Evaluating the Effectiveness of Clinical Training in an Attachment-Based Parent-Child

Intervention

By

Jordan Bate

December 2016

Submitted to the New School for Social Research of The New School in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Dissertation Committee:
Dr. Miriam Steele
Dr. Howard Steele
Dr. Jeremy Safran
Dr. Julia Ott

Abstract

The field of clinical psychology has emphasized the development of evidence-based psychotherapy interventions to promote wellbeing and prevent child abuse and neglect in young children and families. Yet, lack of attention to training clinicians is a significant barrier to delivering these promising interventions. This dissertation aims to explore the impact of training in Group Attachment Based Intervention (GABI), a parent-child psychotherapy for high-risk families. Chapter 1 reviews the existing literature on how to train clinicians and evaluate the effectiveness of psychotherapy training. Based on this review, we developed a new tool, the Applications of Clinical Training Assessment (ACTA), to evaluate the quality of training a parent-child psychotherapy. Chapter 2 reports findings from using ACTA, a video observation task, to evaluate the effectiveness of a one-day workshop and ongoing apprenticeship/supervision in GABI. Ninety-nine students/clinicians participated in a one-day training workshop, and 21 completed an apprenticeship with supervision. Participants completed the ACTA task before and after the training workshop, and at the end of their apprenticeship. Responses were coded to measure trainees' ability to identify and apply the GABI model in evaluations of clinical situations. Results report on reliability and validity. We found that the one-day workshop produced increases in recognition and application of all six core GABI principles, as well as improvement in overall quality of observations. In the subset of participants who completed the apprenticeship, there was evidence of improvement in Overall Quality of Response but improvements on specific GABI principles was more limited.

Table of Contents

Acknowledgements

List of Tables

List of Figures

List of Abbreviations and Acronyms

Interventions
CBT: Cognitive Behavioral Therapy
CT: Cognitive Therapy
IPT: Interpersonal Therapy
SE: Supportive Emotional Therapy
MI: Motivational Interviewing
PCIT: Parent-Child Interaction Therapy
CPP: Child-Parent Psychotherapy
GABI: Group Attachment-Based Intervention

Measures
ACTA: Applications of Clinical Training Assessment

ACTA Variables
RF: Reflective Functioning
EA: Emotional Attunement
AR: Affect Regulation
RET: Reticence
NUR: Nurturance
GRP: Group
OWNRF: Own RF
NONV: Nonverbal
NONJUDG: Non-Judgmental Stance
OVRLL: Overall Quality of Response

Chapter 1: Literature Review

There is currently a strong emphasis in clinical psychology on evidence-based psychotherapy, but the lack of attention to evidence based training and dissemination presents a barrier to the implementation of promising psychotherapeutic interventions (Beidas, Edmunds, Marcus et al., 2012; Kazdin, 2008; Lyon, Stirman, Kerns et al., 2011; McHugh & Barlow, 2010). Implementation science attempts to remedy this problem through the "Scientific study of methods to promote the systematic uptake of clinical research findings and other evidence-based practices into routine practice" (Beidas et al., 2006, pg. 2). Competency, leadership, and organizational factors have been identified as central implementation drivers. According to Beidas et al. (2012), "Perhaps the most daunting challenge to implementation of empirically supported treatments for youths in community settings is training clinicians to administer these complex treatments with fidelity" (pg. 3).

Manual based trainings have become a requirement for receiving funding for intervention studies, yet manuals and other popular training methods, including didactic trainings and supervision, have been adopted largely on the basis of face validity and anecdotal evidence, rather than empirical support (Beidas et al., 2012; Kazak, Hoagwood, Weisz et al., 2012; Luborsky & DeRubeis, 1984; Rounsaville, O'Malley, Foley & Weissman, 1988; Sholomskas, Syracuse-Siewert, Rousaville, Ball, Nuro & Carroll, 2005). Just like treatments should have an evidence base, training practices should be informed by evidence too.

Chapter 1 of this dissertation reviews the existing theoretical literature on how to train clinicians, as well as the extant research evaluating the effectiveness of training in psychotherapeutic interventions. It will conclude by describing a new tool, the Applications of

Clinical Training Assessment (ACTA), which was designed based on this review of the theoretical and empirical literature, to evaluate the quality of training in parent-child psychotherapies through the use of a video observation exercise.

Part I: History of Adherence and Fidelity Research

For the most part, treatment fidelity research has resided within the context of intervention outcome studies, and has been aimed at controlling the independent variable in order to gain a reliable and valid measure of treatment efficacy. In most clinical trials, adherence and competence are evaluated only after training (Morgenstern, Morgan, McCrady et al., 2001). But even then, less than half of studies reviewed by Borrelli and colleagues in 2005 actually reported whether treatment was delivered according to a manual, and less than a quarter of studies reviewed evaluated the impact of training on skill acquisition and maintenance (Borrelli, Sepinwall, Ernst et al., 2005).

Because of the way training and clinical skill have been assessed, the existing research does not say much about whether training was responsible for change, or whether clinicians were highly skilled prior to training (Morgenstern et al., 2001). Within-participant designs, measuring changes from before to after training, are necessary in order to inform clinical practice and dissemination of treatments outside the laboratory setting.

Fortunately, the field has been growing in this direction, with the bulk of the research on psychotherapy training centering on cognitive-behavioral therapy (Milne, Baker, Blackburn et al., 1999; Morgenstern et al., 2001; Sholomskas et al., 2005; Simons, Padesky, Montemarano, et al., 2010), motivational interviewing (Miller & Mount, 2001; Miller, Yahne, Moyers et al.,

2004), and short term psychodynamic therapy (Crits-Christoph, Connolly, Azarian et al., 1996; Crits-Christoph, Siqueland, Chittams et al., 1998; Crist-Christoph, Gibbons, Crits-Christoph et al., 2006; Henry, Schacht, Strupp et al., 1993; Hilsenroth, Defife, Blagys et al., 2006). Importantly, there is a noticeable gap in the area of parent-child psychotherapies, which will be addressed later in this dissertation. However, Parent-Child Interaction Therapy (PCIT), Child-Parent Psychotherapy (CPP), Triple-P for Positive Parenting, and Incredible Years are among those that have published on their training and implementation processes (Herschell, McNeil, Urquiza et al., 2010; Lyon & Budd, 2010; Cicchetti, Rogosch & Toth, 1999, 2000, 2006; Toth, Rogosch, & Cicchetti, 2006; Lieberman, Ghosh Ippen, Van Horn et al., 2005; Lieberman, Van Horn, Ghosh Ippen et al., 2006; Sanders & Kirby, 2014; Seng, Prinz, Sanders et al., 2006; Stern, Alaggia, Waston & Morgan, 2008).

Defining Training Outcomes

A significant complication in the study of training is the lack of consensus around how to define and assess training outcomes and clinical competencies, by which we mean the knowledge, skills, and attitudes needed to deliver a specific treatment (Muse & McManus, 2013). Certain aspects of treatment fidelity have received more attention than others. Three training outcomes are typically evaluated: knowledge, adherence, and competence. Knowledge, the most basic outcome, is not often the focus of research because it is not necessarily correlated with administration of the intervention in clinical practice (Herschell, Kolko, Baumann, et al., 2010). Adherence is a rating of how closely a clinician follows a manualized treatment the way it is intended, and is frequently referenced in outcome trials. Finally, competence is the level of

skill with which the treatment is delivered. Although theoretically a clinician should be both adherent and competent, studies have shown that these constructs may be independent (Crits-Cristoph et al., 1996), and most studies evaluate only adherence or competence.

In an effort to tease apart these constructs empirically, Miller and Mount (2001) compared written responses to hypothetical vignettes (a measure of knowledge) to therapist skill on a role-play assessment and an audiotaped sample of the therapist's work (measures of competence). Participants showed significant changes on both the written and role-play measures, but changes on the role-play were more modest. Most importantly, correlations at follow-up reflected modest convergence between written knowledge and skills in practice. Given the large amount of time and resources it takes to conduct simulated therapy sessions, it is promising that certain types of written evaluations converge with observer-based assessments of competency.

However, research has also suggested that adherence does not necessarily imply competence. For example, the Vanderbilt II study found that manualized training in short-term dynamic therapy successfully changed therapists' technical interventions in line with the protocol, but there were unexpected deteriorations in interpersonal and interactional aspects of therapy (Henry et al., 1993). In contrast, when supervision was included as part of the training process there were positive correlations between adherence and competence (Crits-Cristoph et al., 1996), and between competence and effectiveness of treatment (Rounsaville et al., 1988).

Developing Competency Measures

The question of what construct to measure is only part of the challenge of assessing training outcomes. A second question relates to the issue of how to measure a construct like competence. Miller and Mount (2001) used both written tests and role-plays developed specifically for their study, which are common formats for assessing training outcomes. Other studies have relied on self-report or supervisor report; however, the validity of these methods is questionable. For example, trauma-focused (TF) CBT training was found to impact therapist attitudes according to self-report, but had little impact on therapist behavior when charts were reviewed objectively (Jensen-Doss et al., 2008). In light of the mixed results studies have produced, careful selection of standardized measurement tools is imperative to studying training efficacy, because a lack of competence could limit effectiveness of treatment.

Research in Cognitive Behavioral Therapy (CBT or CT) has perhaps come the furthest in terms of developing reliable and valid measures to assess competence. Yet, in a review of 27 randomized clinical trials (RCTs) of CBT for anxiety and depression, where 25 included mention of recording therapy sessions and monitoring adherence, only seven reported using a standardized measure of competence in their assessment, and two of the studies did not report any checks on adherence or competence (Roth, Pilling & Turner, 2011).

The Cognitive Therapy Scale (CTS)/CTS-Revised (CTS-R) is one example of a measure that has been tested across multiple studies and is considered standardized, to some degree (Blackburn, James, Milne et al., 2001). Though it is only applicable to CBT training, the process of its development provides a model for other competency scales. The CTS was developed by Young and Beck in 1980 and later revised by four expert CT therapists who had experience

using it. It consists of 14 items that are rated on a 7-point Likert scale, defined by 5 levels of skill acquisition. Blackburn et al. (2001) have demonstrated that the scales have good internal reliability, and inter-rater reliability for the individual items ranged from an intraclass correlation (ICC) of .42 to an ICC of .67. Furthermore, the CTS-R has demonstrated discriminant validity by detecting an increase in competence from beginning to later stages of training in CT (Blackburn et al., 2001).

Other studies have utilized competence-rating systems that are specific to their intervention or study. For example, Karlin et al. (2013) used a competency rating scale that rated therapists' use of 15 specific techniques on a scale from 0 (poor) to 4 (Excellent), to show that training in CBT for insomnia in the U.S. Department of Veteran Affairs Healthcare System led to significant increases in clinicians' competency following training. Specifically, results showed those clinicians' scores on the last session of a six-session treatment were significantly higher than on the first session, and that clinicians' scores were significantly higher with the second patient they treated, as compared with their first patient.

Stewart et al. (2014) also used a competency rating form (the IPTRS) to code therapy session tapes in interpersonal psychotherapy (IPT) for depression, before and after a three-day workshop and six months of weekly group consultation, throughout the U.S. Department of Veteran Affairs Healthcare System. They found that therapist competence in IPT increased from their first patient to their second, and intent-to-treat analyses showed a significant reduction in overall depression, significant improvements in quality of life, and improvement in the therapeutic alliance.

Whether considered to be a standardized measure or designed for a specific study, these competence-rating scales were all developed through a similar process that included expert consensus. All involve rating a set of competencies, skills, or techniques on a scale, often from poor to excellent. These three scales also have shown discriminant validity by detecting change in competence over the course of the training process (Blackburn, James, Milne et al., 2001; Karlin, Trockel Taylor et al., 2013; Walser, Karlin, Trockel et al., 2013). The two VA Healthcare System studies also included assessments of patient outcomes alongside competence outcomes, showing empirically that improvements in competence were linked to improved patient outcomes (Karlin, Trockel Taylor et al., 2013; Stewart, Raffa, Steele et al., 2014; Walser, Karlin, Trockel et al., 2013).

The Yale Adherence Competence Scale (YACS) represents a different kind of competence measure, one that is designed as a "generic" rating system that can adapted for a variety of treatments and a wide range of studies (Carroll, Nich, Sifry, et al., 2001). It is a 55-item scale assessing both adherence to and competence in delivering general elements of psychotherapeutic intervention, as well as unique treatment-specific elements. The three "General" scales are Assessment, General Support, Goals of Treatment and the three "Treatment" scales are Clinical Management, Twelve Step Facilitation, and Cognitive-Behavioral Treatment. Similar to other competence measures, items for the YACS were derived from a review of both session videotapes and treatment manuals. Both Adherence, the degree to which the intervention was present in the session, and Quality, the amount of skill the therapist demonstrated in delivery, are rated on Likert scale. A detailed rater's manual was developed, and trained raters are required to complete a didactic seminar and rate 10 practice tapes in order to be

certified. The YACS was validated in an RCT evaluating five substance abuse psychotherapy treatments (Carroll, et al., 2001). The six scales were found to be highly reliable, with intraclass correlations (ICCs) ranging from .71 to .97. The three "Treatment" scales were negatively correlated with each other and positively correlated with the "General" scales, and could be used to discriminate between the different types of therapy. The finding that variability in the treatment-specific scales was accounted for by treatment condition, but the variability in the general scales was related to variability in therapists rather than the treatment condition evidenced this. Results also indicated that therapists might be adherent to a manual, but do so with their own individual style, and that adherence and competence were moderately positively correlated. Of note, there were stronger inter-correlations between competence ratings, compared to adherence ratings, meaning that skilled therapists may use a range of techniques, but are consistent in their level of skillfulness. The authors of the YACS highlight the potential opportunities stemming from the availability of a rating system that is not connected to one specific treatment model, but can instead be used in a wide range of studies and to assess components of many different treatments (Carroll et al., 2001). They also underscore how the development and utilization of measures like the YACS facilitate more cross-study comparisons.

Development of Competence in Psychodynamic Interventions

Findings from studies in one theoretical orientation are not necessarily applicable to training in other forms of therapy, particularly psychodynamic therapies that are typically less structured than short-term cognitive and behavioral treatments. The body of research on training in psychodynamic interventions has varied, leading some to caution that manual based training

may not be optimal for training clinicians in interpersonal and emotional aspects of treatment (Crits-Christoph, Gibbons, Crits-Christoph, Narducci, Schamberger & Gallop, 2006). For example, the Vanderbilt II Study found concerning results that training in short-term dynamic therapy increased adherence but had a negative impact on therapeutic alliance. Other studies, however, have shown that training in short term psychodynamic therapy increased psychodynamic interventions used, and did not have any deleterious effects on therapeutic alliance (Hilsenroth, et al., 2006). The authors suggest that the difference may have been the inclusion of supervision in the training process, highlighting the importance of understanding how the individual components of training contribute to therapist expertise.

Attempts have also been made at comparing the impact of training in different orientations. A comparison of therapists' adherence following training in three different types of therapy - supportive-expressive dynamic therapy (SE), cognitive-behavior therapy (CBT), and individual drug counseling - found that only in the CBT condition did training result in learning that carried over from one case to another (Crits-Christoph, et al., 1998). However, another study found that mean total competence following training was higher for those trained in SE, compared to CBT/CT (Crits-Christoph et al., 1996).

Dennhag et al. (2012) evaluated the impact of training in CT and psychodynamic therapy in terms of patient outcomes, and found that length of training had different effects in CT compared to dynamic therapy. When learning CT, longer duration of training (two years compared to one year) produced better outcomes for patients. In contrast, either one or two years of training in psychodynamic psychotherapy produced similar outcomes for patients. The authors hypothesized that the difference was related to the structure of the learning process. They

suggested that when learning a manualized treatment, like CT, more time is needed before seeing results, because integrating new techniques can be stressful and interfere with therapist competence early on, particularly ability to establish a working alliance with patients. Therefore, they recommend that the early stages of training in any modality focus on building the working alliance, before shifting the focus to learning specific techniques or skills.

These findings highlight the need to more completely understand how trainees internalize a therapeutic model, in both psychodynamic and more structured treatments. It remains unclear whether the ability to articulate principles and technically adhere to protocols means that trainees can monitor their own behavior or fully appreciate the ongoing interactional process as it actually unfolds (Henry, et al., 1993). It is also ambiguous how different components of training impact clinicians. What is evident, though, is that methods for evaluating training should be sensitive to this learning process.

Comparing Training Methods

Many studies tend to evaluate training methods as a whole, rather than the effects of each aspect of training separately (Milne, Baker, Blackburn et al., 1999; Barnfield, Mathieson, & Beaumont, 2007). Yet, different aspects of training may differentially promote knowledge, adherence, and competence over the course of the training process.

Sholomskas et al. (2005) investigated changes in both knowledge, evaluated on a 55-item multiple choice test, and skill, as measured by the Yale Adherence Competence Scale (YACS), before and after different types of training. The three training conditions compared were: 1) review of a manual only, 2) review of a manual plus a training Web site, or 3) review of the

manual plus a didactic seminar followed by supervised casework. Findings showed that there were significant increases in knowledge in all the conditions, regardless of the type of training participants received. However, the seminar-plus-supervision was superior to the manual-only condition in increasing therapists' adherence and competence in role-plays, with the Web training falling in between.

Miller et al. (2004) also compared different training methods for motivational interviewing (MI), looking at competence in role-plays, and comparing self-guided training (manual and training videos) to a workshop plus practice feedback, a workshop plus coaching sessions, and a workshop plus feedback and coaching. The four workshop groups all showed improvement immediately following the workshop, though participants showed no significant improvement with self-guided training. The addition of feedback and/or coaching also significantly improved retention.

Most studies find that supervision or "coaching" is a critical component of training that increases effectiveness, but researchers have just begun to study the supervision process and how it impacts therapists' activity in sessions. In a study of more than 400 clinicians who treated more than one thousand youth patients, Schoenwald et al. (2009) investigated supervisor adherence to a treatment specific protocol (as rated by the trainee therapist), therapist adherence in session (as rated by the patient's caregiver), and long-term post-treatment patient outcomes. Findings showed that *Focus on Adherence to Principles* in supervision sessions predicted *Therapist Adherence* in the next therapy session, and that *Adherence to Structure and Process* and *Focus on Clinician Development* in supervision predicted long-term outcomes in the population of youth treated in this study.

In a follow-up study of 16 clinicians in the Vanderbilt II study that looked at adherence before and after supervision sessions, there was an increase in adherence from the session before to the session after supervision (Anderson, Crowley, Patterson et al., 2012). In terms of supervision content, trainees' discussion of specific techniques and strategies was related to adherence in the session prior to (but not after) supervision. However, supervisors' discussion of specific techniques predicted trainees' adherence in the session following the supervision. Also of note, although supervision influenced therapist adherence to multiple techniques, it did not influence adherence on principles related to the therapeutic relationship (Anderson et al., 2012). This is consistent with other findings from the Vanderbilt II study, as well as Dennhag et al.'s (2012) comparison of training in CT and psychodynamic therapy, and implementation research on child-parent psychotherapy (CPP; Barnett et al., 2014), suggesting that the therapeutic relationship is one aspect of competence that is not being sufficiently addressed in training.

Factors Affecting the Impact of Training

Because most research on the outcomes of training comes from clinical trials it has tended to focus on the impact of rigorous in-depth trainings with highly experienced and competent therapists (Beveridge, Fowles, Masse et al., 2015). The reality in clinical practice, though, is that trainees are often novice clinicians with limited experience, and training consists of widespread distribution of manuals and/or brief didactic training (Herschell et al., 2010; Sburlati, Lyneham, Mufson et al., 2012). Without knowing who is being trained and how the training is working for them, we cannot effectively disseminate evidence-based interventions.

To determine whether Motivational Interviewing (MI) training is effective in a less skilled and less motivated population, Moyers et al. (2008) evaluated outcomes in a diverse group of Air Force behavioral health providers. They found that training was effective in immediately increasing the skill level of these clinicians, but gains had decreased by the four-month follow-up point. Training enrichments (i.e. personalized feedback and phone consultations) did not have an expected, additive effect on clinician skill level. These findings lend support to the concern that a foundation of basic clinical skills may be needed in order to achieve lasting benefits from training in new modalities, and therefore different training processes may be appropriate based on level of trainees' experience.

Other research, however, indicates that less experienced therapists may actually be more likely to change as a result of training. In a study of interpersonal psychotherapy (ITP), more experienced therapists were found to perform at a high level on their first supervised case, but no improvement in competence was found with subsequent cases; this was in comparison to younger and less experienced therapists, who demonstrated acceptable competence on their first case and showed further improvement on their second case (Rounsaville et al., 1986; Crits-Cristoph et al., 1998).

The effectiveness of training may also depend on the extent to which the treatment differs from the therapist's usual practice. Siqueland et al. (2000) found that cognitive therapists (CT) benefited more from training when they had more general CT experience, but those who had more hours of supervision in CT (considered more specific training) demonstrated less change. While experience may help trainees to learn more quickly, those who have more experience in "their own [specific] style" of CT may actually have a harder time adjusting to a new model

(Siqueland, Crits-Christoph, Barber, et al., 2000). Henry et al. (1993) reported similar findings in training in dynamic therapy, showing that more years of general therapy experience predicted more change in training, but that more supervision in a particular modality had a negative effect on training.

Finally, Richards et al. (2011) considered the ways that qualities of the therapist, besides their previous experience, impacted their engagement in training and subsequent changes in their approach as a result of training. They found that therapists who reported higher levels of psychological flexibility had a greater increase in knowledge of Acceptance and Commitment Therapy and self-reported behavior change, though they were not necessarily more engaged in the workshop exercises. These results should be interpreted with caution, since they are based on self-report data, but they indicate the potential value of rigorous research in this area.

The differences observed between more and less experienced trainees in these studies support the need for research evaluating not only the impact of training, but also what forms of training are most effective for different clinicians, depending on their educational and clinical backgrounds (Dodd et al., 2013).

Training and Dissemination of Parent-Child Interventions

The gains made in implementation science have been significant, however, there are still relatively few studies in training and dissemination of parent-child interventions. For example, a 2014 review of treatment fidelity strategies used in evidence-based parent training programs found that 83% of publications described how provider skills were maintained over time through supervision; but only 48% measured provider skill acquisition post-training, 57% reported

standardized training for providers, and only 65% provided a description of how providers were trained (Garbacz, Brown, Spee et al., 2014). Furthermore, their review only assessed parenting interventions, and did not include parent-child interventions or child-only treatments.

Therefore, this author conducted a review of six evidence-based parent-child interventions (Child-Parent Psychotherapy, Attachment and Biobehavioral Catchup, Minding the Baby, UCLA Family Development Project, Circle of Security, and Parent-Child Interaction Therapy), assessing the degree to which they have reported on training methods and effectiveness. The majority of these interventions report using a combination of manuals, workshops and supervision in training, and rely on supervision and adherence checklists to assess fidelity (Lieberman et al., 2006; Cicchetti et al., 2000; Toth et al., 2006; Heinicke et al., 1999, 2000, 2006; Hoffman, Marvin, Cooper et al., 2006; Slade, Grienenberger, Bernbarch et al., 2005; Werba, Eyberg, Boggs, et al., 2006). Some research has also been published on implementation in community settings, but only one of these programs, PCIT, has published an empirical study on training effectiveness.

Herschell and colleagues (2009) assessed whether community practitioners could acquire and master PCIT knowledge and skills from reading a treatment manual, and if there were differences between those who received experiential training versus traditional didactic training. They found that reading a treatment manual was associated with increases in knowledge and skill; but none of the trainees achieved "mastery" in either, from reading the manual alone. Both experiential and didactic trainings were found to significantly increase knowledge and skill, but there were not significant differences depending on the type of training, and a relatively low proportion of trainees obtained mastery. The study also looked at whether particular therapist

characteristics were associated with training success. Degree type (e.g. counseling, psychology, social work), but not theoretical orientation (e.g. cognitive behavioral, family systems, psychodynamic), was associated with mastery of the Child-Directed Interaction component of the intervention.

Lyon and Budd (2010) also published a paper on the implementation of PCIT in a community mental health clinic, which found that therapists in the community achieved 91% adherence on the PCIT fidelity checklist; this is lower than in clinical trials, but considered reasonable in more complex cases (Lyon & Budd, 2010). Importantly, though, they found that only four of the 14 families in this study completed treatment, highlighting the need for training to address barriers to treatment completion. Consistent with this, a later evaluation of trainees' perspectives on PCIT training indicated that trainees also cited a lack of training in how to motivate parents for treatment as a weakness of the training (Christian, Niec, Acevedo-Polokovick et al., 2014).

A case study on the dissemination of PCIT throughout the state of Delaware was published in 2015, which showed that ongoing consultation and consistent onsite presence of a trainer improved long-term implementation. However, only a quarter of those that participated in an initial training completed the full PCIT training (Beveridge et al, 2015). This does not necessarily mean that they did not implement PCIT with fidelity, but it indicates that there may be barriers to implementation. Beyond factors associated with the therapists, there are organizational factors that possibly interfered, including large caseloads and limited ability to see patients weekly, lack of priority for PCIT referrals, lack of young child referrals, inadequate technical support and limited time to participate in ongoing training and videotaping required. In

addition, two papers have reported on training in adaptations of PCIT to a home-based or office-based therapy (Lanier, Kohl, Benz et al., 2011) and for the prevention of child maltreatment (Thomas & Zimmer-Gemback, 2012). These studies primarily addressed issues around attrition and patient outcomes, and did not actually assess outcomes of training. But what this research tells us is that not only do we need to assess training, but we also need to use the findings to improve training and address barriers to successful implementation.

Similar studies have been done looking at the transportability of Child-Parent Psychotherapy (CPP) in other settings, including rural public health agencies (Barnett, Rosenberg, Rosenberg et al., 2014) and in Israel (David & Schiff 2015). Though results do not explicitly provide information on the outcome of training, feedback about barriers to implementation and mechanisms of action was derived from qualitative questions about implementation, session records, and patient outcome records. Responses indicated that a lack of clinician expertise in treating very young children, and poor therapeutic alliance, were barriers to effective implementation of CPP. On the session records, the most commonly endorsed mechanism of change was translating the child's behavior to the caregiver, promoting attachment in session, processing past trauma, providing reflective developmental guidance, and sharing observations of child-parent interactions.

David et al. (2015) used a qualitative interview and focus groups to understand the experiences of trainees implementing CPP in Israel, and found that desire to use CPP was based largely on the quality of the course instructors. Additionally, participants appreciated that beyond theory, there was a heavy emphasis in training on clinical work. Following training, almost all

participants reported they were using CPP-informed treatment, even if they did not fulfill all the criteria for implementing the intervention.

Attachment and Biobehavioral Catchup (ABC) is a parent-child intervention that has a growing evidence base, including publications on its implementation and transportability, and is moving toward empirical studies of training (Dozier, Peloso, Lindheim et al., 2006). In both laboratory and community settings ABC has been found to be effective in decreasing disorganized attachment, increasing secure attachment and normalizing cortisol levels among children (Bernard, Dozier, Bick et al., 2012, 2015). A link between parent-child outcomes and in-the-moment feedback from clinicians has also been demonstrated empirically (Caron, Weston-Lee, Haggerty et al., 2016; Meade, Dozier & Bernard, 2014). Based on this finding, a coding system for rating parent coach (i.e. clinician) comments throughout a session was developed in order to track application of this technique. And a single-subject design showed that training a clinician to code her own sessions led to improvement in the frequency of in-the-moment feedback with patients. This training model must still be evaluated in a larger sample, but the innovation represents a shift in the field towards thinking about how research on training can actually inform and improve training methods.

Finally, although parent-training programs do not include the child in sessions, they tend to be used extensively and much can be learned from research on their implementation. In Triple P for Positive Parenting, standardized training materials have been developed and are managed by Triple P International, an independent organization dedicated to all aspects of the training and dissemination process worldwide, from initial training, to post-training support, and follow-up technical assistance. According to the authors, Triple P International also monitors the quality of

implementation (Sanders & Kirby, 2014). Seng et al. (2006) found that trainees were highly satisfied with training in Triple P, reported few barriers to implementation, and the majority of those who completed training to a minimal or acceptable level of proficiency (67.6%) went on to deliver the program in the community. Furthermore, more than 90% of those delivering Triple P six months following training had completed the training and achieved an acceptable level of competence. Of note, however, these findings were based on therapist and supervisor self-report, which may not be a reliable measure of implementation (Jenson-Doss, 2008).

A description of the implementation of The Incredible Years Parenting Program in community-based early learning centers included details of the training process, quantitative evaluation of adherence following training, and qualitative research on the barriers to and facilitators of adherence (Stern, Alaggia, Watson & Morgan, 2008). Prior to treatment delivery, group leaders participated in a three-day training with a certified Incredible Years trainer and during the implementation they received supervision from the Manager of Early Intervention Services at their agency. This format of training resulted in high adherence rates, which was facilitated by supervisors' emphasis on adherence and empirical support for the approach, as well as their responsiveness, support, problem solving, and individualized approaches. Perhaps more valuable, though, are the findings related to barriers to adherence. For example, "desire to keep parents involved," and concerns about the relevance of certain techniques to specific groups of parents, resulted in more limited use of specific techniques indicated in the protocol, like role-plays, videotape modeling, and buddy calls. Additionally, concern or confusion about time-out procedures not being appropriate for specific groups resulted in the more limited use of that

component of the intervention. Also, time constraints in large groups and English as a second language (ESL) had a negative effect on adherence.

Taking Into Account the Learning Process

The research on both adult and child therapies illustrates that implementation is a complex process that requires training clinicians in both technical skill, as well as understanding the population, and forming a therapeutic alliance to engage patients in treatment. More extensive training tends to be superior to brief trainings (Roth et al., 2011), and workshops and manuals do not tend to be sufficient in improving therapists' skills or patients' outcomes (Rakovshik & McManus, 2010). Multi-component training programs that include a manual, workshop and supervision, have generally been found to be most consistent in producing changes in therapist skills, across orientations (Herschell et al., 2010; Beveridge et al., 2015; Rakovshik & McManus, 2010). Studies have concluded that trainees benefit from the inclusion of both theoretical instruction as well as experiential and interactive training, and that monitoring adherence and providing feedback may be a necessity in the early stages of training (Rakovshik & McManus, 2010).

Measurement of training outcomes needs to be based on this understanding of the clinical learning process. The acquisition of specific technical skills may take place quickly, but the ability to conceptualize cases, identify problems, and make clinical decisions about treatment takes more time to develop (Holloway & Neufeldt, 1995). Lyon et al. (2011) describe an initial survivor phase in which the therapists' existing knowledge is disrupted, followed by a state in which knowledge and skills are successfully applied and adapted.

Training outcome assessments that can evaluate both short-term outcomes (reactions, attitudes, knowledge and skills) and long-term outcomes (use of new skills, changes in client outcomes) are much needed (Becker et al., 2011). However, existing methods tend to focus on either knowledge or skill, but not both, and therefore, have limitations in their utility for evaluating multi-component trainings. For example, tests can provide some evidence of knowledge gains from workshops or manuals, but they do not assess application or skill. In contrast, role-plays and review of session audio/video are valuable for assessing developments in competence over the course of training, but it may take time for those improvements to show. Those measures are also complicated to collect and therefore difficult to use in the community on a large scale.

A thorough review of the literature on clinical training in both adult and parent-child therapies illustrates the progress that has been made to expand psychotherapy research beyond outcome studies, and to address questions about what the mechanisms of action are and how we can ensure trainees are learning and applying the skills and techniques that trainings purport to teach. The following section will describe our efforts to develop a new assessment tool for evaluating outcomes of training that is informed by both theory and empirical research on clinical training.

Part II: Developing a New Method For Evaluating Training in Parent-Child Treatment

Inquiring about how to improve the quality of training so as to maintain the effectiveness and fidelity of interventions when they are disseminated is a necessary next step in continuing to improve treatment for children and families. Research on training clinicians in parent-child

therapies is only in its infancy, and we still know relatively little about the mechanism of actions and psychotherapy process. In order for progress to continue, we need standardized tools for evaluating multiple aspects of clinical knowledge and competence. These assessments should have relevance to the whole training process, while being efficient to employ in a community setting, and appropriate for evaluating principle-based trainings rather than those that teach structured techniques. The Applications of Clinical Training Assessment (ACTA) was, designed to fill this gap.

 ACTA was created originally as part of a research program on the effectiveness of a parent-child intervention, to assess the impact of training on clinicians' understanding and application of core principles of attachment-based, dyadic psychotherapy. It is a promising tool that can be applied to other interventions that share a dyadic frame and theoretical basis. It also provides a model for the development of similar assessments that could be used to monitor training outcomes in other modalities.

Identifying Therapeutic Action in an Attachment-Based Intervention

The Applications of Clinical Training Assessment (ACTA) was developed in the context of a pragmatic clinical trial of Group Attachment-Based Intervention (GABI), which is a psychotherapy designed to meet the needs of socially isolated families who are referred because of concerns about risk for child maltreatment. This risk is often related to parents' own histories of multiple adverse childhood experiences that impact parenting, as well as the stress of ongoing exposure to poverty, and domestic and neighborhood violence (Steele, Murphy & Steele, 2010).

The underlying theoretical context for GABI is informed by attachment, psychodynamic,

social emotional, and developmental theories and research, and studies of trauma. Vulnerability is understood clinically in terms of attachment theory and contextually with regard to families' histories of adverse childhood experiences (ACEs) (Felitti, Vincent, Anda, et al., 1998). As detailed in Murphy, Steele, Bate, et al. (2015), the aim is to ameliorate the impact of early adversity with interventions that promote secure attachment relationships.

Pilot studies suggest that GABI promotes positive changes in the parent-child relationship (Steele et al., 2010), and a pragmatic clinical trial funded by the U.S. Department of Health and Human Services, Health Resources and Services Administration is in the final stages. Given the potential promise of GABI to affect change in the parent-child relationship and the long-term plans to disseminate the intervention to other sites, it is critical that we identify the underlying mechanisms of action, and implement a replicable training process that will allow other clinicians and sites to deliver the GABI model with similar effectiveness.

GABI Competencies. To develop and assess the effectiveness of training, one must have an idea about what makes a therapeutic intervention effective, and what competencies (i.e. knowledge, skills and attitudes) a clinician needs in order to deliver the intervention with the appropriate quality. Within the GABI research program there has been an ongoing effort to identify the critical competencies, following a process similar to the development of competency measures in CBT, described earlier in this dissertation (Sburlati et al., 2012; Vivian, Hershenberg, Teachman et al., 2012).

A group of GABI experts that included this author (JB), as well as Miriam Steele, PhD and Howard Steele, PhD from the Center for Attachment Research, and Anne Murphy PhD, who created GABI at Albert Einstein College of Medicine, reviewed more than 100 hours of

videotape from GABI sessions, with an eye for qualitatively picking out moments where clinical technique appeared to produce a shift in the parent-child dyad. As a result, a set of competencies emerged, which represent principles that are both targets of the intervention and techniques clinicians use in order to facilitate change in the parent-child relationship. These are Reflective Functioning, Emotional Attunement, Affect Regulation, Reticence, Intergenerational Patterns, Nurturance, and Group Context. Taken together, these competencies reinforce the clinician's role as a sensitive, supportive and nurturing figure that serves to 'hold' or 'contain' the families throughout this difficult work, eventually enabling the parents to provide this holding environment for their children. For training purposes, the core competencies are summarized in the acronym, REARING.

Training in GABI

Though manuals teaching structured treatments and prescriptive techniques have become the status quo, criticisms of these approaches have emerged, emphasizing the benefits of principle-based trainings that can be flexibly employed, and integration across orientations and between research and practice (Skovholdt & Starkey, 2010; Chorpita & Daleiden, 2009). As Abramowitz theorized, "one does not need to be totally bound to a treatment manual to be an effective clinician, especially if he or she understands how to apply the experimentally established principles of behavior (including the behavior of cognition) to clinical problems" (Abramowitz, 2006, pg. 164). Adherence research has also corroborated this idea that "flexibility within fidelity" may be the optimal outcome of training (Kendall & Beidas, 2007).

In line with these recommendations, the training program in GABI is designed to deepen clinicians' understandings of the seven REARING principles and how they relate to clinical practice, from the perspectives of attachment theory, developmental research, and trauma research. Descriptions of specific techniques are included, often in the form of video examples and vignettes, to provide clinical trainees with a sense of how an attachment-based approach looks in practice. But, the training is designed to allow for flexibility and creativity within the general overarching principles.

GABI training currently consists of a one-day didactic workshop and introduction to the treatment manual, followed by approximately six-to-eight months of training in an apprenticeship model where trainees work alongside the lead clinician and participate in reflective supervision.

Specific goals of the one-day workshop and manual are to 1) introduce the core REARING concepts, 2) establish a framework for thinking about assessment, pathology and change, and 3) enhance and facilitate the ability to observe clinical material in relation to those concepts (Moras, 1993). These goals are directed at creating a foundation that will allow trainees to maximally benefit from the apprentice model, where training focuses on incorporating these concepts into clinical practice.

While working in the apprenticeship alongside the lead clinician for approximately six to eight months, trainees also participate in a reflective supervision group. Reflective supervision is a specific type of supervision in which the lead clinician practices and encourages the paying of attention to (a) one's own inner experience, (b) the experience of the infant or child, and (c) the experience of the parent (Eggbeer, Shamoon-Shanuck & Clark, 2010; Steele et al., 2010). As we

move toward dissemination of GABI, it is critical that we not repeat the earlier mistakes of assuming that these training methods are effective, and instead try to empirically demonstrate and understand the effects of each component of this model.

Development of a Novel Paradigm

As illustrated in the first half of this chapter, popular methods for evaluating training outcomes include tests, self-report and supervisor report measures, role-plays and review of session audio/video; however, each of these methods has shortcomings. Criteria for selecting a method for evaluating the impact of GABI training on clinicians included the following:

1. The measure would have utility for assessing both short-term outcomes, such as knowledge, and long-term outcomes, including use of new skills (Becker et al., 2011).

2. It would have ecological validity, in that the task itself should require using skills that are relevant to the practice of psychotherapy.

3. It would be easy and efficient to employ, so that it can continue to be used when dissemination is brought to scale.

The Centrality of Observation

Active observation is a critical part of clinical work and a foundation for the implementation of interventions (Sternberg, 2005; Bick, 1976). This is particularly true for parent-child psychotherapies that require the therapist to hold both parent and child, past and present in mind. Honing observational skills that will guide clinical thinking, conceptualization

about the patient, and ultimately decisions about how to intervene, is therefore a central part of

training in GABI (Freud, 1953).

The significance of observation in clinical work originated with developmental

psychologists, most significantly Anne Freud and Esther Bick. Bick introduced infant

observation as a component of psychoanalytic training at the Tavistock Clinic in 1948. She

considered it an important task for helping students better understand the experience of their

child patients and their non-verbal behavior, as well as the infant's development, and

development of the parent-child relationship (Bick, 1964).

Bick also emphasized the benefits of taking an observational stance in the face of intense

emotional impact. "Lacking, as he does, a clear map charting the route and a definite goal to

focus and funnel attention, the analyst is exposed to uncertainty, confusion, anxiety when

bombarded at close quarters by the emotional experience of another person as a mother is

bombarded by the emotional state of the infant" (Harris & Bick, 1976, page 119). Observers

learn to resist judgment and criticism, especially about the parent's behaviors, and cope with

anxieties that may be brought up. The observer has to find a balance between getting close

enough that they can see the details but remaining distanced enough so as not to impose their

own countertransference. This critical clinical skill is also similar to the reticent, reflective,

"watch, wait, and wonder" approach that GABI and other parent-child interventions encourage

parents to take with regard to their children (Cohen, Muir, Lojkasek, et al., 1999). By adopting

this observer stance, clinicians also model for parents a new, often more empathic, way of

thinking about their children's behavior.

Janine Sternberg (2005) empirically showed that infant observation training promotes

skills and capacities considered central to psychoanalytic practice. Though not specifically studying whether the training was successful, she used both a top down and bottom up approach to identify competencies developed in infant observation and those considered essential for psychoanalysis, and then explored the overlap in competencies between these two areas. Her research used qualitative measures (group interviews) analyzed with grounded theory, and showed evidence of considerable development and deepening along certain dimensions of learning considered important to psychoanalysis among those training in infant observation (Sternberg, 2005).

Theoretically, the content of clinicians' observations provides information about how they apply knowledge of the framework in their practice. As Anna Freud wrote, "the material which presents itself is seen and assessed not by an instrument, nor by a blank and therefore unprejudiced mind but on the basis of pre-existent knowledge, preformed ideas and personal attitudes (though these should be conscious in the case of the analyzed observer)" (Anna Freud, 1951, pg. 20). As such, therapists' observations may prove useful in tracking the development of clinical understanding and skill across different phases of training, and provide a snapshot of both the therapist's skills as well as their conceptualization, reflecting their underlying framework.

In GABI, observational skills are sharpened through the extensive use of videos from GABI sessions, in both the one-day workshop and reflective supervision groups. Video is particularly useful in parent-child work because it can capture the important details of interactions, such as movement, rhythm, and facial expression, in a way that other methods cannot (Woodhead, Bland, Baradon, 2006). There is also empirical support for the use of video

as a training resource. Research has found that videotape modeling used in training is equally effective to experiential training methods, such as role-plays (Herschell, 2009).

The ACTA paradigm was therefore designed to use trainees' observations of real-life clinical situations captured on video, as the lens through which to evaluate their understanding and application of the GABI treatment model in a more in-depth way (beyond straight recognition) and in a relatively naturalistic application. Myles and Milne (2004) have used a Video Assessment Task to study the effectiveness of training in CBT for adult patients, and found that clinicians improved significantly in their knowledge of and abilities to identify symptoms and name appropriate CBT strategies, as a result of training. Building on their methodology, ACTA aims to capture some of the nuanced changes in clinical approach and skill that happen earlier on in training, as well as the further development after months of practice.

Applications of Clinical Training Assessment (ACTA).

The ACTA task utilizes 3-minute long video clips edited from footage of GABI from real-life sessions delivered at Albert Einstein College of Medicine. This use of videotape allows for standardization, while aiming to maintain ecological validity and reflect the complexity of this clinical work, by using clips from the intervention as it is actually delivered in practice with patients (Muse & McManus, 2013). After viewing each clip, trainees are asked to respond in writing to three open-ended questions about their observations:

1. What do you see happening in this video? What struck you about this situation?

2. Imagine that you were the clinician in this situation, and were bringing it to supervision for discussion. What would you talk about?

3. Did you see any therapeutic interventions being used? Please describe.

A group of GABI experts that included this author, Dr. Anne Murphy, Dr. Miriam Steele, and graduate students who were involved in identifying the competencies and writing the manual, selected the videos. Videos were chosen based on their coverage of the full range of GABI concepts, with the exception of Intergenerational Transmission, which tends to be apparent in the parent-group part of the intervention only.

Scoring. A coding system was developed by this author in consultation with Dr. Miriam Steele and Dr. Anne Murphy, to assess application of the GABI model to observations and evaluations of these clinical situations. Barber et al. (2007) recommend that training address prescribed (or unique and essential) elements of the treatment and nonspecific elements (either essential or nonessential) that are common to several different therapies. The ACTA paradigm is designed to evaluate both.

With the ACTA coding system, trainees' responses are assigned codes on a Likert-scale from 1 to 5 for their understanding and application of each REARING concept (*Reflective Functioning, Emotional Attunement, Affect Regulation, Reticence, Nurturance, and Group Context*). Intergenerational Patterns was excluded because it tends to arise in the parent group only and is typically overt; therefore identification of this concept may not accurately reflect competencies in observation. On the ACTA scale, a score of 1 indicates no mention or incorrect understanding/application of the concept, 3 signifies basic or probable understanding of the concept, and 5 reflects well developed understanding, internalization of and correct application of the concept.

Responses are also coded on the same Likert-scale for three non-specific aspects of observation: the participant's *Own Use of Reflective Functioning* (separate from ability to identify Reflective Functioning in the video clips), *Attention to Nonverbal Information*, and *Non-Judgmental Stance*.

Finally, responses are assigned a code for *Overall Quality of Response*, based on the number and quality of GABI concepts referenced, combined with demonstration of the other non-specific skills. A score of 1 is given to responses that overall show little or no knowledge of the GABI concepts and lack thoughtfulness and observational skills. A score of 3 is given to responses that overall indicate a basic understanding of some GABI concepts, but not others, and show some thoughtfulness but moderate quality of observational skills. A score of 5 is given to those responses that demonstrate recognition and understanding of many GABI concepts, as well as clinical and observational skills in other non-specific areas. A sample of the coding sheet is included in Appendix 1.

Conclusion

Despite research demonstrating the negative sequelae of adverse childhood experiences (ACEs) (Felitti et al, 1998; Murphy, Steele, Dube et al., 2013) and the intergenerational transmission of child abuse and neglect (Berlin et al., 2011; Fonagy, Steele, Steele et al., 1991; Murphy et al., 2014), there continues to be a disparity between what we know about how to treat abuse and neglect and the implementation of interventions to address the problem within the community. In order to improve treatment for children and families, research must go beyond traditional efficacy studies and move towards implementation science.

Empirical evaluation of the training process should be incorporated into the larger context of therapy outcome research. For this to happen, the field needs a broader range of reliable and valid methods for assessing the acquisition of knowledge and skills required to effectively deliver an intervention. Despite the gains that have been made, there remains a lack of assessment tools that have relevance to the whole training process, while also being efficient to employ in a community setting, and appropriate for evaluating applications of integrative, principle-based trainings rather than those that teaching structured techniques. Such tools are especially needed in the realm of child and dyadic therapies, where we know less about the mechanism of actions and process of psychotherapy as compared to adult treatments.

The Applications of Clinical Training Assessment (ACTA) was developed in an attempt to fill this gap. Families that are considered at-risk of child abuse and neglect present myriad needs and challenges, including trauma, poverty, social isolation, and discrimination. Trainings, like the one in Group Attachment-Based Intervention (GABI), that are focused on broad principles of treatment that can serve as a foundation and be flexibly applied as needed, rather than emphasizing specific protocols, have been recommended to treat such complex problems (Chorpita & Daleiden, 2009). GABI training aims not only to prepare clinicians to deliver a specific intervention, but also to increase their understanding of core principles thought to facilitate the therapeutic action, and promote an understanding of what a trauma-informed intervention and center encompass. In Chapter 2 of this dissertation and in future studies, we aim to demonstrate the validity of ACTA as a tool for evaluating changes in understanding and application of the GABI model, over the course of the training process for clinicians. However, we also hope that this ongoing research will go beyond supplying valuable information about

training in GABI, and test a new approach to the evaluation of clinical training that may be adapted to support the dissemination of other interventions developed for the prevention of child maltreatment.

Chapter II: Empirical Article

Introduction

There is a strong emphasis on the utilization of evidence based psychotherapy treatments for children and families, but lack of attention to evidence based training and dissemination presents a barrier to the implementation of promising psychotherapeutic interventions (Beidas et al., 2012; Kazdin, 2008; Lyon et al., 2011; McHugh and Barlow, 2010). Manual based trainings are a common prerequisite for receiving funding for clinical trials, yet manuals and other popular training methods, such as didactic workshops and supervision, have been adopted largely on the basis of face validity and anecdotal evidence, rather than empirical support (Beidas et al., 2012; Kazak et al., 2012; Luborsky & DeRubeis, 1984; Rounsaville, et al., 1988; Sholomskas et al., 2005). Just like treatments should have an evidence base, training practices should be informed by evidence too. Once a therapy is shown to be effective, the mechanisms of action must be identified, and training methods focused on these principles should be critically evaluated in terms of whether trainees have learned the skills and techniques that trainings purport to teach.

Research on Training in Parent-Child and Parenting Interventions

A few of the well-established parent-child or parenting interventions have published results of dissemination and implementation efforts, including Child Parent Psychotherapy, Parent Child Interaction Therapy, Attachment and Biobehavioral Catchup, The Incredible Years, and Triple P for Positive Parenting. The majority report using a combination of manuals, workshops, supervision, and adherence checklists in their training processes (Herschell, McNeil, Urquiza et al., 2009; Lyon & Budd, 2010; Cicchetti, Rogosch & Toth, 1999, 2000, 2006; Toth,

Rogosch, & Cicchetti, 2006; Lieberman, Ghosh Ippen, Van Horn et al., 2005; Lieberman, Van

Horn, Ghosh Ippen et al., 2006; Sanders & Kirby, 2014; Seng, Prinz, Sanders et al., 2006; Stern,

Alaggia, Waston & Morgan, 2008).

Only PCIT has published research specifically relating to the effectiveness of their

training methods. Herschell and colleagues (2009) assessed whether community practitioners

could acquire and master PCIT knowledge and skills from reading a treatment manual, and if

there were differences between those who received experiential training versus traditional

didactic training. They found that reading a treatment manual was associated with increases in

knowledge and skill, but none of the trainees achieved "mastery" in either from reading the

manual alone. Both experiential and didactic trainings were found to significantly increase

knowledge and skill, but there were not significant differences depending on the type of training,

and a relatively low proportion of trainees obtained mastery. The study also looked at whether

particular therapist characteristics were associated with training success. Degree type (e.g.

clinical, counseling, education), but not theoretical orientation (e.g. psychodynamic, behavioral,

family systems), was associated with mastery of the Child-Directed Interaction component of the

intervention.

When PCIT was implemented in a community mental health clinic, findings showed that

therapists achieved 91% adherence on the PCIT fidelity checklist, which is lower than in clinical

trials, but considered reasonable in more complex cases (Lyon & Budd, 2010). Importantly,

though, they found that only four of the 14 families in this study completed treatment,

highlighting the need for training to address barriers to treatment completion. A later evaluation

of trainees' perspectives on PCIT training further indicated that trainees perceived a lack of

training in how to motivate parents for treatment as a weakness of the training (Christian, Niec, Acevedo-Polokovick, et al., 2014).

Multiple randomized clinical trials have reported on the efficacy of Child-Parent Psychotherapy (Cicchetti et al., 1999, 2000, 2006; Toth et al., 2006; Lieberman et al., 2005, 2006) and its transportability to other settings has also been evaluated, in rural public health agencies and other cultures (Israel), (Barnett et al., 2014; David et al., 2015). Session records, clinician report, and patient outcome records collected in these studies provide valuable feedback about hypothesized mechanisms of action, and barriers to implementation. The most commonly endorsed mechanisms of change in session records were translating the child's behavior to the caregiver, promoting attachment in session, processing past trauma, providing reflective developmental guidance, and sharing observations of child-parent interactions (Barnett et al., 2014). Similar to the findings in PCIT, community clinicians indicated that a lack of expertise in treating very young children and poor therapeutic alliance were barriers to effective implementation of CPP (Barnett et al., 2014). Trainees in Israel who reported on how they experienced the training process highlighted that they considered quality of instructors and a heavy emphasis on clinical work to be highly impactful on their motivation to use CPP in their practice (David et al., 2015).

The effectiveness of Attachment and Biobehavioral Catchup has been demonstrated both in laboratory and community settings, and studies have established a link between parent-child outcomes and clinicians' in-the-moment feedback, one of the hypothesized active ingredients of treatment (Bernard et al., 2012; Bernard, Dozier, Bick, & Gordon, 2015; Caron, Weston-Lee, Haggerty et al., 2016; Meade, Dozier & Bernard, 2014). A coding system for rating parent coach

comments throughout a session has been developed to assess clinician competence. In a single-subject design, a clinician was trained to code her own sessions and results found that this training led to improvement in the frequency of in-the-moment feedback with patients. Although this training model remains to be evaluated in a larger sample, the innovation represents a step towards thinking about how research on training can actually inform and improve the training.

Group Attachment-Based Intervention (GABI)

While the interventions described have garnered empirical support and demonstrated transportability to community settings, the fact remains that child maltreatment continues to be an area of great concern and unmet need. We therefore need a wide range of treatments that can reach families from a variety of backgrounds. This significant demand led to the development of Group Attachment-Based Intervention by Dr. Anne Murphy at Albert Einstein College of Medicine's Rose F. Kennedy Center in Bronx, NY (GABI; Steele, Murphy & Steele, 2010). GABI is specifically designed to meet the needs of parents with zero to three year-old children who are considered at risk for child maltreatment. The families served are struggling to meet the emotional needs of their young children in large part due to the parents' own histories of neglect and abuse, and ongoing exposure to toxic stresses, such as poverty, domestic and community violence, and homelessness. They live in an urban community that is paradoxically crowded and cramped but also socially isolating. To address this problem, clinical principles from dyadic interventions like CPP and attachment and psychodynamic theories were incorporated within a group context. As family history, environmental contexts, and social isolation may negatively affect parents' abilities to form secure attachments with their children, the primary aim of GABI

is to promote secure attachment relationships, which help to disrupt the intergenerational impact of trauma and protect these young children from their own adverse experiences.

GABI meets three times weekly for two hours. Sessions are uniquely designed to target each component of the attachment system – parent, child, and the relationship. The group begins with a parent-child session lasting 45 minutes. In contrast to a traditional group setting, in GABI there are approximately as many clinicians in the group as there are families, and each family has the support and attention of a clinician whose primary objective is to nurture the parent-child interactions.

Parents and children then separate, for a parent-group and a child-group, held simultaneously. In the child group, children have the opportunity to negotiate interactions with their peers, as well as feel the presence of adults who are predictable and make strong attempts to understand their thoughts and feelings. The parent group is a time for parents to receive and ultimately give one another social support, and make links between their past and their current relationship with their child. Parents and children are then reunited, and a goodbye song helps consolidate each family as they leave the cohesion of the group

In collaboration with Miriam and Howard Steele at the Center for Attachment Research, at the New School for Social Research, a research initiative was developed to study the impact of GABI on multiple levels, including child development, parenting stress and mental health, and the parent-child attachment relationship. Based on promising pilot results (Steele, Murphy & Steele, 2010), a pragmatic clinical trial funded by the U.S. Department of Health and Human Services, Health Resources and Services Administration was funded and is nearing completion. Given the potential for GABI to affect change in the parent-child relationship, a critical next step

is to identify the underlying mechanisms of action, and implement a replicable training process that will allow other clinicians and sites to deliver the GABI model with similar effectiveness.

GABI Training. Within the GABI research program, there has been an ongoing effort to identify the therapeutic action and outline principles of the treatment. A group of experts that included this author (JB), Miriam Steele, Howard Steele and Anne Murphy, combined our knowledge and clinical expertise with review of more than 100 hours of videotape from GABI sessions, with an eye for qualitatively picking out moments where clinical technique appeared to produce a shift in the parent-child dyad. As a result of these efforts, we identified a set of core concepts that represent both targets of the intervention and techniques clinicians use in order to facilitate change in the parent-child relationship. These are: reflective functioning, emotional attunement, affect regulation, reticence, intergenerational patterns, nurturance, and group context. Taken together, these principles reinforce the clinician's role as a sensitive, supportive and nurturing figure that serves to "hold" or "contain" the families throughout this difficult work, eventually enabling the parents to provide this holding environment for their children. For training purposes, the core concepts are summarized in the acronym, REARING.

Although manuals teaching structured treatments and prescriptive techniques have become the status quo, criticisms of these approaches have emerged, emphasizing the need for integration across orientations as well as between research and practice, and the benefits of principle-based trainings that can be flexibly employed (Skovholdt & Starkey, 2010; Chorpita & Daleiden, 2009). As Abramowitz theorized, "one does not need to be totally bound to a treatment manual to be an effective clinician, especially if he or she understands how to apply the experimentally established principles of behavior (including the behavior of cognition) to clinical

problems" (Abramowitz, 2006, pg. 164). In line with these recommendations, the training program in GABI is designed to deepen clinicians' understandings of these seven REARING principles and how they relate to clinical practice, from the perspectives of attachment theory, developmental research, and trauma research. While descriptions of specific techniques are included in the form of video examples and vignettes, to provide clinical trainees with a sense of how an attachment-based approach looks in practice, the training is designed to allow for flexibility and creativity within general overarching principles. Training currently consists of a one-day didactic workshop and introduction to the treatment manual, followed by approximately six-to-eight months of training in an apprenticeship model where trainees work alongside the lead clinician and participate in reflective supervision.

Specific goals of the one-day workshop and manual are to 1) introduce the core concepts of GABI, 2) establish a framework for thinking about assessment, pathology and change, and 3) enhance and facilitate the ability to observe and describe clinical material in relation to that framework (Moras, 1993). These goals are directed at creating a foundation that will allow trainees to maximally benefit from the apprentice model, where training focuses on incorporating these concepts into clinical practice.

While working in the apprenticeship model alongside the lead clinician, trainees also participate in a group reflective supervision. Reflective supervision is a specific type of supervision in which the lead clinician practices and encourages the paying of attention to (a) one's own inner experience, (b) the experience of the infant or child, and (c) the experience of the parent (Eggbeer, Shamoon-Shanuck & Clark, 2010; Steele, Murphy & Steele, 2010).

As we move toward dissemination of GABI, it is critical that we not repeat the earlier mistakes of assuming that these training methods are effective, and instead try to empirically demonstrate and understand the effects of each component of this model. Towards this end, the research program is going beyond a pragmatic clinical trial, and incorporating evaluation of the training process.

Assessing Training

Multi-component training programs, like the one employed in GABI, have generally been found to be most consistent in producing changes in therapist skills across orientations (Herschell et al., 2010). However, the common methods for studying training, such as self or supervisor report, multiple choice tests, role-plays, and audio/video of sessions, tend to focus on either knowledge or skill, but not both (Sholomskas et al., 2005; Simons et al., 2010; Morgenstern et al., 2001; Milne et al., 1999; Miller et al., 1999; Miller et al., 2001; Miller et al., 2004; Henry et al., 1993; Crits-Christoph et al., 1996; Crits-Christoph et al., 1998; Crits-Christoph et al. 2006; Hilsenroth et al. 2006). Although competence ratings tend to be perceived as the gold standard, collecting audio/video data can be time consuming and complicated. Therefore, these measures have limitations in their utility for evaluating the full training process. Moreover, there are almost no standardized tools for evaluating training in parent-child treatments. So we designed a new assessment, the Applications of Clinical Training Assessment, that would have relevance to the whole training process, while also being efficient to employ in a community setting, and appropriate for evaluating applications of a psychodynamic, principle-based trainings in parent-child work. The aim of this dissertation is to report on the development

of ACTA and its utility in evaluating the impact of clinical training in Group Attachment-Based Intervention.

Development of ACTA. Active observation is a critical part of clinical work and a foundation for the implementation of other interventions (Sternberg, 2005; Bick, 1976). This is particularly true for parent-child psychotherapies that require the therapist to hold both parent and child, past and present in mind. The significance of observation in clinical work originated with Anna Freud (1953) and Esther Bick (1976), who considered infant observation central to psychoanalytic training, and was also at the heart of John Bowlby and Mary Ainsworth's work on attachment theory. Janine Sternberg (2005) eventually showed empirically that infant observation training promotes skills and capacities considered central to psychoanalytic practice.

Honing observational skills that will guide clinical thinking, conceptualization about the patient, and ultimately decisions about how to intervene, is a central part of training in GABI (Freud, 1953). Observation skills are sharpened through the extensive use of video of sessions, in both the one-day workshop and supervision groups. Video is particularly useful in parent-child therapies because it can capture the details of *interactions*, such as movement, rhythm, and facial expression, which are especially important, in a way that other methods cannot (Woodhead, Bland, Baradon, 2006). There is also empirical support for the use of video as a training resource, as research has found that videotape modeling used in training is equally effective to experiential training methods, such as role-plays (Herschell, 2009).

Theoretically, the content of clinicians' observations provides information about how they apply the knowledge of their framework for intervention and provides a snapshot of both their skills in conceptualization and clinical approach (Anna Freud, 1951). Based on what we

know about the learning process, we might expect to see clinicians change their thinking before they are able to make changes in their practice. Therefore, therapists' observations might prove useful in tracking the development of clinical understanding and skill across different phases of training.

For these reasons, the ACTA paradigm was designed to use trainees' observations of real-life clinical situations captured on video as the lens through which to evaluate their understanding and application of the GABI treatment model in a more in-depth way (beyond straight recognition) and in a naturalistic application. Myles and Milne (2004) have used a Video Assessment Task to study the effectiveness of training in CBT for adult patients, and found that clinicians improved significantly in their knowledge of and abilities to identify symptoms and name appropriate CBT strategies as a result of training. Building on their methodology, ACTA was designed as a video observation task that would assess clinicians' reflections regarding clinical situations rather than focusing on their delivery of specific techniques, with the goal of assessing changes in clinical knowledge and approach throughout training.

Current Study

The ACTA paradigm uses video clips of GABI as it is actually delivered in practice, which allows for standardization, while aiming to maintain ecological validity and reflect the complexity of this clinical work (Muse & McManus, 2013). After viewing video clips, trainees are asked to respond in writing to three open-ended questions about their observations. Responses are then rated based on the quality with which they draw on the GABI framework and general observational skills, to inform their observations. In this dissertation, we used ACTA to

evaluate the effectiveness of a one-day workshop and ongoing apprenticeship/supervision in GABI**.**

Hypotheses

The first part of this study aimed to establish the reliability of the ACTA coding system and provide an understanding of how the three videos functioned as stimuli. We hypothesized the following:

1. ACTA would be a reliable measure of training outcomes, based on both internal reliability and inter-rater reliability.

2. There would be differences in degree to which trainees appied the individual principles across the three video stimuli, but Overall Quality will not differ across three videos.

The second part of this study looked at the utility of ACTA for evaluating the effectiveness of clinical training in GABI. We hypothesized as follows:

3. The one-day workshop would significantly increase participants' understanding and application of core principles of the GABI framework, as measured by ACTA.

4. The one-day workshop would improve trainees' general clinical skills related to observation, as illustrated in their evaluation of clinical situations on ACTA.

5. The gains in application and internalization of core principles and observation skills would be maintained or increase further after ongoing supervision, eight months later.

Method

Participants

The study employed a within-subjects design, with data collection taking place at three time points across training. Data were collected over the course of 5 years, resulting in 5 training cohorts. Following is a description of the trainee population that participated in this study, across all cohorts. These participants were recruited from one-day GABI training workshops. Participants who had extensive experience in GABI (more than two years of training and responsibilities leading the groups) were excluded, because their experience may interfere with the impact of this specific workshop.

GABI One-Day Workshop Training Participants at Baseline. Ninety-nine clinicians/students (91 women, 7 men, 1 did not report gender) as shown below in Table 1.

********** INSERT TABLE 1 ABOUT HERE **********

Table 1 shows the demographic features of those who participated in the GABI workshop and the research component at baseline. Age ranged from 21 years to 74 years (M = 34.10 years, SD=11.43). Seventy of the participants were White/Caucasian (71%), 4 were Black/African-American (4%), 12 were Hispanic/Latin American (12.1%), 6 were Asian/Asian American (6.1%), 5 identified as Other (5.1%), and 2 did not report ethnicity (2.0%). Seven participants were currently in a clinical psychology Ph.D. program, 12 were in a Psy.D. program, 5 were in a masters in social work (MSW) program, 16 were currently enrolled in a general psychology MA program, 1 was in a mental health counseling program, 2 were currently working toward a BA degree, 54 had completed their training, and 2 did not report on their degree in progress. Twenty-

six participants (26.3%) had completed a MA degree, 17 (17.2%) had MSW degrees, twelve (12.1%) had completed a PhD, two (2%) had a PsyD, four (4%) were MD's, 32 (32.3%) had completed their BA degree and 2 were still undergraduates, four did not report their highest degree. Participants also had a wide range of clinical experience in general (M = 52.21 months, SD= 84.25 months), ranging from 0-38 years, and experience in GABI specifically (M = 2.77 weeks, SD = 11.29 weeks), ranging from 0-2 years.

GABI Apprenticeship Training Participants. Those trainees who continued training in the apprenticeship model at Albert Einstein were approached for consent to be assessed following completion of the 6-8 month training period (n=21; 19 female, 2 male). Demographic data for those who completed the apprenticeship is in Table 2.

********** INSERT TABLE 2 ABOUT HERE **********

Eighteen apprenticeship participants (more then 90%) were White/Caucasian, 1 was Black/African-American, 1 was Hispanic/Latin American, and 1 identified as Other ethnicity. Mean age was 29.95 years (SD=3.46). Ten trainees were currently PsyD students, four were PhD students, four were MSW students, one was not enrolled in a program, and two did not respond. At baseline they tended to have more training in GABI on average than other workshop participants (M=9.33 weeks, SD=24.58 weeks), but they had less clinical training in general (M=43.29 months, SD=89.10 months). At end of the apprenticeship, mean experience in GABI was 39.61 weeks (SD=3.45), and ranged from 25 weeks to 40 weeks.

Control Group. The study also used responses from two untrained control groups, one with low experience (Classroom) and the other with high experience (Professional).

The first untrained control group (Classroom) consisted of students in a Master's program

in general psychology who completed a class on developmental psychopathology but had not received any training in GABI (n=27). These participants were recruited verbally during the last class lecture, and viewed and responded to the videos in the classroom, as a group. They were asked to sign the consent forms, if they agreed to their responses being used for this study, and were informed that the professor would not have access to information about who consented or how they responded.

Twenty-seven students (21 female, 5 male; Table 3) participated in the untrained Classroom component of the study. Age ranged from 21 to 43 years (M = 27.15, SD = 5.97). The majority (n=20, 76.9%) was White/Caucasian, one was Black/African-American, one was Hispanic/Latin American, and four were Asian/Asian American. Students' clinical experience in psychotherapy ranged from 0 to 48 months (M=5.04, SD=11.15), and only 5 students had more than 6 months of clinical experience in psychotherapy. As noted, none of the participants in this sample had experience in GABI.

********** INSERT TABLE 3 ABOUT HERE **********

Professional Control Group. The second untrained control group consisted of Professionals who attended the Zero to Three New York Conference and participated in a breakout session on GABI (n=15). The purpose of this control group was to understand how experienced professionals, with a variety of clinical backgrounds, evaluate the situations in two of the videos used in training. Fifteen professionals (all female; Table 3) participated in the Professional control group component of the study. Age ranged from 29 to 72 years (M=53.00, SD=4.81). The majority (93.3%) was White/Caucasian (n=14), and one person reported "Other" as Ethnicity. The professionals had a range of degrees and experience. Four participants (28.6%)

held or were in the process of obtaining Ph.D. degrees, four (28.6%) held or were in the process

of obtaining Psy.D. degrees, three (21.4%) held or were in the process of obtaining MSW

degrees, and three (21.4%) held or were in the process of obtaining another type of MA degree.

Clinical experience in psychotherapy ranged from 0 months to 40 years (M=204 months, or 17

years; SD=14.8 months). By chance, one participant had two years of experience working at the

Center for Babies Toddlers and Families and in GABI previously and was excluded.

Procedure

All participants were provided with informed consent forms, and asked to fill out

background information, including age, ethnicity, education, experience delivering

psychotherapy and experience in GABI. Three videos were used as the stimuli, and they are

labeled according to the toys the children are playing with in the videos – Beads, Dollhouse, and

Scissors.

Participants in the Classroom sample viewed all three videos (Beads, Dollhouse, and

Scissors) in a group setting. Participants in the untrained sample of professionals viewed two of

the videos (Beads and Dollhouse) in a group setting at the GABI presentation.

Participants in the experimental condition were shown one of the videos (referred to as

Video 1) at the start of a one-day GABI workshop marking the beginning of training.

Immediately following the workshop, they were shown the same video (Video 1) and a novel

video (referred to as Video 2), to protect against practice effects. After approximately 6-months

of on-site training and supervision at Einstein, the third assessment was conducted, where

participants were shown the same two videos they had seen at Time 1 and 2 (Video 1 and Video

2) and a novel video (referred to as Video 3). The use of the same video(s) and a novel video after each phase of training was done in order to control for practice effects and for differences between the videos. This procedure is illustrated in Figure 1.

********** INSERT FIGURE 1 ABOUT HERE **********

The order of the three different videos was adjusted across the cohorts of trainees, so that each cohort saw a different video at Time 1. Twenty-two participants saw Beads at Time 1, 76 participants saw Dollhouse at Time 1, and 26 participants saw Scissors at Time 1.

Data were collected across five years of GABI trainings, based on an academic calendar. The one-day workshop was held in September/October of each year, and the Time 3 follow-up assessments were done when most trainees completed their apprenticeship training, usually in May/June of the following year. For the purpose of reporting the results, the five cohorts were each assigned a letter: Cohort A (2011-2012), Cohort B (2012-2013), Cohort C (2013-2014), Cohort D (2014-2015) and Cohort E (2015-2016). Due to logistical reasons, Time 3 data were not collected from Cohort D. Additionally, data from only Time 1 and Time 3 are reported for Cohort E, because this was the first year that they did not participate in a one-day workshop, and instead had an onsite orientation and were given access to an online GABI training website that they completed at their own pace. As such, there was not a clear Time 2 assessment point.

Measures

Applications of Clinical Training Assessment (ACTA). The ACTA task utilizes 3-minute long video clips edited from footage of GABI from real-life sessions delivered at Albert Einstein. This use of videotape allows for standardization, while aiming to maintain ecological validity and reflecting the complexity of this clinical work by using clips from the intervention as it is actually delivered in practice with patients (Muse & McManus, 2013). After viewing each clip, trainees are asked to respond in writing to three open-ended questions about their observations:

1. What do you see happening in this video? What struck you about this situation?

2. Imagine that you were the clinician in this situation, and were bringing it to supervision for discussion. What would you talk about?

3. Did you see any therapeutic interventions being used? Please describe.

The videos were selected by a group consisting of GABI experts (this author, Dr. Anne Murphy and Dr. Miriam Steele) and graduate students involved in viewing over 100 hours of video footage and writing the manual. Videos were chosen based on their coverage of the full range of GABI concepts, with the exception of Intergenerational Transmission. Although a core concept in GABI, Intergeneration Transmission was excluded, because it tends to arise in the Parent Group only and is typically overt; therefore identification of this concept may not accurately reflect competencies in observation.

Scoring. A coding system was developed by this author in consultation with Dr. Miriam Steele and Dr. Anne Murphy, to assess application of the GABI model to observations and evaluations of clinical situations. Barber et al. (2007) recommend that training address prescribed

(or unique and essential) elements of the treatment and nonspecific elements (either essential or nonessential) that are common to several different therapies. The ACTA paradigm is designed to evaluate both prescribed and nonspecific elements of treatment.

With the ACTA coding system, trainees' responses are assigned codes on a Likert-scale from 1 to 5 for their understanding and application of each REARING concepts (*Reflective Functioning, Emotional Attunement, Affect Regulation, Reticence, Nurturance,* and *Group Context*). On the ACTA scale, a score of 1 indicates no mention or incorrect understanding/application of the concept, 3 signifies basic or probable understanding of the concept, and 5 reflects well developed understanding, internalization of and correct application of the concept.

Responses are also coded on the same Likert-scale for three non-specific aspects of observation; these are the participant's *Own Use of Reflective Functioning* (separate from ability to identify Reflective Functioning in the video clips), *Attention to Nonverbal Information*, and the *Non-Judgmental Stance*.

Finally, responses are assigned a code for *Overall Quality of Response*, based on the number and quality of GABI concepts referenced, combined with demonstration of the other non-specific skills. A score of 1 is given to responses that overall show little or no knowledge of the GABI concepts and lack thoughtfulness and observational skills. A score of 3 is given to responses that indicate a basic understanding of some GABI concepts, but not others, and show some thoughtfulness but moderate quality of observational skills. A score of 5 is given to those responses that demonstrate recognition and understanding of many GABI concepts, as well as

clinical and observational skills in other non-specific areas. A sample of the coding sheet is included in Appendix 1.

The aim of this study is to establish the reliability and validity of the ACTA paradigm, and determine whether the current training processes are efficacious in teaching trainees how to employ the GABI framework in their clinical work.

Results

Results reported in this section are organized in three categories 1) Establishing reliable anchors for GABI concepts, 2) Exploring differential responses to video stimuli at baseline, and 3) Evaluating effects of a one-day workshop and apprenticeship training in GABI on application of GABI concepts.

Establishing Reliable Anchors and Internal Reliability

Prior to calculating reliability, two coders coded 27 responses, discussed their codes and came to a consensus on codes for a subset of the sample. Based on their discussions, definitions of codes were refined in the ACTA Coding Manual.

Three coders then completed a reliability set of 24 responses using the ACTA Coding Manual, and their inter-rater reliability was calculated. Table 4 shows that inter-rater reliability (average two-way, random effects ICC $\geq$ 0.70) was achieved for all of the codes with the exception of *Non-Judgmental Stance* (ICC=.48). Despite the lack of reliability on Non-Judgment, we believed this to be an important construct, and therefore kept it in the other analyses.

********** INSERT TABLE 4 ABOUT HERE **********

Two or more raters coded more than twenty-five percent (26.11%) of the responses (n=70). Two-way random effects ICCs from these cases are reported in Table 5.

********** INSERT TABLE 5 ABOUT HERE **********

On the full sample there was some drift in the reliability, which declined slightly. However, according to Cicchetti & Sparrow (1981) ICCs in the range of .40–.59 are considered fair, .60–.74 considered good, and above .75 are considered excellent. When ratings differed by one point, the lead author's rating was used. When ratings differed by two points, the raters discussed and reached a consensus score, which was used in the final analyses. The first author coded the remainder of the sample.

Exploratory analyses were also conducted to understand the internal reliability of the coding system. Cronbach's Alpha at Time 1 was .643. This is an appropriate alpha for this type of coding system, because the GABI concepts are not intended to overlap significantly, but we would expect them to be related, such that if someone has a strong understanding of GABI they score higher on most, and if they have a weak understanding of GABI they score lower on most. We also expected that each of the codes should be correlated with the *Overall Quality* score, and indeed found that all of the codes, except for *Use of Own RF* and *Attention to Nonverbal Detail*, were positively correlated with the *Overall Score* code at all three time points (Time 1 Video 1 and Time 2 Video 2 correlations are reported in Table 6).

Assessing Differences Among Video Stimuli in Control Samples at Baseline

In order to maximize the ecological validity of the study, the video stimuli used were

clips taken from real GABI sessions, which were selected based on the presence of all the GABI concepts, even if subtle, rather than the more obvious presence of just a few concepts. Because the videos were from actual sessions and not staged, we expected there to be differences in the presence of individual GABI concepts among videos, since the content of some videos prompted participants to focus more strongly on certain GABI concepts. We hypothesized, however, that the differences between videos would not result in differences in *Overall Quality of Response*, a measure of overall understanding and application of GABI concepts and non-specific principles.

We used three control samples to identify any differences between videos: an untrained sample of students (Classroom) and an untrained group of professionals (Professionals) with experience working with children zero to three, but without training in GABI, and GABI trainees at baseline (Trainee). We sought to use this information to better interpret the results when using these videos to assess the impact of the GABI training process.

Across all three samples we ran three comparisons (Beads vs. Dollhouse, Beads vs. Scissors, Dollhouse vs. Scissors), for each of the 10 codes, resulting in 90 total comparisons. Of those, 23 were significant, which we report in Tables 7, 8, and 9.

********** INSERT TABLE 7 ABOUT HERE **********

********** INSERT TABLE 8 ABOUT HERE **********

********** INSERT TABLE 9 ABOUT HERE **********

In all three samples, *Reflective Functioning* (RF) ratings were lower on the Beads video, compared to Dollhouse. In the trainee sample, *RF* ratings were also significantly lower on Beads compared to Scissors. In all three samples, *Emotional Attunement* was significantly higher in the Scissors video, compared to the Dollhouse video. *Affect Regulation* was also significantly higher

on Scissors compared to either Dollhouse or Beads, and it was significantly higher on Dollhouse than on Beads. *Nurturance* scores were highest on Beads, compared to the other two videos. And *Group* scores were lowest on Beads, compared to the other two videos. Participants' *Own RF* was highest on Dollhouse compared to the other two videos. There were no significant differences in *Reticence*, *Attention to Nonverbal Details*, *Non-Judgment*, and *Overall Quality of Response*.

The lack of differences among *Overall Quality* ratings is important theoretically, suggesting that as predicted, although there were some differences at the individual principle level, these differences ultimately balanced each other out and resulted in similar *Overall* scores. In other words, although participants recognized some differences in concepts in the three videos, the overall quality of their responses was consistent across videos.

We also investigated whether there were any correlations between therapist factors (months of clinical experience and weeks of experience in GABI) and ACTA ratings. We found that Months of Clinical Experience was only correlated with *Reflective Functioning* at baseline (r= -.197, p=.03). This correlation was weak and cannot be controlled for in paired samples t-tests, but it disappeared when we looked at the relationship between Months of Clinical Experience and *RF* at Time 2.

Evaluating the Impact of One-Day Workshop

In order to evaluate the impact of the one-day training workshop in the sample of trainees, we ran paired samples t-tests to analyze differences in mean scores on the same video before (Time 1 Video 1) and after (Time 2 Video 1) the training workshop, and with a novel

video after the training workshop (Time 2 Video 2).

Full results are reported in Table 10. The following results report only the significant differences in scores for each concept, and for *Overall* rating.

********** INSERT TABLE 10 ABOUT HERE **********

As Table 10 shows, when the same video (Video 1) was used before and after the training workshop, there were significant increases in scores on all GABI-specific principles: *Reflective Functioning, Emotional Attunement, Affect Regulation, Reticence, Nurturance,* and *Group.* There were also significant improvements in *Non-Judgment* and *Overall Quality of Observation.* When a novel video (Video 2) was used after the workshop there were significant differences in the predicted direction, with scores on *Emotional Attunement, Reticence, Nurturance, Group* and *Overall Quality* higher than Video 1 before training (Table 10).

Effect sizes were calculated and are reported in Table 10, as well. On comparisons where t-tests were significant, most effect sizes ranged from moderate (*Affect Regulation, Reflective Functioning*) to very large (*Reticence, Overall Quality*). Only *Group Context* was considered to have a small effect size.

Data from before and after the workshop were collected across four different years, and as we would expect there were differences among the cohorts at baseline, and as a result of training. At baseline, Cohort D had the lowest mean *Overall Quality of Response* (M=2.11, SD=.73), which was significantly lower than the *Overall Quality of Response* for Cohort B, who had the highest mean *Overall Quality* rating at baseline (M=3.00, SD=.76).

As we would expect, there were differences among the cohorts in the changes observed as a result of training, shown in Table 11. Because the order of the videos differed across cohort

years, changes in the novel video could not be considered comparable, so we looked only at how observations changed when trainees saw the same video at both Time 1 and Time 2.

********** INSERT TABLE 11 ABOUT HERE **********

All cohorts except Cohort B, which had a small sample size of only 6, showed improvements in *Overall Quality of Response*. Cohort A improved in *Affect Regulation* and *Nurturance* when the same video was seen a second time. Cohort B improved in *Reticence* when they viewed the same video after training. Cohort C showed improvements in *Emotional Attunement*, *Affect Regulation*, and *Reticence* when the same video was presented after training. Finally, Cohort D demonstrated the greatest number of significant changes. Cohort D's ratings increased on all of the REARING principles when they saw the same video after the workshop.

Evaluating the Impact of Apprenticeship and Supervision

To evaluate the development and maintenance of competencies at follow-up approximately six to eight months later, after the apprenticeship training, we ran paired samples t-tests comparing Video 1 at Time 1 to the same video at Time 3 (Video 1), as well as to Video 2, which they saw at Time 2, and a novel video (Video 3), at the end of the apprenticeship (Time 3). We also ran paired samples t-tests to investigate changes from Time 2 to Time 3 in observations of the same videos (Video 1 and Video 2), and each compared to a novel video (Video 3) at Time 3. Results from the sample that completed a 6-8 month apprenticeship with supervision (n=21) are presented in Table 12, and means are reported in Table 13.

********** INSERT TABLE 12 ABOUT HERE **********

********** INSERT TABLE 13 ABOUT HERE **********

First, we looked at the impact of the one-day workshop on this subset of trainees, as they tended to have less training overall but more training in GABI at the start of training, compared to other participants in the one-day workshop (Table 12). The trainees who participated in both the one-day workshop and apprenticeship training showed significant improvements after the one-day workshop only in *Reticence* and *Overall Quality of Response* when both the same video and novel video were used.

There were, however, improvements observed from Time 2 (after the workshop) to Time 3 (after the apprenticeship). *Reflective Functioning* scores increased significantly when comparing Time 1 Video 1 to Time 3 Video 1 (same video) and Time 3 Video 3 (novel video). *Affect Regulation* scores increased, when comparing Time 2 Video 2 to Time 3 Video 2 (same video) and Time 3 Video 3 (novel video). *Nurturance* scores increased when comparing Time 2 Video 2 to the novel video (Video 3) at Time 3. *Group* scores increased when comparing the same video (Video 1) at Time 2 and Time 3. However, *Overall Quality of Response* did not improve from Time 2 to Time 3, on any of the comparisons.

Table 14 reports effect sizes on the comparisons that were significant.

********** INSERT TABLE 14 ABOUT HERE **********

Most effect sizes ranged from medium (*Reticence*, Time 1 Video 1 compared to Time 3 Video 1), to very large (*Overall*, Time 1 compared to Time 2, Time 1 compared to Time 3 Video 1, Video 2, and Video 3; *Reticence*, Time 1 Video 1 compared to Time 2 Video 2; *Affect Regulation*, Time 1 Video 1 compared to Time 3 Video 1). Only one of the effect sizes was small, *Nurturance* on Time 1 Video 1 compared to time 3 Video 1.

When comparing observations at Time 3 to those from Time 1, there were also some

notable improvements. *Affect Regulation* scores were higher on the novel video at Time 3, compared to Video 1 at Time 1. *Reticence* scores were higher on both Videos 1 and 2 at Time 3, compared to Video 1 at Time 1. And *Nurturance* scores increased when Video 1 was shown again at Time 3. The *Overall Quality of Response* increased significantly from Time 1 to Time 3, regardless of whether they saw Video 1, 2, or 3 at Time 3.

To further understand the lack of changes in multiple GABI concepts in the apprenticeship sample, we conducted independent samples t-tests, comparing the apprenticeship group to the workshop-only group. The apprenticeship group had significantly more experience in GABI prior to training, compared to the workshop-only group ($t(df)$=2.63(104), p=.01). At Time 1, the apprenticeship group also had higher ratings on *Reticence* ($t(df)$=2.16(104), p=.03) and *Overall Quality of Response* ($t(df)$=2.23(104). p=.03). However, after the workshop, there were not significant differences between groups in mean scores on either *Reticence* or *Overall Quality of Response*. In fact, the only significant difference between groups at Time 2 were on *Reflective Functioning* ($t(df)$=2.94(97), p=.00) and *Group* ($t(df)$=3.05(97), p=.00), and workshop-only participants achieved higher ratings, compared to apprenticeship participants.

Discussion

Despite research demonstrating the negative sequelae of adverse childhood experiences (ACEs) (Felitti et al, 1998; Dube et al., 2013) and the intergenerational transmission of child abuse and neglect (Berlin et al., 2011; Fonagy et al., 1998; Murphy et al., 2013), there continues to be a disparity between what we know about treatment and prevention of abuse and neglect and the implementation of targeted interventions to address the problem in the community.

Families that are considered at-risk of child abuse and neglect present myriad needs and challenges, including trauma, poverty, social isolation, and discrimination. Trainings focused on broad principles of treatment that can serve as a foundation and be flexibly applied as needed, rather than emphasizing specific protocols, have been recommended to treat such complex problems (Chorpita, 2009). However, the effectiveness of such trainings should not be assumed. If we are to improve the dissemination of evidence-based interventions, we must empirically test whether training is effectively teaching clinicians the information and skills they need to implement these interventions in practice.

Group Attachment Based Intervention (GABI) was designed to prevent child maltreatment, and its training aims to increase clinicians' understanding of core principles (REARING) thought to facilitate the therapeutic action and help parents and children work toward attachment security, from a trauma-informed perspective.

There are a dearth of measures designed to assess knowledge and competence in parent-child therapies. In particular, training assessments need to have ecological validity, be sensitive to changes across the training (including both short and long-term outcomes), and be efficient to administer and score so that they can be used when interventions are brought to scale. We developed the Applications of Clinical Training Assessment (ACTA) to meet these needs. Recognizing that observation is a foundation for clinical practice, the ACTA task asks participants to view 3-minute videotapes of sessions and record their observations in response to open ended questions. Their responses are then coded for the degree to which they demonstrate understanding and skillfull application of the GABI specific principles (REARING) and general

observation skills, such as mentalization, attention to nonverbal details, and a non-judgmental stance.

This study represents the first time ACTA has been used to evaluate a parent-child training. Results from 99 clinicians who participated in a one-day workshop and 21 who completed an apprenticeship following the workshop provide evidence that training in GABI has an impact on overall quality of observations of clinical situations, as well as understanding and application of key principles specific to the intervention.

Furthermore, results demonstrated that independent coders could achieve reliability (ICC $\geq .70$) on the individual concepts central to GABI and overall ratings. One code, *Non-Judgmental Stance*, does need further refinement in order to improve inter-rater reliability. Identification of anchors for the GABI principles that can be reliably coded is an important first step in the development of more complex adherence and skill measures that can be used in future studies of dyadic therapies.

Part one of this study aimed to understand how the three video stimuli used in ACTA functioned in three untrained control groups: psychology MA students (Classroom), professionals (Professionals), and participants in a GABI one-day workshop at baseline (Trainee). Because the video clips were taken from actual GABI sessions and were not staged, we expected there to be individual differences among videos in terms of the individual GABI concepts observers focused on. However, we also hypothesized that these differences on individual items would not result in differences in *Overall Quality of Response* score, a measure of overall understanding and application of GABI concepts. Indeed, results indicated that as expected there were individual differences in how participants included the individual concepts,

but these differences did not amount to differences in *Overall Quality* scores across the three videos.

Furthermore, the differences were consistent across all three samples. *Nurturance* was highest in the Beads video, *Reflective Functioning* was highest in the Dollhouse video, and *Emotional Attunement* and *Affect Regulation* were highest in the Scissors video. There were no significant differences in *Reticence, Group*, or *Overall Quality* scores across the three videos. The lack of differences in *Overall Quality* scores is critical because it indicates that any difference in *Overall* score seen in the experimental arm of the study could be interpreted as due to the training and not differences between the videos.

The second part of this study evaluated the impact of a one-day workshop on understanding and application of the GABI concepts and non-specific observational skills. Results showed that when trainees were shown the same video before and after the workshop, their ability to skillfully apply all of the GABI-specific principles increased as a result of the training. They also increased in their ability to take a *Non-Judgmental Stance*, but their use of other general observational skills did not change. When participants saw a novel video after the workshop, their observations improved in terms of recognizing *Emotional Attunement*, *Reticence, Nurturance*, and *Group Context*, but not on the other GABI or general principles. Perhaps most importantly, the *Overall Quality* of trainees' responses increased significantly, when both the same video and the novel video were shown after training, to a 3.69 on the same video and a 3.56 on the novel video. As a score of 3 represents a basic understanding, this rating is reassuring that trainees have a satisfactory level of competence following the training.

When we looked at the sample in terms of individual cohorts, we found that *Overall*

Quality of Response consistently increased as a result of the one-day workshop (with the exception of Cohort B, which had a small sample size of only six). By the fourth training cohort (Cohort D), trainees showed improvements on all of the REARING principles when the same video was seen after the workshop, though this had not been true for previous cohorts.

The third part of this study evaluated the effectiveness of the apprenticeship component of GABI training. Twenty-one trainees completed ACTA at baseline and after the apprenticeship. At Time 3 they saw the same videos they had been shown at Time 1 and Time 2, as well as a novel video. Across all three videos, ratings on *Overall Quality of Response* increased, compared to their baseline responses. However, there were few significant increases in ratings on GABI specific and general principles (only *Affect Regulation*, *Reticence*, and *Nurturance*, and these were not consistent across videos). Of note, trainees who participated in the apprenticeship (n=21) actually demonstrated fewer significant increases in understanding and application of GABI principles following the one-day workshop, compared to the sample as a whole. They tended to have more GABI specific experience but less general experience than the sample as a whole, coming into training. Consistent with what we would expect, then, they also had higher ratings on *Reticence* and *Overall Quality of Response* before training. This finding is similar to other studies that have shown that participants with more specific experience in a treatment modality tend to show fewer effects from the training itself (Rounsaville et al., 1986; Crits-Christoph et al., 1998; Henry et al., 1993; Siqueland, Crits-Christoph, Barber, et al., 2000).

Gaining Valuable Feedback

Beyond demonstrating positive effects of training, empirical study of training offers us the valuable opportunity to identify areas for improvement in training. The current study demonstrated that the one-day workshop and apprenticeship training impacted overall understanding and application of the framework, but the results regarding individual components of the framework were less conclusive when cohorts were looked at individually. Results of training outcome studies should be explored and used to inform future trainings. In this regard, the fact that Cohort D - the most recent and the largest cohort - had significant improvements on all of the GABI specific principles is hopeful.

The results raise questions, though, about how training can help those who are coming in with some experience, but have not yet achieved a high level of knowledge or skill, to improve in these areas. Looking at mean ratings from Time 3, it appears that high *Overall Quality of Response* ratings were driven by high ratings in *Reflective Functioning*, *Reticence*, and *Non-Judgment*. Greater attention to the principles of *Emotional Attunement*, *Affect Regulation*, *Nurturance* and *Group Context* may represent areas of opportunity for training to help clinicians improve.

Additionally, though not a core GABI concept, results from the workshop and the apprenticeship components support the need for stronger emphasis on nonverbal behavior. Heightened focus on this area may improve not only scores on the *Nonverbal* code, but also abilities to recognize the core GABI concepts within nonverbal behavior and interactions.

Finally, it is worth noting that new developments were being made in the GABI training process at the time this study was nearing completion. First, the workshop format has been

expanded to multiple days in order to cover more material. Additionally, the written manual that was used to guide training during the period of this study has been updated and translated into an online training curriculum, hosted on Udemy. The online platform allows for teaching to occur through written materials, as well as lectures, and most crucially, video of clinical material. In future research we will use ACTA to assess the unique contributions of this training tool, as a stand-alone guide as well as in combination with other training activities (i.e. workshops, supervision, apprenticeship).

Limitations and Areas for Future Research

There are some limitations to this study and areas for future research that should be noted. First, the training evaluation method used assumes that knowledge about the theory and techniques comprising an intervention is a prerequisite for effective intervention use (Becker et al., 2011). However, we recognize that quality of observations, particularly in response to video, is not a substitute for other types of training outcomes, such as skill in clinical practice and ultimately patient outcomes. Additionally, participants in this study were limited to a set amount of time (3 minutes) to respond to each question. While this standardized the procedure and was aimed at efficiency and preventing fatigue, it is possible that responses may have been different if more time was provided for the responses or if participants could say their responses out loud as opposed to write them out. Although the use of a video-observation paradigm and written responses are expedient, further research is needed to show that it is also valid and related to outcomes in clinical practice.

Towards this end, we have begun development of a competence measure that applies ratings of clinicians' competence in utilizing the core REARING principles in practice, by coding video of GABI sessions. Future studies will explore the links between performance on ACTA and competence ratings on sessions. It will also be important for us to show that competence on ACTA and in practice is related to patient outcomes.

A second limitation was the small sample size at follow-up. This was a naturalistic study of training, and at this time, GABI training is only taking place at one institution. As GABI is disseminated to other institutions, there will be opportunities for more clinicians to complete apprenticeships or implement GABI with close supervision. It will be especially important for us to use ACTA when lead clinicians at other institutions are starting a group with new trainees, to ensure that their training processes are also effective and that trainees are continuing to internalize the key principles of GABI.

Third, while an attempt was made to counterbalance the videos, this procedure in a naturalistic setting could not be precise; more data is needed in order to completely control for the effect of the different videos. We are currently in the process of creating an online version of ACTA that will allow us to more precisely counter-balance the videos in the administration of the task.

Lastly, the GABI treatment model contains three separate components, only two of which were employed in this study: a parent-child component, a parent-only component and a child-only component. Videos 1 and 3 were of the parent-child group, and Video 2 was of the child-only group. While the GABI concepts and clinical and observational skills required for each component are theoretically the same, the groups do "look" different and it remains unclear

whether the clinical competencies observed in responses to the child-only and parent-child videos are the same as what would be observed in responses to the parent-only group. Future research will incorporate the parent-group into the training and competency research.

Conclusion

Findings from this dissertation indicate that the one-day GABI training workshop and the apprenticeship training increase trainees' ability to understand and apply the clinical framework to their evaluations of new clinical situations. We identified reliable anchors for the GABI concepts, which can inform future measures of adherence and skill, and also showed that training does have an impact on how trainees' observe and understand the intervention.

By empirically measuring the impact of training, we are aiming to not only demonstrate that training is effective, but to also identify where changes are needed to improve training. A large area for future research will be how the apprenticeship model and reflective supervision can further impact trainees' clinical skills in practice. As the ultimate goal is improving treatment for children and families, ongoing research will continue to assess whether the therapeutic interventions identified, and the training processes in place, effectively produce the desired outcomes for families.

This research marks an important step in laying the foundation for continued study of the impact of training in therapies for children and families. Beyond its utility in ensuring that the quality of the treatment is maintained throughout the dissemination process, ACTA is a tool that we hope can be adapted and used to evaluate other interventions that target the needs of parents and children.

References

Abramowitz, J. S. (2006). Toward a functional analytic approach to psychologically complex patients: A Comment on Ruscio and Holohan (2006). *Clinical Psychology: Science and Practice*, *13*(2), 163-167.

Barber, J. P., Triffleman, E., & Marmar, C. (2007). Considerations in treatment integrity: Implications and recommendations for PTSD research. *Journal of Traumatic Stress*, *20*(5), 793-805.

Barnett, E. R., Rosenberg, H. J., Rosenberg, S. D., Osofsky, J. D., & Wolford, G. L. (2014). Innovations in Practice: Dissemination and implementation of child–parent psychotherapy in rural public health agencies. *Child and Adolescent Mental Health*, *19*(3), 215-218.

Barnfield, T. V., Mathieson, F. M., & Beaumont, G. R. (2007). Assessing the development of competence during postgraduate cognitive-behavioral therapy training. *Journal of Cognitive Psychotherapy: An International Quarterly*, *21*(2), 140-148.

Becker, K. D., & Stirman, S. W. (2011). The science of training in evidence-based treatments in the context of implementation programs: current status and prospects for the future. *Administration and Policy in Mental Health*, *38*(4), 217-22. doi:10.1007/s10488-011-0361-0

Beidas, R. S., Edmunds, J. M., Marcus, S. C., & Kendall, P. C. (2012). Training and consultation to promote implementation of an empirically supported treatment: A randomized trial. *Psychiatric Services*, *63*(7), 660-665.

Beveridge, R. M., Fowles, T. R., Masse, J. J., McGoron, L., Smith, M. A., Parrish, B. P., ... &
Widdoes, N. (2015). State-wide dissemination and implementation of parent–child
interaction therapy (PCIT): Application of theory. *Children and Youth Services
Review,* 48, 38-48.

Berlin, L. J., Appleyard, K., & Dodge, K. A. (2011). Intergenerational continuity in child
maltreatment: Mediating mechanisms and implications for prevention. *Child
Development*, *82*(1), 162-176.

Bernard, K., Dozier, M., Bick, J., & Gordon, M. K. (2015). Intervening to enhance cortisol
regulation among children at risk for neglect: results of a randomized clinical trial.
Development and Psychopathology, 27(03), 829-841.

Bernard, K., Dozier, M., Bick, J., Lewis-Morrarty, E., Lindhiem, O., & Carlson, E. (2012).
Enhancing attachment organization among maltreated children: Results of a randomized
clinical trial. Child Development, 83(2), 623-636.

Bick, E. (1964). Notes on infant observation in psychoanalytic training. *International Journal of
Psychoanalysis*, *45*, 558-566.

Blackburn, I. M., James, I. A., Milne, D. L., Baker, C., Standart, S., Garland, A., & Reichelt, F.
K. (2001). The revised cognitive therapy scale (CTS-R): psychometric properties.
Behavioural and CognitivePpsychotherapy, 29(04), 431-446.

Borrelli, B., Sepinwall, D., Ernst, D., Bellg, A. J., Czajkowski, S., Breger, R., DeFrancesco, C.,
et al. (2005). A new tool to assess treatment fidelity and evaluation of treatment fidelity
across 10 years of health behavior research. *Journal of Consulting and Clinical
Psychology*, *73*(5), 852-60. doi:10.1037/0022-006X.73.5.852

Caron, E. B., Weston-Lee, P., Haggerty, D., & Dozier, M. (2016). Community implementation outcomes of Attachment and Biobehavioral Catch-up. *Child Abuse & Neglect*, 53, 128-137.

Chorpita, B. F., & Daleiden, E. L. (2009). Mapping evidence-based treatments for children and adolescents: application of the distillation and matching model to 615 treatments from 322 randomized trials. *Journal of Consulting and Clinical Psychology*, *77*(3), 566.

Christian, A. S., Niec, L. N., Acevedo-Polakovich, I. D., & Kassab, V. A. (2014). Dissemination of an evidence-based parenting program: Clinician perspectives on training and implementation. *Children and Youth Services Review*, *43*, 8-17.

Cicchetti, D., Rogosch, F.A., & Toth, S.L. (2000). The efficacy of Toddler-Parent Psychotherapy for fostering cognitive development in offspring. *Journal of Abnormal Child Psychology. 28,* 135-148.

Cicchetti, D., Rogosch, F.A., & Toth, S.L. (2006). Fostering secure attachment in infant in maltreating families through preventive interventions. *Development and Psychopathology, 18,* 623-650.

Cicchetti D., Toth S.L., Rogosch F.A. (1999). The efficacy of toddler-parent psychotherapy to increase attachment security in offspring of depressed mothers. *Attachment and Human Development 1,* 34-66.

Cohen, N. J., Muir, E., Lojkasek, M., Muir, R., Parker, C. J., Barwick, M., & Brown, M. (1999). Watch, wait, and wonder: Testing the effectiveness of a new approach to mother-infant psychotherapy. *Infant Mental Health Journal*, *20*(4), 429-451.

Crits-Christoph, P, Siqueland, L., Chittams, J., Barber, J. P., Beck, A. T., Frank, A., Liese, B., et al. (1998). Training in cognitive, supportive-expressive, and drug counseling therapies for cocaine dependence. *Journal of Consulting and Clinical Psychology*, *66*(3), 484-92. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3330597&tool=pmcentrez&rendertype=abstract

Crits-Christoph, P., Connolly, M. B., Azarian, K., Crits-Christoph, K., & Shappell, S. (1996). An open trial of brief supportive-expressive psychotherapy in the treatment of generalized anxiety disorder. *Psychotherapy*, *33*(3), 418-430.

Crits-Christoph, P., Gibbons, M. B. C., Crits-Christoph, K., Narducci, J., Schamberger, M., & Gallop, R. (2006). Can therapists be trained to improve their alliances? A preliminary study of alliance-fostering psychotherapy. *Psychotherapy Research*, *16*(3), 268-281. doi:10.1080/10503300500268557

David, P., & Schiff, M. (2015). Learning from bottom-up dissemination: Importing an evidence-based trauma intervention for infants and young children to Israel. *Evaluation and Program Planning*, 53, 18-24.

Dennhag, I., & Armelius, B. Å. (2012). Baseline training in cognitive and psychodynamic psychotherapy during a psychologist training program. Exploring client outcomes in therapies of one or two semesters. *Psychotherapy Research*, 22(5), 515-526.

Dodd, K., Austin, K., Baxter, L., Jennison, J., Kenny, M., Lippold, T., ... & Webb, Z. (2013). Effectiveness of brief training in cognitive-behaviour therapy techniques for staff

working with people with intellectual disabilities. *Advances in Mental Health and Intellectual Disabilities*, *7*(5), 300-311.

Dozier, M., Peloso, E., Lindhiem, O., Gordon, M. K., Manni, M., Sepulveda, S., Ackerman, J., et al. (2006). Developing evidence-based interventions for foster children: an example of a randomized clinical trial with infants and toddlers. *Journal of Social Issues*, *62*(4), 767-785. doi:10.1111/j.1540-4560.2006.00486.x

Eggbeer, L, Shahmoon-Shanok, R., Clark, R. (2010). Reaching towards an evidence base for reflective supervision. *Zero to Three*, Vol. 31 (2) 39-45, Washington DC.

Felitti, M. D., Vincent, J., Anda, M. D., Robert, F., Nordenberg, M. D., Williamson, M. S., et al. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults: The Adverse Childhood Experiences (ACE) Study. *American Journal of Preventive Medicine*, *14*(4), 245-258.

Fonagy, P., Steele, M., Steele, H., Moran, G. S., & Higgitt, A. C. (1991). The capacity for understanding mental states: The reflective self in parent and child and its significance for security of attachment. *Infant Mental Health Journal*, *12*(3), 201-218.

Freud, A. (1951). Observations on child development the phases of libido development reflected in the child's behavior. *Psychoanalytic Study of the Child*, *6*, 18-30.

Freud, A. (1953). Some remarks on infant observation. *Psychoanalytic Study of the Child*, *8*, 9-18.

Garbacz, L. L., Brown, D. M., Spee, G. A., Polo, A. J., & Budd, K. S. (2014). Establishing Treatment Fidelity in Evidence-Based Parent Training Programs for Externalizing

Disorders in Children and Adolescents. *Clinical Child and Family Psychology*

Review*, 17*(3), 230-247.

Harris, M., & Bick, E. (1987). The contribution of observation of mother-infant interaction and

development to the equipment of a psychoanalyst or psychoanalytic psychotherapist. In

M. H. Williams (Ed.), *Collected Papers of Martha Harris and Esther Bick* (pp. 117-132).

Perthshire: The Clunie Press.

Heinicke, C. M., Fineman, N. R., Ruth, G., Recchia, S. L., Guthrie, D., & Rodning, C. (1999).

Relationship-based intervention with at-risk mothers: outcome in the first year of life.

Infant Mental Health Journal, 20(4), 349-374.

Heinicke, C. M., Goorsky, M., Levine, M., Ponce, V., Ruth, G., Silverman, M., & Sotelo, C.

(2006). Pre- and postnatal antecedents of a home-visiting intervention and family

developmental outcome. *Infant Mental Health Journal, 27*(1), 91-119. doi:10.1002/imhj.

Heinicke, C. M., Goorsky, M., Moscov, S., Dudley, K., Gordon, J., Schneider, C., & Gurthrie, D.

(2000). Relationship-based intervention with at-risk mothers: factors affecting variations

in outcome. *Infant Mental Health Journal, 21*(3), 133-155.

Henry, W. P., Schacht, T. E., Strupp, H. H., Butler, S. F., & Binder, J. L. (1993). Effects of

training in time-limited dynamic psychotherapy: mediators of therapists' responses to

training. *Journal of Consulting and Clinical Psychology, 61*(3), 441-7. Retrieved from

http://www.ncbi.nlm.nih.gov/pubmed/8326045

Herschell, A. D., Kolko, D. J., Baumann, B. L., & Davis, A. C. (2010). The role of therapist

training in the implementation of psychosocial treatments: A review and critique with

recommendations. *Clinical Psychology Review*, *30*(4), 448-466. doi:10.1016/j.cpr.2010.02.005.

Herschell, A. D., McNeil, C. B., Urquiza, A. J., McGrath, J. M., Zebell, N. M., Timmer, S. G., & Porter, A. (2009). Evaluation of a treatment manual and workshops for disseminating, parent–child interaction therapy. *Administration and Policy in Mental Health and Mental Health Services Research*, *36*(1), 63-81.

Hilsenroth, M. J., Defife, J. A., Blagys, M. D., & Ackerman, S. J. (2006). Effects of training in short-term psychodynamic psychotherapy: changes in graduate clinician technique. *Psychotherapy Research*, *16*(3), 293-305. doi:10.1080/10503300500264887

Hoffman, K. T., Marvin, R. S., Cooper, G., & Powell, B. (2006). Changing toddlers' and preschoolers' attachment classifications: the Circle of Security intervention. *Journal of Consulting and Clinical Psychology*, *74*(6), 1017-26. doi:10.1037/0022-006X.74.6.1017

Holloway, E. L., & Neufeldt, S. A. (1995). Supervision: its contributions to treatment efficacy. *Journal of Consulting and Clinical Psychology*, *63*(2), 207.

Jensen-Doss, A., Cusack, K. J., & de Arellano, M. a. (2008). Workshop-based training in trauma-focused CBT: an in-depth analysis of impact on provider practices. *Community Mental Health Journal*, *44*(4), 227-44. doi:10.1007/s10597-007-9121-8

Karlin, B. E., Trockel, M., Taylor, C. B., Gimeno, J., & Manber, R. (2013). National dissemination of cognitive behavioral therapy for insomnia in veterans: therapist-and patient-level outcomes. *Journal of Consulting and Clinical Psychology, 81*(5), 912.

Kazak, A. E., Hoagwood, K., Weisz, J. R., Hood, K., Kratochwill, T. R., Vargas, L. A., & Banez,

    G. A. (2010). A meta-systems approach to evidence-based practice for children and

    adolescents. *American Psychologist*, *65*(2), 85.

Kazdin, A. E. (2008). Evidence-based treatment and practice: new opportunities to bridge

    clinical research and practice, enhance the knowledge base, and improve patient care.

    American Psychologist, *63*(3), 146-59. doi:10.1037/0003-066X.63.3.146

Kendall, P. C., & Beidas, R. S. (2007). Smoothing the trail for dissemination of evidence-based

    practices for youth: flexibility within fidelity. *Professional Psychology: Research and

    Practice*, *38*(1), 13.

Lanier, P., Kohl, P. L., Benz, J., Swinger, D., Moussette, P., & Drake, B. (2011). Parent–child

    interaction therapy in a community setting: examining outcomes, attrition, and treatment

    setting. *Research on social work practice*, *21*(6), 689-698.

 Lieberman, A. F., Ghosh Ippen, C., & Van Horn, P. (2006). Child-parent psychotherapy: 6-

    month follow-up of a randomized controlled trial. *Journal of the American Academy of

    Child and Adolescent Psychiatry*, *45*(8), 913-8. doi:10.1097/01.chi.0000222784.03735.92

Lieberman, A. F., Van Horn, P., & Ippen, C. G. (2005). Toward evidence-based treatment:

    Child-parent psychotherapy with preschoolers exposed to marital violence. *Journal of the

    American Academy of Child & Adolescent Psychiatry*, *44*(12), 1241-1248.

Luborsky, L., & DeRubeis, R. J. (1984). The use of psychotherapy treatment manuals: A small

    revolution in psychotherapy research style. *Clinical Psychology Review*, *4*(1), 5-14.

Lyon, A. R., & Budd, K. S. (2010). A community mental health implementation of Parent-Child

      Interaction Therapy (PCIT). *Journal of Child and Family Studies*, *19*(5), 654-668.

      doi:10.1007/s10826-010-9353-z

Lyon, A. R., Stirman, S. W., Kerns, S. E. U., & Bruns, E. J. (2011). Developing the mental

      health workforce: review and application of training approaches from multiple

      disciplines. *Administration and Policy in Mental Health*, *38*(4), 238-53.

      doi:10.1007/s10488-010-0331-y

Meade, E. B., Dozier, M., & Bernard, K. (2014). Using video feedback as a tool in training

      parent coaches: Promising results from a single-subject design. *Attachment & Human

      Development, 16*(4), 356-370.

McHugh, R. K., & Barlow, D. H. (2010). The dissemination and implementation of evidence-

      based psychological treatments. A review of current efforts. *The American Psychologist*,

      65(2), 73-84. doi:10.1037/a0018121

Miller, W. R., & Mount, K. A. (2001). A small study of training in Motivational Interviewing:

      Does one workshop change clinician and client behavior? *Behavioural and Cognitive

      Psychotherapy*, *29*(2001), 457-471.

Miller, W. R., Yahne, C. E., Moyers, T. B., Martinez, J., & Pirritano, M. (2004). A randomized

      trial of methods to help clinicians learn motivational interviewing. *Journal of Consulting

      and Clinical Psychology*, *72*(6), 1050-62. doi:10.1037/0022-006X.72.6.1050

Milne, D. L., Baker, C., Blackburn, I. M., James, I., & Reichelt, K. (1999). Effectiveness of

      cognitive therapy training. *Journal of Behavior Therapy and Experimental Psychiatry*,

      30(2), 81-92. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10489085

Moras, K. (1993). The use of treatment manuals to train psychotherapists: Observations and

    recommendations. *Psychotherapy: Theory, Research, Practice, Training*, *30*(4), 581.

Morgenstern, J., Morgan, T. J., McCrady, B. S., Keller, D. S., & Carroll, K. M. (2001). Manual-

    guided cognitive-behavioral therapy training: A promising method for disseminating

    empirically supported substance abuse treatments to the practice community. *Psychology*

    of Addictive Behaviors, *15*(2), 83-88. doi:10.1037//0893-164X.15.2.83

Moyers, T. B., Manuel, J. K., Wilson, P. G., Hendrickson, S. M. L., Talcott, W., & Durand, P.

    (2007). A randomized trial investigating training in motivational interviewing for

    behavioral health providers. *Behavioural and Cognitive Psychotherapy*, *36*(02), 149-162.

    doi:10.1017/S1352465807004055

Murphy, A., Steele, H., Bate, J., Nikitiades, A., Allman, B., Bonuck, K., ... & Steele, M. (2015).

    Group attachment-based intervention: trauma-informed care for families with adverse

    childhood experiences. *Family & Community Health*, 38(3), 268-279.

Murphy, A., Steele, M., Dube, S. R., Bate, J., Bonuck, K., Meissner, P., ... & Steele, H.

    (2014). Adverse Childhood Experiences (ACEs) Questionnaire and Adult Attachment

    Interview (AAI): Implications for parent child relationships. *Child Abuse & Neglect*,

    38(2), 224-233.

Muse, K., & McManus, F. (2013). A systematic review of methods for assessing competence in

    cognitive–behavioural therapy. *Clinical Psychology Review, 33*(3), 484-499.

Rakovshik, S. G., & McManus, F. (2010). Establishing evidence-based training in cognitive

    behavioral therapy: A review of current empirical findings and theoretical

    guidance. *Clinical Psychology Review, 30*(5), 496-516.

Rounsaville, B J, O'Malley, S., Foley, S., & Weissman, M. M. (1988). Role of manual-guided

training in the conduct and efficacy of interpersonal psychotherapy for depression.

Journal of Consulting and Clinical Psychology, *56*(5), 681-8. Retrieved from

http://www.ncbi.nlm.nih.gov/pubmed/3057006

Richards, R., Oliver, J. E., Morris, E., Aherne, K., Iervolino, A. C., & Wingrove, J. (2011).

Acceptance and Commitment Therapy training for clinicians: an evaluation. *The*

Cognitive Behaviour Therapist, *4*(03), 114-121.

Roth, A. D., Pilling, S., & Turner, J. (2010). Therapist training and supervision in clinical trials:

Implications for clinical practice. *Behavioural and Cognitive Psychotherapy, 38*(03),

291-302.

Sanders, M. R., Kirby, J. N., Tellegen, C. L., & Day, J. J. (2014). The Triple P-Positive

Parenting Program: A systematic review and meta-analysis of a multi-level system of

parenting support. *Clinical Psychology Review*, 34(4), 337-357.

Sburlati, E. S., Lyneham, H. J., Mufson, L. H., & Schniering, C. a. (2012). A model of therapist

competencies for the empirically supported interpersonal psychotherapy for adolescent

depression. *Clinical Child and Family Psychology Review*, *15*(2), 93-112.

doi:10.1007/s10567-012-0111-1.

Seng, A. C., Prinz, R. J., & Sanders, M. R. (2006). The role of training variables in effective

dissemination of evidence-based parenting interventions the role of training variables in

effective dissemination of parenting interventions. *International Journal of Mental*

Health, *8*(4), 20-28.

Sholomskas, D. E., Syracuse-Siewert, G., Rounsaville, B. J., Ball, S. A., Nuro, K. F., & Carroll, K. M. (2005). We don't train in vain. *Journal of Consulting and Clinical Psychology*, *73*(1), 106-115.

Simons, A. D., Padesky, C. A., Montemarano, J., Lewis, C. C., Murakami, J., Lamb, K., DeVinney, S., et al. (2010). Training and dissemination of cognitive behavior therapy for depression in adults: a preliminary examination of therapist competence and client outcomes. *Journal of Consulting and Clinical Psychology*, *78*(5), 751-6. doi:10.1037/a0020569

Siqueland, L., Crits-Christoph, P., Barber, J. P., Butler, S. F., Thase, M., Najavits, L., & Onken, L. S. (2000). The role of therapist characteristics in training effects in cognitive, supportive-expressive, and drug counseling therapies for cocaine dependence. *The Journal of Psychotherapy Practice and Research*, *9*(3), 123-30. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3330597&tool=pmcentrez&rendertype=abstract

Skovholt, T. M., & Starkey, M. T. (2010). The three legs of the practitioner's learning stool: Practice, research/theory, and personal life. *Journal of Contemporary Psychotherapy*, *40*(3), 125-130.

Slade, A., Grienenberger, J., Bernbach, E., Levy, D., & Locker, A. (2005). Maternal reflective functioning, attachment, and the transmission gap: a preliminary study. *Attachment & Human Development*, *7*(3), 283-98. doi:10.1080/14616730500245880

Steele, M., Murphy, A., & Steele, H. (2010). Identifying therapeutic action in an attachment-centered intervention with high risk families. *Clinical Social Work Journal*, *38*(1), 61-72. doi:10.1007/s10615-009-0257-6

Stern, S. B., Alaggia, R., Watson, K., & Morton, T. R. (2007). Implementing an evidence-based parenting program with adherence in the real world of community practice. *Research on Social Work Practice, 18:* 543-554. doi:10.1177/1049731507308999

Sternberg, J. (2005). *Infant Observation at the Heart of Training*. London: Karnac.

Stewart, M. O., Raffa, S. D., Steele, J. L., Miller, S. A., Clougherty, K. F., Hinrichsen, G. A., & Karlin, B. E. (2014). National dissemination of interpersonal psychotherapy for depression in veterans: Therapist and patient-level outcomes. *Journal of Consulting and Clinical Psychology, 82*(6), 1201.

Thomas, R., & Zimmer-Gembeck, M. J. (2012). Parent-child interaction therapy: an evidence-based treatment for child maltreatment. *Child Maltreatment*, 1077559512459555. doi:10.1177/1077559512459555

Toth, Sheree L, Rogosch, F. a, Manly, J. T., & Cicchetti, D. (2006). The efficacy of toddler-parent psychotherapy to reorganize attachment in the young offspring of mothers with major depressive disorder: a randomized preventive trial. *Journal of Consulting and Clinical Psychology*, *74*(6), 1006-16. doi:10.1037/0022-006X.74.6.1006

Vivian, D., Hershenberg, R., Teachman, B. A., Drabick, D. A., Goldfried, M. R., & Wolfe, B. (2012). A translational model of research-practice integration. *Psychotherapy*, *49*(2), 143.

Werba, B. E., Eyberg, S. M., Boggs, S. R., & Algina, J. (2006). Predicting outcome in parent-child interaction therapy: success and attrition. *Behavior Modification*, *30*(5), 618-46. doi:10.1177/0145445504272977

Woodhead, J., Bland, K., & Baradon, T. (2006). Focusing the lens: The use of digital video in the practice and evaluation of parent–infant psychotherapy. *Infant Observation*, *9*(2), 139-150. doi:10.1080/13698030600810318

Tables and Figures

TABLE 1.

Demographic Characteristics of GABI One-Day Workshop Participants at Baseline (n=99)

	% (n)	Mean (SD)
Gender		
Female	92.9 (91)	-
Male	7.1 (7)	-
Age	-	34.08 (11.43)
Race/Ethnicity		
White/Caucasian	72.2 (70)	-
Black/African American	4.1 (4)	-
Hispanic/Latin American	12.4 (12)	-
Asian/Asian American	6.2 (6)	-
Other	5.2 (5)	-
Highest Degree Complete		
BA	33.7 (32)	-
MA	27.4 (26)	-
MSW	17.9 (17)	-
Psy.D.	2.1 (2)	-
Ph.D.	12.6 (12)	-
MD	4.2 (4)	-
None	2.1 (2)	-
Degree in Progress		
BA	2.1 (2)	-
MA	16.5 (16)	-
MSW	5.2 (5)	-
PsyD	12.4 (12)	-
PhD	7.2 (7)	-
MHC	1 (1)	-
None	55.7 (54)	-
General Experience (Months)	-	56.20 (90.46)
GABI Experience (Weeks)	-	3.30 (12.95)

TABLE 2.

Demographic Characteristics of GABI Apprenticeship Training Participants at Baseline (n=21)		
	% (n)	Mean (SD)
Gender		
Female	90.5 (19)	-
Male	9.5 (2)	-
Age	-	32.10 (9.73)
Race/Ethnicity		
White/Caucasian	85.7 (18)	-
Black/African American	4.8 (1)	-
Hispanic/Latin American	4.8 (1)	-
Asian/Asian American	-	-
Other	4.8 (1)	-
Highest Degree Complete		
BA	52.6 (10)	-
MA	47.4 (9)	-
Degree in Progress		
MSW	26.3 (5)	-
Psy.D.	47.4 (9)	-
Ph.D.	21.1 (4)	-
None	5.3 (4.5)	-
General Experience (Months)	-	43.29 (89.7)
GABI Experience (Weeks)	-	32.10 (9.73)

TABLE 3.

Demographic Characteristics of Control Groups: Classroom (n=27) and Professional (n=15)				
	Classroom % (n)	Mean (SD)	Professional % (n)	Mean (SD)
Gender				
Female	77.8 (21)	-	100 (15)	-
Male	18.5 (5)	-	0 (0)	-
Age	-	27.15 (5.97)	-	53.00 (14.81)
Race/Ethnicity				
White/Caucasian	74.1 (20)	-	93.3 (14)	-
Black/African American	3.7 (1)	-	-	-
Hispanic/Latin American	3.7 (1)	-	-	-
Asian/Asian American	14.8 (4)	-	-	-
Other	0 (0)	-	6.7 (1)	
Degree Type				
BA	-	-	-	-
MA	-	-	21.4 (3)	-
MSW	-	-	21.4 (3)	-
Psy.D.	-	-	28.6 (4)	-
Ph.D.	-	-	-	-
General Experience (Months)	-	5.04 (11.15)	-	204 (14.8)

FIGURE 1

ACTA Study Design
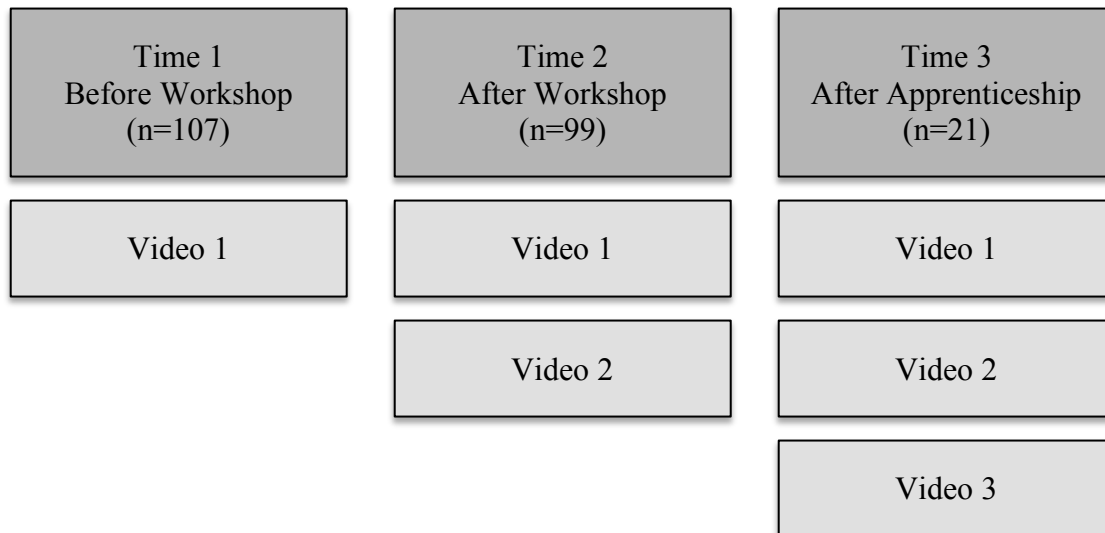
TABLE 4

Average ICCs Among Raters on Reliability Set (n=24 cases)

Code	Average Measures ICC	Single Measures ICC
GABI Principles		
Reflective Functioning	.80*	.57
Emotional Attunement	.75*	.50
Affect Regulation	.77*	.53
Reticence	.88*	.70*
Nurturance	.86*	.67
Group	.77*	.53
Nonspecific Principles		
Own RF	.73*	.46
Nonverbal	.82*	.58
Non-Judgment	.48	.23
Overall Quality	.82*	.60

* ICC ≥ .70

TABLE 5.

Average ICCs Among Raters on Subset of Experimental (Trainee) Group (n=70)

Code	Average Measures ICC	Single Measures ICC
GABI Principles		
Reflective Functioning	.68	.51
Emotional Attunement	.57	.40
Affect Regulation	.80*	.66
Reticence	.85*	.74*
Nurturance	.80*	.66
Group	.64	.47
Nonspecific Principles		
Own RF	.45	.29
Nonverbal	.67	.50
Non-Judgment	.47	.31
Overall Quality	.75*	.60

* ICC ≥ .70

TABLE 6.

Correlations Between Individual ACTA Codes and Overall Quality of Response

	Time 1 Video 1 Overall Quality of Response	Time 2 Video 2 Overall Quality of Response
Reflective Functioning	.58**	.44**
Emotional Attunement	.32**	.56**
Affect Regulation	.42**	.48**
Reticence	.46**	.62*
Nurturance	.45**	.33**
Group Context	.30**	.38**
Own RF	.16	.21*
Nonverbal	.11	.22*
Non-Judgment	.23**	.34*

$* p \leq .05$ $** p \leq .01$

TABLE 7.

Comparison of Mean Ratings on ACTA Across Three Video Stimuli in a Trainee Sample at Baseline

Code	Beads	Dollhouse	Scissors	F-Value	Sig. Post-Hoc
Reflective Functioning	1.82(1.05)	2.71(1.08)	2.84(.83)	7.40**	B<D
					B<S
Emotional Attunement	2.23(1.07)	1.67(.77)	2.38(.90)	6.94**	S>D
Affect Regulation	1.13(.47)	1.63(.80)	2.58(.95)	21.85**	S>D>B
Reticence	2.11(.95)	1.90(.83)	1.85(.92)	.617	None
Nurturance	2.86(.94)	1.65(.91)	1.96(1.00)	12.85**	B>D
					B>S
Group	1.27(.55)	2.12(.82)	2.42(.81)	14.56**	B<D
					B<S
Own RF	2.05(1.05)	2.80(1.17)	1.92(1.06)	6.87**	D>B
					D>S
Nonverbal	2.05(1.13)	2.10(.85)	1.65(1.06)	1.87	None
Non-Judgement	4.59(.67)	4.59(.78)	4.73(.67)	.364	None
Overall Quality	2.66(.68)	2.29(.81)	2.58(1.06)	1.78	None

$* p \leq .05$ $** p \leq .01$ B: Beads   D: Dollhouse   S: Scissors

TABLE 8.

Comparison of Mean Scores on ACTA Across Three Video Stimuli in an Untrained Classroom Sample

Code	Beads	Dollhouse	Scissors	F-Value	Sig. Post-Hoc
Reflective Functioning	1.30 (.72)	2.05 (1.32)	1.69 (1.03)	3.52*	B<D
Emotional Attunement	1.81 (1.04)	11.65 (1.02)	2.44 (1.48)	3.33*	S>D
Affect Regulation	1.11 (.42)	1.41 (.80)	2.33 (1.00)	18.14**	S> B
Reticence	1.59 (.97)	2.19 (1.44)	1.63 (.79)	2.45	None
Nurturance	2.86(.94)	1.26 (.53)	1.30 (.72)	10.23**	B>D
					B>S
Group	1.52 (.98)	1.33 (.73)	1.11 (.42)	2.02	None
Own RF	1.50 (.69)	2.33 (1.33)	1.81 (1.14)	4.03*	D>B
Nonverbal	1.33 (.78)	1.59 (1.01)	1.15 (.60)	2.02	None
Non-Judgment	3.00 (.78)	3.00 (.68)	3.26 (.66)	1.20	None
Overall Quality	2.17 (.73)	2.31 (.76)	2.50 (.66)	1.45	None

* $p \leq .05$ ** $p \leq .01$ B: Beads   D: Dollhouse   S: Scissors

TABLE 9.

Comparison of Mean Scores on ACTA Across Two Video Stimuli in an Untrained Professional Sample

Code	Beads	Dollhouse	F-value	Sig. Post-Hoc
Reflective Functioning	1.60 (.99)	2.53 (1.19)	5.49*	D>B
Emotional Attunement	3.20 (1.08)	2.13 (1.25)	6.27*	B>D
Affect Regulation	1.00 (0.00)	2.60 (1.24)	24.89**	D>B
Reticence	1.93 (1.16)	2.87 (1.41)	3.92	None
Nurturance	2.73 (1.03)	1.33 (.62)	20.31**	B>D
Group	1.80 (.94)	1.53 (.92)	.62	None
Own RF	1.93 (1.28)	2.53 (1.36)	1.55	None
Nonverbal	2.27 (1.22)	1.33 (.62)	6.96*	B>D
Non-Judgment	3.47 (.64)	3.13 (.83)	1.51	None
Overall	2.90 (.71)	3.07 (.88)	.32	None
* p ≤ .05__** p ≤ .01	B: Beads	D: Dollhouse		

TABLE 10.

Paired Samples T-Tests and Effect Sizes: Comparing Mean ACTA Ratings Before and After One-Day Workshop, Using the Same Video and a Novel Video

	Time 1 Video 1	Time 2 Video 1	t	df	Effect Size Cohen's d	Time 2 Video 2	t	df	Effect Size Cohen's d
GABI Principles									
Reflective Functioning	2.55 (1.08)	3.19 (1.19)	4.50**	98	.56	2.71 (1.28)	.97	98	.14
Emotional Attunement	1.98 (.93)	2.86 (1.23)	5.94**	98	.81	2.82 (1.26)	5.45**	98	.76
Affect Regulation	1.77 (.93)	2.40 (1.24)	4.91**	98	.57	1.91 (1.12)	1.16	98	.14
Reticence	1.93 (.88)	3.26 (1.27)	9.71**	98	1.22	3.20 (1.17)	9.13**	98	1.23
Nurturance	2.00 (1.05)	2.70 (1.32)	5.28**	98	1.08	2.74 (1.21)	4.57**	98	.65
Group	2.01 (.86)	2.32 (1.04)	2.98*	98	.32	2.30 (1.11)	2.29*	98	.70
Non-Specific Principles									
Own RF	2.40 (1.18)	2.61 (1.22)	1.41	98	.17	2.29 (1.23)	.65	98	.09
Nonverbal	1.97 (.98)	2.02 (1.22)	.36	98	.05	2.10 (1.04)	.92	98	.13
Non-Judgment	4.63 (.72)	4.86 (.45)	3.46*	98	.38	4.69 (.66)	.73	98	.09
Overall Quality	2.45 (.86)	3.69 (.81)	12.43**	98	1.48	3.56 (.81)	11.23**	98	1.33

* p ≤ .05   ** p ≤ .01

TABLE 11.

Significant Paired Samples T-Tests Comparing Mean ACTA Ratings on the Same Video Before and After Training, By Cohort Year

Cohort t(df)	A 2011-2012	B 2012-2013	C 2013-2014	D 2014-2015
Reflective Functioning	-	-	-	5.60(43)*
Emotional Attunement	-	-	2.47(25)*	6.60(43)*
Affect Regulation	2.42(21)*	-	2.68(25)*	3.76(43)*
Reticence	-	3.27(6)*	7.55(25)*	9.48(43)*
Nurturance	5.70(21)*	-	-	3.12(43)*
Group	-	-	-	2.83(43)*
Overall	8.77(21)*	-	5.21(25)*	9.66(43)*

* p ≤ .05   ** p ≤ .01

TABLE 12.

Significant Paired Samples T-Tests Showing Changes in Mean ACTA Ratings at Three Time Points Across Apprenticeship Training, Using Same Videos and Novel Videos

Code	TIME 1 → TIME 2		TIME 2 → TIME 3				TIME 1 → TIME 3		
			Video 1 Comparisons		Video 2 Comparisons		Same	Novel	Novel
	Same	Novel	*Same*	*Novel*	*Same*	*Novel*	(Video 1)	(Video 2)	(Video 3)
RF	-	-	-	2.52(14)	-	-	-	-	-
EA	-	-	-	-	-	-	-	-	-
AR	-	-	-	-	2.23(14)	2.52(14)	-	-	3.48(20)
RET	2.67(13)	3.73(13)	-	-	-	-	2.65(20)	2.91(20)	-
NUR	-	-	-	-	-	2.70(14)	2.28(20)	-	-
GRP	-	-	4.58(14)	-	-	-	-	-	-
OWNRF	-	-	-	-	-	-	-	-	-
NONV	-	-	-	-	-	-	-	-	-
NONJUDG	-	-	-	-	-	-	-	-	-
OVRALL	5.10(13)	4.16(13)	-	-	-	-	3.70(20)	3.93(20)	5.12(20)

RF: Reflective Functioning   GRP: Group
EA: Emotional Attunement   OWNRF: Own RF
AR: Affect Regulation   NONV: Nonverbal
RET: Reticence   NONJUDG: Non-Judgmental Stance
NUR: Nurturance   Overall Quality

TABLE 13.

Means and Standard Deviations of ACTA Ratings in Apprenticeship Sample, Across Three Videos and Three Time Points

	Time 1 Video 1 (n=21)	Time 2 Video 1 (n=14)	Time 2 Video 2 (n=14)	Time 3 Video 1 (n=21)	Time 3 Video 2 (n=21)	Time 3 Video 3 (n=21)
GABI Principles						
Reflective Functioning	2.90(1.14)	2.36(1.15)	3.36 (2.27)	3.42(1.16)	3.29(1.01)	3.52(1.12)
Emotional Attunement	2.23(.89)	2.57(1.45)	2.43(1.16)	2.57(1.12)	2.09(1.26)	2.76(1.13)
Affect Regulation	1.71(.85)	1.93(1.54)	1.64(1.01)	2.00(1.27)	2.24(1.09)	2.90(1.22)
Reticence	2.38(1.11)	3.21(1.48)	3.64(1.01)	3.19(1.21)	3.33(1.11)	2.81(1.29)
Nurturance	2.10(1.09)	3.00(1.47)	1.92(1.14)	2.71(1.31)	2.76(1.41)	2.29(1.19)
Group	2.05(.86)	1.57(.94)	2.07(1.21)	2.38(.67)	2.05(.86)	2.13(.91)
Non-Specific Principles						
Own RF	2.90(1.3)	2.86(1.46)	2.25(1.45)	2.71(1.45)	2.38(1.28)	2.00(1.34)
Nonverbal	2.00(.89)	2.07(1.44)	2.07 (1.00)	1.86(1.01)	2.00(1.14)	1.62(.22)
Non-Judgment	4.57(.68)	4.93(.27)	4.71(.61)	4.81(.51)	4.86(.34)	4.90(.30)
Overall Quality	2.86(.79)	3.68(.58)	3.64(.63)	3.62(.74)	3.62(.74)	3.86(.73)

    * p ≤ .05   ** p ≤ .01

TABLE 14.

Effect Sizes (Cohen's D) for Significant Paired-Samples T-Tests in Apprenticeship Sample, Across Three Time Points and Videos

Code	TIME 1 → TIME 2		TIME 2 → TIME 3				TIME 1 → TIME 3		
			Video 1 Comparisons		Video 2 Comparisons		Same	Novel	Novel
	Same	Novel	*Same*	*Novel*	*Same*	*Novel*	(Video 1)	(Video 2)	(Video 3)
RF	-	-	-	1.09	-	-	-	-	-
EA	-	-	-	-	-	-	-	-	-
AR	-	-	-	-	.79	1.07	-	-	1.13
RET	.84	1.47	-	-	-	-	.69	.85	-
NUR	-	-	-	-	-	.68	.24	-	-
GRP	-	-	.92	-	-	-	-	-	-
OWNRF	-	-	-	-	-	-	-	-	-
NONV	-	-	-	-	-	-	-	-	-
NONJUDG	-	-	-	-	-	-	-	-	-
OVRALL	1.27	1.18	-	-	-	-	.99	.99	1.31

Appendix 1

APPLICATIONS OF CLINICAL TRAINING ASSESSMENT (ACTA)

I. APPLICATION OF GABI CONCEPTS

Reflective functioning

Reflective Functioning (RF) is the ability to think about the thoughts, feelings, and intentions of another person. It is the hallmark objective of GABI, to which all of the clinical goals and tools are linked.

In order to achieve a score of 3 or higher, response must make clear an effort on behalf of clinician, parent, or child to understand their own mind or the mind of another. Most responses scoring this high will comment on the therapists' attempts to help the parent understand the mind of his/her child. Alternatively, noting therapists' attempts to help children better understand their own thoughts and feelings also scores a 3 or higher.

Mentions of mentalization and theory of mind would get a minimum score of a 2/3, and with more description could get a higher score.

1	2	3	4	5
no mention or incorrect application of concept. Participant does not understand or apply concept.		alludes to concept but does not provide details or demonstrate definitive understanding. Think participant understands concept.		demonstrates thorough understanding and describes concept in appropriate detail (can be achieved by a single score of 5, or multiple scores of 3/4). Certain participant understands concept.

Example of high score:
The therapist did not intervene physically, but instead tried to get the boy to recognize cues from the other children, better understand what his behavior looks like to others (mad), and to recognize a causal link between his behavior and others negative reactions.

Emotional mirroring

Emotional Mirroring is a critical skill through which therapists try to engage parents in a way that facilitates recognition and understanding of their children's emotional states, conveying to the child a sense of being understood.

This code has more to do with affectively matching the child or parent or meeting them where they are at emotionally, rather than purely labeling feelings, which falls under Reflective Functioning.

In order to achieve a score of 3 or higher, response must describe a parent, therapist, or child, attempting to emotionally relate to another. Responses that mention labeling affect score around a 3, but with more detail can score higher. Responses that note shared joy or sadness, where one person is reflecting the emotions of another, would score 3 or higher, depending on level of detail. References to therapist validating patients' emotions may receive a score of 3, depending on the amount of detail.

. 1	2	3	4	5 . .
no mention or incorrect application of concept. Participant does not understand or apply concept.		alludes to concept but does not provide details or demonstrate definitive understanding. Think participant understands concept.		demonstrates thorough understanding and describes concept in appropriate detail (can be achieved by a single score of 5, or multiple scores of 3/4). Certain participant understands concept.

Affect regulation

Affect Regulation is achieved by therapists who are sensitive to the expression of volatile feeling states. Parents are able to develop an understanding of themselves and their children, turning volatile expression into emotional states that can be more easily understood.

In order to achieve a score of 3 or higher response must reference a heightened emotional state (though not necessarily negative), that gets "turned down," to a more manageable level.

. 1	2	3	4	5 . .
no mention or incorrect application of concept. Participant does not understand or apply concept.		alludes to concept but does not provide details or demonstrate definitive understanding. Think participant understands concept.		demonstrates thorough understanding and describes concept in appropriate detail (can be achieved by a single score of 5, or multiple scores of 3/4). Certain participant understands concept.

Reticence

Reticence involves waiting to intervene, gives parents and children the space to discover their own feeling states and enhance self-efficacy. Both the therapists and the parents practice reticence in order to have access to important information that would otherwise be lost.

In order to achieve a score of 3 or higher, response must note that a parent or therapist holds themselves back, is non-interfering, or some other form. Higher scores show a clear valuing of holding back, often accompanied by what it allowed patients to achieve.

. 1	2	3	4	5 . .
no mention or incorrect application of concept. Participant does not understand or apply concept.		alludes to concept but does not provide details or demonstrate definitive understanding. Think participant understands concept.		demonstrates thorough understanding and describes concept in appropriate detail (can be achieved by a single score of 5, or multiple scores of 3/4). Certain participant understands concept.

Nurturance

Nurturance involves nurturing both the parents and children, and promotes the nurturance of the children by their parents. Talking about safety, providing for children's basic physical needs also falls under nurturance, but depending on whether it is elaborated on may score on the lower end.

To achieve a score of 3 or higher, response should indicate therapists' efforts to foster the parent/child relationship, or the parents' nurturing of their children (i.e. providing a safe haven). However, the use of attachment terms such as "secure" or "safe haven" without clear description of the situation would not score high.

. 1	2	3	4	5 . .
no mention or incorrect application of concept. Participant does not understand or apply concept.		alludes to concept but does not provide details or demonstrate definitive understanding. Think participant understands concept.		demonstrates thorough understanding and describes concept in appropriate detail (can be achieved by a single score of 5, or multiple scores of 3/4). Certain participant understands concept.

Example of high scores:
Therapist includes mother, turns toward her, encourages child to ask mom to help- as a way to include mom in child connection.

Group

Group Context provides important sources of social support to the parents and facilitates peer relationships amongst the children, combating the inherent social isolation faced by the participants.

Recognition that a key therapeutic moment involved inter-family interactions, rather than strictly parent-child or patient-therapist interactions.

Mention of children playing well together may receive a score of 2. However, in order to receive higher scores, response must show understanding of the therapeutic benefit of the group or what could be achieved only through the presence of peers.

. 1	2	3	4	5 . .
no mention or incorrect application of concept. Participant does not understand or apply concept.		alludes to concept but does not provide details or demonstrate definitive understanding. Think participant understands concept.		demonstrates thorough understanding and describes concept in appropriate detail (can be achieved by a single score of 5, or multiple scores of 3/4). Certain participant understands concept.

Examples of high scores:
The children support each other, share toys, laugh together, model actions for each other, copy from one another.

III. OVERALL EVALUATION

Valence and Non-Judgment

. 1	2	3	4	5 .
describes scenario with evaluations of "good" and "bad"		some references to evaluation but these are minimal and scenario still described mostly in terms of observations		describes scenario in terms of GABI concepts no evaluation of good/bad

Discussing interactions in group in terms of manners, safety, politeness, hygiene, also may fall under the category of assigning valence/judgment.

Emphasis on Nonverbal behavior

. 1	2	3	4	5 .
No mention of any nonverbal behavior		Mentions of nonverbal behavior, without attempt to understand meaning		Nonverbal behavior is the focus of the description, and linked to underlying thoughts, feelings, and intentions.

Note
This can be obtained in 2 ways: 1.) description includes a lot of nonverbal "clinician was on the ground";
2.) explanation shows understanding of the meaning of nonverbal behavior, "nonverbal behavior showed alliance with parent."

Ability to Use RF in Evaluation of Situation

. 1	2	3	4	5 .
No attempts to put self in shoes of parent, child or clinician AND no reflection on own mental state while watching video.		Some brief attempts to consider others' mental states, or reflection on own mental state while watching video (e.g. I felt frustrated...)		Trying to understand mental states of people in the video, or careful reflection on own thoughts or feelings.

Note: Higher scores may demonstrate developmental change in perspective (i.e. The first time I saw this video I thought the therapist was not active enough, but now I see her holding back as purposeful), or acknowledging conflicting feelings (i.e. on the one hand I felt bad for the mom because she seemed so anxious, on the other hand I was mad at her for not being more engaged). Credit is also given to more complex reflection on the mental states of children or parents in the video (i.e. linking mental states to behavior, acknowledging mental states as opaque or susceptible to disguise).

Example of high score:
The girl with the scissors may have had an unfair advantage sitting in the therapist's lap, as it enjoying her "protection." She behaved very passively after saying she would hand them over when she was done and was not using the scissors. Couldn't tell if she was scared of the other little girl?

Overall Quality of Observation/Evaluation

. 1	2	3	4	5 .
Extremely brief - close to left blank		Answered the questions using some GABI concepts but with minimal insight. Think participant understands GABI concepts, but lacks detailed descriptions, so may be less certain.		Provides full description of observations, which often include both verbal and nonverbal, and some RF. Responses are vivid, involve explanations, and are easily understood by the reader. Certain participant understands and values GABI concepts.