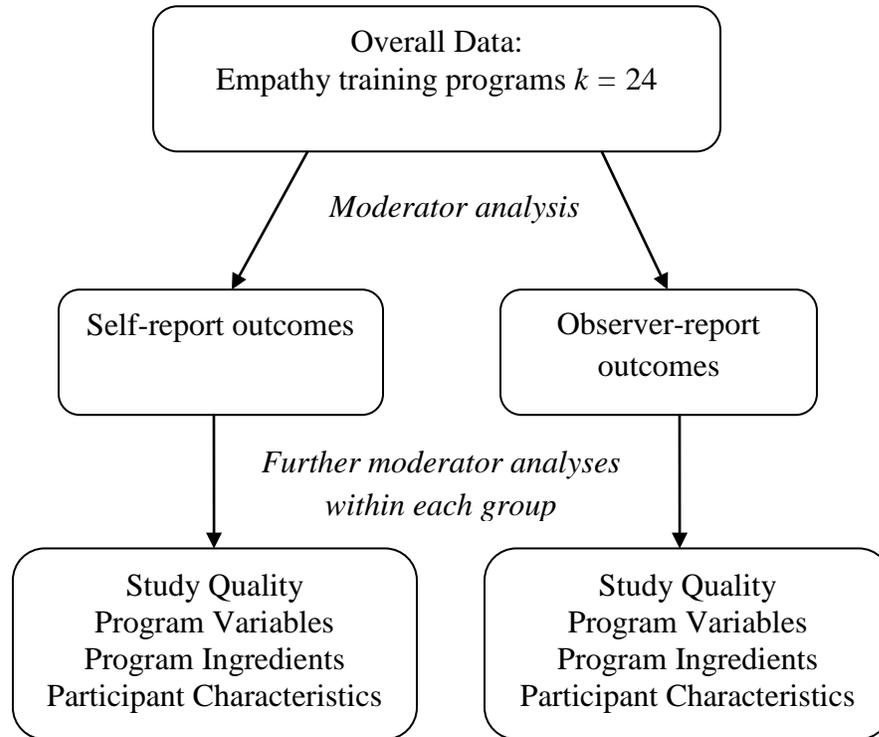


Figure 4. Partitioning Data by Outcome Type

### Partitioned Data Analysis

One of the most significant findings of the moderator analysis is the difference between empathy programs using self-report outcome measures and programs using observer reported outcomes. As can be seen in Table 6, the data grouped by outcome scale has significant heterogeneity, and even when the outcomes are partitioned into self-report scales and observer reported scales, group difference persists in the data. The difference in measurement scale likely accounts for a significant portion of the variance present in the combined data. Because of this difference, further moderator analyses are conducted with the data partitioned into self-report and observer outcomes and tested for heterogeneity. This section will report on the findings from the analysis of the partitioned data in the same format as the previous section, by study quality, program variables, program ingredients, and participant characteristics. The overall effect of studies using self-report scales is  $g = 0.386$ ,  $p = 0.000$ ,  $k = 13$ ,  $I^2 = 75.667$  and the overall effect of studies using observer report scales is  $g = 1.541$ ,  $p = 0.000$ ,  $k = 12$ ,  $I^2 = 93.889$ . We can see that studies using observer-reported scales report a nearly 400% higher treatment effect than studies using self-report scales.

Once the data are sorted by self-report and observer-report, no moderators show significant heterogeneity in both partitioned data sets. The lack of a robust variable that remains significant in both sets of data supports the decision to view the data separately by outcome type and that the overall results reported in the previous section should be interpreted with caution. Within each set of outcome data there are some moderators that show group differences in either the self-report or the observer-reported data, but not in both.

### Partitioned Data by Variable Type

Study variables: Partitioned data. Study quality is explored within the self-report (Table 9) and observer data (Table 10). Of all the variables tested for heterogeneity, the only significant difference between groups is in the observer-reported group assignment variable. The heterogeneity reveals that among studies using observer reported data, studies with random assignment or matched cases have a larger treatment effect than studies using no randomization,  $Q_b(3, k = 12) = 62.625, p = 0.001$ .

Program variables: Partitioned data. There are eight program variables in each set of partitioned data. Within the self-report program variables, none show significant heterogeneity. Even among the nine distinct types of self-report scales, the effects are statistically homogenous. These results are shown in Table 11.

However, within the observer-reported outcomes, two of the variables show significant heterogeneity, the specific type of empathy scale used  $Q_b(3, k = 43) = 16.205, p = 0.001$ , and trainer modeling empathy  $Q_b(1, k = 12) = 7.362, p = 0.007$ . In observer-report studies in which the trainer models empathy, the participants have more conservative gains in empathy ( $g = 0.815, k = 5$ ) than when trainers that do not model empathy ( $g = 2.034, k = 7$ ). These results are shown in Table 12.

Two other variables show statistical heterogeneity within the observer-report data, but each of them has only one study in the comparison group precluding any conclusions from the group difference statistic.

Program ingredients: Partitioned data. Among self-report scales, the only program ingredient showing significant heterogeneity in Table 13 is the use of psycho-education  $Q_b(1, k = 47) = 4.157 p = 0.041$ . The analysis shows that programs with a

Table 9

## Self-report Outcomes: Study Variables

Group	k	Hedge's Std.		95%		Q <sub>w</sub>	Q <sub>b</sub>	df	Q <sub>b</sub>
		g	Error	conf. interval					
Group Assignment	13						1.856	2	0.395
Random	5	0.268	0.143	-0.012	0.548	0.061			
Matched Case	2	0.313	0.139	0.040	0.585	0.024			
Tested for equivalence	0								
No randomization	6	0.482	0.102	0.283	0.681	0.000			
Study Quality (ordinal)	13						1.410	2	0.494
High	2	0.090	0.325	-0.547	0.727	0.781			
Moderate	7	0.410	0.060	0.004	0.292	0.003			
Low	4	0.529	0.178	0.032	0.179	0.878			
Study Quality (meta-regression)	13	<i>b</i> =0.051	0.035	-0.121	0.019	0.155	2.025	1	0.155
Comparison groups	13						0.510	2	0.775
None: pre-post	5	0.419	0.129	0.218	0.725	0.000			
Treatment to Control	3	0.422	0.086	0.253	0.590	0.000			
Treatment to Treatment	5	0.320	0.169	-0.012	0.651	0.059			

Note: *k* = number of studies or unique effects contributing to an effect size. Variables that are consistent across effects within for studies will have a total *k* = 13. Variables reporting different levels with studies due to subgroups or outcome variables will have a total *k* = 47.

Table 10

## Observer-report Outcomes: Study Variables

Group	k	Hedges' Std.		95%		Q <sub>w</sub>	Q <sub>b</sub>	df	Q <sub>b</sub>
		g	Error	conf. interval					
Group Assignment	12	1.220	0.138	0.958	1.498	0.000	62.625	3	0.000*
Random	4	1.241	0.524	0.214	2.268	0.018			
Matched Case	5	1.806	0.403	1.016	2.596	0.000			
Tested for equivalence	1	3.390	0.327	2.749	4.032	0.000			
No randomization	2	0.517	0.173	0.179	0.856	0.003			
Study Quality (ordinal)	12						0.012	2	0.994
High	3	1.540	0.551	0.460	2.621	0.005			
Moderate	6	1.579	0.362	0.870	2.287	0.000			
Low	3	1.481	0.867	-0.217	3.180	0.087			
Study Quality (meta-regression)	12	<i>b</i> =0.014	0.119	-0.220	0.249	0.904	0.014	1	0.904
Comparison groups	12						1.723	2	0.423
None: pre-post	1	1.033	0.429	0.192	1.874	0.016			
Treatment to Control	8	1.759	0.437	0.903	2.615	0.000			
Treatment to Treatment	3	1.151	0.295	0.573	1.729	0.000			

Note: *k* = number of studies or unique effects contributing to an effect size. Variables that are consistent across effects within for studies will have a total *k* = 12. Variables reporting different levels with studies due to subgroups or outcome variables will have a total *k* = 43.

Table 11

## Self-report Outcomes: Program Variables

Group	k	Hedge's Std.		95%		Q <sub>w</sub> p-value	Q <sub>b</sub>	df	Q <sub>b</sub> p-value
		g	Error	conf. interval					
Self-Report Overall	47	0.382	0.051	0.283	0.481	0.000	13.086	8	0.109
BEES	3	0.530	0.230	0.078	0.982	0.021			
CMEM	12	0.473	0.143	0.192	0.753	0.001			
Empathy-other	3	0.605	0.226	0.163	1.047	0.007			
HES	4	0.142	0.146	-0.143	0.427	0.329			
IRI	11	0.511	0.136	0.245	0.777	0.000			
ODPT	2	0.369	0.116	0.141	0.597	0.001			
QMEE	2	0.248	0.190	-0.125	0.622	0.193			
RES	8	-0.037	0.177	-0.385	0.310	0.834			
SDPT	2	0.533	0.120	0.298	0.768	0.000			
Type of empathy measured	47						3.996	2	0.136
Cognitive	31	0.304	0.062	0.183	0.424	0.000			
Affective	11	0.547	0.162	0.229	0.865	0.001			
Both	5	0.756	0.290	0.187	1.325	0.009			
Dosage of training (meta-regression)									
Total Hours	10	<i>b</i> = -0.001				0.139	2.109	1	0.139
Total Hrs. minus outlier	9	<i>b</i> = 0.020				0.099	2.712	1	0.099
Number of sessions	7	<i>b</i> = 0.033				0.195	1.682	1	0.398
Number of weeks	9	<i>b</i> = 0.021				0.516	0.421	1	0.516
Setting for training	13						5.040	4	0.169
College	4	0.166	0.140	-0.109	0.442	0.236			
Group Home/Residential	3	0.404	0.158	0.094	0.714	0.011			
Jail/prison/secure facility	3	0.819	0.266	0.298	1.340	0.002			
Outpatient	3	0.388	0.085	0.222	0.553	0.000			
Positive Empathy	47						0.174	1	0.677
Yes	27	0.380	0.052	0.279	0.482	0.000			
No	20	0.436	0.122	0.196	0.676	0.000			
Trainer models empathy	13						2.704	1	0.100
Yes	2	0.184	0.115	-0.041	0.410	0.110			
No	11	0.413	0.078	0.260	0.567	0.000			
Modality of training	13						2.432	2	0.296
Couples	3	0.388	0.085	0.222	0.553	0.000			
Individual	1	0.560	0.087	0.389	0.732	0.000			
Groups	9	0.385	0.118	0.154	0.616	0.000			
Independent Empathy Training	13						1.092	1	0.296
Empathy training only	9	0.361	0.114	0.136	0.585	0.002			
Empathy plus other	4	0.496	0.060	0.377	0.614	0.000			

Note: *k* = number of studies or unique effects contributing to an effect size. Variables that are consistent across effects within for studies will have a total *k* = 13. Variables reporting different levels with studies due to subgroups or outcome variables will have a total *k* = 47.

Table 12

## Observer-report Outcomes: Program Variables

Group	k	Hedge's Std.		95%		Q <sub>w</sub> p-value	Q <sub>b</sub>	df	Q <sub>b</sub> p-value
		g	Error	conf. interval					
Observer Overall	43	1.220	0.135	0.995	1.485	0.000	16.205	3	0.001*
Carkhuff	22	1.206	0.192	0.830	1.582	0.000			
Empathy-other	9	2.014	0.439	1.154	2.874	0.000			
Feshbach & Roe	4	0.543	0.257	0.543	0.257	0.034			
Truax	8	2.136	0.375	1.401	2.870	0.000			
Type of empathy measured	12						0.000	0	1.000
Cognitive	12	1.541	0.272	1.009	2.074	0.000			
Dosage of training (meta-regression)									
Total Hours	10	<i>b</i> =-0.007				0.865	0.029	1	0.865
Number of sessions	10	<i>b</i> =-0.015				0.718	0.130	1	0.718
Number of weeks	11	<i>b</i> =-0.098				0.164	1.941	1	0.164
Setting for training	12						1.948	3	0.583
College	3	1.625	0.465	0.714	2.536	0.000			
Jail/prison/secure facility	2	1.009	0.275	0.470	1.548	0.000			
Outpatient	2	1.883	1.507	-1.071	4.837	0.212			
School preK-12	5	1.560	0.520	0.541	2.580	0.003			
Positive Empathy	43						0.461	1	0.497
Yes	10	1.276	0.351	0.587	1.964	0.000			
No	33	1.546	0.188	1.177	1.915	0.000			
Trainer models empathy	12						7.362	1	0.007*
Yes	5	0.815	0.157	0.507	1.122	0.000			
No	7	2.034	0.421	1.209	2.860	0.000			
Modality of training	12						7.430	1	0.006*
Individual	1	2.503	0.270	1.975	3.032	0.000			
Groups	11	1.452	0.276	0.912	1.992	0.000			
Independent Empathy Training	12						42.302	1	0.000*
Empathy training only	11	1.320	0.232	0.865	1.778	0.000			
Empathy plus other	1	4.092	0.357	3.391	4.792	0.000			

Note: *k* = number of studies or unique effects contributing to an effect size. Variables that are consistent across effects within for studies will have a total *k* = 13. Variables reporting different levels with studies due to subgroups or outcome variables will have a total *k* = 47.

Table 13

## Self-report Outcomes: Program Ingredients

Group	k	Hedge's Std.		95%		Q <sub>w</sub>	Q <sub>b</sub>	df	Q <sub>b</sub>
		g	Error	conf. interval		p-value			p-value
Audio tape	13						0.110	1	0.740
Yes	3	0.551	0.456	-0.343	1.444	0.227			
No	10	0.398	0.056	0.289	0.507	0.000			
Video tape	13						0.684	1	0.408
Yes	10	0.440	0.079	0.284	0.595	0.000			
No	3	0.242	0.225	-0.199	0.684	0.282			
Role-play	13						0.979	1	0.322
Yes	9	0.448	0.087	0.277	0.620	0.000			
No	4	0.263	0.166	-0.061	0.588	0.122			
Exposure to victim	47						3.554	1	0.059
Yes	15	0.586	0.134	0.323	0.849	0.000			
No	32	0.306	0.063	0.182	0.431	0.000			
Visualization	13						0.183	1	0.669
Yes	1	0.559	0.484	-0.350	1.549	0.216			
No	12	0.390	0.075	0.242	0.537	0.000			
Homework	13						0.210	1	0.647
Yes	4	0.447	0.074	0.302	0.591	0.000			
No	9	0.383	0.119	0.150	0.616	0.001			
Written assignments	13						1.865	1	0.172
Yes	3	0.654	0.228	0.207	1.101	0.004			
No	10	0.325	0.077	0.174	0.477	0.000			
Reading Materials	13						1.029	1	0.310
Yes	3	0.601	0.240	0.131	1.071	0.012			
No	10	0.344	0.080	0.187	0.502	0.000			
Psycho education	47						4.157	1	0.041*
Yes	30	0.297	0.065	0.169	0.424	0.000			
No	17	0.584	0.125	0.339	0.829	0.000			
Psychodrama	13						0.564	1	0.453
Yes	2	0.267	0.174	-0.075	0.608	0.126			
No	11	0.411	0.080	0.254	0.567	0.000			
Total number of training ingredients									
Meta-regression:	47	<i>b</i> =0.0387	0.041	-0.041	0.119	0.344	0.895	1	0.344

Note: *k* = number of studies or unique effects contributing to an effect size. Variables that are consistent across effects within for studies will have a total *k* = 13. Variables reporting different levels with studies due to subgroups or outcome variables will have a total *k* = 47.

psycho-educational component have lower gains ( $g = 0.297$ ,  $k = 30$ ) in empathy than programs without psycho-education ( $g = 0.584$ ,  $k = 17$ ).

Among observer reported data in Table 14, the use of an audio tape, reading materials, and psychodrama all show significant groups differences. However, all these statistically significant differences have very small numbers of studies and preclude any meaningful interpretation of the group difference.

Participant variables: Partitioned data. Among self-report outcomes, the only participant variable in Table 15 that shows significant heterogeneity is voluntary status. This difference shows that among client involved in empathy training using self-report outcome scales, involuntary clients ( $g = 0.574$ ,  $k = 6$ ) have significantly higher gains in empathy than voluntary clients ( $g=0.272$ ,  $k = 7$ ).

Among observer reported data in Table 16, gender and age show group differences. The gender of the participant is significantly related to effects of empathy training program with females ( $g = 2.248$ ,  $k = 6$ ) showing higher gains in empathy than males ( $g = 1.411$ ,  $k = 3$ ). Age also shows significant heterogeneity,  $Q_b(4, k = 12) = 11.581$ ,  $p = 0.021$  with older or younger clients showing more conservative gains in empathy, while adolescents, young adults, and adults demonstrate more robust gains. The age data was transformed into a continuous scale and a meta-regression was conducted to identify any trends in the data. The regression line was nearly flat and insignificant ( $b = 0.014$ ,  $p = 0.952$ ), showing no significant trend in empathy by age of participant.

This concludes the presentation of the partitioned data and leads to the final section of the chapter, the analysis of potential publication bias of the 24 included studies.

Table 14

## Observer-report Outcomes: Program Ingredients

Group	k	Hedge's Std.		95%		Q <sub>w</sub> p-value	Q <sub>b</sub>	df	Q <sub>b</sub> p-value
		g	Error	conf.	interval				
Audio tape	12						10.776	1	0.001*
Yes	2	2.927	0.443	2.059	3.796	0.000			
No	10	1.260	0.249	0.773	1.747	0.000			
Video tape	12						0.934	1	0.334
Yes	4	1.292	0.167	0.965	1.619	0.000			
No	8	1.713	0.402	0.925	2.501	0.000			
Role-play	43						0.295	1	0.587
Yes	31	1.424	0.183	1.066	1.782	0.000			
No	12	1.644	0.363	0.934	2.355	0.000			
Exposure to victim	12						0.000	0	1.000
No	12	1.541	0.272	1.009	2.074	0.000			
Visualization	43						0.000	1	0.993
Yes	3	1.494	0.512	0.262	0.492	0.003			
No	40	1.490	0.173	1.150	1.830	0.000			
Homework	12						0.000	0	1.000
No	12	1.541	0.272	1.009	2.074	0.000			
Written assignments	43						2.000	1	0.157
Yes	3	2.989	1.136	0.763	5.216	0.009			
No	40	1.368	0.155	1.065	1.671	0.000			
Reading Materials	12						23.717	1	0.000*
Yes	1	3.390	0.327	2.749	4.032	0.000			
No	11	1.376	0.253	0.880	1.872	0.000			
Psycho education	43						0.020	1	0.888
Yes	25	1.469	0.220	1.039	1.899	0.000			
No	18	1.516	0.256	1.015	2.018	0.000			
Psychodrama	12						11.362	1	0.001*
Yes	1	0.421	0.220	-0.011	0.854	0.056			
No	11	1.648	0.289	1.081	2.215	0.000			
Total number of training ingredients									
Meta-regression:	43	<i>b</i> =0.0973	0.157	-0.211	0.406	0.537	0.381	1	0.537

Note: *k* = number of studies or unique effects contributing to an effect size. Variables that are consistent across effects within for studies will have a total *k* = 12. Variables reporting different levels with studies due to subgroups or outcome variables will have a total *k* = 43.

Table 15

## Self-report Outcomes: Participant Variables

Group	k	Hedges' g	Std. Error	95% conf. interval		Q <sub>w</sub> p-value	Q <sub>b</sub> df	Q <sub>b</sub> p-value
Age	13						3.999 2	0.135
13-18 years	3	0.404	0.158	0.094	0.714	0.011		
19-30 years	5	0.194	0.132	-0.064	0.452	0.140		
31-60 years	5	0.531	0.106	0.324	0.739	0.000		
Gender	47						0.280 2	0.870
All Female	5	0.438	0.180	0.086	0.790	0.015		
All Male	34	0.357	0.084	0.192	0.523	0.000		
Both genders	8	0.405	0.061	0.285	0.525	0.000		
Level of Prevention	13						4.156 2	0.125
Primary	4	0.366	0.072	0.224	0.508	0.000		
Secondary	3	0.134	0.200	-0.257	0.525	0.502		
Tertiary	6	0.574	0.118	0.014	0.806	0.000		
Presenting Problem	13						4.919 4	0.296
No presenting problem	1	0.240	1.159	-0.072	0.552	0.131		
Couples	3	0.388	0.085	0.222	0.553	0.000		
Prim/secondary prevention	3	0.134	0.200	-0.257	0.525	0.502		
Delinquency/criminal	2	0.562	0.247	0.078	1.045	0.023		
Sexual offender	4	0.582	0.139	0.309	0.855	0.000		
Voluntary Status	13						4.316 1	0.038*
Involuntary	6	0.574	0.118	0.342	0.806	0.000		
Voluntary	7	0.272	0.084	0.106	0.437	0.001		
Ethnicity Reported	13						4.756 1	0.029*
Yes	6	0.252	0.101	0.054	0.450	0.013		
No	7	0.552	0.094	0.369	0.736	0.000		
% Caucasian (meta-regression)	6	<i>b</i> = -0.004	0.002	-0.009	0.002	0.211	1.565 1	0.211

Note: *k* = number of studies or unique effects contributing to an effect size. Variables that are consistent across effects within for studies will have a total *k* = 13. Variables reporting different levels with studies due to subgroups or outcome variables will have a total *k* = 47.

Table 16

## Observer-report Outcomes: Participant Variables

Group	k	Hedge's Std.		95%		Q <sub>w</sub> p-value	Q <sub>b</sub>	df	Q <sub>b</sub> p-value
		g	Error	conf. interval					
Age	12						11.581	4	0.021*
0-12 years	2	0.517	0.173	0.179	0.856	0.003			
13-18 years	2	2.828	1.249	0.381	5.275	0.024			
19-30 years	4	1.480	0.376	0.743	2.218	0.000			
31-60 years	3	1.860	0.771	0.349	3.372	0.016			
60+ years	1	0.376	0.324	-0.259	1.011	0.246			
Meta-regression	12	<i>b</i> =0.014	0.238	-0.451	0.481	0.952	0.004	1	0.952
Gender	43						34.092	3	0.000*
All Female	6	2.248	0.210	1.837	2.659	0.000			
All Male	3	1.411	0.423	0.582	2.240	0.000			
Both genders	21	1.747	0.131	1.133	2.360	0.000			
Not reported	13	0.860	0.130	0.605	1.114	0.000			
Level of Prevention	12						2.332	1	0.127
Primary	10	1.639	0.308	1.037	2.242	0.000			
Tertiary	2	1.009	0.275	0.076	0.470	0.000			
Presenting Problem	12						2.106	2	0.349
No presenting problem	8	1.491	0.328	0.848	2.134	0.000			
Parenting skills	2	2.256	1.122	0.057	4.455	0.044			
Delinquency/criminal	2	1.009	0.275	0.470	1.548	0.000			
Voluntary Status	12						2.332	1	0.127
Involuntary	2	1.009	0.275	0.470	1.548	0.000			
Voluntary	10	1.639	0.308	1.037	2.242	0.000			
Ethnicity Reported	12						0.0000	0	1.000
No	12	1.541	0.272	1.009	2.074	0.000			
Percent Caucasian (meta-regression)		No Data							

Note: *k* = number of studies or unique effects contributing to an effect size. Variables that are consistent across effects within for studies will have a total *k* = 12. Variables reporting different levels with studies due to subgroups or outcome variables will have a total *k* = 43.

### Publication Bias

One of the largest threats to the validity of a meta-analysis is publication bias. This threat has the potential to produce higher overall effects because studies with positive findings tend to be published more frequently (Epstein, 2004), and are more likely to be included in the analysis. Four separate tests were used to assess for publication bias and are described in this section.

First, the possibility of publication bias is assessed using a Rosenthal Fail-safe  $N$  (Rosenthal, 1979). This test offers an estimate of the number of additional studies *not* included in the analysis that would be needed to make the p-value nonsignificant. For this study, 3183 studies with null result would be needed make the finding non-significant to in this study. This is a large number considering the number of studies ( $k = 24$ ) included in the analysis. It is highly unlikely that over 3000 additional studies on empathy training exist.

Orwin's Fail-safe  $N$  is an alternate test for publication bias (Borenstein et al., 2009) and was also calculated to assess for publication bias. The value of Orwin's Fail-safe  $N$  is 267 ( $k = 24$ , trivial  $g = 0.05$ ,  $g = 0.00$  in missing studies), indicating that 267 additional studies with trivial or zero effect would be needed to reduce the effect size for this study to  $g = 0.05$ . No evidence of publication bias is indicated by these tests.

An alternate assessment of publication bias may be conducted by plotting effect size by the standard error of the studies. If there is no bias or missing studies, the distribution of studies should look like an inverted funnel, with studies clustered around the mean symmetrically. This is assessed through a visual inspection of the graph. In

Figure 5 we see that the distribution of effect is generally in an inverted funnel shape, although a few outliers create a somewhat unbalanced distribution.

To address the outliers and balance the distribution, Duval and Tweedie's trim and fill (Duval and Tweedie, 2000) was conducted. We can see in Figure 5 that this Trim and Fill process reduced the overall effect only slightly and reinforces our previous findings that there is no significant publication bias present in this meta-analysis.

### Conclusion

The meta-analysis of 24 studies of empathy training programs reveal that overall, empathy training programs are associated with increases in empathy ( $g = 0.911$ ). The studies in the analysis were examined for potential sources of bias including an assessment of study quality and publication bias. The analysis does not indicate a significant study quality effect or publication bias in the studies.

Tests of heterogeneity of the effects reveal significant differences in effect between programs. A moderator analysis reveals a significant difference in the effectiveness of the training related to the way empathy is measured. Empathy training programs in which an observer rates empathy show a much larger effect (1.488) than programs using self-report outcomes (0.386). This finding leads to partitioning the data by observer-report and self-report outcomes. This grouping by outcome type eliminates a large portion of the variance in the data and allowed for further exploration of group differences. Once the data are partitioned, no variables show significant group differences in both the self-report and observer-report groups of data. Several variables do show significant group differences in either the self-report or observer-report data and these differences warrant further discussion the next chapter.



## CHAPTER 5

### DISCUSSION

This chapter discusses the meta-analysis of empathy training programs for client populations. The goals of this study were to investigate empathy training programs and understand the variables that contribute to effective programs. A better understanding of these programs will assist practitioners in program development and inform future research on empathy training. The major findings of the study are summarized here and explored in more detail in the subsequent sections.

#### Summary of Results

The results show that nearly all empathy training programs are generally effective in increasing empathy. However, a simple and straight-forward statement that empathy training increases empathy is not fully supported by the moderator analysis. In addition, the analysis includes a modest number of studies that are of mixed quality and varied in their methodology. These limitations are discussed in further detail later in the chapter.

The primary goals of this study were to investigate whether empathy training is effective and if certain features of the programs lead to differential increases in empathy. For the most part, few significant moderators were found. This suggests that empathy training is robust against variations in how the training is delivered or who receives the training. It appears that empathy may be increased through a variety of training

techniques including role-plays, homework, visualization, exposure to victims, or watching a video. In addition, the length and dosage of the empathy training are not linked to program effectiveness; suggesting that increases in empathy can be realized in shorter, cost-effective programs. While the aforementioned variables did not significantly influence outcomes, the method used to measure empathy did moderate outcomes. Specifically, effects varied based on whether self-report and observer-report empathy scales were used to measure empathy. This finding raises interesting questions about how empathy is measured and the construct of empathy.

Prior to further discussion about how empathy is measured and understood, other important findings from the study are presented. The interpretation of results considers both the straightforward interpretation of the data and the measurement effect found in the moderator analysis. The chapter concludes by discussing implications of the study for practice and research, the limitations of the study, and future directions for empathy research.

The first research question is simply, “Do empathy training programs increase empathy?” It appears the answer is yes, as evidenced by 23 of 24 empathy training programs showing a large, positive effect from training ( $g = 0.911$ ). This is an important finding for direct practice workers and administrators because empathy is correlated to many psychosocial variables and social problems commonly targeted in social work. If we increase empathy through training we may be able to decrease child abuse (Weihe, 1997), aggression towards others (Jolliffe, 2006a), sexual assault (Anderson & Whiston, 2005), child molestation (Marshall et al., 1996), interpersonal violence (O’Donohue, 2003), and criminal recidivism (Jackson, 2009). Increases in empathy also have a

positive impact on social functioning (Del Barrio et al., 2004), parenting practices (Kaminski, Valle, Filene, & Boyle, 2008), family relationships (Guerney, 1988), and moral development (Kagan, 1984).

Of course, interventions designed to increase empathy may not necessarily achieve these beneficial effects reported in the literature. It may be that an increase in empathy from a training program is largely a skill, closely linked to a cognitive or behavioral change, rather than a more salutary change in affective empathy. This is an important distinction that is worthy of further discussion and consideration throughout the discussion.

The question of whether empathy training programs work cannot simply be answered with a *yes* or *no*. The second research question asks if variation in outcomes exists in empathy training programs. In fact, significant variation exists in the outcomes of empathy training ( $I^2 = 93.583$ ). This suggests that the overall effect size is not reflective of all empathy training programs. This leads to exploration of this variation through the moderator analysis.

The next question asks what variables contribute to empathy training program outcomes and what is the relative strength of these variables. The moderator analysis reveals a significant difference between empathy programs that use client self-report outcome scales and those that rely on observers to rate empathy. On average, self-report scales show moderate gains in empathy from training ( $g = 0.386$ ), while studies using an observer to rate empathy show much larger effect ( $g = 1.488$ ). The outcomes derived from the observer-reported ratings are nearly 400% larger than from self-report scales. Although the study did not initially set out to investigate the difference between these

outcomes, this moderator effect emerged from the analysis and will be important as I discuss the finding of the study.

Finding such a significant variable that moderates the overall outcome leads to subdividing the data into self-report and observer-report outcomes. Once the data are separated into these groups, not a single moderator variable that is significant in the overall data continues to be significant in the self-report and observer-report data. The lack of a robust variable that persists across self-report and observer-report groups suggests that the way empathy is measured is important.

Plausible explanations for the difference between observer and self-reported outcomes could be the effects of *research reactivity* or *social desirability bias* (Rubin, 2008). This effect could be an *artifact*, or an uncontrolled bias in the studies that can limit the confidence of results (David, 2008). These effects could artificially inflate empathy scores because the observer-reports may capture increases in empathy from the training plus an increase from being observed or from wanting to please the observers. An example of research reactivity would be that a participant would feel pressure to display empathy because they know that they are being observed. The desire to appear socially desirable could also impact the outcomes as participants model more prosocial, empathic behaviors. On the other hand, the self-report instruments may not be as vulnerable to these effects and may reflect less bias and more direct training effects. A self-report paper and pencil test may not have the same performance issues as the observed test and client may feel more comfortable to report genuine empathy changes and maybe even changes in feelings, or affective empathy.

The moderator analysis of program qualities, training ingredients, and participant variables further informs the inquiry into what makes empathy programs effective. The moderator analysis suggests several important group differences and interaction effects important to practitioners and researchers that are presented in question form in the next sections.

### What Specific Interventions Promote Empathy?

It was hypothesized that certain interventions may increase empathy more than others, yet the results suggest that empathy can be promoted equally using many different interventions. The specific interventions explored as potential moderators in this study include the use of: audiotape, videotape, role-play, exposure to victims, visualization, homework, written assignments, reading materials, psycho education, and psychodrama. Across all three groups of data; overall, self-report, and observer-report, only the use of psycho education moderated outcomes. Among self-report outcomes, training that included psycho education showed a lower training effect compared to programs that did not. This finding is surprising for two reasons. First, psycho education is often part of any psycho-therapeutic intervention (Hepworth, et al., 2010) and second, that psycho education is the most frequently used intervention in this study (61% of studies). A possible explanation for this diminished effect may be that time spent doing psycho education may be wasted in comparison to other more effective empathy interventions. It may be that educating clients about empathy may increase an intellectual understanding of empathy but this may not necessarily lead to empathic understanding. It may be that interventions like role-playing or visualization that require clients to engage both cognitively and emotionally lead to higher self-reported empathy.

### Does Training Increase Cognitive or Affective Empathy, or Both?

The literature on empathy continues to debate the true nature of empathy. According to most, an empathic response consists of two main components: cognitive empathy, or skills in recognizing and communicating about the emotional states of others (Carkhuff & Truax, 1967); and affective empathy, or the emotional responsiveness to another persons' experience (Mehrabian & Epstein, 1972). The findings of this study suggest that training can increase both cognitive and affective empathy. It is worth noting that 82% of the outcomes of the study assessed only cognitive empathy, and that all 12 studies using observer-report outcomes assessed only cognitive empathy (Table 12). This may be explained by the nature of the scoring process; as observers rate empathy according to behavioral or verbal indicators, this may simply reflect the assessment of cognitive empathy skills.

### Does the Location or Setting of the Training Make a Difference?

There is some controversy as to whether treatment should be conducted only in specialized treatment settings or if treatment can work in institutional or prison settings. The results of this study suggest that empathy training is equally effective in various settings whether it is a school, outpatient, inpatient, jail, or prison setting. This finding dispels the outdated myth that "nothing works" (Martinson, 1974, p. 50) in correctional treatment. The perpetuation of this erroneous finding results in tens of thousands of inmates receiving no treatment during their incarceration. If we can increase the empathy of criminal offenders, in conjunction with other rehabilitation services, we may reduce both the real and social costs of crime.

### Is Empathy Training More Effective in Groups?

The choice to deliver an intervention in a group, individual, family, or couples format is influenced by both theoretical models and economic pressures. Because of the interactive nature of empathy, one could predict that empathy training would be more effective in groups. The results of this study suggest that all these modalities are equally effective. Although empathy training can be delivered using any of these modalities, group delivery of empathy training may offer other advantages. First, group is generally a more cost-effective modality, so the training could be provided to a larger number of clients than training delivered in the other modalities. In addition, empathy is largely an interactive process (Freedberg, 2007). The presence of others to model after, practice with, and experiment with new ways of relating may have additional social benefits common to groups like improving communication, social skills, and support (Hepworth et.al, 2010). Given these additional social and economic advantages, it is no surprise that groups are the most common treatment modality in the study, with 19 of 24 using a group setting for empathy training.

### Are Results Durable Over Time?

Understanding the lasting effects of any intervention is certainly important. Unfortunately this meta-analysis cannot provide a reliable estimate of the durability of effect because only 5 of the 24 studies report follow-up data precluding reliable analysis. The descriptive data are never-the-less promising. The follow-up times ranged from 2 to 15 months and 62.5% of the empathy scores continued to show significant positive

change from baseline empathy scores. Clearly, further research is needed to ascertain the effect of empathy training over time.

### Does the Dosage of the Training Influence Outcomes?

Results suggest that significant empathy gains can be realized in very short time periods and higher dosage programs are no more effective than shorter programs. Three indicators of dosage were used: total hours of training, the length in weeks of the training, and the total number of sessions. The meta-regression of all three dosage variables by effect is non-significant and the regression line is flat. The results suggest that programs with as little as an hour of empathy training are as effective as the higher dosage programs. Although the analysis failed to discriminate between different dosages, it seems unlikely that empathy training effectiveness is completely unrelated to dosage.

One possible explanation is that empathy training may trigger a large initial jump in empathy scores, but may reflect simple skill acquisition more than empathy. It is possible that empathy training programs, especially programs using observer-scored outcomes, may teach to the test. In this scenario, clients may display highly empathic responses immediately following training, but changes in future empathy and the internalization of empathy is unknown. The learned responses may look like empathy, and may actually reflect an increase in cognitive empathy, but really tell us nothing about changes in the emotional responsivity or affective empathy of the client. A second explanation is that a large initial increase in empathy may reflect a pre-disposition and a natural affinity for empathy that is rooted in our neuroanatomy (Frith & Frith, 2006) as well as early socialization experiences (Bowlby, 1982). It may be that we are *wired* for empathy and that clients respond readily to even low-dosage interventions.

Whatever the reason, it does appear that empathy can be increased relatively quickly and efficiently. Low-dosage programs can be easier to implement in a variety of social service settings, can be easier to add to existing treatment, and could be a good intervention for populations that may be hard to reach or retain in treatment.

#### Does the Number of Training Ingredients Influence Outcomes?

In general, psychological and psychosocial interventions that are multi-modal and multi-faced have better outcomes than programs that use a single modality (Lipsey & Cullen, 2007; Wilson & Lipsey 2007). The findings of the current study are not consistent with this trend. In fact, the results suggest that using a larger number of training activities is not necessarily better. Programs that provide only one or two empathy activities seem to be as effective as programs that have five or more activities. The lack of a relationship between the number of training ingredients and treatment effect suggests that programs need not be complicated and multifaceted to be effective. It could be that one or two empathy activities with depth, rather than breadth, may be most effective in increasing empathy.

#### Does the Age of the Client Influence Outcomes?

Client age is an important variable for psycho-therapeutic interventions as it may suggest developmental readiness and cognitive flexibility. In empathy training programs, previous studies suggest that older adults may not respond as readily to empathy training (Steibe, 1979). In contrast, this study finds no relationship between client age and treatment effect in both the moderator analysis and the meta-regression of age group and

treatment effect. Age does not moderate the effect of empathy training and clients of all ages can show similar increases from training.

### Does the Gender of the Client Influence Outcomes?

There is persuasive evidence in the literature that females have more empathy than males (Owen-Anderson et al. 2008; Bohart & Greenberg, 1997), and they benefit more from empathy training programs (Baron-Cohen, 2003). Although another study reports that while females self-report higher levels of empathy, if empathy is rated unobtrusively, there is no difference by gender (Norris, 2006). The results of the current meta-analysis suggest that both males and females can benefit from empathy training, and that females may realize greater gains than males in certain conditions. The differential impact has provocative program and practice implications.

The overall results do show a strong gender-treatment effect relationship with female-only groups ( $g = 1.363$ ) and mixed-gender groups ( $g = 1.298$ ) showing the greatest gains and male-only groups ( $g = 0.403$ ) with the lowest effect. The female-only and mixed-gender group effects are statistically homogenous. Gender was also explored through meta-regression that suggests groups with larger percentages of males have lower empathy gains. These findings suggest that mixed-gender groups may be preferred to single-gender groups if the goal is to increase empathic behaviors in males.

Interestingly, the *gender X treatment* effect disappears within the self-report only data, revealing that when people self-score their empathy, males and females benefit equally from training. This effect may be due to one of the following explanations. First, the stereotypical view of females as more empathic could introduce a bias in the way observers rate empathy, artificially inflating empathy scores for females in the overall

and observer-reported data. Alternately, male participants may not respond as readily to the largely skill-based or cognitive empathy training, but realize similar gains as females in feeling, or affective, empathy reflected in self-report data.

#### Is Empathy Training More Effective at Certain Levels of Prevention?

The prevention level of the program was not related to differential treatment effects. The effect of empathy training for nonindicated populations (primary prevention) is the same for at-risk populations (secondary prevention) and for indicated populations (tertiary prevention). Thus, empathy training programs could be promising for various prevention and intervention programs.

#### Does Empathy Training Work Across Client Populations?

One of the most common explanations for why people hurt others is that they lack empathy (Finkelhor & Lewis, 1988; Barbaree et al., 1979). In fact, good studies on serial criminals and sexual offenders find this explanation to be false (Varker et al., 2008; Fernandez et al., 1999; Marshall et al., 1995). The present study also suggests that empathy training works equally well with criminal offenders and sex offenders as it does for college students, couples, and parents.

Client presenting problem was analyzed both categorically as well as on a scale of dangerousness. Both analyses suggest that in populations ranging from non-indicated, non-clinical clients through adult sexual offenders, empathy training can be effective. This finding is contrary to more traditional ideas about criminals lacking or being incapable of empathy and confirms more recent findings about criminals being generally empathic, but lacking in certain types of specific empathy that may be attributed to

cognitive distortions (Jolliffe, 2004). This finding does not show that if we increase empathy in dangerous populations that we necessarily reduce the recidivism of an individual offender. It does show that empathy training could be part of a larger treatment program to reduce recidivism, even for dangerous offenders.

While we seem to be able to increase empathy through training, this doesn't mean individuals will choose to use those skills in the future. While most people receive positive social reinforcement for empathic responses (Guernsey, 1988) clients residing in violent communities or institutional settings may not derive much social benefit from empathy. In fact, empathy could be a liability in violent or life-or-death circumstances. In essence, we may be able to improve empathy, but this may not mean that the client will be motivated to use these skills in the future. Future research is needed to investigate the relationship between empathy, social reinforcement, and motivation to be empathic so we can further understand how to maximize gains from empathy training programs.

#### Does Voluntary Status Influence Outcomes?

Compulsory treatment may seem to be an oxymoron because internal motivation is such an important clinical variable (Miller & Rollnick, 2002). One could imagine that involuntary populations may be less receptive to treatment and have lower response to treatment. The results of the present study suggest that both involuntary and voluntary populations can benefit from empathy training. The old clinical wisdom that clients have to *want to change*, which may be just an excuse not to treat difficult clients, is challenged by these findings. Intention to change and treatment engagement are the two factors most closely linked to motivation (Lundahl, Kunz, Brownell, Tollefson, & Burke, 2010). It may be that involuntary treatment starts out being motivated by external forces like the

threat of legal sanctions or other coercive forces. Once involuntary populations are engaged in treatment, the source of the motivation may shift from external to internal. This phenomenon is similar to *equifinality*, or the systems theory concept that similar outcomes can be achieved from different starting points (Hepworth et al., 2010). Empathy training can and should be implemented with involuntary populations, especially with clients who have a history of violence or potential to harm others.

#### Do Certain Ethnic Populations Respond Better to Empathy Training?

There is limited data on the ethnicity of the clients in the studies. In fact only 13 of the 24 studies report any ethnicity data, and most report insufficient data for a thorough analysis. Although the data are limited, a meta-regression was conducted although the power of the test is limited. The meta-regression of effect by percent Caucasian in the sample is possible and reveals no relationship between ethnicity and effect of empathy training.

#### Contribution to the Literature on Empathy

This is the first meta-analysis of empathy training programs for client populations. Although there is a large body of theoretical literature and correlational studies on empathy, there has been a significant gap in the literature on empathy training and interventions for client populations. This present study is consistent with results of three prior meta-analyses on empathy training with psychotherapists and helping professionals (Baker & Daniels, 1989; Baker, 1990; Hill & Lent, 2006) and one meta-analysis of empathy training with medical students (Stepien & Baernstein, 2006). All found that

empathy training has moderate to large effect sizes. In addition, all the previous studies note significant heterogeneity of effects.

The most recent meta-analysis (Hill & Lent, 2006) included 14 studies and 526 participants, the majority of whom were undergraduate or graduate students in the social sciences. Hill and Lent (2006) found that multimethod interventions were more effective than single methods, while the present study found no relationship between number of methods and treatment effect. Hill and Lent (2006) also found that modeling empathy was the most effective training method, while the present study found no relationship between a trainer modeling empathy and effect.

In a previous meta-analysis of empathy training with medical students (Stepien & Baernstein, 2006) that includes 13 quantitative and qualitative studies found a positive effect from training in seven of eight quantitative studies, although no overall effect is reported.

The present study on empathy training differs from these prior meta-analyses on empathy in that it explored client, not professional populations, explored a larger number of moderating variables, and includes assessments of study quality and publication bias that are absent from the previous studies. In addition, the present study includes a larger number of studies using self-report outcomes (13). In the four previous meta-analyses, only two of the 166 total studies in these analyses used self-report outcomes.

Clearly more primary and synthesis research on empathy training programs is needed, especially on client populations. The present study fills a gap in the extant literature on empathy while provoking additional questions about the measurement of outcomes and moderating variables that influence outcomes.

### Implications of the Study

The study found that empathy training can increase empathy, and that the way in which empathy is measured is a powerful moderator of the strength of the effect. These findings, based on a large difference between self-report and observer-report outcomes, generates several hypothesis with implications for theory, research, and practice.

First, this measurement bias could simply be a reflection of the complex nature of empathy. The debate in the empathy literature on the very construct and definition of empathy that centers around five main questions; 1) is empathy a cognitive or affective process, or both (Thornton & Thornton, 1995), 2) what constitutes “true” empathy (Thwaites, 2007), 3) does an empathic response proceed through a sequence of cognitive and affective responses (Davis, 1983), 4) how do we accurately measure empathy (Jolliffe & Farrington, 2006b), and 5) can empathy be taught (Monroe, 2006)? This study really only answers the last question adequately, and informs further inquiry into the fourth question, but the remaining questions are still largely unanswered and likely have an impact on the results of the study. There are still important questions about empathy remaining that need further investigation to inform better training programs.

Second, gains in the self-report outcomes may more accurately reflect internal emotional states, or affective empathy, and observer-report outcomes may reflect changes in cognitive empathy. Most empathy training programs are largely psycho education and skill-based, so as clients incorporate empathy knowledge and skills from the training, it is manifested in their observable behaviors. When observers rate empathy based on these verbal or behavioral indicators, this score may reflect changes in cognitive empathy (Hill & King, 1976). We may only see moderate gains in affective empathy because most

programs do not necessarily target this in the training. Further, this observed behavior-based empathy may be an additional type of empathy divergent from affective or cognitive empathy altogether and worthy of further investigation.

For example, it may be that empathy training programs are highly effective in teaching empathy-like social skills, but the training may have little impact on the development of a deeper empathic orientation towards others that some have called a *true* empathic response (Thwaites, 2007). This motivation to be empathic could be a significant factor in the long-term outcomes from training (Terry & Mitchell, 2001). In fact, a possible explanation of why empathy training seems to be effective across a variety of settings and with a variety of clients may be that at the moment of assessment the clients are motivated to be empathic. If the motivation to be empathic is a factor in empathic behaviors, it is important that empathy enhancement programs focus on motivational factors in addition to empathy skills.

Behavior-based empathy may merely be a social skill that makes people more effective interpersonally, but lacks the important benefits of *true* empathy, like deterring hurtful behavior towards others. Alternately, teaching behavioral empathy may have positive long-term outcomes. As clients realize gains in social interactions they are less likely to engage in harmful or anti-social behaviors because their needs are being met in prosocial ways (Glasser, 1975).

A third hypothesis is empathy really does proceed through a sequence of cognitive and affective responses (Davis 1980; Graziano, 2007; Marshall, 1995) and that interventions should be tailored to these stages. This process is similar to the transtheoretical model (Prochaska & DiClemente, 1982) that assesses a client's readiness

for change and then targets interventions with the goal of moving them towards the next level. For example if a client is in the pre-contemplation stage, the most effective interventions would assist them in moving towards contemplation. Effective empathy training programs could identify the stage at which the empathic response stalls and target interventions that encourage the progression of the response.

A fourth hypothesis generated from the study is that the self-report and observer-report empathy outcomes are completely divergent scales, especially considering their sensitivity and scoring procedures. The scoring on the Carkhuff Empathic Understanding Scale (Carkhuff, 1969), the most common observer scale used in this analysis, is accomplished by observers rating a participant on a 1- to 5-point scale. Using this limited range of scoring options, even a one point increase represents a large (20%) change in empathy. This results in a very sensitive empathy instrument that may yield large quantitative increases in empathy for relatively small observable changes. This is in contrast to the way most self-report scales are scored. For example, the most commonly used self-report scale, the Interpersonal Reactivity Index (IRI; Davis, 1980) has 28 questions each rated on a 5-point scale. The larger number of items on the scale may yield a more stable measure of empathy and increase the specificity of the IRI and other self-report measures, while decreasing sensitivity to change. These differences may contribute to the more conservative gains found in the self-report compared to observer-report outcomes. This hypothesis may inform future research on empathy instruments and selection of the appropriate outcome scales for intervention research.

Fifth, the potential bias from using nonblinded observers in many of the studies could artificially inflate the outcomes for these groups. Future studies on empathy must

be aware of the potential for bias by participants as well as researchers, and design studies to protect against these threats to validity. The ubiquitous nature of empathy may require even higher standards of methodological rigor to protect against contamination of the results.

A final implication is that while this study suggests that empathy training is effective, the results cannot predict that these changes will necessarily result in some of the positive outcomes seen in the correlational studies on empathy. The use of observer-report outcomes in empathy training programs, that may correspond to cognitive empathy, represents a disconnect from the way empathy is assessed in correlational studies on empathy. The correlational studies on empathy utilize almost exclusively self-report data (Varker et al., 2008) that may more accurately reflect affective empathy. In essence, empathy gains from training observed in the lab may not necessarily be associated with many of the positive outcomes from high empathy in correlational studies. Clearly, we need additional investigation of empathy training and social outcomes that use a consistent measurement and construct.

#### Limitations of Study

In spite of substantial efforts to maintain the scientific integrity of the research project and conduct the research according to established protocols, the study has several limitations.

The first is the limitation based on sampling frame common to most research projects and especially important in a meta-analysis. The study relies on primary studies to report accurate data. Although an assessment of study quality was conducted, this is no assurance that the data in the primary studies are accurate. In addition, it is unlikely

that the search strategy employed in the study captured all studies on empathy training. In fact, because the construct of empathy overlaps with many other related skills and feelings like *perspective taking, concern, sympathy, understanding, compassion, identification, and responsiveness*, there are surely studies that were missed in the search or eliminated by the exclusionary sorting process. The tests for publication bias are one way that I attempted to compensate for this inherent bias. None-the-less, the results must be viewed with this caveat.

The statistical tests used in the analysis present another limitation. Specifically, the sensitivity of the  $Q$  statistic and meta-regression are limited. This may limit the test's ability to discriminate between groups, when there may in fact be a group difference, leading to a type II error. The use of meta-regression with fewer than 10 high-quality studies is not recommended (Higgins & Green, 2009) so meta-regression results generated from this analysis are really exploratory in nature. The statistical tests are designed to provide evidence that the true effects vary and variation is indicated by a significant  $p$ -value. While a significant  $p$ -value indicates that there is a group difference, a nonsignificant value does not indicate homogeneity because the nonsignificant finding may be due to low power (Borenstein et al., 2009). In this study low power could result from two factors: sample size and large within- and between-study variance.

The relatively small number of studies that met final inclusion criteria limits the power of the statistical tests. The small number in the study is surprising because a preliminary search of PsycINFO, MEDLINE, and ERIC using the term *empathy* returned 15,766 articles. Further searching using the terms *empathy and training* returned 2589 articles. This large number of studies on empathy training led to estimating that at least

100 studies would be included in the final analysis and this larger number of studies would allow for more precise detection of groups differences. Unfortunately only 24 studies are included in the analysis, limiting power. In addition, the moderator effect that led to partitioning the data by outcome type further reduced the sample size and reduced the sensitivity of the analysis. Dividing the sample limited our ability to draw some conclusions with confidence, even when there were significant findings. For example, in a few cases even when the  $p$ -values are significant, the number of studies in one of the groups is so small that drawing conclusions is impossible.

The large variation both within and between studies as indicated by the large  $I^2$  and  $Q$  values indicate substantial variance in the studies. With variance indicators this high it could be difficult to detect group differences and so there may still be moderator effects that could not be assessed effectively in this study.

Another limitation of the study is that a large number of variables were explored through moderator analysis and this may have resulted in a type I error for some of the variables. Type I error can result from the probability error inherent in inferential statistics. A likelihood of a type I error increases from testing a large number of variables and can be analogous to *fishing for results*. This study tested for interaction effects among a number of moderators, and so there is an increased likelihood that some yield a significant  $p$ -value when there is no real difference. The analysis of multiple moderators within an individual study has the potential to violate the assumption of independence. Therefore, results emerging from the moderator analyses using the 90 unique effect sizes, compared to the 24 study-level effect sizes, is more exploratory in nature. I attempted to control for type I error by formulating research questions and variables *a priori*, by

limiting the number of variables tested in the analysis, and by using study-level effects for analysis whenever possible.

The analysis of multiple moderators within an individual study has the potential to violate the assumption of independence and therefore results emerging from the moderator analyses using the 90 effect sizes, compared to the 24 study-level effect sizes, is more exploratory in nature. I attempted to control for type I error by formulating research questions and variables a priori, by limiting the number of variables tested in the analysis, and by using study-level effects for analysis whenever possible. Nevertheless, type I error is always a threat to the findings and why results must be both reported and viewed with appropriate caution.

#### Future Directions

Practitioners should consider implementing empathy training programs to increase empathy. Although it is unknown if the training results in the simple acquisition of additional prosocial skills or actual changes in empathy, it appears promising. There are a large number of theoretical and descriptive empathy studies in the literature and relatively few that provide practical, practice-based guidance about changing empathy. Future research should focus more on effectiveness and outcome research rather than on discourse and development of theoretical models.

The study found few high-quality research articles on empathy training which identifies the need for practitioners and researchers to design and publish rigorous effectiveness studies on empathy training programs. These studies should include the use of randomized control and comparison groups, and collect pre-post as well as follow-up data to assess the durability of effect. Practice-based researchers should use validated

outcome instruments and consider using multiple outcome scales to better capture the full range of empathic responsiveness. Based on the findings that empathy changes varies as a function of the outcome instrument used, researchers should investigate and contrast the psychometric qualities and validity of empathy scales. This may lead to the modification of existing empathy instruments and the development of new ones.

One of the goals of the study was to identify variables and practices which may contribute to effective training programs. Unfortunately, the study was unable to adequately discriminate between most program and client variables and does not adequately answer the classic question; “What treatment, by whom, is most effective for this individual, with that specific problem, and under which set of circumstances?” (Paul, 1967, p. 111). This leaves several questions about empathy training unanswered. Further evaluation of the variables within empathy programs is needed to inform the development of best-practice guidelines.

Finally, the ongoing evaluation of empathy training programs requires reliable and valid outcome instruments. Further research that unifies the construct, operationalization, and measurement of empathy is vital to program development. This is a worthy project that will require ongoing collaboration between theorists, researchers, and practitioners.

### Conclusion

This study is the first meta-analysis on empathy training programs for client populations. The findings suggest that empathy training increases empathy, that the effect is robust across program and client types, and that the strength of the effect seems to be associated with how empathy is measured. The study finds that when groups of clients

self-report empathy, the effects from training are moderate; and that when empathy is rated by observers, the training effects are large. These findings lead to additional questions about the construct of empathy, the measurement of empathy, and if empathy training programs increase empathy-like skills or true empathy. The findings of the study lead to recommendations for practitioners and implications for further research.

## APPENDIX A

### STUDY SEARCH AND SORTING STRATEGY

Use Advanced Searches in all Databases:

Search terms: *empathy and (train\* or interven\* or treat\* or educ\* or program\*)*

Limiters selected in advanced search menus: human, English language,

Date of Search: 7-16-09

#### SEARCH

1. Terms: empathy and train\* or victim and empathy
2. Databases to search using EBSCO:
  - a. PsycINFO
  - b. MEDLINE
  - c. CINAHL
  - d. ERIC
3. Social Work Abstracts (Ovid): *separate search*
4. CSW Illumina: *separate search*
  - a. Criminal Justice Abstracts
  - b. Social Service Abstracts
  - c. Sociological abstracts
5. Method
  - a. Use EBSCO login to track articles and search strategy, criteria (if available)
  - b. Print off results of search for other, non-EBSCO databases
  - c. Convert abstract lists into Adobe to maintain integrity of document and to print, screen and vet.
  - d. Organize articles by database

#### SCREENING AND VETTING: MAKING DECISIONS

1. At the abstract level:
  - a. Did it manipulate empathy levels?
    - i. If yes, get the full text
    - ii. If no, reject
2. If it manipulated empathy levels: (DOCUMENT HOW MANY ARTICLES REMAIN)
  - i. Does it deal with clients
    1. If yes, go to step 3
    2. If no (professional related studies), then abort
3. Review the full text to determine if study reports numbers adequate for statistical analysis (look at PDF or request thru ILL)

- i. If there are numbers in the results section (e.g., frequencies, means, standard deviations, etc.) or statistics in the results section (e.g., t-value,  $\chi^2$ , f-value, p values, etc) KEEP THE ARTICLE
- ii. If no to #3i, abort
- iii. If in doubt, keep articles for now and decide later

#### ARTICLE MANAGEMENT FOR KEPT ARTICLES

1. Import all citations into endnote
2. maintain electronic copy of article on cd and thumb drive
3. Print and alphabetize
4. ACCOUNTING: Track how many articles we start with and are eliminated at each step.

## APPENDIX B

### PROTOCOL FOR SCREENING, SORTING, AND CODING

#### ARTICLES FOR INCLUSION

1. 8-24-09 GRA Training and Orientation
  - a. Attend initial training to review inclusion criteria and coding strategy
  - b. Meet bi-weekly to review progress (either in person or via phone)
2. Screening part 1: Pre-Screen:
  - a. screen found articles for preliminary inclusion by reading abstracts according to following criteria:
    - i. Research design: Study must be: a pre-post design, comparison groups, or treatment vs. control
    - ii. Must be an empathy intervention study
    - iii. Qualitative studies excluded
    - iv. Study must have at least 5 subjects in each treatment group
    - v. Research subjects must be a client population-
      1. exclude any professional training program (med students, nurse, counselors) or program that targets vocational skills (training of para-professionals and volunteer staff are excluded)
      2. General college students would be included unless the study is targeting students in a professional training program (i.e. SW students would be excluded, but a sample of fraternity or psych 101 students would be included)
  - b. If in doubt about design, study, data, or population, keep study for further investigation.
  - c. Report articles kept at this step by listing record #, name of first and second author, date of publication, and database retrieved from.
  - d. GRA to screen 20% (600 articles)
  - e. Email list to rob to calculate reliability
3. Screening part 2: Full text article screen
  - a. Read full text of articles that passed through first screening of abstracts.
  - b. Specifically focus on method section to verify criteria in step 3 above.
    - i. data reported-must report quantitative data for the empathy DV
  - c. Must report some empathy scale or subscale as outcome variable (DV)
    - i. Ensure there are at least 5 subjects in each treatment group (cell size).
    - ii. Must report quantitative data-qualitative studies are excluded

1. Quantitative statistics reported-Means, sd, frequencies
  2. Need enough raw data to calculate group differences or have statistical test reported.
  - d. List articles kept through this screening step in similar format as 1h above.
  - e. GRA to sort 20% (30)
  - f. Email to Rob for reliability calculation
4. Code retained articles:
- a. GRA to double-code 100%
  - b. code articles using coding sheet and code book
  - c. code study quality using ETP study quality rating form
  - d. return coded study, code sheet, and SQRF to Rob
  - e. reliability of coding checked, discrepancies resolved through discussion

## APPENDIX C

### CODE SHEET AND CODE BOOK

#### *Program Information*

1. Name of Author(s) \_\_\_\_\_

2. Date (year) of Publication \_\_\_\_\_

3. Modality of treatment: (circle)	Tx group 1	Tx group 2	Control group
a. individual	Y	Y	Y
b. group	Y	Y	Y
c. family	Y	Y	Y
d. couple	Y	Y	Y

Code a if primary intervention is one-on-one clinical intervention

Code b if primary intervention takes place in groups of 3 or more unrelated clients

Code c if primary interventions take place in therapy with one or more family members (family may be relatives or other members of natural support, not professional support systems)

Code d if primary interventions take place in therapy with an intimate partner

4. Program dosage (continuous variable)	Tx group 1	Tx group 2	Control
a. number of weeks start to finish	_____	_____	_____
b. number of sessions total	_____	_____	_____
c. total contact hours	_____	_____	_____
d. unable to determine dosage	_____	_____	_____

5. Role of interventions	Tx group 1		Tx group 2		Control group	
a. empathy training alone	Y	N	Y	N	Y	N
b. ETP PLUS other program	Y	N	Y	N	Y	N

List other program components \_\_\_\_\_

6. Training ingredients to increase empathy	Tx group 1	Tx group 2	Control
a. role play	Y	Y	Y

b. listen to audio tape	Y	Y	Y
c. watch video	Y	Y	Y
d. exposure to victim	Y	Y	Y
e. letter to victim	Y	Y	Y
f. visualization	Y	Y	Y
g. independent reading materials	Y	Y	Y
h. psycho-education	Y	Y	Y
i. positive empathy	Y	Y	Y
j. cultural component	Y	Y	Y
k. spiritual component	Y	Y	Y
l. other _____	Y	Y	Y

Total number: (# of Y in each category)      \_\_\_\_\_      \_\_\_\_\_      \_\_\_\_\_

Operational definitions:

- a. acting or taking on another's role experientially
- b. listening to an audio stimulus
- c. watching a movie or video clip of any sort
- d. in-person exposure to victim or victim panel
- e. writing a letter or apology or any writing exercise or journaling
- f. guided or individual exercise by which participant visualizes an event
- g. given reading material about victimization, impact of crime, statistics about consequences of crime or victimization
- h. live presentation about victimization, impact of crime, statistics about consequences of crime or victimization-not personalized processing
- i. program provides intervention that promotes a fuller range of emotional responsiveness i.e. happiness, joy, excitement for another
- j. component drawn from a religious or spiritual source
- k. component culturally based or drawn from a cultural ritual
- l. other significant program component not listed above (name)

7. Interactive empathy-presence of trainer/therapist empathic responding

- a. yes-explicit
- b. yes-present but not explicitly stated.
- c. no

Operation definitions:

- a. Code a if program explicitly states that trainer/therapist will engage in empathic responding with client.
- b. Code b if program does not explicitly state that trainer/therapist will model empathy but the training ingredients require the trainer/therapist to engage in empathic activities towards client.

- c. Code c if program shows no evidence of trainer therapist modeling or engaging in empathic responding.
8. Empathy Rating Scale (outcome DV):
- IRI
  - (M)BEES
  - HES
  - QMEE
  - BES
  - Carkhuff/Truax
  - RES (Rape empathy scale)
  - Other empathy scale: name \_\_\_\_\_

*Participant Information (circle letter)*

1. Reported mean age in years:

<u>Tx Group 1</u>	<u>Tx group 2</u>	<u>Control Group</u>
a. Child <12	a. Child <12	a. Child <12
b. Adolescent 13-18 y/o	b. Adolescent 13-18 y/o	b. Adolesc.13-18 y/o
c. Young adult 19-29	c. Young Adult 19-29	c. Young Adult 19-29
d. Adult 30-60	d. Adult 30-60	d. Adult 30-60
e. Senior 61+	e. Senior 61+	e. Senior 61+

If no age reported, but not a child, code as d, Adult

If no age reported but not adult, code as b, Adolescent

Code College students as c, young adult

2. Cognitive impairment

- yes-impairment identified
- no-not reported

Code a if study explicitly states clients have some cognitive impairment, learning disability, mental retardation, or reports IQ lower than 80-*not any axis 1 diagnosis (ADHD)*

3. Specific diagnosis of participants
- |                         | <u>Tx group 1</u> | <u>Tx group 2</u> | <u>Control group</u> |
|-------------------------|-------------------|-------------------|----------------------|
| a. Yes                  | Diagnosis _____   | _____             | _____                |
| b. No or none indicated |                   |                   |                      |

Code a if study explicitly reports a diagnosis of participants

4. Client presenting problem (ordinal scale for severity/dangerousness)

- life enhancement/not clinically indicated
- primary/secondary prevention program

- c. couples counseling/intimate relationships
  - d. parenting skills
  - e. child or youth aggressiveness/social skills/bullying
  - f. domestic violence/interpersonal violence perpetrator
  - g. delinquency/criminal behavior-property crime(not sex offending)
  - h. delinquency/criminal behavior-violence against person crime
  - i. sexual offender
  - j. not reported
5. Prevention/timing of program (ordinal for level of intervention)
- a. primary prevention-target to general population
  - b. secondary prevention-targeted to at-risk population
  - c. tertiary prevention-indicated post problem
6. Voluntary status
- a. voluntary
  - b. mandated-legally (under some coercive control, i.e. court or incarcerated)
  - c. not reported/unable to determine
7. Setting of program
- a. school (pre-k-12)
  - b. college
  - c. outpatient setting
  - d. day treatment
  - e. group home/residential community program
  - f. jail/prison/secure psychiatric facility
8. Gender
- |                             | Tx group 1 | Tx group 2 | Control group |
|-----------------------------|------------|------------|---------------|
| a. male only                | Y          | Y          | Y             |
| b. female only              | Y          | Y          | Y             |
| c. mixed gender             | Y          | Y          | Y             |
| d. If provided %male/female | ___/___    | ___/___    | ___/___       |
9. Ethnicity
- |  | Control % | Treatment % | Combined % |
|--|-----------|-------------|------------|
| (Circle one or more, give percentages if provided) |           |             |            |
| a. White/Caucasian                                 | _____     | _____       | _____      |
| b. Hispanic/Latino                                 | _____     | _____       | _____      |
| c. African American                                | _____     | _____       | _____      |
| d. Asian   | _____     | _____       | _____      |
| e. Native American                                 | _____     | _____       | _____      |

f. Pacific islander	_____	_____	_____
g. Other	_____	_____	_____
h. Not reported	_____	_____	_____

Effect Sizes

<u>Outcome Name &amp; Description</u>	<u>Effect Size Value</u>	<u>Number of participants in study (Treatment groups &amp; control)</u>	<u>Notes: table that data reported, page number, other ID info</u>

## APPENDIX D

### STUDY QUALITY RATING FORM-EMPATHY TRAINING PROGRAMS

---

Study # plus last name of first author and year of  
publication

#### Selection Bias-4

- Group Assignment:
  - 3 = True randomization and/or propensity score reported
  - 2 = Matched group or case control
  - 1 = Tested for pre-treatment equivalence
  - 0 = Violated randomization, nonequivalent
- Indicators of population demographics
  - 1 = Reported 3 or more
  - 0 = Reported 2 or less

#### Performance Bias-5

- Fidelity in ETP Delivery, manualization of program
  - 1 = ETP standardized: manual, specific training
  - 0 = No evidence of standardized delivery
- Fidelity and Adherence to ETP
  - 2 = Assessed: High Fidelity or *Supervision*
  - 1 = Assessed: Moderate fidelity
  - 0 = No mention of fidelity or supervision
- Blinding
  - 2 = Participants and researchers
  - 1 = Participants or researchers or coders blind to condition
  - 0 = No blinding of participants, researchers, or coders

#### Attrition Bias-4

- Does study assess attrition or conduct intent to treat analysis?
  - 1 = yes
  - 0 = no

- Completion rate of participants
  - 2 = >85%
  - 1 = 70% to 84.9%
  - 0 = Less than 69.9% or not mentioned
- Specific measures taken to minimize loss of subjects from study
  - 1 = Yes
  - 0 = No

#### Detection Bias-4

- Measurement of Outcomes: Informant
  - 3 = Blind or independent observer
  - 2 = Participant + Collateral (e.g., significant other)
  - 1 = Collateral Only
  - 0 = Participant only (self report only)
- Validated empathy or psychometric scale used to measure main effect
  - 1 = Commonly accepted scale or subscale ((M/BEES, IRI, HES, QMEE, RES, BEES, Carkhuff/Truax)
  - 0 = Proprietary or researcher created scale

#### Reporting Bias-3

- Data to calculate Effect Size
  - 2 = Means, SD, # participants; or percent; or frequency
  - 1 = exact statistic reported: t-test, F-test, Odds ratio, etc. OR reports exact probability (.05, .01)
  - 0 = p values only or indication of “significant” or not
- Data for all assessed Dependent Variables reported?
  - 1 = Yes
  - 0 = No

Total Score: /20      Overall Study Quality:      High      Moderate      Low

Comments:

## REFERENCES

*References marked with an asterisk "\*" indicate studies included in the meta-analysis.*

- Ainsworth, M. & Bowlby, J. (1991). An ethological approach to personality development. *American Psychologist*, 46(4), 333-341.
- Albiero, P., Matricardi, G., Speltri, D., & Toso, D. (2009). The assessment of empathy in adolescence: A contribution to the Italian validation of the Basic Empathy Scale. *Journal of Adolescence*, 32, 393-408.
- Anderson, L., & Whiston, S. (2005). Sexual assault education programs: A meta-analytic examination of their effectiveness. *Psychology of Women Quarterly*, 29(4), 374-388.
- Andrew, J., Cooke, M., & Muncer, S. (2008). The relationship between empathy and Machiavellianism: An alternative to empathizing–systemizing theory. *Personality & Individual Differences*, 44(5), 1203-1211.
- \*Angera, J. and E. Long (2006). Qualitative and quantitative evaluations of an empathy training program for couples in marriage and romantic relationship. *Journal of Couple & Relationship Therapy*, 5(1), 1-26.
- Apperly I.A., Samson D., & Humphreys G.W. 2009. Studies of adults can inform accounts of theory of mind development. *Developmental Psychology* 45(1), 190-201.
- Aristotle. (1976). *Ethics* (J. Thompson, Trans.). London: Penguin Books.
- Baker, S., & Daniels, T. (1989). Integrating research on the microcounseling program: A meta-analysis. *Journal of Counseling Psychology*, 36(2), 213-222.
- Baker, S., Daniels, T., & Greeley, A. (1990). Systematic training of graduate-level counselors: Narrative and meta-analytic reviews of three major programs. *The Counseling Psychologist*, 18(3), 355-421.
- Bandura, A. (1969). *Social foundations of behavioral modification*. New York: Holt, Rhinehart, & Winston.
- Bandura, A. (1999). Moral disengagement in perpetration of inhumanities. *Personality and Social Psychology Review*, 3, 193-209.

- Barbaree, H., Marshall, W., & Lanthier, R. (1979). Deviant sexual arousal in rapists. *Behaviour, Research and Therapy*, 17(3), 215-222.
- Baron-Cohen, S. (2006). The hyper-systemizing, assortative mating theory of autism. *Progress in Neuropsychopharmacology and Biological Psychiatry*, 30, 865-872
- Baron-Cohen, S. (2003). *The essential difference: Men, women and the extreme male brain*. London: Penguin Books.
- \*Berg, D., Lonsway, K., & Fitzgerald, L. (1999). Rape prevention education for men: The effectiveness of empathy-induction techniques. *Journal of College Student Development*, 40(3), 219-234.
- Blair, R. J. R. (2008). Fine cuts of empathy and the amygdala: Dissociable deficits in psychopathy and autism. *The Quarterly Journal of Experimental Psychology*, 61, 157-170.
- Bock, G., & Goode, J. (2007). *Empathy and fairness*. Chichester, England: John Wiley and Sons.
- Bohart, A. C., & Greenberg, L. S. (Eds.). (1997). *Empathy reconsidered: New directions in psychotherapy*. Washington, DC: American Psychological Association.
- Borenstein, M., Hedges, L.V., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis*. West Sussex, England: Wiley & Sons.
- Borenstein, M., Hedges, L.V., Higgins, J., & Rothstein, H. (2005). *Comprehensive Meta-Analysis (Version 2) [Computer software]*. Englewood, NJ: Biostat.
- Bowlby, J. (1982). *Attachment and loss: Vol. 1: Attachment*. New York: Basic Books.
- Broom, E. (2002). An examination of factors related to the cognitive and affective empathy levels of adjudicated youth. *Dissertation Abstracts International Section A*, 63.
- Burke, D. M. (2001). Empathy in sexual offending and nonoffending adolescent males. *Journal of Interpersonal Violence*, 16, 222-233.
- Carkhuff, R.R. & Truax C.B. (1965). Training in counseling and psychotherapy: An evaluation of integrated didactic and experiential approach. *Journal of Consulting Psychology*, 29, 333-336.
- Carkhuff, R. P. (1969). *Helping and human relations (Vols. 1 and 2)*. New York: Holt, Rinehart & Winston.
- Carkhuff, R. P. (1980). *The art of helping (IV)*. Amherst, MA: Human Resources Development Press.
- Cohen, J. (1988). *Statistical power for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

- Cooper, H., & Hedges, L. (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.
- Cliffordson, C. (2002). Parent's judgments and students' self-judgment of empathy. *European Journal of Psychological Assessment, 17*, 36–47.
- D'Orazio, D.M (2004) Letter to the editor, *Sexual Abuse: A Journal of Research and Treatment 16* (2), pp. 173–174.
- Dapretto, M., Davies, M., Pfeifer, J., Scott, A., Sigman, M., Bookheimer, S., et al. (2006). Understanding emotions in others: Mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience, 9*(1), 28-30.
- David, B. S. (2008). Research artifacts and the social psychology of psychological Experiments. *Social and Personal Psychology Compass, 2*, 861-877.
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology, 10*, 85.
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*, 113-126.
- Davis, M. H. (1994). *Empathy: A social psychological approach*. Madison, WI: Brown & Benchmark.
- Davis, M. H., & Oathout, H. A. (1987). Maintenance of satisfaction in romantic relationship: Empathy and relational competence. *Journal of Personality and Social Psychology, 53*, 397-410.
- Del Barrio, V., Aluja, A., & Garcia, L. F. (2004). Relationship between empathy and the big five personality traits in a sample of Spanish adolescents. *Social Behavior and Personality, 32*, 677-682.
- Denzin, N. K., & Lincoln, Y.S. (Eds.). (2000). *Handbook of qualitative research*. Thousand Oaks, CA: Sage.
- Di Chiara, G. (1982). *L'assetto mentale dello psicoanalista quale invariante tra terapia e conoscenza*. Paper presented at the Fifth National Conference of the Societa Psicoanalista Italiana, Rome, Italy.
- \*Dorfman, F. D. (1993). The effects of empathy training on adolescent sex offenders' level of empathy. *Dissertation Abstracts International Section A, 65*
- Durlak, J. A., & Lipsey, M.W., (1991). A practitioner's guide to meta-analysis. *American Journal of Community Psychology, 19*, 291-332.
- Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455-463.

- Eisenberg, N., & Strayer, J. (Eds.). (1987). *Empathy and its development*. New York: Cambridge University Press.
- Eisenberg, N., & Fabes, R. A. (1990). Empathy: Conceptualization, measurement, and relation to prosocial behavior. *Motivation and Emotion, 14*, 131-149.
- Eisenberg, N., & Fabes, R. A. (1998). Prosocial development. In Eisenberg, N., & Damon, W. (Eds.), *Handbook of child psychology* (pp. 701-778). New York: Wiley.
- Eisenberg, N., & Miller, P. A. (1987). The relation of empathy to prosocial and related behaviors. *Psychological Bulletin, 101*, 91-119.
- Eisenberg, N., Fabes, R. A., Murphy, B., Karbon, M., Maszk, P., Smith, M., et al. (1994). The relations of emotionality and regulation to dispositional and situational empathy related responding. *Journal of Personality and Social Psychology, 66*, 776-797.
- Endresen, I., & Olweus, D. (2001). Self-reported empathy in Norwegian adolescents: Sex differences, age trends, and relationship to bullying. *Constructive & destructive behavior: Implications for family, school, & society* (pp. 147-165). Washington, DC: American Psychological Association.
- Epstein, W. (2004). Confirmational response bias and the quality of the editorial processes among American social work journals. *Research on Social Work Practice, 14*, 450-458.
- Erikson, E. (1963). *Childhood and society* (2nd ed.). New York: W. W. Norton.
- Eyberg, S. M., Nelson, M. M., & Boggs, S. R. (2008). Evidence-based psychosocial treatments for children and adolescents with disruptive behavior. *Journal of Clinical Child & Adolescent Psychology, 37*, 215-237.
- Fairbairn, W. R. D., (1952). *An object-relations theory of the personality*. New York: Basic Books.
- Feldman, D.H., & Goldsmith, L. (1986). *Nature's gambit*. New York: Basic Books.
- Fernandez, Y. M., Marshall, W. L., Lightbody, S., & O'Sullivan, C. (1999). The child molester empathy measure: description and examination of its reliability and validity. *Sexual Abuse: A Journal of Research and Treatment, 11*, 17-31.
- Feshbach, N., & Roe, K. (1968). Empathy in six- and seven-year-olds. *Child Development, 39*, 133-145.
- Feshback, N. D. (1987). Parental empathy and child adjustment/maladjustment. In N. Eisenberg, & J. Strayer (Eds.), *Empathy and its development* (pp. 271-291). New York: Cambridge University Press.

- Finkelhor, D., & Lewis, I. (1988). An epidemiologic approach to the study of child molestation. *Human sexual aggression: Current perspectives* (pp. 64-78). New York: New York Academy of Sciences.
- Fisher, D., Beech, A., & Browne, K. (1999). Comparison of sex offenders to nonoffenders on selected psychological measures. *International Journal of Offender Therapy and Comparative Criminology*, 43(4), 473-491.
- Foubert, J. D. (2000). The longitudinal effects of a rape-prevention program on fraternity men's attitudes. *Journal of American College Health*, 48, 158.
- \*Foubert, J. D. and Newberry J. T. (2006). Effects of two versions of an empathy-based rape prevention program on fraternity men's survivor empathy, attitudes, and behavioral intent to commit rape or sexual assault. *Journal of College Student Development* 47(2), 133-148.
- Freedberg, S. (2007). Re-examining Empathy: A relational-feminist point of view. *Social Work*, 52(3), 251-259.
- Freeman-Longo, R.E., Bird, S.L., Stevenson, W.F., & Fiske, J.A. (1994). 1994 Nationwide survey of treatment programs and models. Brandon, VT: Safer Society Press.
- Frith, C. D., & Frith, U. (1999). Interacting minds: A biological basis. *Science*, 286, 1692-1695.
- Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, 50, 531-534.
- Gallese, V. (2003). The roots of empathy: The shared manifold hypothesis and the neural basis of intersubjectivity. *Psychopathology*, 36(4), 171-180
- Garber, J., Robinson, N., & Valentiner, D. (1997). The relation between parenting and adolescent depression: Self-worth as a mediator. *Journal of Adolescent Research*, 12, 12-33.
- Gini, G., Albiero, P., Benelli, B., & Altoe, G. (2007). Does empathy predict adolescents' bullying and defending behavior? *Aggressive Behavior*, 33, 1-10.
- Gladding, S. (1978). Empathy, gender, and training as factors in the identification of normal infant cry-signals. *Perceptual and Motor Skills*, 47(1), 267-270.
- Glasser, W. (1975). *Reality therapy*. New York: Harper & Row.
- Goldstein, E. G. (2001). *Object relations theory and self-psychology in social work practice*. New York: Free Press.
- Goleman, D. (1995). *Emotional intelligence: Why it can matter more than IQ*. New York: Bantam.

- Grady, M. (2005). Effects of an empathy-focused group treatment program for incarcerated sex offenders. *Dissertation Abstracts International Section A*, 65,
- Graziano W.G., Habashi M.M., Sheese B.E., and Tobin R.M. (2007). Agreeableness, empathy, and helping: a person x situation perspective. *Journal of Personality And Social Psychology* 93(4), 583-599
- Grskovic, J. A., & Goetze, H. (2008). Short-term filial therapy with German mothers: Findings from a controlled study. *International Journal of Play Therapy*, 17, 39-51.
- Guernsey, B. G., Jr. (1988). Family relationship enhancement: A skill training approach. In L. A. Bond & B. M. Wagner (Eds.), *Families in transition: Primary prevention programs that work* (pp. 99-134). Beverly Hills, CA: Sage.
- \*Guzzetta, R. (1976). Acquisition and transfer of empathy by the parents of early adolescents through structured learning training. *Journal of Counseling Psychology*, 23(5), 449-453.
- Hartmann, H. (1958). *Ego psychology and the problem of adaptation* (David Rapaport, Trans.). New York: International Universities Press.
- Hastings, P., Zahn-Waxler, C., Robinson, J., Usher, B., & Bridges, D. (2000). The development of concern for others in children with behavior problems. *Developmental Psychology*, 36(5), 531-546.
- Hay, D. F. (1994). Prosocial development. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 35, 29-71.
- \*Haynes, L. A., & Avery, A.W. (1979). Training adolescents in self-disclosure and empathy skills. *Journal of Counselling Psychology*, 26, 526-30.
- Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.
- Henry, C. S., Sager, D. W., & Plunkett, S. W. (1996). Adolescents' perceptions of family system characteristics, parent adolescent dyadic behaviors, adolescent qualities and adolescent empathy. *Family Relations*, 45, 283-292.
- Hepworth, D.H., Rooney, R.H., Rooney, G.D., Strom-Gottfried, K., Larsen, J. (2010). *Direct social work practice: Theory and skills* (8<sup>th</sup> ed). Belmont, CA: Brooks/Cole.
- \*Heran, W. J. (2005). The effects of global empathy training on attachment styles, social competencies and empathy deficits with male adolescent sex offenders in court-ordered residential treatment. *Dissertation Abstracts International*, 66, 3463A.
- Higgins, L., Thompson, S. G., Deeks, J.J., & Altman, D.G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, 237, 557-560.

- Higgins J.P.T., Altman D.G., Chapter 8: Assessing the risk of bias in included studies. In: Higgins, J.P.T. Green S. (Editors). *Cochrane handbook for systematic reviews of interventions*. Version 5.0.1 [updated September 2008]. The Cochrane Collaboration, 2008. Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org)
- Higgins J.P.T., Green S. (Eds). *Cochrane handbook for systematic reviews of interventions* Version 5.0.2 [updated September 2009]. The Cochrane Collaboration, 2009. Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org)
- Hill, C., & King, J. (1976). Perceptions of empathy as a function of the measuring instrument. *Journal of Counseling Psychology*, 23(2), 155-157.
- Hodges, S.D., & Myers, M.W. (2007). Empathy. *Encyclopedia of Social Psychology*, 1, 296-298.
- Hoffman, M. L. (1987). The contribution of empathy to justice and moral judgment. In N. Eisenberg & J. Strayer (Eds.), *Empathy and its development* (pp. 47-80). New York: Cambridge University Press.
- Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice*. New York: Cambridge University Press.
- Hoffman, M. L. (2001). Toward a comprehensive empathy-based theory of prosocial moral development. In A. C. Bohart & D. J. Stipek (Eds.), *Constructive and destructive behavior implications for family, school and society* (pp. 61-86). Washington DC: American Psychological Association.
- Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology*, 33, 307-316.
- Hume, D. (1967). *A treatise of human nature*. Oxford, UK: Oxford University Press. (Original work published 1740).
- Iacoboni, M., & Lenzi, G. (2002). Mirror neurons, the insula, and empathy. *Behavioral & Brain Sciences*, 25(1), 39-40.
- \*Isquick, M.F. (1981). Training older people in empathy: Effects on empathy, attitudes, and self-exploration. *International Journal of Aging and Human Development*, 13, 1-14.
- Jackson, A. L., & Bonacker, N. (2006). The effect of victim impact training programs on the development of guilt, shame and empathy among offenders. *International Review of Victimology*, 13, 301-324.
- \*Janoka, C., & Scheckenbach, A. (1978). Empathy training with inmates and staff utilizing the Carkhuff model. *Corrective and Social Psychiatry and Journal of Behavior Technology, Methods and Therapy*, 24, 6-12.
- Jolliffe, D., & Farrington, D. P. (2004). Empathy and offending: A systematic review and meta-analysis. *Aggression and Violent Behavior*, 9, 441-476.

- Jolliffe, D., & Farrington, D. P. (2006a). Examining the relationship between low empathy and bullying. *Aggressive Behavior*, 32, 540-550.
- Jolliffe, D., & Farrington, D. P. (2006b). Development and validation of the Basic Empathy Scale. *Journal of Adolescence*, 29, 589-611.
- Kagan, J. (1997). Temperament and the reactions to unfamiliarity. *Child Development*, 68, 139-143.
- \*Kalliopuska, M., & Tiitinen, U. (1991). Influence of two developmental programmes on the empathy and prosociability of preschool children. *Perceptual and Motor Skills*, 72, 323-328.
- \*Kalliopuska, M., & Ruokonen, I. (1993). A study with a follow-up of the effects of music education on holistic development of empathy. *Perceptual and Motor Skills*, 76, 131-137.
- Kaminski, J., Valle, L., Filene, J., & Boyle, C. (2008). A meta-analytic review of components associated with parent training program effectiveness. *Journal of Abnormal Child Psychology: An official publication of the International Society for Research in Child and Adolescent Psychopathology*, 36(4), 567-589.
- Keller, E. F. (1983). *A feeling for the organism: The life and works of Barbara McClintock*. New York: Freeman.
- Kennett, J. (2002). Autism, empathy and moral agency. *The Philosophical Quarterly*, 52, 340-357.
- Kerlinger, F.N. & Lee, H.B. (2000). *Foundations of behavioral research*. Orlando, FL: Harcourt.
- \*Kimberlin, C., & Friesen, D. (1977). Effects of client ambivalence, trainee conceptual level, and empathy training condition on empathic responding. *Journal of Counseling Psychology*, 24(4), 354-358.
- \*Kipper, D. A., & Ben-Ely, Z. (1979). The effectiveness of the psychodramatic double method, the reflection method, and lecturing in the training of empathy. *Journal of Clinical Psychology*, 35, 370-375.
- Kohlberg, L. (1981). *Essays on moral development*. San Francisco: Harper & Row.
- Kohut, H. (1959). Introspection, empathy and psychoanalysis. *Journal of the American Psychoanalytic Association*, 7, 459-483.
- \*Kremer, J. F., & Dietzen, L. L. (1991). Two approaches to teaching accurate empathy to undergraduates: Teacher-intensive and self-directed. *Journal of College Student Development*, 32, 69-75.
- Landis, J.R., & Koch, G. G. (1977) The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.

- Larsen, J., & Hepworth, D. (1978). Skill development through competency-based education. *Journal of Education for Social Workers, 14*, 73-81.
- Lee, L. A. (1987). Rape prevention: Experiential training for men. *Journal of Counseling & Development, 66*, 100-101.
- Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- Lipsey, M., & Cullen, F. (2007). The effectiveness of correctional rehabilitation: A review of systematic reviews. *Annual Review of Law and Social Science*. Vol.3.
- Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. Oxford, England: Oxford University Press.
- Litvack-Miller, W., McDougall, D., & Romney, D. M. (1997). The structure of empathy during middle childhood and its relationship to prosocial behaviour. *Genetic, Social and General Psychology Monographs, 123*, 303-325.
- \*Lomis, M. J., and L. L. Baker (1985). Microtraining of forensic psychiatric patients for empathic counseling skills. *Journal of Counseling Psychology 32*(1), 84-93.
- \*Long, E. C. J., Angera, J. J., Carter, S. J., Nakamoto, M., & Kalso, M. (1999). Understanding the one you love: A longitudinal assessment of an empathy training program for couples in romantic relationships. *Family Relations, 48*, 235-242.
- \*Long, E. C. J., Angera, J. J., & Hakoyama, M. (2006). Using videotaped feedback during intervention with married couples: A qualitative assessment. *Family Relations, 55*, 428-438.
- Long, E. C. J., Angera, J. J., & Hakoyama, M. (2008). Transferable principles from a formative evaluation of a couples' empathy program. *Journal of Couple & Relationship Therapy, 7*, 88-112.
- Lovecky, D. V. (1993). The quest for meaning: Counseling issues with gifted children and adolescents. In L. K. Silverman (Ed.), *Counseling the gifted and talented* (pp. 29-50). Denver: Love Publishing.
- Lovecky, D.V. (1992). Exploring social and emotional aspects of giftedness. *Roeper Review, 15*, 18-25.
- Lundahl, B., & Yaffe, J. (2007). Use of meta-analysis in social work and allied disciplines. *Journal of Social Service Research, 33*, 1- 11.
- Lundahl, B.W., Kunz, C., Brownell, C., Tollefson, D., & Burke, B. L. (2010). A meta-analysis of motivational interviewing. *Research on Social Work Practice Online First*. Retrieved January 11, 2010 from <http://rswp.sagepub.com>.

- Magen, Z. (2004). Commitment beyond self and adolescence: The issue of happiness. *Social Indicators Research*, 37(3), 235–267.
- Mahler, M. (1968). *On human symbiosis and the vicissitudes of individuation*. New York: International Universities Press.
- Mahler, M. S., Pine, F., & Bergman, A. (1975). *The psychological birth of the human infant*. New York: Basic Books.
- Mahon, N. E., Yarcheski, A., & Yarcheski, T. J. (2005). Happiness as related to gender and health in early adolescents. *Clinical Nursing Research*, 14, 175-190.
- Marshall, W. L., Hudson, S. M., Jones, R., & Fernandez, Y. M. (1995). Empathy in sex offenders. *Clinical Psychology Review*, 15, 99-113.
- \*Marshall, W. L., O’Sullivan, C., & Fernandez, Y. M. (1996). The enhancement of victim empathy among incarcerated child molesters. *Legal and Criminological Psychology*, 1, 95-102.
- McFarland, S. (1998). Communism as religion. *International Journal for the Psychology of Religion*, 8(1), 33-48.
- McMahon, S., & Washburn, J. (2003). Violence prevention: An evaluation of program effects with urban african american students. *Journal of Primary Prevention*, 24(1), 43-62.
- Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality*, 40, 525-543.
- Mehrabian, A. (1996). *Manual for the Balanced Emotional Empathy Scale (BEES)*, unpublished data.
- Mehrabian, A. (1997). Relation among personality scales of aggression, violence, and empathy: Validation evidence bearing on the risk eruptive violence scale. *Aggressive Behavior*, 23, 433-445.
- Miller, P. A., & Eisenberg, N. (1988). The relation of empathy to aggressive and externalizing/antisocial behavior. *Psychological Bulletin*, 103, 324-344.
- Miller, W.R., & Rollnick, S. (2002). *Motivational interviewing: Preparing people to change addictive behavior* (2<sup>nd</sup> ed.). New York: Guilford Press.
- Miller, J. B., & Stiver, I. P. (1997). *The healing connection: How women form relationships in therapy and in life*. Boston: Beacon Press.
- Moher D., Liberati A., Tetzlaff J., Altman D.G. The PRISMA Group (2009) Preferred Reporting Items for Systematic Reviews and Meta-Analyses: *The PRISMA Statement*. PLoS Med 6(7):0 e1000097.
- Monroe, K. (2006). Can empathy be taught? *Academe*, 92(4), 58-63

- National Association of Social Workers. (approved 1996, revised 1999). *Code of ethics of the National Association of Social Workers*. Washington, DC: Author.
- Natvig, G. K., Albrektsen, G., & Qvarnstrøm, U. (2003). Associations between psychosocial factors and happiness among school adolescents. *International Journal of Nursing Practice*, 9, 166-175.
- Norris, V. (2006). Sex differences. *Encyclopedia of Human Development*, 3, 1147-1154.
- O'Donohue, W., Yeater, E. A., & Fanetti, M. (2003). Rape prevention with college males: The roles of rape myth acceptance, victim empathy, and outcome expectancies. *Journal of Interpersonal Violence*, 18, 513-531.
- \*Ogle, N. (2008). Cultivating empathy among undergraduate students through the implementation of personalization exercises. *Dissertation Abstracts International Section A*, 69.
- Olden, C. (1958). Notes on the development of empathy. *Psychoanalytic Study of the Child*, 13, 505-518.
- Owen-Anderson, A. (2006). 'I know what he is feeling because it is like i am inside of him.' Examining sensory sensitivities, empathy, and expressed emotion in boys with gender identity disorder and their mothers: A comparison to clinical control boys and community control boys and girls. *Dissertation Abstracts International*, 67.
- Owen-Anderson, A. F. H., Jenkins, J. M., Bradley, S. J., & Zucker, K. J. (2008). Empathy in boys with gender identity disorder: A comparison to externalizing clinical control boys and community control boys and girls. *Child Psychiatry & Human Development*, 39, 67-83.
- Paul, G. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting Psychology*, 31, 109-118
- \*Pecukonis, E. V. (1990). A cognitive/affective empathy training program as a function of ego development in aggressive adolescent females. *Adolescence*, 25, 59-76.
- Pett, M. A. (1997). *Nonparametric statistics for health care research: Statistics for small samples and unusual distributions*. Thousand Oaks, CA: Sage Publications
- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Oxford: Blackwell.
- Piaget, J. (1932). *The language and thought of the child, 1926; Judgment and reasoning in the child, 1928; The child's conception of the world, 1929; The child's conception of physical causality, 1930; The moral judgment of the child, 1932*. Oxford: Harcourt.

- Piechowski, M. M. (1991) Emotional Development and Emotional Giftedness. In N. Colangelo & G. Davis (Eds.), *Handbook of gifted education* (pp. 285-306). Needham Heights, MA: Allyn & Bacon
- Pithers, W. D. (1994). Process evaluation of a group therapy component designed to enhance sex offenders' empathy for sexual abuse survivors. *Behaviour Research and Therapy*, 32, 565-570.
- Pithers, W., & Gray, A. (1996). Utility of relapse prevention in treatment of sexual abusers. *Sexual Abuse: Journal of Research and Treatment*, 8(3), 223-230.
- \*Pithers, W. D. (1999). Empathy: Definition, enhancement, and relevance to the treatment of sexual abusers. *Journal of Interpersonal Violence*, 14(3), 257-284
- Poulos, S., Elison, J., & Lennon, R. (2004). The hierarchical structure of the Interpersonal Reactivity Index. *Social Behavior and Personality*, 32, 355-360.
- Prochaska, J. O., & DiClemente, C.C. (1982). Transtheoretical therapy: toward a more integrative model of change. *Psychotherapy*, 19, 276-288.
- Richmond, Mary E. (1922). *What is social case work? An introductory description*. New York: Russell Sage Foundation.
- Roberts W, & Strayer, J. (1996). Empathy, emotional expressiveness, and prosocial behavior. *Child Development*, 67, 449-470.
- Rogers, C. R. (1959). A theory of therapy, personality and interpersonal relationships, as developed in the client-centered framework. In S. Koch (Ed.), *Psychology: A study of a science: Vol. 3. Formulation of the person and the social context* (pp. 184-256). New York: McGraw-Hill.
- Root-Bernstein, R.S. (1987). Tools of thought: Designing an integrated curriculum for lifelong learners. *Roeper Review*, 10, 17-21.
- Rosenthal, R. (1979). The file drawer problems and tolerance for null results. *Psychological Bulletin*, 86, 638-641.
- Rothstein H.R., Sutton A.J., & Borenstein M. (Eds.). (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Sussex: John Wiley and Sons.
- Rubin, A. (2008). *Practitioner's guide to using research for evidence-based practice*. Hoboken, NJ: John Wiley & Sons.
- Samenow, S. E. (2002). *Straight talk about criminals*. Lanham, MD: Rowman & Littlefield.
- \*Schewe, P. A., & O'Donohue, W. (1993). Sexual abuse prevention with high-risk males: The roles of victim empathy and rape myths. *Violence and Victims*, 8, 339-351.

- Schulte-Rüther, M., Markowitsch, H. J., Shah, N. J., Fink, G. R., & Piefke, M. (2008). Gender differences in brain networks supporting empathy. *NeuroImage*, *42*, 393-409.
- Silverman, L.K. (1993). A developmental model for counseling the gifted. In L.K. Silverman (Ed.), *Counseling the gifted and talented* (pp. 51- 78). Denver, CO: Love Publishing.
- Steibe, S., Boulet, D., & Lee, D. (1979). Trainee trait empathy, age, trainer functioning, client age and training time as discriminators of successful empathy training. *Canadian Counsellor*, *14*(1), 41-46
- Stepien, K.A., & Baernstein, A. (2006). Educating for empathy: A review. *Journal of General Internal Medicine* *21*(5), 524-530.
- Strayer, J., & Roberts, W. (2004). Empathy and observed anger and aggression in 5-year-olds. *Social Development* *13*, 1-13.
- Terry, K. J., & Mitchell, E. W. (2001). Motivation and sex offender treatment efficacy: Leading a horse to water and making it drink? *International Journal of Offender Therapy and Comparative Criminology*, *45*, 663-672.
- \*Therrien, M. E. (1979). Evaluating empathy skill training for parents. *Social Work*, *24*, 417-418.
- Thornton, S., & Thornton, D. (1995). Facets of empathy. *Personality and Individual Differences*, *19*, 765-767.
- Thwaites, R., & Bennett-Levy, J. (2007). Conceptualizing empathy in cognitive behaviour therapy: Making the implicit explicit. *Behavioural and Cognitive Psychotherapy*, *35*(5), 591-612.
- Truax, C.B. & Carkhuff, R.R. (1967). *Toward effective counseling and psychotherapy: Training and Practice*. Chicago: Aldine Publishing Company.
- \*Uhlemann, M. R. & Koehn, C. V. (1989). Effects of covert and overt modeling on the communication of empathy. *Canadian Journal of Counselling*, *23*, 372-381.
- Varker, T., Devilly, G. J., Ward, T., & Beech, A. R. (2008). Empathy and adolescent sexual offenders: A review of the literature. *Aggression and Violent Behavior*, *13*, 251-260.
- Ward, T., Keenan, T., & Hudson, S. M. (2000). Understanding cognitive, affective, and intimacy deficits in sexual offenders: A developmental perspective. *Aggression and Violent Behavior*, *5*, 41-62.
- Webster-Stratton, C. & Reid, M. J. (2004). Strengthening social and emotional competence in young children--the foundation for early school readiness and success. *Infants & Young Children: An Interdisciplinary Journal of Special Care Practices*, *17*, 96-113.

- Weisburd, D., Lum, C. M., & Petrosino, A. (2001). Does research design affect study outcomes in criminal justice? *Annals of the American Academy of Political and Social Science*, 578, 50–70.
- Wiehe, V. (1997). Approaching child abuse treatment from the perspective of empathy. *Child Abuse & Neglect*, 21(12), 1191-1204.
- Wilson, S. & Lipsey, M. (2007). School-based interventions for aggressive and disruptive behavior: Update of a meta-analysis. *American Journal of Preventive Medicine*, 33(2), 130-143.
- Zaki, J., Bolger, N. & Oschner, K. (2008). It takes two: The interpersonal nature of empathic accuracy. *Psychological Science*, 19, 399-404.
- Zucker, P. J., Worthington, E. L., & Forsyth, D. R. (1985). Increasing empathy through participation in structured groups: Some attributional evidence. *Human Relations*, 38, 247-255.

Changes in Dispositional Empathy in American College Students Over Time: A Meta-Analysis. Sara H. Konrath. 1,2.Â The IRI is an ideal measure of empathy to use for a cross-temporal meta-analysis. One major strength of the scale is that it assesses both cognitive and affective components of empathy, which could theoretically be changing at different rates over time. In addition, the IRI is reliable, well validated, and widely used.