CONVERGENT VALIDITY AND UTILITY OF EMPATHIC ACCURACY MEASURES ACROSS THERAPISTS AND NON-THERAPISTS

By

Lia Stern

Submitted to the

Faculty of the College of Arts and Sciences

of American University

in Partial Fulfillment of

the Requirements for the Degree of

Master of Arts

In

Psychology

Chair:

David A.F. Haaga, Ph.D.

Anthony H. Ahrens, Ph.D.

Nathaniel R. Herr, Ph.D.

Dean of the College of Arts and Sciences

April 29, 2014

Date

2014

American University

Washington, D.C. 20016

CONVERGENT VALIDITY AND UTILITY OF

EMPATHIC ACCURACY MEASURES ACROSS

THERAPISTS AND NON-THERAPISTS

BY

Lia Stern

ABSTRACT

Empathic accuracy (EA) is the therapist variable most frequently associated with positive client outcomes, and is a skill that can improve with training (Barone et al., 2005; Forrester, Kershaw, Moss, & Hughes, 2008). EA is heterogeneous in nature, comprising both affective and cognitive components, yet very little research has compared the ability of various assessments to accurately reflect these nuances. Theoretically, therapists should be more accurate on self-report measures of EA than non-therapists, and the convergent validity of these subjective measures should be higher with objective measures as a result. This study compared the scores of 36 therapists (clinical psychology students) and 36 "non-therapists" (age and gender matched controls) on self-report measures of EA as well as a video-based performance task in which participants' heart rates were measured as they attempted to infer the thoughts and feelings of a videotaped woman while she narrated a prior traumatic experience. It was hypothesized that therapist scores would show stronger convergent validity than non-therapist controls, and that this would strengthen with clinical training. Analyses revealed predominantly small and non-significant correlations across all indices-pairings in the sample as a whole. However, correlations of self-report and physiological assessments of affective EA, as well as global and study-specific measures of cognitive EA were significantly stronger among therapists than non-therapists. These results indicate that EA measures may reflect higher convergent validity among therapists than non-therapist controls.

TABLE OF CONTENTS

LIST OF TABLES

Table

CHAPTER 1

INTRODUCTION

All else being equal, empathic individuals are likely to be the most insightful therapists, and those most successful at developing a healthy client-therapist relationship (Gesn & Ickes, 1999). Across studies, practitioner empathy is the therapist variable most frequently and substantially associated with positive client outcome, even when compared to other possible contributors (Forrester et al., 2008; Marangoni, Garcia, Ickes, & Teng, 1995). A 2005 review of 52 child psychotherapy treatment studies identified that therapist empathy, attention, and positive regard were all essential to progress in therapy settings (Jensen, Weersing, Hoagwood, & Goldman, 2005), corroborating Gurman's pivotal 1977 study. Gurman's research found that, on average, empathy correlated .62 with congruence, .53 with positive regard, and .28 with unconditionality, suggesting that while some construct overlap exists, empathy still accounts for unique and meaningful contributions to clinical progress above and beyond the other variables (Gurman, 1977). Further, practitioners who are highly empathic are more likely to accurately attune to their clients' needs, correctly interpret the level of expressed empathy they will find most useful, and subsequently develop a strong dyadic relationship through their tactful expression of this understanding (Elliott, Bohart, Watson, & Greenberg, 2011).

Impressively, Elliott and colleagues (2011) found that even beyond a therapist's ability to demonstrate empathy, a client's *perception* of therapeutic empathy accounts for approximately 9% of variance in clinical outcomes. Interrupting, poor eye contact, and dismissal of client input in favor of promoting practitioner goals were all characteristics that clients associated with low-empathy therapists (Myers, 2000). Alternately, attentiveness, open discussion across a range of topics, and attention to detail were perceived as empathic in client reports (Myers, 2000). These findings indicate that while clients may differ in the degree to which they desire expressed

empathy from therapists, some generalizable benchmarks do exist. Regardless of the exact manner in which a therapist expresses empathy, it is likely that the integral component of responsiveness contributes to a client's experience of commitment and interest, which provides a foundation for feeling understood and validated in a therapeutic dyad (Seehausen, Kazzer, Bajbouj, & Prehn, 2012).

Empathic insights are especially interesting because they are a cornerstone of daily interactions, and as such, are often overlooked. The exception to this is when they lapse in noticeably dramatic or consequential ways. Such empathic disintegration could manifest itself in a therapist's inability to accurately interpret client emotions, or to temper their own responses to clients. Such a failure would be especially detrimental in a clinical setting where such a mistake could distinguish between a close therapeutic relationship and an ostracized client (Seehausen et al., 2012).

Defining Empathy

In the literature, empathic skill is objectively measured under the title Empathic Accuracy (EA). The precise definition of EA varies across studies, and is often dependent on the theoretical framework employed by the researchers. Consequentially, while many lay people have conceptions of empathy phenotypes, researchers have yet to identify a consolidated interpretation for scientific inquiry. However, most conceptualizations parallel a well-known and respected definition supplied by William Ickes, a pioneer in empathy research and assessment methodology. According to Gesn & Ickes (1999), EA consists of multiple stages in which an observer develops an empathic understanding of a target experience, experiences physiological or affective empathic arousal, and subsequently expresses this understanding (Gesn & Ickes, 1999). The level of skill with which an individual can accurately interpret, experience, and

communicate incoming emotional stimuli is then operationalized as Empathic Accuracy. For example, rather than simply recognizing that a client is feeling “anger,” a therapist high in empathic accuracy should accurately identify the nuances of an emotion (e.g. recognizing the emotion’s intensity, origin, and triggers), link these with other relevant emotional experiences, and subsequently express the amalgamation to their client (Gerdes & Segal, 2011). This communicative practice is likely a cornerstone of the effective dyadic relationship necessary for positive clinical outcomes and client growth that therapists strive for.

Ickes’ definition of EA is frequently understood and researched as having both affective and cognitive components. Davis et al. (2004) described the affective component as the capacity to be affected (emotionally or physiologically) by someone else’s feelings, whereas cognitive empathy required the subsequent expression of those identified feelings into an interpretable and usable narrative (Davis et al., 2004). Davis explained the distinction as an individual’s ability to experience vicarious emotional or physiological arousal as a result of someone’s expressed emotion (affective empathy) versus adopting someone else’s psychological viewpoint (cognitive empathy) (Davis, 1994). While variation exists in the operationalization of empathic accuracy and its related sub-components, it is widely accepted that empathy is a larger umbrella term comprised of both affective and cognitive substrates.

Neurophysiological research has reinforced the distinction between cognitive and affective components, identifying an array of processes and brain structures involved in both the experience and expression of empathy. According to a 2009 study by Decety and Lamm, “Emotional stimulation” is the first of three primary neurological processes involved in empathy. This process entails physiological arousal on the part of the listener, and is caused by memories of similarly arousing experiences to those the orator is speaking about. This process is largely

housed in the limbic system as it is integrated in both emotion and memory systems (Decety & Lamm, 2009). The second element is a "perspective-taking process" which originates in the prefrontal and temporal cortex. Perspectival ability encompasses the capacity to cognitively and metacognitively interpret the experience of shared emotionality involved in the "emotional stimulation" process (Shamay-Tsoory, 2009). Finally, an "emotional-regulation" capacity permits individuals to self-soothe in response to the stimuli, process the emotional experience and subsequently help the discloser do the same. This final process likely originates in the prefrontal and orbitofrontal cortex (Decety & Lamm, 2009). Clearly, while some differences exist in definitions of EA both within and across social psychology, clinical psychology, and behavioral neuroscience, some consistencies exist as well. The resulting consensus is that empathy consists of both cognitive and affective components, which produce the experience, interpretation, and expression of a dyadic partner's emotional experience.

Despite the aforementioned studies concerning neuropsychological origins of EA, there is very little research pertaining to trait empathy in the general population, or clinical populations specifically. However, empathic ability is often considered a relatively stable personality trait (Leiberg & Anders, 2006) with women typically exhibiting higher levels than men (Baron-Cohen & Wheelwright, 2004). Further, recent research suggests that empathy is largely responsive to training, and recent studies exploring EA trainability have risen dramatically as a result (Barone et al., 2005). However, as mentioned previously, a noticeable problem with EA research is that very little consensus exists in operational definitions across studies. Consequentially, "empathy" measurements vary widely, making it challenging to draw meaningful comparisons or conclusions across studies (Cliffordson, 2001).

Quantifying Empathy: Self-report Measures

Research of EA has predominantly centered around self-report measures due to their expediency and the ease with which they allow for aggregation of data across a wide-range of subjects (Gesn & Ickes, 1999). However, over the past two decades researchers have debated the accuracy of self-report measures as a general assessment tool, and as they relate to EA specifically. Globally, this resistance stems from concerns that social desirability biases or motivational issues may reduce the predictive validity of participant responses (Blair, 2001). This is especially true in assessment scenarios that allot excessive time for participants to engage in introspection or internal debate, and may lead to altered responses based on perceived levels of social desirability in the study materials (Hoffman et. al., 2005). Several studies now compare implicit measures (which aim to circumnavigate such biases) and explicit measures, in hopes of exploring this challenge further. The underlying theory of these studies is that weak correlations between explicit and implicit measurements would expose motivational and cognitive factors that impact explicit, but not implicit assessment techniques. However, a meta-analytic study by Hoffman & colleagues found that correlations between implicit and explicit measures function independent of social desirability or introspection biases for self-reported representations (Hoffman et al., 2005). It is worth noting however, that this study was conducted at a meta-analytic level, and differences in participant biases may be identified in closer levels of analysis. Despite these concerns, self-report measures remain the most frequently employed assessment tool in empathy research, and a common denominator across many psychological studies.

The literature that exists on EA assessment has focused, largely, on the impact of the multifaceted and nebulous construct definition on the development of assessment tools. The variation in specific focal points across empathy measures (cognitive vs. affective, interpretive vs. expressive, etc.) has lead to low levels of reliability and validity within this pool of research,

and produced data that are hard to compare at a meta-analytic level (Levenson & Ruef, 1992). Further, the literature suggests that individuals are typically poor predictors of performance ability in relation to their empathic capacity, which would necessarily be reflected in their self-report scores (Marangoni et al., 1995). Due to their ease of administration many researchers argue that self-report measures still serve a practically valid purpose and should be used in EA studies regardless of conflicting or disparate operational definitions across studies (Lawrence, Shaw, Baker, Baron-Cohen, & David, 2004).

The most frequently used self-report measures of empathy are the Hogan Empathy Scale (HES; Hogan, 1969), Empathy Quotient (EQ; Baron-Cohen, Richler, Bisarya, Gurunathan, & Wheelwright, 2003), Epstein Feelings Inventory (EFI; Mehrabian & Epstein, 1972) and the Interpersonal Reactivity Index (IRI; Davis, 1983). These measures all reflect unique facets of empathy, with only the IRI recognizing both cognitive and affective components. For example, Hogan (1969) failed to differentiate between cognitive and affective empathy, instead conceptualizing empathy as a largely cognitive structure with emphasis on the "imaginative apprehension of another's condition or state of mind" (p. 307; Hogan, 1969). The EQ similarly ignores such a distinction, focusing instead on the motivation of individuals to identify a dyadic partner's emotions and respond appropriately. Both the HES and EQ provide valuable information on cognitive contributors to empathy, but fail to include the affective components that both neuroscience and social psychology have identified as pivotal contributors to empathy, subsequently reducing the content validity of these tools.

The EFI was developed in response to the dearth of literature—and assessment tools—available for targeted analysis of vicarious emotional arousal and affective response (Chlopan, Mccain, Carbonell, & Hagen, 1985). The test has seven intercorrelated subscales, entitled

Susceptibility to Emotional Contagion, Appreciation of the Feelings of Unfamiliar and Distant Others, Extreme Emotional Responsiveness, Tendency to be Moved by Others' Positive Emotional Experiences, Tendency to be Moved by Others' Emotional Negative Experiences, Sympathetic Tendency, and, *Willingness to be in Contact with Others Who Have Problems.* In developing these subscales and their composite items, Mehrabian and Epstein (1972) established several requirements: first, the items should produce an insignificant correlation with social desirability measures (Crowne & Marlowe, 1960). Second, items must correlate with the overall test score at a significance level of .01. Finally, each item's content validity was determined through factor analysis of the total pool of items. Interestingly, convergent validity studies have linked the EFI with arousability, such that high ratings of empathy correlate with higher levels of general arousal (Chlopan et al., 1985). The EFI was positively correlated with the Eysenck Personality Questionnaire subscale of neuroticism (.35) (Eysenck & Eysenck, 1978), as well as the Social Interest Scale (.40) (Crandall & Harris, 1976). These mid-level correlations suggest that the EFI does tap some level of generalized arousability, but also measures an additional construct, which is, presumably, empathic direction. It should be noted that the Epstein Feelings Inventory was previously entitled "The Questionnaire Measure of Emotional Empathy" and related literature may refer to the former or present label interchangeably (Mehrabian & Epstein, 1972).

One way to overcome affective or cognitive psychometric slants would be to administer two self-reports measures, one for each component, or a hybrid test encompassing both. The IRI was crafted to do the latter. This often-utilized measure of empathic accuracy attempts to capture differences between the affective emotional responses and cognitive perspective taking capacity. Of the four subscales in this measure, two contribute to affective components (Personal Distress

and Empathic Concern), while two contribute to cognitive components (Perspective Taking and Fantasy subscales; Davis, 1980). For this reason, the IRI is viewed as having high content validity as a self-report assessment of empathy. In support of this, confirmatory factor analysis of the IRI has found Empathic Concern at the apex, but that this dimension had substantial overlap with Perspective Taking and Fantasy factors (Spreng et al., 2009). Empathic Concern embodies the affective component of empathy, whereas Perspective Taking and Fantasy account for cognitive components. As a result, the IRI factor composition reflects the associated literature in which affective and cognitive components are both integral components of empathic expression (Spreng et al., 2009). Importantly, the IRI has deficiencies as well. For instance, The Personal Distress Scale is often omitted as it has been posited to more accurately reflect sympathy than empathy (Reniers, et al., 2013). Clearly, even within the domain of self-report measures, EA assessment is a highly individuated and heterogeneous field.

Psychometric studies of empathy typically explore the validity and reliability of indices by comparing scores across existing measures. For example, in constructing the Toronto Empathy Questionnaire, Spreng, McKinnon, Mar, and Levine “attempted to formulate a consensus among the many scales in use to gauge the empathy construct” (p. 63), forcing items to load onto one single factor, and attempting to unify nine pre-existing measures in order to create one unidimensional construct. In addition to the 95 questions that were selected from various surveys, the authors constructed two additional items based on relevant research findings. Scores on this newly constructed measure were then contrasted with the IRI to assess convergent validity and the 50-item Autism Quotient as an indicator of discriminant validity. This process completed by Spreng and colleagues is fairly standard for both aggregating foundational questions, as well as assessing a measure’s psychometric uniqueness and

importance. However, it is rare that researchers include external criterion variable analysis in the construction of their measures, which would provide a valuable measurement of application above and beyond reliance on previous definitions or preexisting measures (Roberts, Solomon & Langenau, 2011).

Quantifying Empathy: Performance-based Measures

Ickes (1993) proposed that the most effective way to assess EA is by "compar[ing] the content of a target person's actual thoughts with the content of the corresponding inferred thoughts and feelings reported by the perceiver" (p. 591). This process requires a dyadic interaction either with the assistance of video recordings or in-person exchanges where one or more people produce emotional stimuli and additional participants observe the recordings. There are currently two research designs that actualize this process (Ickes, Gesn, & Graham, 2000). An *unstructured dyadic interaction paradigm* involves a recorded interaction between two research participants, who are unaware that they are being filmed. Participants subsequently review the tape, and try to make inferences about what their research partner was experiencing throughout the conversation (Ickes & Tooke, 1998). The second framework is a *standard stimulus paradigm* and operates similarly to the unstructured type, but involves recording one standard set of interviews that serve as the stimulus across all study participants. Participants must state what they think the target of the recording was thinking or feeling at designated stop points throughout the film rather than their former dyadic partner (Gesn & Ickes, 1999). The participants' responses are then compared to what the recorded target previously reported experiencing at those times. Responses can either be multiple-choice or open-ended in nature and are compared to the explicitly reported emotional experience of the target in the film.

The inherent comparison between expressed and interpreted emotionality (rather than simply recording one or the other) provides some objective measure of the observer's correctness. It is this "objective" standard that has the potential to increase the validity of performance-based methods above and beyond that of subjective measures like self-report, for which there are no comparable standards of accuracy (Wilson & Griswold, 1985). Additional support for this process was demonstrated in Maragoni et al.'s (1995) study, which found that participants' EA assignment was consistent across three separate target videos. This cross-situational consistency implies that EA is indeed a quantifiable and relatively stable construct, and that measurements produced using this tool are largely reliable and replicable.

Quantifying Empathy: Physiological Measures

A final method for assessing EA is the use of physiological measurements and arousal monitoring. Physiological arousal in response to emotionally stimulating material has been measured for almost a century (Buck, Savin, Miller, & Caul, 1972). Researchers across a range of studies have found that emotionally charged stimuli elicit changes in heart rate, skin conductance, respiration patterns and blood pressure. Because of the relative ease of data collection and the inherent objectivity of physiological assessments (Shortt & Pennebaker, 1992), these measures have become particularly appealing to researchers, especially when compared to self-report or performance based measurements in isolation (Marangoni et al., 1995). Across recent research studies, it appears the emerging 'gold standard' in empathy assessment includes physiological measures taken alongside performance-based assessments. In doing so, researchers are more likely to tap all three sub-components of empathy, including the interpretive, expressive and physiological sub-sets at both cognitive and affective levels.

At the broadest level, relevant research has shown that people high in empathic accuracy tend to become physically and emotionally aroused when exposed to emotionally charged narrative stimuli (Mehrabian & Epstein, 1972). A noteworthy study by Shortt and Pennebaker (1992) demonstrated varied physiological arousal among people either high or low in trait empathy when exposed to arousing narratives of trauma survivors. The 1992 study measured physiological fluctuations of participants watching recordings of Holocaust survivors as they recounted several traumatizing experiences. Researchers subsequently compared those fluctuations to participants' self-reported levels of trait empathy (Shortt & Pennebaker, 1992). The data revealed that skin conductance levels, heart rate, and blood pressure were all indicators of EA and fluctuated with self-reported empathy levels. High levels of EA were reflected in stronger physiological responses, whereas low levels of EA were associated with mild (if any) physiological arousal. A second, and intriguing finding of the study was that participants who identified themselves as high in EA had physiological responses more similar to those of the disclosing participants (Shortt & Pennebaker, 1992). The researchers suggested that this may be due to the "empathy hypothesis" whereby observing participants ostensibly feel what the discloser feels, eliciting the physiological arousal necessary for EA.

A study conducted by Levenson and Ruef (1992) found similar results in their standard stimulus paradigm which utilized staged marital conflict video recordings. Again, heart rate and skin conductance levels fluctuated directly with observer EA levels (Levenson & Ruef, 1992). Consistent with Shortt and Pennebaker (1992), this study provides support for the argument that individuals high in EA may actually 'feel' what their dyadic partner is expressing, at both a psychological and physiological level. Another pertinent finding of the study was that negative stimuli encouraged much stronger physiological fluctuations than positive stimuli. The

researchers hypothesized that such differentiation may be the result of higher levels of autonomic arousal elicited by negative rather than positive emotions. A noticeable benefit of physiological measurement is the ability to tap automatic or involuntary emotional arousal (Vellante, et al., 2013). However, this "benefit" necessarily assumes that emotional arousal is an accurate indicator of empathic response. Rather than serving as indices of empathic accuracy, these measures could simply assess arousability, or the sensitivity with which an individual reacts (or controls their reactions) to incoming stimuli of an emotionally charged nature.

Quantifying Empathy: Cross-Modal Analysis

To date, research has largely assumed that self-report, performance-based and physiological assessments tap the same underlying construct. However, due to the variegated nature of empathy definitions, this may not be the case. An example of this is The Eyes Test, a performance-based measure which is used to assess a participants' ability to interpret emotional expression based on "expressive" still-frame pictures of eyes. The Eyes test correlates with self-report measures of empathy and is discriminant with measures of autism (Vellante et al., 2013). However, several studies failed to find such a correlation with self-report measures of empathy and the performance task. The Eyes Test scores was not linked with IRI scores in a study by Muller and colleagues (2010), and only correlated modestly with the social skills subsection of the EQ (Lawrence et al., 2004). This inconsistency leads to questions if another factor, for example, IQ may be a more direct contributor to these performance-based scores. There is a paucity of literature contrasting the various types of empathy assessment. However, it is only through the exploration of the various strengths and weaknesses inherent to each style that researchers and clinicians will be able to determine which measures are most appropriate in given settings, and which components of empathy they most accurately tap.

A 2011 study by Roberts, Solomon and Langenau explored cross-method analysis in the medical realm, testing the construct validity of "humanistic clinical skills" (largely interpretive and expressive components of empathy), as measured in medical licensure performance examinations. One of the first studies of its kind in the field, this research explored the construct validity of six humanistic skills measured by the Global Patient Assessment Tool, including eliciting information, listening skills, respectfulness, empathy, giving information and professionalism. These skills were assessed using both a standardized exam format as well as patient care paradigm, in which their "humanistic clinical skills" were subjectively rated and subsequently analyzed. Their results showed strong convergent validity across all six categories and both measurement types, with validities ranging between .48-.62, and were statistically different from zero (Roberts, Solomon, & Langenau, 2011). Additionally, they found a moderate correlation across assessment types, indicating that although some variation did exist, both the written and performance-based measures of licensure assessment provided similar results.

Studies similar to the one conducted by Roberts and colleagues (2011) are important in that they provide evidence that measurement style does not confound scoring, but provides multiple valid methods of assessment. To date, no research has analyzed the convergent validity across empathy measurement types in order to facilitate an understanding of self-report, performance-based or physiological tools and their respective uses. Vast differentiation in empathic identification across measurement tools would indicate either temperamental indices (i.e. physiological measurement may measure overall arousal as opposed to empathic expression specifically), or potentially the assessment of differential root constructs.

Present Study

As a result of the aforementioned research, the present study utilizes three assessment styles (self-report, performance-based, and physiological) in an attempt to access both the cognitive and affective components of empathy, as well as to further understand the nuances intrinsic to each measurement modality. Additionally, we asked participants to rate how accurately they believed they completed performance-based measures, as we were interested to see if those with more experience attempting to convey empathic accuracy would demonstrate a closer connection between self-perceived and performance-based assessments of EA. This could potentially offer a barometer as to how accurate therapists felt they were compared to non-therapists in the assessment of their strengths or shortcomings with EA. Although self-report measures have been described—at best—as "fair," research has also shown that EA can improve with training (Barone et al., 2005), and is consequentially a primary aim of clinical programs. Accuracy of self-report measures among therapists may improve accordingly then, at a rate paralleled to the heightened metacognitive abilities reinforced throughout doctoral training. Although some literature has suggested that self-assessed EA ability tends to be inaccurate (Marangoni et al., 1995), therapists, regardless of their level of training, may be better judges of their own empathic abilities than non-therapists, as emotional attunement and adjustment are pivotal components of their occupation.

Hypotheses

Given the literature, we hypothesized that therapists will be more accurate than non-therapists on self-report measures of EA. Further, this accuracy is hypothesized to increase with clinical training, due to the metacognitive abilities reinforced during those programs, and this accuracy should translate across measurement modalities. In support of this we expect: 1)

Convergent validity of performance-based and self-report measures of empathy should be moderate—across the entire sample—as both primarily target cognitive empathy. 1a) Convergent validity of physiological measures will be low with both self-report and performance-based measures—among the entire sample—as the first primarily taps affective empathy while the latter measures cognitive empathy. 2) Correlations between self-report, performance-based, and physiological measures should be stronger among therapists than non-therapists. 3) Correlations among self-report measure of EA, performance-based measures and physiological measures should be stronger among seasoned therapists than novice therapists. The literature suggests that subscales of the IRI may reflect different components of EA (Spreng et al., 2009). As such, exploratory analyses were conducted to examine the degree to which subscales of the IRI impacted correlations between that measure and all other indices across conditions.

CHAPTER 2

METHOD

Phase 1 – Stimulus Development

One initial participant was recruited from the local community via Craigslist advertisement to create the video recording used in the standard stimulus paradigm. The advertisement explained the purpose of the study, the procedures and the specific requirements necessary for participation. After phone screenings, the selected target was a 35-year-old Caucasian female who consented to discuss a real-life traumatic event on camera, acknowledging that the resulting product would be shown to subsequent participants. In previous studies utilizing this standard stimulus paradigm, female targets have been chosen over males due to the belief that they may "disclose personally meaningful, intimate concerns, and do so in a more expressive fashion" than their male counterparts (Marangoni et al., p.858). This alacrity is essential to performance-based measures of EA as the target's expressed emotion is intended to elicit similar responses from observing participants (Maragoni et al, 1995). Further, negative valence stimuli have been shown to produce higher levels of arousal when compared to positive valence stimuli, and as such, is common content across standard stimulus paradigm interviews (Maragoni et al, 1995). The target's heart rate was measured at 60-second intervals throughout the recording using a finger-pulse oximeter; a total of 19 measurements were taken.

The recording spans 20 minutes, and focused on the face of the target while she spoke. At the onset of the tape, the target was verbally prompted by the researcher with the following directions: "For about 30 minutes I'd like you to talk about a traumatizing event that you've experienced in your life. This won't be like a therapy session, so I won't be responding to you, but I'll be recording your physiological measurements as you talk. Talk as openly as you can as if you were thinking it through, or writing it down. Start from the beginning, and include as

many details as you can remember." The target chose to speak about two separate traumatic experiences: the first was a scenario from age 11 in which she was sexually molested by a group of adolescent boys. She discussed the incident itself as well as the subsequent impact it had on her adolescence and adulthood. The second experience the target chose to discuss was the sudden death of her father, which occurred later in adulthood.

After filming, the target reviewed the recording and was asked to pause the clip at moments during which she recalled experiencing particularly poignant thoughts or feelings, the contents of which she detailed on a provided response sheet. In order to deter reporting bias or the overlay of new thoughts and feelings on her initial experience, the first several minutes of the film were treated as an "acclamation period" and included no emotional content. This time period was included so that the participant could become acclimated to viewing herself on camera, allowing her to "re-experience" the content of the recording. The participant was given the freedom to delete any part of the recording she felt was too intimate for empirical purposes, but chose not to do so. After the participant consented and recorded the video, she was compensated $100 for her involvement, and given the option of requesting referral to a local outpatient clinic, although she did not request such information.

Phase 2 – Participant Recruitment

The 72 additional participants involved in this study were divided into therapist (clinical doctoral candidates) and non-therapist (age and gender-matched controls) conditions, with 36 participants per category. "Therapist" participants were recruited from Clinical Psychology Doctoral programs and divided into beginning (first year) and advanced (fourth or fifth year) categories. "Non-therapists" were recruited through Craigslist advertisements and included only if they had no history of psychological training or employment. Non-therapist participants were

age-matched within five years of their clinical counterparts. There were 18 participants in each of the resulting four conditions.

Participant Demographics

The total sample included 72 participants with 36 individuals in each of the therapist and non-therapist groups. Therapist participants were equally divided between beginning and advanced trainees, with 18 participants per condition. The demographic survey provided information on gender, age, and race across all participants (see Table 1 for details). The sample was 71% female. The average age of participants was 27.22 (SD=4.98) ranging from 22 to 50 years old. The average age of the therapist group was 27.53 (SD=4.72) and ranged from 22-49. Beginning therapists averaged 25.89 years of age (SD=3.31) and ranged from 22 to 35 years, while advanced therapists averaged 29.17 years of age (SD=5.40) and ranged from 24 to 49 years. Non-therapists, who were age-matched to therapists averaged 26.92 years of age (SD=5.23) and ranged from 22 to 50 years old.

The sample was largely Caucasian (71 percent), with 21 percent of participants identifying as African-American, 4 percent as Asian/Pacific Islander, 3 percent as mixed race, and one participant as "other." Three percent of participants identified as Hispanic/Latino. Unlike gender and age, race was not matched across therapist group, and a statistically significant difference emerged in the representation of African-American participants across therapist and non-therapist conditions $\chi(1) = 5.389$, $p = 0.020$.The non-therapist condition included 11 fewer Caucasian participants and 9 additional African American participants compared to the therapist condition.

There was a range of trauma-related experience among participants in the therapist condition. Ten participants reported working with trauma-focused therapy, 8 reported completing

trauma-related intakes or assessments, and 4 participants had experience working with crisis intervention hotlines. Participants in the advanced therapist group reported seeing an average of 33.17 clients with individual reports ranging from 8-100.

Table 1. Participant Demographic Descriptive Statistics

	Therapists (n=36)	Nontherapists (n=36)	Beginning therapists (n=18)	Advanced therapists (n=18)
Gender				
Male	11(31%)	10 (28%)	4 (22%)	7 (39%)
Female	25 (69%)	26 (72%)	14 (78%)	11 (61%)
Age				
Range	22-49	22-50	22-35	24-49
Mean(SD)	27.53 (4.72)	26.92 (5.28)	25.89 (3.31)	29.17 (5.40)
Race				
African-American	3 (8%)	12 (33%)	2 (11%)	1 (6%)
White	31 (86%)	20 (56%)	14 (78%)	17 (94%)
Asian/Pacific Islander	1 (3%)	2 (6%)	1 (6%)	0
Other	0	1 (3%)	0	0
Mixed Race	1 (3%)	1 (3%)	1 (6%)	0
Hispanic/Latino?				
Yes	1 (3%)	1 (3%)	1 (6%)	0
No	35 (97%)	35 (97%)	17 (94%)	18 (100%)

Experimental Procedure

Experimenters met with participants at scheduled times and escorted them to a private office at American University. Participants were given informed consent documentation, which acknowledged the potentially upsetting content encapsulated in the study design. After consenting to participate, participants were administered a demographic questionnaire, self-

report measure of empathy (The Epstein Feelings Inventory; Mehrabian & Epstein, 1972) and a state-anxiety measure. Participants were then given the following standardized instructions for the paradigm procedure, “While you are watching the following videotape, try your best to put yourself in the shoes of the person on the tape. The videotape will be paused at several points throughout the tape. When there is a pause, please write on the form below the time you see displayed on the tape, and what you think the target person was thinking or feeling at this point. You will then be shown four multiple-choice options of possible response options that the target person actually wrote as the thought and/or feeling that she recalled while talking during the taping. Please select which option you think that she wrote.” The tape was paused at eight predetermined times identified by the target, and participants were asked to write in an open-ended format the content of what they believed she was thinking or feeling at that point in her narrative. The experimenter transcribed the provided answer of the target participant into a multiple-choice question, with one option summarizing the target’s reported experience, and three distractors providing various incorrect choices.

After finishing the standard stimulus paradigm, participants were given a self-report measure of state-anxiety, a questionnaire measuring self-assessed EA on the paradigm task, debriefed, and compensated for their time. The experimental procedure lasted approximately one hour, and all participants were compensated $30 in cash. Participants were offered referrals to a local outpatient clinic to address anxiety or discomfort that resulted from their involvement; no participants requested such a referral.

MEASURES

Self-Report Measures

Demographic Survey

Two versions of the demographic survey were crafted, one for each experimental category of the study. Both therapists and non-therapists were asked to indicate their age, gender, race/ethnicity and marital status. Non-therapists were additionally asked to describe what, if any, previous experience they had with the mental-health field.

Therapists were asked to provide information about their doctoral training, including theoretical orientation. Therapists were also asked to provide details about their clinical experience, including any history of involvement with traumatized patients, the number of clients seen through their doctoral program, and any additional clinical experience they may have had either prior to, or concurrent with enrollment in their program.

Empathy

Self-reported empathy was first measured using the Epstein Feelings Inventory (EFI: Mehrabian & Epstein, 1972). Participants rated each of 33 items on a self-reflective scale from +4 (very strong agreement) to -4 (very strong disagreement). To calculate an overall empathy score, reverse-keyed items were changed from negative to positive and all 33 responses were summed. Epstein and Mehrabian reported, females, M=44, SD=21; males, M=23, SD=22. All items were used in the present study. There are seven subscales: *susceptibility to emotional contagion; appreciation of the feelings of unfamiliar and distant others; extreme emotional responsiveness; tendency to be moved by others' positive and emotional experiences; tendency to be moved by others' negative emotional experiences; sympathetic tendency;* and *willingness to be in contact with others who have problems*, all of which were found significant at the .01 level (Mehrabian & Epstein, 1972). The EFI has also exhibited high reliability (split-half reliability of

the measure is .84) and discriminant validity (Mehrabian & Epstein, 1972). The current sample also showed high internal consistency as well (α=.77).

Participants' perceptions of their own Empathic Accuracy abilities were subsequently measured using a cross-section of items from the Interpersonal Reactivity Index (IRI; Davis, 1980). Davis (1980) reported retest reliability from .61-.81 in a 60-75 day time span, as well as internal consistency from .68-.79 for the subscales. This frequently used measure of dispositional empathy has 28-items, 10 of which were utilized in the current study to approximate self-perceived accuracy on the performance-based task. This abridged version was created in order to obtain representative data without significantly increasing the duration of the study. The full measure is considered a balanced assessment of EA, with a slight cognitive leaning. However, the modifications that were made for this study—specifically which questions were included or omitted—made the cognitive orientation somewhat more pronounced. Questions were selected that could best reflect the video paradigm. For example, the original question, "When I see someone who badly needs help in an emergency, I go to pieces" was altered to "While watching the person in the tape, I felt like I could 'go to pieces'" (see Appendix D for full list of questions). Participants were asked to rate statements from 0 "does not describe me well" to 4 "describes me very well." Responses were summed to create a composite IRI score.

Two study-specific items were added to assess participants' self-perceived accuracy in assessing the target's responses, as well as how their own emotions did (or did not) fluctuate in accordance with the stimulus participant. Responses to these items ranged from 1 "not at all accurate" to 10 "completely accurate."

Performance-Based Measures of Empathy

Performance-based scores were computed using open-ended and multiple-choice responses to the standard paradigm stimulus tape. In order to equally assess the open-ended and multiple-choice response components, raw scores were converted to z-scores and summed to create an aggregate "performance score."

Open-Ended Response

Participants' open-ended responses to the video stimulus were coded for similarity of overall meaning to the target's written response. There was no required length for these responses, and they varied significantly across participants as a result. Two coders who were masked to participant condition assigned values of 0 (not at all similar), 1 (somewhat similar) or 2 (highly similar) to participant responses. Both ratings were averaged within and across items to produce an accuracy score for participants with possible scores spanning 0 (no correct responses) to 16 (both raters agree that all answers are correct); interrater reliability was high (Kappa = .882).

Multiple-Choice Response

The number of correct multiple-choice answers for each participant was summed to create a multiple-choice response score. The index ranged from 0 (no correct responses) to 8 (all correct responses).

Physiological Measure of Empathy

Heart Rate

While watching the stimulus video, a fingertip oximeter was used to measure participants' heart rates in a manner consistent with prior studies of physiological demarcations of EA (Shortt & Pennebaker, 1992). A total of 21 measurements were taken at one-minute

intervals mirroring data collection times of the target participant. The first two measurement points were omitted from subsequent analyses in order to reduce the likelihood that anticipatory anxiety or unfamiliarity would impact either the heart rate of the target participant or any subsequent participants. It is important to note that there were eight points throughout the paradigm task during which participants were interrupted from watching the stimulus in order to think about—and report on—what they believed the target person was thinking or feeling. These interruptions caused participants to shift their attention from the target participant to their own interpretations of her experience, and to the study procedures themselves. It is likely that these interruptions led to alterations in participants' physiological arousal near the stop points, which may not have occurred had participants watched the film straight through in its entirety. Similarity, interpreted as empathic accuracy in this context, between observer and target participant heart rates was assessed using correlations. A participant's physiological score was defined as the correlation statistic between their heart rate and the target's heart rate across the 19 included time points.

Convergent Validity Analyses

Convergent validity of self-report and performance-based measures was assessed using Pearson's product-moment correlation coefficients. Given the aforementioned heterogeneity of EA self-report measures, scores from the IRI and EFI were not aggregated into one composite total. I believed that doing so could be overly reductive, with the potential to obscure interesting nuances in the data that might otherwise be apparent. As such, IRI and EFI scores were compared independently in all subsequent analyses. Unlike self-report measures, the multiple-choice and free response performance-based components all related to a single, unifying paradigm. As such, it was determined that no unique information would be lost in an aggregate

score. The resulting “performance score” included a summation of the z-scores for the 8 performance-based multiple-choice questions and 8 associated open-ended response components. Finally, the two study-specific self-report measures of EA were each analyzed independently given their orientation to either affective or cognitive components of EA.

CHAPTER 3

RESULTS

Means and standard deviations across all indices are reported for non-therapists, novice therapists and advanced therapists in Table 2. Raw scores for the EFI, IRI, Accuracy 1 and Accuracy 2 assessments are reported. Performance scores for the paradigm task comprise aggregated z-scores from the relevant multiple-choice and open-ended paradigm questions. Physiological scores reflect correlations between the target and subsequent participants from the paradigm task.

Table 2. Means and Standard Deviations for all Indices across Non-therapists, Novice Therapists and Advanced Therapists

	EFI	IRI	Performance	Physiological	Accuracy 1	Accuracy 2
Non-Therapists						
Mean	74.58	19.08	-.39	-.05	6.58	5.53
SD	15.62	6.22	1.47	.29	1.73	2.17
Novice Therapists						
Mean	81.06	17.50	.152	.04	6.17	5.61
SD	13.25	3.54	1.75	.24	1.62	1.72
Advanced Therapists						
Mean	78.17	18.17	.586	-.042	6.44	5.61
SD	9.70	4.87	1.55	.265	1.20	1.34

Note: The Epstein Feelings Inventory (EFI: Mehrabian & Epstein, 1972) is an affective measure of EA, whereas the Interpersonal Reactivity Index (IRI; Davis, 1983) is cognitively oriented, raw scores for both are reported here. The performance measure comprised a summation of multiple choice and free response answers related to the standard-stimulus paradigm, and is reported as an aggregate z-score. Physiological scores reflect heart rate data collected during the same exercise and is reported as the correlation between participants and the paradigm-target. Accuracy 1 scores assess self-reported cognitive EA, whereas Accuracy 2 scores assess self-reported affective EA in relation to the paradigm task; both are reported as raw scores.

Correlations across all pairings in the sample as a whole were small (*r*'s ranged from -.091 to .200) and non-significant (see Table 3). However, two significant correlations emerged. The first occurred between scores on the IRI and the second accuracy question—"How much do you think your own feelings matched that of the person on the tape?" (r = .280, $p < .05$). Scores on the second accuracy question were also significantly correlated with scores from the first accuracy question—"How accurate do you think you were when trying to guess what the person on the tape was thinking and/or feeling?" ($r = .350$, $p < .01$), indicating that individuals who perceived themselves as more emotionally reflective of the person on the tape also perceived themselves as better able to interpret the target's thoughts or feelings.

A multiple regression was conducted using the three subscales of the IRI and the total score of the EFI as predictors, with the multiple choice score of the performance-based measures as the dependent variable. The model as a whole was nonsignificant (R-square = .077, F [4, 67] = 1.40, p = .25), with each predictor demonstrating nonsignificance as well, suggesting that self-reported empathy did not predict empathic accuracy in the paradigm task as indicated by multiple-choice scores. The same was true when the subscales of the IRI and the full score of the EFI were used as predictors of free-response questions (R-square = .076, F [4, 66] = 1.36, p = .259) and the combined "performance score" (R-square = .056, F [4, 67] = .997, p =.415) in the paradigm task. These three regression analyses indicate that self-report measures were not predictive of Empathic Accuracy in the form of either the multiple choice component or the open-ended free response section of the performance paradigm, nor of the aggregated score which combined these two components, corroborating the initial findings of the bivariate correlations.

Hypothesis 1a: Convergent Validity of Physiological Measures will be Low with Both Self-report and Performance-based Measures Regardless of Participant Condition

As with self-report and performance-based measures, convergent validity was assessed using Pearson's product-moment correlations. Analyses revealed small, non-significant correlations (*r*'s ranged from -.082 to .164) between physiological measures and all other indices. A multiple regression corroborated this, using all three subscales of the IRI and the total score of the EFI as predictors, and the physiological EA measure as the dependent variable. The model as a whole was nonsignificant (R-square = .057, F [4, 67] = 1.01, p = .41), as was each individual predictor. Thus, consistent with the bivariate correlations, self-rated empathy did not predict physiological EA in the form of heart rate correlation between participant and target during the trauma disclosure.

Table 3. Correlations of Self-report, Performance-based and Physiological Empathy Assessments Across all Participants

	EFI	IRI	Performance	Physiological	Accuracy 1
IRI	-.041				
Performance	.200	-.091			
Physiological	.164	.017	-.057		
Accuracy 1	-.060	.076	-.036	-.082	
Accuracy 2	.007	.280*	-.043	.053	.350**

Note: The Epstein Feelings Inventory (EFI: Mehrabian & Epstein, 1972) is an affective measure of EA, whereas the Interpersonal Reactivity Index (IRI; Davis, 1983) is cognitively oriented. The performance measure comprised a summation of multiple choice and free response answers related to the standard-stimulus paradigm. Physiological scores reflect heart rate data collected during the same exercise. Accuracy 1 scores assess self-reported cognitive EA, whereas Accuracy 2 scores assess self-reported affective EA in relation to the paradigm task.

*p<.05. **p<.01

Hypothesis 2: Correlations Between Self-report, Performance-based, and Physiological Measures will be Stronger Among Therapists than Non-Therapists.

Correlations were not consistently stronger for therapist than for non-therapist conditions (see Table 4). An independent samples t-test revealed that therapists (M=5.94, SD=1.62) did not score higher than non-therapists (M=5.87, SD=1.61) on the open-ended response section t(69)=.190, p=.85 of the performance-based paradigm. However, therapists (M=2.42, SD=1.08) did score higher on the multiple choice section of the performance paradigm than non-therapists (M=1.58, SD=1.05), t(70)=3.32, p<.01, as well as the aggregate performance score overall t(70)=2.06, p<.05. However, it is important to acknowledge that there were substantial differences in the standard deviations of self-report measures in association with therapist status. On the EFI non-therapists had a more variable score distribution (M=74.59, SD=15.62) than advanced therapists (M=78.17, SD=9.70), and the same was true with scores of the IRI across non-therapists (M=19.08, SD=6.22) and advanced therapists (M=18.17, SD=4.87). The decreased spread of scores among advanced therapists self-report measures (as demonstrated by smaller standard deviations) indicates a restriction of range, and a decreased likelihood of finding significant correlations across this data as opposed to novice therapist or non-therapist scores. This reduction in self-reported significance could consequentially impede correlations between the self-report measures and other indices. For example, it may be especially challenging to show a significant correlation between both self-report measures in advanced therapists, given that both measures seem to be operating in conjunction with a noticeably restricted range.

Among non-therapists, the second accuracy question—"How much do you think your own feelings matched that of the person on the tape?"—was correlated with the IRI (r =.366, p<.05). For non-therapists, the second accuracy question also correlated with the first accuracy question—"How accurate do you think you were when trying to guess what the person on the

tape was thinking and/or feeling?" (r =.448, p<.01). However, when compared to the therapist condition, neither the correlation between the IRI and the second accuracy question (z=.448, p=.121) nor the accuracy questions correlated with each other (z=1.27, p=.102) were significantly different.

Interestingly, among therapists, the second accuracy measure did not correlate with the IRI (r=.070, p=.684) or with the first accuracy measure (r=.191, p=.265). However, these correlations were not significantly different from those of the non-therapist group (z=1.27, p=.102) and (z=1.17, p=.121) respectively. A moderate correlation emerged between physiological measurements and the EFI (r=.531, p<.01). This correlation between affective empathy and physiological arousal occurred at a level above and beyond non-therapist participants (r=.531, p<*.05*; z = -2.77, p<.01).

Table 4. Correlations of Self-report, Performance-based, and Physiological Empathy Assessments Among Therapists and Non-Therapists

	EFI	IRI	Performance	Physiological	Accuracy 1
IRI					
Non-therapist	.021				
Therapist	-.100				
Performance					
Non-therapist	.253	-.105			
Therapist	.065	-.018			
Physiological					
Non-therapist	-.090	.099	-.035		
Therapist	.531**	-.095	-.126		
Accuracy 1					
Non-therapist	-.026	-.021	-.011	-.123	

Therapist	-.077	.226	-.020	-.077	
Accuracy 2					
Non-therapist	.021	.366*	.047	-.050	.448**
Therapist	-.015	.070	-.177	.220	.191

Note: The Epstein Feelings Inventory (EFI: Mehrabian & Epstein, 1972) is an affective measure of EA, whereas the Interpersonal Reactivity Index (IRI; Davis, 1983) is cognitively oriented. The performance measure comprised a summation of multiple choice and free response answers related to the standard-stimulus paradigm. Physiological scores reflect heart rate data collected during the same exercise. Accuracy 1 scores assess self-reported cognitive EA, whereas Accuracy 2 scores assess self-reported affective EA in relation to the paradigm task.

*p<.05. **p<.01

Hypothesis 3: Correlations of Self-report, Performance-based, and Physiological EA Measures will be Stronger Among Advanced Therapists than Novice Therapists

Contrary to expectations, advanced therapists did not consistently show stronger correlations than their less-experienced counterparts (see Table 5). The correlation between EFI scores and physiological measurements were significant across both advanced therapists (r=.555, p<.05) and beginning therapists (r=.514, p<.05), and a Fisher's Transformation suggests that the differential strengths of these correlations did not exceed chance levels (z=.23, p=.81). As such, therapists' self-reported level of affective EA was moderately related to their actual physiological symmetry with the paradigm target, independent of clinical training status. Using a Spearman correlation to account for response outliers in the advanced therapist accuracy 1 data, a moderate correlation emerged between the IRI and the first accuracy question (r=.520, p<.05). This correlation was noticeably different from beginning therapists, who showed no such relationship (r=.015, p=.952; z=2.08, p=.038).

A significant inverse correlation emerged between beginning therapists' scores on the IRI and physiological assessments (r=-.508, p<.05), suggesting that novice therapists who believed they were more cognitively empathic showed physiological responses that were less

synchronized with the performance-based paradigm target. This correlation differed significantly from advanced therapists (Z=2.08, $p<.05$), who did not demonstrate a comparable relationship between the IRI and physiological measures ($r=.198$, $p=.431$). Additionally, among beginning therapists, scores on the EFI and IRI were inversely related at a non-significant level ($r=-.390$, $p=.110$). This correlation suggests that among novice therapists, increased cognitive empathy may be related to decreased affective empathy and vice versa.

Table 5. Correlations of Self-report, Performance-based and Physiological Empathy Assessments Among Advanced and Novice Therapists

	EFI	IRI	Performance	Physiological	Accuracy 1
IRI					
Advanced	.197				
Beginning	-.390				
Performance					
Advanced	.115	-.204			
Beginning	.063	.182			
Physiological					
Advanced	.555*	.198	.120		
Beginning	.514*	-.508*	-.329		
Accuracy 1					
Advanced	.206	.440	.130	-.099	
Beginning	-.212	.015	-.141	.096	
Accuracy 2					
Advanced	-.086	.210	-.311	.038	.078
Beginning	.024	-.073	-.089	.384	.257

Note: The Epstein Feelings Inventory (EFI: Mehrabian & Epstein, 1972) is an affective measure of EA, whereas the Interpersonal Reactivity Index (IRI; Davis, 1983) is cognitively oriented. The performance measure comprised a summation of multiple choice and free response answers related to the standard-stimulus paradigm. Physiological scores reflect heart rate data collected during the same exercise. Accuracy 1 scores assess self-reported cognitive EA, whereas Accuracy 2 scores assess self-reported affective EA in relation to the paradigm task.

*p<.05. **p<.01

Exploratory Analyses

Relevant literature suggests that the IRI may be one of the only measures of EA that includes both cognitive and affective components (Spreng et al., 2009; Davis, 1980). However, as mentioned, the modifications that were made to the IRI for the present study resulted in a somewhat cognitive redistribution of questions. As such, exploratory analyses were done to discern if certain subscales were more strongly correlated with various indices than others, and if so, how this was impacted by therapist or non-therapist status. The subscales showed significantly different relationships with three indices: the Accuracy 2 question, EFI, and physiological data (see Table 6).

Table 6. Correlations of IRI Subscales with Various Measures Across Non-Therapists, Novice Therapists and Advanced Therapists

	Perspective Taking	Fantasy	Personal Distress
Accuracy 2			
Non-Therapists	*.377**	*.438***	*.119*
Beginning Therapists	.122	-.066	-.101
Advanced Therapists	*-.094*	*.356*	*.044*
EFI			
Non-Therapists	*.084*	*.086*	*-.090*
Beginning Therapists	.047	-.103	-.565*

Advanced Therapists	*.682***	*.062*	*.107*
Physiological			
Non-Therapists	*-.383**	*.119*	*.142*
Beginning Therapists	-.112	-.285	-.468*
Advanced Therapists	.489*	.239	-.009

Note: Accuracy 2 scores assess self-reported affective EA in relation to the paradigm task. The Epstein Feelings Inventory (EFI: Mehrabian & Epstein, 1972) is an affective measure of EA, whereas the Interpersonal Reactivity Index (IRI; Davis, 1983) is cognitively oriented. Physiological scores reflect heart rate data collected during the performance paradigm.

*p<.05. **p<.01

Among non-therapist controls the IRI was correlated with the second accuracy question—"How much do you think your own feelings matched that of the person on the tape?" (r=.366, p<.05). The Fantasy subscale (r=.438, p<.01) and the Perspective Taking subscale (r=.377, p<.05) were moderately associated with the second accuracy question, while the Personal Distress scale did not demonstrate such a relationship (r=.119, p=.489). Dunn and Clark's (1969) method for comparing correlated correlations revealed that the three subscales all factored into the correlation with the second accuracy question at a comparable level. However, scores of the Fantasy subscale were strongly correlated with Accuracy 2 status, almost above and beyond the Personal Distress subscale, the comparison of which approached significant (z=1.85, p=.06).

Among therapists, the IRI no longer correlated with accuracy 2 scores (r =-.073, p=.775), and all subscales reflected this lack of relationship (r's ranged from -.101 to .122). Among advanced therapists, neither the Perspective Taking subscale (r=.094, p=.712) nor the Personal Distress subscale (r=.044, p=.863) showed a directional relationship with Accuracy 2 scores. However, scores on the Fantasy subscale demonstrated a sizeable increase in association with

Accuracy 2 scores (r=.356, p=.147), although not at a level that differentiated this subscale from the Perspective Taking (z=1.41, p=.16) or Personal Distress subscales (z=1.19, p=.23).

The subscales of the IRI also showed significant variance with the physiological data. Among controls, the Perspective Taking scale was inversely correlated with the participants heart rate correlations (r=-.383, p<.05), at a level that above and beyond the Personal Distress scale (r=.142, p=.407; z=2.27, p<.05) and the Fantasy Scale (r=.119, p=.490; z=2.17, p<.05). Among novice therapists, the Personal Distress scale became moderately inversely correlated with physiological data (r=-.468, p=.05) more so than for non-therapist controls (r=.142, p=.407; z=-2.09, p<.05). Among advanced therapists, scores on the Perspective Taking scale become more strongly correlated with physiological data (r=.489, p<.05) than for novice therapists (r=-.122, p=.657; z=1.8, p<.05) or non-therapists (r=-.383, p<.05; z=3.01, p=.001). While this was not at a level significantly different than the Fantasy scale (r=.239, p=.339; z=.87, p=.*38*), it was differentiated from the Personal Distress Scale (r=-.009, p=.971; z=1.37, p=.17).

Although the IRI did not correlate significantly with the EFI at the level of the full sample (r=-.041, p=.753), interesting differences came to light through targeted analyses of the included subscales. Non-therapist controls showed almost no relationship between the subscales of the IRI and EFI as a whole (r's ranged from -.090 - .086). However, among novice therapists the Personal Distress subscale became strongly inversely associated with the EFI (r=-.565, p<.05) and more predictive of EFI scores than the Perspective Taking subscale (r=.047, p=.852; z=-2.15, p<.05), but not the Fantasy Subscale (r=-.103, p=.685; z=-1.28, p=.20). Interestingly, this strong relationship dissipated among advanced therapists (r=.107, p=.673). Advanced therapists exhibited a strong direct relationship between the Perspective Taking subscale and the EFI (r=.682, p=.002) above and beyond both novice therapists (r=.047, p=.852; z=2.15, p<.05)

and non-therapist controls (r=.084, p=.627; z=2.40, p<.01). Further, among advanced therapists, the Perspective Taking subscale was more predictive of EFI scores (r=.682, p<.01 than the Fantasy Subscale (r=.062, p=.806; z=2.10, p<.05), but not the Personal Distress subscale (r=.107, p=.673; z=1.77, p=.08).

CHAPTER 4

DISCUSSION

Interestingly, very little research has compared the utility of various measures of Empathic Accuracy, despite known heterogeneity in the construct and among the measures most frequently used to assess it. There is a paucity of literature on the convergent validity of self-report, performance-based and physiological measures of EA, despite their frequent interchange and overlap in the relevant literature. Additionally, there is little insight into the impact that sample characteristics may have on the relevance of these measures, and if various assessments may demonstrate higher accuracy among certain demographics than others. Theoretically, therapists should be more accurate in self-reported measures of EA, due to the centrality of emotional awareness in clinical psychology and the extensive training clinicians receive in preparation for this field. As a result, convergent validity across these subjective indices and more objective assessments should be stronger in a therapist sample. The current study attempted to elucidate these psychometric issues, assessing the convergent validity of self-report, performance-based and physiological assessment modalities among therapist and non-therapist samples.

Because both self-report and performance-based measures of EA primarily assess cognitive empathy, it was expected that they would correlate moderately well across the entire sample. It was also predicted that these measures would correlate poorly with physiological measures of EA, which have a known affective orientation. Across the sample as a whole, correlations of self-report, performance-based and physiological modalities were small, with only two achieving significance. The second accuracy question—"How much do you think your own feelings matched that of the person on the tape?" was moderately correlated with The Interpersonal Reactivity Index (r=.280, p< .05), as well as with the first accuracy question—

"How accurate do you think you were when trying to guess what the person on the tape was thinking and/or feeling?" (r = .350, p<.01). The first correlation suggests that individuals who rated themselves as more cognitively empathic at a categorical level also rated themselves as more affectively aligned with the target participant throughout this specific paradigm.

The second significant correlation between the two accuracy questions suggests that individuals who perceived themselves as having affective experiences more similar to the paradigm participant also felt they were more accurate in interpreting her emotions. This could be reflective of the reciprocal relationship that Ickes suggested throughout his research, in that cognitive interpretations inform and are informed by affective experience. If a participant feels unable to identify with a person's emotional presentation then they may also experience less physical arousal in response to the interaction, and report lower levels as a result. Conversely, if an individual is less affectively stimulated, they may experience less cognitive similarity. This does not reflect actual performance abilities, but as self-reporters of EA ability, it is important to understand the trends that occur in reporting perceived ability and accuracy among participants, as such nuances often influence data in unexpected ways. Studies involving populations that might be especially skilled—or lacking—in one area of EA versus another may help elucidate any differences in the importance of cognitive versus affective components, and better clarify the overlap and relationship between the two. For example, understanding client's affective or cognitive abilities among individuals with borderline personality disorder (BPD) may help clarify if this specialized population is more well-versed in cognitive EA rather than affective EA, or vice versa. If it is the case that individuals with BPD excel in one area of EA while exhibiting deficiencies in another, scores on generalized assessments may be inaccurate. Given that deficiencies in affective abilities would deflate the impact of higher scores on cognitive

ability, composite scores may reflect inaccurately low levels of EA among this specific population. It is not currently understood to what degree affective or cognitive abilities factor into daily interactions, and more nuanced assessments may facilitate this understanding. If study designs simply utilize one measurement of EA (which is common practice among researchers not well-versed in this literature) results could range from misleading to inaccurate. As such, it is important that researchers acknowledge exactly what they are measuring when using indices of EA, and report their findings accordingly.

The unexpected lack of relationships between self-report and performance-based measures of EA across the full sample may be indicative of larger variations in psychometric construct validity than are currently assumed in the literature. Given the heterogeneity of the construct (Gesn & Ickes, 1999), there may be variations in construct representation across measurement modality (i.e. between self-report and performance-based measures) as well as within these discrete categories (i.e. between the IRI and EFI). These deviations could reflect the cognitive and affective branches of EA as previously mentioned, or the various stages of the empathic process of emotional experience, interpretation and expression (Gerdes & Segal, 2011). It is also possible that some third, unidentified variable—such as participant demographics—may factor into the correlations in a way that has been, as of yet, undetected. Given the relatively large sample (72 participants overall) it is unlikely that null-findings were the result of insufficient sample size. Further, given the number of analyses run on these data it would be more likely to suffer a Type I error than a Type II as indicated here. However, nonsignificance could be a result of the measures that were selected and the ways in which they were modified for the present study. Alterations to the wording, length, and subscale representation of the IRI may have caused some changes to the content validity of the measure, which could be reflected

in reduced convergent validity with performance-based measures. The EFI and IRI were specifically selected for their somewhat distinct representations of EA (cognitive vs. affective) in order to determine if they could overcome such differentiation and correlated based on the "underlying" root construct of EA. Given this, null findings are not surprising, but are important. Analyzing EA measures at a factor analytic level may help clarify this noticeable deviance.

The target participant's physiological data was not tremendously variable (her heart rate spanned 69 to 84 beats per minute) in the paradigm recording. However, a descriptive look at the data suggested two potentially meaningful phases—consisting of the first and second halves of the recording. A correlation comparing these two components revealed that the first and second sections were not significantly related to each other (r=.002, p=.990). However, heart rate data collected in the first ten minutes of the film were moderately associated with participants' overall physiological similarity to the target participant (r=.455, p<.01), and data from the second half of the film were significantly more strongly associated with participants' overall physiological similarity to the target (r=.766, p<.01; z=-3.05, p<.01). It may be that participants' own physiological responses to completing the paradigm task interfered with their similarity to the target participant in the start of this film. Any nervousness or anxiety from being involved in a novel task may have been exposed in participants' physiological data, reflecting their own general arousal rather than their ability to reflect the target participant. Further, participants knew that there would be a narrated account of sexual trauma in the film, but not the intensity with which it would be presented. It may have been that anticipatory anxiety altered participants' physiological arousal, making it less similar to the target participant. By the second half of the film the novelty of the task itself may have diminished, and the content of the film was largely disclosed. As such, participants' scores in the second half of the film may be more accurate

reflections of their physiological similarity to the target. To test this, subsequent analyses were conducted comparing correlations from the second-half of the performance paradigm and all other measures, but no significant correlations emerged.

Interestingly, correlations between EA measures were not consistently stronger for therapists or non-therapists. Non-therapists exhibited correlations between the second accuracy question and the IRI as well as between both accuracy questions. The Fantasy and Perspective Taking subscales largely drove the correlation between the IRI and the first accuracy question. It may be that cognitive components of empathic accuracy are especially important among non-therapists in determining their affective arousal levels and accurately reporting these. In support of the initial hypothesis, a moderate correlation emerged among therapists between physiological measurements and the EFI, such that higher levels of self-reported affective response were associated with greater physiological arousal. This occurred at a level above and beyond the comparable correlation among non-therapists. Theoretically, therapists should have a more developed understanding of affective arousal at a conceptual level, their own affective tendencies, and the ways in which they use these with clients. For example, it is possible that therapists may use affective synchronicity with their clients to develop a strong therapeutic alliance. By truly "experiencing" a clients' emotion, therapists may be more capable of verbalizing them in a useful and intimate way. Conversely, it may be that therapists who are especially sensitive to affective arousal need to consciously become emotionally desynchronized with, or distanced from, clients in order to protect against compassion fatigue and burnout caused by chronic affective overstimulation. In either case, a therapist should be more capable of reporting their tendency towards either end of the spectrum, and scores on objective measures that assess these tendencies should be more closely correlated with the subjective measures as a

result. This data does reflect such similarity, and suggests that therapist reports of affective arousal are more strongly correlated than non-therapists, and provides some initial indication that the EFI may be differentially accurate across participants.

Among the therapist sample, the second accuracy measures were not correlated with the IRI or the first accuracy measure. As such, participants' reported ability to identify emotions and express them on a daily basis did not have a relationship with the affective closeness they experienced with the target in the current paradigm. These findings may have emerged for several reasons. First, therapists may be better equipped to affectively seclude themselves in empathically arousing situations, as with the paradigm task. Given the potential emotional drain inherent to clinical work, most clinical training involves some aspect of self-care and preservation; this often relates to emotional susceptibility. As such, limiting affective arousal may be especially important among this sample, which could be supported through consistently low levels of affective arousal among therapists. However, it is also possible that this target was especially challenging to identify with emotionally. This is supported by the fact that while therapists might rate themselves as Empathically skilled on a general level; this belief did not translate into the context-specific paradigm. This is especially likely given that there are no current "standards of excellence" for paradigm construction, and as such, the crafting of such stimuli is left to the discretion of a given research team. As a result, it is possible that certain paradigms may be less straightforward than others. Because the target participant had completed her own therapy prior to her participation, and the substantial time lapse between the traumatic event itself and her report, the valence of her affective report may have been tempered to a degree that made it challenging for participants to interpret.

Therapists' views of their own physiological similarity to the target (as assessed by the second accuracy question) were unrelated to their perceived ability at cognitive empathy in this setting. This suggests that despite having a distanced affective experience with the target, their perceived relationship or emotional reciprocity was unrelated to their ability to interpret, conceptualize and state what the target person was thinking or feeling. It is possible that therapists develop an understanding of their overarching EA abilities, and when a situation occurs in which they vary from their typical performance therapists are more able to recognize that. However, it could also be the case that continuous clinical involvement creates an over-inflated report of global EA ability, which would then lead to weaker correlations between more general reports and paradigm-restricted ratings. Future studies would benefit from examining the impact that clinical practicum has on meta-cognition and self-awareness among training therapists, as this may have a substantial impact on the way that psychometric tools reflect their responses. More adept EA psychometric measures may be especially relevant for this setting, given that these might aid in training therapists in attempting to avoid emotional contagion.

Contrary to expectations, advanced therapists did not consistently show stronger correlations than their novice counterparts. Correlations between the EFI and physiological measurements were significant across all therapists, regardless of clinical training level. As such, therapist's perceived level of affective EA was moderately related to physiological arousal, independent of their level of clinical training. If this is the case, then self-reported levels of affective EA may be reasonably employed among therapists. Because of the tighter correlation between physiological arousal and affective self-report, the former may not be necessary in more perfunctory assessments of therapists.

A unique challenge of physiological measurements is that they may be more accurate reflections of generalized arousability as opposed to empathic accuracy specifically. As physiological arousal was moderately associated with therapists' self-reported affective EA, it may be that these measures are more appropriate for those with training in the academic and practical application of the variable itself. This theory is bolstered by the fact that the relationship between affective EA and physiological arousal was not mirrored among the non-therapist sample. This lack of significance implies that physiological assessments may be inappropriate for general administration—at least at this stage in their development. Further work with physiological measurements would benefit from analyses on which types of assessment (i.e. heart rate, skin conductance, respiratory rate, etc.) and in which combinations are most appropriate for various populations.

Beginning therapists demonstrated a significant inverse relationship between scores on the IRI and physiological assessments, suggesting that those individuals who reported lower levels were less cognitively empathic showed physiological responses that were more synchronized with the target in the performance-based paradigm. This relationship was largely driven by the Personal Distress subscale of the IRI. It may be that metacognitive awareness of personal distress leads to intentional distancing from clients at an affective level, which is reflected in physiological data, here assessed as heart rate similarity. However, among advanced therapists only the Perspective Taking subscale of the IRI was significantly and positively correlated with physiological arousal, rather than the IRI as a whole. For advanced therapists then, the ability to cognitive adopt the perspective of their clients, rather than experience the same level of cognitive distress may be the basis of affective synchronicity. As therapists become more familiar with common affective reactions across clients, it may be less germane for

them to explicitly process the experience of a given emotion; they may instead be able to implicitly gain perspective and consequentially reach physiological similarity. In support of this, the Perspective Taking subscale of the IRI was more strongly correlated with physiological scores than the Personal Distress scales among advanced therapists, and this was not apparent in novice therapists or non-therapists. Interestingly, non-therapists actually demonstrated an inverse relationship between the Perspective Taking subscale and physiological scores. For non-therapists, the cognitive processing may hinder their ability to relate to others at an affective level. This finding was corroborated at a descriptive level through an inverse relationship between scores on the EFI and IRI among beginning therapists. Relatedly, beginning therapists physiological measurements were loosely associated with Accuracy 2 scores, such that those participants whose heart-rate data more closely aligned with the target believed they were more able to adopt her perspective.

When taken together, the data suggest that among beginning therapists, affective and cognitive EA may range from uncorrelated to inversely related. At a practical level, this is concerning, given that both affective and cognitive facets of EA are imperative to a supportive clinical setting (Myers, 2000; Seehausen et al., 2012). As such, it is important for training clinicians to obtain an accurate rating of their relevant levels of each facet, in order to better focus on areas in need of improvement. If only one assessment it administered, it is likely that such training clinicians will obtain an inaccurate global rating that may over- or underestimate their actual scores. It would be important to corroborate this fluctuation in cognitive and affective ability among training therapists in a more robust sample. It may also be useful to explore the relevant impact of various therapeutic approaches (for example, Acceptance and Commitment Therapy versus psychodynamic therapy) on therapist ratings and reflections in EA

assessments. If such tools are to be used on a practical, clinical training level, it is important to obtain an accurate understanding of base rates and norms among relevant clinical populations.

Advanced therapists demonstrated a moderate correlation between scores on the IRI and the first accuracy question, such that those who interpreted themselves as higher in cognitive empathy at a global level also felt that they demonstrated higher levels of cognitive EA in this specific paradigm task. This correlation, versus those of non-therapists who demonstrated a significant correlation between the IRI and the second accuracy question—which refers to a more affective skillset—supports our hypothesis that therapist measures should correlate better than non-therapist scores. Because both the IRI and the first accuracy questions are cognitively skewed, individuals who perceived themselves as cognitively attuned on an expansive level should also find that they were cognitively proficient in this specific instance as well. Whereas non-therapist participants showed a link between affective self-report and global cognitive EA capacity, therapists demonstrated a strong correlation between both global and local measures of cognitive EA. It may be that therapists recognize and report more strongly on measures of cognition—reflecting the training they have received, whereas non-therapists rely upon and consequentially reflect affective components of EA. If this is the case, then self-report measures directed at cognitive capacities may be missing valuable affective information from non-therapist participants, which are the more likely sample population for such studies.

Limitations

Several methodological limitations should be mentioned in reviewing this study. The EFI and IRI were intentionally included for their constructive slants (i.e. affective versus cognitive orientation), as well as the frequency with which these measures are utilized in the relevant literature. However, for this study the IRI was modified from its original version to more

accurately reflect the aims of the project. Only 10 of the 28 questions were utilized in this study, representing the three subscales of Perspective-taking, Fantasy and Personal Distress, while omitting questions from the Empathic-Concern scale (Davis, 1980). This was done in an attempt to reduce the duration of the study, while still obtaining representative data. While the full measure is typically considered relatively balanced with a slight cognitive tendency, the modifications that were made for this study—specifically which questions were included or omitted—made the cognitive orientation somewhat more pronounced. As such, readers should be cautious in directly translating the findings from this study to the IRI in general, and follow-up analyses using the measure in its entirety are warranted.

Additionally, performance-based measures do not have a "standard" template for construction; instead, each rendition utilizes a novel, unvalidated stimulus. One important source of variance in the paradigm is the target participant who narrates their traumatic experience on film. For example, an individual's meta-cognitive abilities, or the accuracy with which they interpret and report on their own traumatic experience, can have a sizeable influence on related performance scores. The target participant in this study had participated in therapy post-trauma, and the filming took place more than two decades after her traumatic experience occurred. As such, the effusiveness with which she narrated her story may have been somewhat dampened, which is of primary importance in the standard-stimulus paradigm (Marangoni et al., 1995). Thus, subsequent participants may have found it particularly challenging to identify with her throughout the film.

Additionally, scores on the performance-based paradigm seem to suggest that our version of this measure may have been exceptionally challenging. Across the entire sample, the highest score of the multiple choice portion was 5 of the total possible 8 points, and the mean score was

significantly lower (M=2.00, SD=1.14), revealing a significant floor effect. The spread of the open-ended portion of this paradigm was also somewhat restricted, with scores ranging from 2.5 to 9.5, (M=5.908, SD=1.61) out of a total possible 16 points. It is important to acknowledge that this limitation is not restricted to this study alone, but inherent to the measure and its construction in general. This conflict exposes a paradox inherent to the paradigm format: eventually affective expression crosses the threshold from emotionally evocative stimuli to potentially harmful exposure. If such measures are continually utilized in the relevant literature, it will be important to further refine their construction, in hopes that a more standardized and generalizable performance-based format can be developed. An important improvement may be the inclusion of a pilot film prior to the full-study procedure, despite the added fiscal and temporal burden such modifications could incur.

A third limitation to this study was the single source of physiological data that was collected. While heart rate is a common indicator of affective arousal, it may be that other assessments or combinations of assessments may provide more accurate reflections of EA induced physiological fluctuations. For example, Shortt and Pennebaker (1992) utilized both heart rate (HR) and skin conductance level (SCL) to assess physiological arousal, and found that SCL was more sensitive to affective changes induced by EA. It is possible that using SCL in this study, or a combination of both HR and SCL may have elicited differential strengths or directions across correlations of the chosen measures, and any follow up studies should consider such an approach.

There are certain demographic variables that may have produced ambiguity in this study as well. First, the distinction between advanced and beginning therapists spans four years of clinical training—during which a great amount of information will be conferred to doctoral

students, but may not be reflected in dramatic shifts in performance. Subsequent studies should include beginning therapists (doctoral training) with truly "advanced therapists"—those practitioners with numerous years of hands-on clinical experience. If as, the literature suggests, EA is susceptible to training, differentiation between beginning and advanced therapists may be better exposed with a greater maturation period than four to five years (Barone et al., 2005). Conversely, if the measures truly correlate to the same degree regardless of therapeutic training, this should be clearly recognized in a more variegated therapist sample.

The ethnic distribution across therapist and non-therapist conditions may have also confounded the results. As noted (Table 1) the non-therapist condition included 11 fewer Caucasian participants and 9 additional African American participants compared to the therapist condition. As such, there may be differences across ethnic categories reflected in the data, rather than simply those determined by therapist identification. An independent samples t-test was conducted to determine if scores on the performance-based measure differed significantly based on participants' ethnic match (Caucasian) or mismatch (not Caucasian) to the target participant. There was not a significant difference in Caucasian participants' scores (M=.06, SD=1.70) or all other participants' scores (M=-.30, SD=40) in the performance-based paradigm; t(70)=-.726, p =.338. Additionally, it has been documented that across EA assessments, women typically exhibit higher levels than men (Baron-Cohen & Wheelwright, 2004). However, an independent samples t-test revealed that there was not a significant difference between male (M=-.01, SD=2.00) and female (M=-.01, SD=1.42) scores on this paradigm measure t(70)=-.001, p=.301). Finally, a correlation comparing participant age and performance-based scores revealed that age, as with race and gender, were not related to scores on the performance-based task (r=-.099, p=.408).

A final limitation of this study is that the sample size may have clouded the findings and emergent correlations. While the overall analyses included 72 participants—a fairy robust sample for generalized comparisons—the subsequent division of 36 participants in therapist and non-therapists conditions, and 18 participants in advanced and beginning therapist conditions may have resulted in an inflated likelihood of type 2 errors and insufficient power. Subsequent analyses would benefit from a larger sample size across sub-conditions in order to confirm or reject the preliminary data from the present study.

Implications

Empathic accuracy is a known cornerstone of social interactions, and may be especially important among clinical psychologists, who frequently use EA techniques to establish and strengthen a therapeutic dyad (Seehausen et al., 2012). However, given the multifaceted nature of EA and the more unilateral assessment tools currently employed in the literature, it may be that researchers are failing to ascertain participants' true, nuanced capabilities. This could be detrimental and misleading for researchers and clinicians hoping to better understand their own, or others experiences of EA. One way to overcome such empirical restrictions may be through the administration of multiple assessments, each comprising unique perspectives on the empathic accuracy construct. If cognitive and affective components of EA are as integral to social interactions as the literature suggests, then it will be important for these two branches of research to integrate—mirroring the importance of each component independently and as it interacts with its counterpart.

A possible avenue for future research would be the inclusion of factor analytic methods and external criterion variable analysis in the revision and modification of existing measures. This process could aid in developing more comprehensive and valid tools, suited for this

complex construct. Such revisions may be especially necessary for those researchers hoping to assess therapist or caregiver EA—a sample of individuals whom our data suggest may respond uniquely to such indices. While both cognitive and affective variables intercorrelated across therapist and non-therapist samples, they failed to do so at the most general level of analysis, suggesting that demographic variables may factor into composite scores in a more substantive way than currently assumed in the literature. It will be important to identify which individuals excel in various aspects of empathy, and which methods are best equipped at improving deficits. If this is to be done, those measures purported to tap into the construct will need to become more adept at doing so, allowing both researchers and clinicians to accurately pursue the construct in empirical and clinical pursuits.

APPENDIX A

MULTIPLE CHOICE MEASURE WITH CORRECT RESPONSES UNDERLINED

Film clip stop points are indicated in parentheses.

1. (3:27)
    a. <u>Anxiety, sadness, some remembering fear</u>
    b. Confusion about what was happening
    c. Anger for what they were doing to me
    d. Violated, afraid, wishing I fought back

2. (4:02)
    a. <u>Feeling lots of anxiety, pain in remembering</u>
    b. Feeling very violated, anger in remembering
    c. Sadness in remembering
    d. Feeling detached from what had happened, thinking how long ago it seems now

3. (5:23)
    a. <u>Anger at my mother, strong hurt</u>
    b. Wishing my mother would have been more understanding
    c. Sadness at how unfair my mother was to me
    d. Anxiety about what had happened to me

4. (7:37)
    a. <u>Feeling depth of loss, thinking of surviving</u>
    b. Feeling sad for what I had been through
    c. Worry that I would be scarred for life
    d. Thinking of how no one understood what I was going through

5. (14:12)
    a. <u>Deep pain, anger at my mother, hurt</u>
    b. Deep sadness, anger at the boys who hurt me
    c. Feeling betrayed by my mother, hopeless
    d. Thinking what might have been different if my mother was more understanding

6. (:34) (second segment)
    a. Remembering at that time how happy I was
    b. Remembering how god it felt to move on
    c. Remembering how difficult it was to try to act normal
    d. Remembering how sad I still felt

7. (1:27)
    a. Remembering how painful, like being kicked in the stomach
    b. Remembering how surprised I was to hear the news
    c. Remembering how I felt like all of my progress had gone out the window
    d. Remembering being scared for my father

8. (3:00)
    a. The darkest day, intense sadness, loss
    b. Guilt that I hadn't been there enough for him before
    c. Scared that I was feeling everything coming back from my past
    d. I felt vulnerable, afraid, sad

APPENDIX B

TARGET PARTICIPANT'S HEART RATE TRACKING DATA FROM STANDARD-STIMULUS PARADIGM

Minute	Heart Rate
1	77
2	78
3	75
4	76
5	75
6	73
7	72
8	71
9	69
10	72
11	72
12	70
13	73
14	78
15	81
16	84
17	77
18	74

APPENDIX C

SELF-REPORT ACCURACY QUESTIONS FOR STANDARD-STIMULUS PARADIGM

Please answer the following questions on a scale from 1-10:
1=Not at all accurate, 5=neither accurate or inaccurate, 10=completely accurate

1. ___ How accurate do you think you were when trying to guess what the person on the tape was thinking and/or feeling?

2. ___ How much do you think your own feelings matched that of the person on the tape?

APPENDIX D

MODIFIED IRI QUESTIONNAIRE (ADAPTED FROM THE INTERPERSONAL REACTIVITY INDEX; DAVIS, 1980)

Please rate each statement from 0 (does not describe me well) through 4 (describes me very well).

1. ___ I felt like I was able to remain calm while watching the person on the tape.
2. ___ If I was in the room with the person on the tape while she was telling her story, I would feel helpless.
3. ___While watching the person in the tape, I felt like I could "go to pieces."
4. ___While watching the person in the tape, I felt scared.
5. ___ While watching the tape, I imagined how I would feel if I were the person in the tape.
6. ___ I really got involved with the feelings of the person in the tape.
7. ___After watching the tape, I felt as if I were the person in the tape.
8. ___ I felt like it was difficult to feel very involved with the story of the person in the tape.
9. ___ I remained objective and didn't get completely caught up in the person on the tape's story.
10. ___I found it difficult to see things from the point of view of the person on the tape.

APPENDIX E

EPSTEIN FEELINGS INVENTORY (MEHRABIAN & EPSTEIN, 1972)

Please rate each item from +4 (very strong agreement) to -4 (very strong disagreement).

+4 +3 +2 +1 0 -1 -2 -3 -4
Very strong agreement — Neutral — Very strong disagreement

1. ___ It makes me sad to see a lonely stranger in a group
2. ___ People make too much of the feelings and sensitivity of animals.
3. ___ I often find public displays of affection annoying
4. ___ I am annoyed by unhappy people who are just sorry for themselves.
5. ___ I become nervous if others around me seem to be nervous.
6. ___ I find it silly for people to cry out of happiness
7. ___ I tend to get emotionally involved with a friend's problems
8. ___ Sometimes the words of a love song can move me deeply
9. ___ I tend to lose control when I am bringing bad news to people
10. ___The people around me have a great influence on my moods.
11. ___ Most foreigners I have met seemed cool and unemotional
12. ___ I would rather be a social worker than work in a job training center
13. ___ I don't get upset just because a friend is acting upset
14. ___ I like to watch people open presents.
15. ___Lonely people are probably unfriendly. (Continued on next page)

+4 +3 +2 +1 0 -1 -2 -3 -4
Very strong agreement — Neutral — Very strong disagreement

16. ___ Seeing people cry upsets me

17. ___ Some songs make me happy

18. ___ I really get involved with the feelings of the characters in a novel

19. ___ I get very angry when I see someone being ill-treated

20. ___ I am able to remain calm even though those around me worry.

21. ___ When a friend starts to talk about his problems, I try to steer the conversation to something else

22. ___ Another's laughter is not catching for me

23. ___ Sometimes at the movies I am amused by the amount of crying and sniffling around me

24. ___ I am able to make decisions without being influenced by people's feelings

25. ___ I cannot continue to feel OK if people around me are depressed.

26. ___ It is hard for me to see how some things upset people so much

27. ___ I am very upset when I see an animal in pain.

28. ___ Becoming involved in books or movies is a little silly

29. ___ It upsets me to see helpless old people

30. ___ I become more irritated than sympathetic when I see someone's tears.

31. ___ I become very involved when I watch a movie

32. ___ I often find that I can remain cool in spite of the excitement around me

33. ___ Little children sometimes cry for no apparent reason.

REFERENCES

Barone, D., Hutchings, P.S., Kimmel, H.J., Traub, H.L., Cooper, J.T., & Marshall, L.M (2005). Increasing empathic accuracy through practice and feedback in a clinical interviewing course. *Journal of Social and Clinical Psychology, 24*(2), 156-171.

Buck, R., Savin, V., Miller, R., & Caul, W. . (1972). Communication of affect through facial expression in humans. *Journal of Personality and Social Psychology, 23*(3), 362-371.

Chlopan, B. E., Mccain, M. L., Carbonell, J. L., & Hagen, R. L. (1985). Empathy - Review of Available Measures. *Journal of Personality and Social Psychology, 48*(3), 635-653. doi: Doi 10.1037//0022-3514.48.3.635

Crandall, J. E. , & Harris, M. D. (1976). Social interest, cooperation, and altruism. . *Journal of Individual Psychology, 32*, 50-54.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *J Consult Psychol, 24*, 349-354.

Davis, M.H. (1994). *Empathy: A social psychological approach*. Dubuque: Brown and Benchmark.

Davis, M.H., Soderlund, T., Cole, J., Gadol, E., Kute, M., & Myers, M. (2004). Cognitions associated with attempts to empathize: How do we imagine the perspective of other? *Personality and Social Psychology Bulletin, 30*, 1625-1635.

Decety, J., & Lamm, C. (2009). Empathy versus personal distress: Recent evidence from social neuroscience. . In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy*. Cambridge, MA: MIT Press.

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why People Fail to Recognize Their Own Incompetence. *Current Direction in Psychological Science, 12*(3), 83-87.

Dunning, D., Meyerowitz, J.A., & Holzberg, A.D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Joural of Personality and Social Psychology, 57*, 1082-1090.

Elliott, R., Bohart, A. C., Watson, J. C., & Greenberg, L. S. (2011). Empathy. *Psychotherapy (Chic), 48*(1), 43-49. doi: 10.1037/a0022187

Everson, H.T., & Tobias, S. (1998). The ability to estimate knowledge and performance in college: A metacognitive analysis. *Instructional Science, 26*, 65-79.

Eysenck, S. B. J., & Eysenck, H. J. (1978). Impulsiveness and venturesomeness: Their position in a dimensional system of personality description. *Psychological Reports, 43*, 1247-1255

Forrester, S., Kershaw, S., Moss, H. , & Hughes, L. (2008). Communication skills in child protection: How do social workers talk to parents? . *Child and Family Social Work, 13*, 41-51.

Gerdes, K.E., & Segal, E. (2011). Importance of Empathy for Social Work Practice: Integrating New Science. *Social Work, 56*(2), 141-148.

Gesn, P.R., & Ickes, W. (1999). The Development of Meaning Contexts for Empathic Accuracy: Channel and Sequence Effects. *Journal of Personality and Social Psychology, 77*(4), 746-761.

Gurman, A.S. (1977). The patient's perception of the therapeutic relationship. In G. A.S & R. A.M (Eds.), *Effective psychotherapy: A handbook of research* (pp. 503-543). New York: Pergamon Press.

Ickes, W., Gesn, P.R., & Graham, T. . (2000). Gender differences in empathic accuracy: Differential ability or differential motivation? *Personal Relationships, 7*, 95-109.

Ickes, W., & Tooke, W. . (1998). The observational method: Studying the interactions of minds and bodies. In S. Duck, D. F. Hay, S. E. Hobfoll, W. Ickes & B. Montgomery (Eds.), *Handbook of personal relationships: Theory, research, and interventions* (pp. 79-97). Chincheser, UK: Wiley.

Jensen, P., Weersing, R., Hoagwood, K.E., & Goldman, E. (2005). What is evidence for evidence-based treatments? A hard look at our soft underbelly. *Mental Health Services Research, 7*, 53-74.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J Pers Soc Psychol, 77*(6), 1121-1134.

Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S., & David, A. S. (2004). Measuring empathy: reliability and validity of the Empathy Quotient. *Psychol Med, 34*(5), 911-919.

Levenson, R.W., & Ruef, A.M. (1992). Empathy: A physiological substrate. *Journal of Personality and Social Psychology, 63*(2), 234-246.

Marangoni, C., Garcia, S., Ickes, W., & Teng, G. . (1995). Empatic accuracy in a clinically relevant setting. *Journal of Personality and Social Psychology, 68*(5), 854-869.

Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality, 40*(4), 525-543.

Myers, S. (2000). Empathic listening: Reports on the experience of being heard. *Journal of Humanistic Psychology, 40*, 148-173.

Seehausen, M., Kazzer, P., Bajbouj, M., & Prehn, K. (2012). Effects of empathic paraphrasing - extrinsic emotion regulation in social conflict (Publication no. 10.3389/fpsyg.2012.00482). (482).

Shamay-Tsoory, S. (2009). Empathic Processing: Its Cognitive and affective dimensions and neuroanatomical basis. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 215-232). Cambridge, MA: MIT Press.

Shortt, J.W., & Pennebaker, J.W. (1992). Talking Versus Hearing About Holocaust Experiences. *Basic and Applied Social Psychology, 13*(2), 165-179.

Wilson, F.R. , & Griswold, M.L. (1985). The effects of method and comprehensiveness of training on the reliability and validity of ratings of counselor empathy. *Measurement and Evaluation in Counseling and Development, 18*, 3-11.