

# Impact of Therapists' Skill on Effectiveness of Parenting Groups for Child Antisocial Behavior

January 15 2008

Stephen Scott<sup>1</sup>, Anne Carby<sup>1</sup>, and Alison Rendu<sup>2</sup>

*1 King's College London, Institute of Psychiatry  
2 University College London*

## Abstract

This study investigated the influence of therapist skill on the effectiveness of a manualized parenting program for child antisocial behavior. A reliable instrument with 10 scales was developed to measure skill. 15 parenting groups were observed, from a controlled trial involving 90 clinically referred, disadvantaged children aged 3-8 with severe antisocial behavior. Child outcome was assessed by interview. The skill level of therapists running the groups was found to be a significant predictor of effectiveness ( $p < 0.05$ ); each increase in skill of one point on a five point scale led to an improvement of 0.58 sd in child outcome. The impact was not reduced after controlling for severity of child problems, therapists' degree of organization, quality of alliance with parents, or number of sessions attended. In subsidiary analyses, nurses and therapists with mental health training showed more skill; therapist gender and age had no effect. In conclusion, therapist skill appears to have a major influence on outcomes. Clinicians and service managers need to ensure that training and supervision procedures fully develop staff skill.

**KEYWORDS** therapist skill, effectiveness, therapeutic alliance, parent training, child antisocial behavior

Discovering what makes psychological therapies effective is important theoretically in order to understand mechanisms of change in human relationships and behavior, and practically in order to enable the successful delivery of services. Two key aspects of the therapeutic process already identified as contributing to effectiveness are the quality of the alliance with the client, and the therapist's degree of adherence to the treatment protocol. However, many have argued (eg the classic paper by Waltz, Addis, Koerner and Jacobson, 1993) that the process of successful therapy goes beyond the sympathetic appliance of procedures according to a manual, and involves a considerable element of skill. This study aimed to develop a reliable measure of skill and investigate the extent to which it affected outcomes in a group-based parenting program for severe antisocial behavior in children aged 3-8 years.

Antisocial behavior is an important target for psychological therapies since it is the commonest psychological disorder in children and adolescents (Costello, Egger and Angold, 2005; Ford, Goodman and Meltzer, 2003). Indeed violence in young people has been termed an epidemic by the Surgeon General (Surgeon General, 2001); both he and the National Institute for Mental Health (NIMH, 2000) have called for more research into what makes interventions work. Parenting programs offer the most promising intervention, and are widely used (Kazdin, 2005). Whilst there are literally hundreds of randomized controlled trials attesting to their effectiveness (Kazdin, 2005), there are rather few studies of what contributes to effective delivery.

Organizing services to be as effective as possible is important for the success of individual clinics, and for the wellbeing of total populations. With the prevalence of moderate or severe psychological disorders at around 10-15% of the population in adulthood and childhood (Kessler, Demler, Frank, Olfson, Pincus et al 2005, Costello et al 2005), getting effective services to the majority of those in need presents a major challenge. With the

increasing emphasis on dissemination of effective therapies, how much training therapists need to be skilful and effective is a pressing question. In the USA, the proportion of adolescents and adults with a mental health disorder receiving treatment in 1990-1992 was 20%, and this rose to 33% by 2001-2003 (Kessler et al 2005). In the UK 25% of children get any specialist treatment (Ford et al 2003) - thus in both countries the majority make do without treatment. This has led to calls to expand the mental health workforce for both children and adults. For example, in the UK, the gross shortfall of therapists to deliver government recommended guidelines for CBT for depression and anxiety has led to the proposal to set up a central agency to train and oversee 10,000 new staff (Layard, 2006), and the lack of suitable workforce to deliver evidence-based programs for child antisocial behavior has led the Prime Minister to announce the setting up of a Parenting Academy to train professionals (Respect Task Force 2006). However, such proposals beg the question as to what level of skill therapists need. Currently, when children and adolescents do receive treatment in real life conditions, it is often ineffective (Weisz, Donenberg, Hans, and Kauneckis, 1995; Bickman, Lambert, Andrade, and Penaloza, 2000). Therefore simply to expand the workforce as they are may not improve outcomes much, if at all. Understanding the role of therapist variables may help, although in the past they have not been given much attention in reports of therapy: Moncher and Prinz (1991) reviewed 359 therapy outcome papers published up until 1989 and found that only a quarter gave any evidence of using a protocol and of these only half reported how much training therapists received. To discover what enables the dissemination of more effective treatments, Weisz and colleagues have called for better collaboration between researchers and therapists (Weisz, Chu & Polo 2004).

*Conceptual issues* There are three concepts that are prominent in theories of how therapy is delivered by particular therapists: first, the therapeutic alliance, second fidelity/adherence (hereafter called fidelity) and third, skill /competence (hereafter called

skill). The alliance could be defined as how well, both personally and collaboratively, client and therapist get on together and agree on goals. The ability to form an alliance with clients has been proposed to be a general ability that may apply to some extent across a number of therapeutic modalities. It can be measured a number of ways, varying from simple client measures of trust in and liking for the therapist to more elaborate questions about shared goals assessed by both clients and therapists. These methods have been reviewed in relation to child therapies by Green (2006). Fidelity concerns the extent to which the therapist follows the actions prescribed in the manual. Like the alliance, it can be relatively easily measured by using a checklist or questionnaire, and helps determine whether or not the therapy is being given the therapy as specified or whether the therapist is deviating and doing something different. However, as argued by Waltz et al (1993), skill is a distinct concept, and we argue that it can be conceptualized as subsuming adherence. Put simply, whereas fidelity concerns *whether* the therapist carried out the prescribed actions, skill concerns *how well* the therapist performs the actions. We argue that it is impossible to do a complex task skillfully without first carrying out the right steps: it would be impossible to dance the waltz expertly without first moving ones feet in the prescribed way; a heart surgeon could not carry out a transplant expertly without following correct procedures. In the psychological domain, two therapists for example giving CBT might both explore cognitions, challenge them, reframe them, and issue homework. However, the more skilled one might do this in a more sensitive way with greater complexity, and so characterize the client's mental state more accurately and be more proficient in overcoming barriers to doing homework, thus leading to more change. Skill implies being able to deliver the intervention well in a range of conditions and contexts, including challenging ones. In this conceptualization, skill is the overarching concept that is key to how well therapy is delivered, and therefore could, at a service level, be the ultimate goal of professional training. If this is so, it should relate well to the effectiveness of therapy.

*Empirical studies relating therapist variables to client outcomes* There are very few studies of skill and outcomes in children and adolescents, perhaps because it is harder to measure than fidelity or alliance. Forgatch, Patterson, & DeGarmo (2005) developed an observer-based instrument to assess therapist variables that included skill in a parenting program for recently remarried mothers. They carefully measured knowledge, structure, teaching skill, clinical skill and overall effectiveness. The outcome they assessed was the proximal one of observed parenting practices, rather than child behavior; greater therapist skill led to more change in parenting. In the adult literature, Shaw, Elkin, Yamaguchi, Olmstead, Vallis et al (1999) found that in for clinician rated outcome of depression, skill accounted for 15% of outcome variability, alliance quality accounted for 5 %, but fidelity had no effect. Another study of cognitive therapy outcomes for depression found that both alliance and skill were related to outcomes (Trepka, Rees, Shapiro, and Barkham 2004), and a parasuicide trial (Davidson, Scott, Schmidt, Tata, Thornton, et al, 2004) therapist competence predicted better outcomes on depression, anxiety and functioning as rated by clinicians. Thus in the relatively few studies that have measured therapist skill, it has proven an important determinant of client outcome.

There have been far more studies of the impact of alliance quality. A meta-analysis of children and adolescent research found an overall effect size of 0.21 standard deviations (sd) which held up across treatment types, and across youth, parent, or family approaches (Shirk and Karver 2003). Kazdin, Marciano and Whitley (2006) found about 7% of the variance in outcome of treatment for antisocial children was related to the alliance, though findings varied by informant. Likewise, in a comparison of Cognitive Behavioral Therapy (CBT) and Family Therapy for adolescent behavior problems, youth rated alliance had no effect on outcomes when using CBT, but that parent rated alliance was associated with better outcomes

when using family therapy (Hogue, Dauber, Stambaugh, Cecero, and Liddle, 2005). Meta-analyses of adult studies (Horvath and Symonds, 1991; Martin, Garske and Davis, 2000; Beutler, Malik, Alimohamed, Harwood, Talebi et al, 2004) found that variations in the quality of the alliance accounted for effect sizes from minimal to 0.22 sd in outcome across several disorders and treatment types. Zuroff and Blatt 2006) in a careful study confirmed the importance of the alliance across four different treatment conditions for depression.

In contrast to the clear impact of therapist skill and alliance quality, studies of fidelity have found more equivocal effects. With children and adolescents, an early study of Multisystemic Therapy (MST) for delinquency by Henggeler, Melton, Brondino, Scherer, and Hanley (1997) compared a total of 15 parent, therapist, and youth rated fidelity scales with 7 youth outcomes, and found significant effects for fidelity for only 7 out of 42 parent rated associations, 4 out of 35 therapist rated, and 1 out of 28 adolescent rated; moreover some effects went the 'wrong' way, with better fidelity leading to worse outcomes. The same group (Huey, Henggeler, Brondino, and Pickrel, 2000) found that when they used a latent variable approach, therapist rated fidelity improved family functioning and parent monitoring, both of which in turn reduced youth delinquency, but that parent and youth rated fidelity had no effect. This last finding could be because it requires a therapist to appreciate the complexity of fidelity, and also because therapists working across cases will be more consistent in their ratings than parents and youths, who may differ widely in their rating of the same phenomena. Studies with adult clients have also been equivocal. Thus one study found no relationship between treatment fidelity and outcome for family therapy for bipolar disorder (Weisman, Tompson, Okasaki, Gregory, Goldstein et al, 2002), and another found no relation between adherence and purging frequency for either CBT or IPT for bulimia (Loeb, Wilson, Labouvie, Pratt, Hayaki et al, 2005). These somewhat modest findings for the role of fidelity raise the question whether applying the treatment according to the manual

alone is sufficient to bring about change and add some empirical support to the conceptual notion that skill may be more central to bringing about change.

A final important consideration when examining what influences therapist behavior is client characteristics. Waltz et al (1993) argued strongly that skill is not a static characteristic, but is influenced by context. A calm client agreeing with the therapist may be much easier to work with skillfully than one who is challenging, angry or withdrawn. Schoenwald, Letourneau and Halliday-Boykins (2005) examined several youth characteristics but found only one, psychosocial functioning, affected therapist adherence. Nonetheless, this domain merits further exploration in child therapies.

For completeness, we should note that a number of general factors that need to be in place before a therapy can be most effective; if they are not present, even an intervention proven in trials may not work. These include (1) having good management structures that support an outcome driven culture (Kam, Greenberg and Walls 2003); (2) choosing interventions that are based on experimentally confirmed theories, and (3) that are empirically supported by evidence of effectiveness from controlled trials (Kendall and Southam-Gerow 1995, Chambless & Hollon 1998); (4) ensuring adequate exposure or ‘dosage’ of therapy (Mihalic, 2000).

*The current study* The present study aimed to build on and expand the studies cited above by investigating the impact of skill on child outcomes using independently observed ratings. As far as we are aware no empirical study has yet addressed this issue. There were four specific aims. First, to develop a measure of how skillfully the program was delivered that subsumed adherence, based on the notion elaborated above that one cannot deliver a treatment skillfully if its key elements are missing. Second, to examine the relative contribution of therapist skill and alliance to outcome success. Third, to examine whether client characteristics predicted

therapist skill, and fourth to investigate whether therapist age, gender or professional training predicted their degree of skill. Consistent with recommendations by Waltz et al (1993) the theoretical underpinnings of the instrument development were 1) that the instrument should be developed by clinicians with a deep understanding of the nuances of the therapy in question 2) skill should therefore be measured by raters who are clinicians rather than clients 3) skills should be directly observed by independent raters rather than self-rated by therapists 4) skill should be observed across a reasonably wide range of therapeutic encounters 5) characteristics specific to the therapy modality should be emphasized alongside more general skills. Additionally 6) the instrument should reasonably short so it could be used in further evaluations of therapy. Finally, 7) as the therapy was given by two therapists working together, it seemed logical to measure the skill of the dyad in delivering the therapy. During a group session, the final expression of the therapy is determined by a combination of individual therapists skills and how well they work together: if one therapist was working skillfully but the other got it wrong, this would be less expert than if they worked together harmoniously to deliver a seamless performance.

## Method

*Trial* The controlled trial compared a parenting program, the Incredible Years (Webster-Stratton and Reid 2003), with wait-list controls, and found that the effect size was 1.1 sd on an intention to treat analysis (reference withheld for anonymity). Parents of 73 children were allocated to intervention and 37 to wait list, of whom 21 also subsequently received the intervention, giving a total of 94 allocated parenting groups; 90 of these were successfully followed up. The children were aged 3-8 years (mean 5.5) and were referred with antisocial behavior to 5 Child and Adolescent Mental Health Service clinics in Southern England. Their mean antisocial behavior score was on the 98<sup>th</sup> percentile, ADHD symptoms on the 90<sup>th</sup>

percentile, and emotional symptoms on the 78<sup>th</sup> percentile, using the outcome measures described below. The families were disadvantaged: 45% were single parents (England mean 9%), 20% ethnic minority (9%), 51% mother gained no qualifications at school or after graduating (13%), 22% had a total household weekly income < \$300/£175 (5%), and 50% of the children were eligible for free school meals (18%). Using a permuted block design, depending on when the children were referred, the children were either allocated to the treatment arm of the trial immediately, or were allocated to the control arm, in which case they received treatment after a 4-6 month wait-list period (reference withheld for anonymity).

Parents of the children attended an ‘Incredible Years’ parenting group. This program has one of the strongest evidence bases in the field for its efficacy and effectiveness and is designated a model program by the Blueprints series (Webster-Stratton and Reid 2003). The principles are described in Table 1. The program covers play, praise and rewards, limit setting, and handling misbehavior. In each session, the two group leaders showed videotaped scenes of parents and children together, which depict “right” and “wrong” ways of handling children. Parents discussed their own child’s behavior and were supported while they practiced alternative ways of managing it. Each week tasks were set for parents to practice at home and telephone calls made to encourage progress. 15 groups were held over 14-18 weeks. The parents of four to nine children were seen for two hours each week; the children did not take part, and no other treatment was given to them. Mean attendance was 10.3 sessions (range 0-18 sessions), 76/90 participants (84%) attended at least 50% of the sessions offered.

(TABLE 1 ABOUT HERE)

*Therapists* Each group was run by two leaders who stayed the same for the duration of the group. There were 13 therapists, who over the three years of the trial worked in various combinations to deliver the 15 groups. All held jobs in their local mental health clinic service.

There were four child psychiatrists, three child mental health nurses, and one family therapist, all with over 15 years of experience; one child clinical psychologist and one occupational therapist with 5 years experience each; and one speech therapist, one developmental child psychologist, and one psychology graduate with no child mental health training or experience. The age range was from 28 to 48 years old (mean 40). Most of the centers had a good reputation, and in general the staff would be as good as or better than the norm.

Therapists were given initial training over three months, then had to observe a group and attend weekly supervision before running a group. Three therapists were accredited as group leaders by the program developer during the trial, and each subsequently went on to achieve higher level ('mentor') accreditation denoting a higher level of expertise. Intervention sessions were videotaped and weekly supervision meetings were continued to develop skill.

*Measures* Because of our interest was in the effectiveness of treatment delivery, we chose as our outcome measure child antisocial behavior as this was the target of the intervention.

The primary outcome was the Parent Account of Child Symptoms interview (Taylor, Schachar, Thorley and Wieselberg, 1985), a well validated semi-structured interview that uses investigator-based criteria to assess the frequency and severity of disruptive behaviors such as fighting, destruction, and disobedience; scores are strongly predictive of later psychosocial outcome (Taylor, Heptinstall, Chadwick, and Danckaerts, 1996). The kappa inter-rater reliability statistic on 20 randomly selected interviews was 0.84 for the antisocial behavior/conduct problems scale. Interviews were held with mothers one to two months before commencing the intervention, and after it ended, five to seven months later. Change in child antisocial behavior, rather than in parenting, was chosen as the outcome of interest for the original study and for this one, since it was the over-riding objective of the trial and of the families who sought help. The focus of this paper concerns the main outcomes of concern to

commissioners of services; it is a separate question beyond the scope of this paper whether parenting was changed and if in turn it mediated child behavior change.

Alliance was measured using by selecting five questions that tap into this construct from a Consumer Satisfaction Questionnaire (Webster-Stratton 1989). Each was rated on a seven point scale. The questions covered the client's assessment of the therapist's interest and concern in the client and the client's child, the therapist's helpfulness, and the client's feelings for the therapist, from dislike him/her very much, up to like him/her very much. Cronbach's alpha was 0.84. Range 4.8 -7.0, mean 6.2, sd 0.56.

Attendance was measured as the percent of sessions attended (range 6-100, mean 72, sd 26).

### *Development of Therapist Skill Scale*

The authors work in a unit that has been using the IY program for over 15 years. The first author is accredited at the highest level ('trainer') of whom at the time of the study there were five in the USA and three in Europe. This entailed extensive meetings with the program developer to ensure deep understanding of the nuances of the program gained from repeated training sessions where videotapes of a wide range of therapists were discussed and supervision given to participants. The aim of the study however was to develop an instrument that could be applied by a wider range of clinicians. The other authors who coded the tapes are both clinical child psychologists who undertook the authorized training for the IY parenting program, and then ran groups as part of a subsequent trial over three years, during which they received weekly supervision from the first author.

The new coding instrument was derived from concepts in the manuals for the IY program and forms for the evaluation of group leaders (Webster-Stratton 2004). There have been a number versions of the forms over time and they cover four overarching domains (A) personal skills displayed in relation to participants (group process skills, leadership skills,

relationship building skills, knowledge); (B) skill in the use of specific techniques (therapist methods); (C) degree of organization; and (D) observed impact on participants (parents' responses).

The first step was to devise criteria to cover constructs in each of the domains. Some constructs from the program had to be excluded as they could not be measured from tape, for example whether the groups started and finished on time, or whether leaders were advocates for the parents outside of the sessions. Then program criteria were combined if they appeared too similar to distinguish during coding or seemed to be tapping a single construct. For example, "build rapport with each member of the group", "encourage everyone to participate" and "view every member of the group as equally important and valued" were combined to create one variable that was labeled "encourage participation".

Over a period of four months all three coders met on a weekly basis to identify and define behaviors relevant to the skill constructs that would be clearly observable during treatment sessions. Preliminary testing of variables was carried out using videotapes of sessions from a different study. Further adaptations were made during this time as it became clearer which variables could not be coded through video analysis and which constructs could be combined as they were measuring similar criteria. This led to 14 constructs in the version that was taken forward for reliability testing. A glossary of terms was then constructed, which was explicit in its definition of optimal desired therapist behavior. There were seven consecutive versions of the glossary. To give an idea of criteria, under the item 'Encourage Participation' optimal performance is defined as giving every parent a chance to feedback; asking what people think; bringing everyone into the discussion; having eye contact with all; using inclusive language and non-verbal behavior; building rapport with each member of the group, and managing dominant or reticent parents.

The definitions in the glossary were about skill quality, e.g. the efficient use of vignettes to bring out key parenting issues and get participants to relate these to their own predicaments, but where necessary also covered issues of quantity, e.g. whether sufficient vignettes were shown during the session. It took into account the difference in leadership style expected at different stages of the program. For example, during most sessions group discussion should be facilitated by using open ended questions to encourage parents to contribute, but certain topics require more didactic teaching to get detailed ideas across, e.g. how to use Time Out. The overall combination of the two leaders' skill was coded, rather than assess each separately, since tasks were divided during sessions, and the combination reflected the quality of the overall delivery of the program as experienced by the participants. The leaders' behavior was coded independently from the participants' reactions, as recommended by Hogue, Liddle and Rowe (1996). To guard against halo effects, examples of observed practice were written down as evidence to support ratings.

A five point rating scale was devised for each variable, with each point defined in the glossary. The five performance levels were 1 = not at all, 2 = poor, 3 = satisfactory, 4 = good, 5 = excellent. For example, for the variable 'Review homework and give feedback and praise,' a therapist would score '1' if they made no acknowledgement of the effort made by parents over the previous week. A rating of '2' would represent a review of homework, but only rare attempts to give feedback or praise parents' effort. The therapist may fail to accept responsibility for any homework misunderstanding, or makes no attempt to solve homework difficulties. A rating of '3' would apply if the therapist gave feedback and reasonably sensitive used praise to most parents; they may also have used parent experiences to highlight key principles, or attempted to problem solve homework difficulties. A '4' would reflect a good level of detail in responsive probing what happened during the week, thoughtful acknowledgement of feelings involved, appropriate positive comments and praise offered to

each parent, and good linking of parents' experiences to program principles. A maximum rating of '5' would mean that the therapist reviewed each parent's week thoughtfully, using "show me" role plays to clarify matters where necessary, showed understanding of the parents' predicaments, and used this in a personal way to offer feedback and help devise apt strategies for them, with encouragement; made insightful and clear links between parent experiences and key program principles; and developed alternative strategies to unblock barriers so parents could complete homework and enact the principles at home.

*Reliability* Inter-rater reliability was tested using 16 tapes from an independent sample drawn from the same population, using the Intra-Class Correlation Coefficient (ICC), with a one-way random effects model taking a single measure. Of the 14 constructs, ten variables met the criterion for reliability (ICC >0.5) as shown in Table 2. Four constructs were not reliable and were dropped: theoretical model explained (referring to the child's viewpoint, making child development principles explicit, and showing a coping model;  $r = .42$ ), being directive when required (leading discussion, confronting and challenging negative beliefs, and predicting relapse;  $r = .37$ ), response to client's issues (validate feelings, reframe ideas, use open-ended questions,  $r = 0.39$ ) and structuring beginning of session (restate ground rules, set agenda;  $r = 0.31$ ).

For the main study, four sessions from each group were rated to ensure wide coverage of therapeutic contexts, one from each stage of the program: play, praise and rewards, limit setting, and time out. The mean score for the four sessions was used, which will have increased the reliability of the measure in comparison to the values for single sessions given in Table 2. Overall, there was good representation of key concepts from each domain of the program. To prevent coder drift, regular meetings were held to monitor inter-rater reliability by duplicate coding of every fifth videotape. Both coders were blind to the treatment outcome data.

(TABLE 2 ABOUT HERE)

*Factor structure* Results of the factor analysis of the ten therapist skill scales are presented here in the methods section as they pertain to the instrument. Principle Components Analysis with Varimax rotation using Kaiser normalization was undertaken using SPSS 12 (Norusis Corporation 2003) on 76 group sessions, the 60 sessions rated from the trial plus the 16 used for reliability testing. The Kaiser-Meyer-Olkin measure of sampling adequacy was acceptable at 0.57, and Bartlett's Test of Sphericity was highly significant ( $p < 0.001$ ). Two factors with Eigenvalues greater than 1 were extracted and accounted for 50% and 21% of the variance respectively (Table 2). Factor one was named Therapist skill; factor two was named Therapist organization.

The therapist skill factor included all the four scales in domain A, skills in personal relationships, and both the scales in domain B, skills in using specific methods. These six scales all measure the leaders' ability to run the therapy sessions in a collaborative, but controlled way that is emotionally sympathetic to the individual needs of participants while still covering the methods and content of the program in a way that maximizes their chances of changing how they relate to their children. The factor also included the scale from domain D, participant reaction. Because the reaction of participants to the program might be considered conceptually different from therapist skill, as it is in a sense a measure of therapy outcome rather than process, the factor analysis was re-run without it. However, a similar two factor solution was obtained, and in subsequent analyses the skill factor had similar associations whether participant reaction was included or not, so it was left in. The therapist organization factor included all three scales from domain C. The scale called 'practical arrangements' included whether parents were seated in a semi-circle with both group leaders among them, rather than bunched together at one end of the room; the 'materials prepared' scale covered evidence that before the group began leaders had set chairs up, got flip charts

ready, had handouts to hand etc; the 'issue homework' scale covered time spent at the end of the setting tasks to be carried out at home during the week.

*Statistical analysis* Results from the original clinical trial indicated that child age and hyperactivity predicted outcome independent of treatment; these variables were therefore included as covariates in analyses testing therapist effects. We also include parental attendance as a covariate because of the previously reported links between number of sessions and treatment effects. Family factors (maternal education, maternal depression and household income) did not predict outcome independent of treatment and are therefore not included as covariates.

The intervention consisted of two group therapists who worked together closely as a team during the treatment period. As a result, the measure of skill described above assessed the skill level of the therapists working together. For the purposes of this analysis, therefore, the abilities of individual therapists in the dyad are not separated but taken together; we consider the implications of this in the discussion.

The first set of analyses tested the hypothesis that Therapist Skill and Organization (from direct observation), and alliance (from parent report) predicted individual change in disruptive behavior using regression analyses. Therapist skill, Organization, and alliance were entered together in a regression model because there was not a strong *a priori* rationale for prioritizing or excluding any of these three indicators. While alliance with the parents was reported by each parent, therapist skill and organization were group-level variables and not individual variables particular to the parent or therapist. That is because all participants in a group received the same score for therapist. Consequently, we needed to account for this non-independence or nested nature of the data, i.e., all individuals within a particular group had the same scores for therapist skill and organization. We did this by using multilevel

modeling with the MLWin statistical package (Draper, Langford, and Lewis 2001). We report the unstandardized regression coefficients and their standard errors from the multilevel model because they provide a clear basis for interpreting the finding and its effect size. The main analyses examine individual child change using multilevel modeling (n=90). Additionally, supplementary analyses are reported using group-level change (n=15) to assess a) the change in antisocial behavior for whole groups in relation to skill and b) therapist characteristics in relation to skill.

## Results

*Prediction of individual outcome, allowing for group membership* Correlations among therapist skill, organization and alliance were modest and non-significant; the strongest association was found between therapist skill and alliance ( $r(15) = .34$ ); the weakest was between skill and Organization ( $r(15) = .14$ ). Results from the regression model in which skill, Organization, and alliance were entered on an a priori basis are presented in Table 4. In the first model predicting post-treatment antisocial behavior, we include as covariates pre-treatment antisocial behavior, child age and pre-treatment hyperactivity, and parental attendance. The second model in Table 3 included parental attendance plus the three group therapist variables, skill, organization, and alliance. Results show that group therapist skill significantly predicted post-treatment antisocial behavior independent of pre-treatment antisocial behavior, other child covariates, therapist organization and quality of alliance with parents. The effect indicated that a 1-unit change in group therapist skill (the range in skill scores was 3.0 to 4.61, SD 0.56) was associated with a reduction of .58 SD in antisocial behavior (the raw B score change was 0.25, the measure had a mean of 1.58 and a SD of 0.43), a substantial effect.

*Supplementary Analyses* For illustrative purposes, we next show the average change in antisocial behavior in the 15 groups according to group therapists' skill (Figure 1). The

correlation between these variables was .71, indicating a strong ordering and linearity between therapist skill and treatment outcome, both measured at the group level (n=15).

(TABLE 3 AND FIGURE 1 ABOUT HERE)

Client and therapist factors associated with skill were tested. Because in the original study the initial level of child antisocial behavior, hyperactivity, and child age predicted outcome, we wished to check whether they also were associated with therapist skill level. Bivariate analyses indicated these were weak (.03 for antisocial behavior, -.27 for hyperactivity, and .33 for age, all non-significant).

All groups were led by two therapists, in various combinations, and the skill ratings were made for their combined effort, rather than individually. However, bearing this limitation in mind, it is possible to calculate the mean skill level of therapists across the groups they worked in. Age and gender had no effect on skill level, but being trained in mental health, being a nurse, and going on later to become a certified “mentor” (certified to give trainings and supervision) in the program all were associated with significantly higher skill levels (Table 4). There was a trend towards the skill level going up after the first six months of the trial. These findings were similar when the differences were re-run controlling for other influences on skill.

(TABLE 4 ABOUT HERE)

## Discussion

This study investigated the relationship between the level of therapist skill and the effectiveness of parent training, which is the most widely used empirically validated psychological treatment for childhood antisocial behavior. The association between the level of the therapists’ skill in running the groups and the child behavior outcome was strong, a one point increase in skill level (the range in therapists’ skill level in this study was 1.6 points)

was associated with a 0.58 SD increase in the effectiveness of the intervention. This relationship held up after controlling for the individual characteristics of the children that affected outcome, hyperactivity and age, and other aspects of the therapeutic process that might have influenced effectiveness, namely the level of therapist organization, the quality of the therapeutic alliance as judged by the parent, and the percentage of sessions they attended. The findings appear to uphold the notion that skill is an important contributor to therapy outcomes, thus adding to the literature on what makes interventions effective. These results are consistent with the findings of Forgatch et al (2004), who found that skill was significantly associated with change in parenting, rather than child antisocial behavior.

Interpretation of the supplementary analyses relating therapist characteristics to skill level has to be cautious due to three constraints. Firstly, numbers were relatively small. Secondly, the selection process was non-random, the principal investigator approached clinics, and interested personnel expressed an interest in training in the method, and were taken on if they seemed warm, enthusiastic, and committed, and had a good reputation in their clinics. Thirdly, skill ratings were made for the therapy as given by pairs of therapists, and not by individuals; individual scores were calculated according to their average with other people, since most ran several groups. Bearing these constraints in mind, it is nonetheless interesting that those therapists who had the motivation and ability to go on to seek and achieve accreditation in the approach had significantly higher skill levels. Therapists with mental health training or experience were judged to be more skilled than those without it, which could be due to more competent people taking or the training, the training leading to better skills, or both. If replicated, this finding could have important implications for workforce expansion, since it could be argued that more challenging cases should be seen by therapists with mental health training, and recruitment should not be only of counselors and others without such training. Nurses delivered the program more skillfully than other

professionals in this study. This is in keeping with Olds' findings in his studies of nurse home visitation, where the effectiveness found with nurses was not replicated when the same program was delivered by volunteers, despite receiving the same training in the program (Korfmacher, O'Brien, Hiatt, & Olds (1999).

Since in the trial nearly all parents were mothers, it was possible that women therapists would be more understanding and skilled, but therapist gender had no effect. Likewise it was possible that older therapists would be more skilled, given their greater range of experience of clients and life, and greater opportunity to hone their therapeutic abilities. However, there was no difference between older and younger therapists in skill level. There was however an almost significant trend towards increased skill after the first six months of the study, suggesting that the supervision process was being effective; when interventions are disseminated amongst services, it may take time for staff to become competent.

The study had a number of methodological strengths. The sample studied was reasonably large and was typical of 'real life' clinical referrals, characterized by severe antisocial behavior, disadvantaged backgrounds, and high comorbidity. The therapists were also reasonably typical, being generalists employed in regular clinics, rather than being university staff who specialized in this intervention only. Therefore effects seen here are likely to generalize to everyday practice. Skill was rated by professionals well trained in the intervention, who observed parenting groups for extensive periods for the trial, sampled across four different stages. Raters were blind to outcome, which was measured using a high quality investigator-based interview. The statistical analysis took the effects of clustering into account, as proposed by Baldwin, Murray, and Shadish (2005). This is important, for example, trials of Cognitive Behavioral Therapy often use a relatively small number of therapists who see a larger number of clients. Within clusters, variance in outcomes may be reduced due to therapist effects, or where there are groups, due to within group effects, such

as members of a particular group getting on well together and supporting each other outside sessions.

A limitation of the study was that its main aim was to investigate the overall relevance of skill, and compare its relevance to the therapeutic alliance, rather than specifically to compare skill with fidelity as skill was conceptualized as subsuming fidelity. Alliance was measured fairly simply from the client's point of view, and it is possible that adding a more sophisticated measure of alliance completed by both clients and therapists could have contributed to outcome.

If other studies confirm the strong impact of therapist skill on child outcomes found here, the implications will be considerable. The need to improve the training, supervision, and evaluation of therapists will become more pressing (Calhoun, Moras, Pilkonis, and Rehm, 1998). Currently in the training of many professions that work in mental health, for example clinical psychologists, evaluation of skill is usually by an exit examination comprising of a mixture of written examinations and observed practice. For the rest of the professional's career, there may be no further evaluation of competence. In contrast, instruments such as the one described here would allow direct judgments to be made about whether individuals have attained competence in particular therapies. Accreditation procedures are already in place for some proprietary programs, including the program used in this study, the Incredible Years; other programs require careful monitoring of treatment fidelity, for example Multisystemic Therapy (Henggeler et al 1997). However, it is often not required by employers that staff be accredited in such programs, and indeed the majority of proprietary parent training programs do not have any accreditation system. In the wider world of adult and child psychotherapies, many do not require supervisors to observe therapists at work with clients at all, relying instead on therapists' accounts of what they did. Perhaps in future it should be a requirement for therapists to pass an examination that includes

assessment of observed work with clients before being allowed to work alone in whatever specific treatments they offer, and to continue getting supervision. This would be part of the general movement towards greater transparency and public accountability in health and counseling provision.

Future research could include the development of a shorter, more easily used instrument for assessing skill. Questions to address include determining the extent to which skill is relevant across types of disorder, types of therapy, and level of symptom severity. For difficulties such as anxiety and depression, which are arguably primarily about the client's inner world and mental state, general therapist warmth and the feeling of being understood through the alliance might be more important than skill in a particular approach. For behavior problems such as child aggression, however, observational studies of causation (Patterson 1982) suggest that for child improvement to occur, parents must change their own practices, in addition to their mindset. It is possible therefore that therapist skill has a more major role in addition to the alliance that for externalizing rather than internalizing disorders, because the client needs not only to be understood and change their beliefs, but also to learn a specific set of behaviors and practice them in relation to their child. For this to come about in difficult cases may require the therapist to inculcate the new behaviors with considerable skill. The meta-analysis by Shirk and Karver (2003) found the therapeutic alliance was more important in youth for externalizing than for internalizing disorders, and this may also be true for skill. Less severe cases might require less skilled workers, and future trials could test this interaction experimentally, for example in a two-by-two design. In staff development, different types and lengths of training could be compared, and the supervision process could be investigated, for example comparing the agreement and effectiveness of direct observation

versus therapist accounts. It would be interesting to conduct a trial within a service formally rewarding therapists for increasing their level of skill, and observe the results on outcomes.

### Acknowledgements

We thank Thonas O'Connor of the University of Rochester for statistical analyses, and Carolyn Webster-Stratton of the University of Washington for continuing encouragement and skill development.

### References

- National Institute of Mental Health (2000) *Child and adolescent violence research at the NIMH. Report number 00- 4706* Bethesda, Maryland: National Institutes of Health.
- Baldwin, S. A., Murray, D. M., & Shadish, W. R. (2005). Empirically supported treatments or type I errors? Problems with the analysis of data from group-administered treatments. *Journal of Consulting & Clinical Psychology, 73*, 924-935.
- Beutler, L. E., Malik, M., Alimohamed, S., Harwood, T. M., Talebi, H., Noble, S. et al. (2004). Therapist variables. In M.J.Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior* (pp. 227-306). New York: John Wiley and sons
- Bickman, L., Lambert, E. W., Andrade, A. R., & Penaloza, R. V. (2000). The Fort Bragg continuum of care for children and adolescents: mental health outcomes over 5 years. *Journal of Consulting & Clinical Psychology, 68*, 710-716.
- Bourdon, K. H., Goodman, R., Raie, D. S., Simpson, G., & and Koretz D.S. (2005). The Strengths and Difficulties Questionnaire: U.S. Normative Data and Psychometric Properties. *Journal of the American Academy of Child and Adolescent Psychiatry, 44:6*, 557.

- Calhoun, K. S., Moras, K., Pilkonis, P. A., & Rehm, L. P. (1998). Empirically supported treatments: implications for training. *Journal of Consulting & Clinical Psychology*, *66*, 151-162.
- Chambless, D. L. & Hollon, S. D. (1998). Defining Empirically Supported Therapies. *Journal of Consulting and Clinical Psychology*, *66*, 7-18.
- Costello, E. J., Egger, H., & Angold, A. (2005). 10-year research update review: the epidemiology of child and adolescent psychiatric disorders: I. Methods and public health burden. *Journal of the American Academy of Child & Adolescent Psychiatry*, *44*, 972-986.
- Davidson, K., Scott, J., Schmidt, U., Tata, P., Thornton, S., & Tyrer, P. (2004). Therapist competence and clinical outcome in the Prevention of Parasuicide by Manual Assisted Cognitive Behaviour Therapy trial: the POPMACT study. *Psychological Medicine*, *34*, 855-863.
- Draper, D., Lanford, I., & Lewis T. (2001) *A user's guide to MLWin*. London: Centre for Multilevel Modeling, Institute of Education.
- Ford, T., Goodman, R., & Meltzer, H. (2003). The British Child and Adolescent Mental Health Survey 1999: The prevalence of DSM-IV disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *42*, 1203.-1211.
- Forgatch, M. S., Patterson, G. R., & DeGarmo, D. S. (2005). Evaluating Fidelity: Predictive Validity for a Measure of Competent Adherence to the Oregon Model of Parent Management Training. *Behavior Therapy*, *36*, 3-13.

- Henggeler, S. W., Melton, G. B., Brondino, M. J., Scherer, D. G., & Hanley, J. H. (1997). Multisystemic therapy with violent and chronic juvenile offenders and their families: The role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology, 65*, 821-833.
- Hogue, A., Liddle, H. A., & Rowe, C. (1996). Treatment Adherence Process Research in Family Therapy: A Rationale and some Practical Guidelines. *Psychotherapy, 33*, 332-341.
- Hogue, A., Dauber, S., Stambaugh, L. F., Cecero, J. J., & Liddle, H. A. (2006). Early Therapeutic Alliance and Treatment Outcome in Individual and Family Therapy for Adolescent Behavior Problems. *Journal of Consulting and Clinical Psychology, 74*, 121-129.
- Horvath, A. O. & Symonds, B. D. (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. *Journal of Counseling Psychology, 38*, 139-149.
- Huey, S. J., Henggeler, S. W., Brondino, M. J., & Pickrel, S. G. (2000). Mechanisms of change in multisystemic therapy: Reducing delinquent behavior through therapist adherence and improved family and peer functioning. *Journal of Consulting and Clinical Psychology, 68*,; 451-467.
- Kam, C. M., Greenberg, M. T., & Walls, C. T. (2003). Examining the role of implementation quality in school-based prevention using the PATHS curriculum. *Prevention Science, 4*, 55-63.
- Kazdin, A. E. (2005). *Parent Management Training*. Oxford University Press.

- Kazdin, A. E., Whitley, M., & Marciano, P. L. (2006). Child-therapist and parent-therapist alliance and therapeutic change in the treatment of children referred for oppositional, aggressive, and antisocial behavior. *Journal of Child Psychology and Psychiatry*, 47, 436-445.
- Korfmacher J. O'Brien R. Hiatt S. Olds D. (1999) Differences in program implementation between nurses and paraprofessionals providing home visits during pregnancy and infancy: a randomized trial. *American Journal of Public Health*. 89:1847-51
- Kendall, P. C. & Southam-Gerow, M. A. (1995). Issues in the transportability of treatment: the case of anxiety disorders in youths. *Journal of Consulting & Clinical Psychology*, 63, 702-708.
- Kessler, R. C., Demler, O., Frank, R. G., Olfson, M., Pincus, H. A., Walters, E. E. et al. (2005). Prevalence and treatment of mental disorders, 1990 to 2003. *New England Journal of Medicine*, 352, 2515-2523.
- Krupnick, J. L., Sotsky, S. M., Simmens, S., Moyer, J., Elkin, I., Watkins, J. et al. (1996). The role of the therapeutic alliance in psychotherapy and pharmacotherapy outcome: findings in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting & Clinical Psychology*, 64, 532-539.
- Layard, R. (2006) The case for psychological treatment centres. *BMJ* 332, 1030 - 1032
- Loeb, K. L., Wilson, G. T., Labouvie, E., Pratt, E. M., Hayaki, J., Walsh, B. T. et al. (2005). Therapeutic alliance and treatment adherence in two interventions for bulimia nervosa: a study of process and outcome. *Journal of Consulting & Clinical Psychology*, 73, 1097-1107.

- Martin, D. J., Garske, F. P., & Davis, M. K. (2000). Relationship of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting & Clinical Psychology* 68, 438-450.
- Mihalic, S. (2002). *The Importance of Implementation Fidelity*. Chapter in: Blueprints for violence prevention replications: factors for implementation success. Pp 28-41. University of Colorado at Boulder: Centre for the study and prevention of violence.
- Moffit, T. E. & Caspi, A. (2001). Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Development and Psychopathology*, 13(2), 355-375.
- Moncher and Priz R (1991) Treatment fidelity in outcome studies. *Clinical psychology review* 247-266
- Patterson, G. R. (1982). *Coercive Family Process*. Oregon, USA: Castalia Publishing Company.
- Respect Taskforce (2006) *The Prime Minister's Respect Action Plan*. London: The Stationery Office
- Schoenwald, Letourneau and Halliday-Boykins (2005) Predicting Therapist Adherence to a transported Family-Based Treatment for Youth. *J Clin Child & Adol Psychol* 34, 658-670
- Shaw, B. F., Elkin, I., Yamaguchi, J., Olmsted, M., Vallis, T. M., Dobson, K. S. et al. (1999). Therapist competence ratings in relation to clinical outcome in cognitive therapy of depression. *Journal of Consulting & Clinical Psychology*, 67, 837-846.

- Shirk, S. R. & Karver, M. (2003). Prediction of treatment outcome from relationship variables in child and adolescent therapy: a meta-analytic review. *Journal of Consulting & Clinical Psychology, 71*, 452-464.
- Surgeon General (2001). *Youth violence: a report of the Surgeon General*. Washington: Department of Health and Human Services.
- Taylor, E., Schachar, R., Thorley, R., & Wieselberg, M. (1986). Conduct disorder and hyperactivity: I. Separation of hyperactivity and antisocial conduct in British child psychiatric patients. *British Journal of Psychiatry, 149*: 760-767.
- Taylor, E., Chadwick, O., Heptinstall, E., & Danckaerts, M. (1996). Hyperactivity and conduct problems as risk factors for adolescent development. *Journal of the American Academy of Child & Adolescent Psychiatry, 35*, 1213-1226.
- Trepka, C., Rees, A., Shapiro, D. A., Hardy, G. E., & Barkham, M. (2004). Therapist Competence and Outcome of Cognitive Therapy for Depression. *Cognitive Therapy & Research, 28*, 143-157.
- Waltz J, Addis M, Koerner K and Jacobson N (1993) Testing the integrity of a psychotherapy protocol; assessment of adherence and competence *Journal of Consulting & Clinical Psychology* 61, 620-630
- Webster-Stratton, C. (1989). Systematic comparison of consumer satisfaction of three cost-effective parent training programs for conduct-problem children. *Behavior Therapy, 20*, 103-115.

Webster-Stratton, C., Reid, M. J., & Hammond, M. (2004). Treating Children With Early-Onset Conduct Problems: Intervention Outcomes for Parent, Child and Teacher Training. *Journal of Clinical Child and Adolescent Psychology, 33*, 105-124.

Webster-Stratton, C. (2004). Quality Training, Supervision, Ongoing Monitoring, and Agency Support: Key Ingredients to Implementing The Incredible Years Programs with Fidelity. Seattle, Washington, US, University of Washington.

Weisman, A., Tompson, M. C., Okazaki, S., Gregory, J., Goldstein, M. J., Rea, M. et al. (2002). Clinicians' fidelity to a manual-based family treatment as a predictor of the one-year course of bipolar disorder. *Family Process, 41*, 123-131.

Weisz, J. R., Donenberg, G. R., & Han, S. S. (1995). Bridging the Gap Between Laboratory and Clinic in Child and Adolescent Psychotherapy. *Journal of Consulting and Clinical Psychology, 63*, 688-701.

Weisz, J. R., Chu, B. C., & Polo, A. J. (2004). Treatment dissemination and evidence-based practice: Strengthening intervention through clinician-researcher collaboration. *Clinical Psychology: Science and Practice 11*, 300-307.

Zuroff, D. C. B. S. J. (2006). The therapeutic relationship in the brief treatment of depression: contributions to clinical improvement and enhanced adaptive capacities. *Journal of Consulting & Clinical Psychology, 74*, 130-140.

TABLE 1. PARENTING PROGRAM CONTENT AND DELIVERY

*Content*

- Structured sequence of topics, introduced with video clips of parents with children
- Topics include play, praise, incentives, setting limits, and discipline
- Emphasis on promoting sociable, self-reliant child behavior, and calm parenting
- Constant reference to parent's own experience and predicament
- Theoretical basis informed by extensive empirical research and made explicit
- Detailed manual available to ensure treatment fidelity and to enable replicability

*Delivery*

- Collaborative approach acknowledging parents' feelings and beliefs
- Difficulties normalized, humor and fun encouraged
- Parents supported to practice new approaches during session and through homework
- Written feedback from parents after every session
- Crèche, good quality refreshments, and transport provided
- Group leaders supervised weekly to ensure treatment fidelity and to develop skills, using videotape of last session to rehearse therapeutic approach

TABLE 2. FACTOR LOADINGS AND RELIABILITY OF SCALES

<u>Scale ( ICC reliability)</u>	<u>Factor</u>	
	<i>Therapist Skill</i>	<i>Therapist Organization</i>
Skill in highlighting the positives (0.68)	.91	
Skill in encouraging participation (0.59)	.81	
Skill in reviewing homework & praising achievement (0.69)	.92	
Skill in summarizing main points (0.51)	.85	
Skill in using vignettes (0.80)	.75	
Skill in using role play to enhance learning (0.90)	.69	
Parents use of skills (0.72)	.66	
Practical arrangements (0.76)		.68
Issue homework (0.90)		.76
Materials prepared in advance (0.50)		.53
<i>Variance accounted for after rotation</i>	<i>50%</i>	<i>21%</i>

TABLE 3 PREDICTION OF ANTISOCIAL BEHAVIOR AFTER TREATMENT (MULTILEVEL MODEL)

<i>Predictors</i>	<i>Model 1 B (SE)</i>	<i>Model 2 B (SE)</i>
Child antisocial behavior before	.46 (.11)**	.48 (.10)**
Child age	-.06 (.31)	-.04 (.03)
Child hyperactivity	.16 (.07)*	.10 (.07)
Parental Attendance	.002 (.002)	.001 (.002)
Therapeutic alliance		.14 (.09)
Therapist organization		-.02 (.12)
Therapist skill		-.25 (.09)*
-2 log likelihood	101.94	93.92

\* p< 0.05 \*\*p<0.001

TABLE 4. THERAPIST CHARACTERISTICS AND SKILL LEVEL

<i>Therapist Characteristic</i>	<i>n</i>	<i>Skill level Mean (sd)</i>	<i>Difference between groups*</i>	
			<i>t value</i>	<i>significance</i>
Trained in mental health	9	4.2 (.34)	2.5	.03
Not trained in mental health	4	3.6 (.52)		
Became accredited Mentor	3	4.3 (.03)	2.3	.048
Did not become Mentor	10	3.9 (.52)		
Occupation Nurse	3	4.4 (.07)	3.1	.01
Occupation Other	10	3.9 (.49)		
Female	9	4.0 (.50)	.15	.88
Male	4	4.1 (.49)		
Age under 40	6	4.1 (.39)	.36	.72
Age over 40	7	4.0 (.56)		
<i>Time trends in skill<sup>§</sup></i>				
Skill at beginning of trial <sup>^</sup>	3	3.4 (.40)	2.0	.07
Skill during mid & later stages	12	4.1 (.53)		

\*two sided independent t test <sup>^</sup>groups run in first six months of trial vs. run in next 30 months

<sup>§</sup>irrespective of therapist, level of skill seen in groups as trial progressed

Figure 1: Impact of therapist skill on effectiveness

