Treatment adherence in a behaviourally based parenting program

Alison Rendu

D.Clin.Psy.

University College London

Volume I

ProQuest Number: U643604

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

ProQuest.

ProQuest U643604

Contents

List of tables

Abstract

Behaviourally based parenting programs have been shown in randomised controlled trials, set in university settings under highly controlled conditions, to reduce conduct problems in children. In contrast, studies in regular clinical practice have found less positive results, questioning the effectiveness of these programs when applied to real world settings. One possible reason for this is that once the controls of the laboratory are removed, programs are not delivered as they are intended, reducing the positive effects observed in the efficacy trials.

This study had two aims. The central aim was to design an instrument that could be used by trained raters to rate treatment adherence to the Webster-Stratton BASIC parenting program. A secondary aim was to explore the relationship between treatment adherence and outcome, as measured by a reduction in child conduct difficulties, using the developed instrument to measure adherence.

The instrument was designed to rate adherence using global rating scales of leaders' behaviour, based on videotapes of parenting groups. Inter-rater reliability between two raters was found to be acceptable for nine of the 27 variables included in the instrument. A factor analysis found that these variables loaded onto two separate factors. Regression analyses showed that one of the factors, which included aspects of leaders' behaviour concerned with group facilitation, was a significant predictor of treatment outcome, but this finding was no longer significant when age of child and hyperactivity were taken into account.

Future research in this area needs to further develop the instrument, increasing its reliability on specific variables and accounting for the validity of the measure. The instrument could then be used to assess treatment adherence outside of the research setting, where group leaders are potentially more likely to depart from the intended delivery of the program. This area of research is important as there is currently a need to bridge the gap between the laboratory and the clinic, enabling clinicians to provide treatments that are as effective as they are efficacious.

Acknowledgements

1. Introduction

Parent training programs have been shown in randomised control trials to increase parental skill, produce more positive parental attitudes and significantly reduce the level of child conduct difficulties (Kazdin & Kendall, 1998). They have been recommended as one of the treatments of choice for families with children with conduct problems (Barlow, 1999; Roth & Fonagy, 1998).

The Webster-Stratton BASIC parenting program (Webster-Stratton & Hancock, 1988) has been shown in both the US and UK to reduce child conduct difficulties in university based research trials (Webster-Stratton, 1984). However these studies include highly sophisticated checks of treatment adherence which do not occur outside of the research setting. As the programs are more widely disseminated throughout the United Kingdom it is questionable as to whether they will be delivered as intended in the manual and consequently whether the positive effects found in the research will be retained.

The present study therefore aimed to design a tool for measuring treatment adherence to the Webster-Stratton BASIC parenting program. This tool was then used to explore the relationship between treatment adherence and outcome.

This chapter firstly provides a brief overview of the need for and development of behaviourally based parenting programs and their effectiveness in reducing childhood conduct difficulties. It then draws from the more general psychological therapies literature to outline the differences between effectiveness and efficacy studies and

looks at the arguments for and against the subsequent development of psychotherapy treatment manuals. It discusses the literature on treatment adherence and presents a brief review of the debate between adherence and competence, stressing the importance of factors that are common to all forms of psychotherapy. Finally the chapter looks at these factors in regard to the present study, outlining the gaps in the research and the specific aims of this study.

1.1. Parenting programs

1.1.1. The rise of childhood antisocial behaviour and its consequences

Childhood antisocial behaviour encompasses a broad range of externalising behaviours ranging from yelling and temper tantrums to aggression, physical destruction and stealing. According to the DSM-IV (Francis, First & Pincus, 1995) these can be divided into two disorders, oppositional defiant disorder and conduct disorder. Oppositional defiant disorder refers to a recurrent pattern of negativistic, defiant, disobedient and hostile behaviour towards authority figures; this must be apparent for at least six months, be of higher frequency than expected for their age and sex and significantly impair social and academic functioning. Conduct disorder is defined as a persistent pattern of behaviour in which age appropriate societal norms or rules and rights of others are violated; this must be apparent for at least twelve months and can either have its onset in childhood (before the age of 10) or adolescence (after the age of 10).

Recent studies have found that antisocial behaviour and aggression in children seems to be rising and at an earlier age (Rutter, Hagell & Giller, 1998). Over 5% of children are demonstrating antisocial behaviour in Great Britain (Meltzer, Gatman, Goodman

& Ford, 1990) and between 7-25% of pre-school and early school age children meet the criteria of oppositional defiant disorder and conduct disorder in the US (Campbell, 1990). Whilst this pattern is worrying in itself, it also has disturbing individual, family and societal consequences.

Research shows that early onset conduct problems have been found to be associated with hyperactivity, emotional problems, language disorders, poor attention and lower IQ (Silberg et al., 1996). Conduct problems are also predictive of serious health and behavioural problems in adolescence – drug abuse, depression, juvenile delinquency and school dropout (Loeber, 1985). Most children with conduct problems leave school with no qualifications and over a third become recurrent juvenile offenders (Farrington, 1995). Childhood conduct problems are also associated with life long antisocial behaviour (Moffitt, 1993). For example, Kazdin (1985,1995) found that adults with a history of chronic aggression beginning in childhood are at an increased risk of committing murder, rape and robbery as well as other criminal offences, and are more likely to develop antisocial personality disorder, interpersonal problems, poor physical health and marital disruption.

Childhood conduct problems also affect the process of socialisation by increasing the risk of peer rejection (Coie, 1990) and the levels of parental criticism, hostility and abuse (Farrington, 1995, Reid, Taplin & Loeber, 1981). Studies have found parents of conduct-problem children are more likely to be depressed, stressed, angry and inclined to use physical forms of discipline (Webster-Stratton, 1994). Results from a qualitative study showed how a child's conduct difficulties affect the family's relationships, increasing marital stress and tension within the extended family. Child

conduct difficulties were also shown to affect the family's relationships with professionals, teachers, and other parents in their community, being characterised by negative feedback, stigmatisation, social isolation and rejection (Webster-Stratton & Spitzer, 1994).

Research has shown that there are a number of family characteristics that heighten the risk of a child developing conduct problems. These include low income, low education, teenage pregnancy, isolation high stress levels, single parenthood, parental psychiatric illness, parental criminal history or substance abuse, high levels of marital discord and depression (Webster-Stratton, 1990). Conduct difficulties also increase in children whose parents are inconsistent in their discipline, physically abusive, highly critical or hostile (Farrington, 1995, Patterson & Strouthamer-Loeber, 1984). Furthermore the risk of developing conduct difficulties appears to increase exponentially with each additional risk factor (Coie et al., 1993, Rutter, 1980).

1.1.2. The development of behaviourally based parenting programs

The pioneering work of Patterson (1982) showed that parents had a causal role in the maintenance of antisocial behaviour in their children. He was able to show that parents, by attending to their child's antisocial behaviour helped maintain this negative behaviour through negative reinforcement, whilst ignoring positive behaviour had an extinguishing effect on the desirable behaviours. Patterson's coercion hypothesis proposes that over time the rate and intensity of the parent and child's negative interactions increase; furthermore the parents' increasing use of negative discipline provides the child with opportunities to learn more disruptive behaviours through observational learning. This hypothesis, therefore, suggests that

11

by changing parental behaviour, teaching parents to deal effectively with children's misbehaviour and modelling effective problem solving and discipline strategies, the coercive cycle can be broken and the child's disruptive behaviours will be reduced. Patterson's success in reducing conduct problems by training parents using these social learning techniques has led to the development of behaviourally based training programs for parents (Patterson, 1986). These aim to reduce the risk of developing conduct problems and alter the poor long-term prognosis for these children.

Since the early 1980's Carolyn Webster-Stratton has been developing parenting training programs for families of children with conduct problems. The programs use group discussion, videotaped modelling, homework activities and rehearsal intervention techniques to help teach effective parenting practices. The original 12-week BASIC program was guided by the modelling literature and focuses on teaching parents interactive play skills, reinforcement skills, non-violent discipline techniques, logical and natural consequences and problem solving strategies (Webster-Stratton & Hancock, 1988). The groups are facilitated by two leaders who foster a collaborative approach to learning, working in partnership with the parents in their group. This is done through reflection, summary of points made by parents, reframing, reinforcement, support and acceptance, humour and optimism, encouragement of participation, teaching of important concepts and role-play exercises (Webster-Stratton, 2000).

The program has been designed to be widely disseminated, with the developers providing a detailed leadership manual, implementation training, technical assistance, standardised curriculum and evaluation material, videos and fidelity instrumentation.

The manual provides a detailed prescription of the material that should be covered in each session as well as the techniques the leaders should adopt to engage the parents with the work, help facilitate a good relationship and provide a non-judgemental supportive group atmosphere.

1.1.3. Parent training outcome studies

Parent training interventions have been recommended as one of the treatments of choice for families with children with conduct problems (Roth & Fonagy, 1998). Literature reviews have highlighted that parent training interventions result in increased parental skill, more positive parental attitudes and a significant reduction in conduct problems (Kazdin & Kendall, 1998). The effects have been shown to remain after several months (Miller & Prinz, 1990) and have been found to be evident up to 14 years after treatment (Long, Forehand Wierson & Morgan, 1994). Indirect benefits have also been reported with studies showing a decrease in family conflict (Dishion & Andrews, 1995), decreases in parent depression and stress (Webster-Stratton & Hammond, 1990) and a reduction in the likelihood of similar behaviours in siblings (Kazdin, 1985).

The Webster-Stratton BASIC parenting program has been researched over a series of randomised studies in the US, involving over 600 children referred for conduct problems. The program has been shown to improve positive parental attitudes and parent child interactions, to reduce parents use of violent forms of discipline and to reduce child conduct problems (Webster-Stratton, 1982, 1984, 1990, 1994; Webster-Stratton, Hollinsworth & Kolpacoff, 1989). It has also been shown to be as effective

as high cost one-to-one therapy both at the end of treatment and at one-year follow-up, being five times more cost effective (Webster-Stratton, 1984, 1985).

These studies, however, have been set in specialised university clinics, using highly motivated and extensively trained leaders dealing with a highly select sample of parents. Many of the trials use volunteers or people selected by referrers to take part in the parenting project, therefore often excluding unmotivated, disorganised or disadvantaged families who may have the most severely antisocial children. They have also included multiple checks of treatment adherence. For example, in a study looking at the preventative effects of parenting programs, group leaders completed a 4 day training workshop, followed the detailed manual for each session, were videotaped for feedback and review by the project director, conducted their first parent group with one of the trained clinic staff members, had weekly supervision and feedback and kept detailed weekly checklists of group process (Webster-Stratton, 1998). It is therefore unclear as to whether the results of these studies will transfer into regular clinical practice when working with more complex, multi-disadvantaged families and with less opportunity to monitor treatment adherence.

1.1.4. Parent training in everyday clinical practice

A look at the general child outcome literature suggests that the positive benefits found in specialised university settings are not replicated in everyday clinical practice. A review of meta-analyses of published trials of psychological treatments for childhood disorders set in university settings, found effects sizes of between 0.71 and 0.84 (Weisz, Weiss & Donenberg 1992). This contrasted with studies in regular clinical practice where no significant effects were found (Andrade, Lambert & Bickman,

2000; Weisz et al., 1992). Reasons suggested for the poor outcome in clinical practice include that these studies involve children with more severe problems and more distressed families, involve less empirically supported interventions from staff with heavier case loads or receive empirically supported interventions that are poorly adhered to and therefore bear little resemblance to the original model.

However two studies provide some evidence that parenting groups can be effective at reducing conduct problems in clinical practice. Taylor, Schmidt, Pepler and Hodgins (1998) compared the Webster-Stratton parenting program to treatment as usual and a waiting list control, using standard referrals to a children's mental health centre and regular clinical staff with large and diverse case loads. They found that mothers in the Webster–Stratton parenting group reported fewer behavioural problems and greater satisfaction with treatment than those in the treatment as usual and waiting list control conditions.

In the UK, Scott, Spender, Doolan, Jacobs and Aspland (2001) delivered the Webster-Stratton BASIC parenting program to parents in regular clinical practice, with standard referrals to child mental health services and regular clinic staff to carry out the intervention. They found that children in the intervention group showed a significant reduction in antisocial behaviour compared to those in the waiting list control group (effect size between groups 1.06). Parents in the intervention group also showed an increase in the proportion of praise to ineffective commands compared to control group parents (effect size between groups 0.76). This research supports the use of such groups in community settings, providing a cost effective treatment for antisocial behaviour. Both studies, however, were still conducted under

conditions of high internal validity. For example, the Scott et al. study used several exclusion criteria including major developmental delay, hyperkinetic syndrome, or any other condition requiring separate treatment, to create a more homogeneous sample than would be expected in regular clinical practice. Both studies also built in regular checks of treatment adherence: providing weekly supervision meetings for the therapists, videotaping all intervention sessions and ensuring leaders kept detailed weekly checklists of group process. Therefore, whilst the studies was carried out in clinical practice they were still highly controlled in many ways, questioning the extent to which they resembled everyday clinical practice and highlighting the need to consider the issues that arise from moving research based therapies from the laboratory into the field.

The difference between highly controlled university based research and regular clinical practice has received a great deal of attention within the field of psychotherapy research. The next section will look at this debate in more detail, drawing on the more general area of psychotherapy research to clarify the difference between efficacy and effectiveness studies and to address the issue of treatment manuals and the role of treatment adherence and common factors in treatment outcome.

1.2. Moving from the laboratory into the field

This section draws on literature from the broader area of psychotherapy research and is therefore not specific to the parenting or child treatment literature.

1.2.1. Effectiveness vs. efficacy

There has been a recent emphasis, within clinical research, on the difference between efficacy and effectiveness studies, questioning the extent to which results can be transferred from the laboratory into everyday clinical practice (Conner-Smith & Weisz, 2003). Efficacy studies are set up under conditions of high internal validity, maximising the likelihood of detecting treatment effects by removing those factors that might obscure them and maximising replicability. They use highly restrictive inclusion and exclusion criteria creating a homogenous treatment group who have the same diagnosis, no comorbidity and commonly less severity. They use appropriate control conditions to which participants are randomly assigned and clearly defined treatments in which the therapists are highly trained and supervised. They usually take place in a laboratory or university setting and the therapists generally have small caseloads (Nathan, Stuart & Dolan, 2000; Weisz, Weiss & Donenberg, 1992; Weisz, Donenberg, Weiss & Han, 1995).

Effectiveness studies conversely emphasise external validity and generalisability. They are concerned with the feasibility of treatments in real world settings and take place in the clinic under everyday conditions without the same number of internal validity checks. They typically use participants who have been referred to the clinic and are in need of treatment, regardless of their diagnosis, severity or comorbidity; thus the treatment group is heterogeneous. Therapists tend to use multiple methods, dictated by the clinical circumstances rather than the research design; they often have limited supervision or specific training and have large caseloads (Nathan, Stuart & Dolan, 2000; Weisz, Weiss & Donenberg, 1992; Weisz, Donenberg, Weiss & Han, 1995).

Researchers commonly claim that efficacy research is the gold standard approach, showing that specific therapies have proven efficacy in treating specific problems (Norquist, Lebowitz & Hyman, 1999). Their rigorous control of factors that help maximise internal validity limits the number of alternative explanations for their findings and enables researches to makes claims about the efficacy of different treatments. Clinicians, however, often question the usefulness of these findings when applied to real world settings, claiming that they do not generalise to the very different conditions of the clinic. Whilst a specific therapy may have good outcomes in the laboratory, this may not be true when applied to a heterogeneous clinic based population who have comorbid diagnoses, increased severity and multiple social problems.

Weisz et al. (1992) reported on four recent meta-analyses, involving more than two hundred controlled outcome studies, finding consistent evidence for the beneficial effect of psychotherapy with children and adolescents, with mean effect sizes ranging from 0.71 to 0.84. Whilst these may be exciting findings pointing towards the usefulness of psychotherapy, all four meta-analyses included only the use of laboratory based efficacy studies, which may not represent conventional clinic treatment. In fact when Weisz, Weiss, Donenberg, and Han (1995) compared these findings against studies that involved (a) treatment of clinic referred youngsters, (b) treatment in service oriented clinic or clinical agencies, not research settings, (c) therapy conducted by practicing clinicians or (d) therapy conducted as a part of the regular service related program of the clinic, not primarily for research, they found markedly poorer outcomes, with effect sizes ranging from –0.40 to 0.29. Whilst the methodology in many of these studies was not ideal and the majority were conducted

some time ago, they still cast doubt on the assumption that the positive effects of efficacy studies is being replicated in the every day setting of the clinic, highlighting the need to bridge the gap between the laboratory and the clinic.

In their 1995 paper, Weisz et al. suggest three useful areas of research to help link laboratory findings with clinical practice: 1) an increase in the amount of outcome research completed in the clinic; 2) the identification of factors that may account for the superior effects of efficacy studies; 3) the exporting of well developed research therapy programs from the lab into the clinic.

Henggeler, Schoenwald and Pickrel (1995) have already commenced such work, looking into the positive effects of multisystemic therapy when implemented in community settings. They have found that positive results are possible with community based client populations when therapists are given sufficient training and supervision, similar to that found in laboratory based studies and when the therapy is consistent with empirically based interventions and delivered in a highly individualised, flexible and comprehensive fashion. Henggeler et al. also suggest that researchers should be more aware of the social systems in which their treatments will be applied, enabling research to approximate real social living, thus aiding the generalisability of research findings to the clinical setting.

Another area of research that has been widely debated and seeks to aid the dissemination of research-based therapies into the clinic is the development and use of treatment manuals.

1.2.2. Manualisation

Psychotherapy manuals have evolved from efficacy research with the need to specify exactly what treatment is provided, in order to enable replication. Manuals aim to provide the therapist with strategic and technical guidelines, so that all patients receive the same active therapeutic ingredients, thus increasing the internal validity of the study. They have now become an important medium for disseminating efficacious treatments in real-world clinical settings, in the hope of improving clinical outcomes and bridging the lab-clinic divide (Addis, Wade & Hatgis, 1999; Wilson, 1996, 1997).

In 1993 the division of clinical psychology of the American Psychological Association set up a task force to report on the promotion and dissemination of psychological procedures (Chambless et al., 1993). Their report cited 18 specific treatment interventions as "empirically validated interventions" emphasising the necessity of manualising treatments to maintain specificity and enable the training of therapists in empirically validated treatments. The report, however, was a catalyst for a heated debate on the use of manuals within psychotherapy with a number of advantages and disadvantages being proposed.

According to Addis et al. (1999), practitioners' most common concerns regarding the use of treatment manuals are that: (a) they affect the therapeutic relationship, (b) they do not meet the needs of all clients, (c) they restrict clinical innovation, (d) they are not superior to current treatments, (e) they are not feasible within clinical practice and (f) practitioners will not be competent in their administration or will lose job satisfaction. Each of these points is discussed in turn below.

1.2.2.1. The Therapeutic Relationship

A common concern amongst clinicians is that they feel less able to develop an effective therapeutic relationship when using structured manuals. Both Castonguay et al. (1996) and Henry et al. (1993) found that therapists' relationship skills declined when using treatment manuals despite an improvement in their technical skills. They also observed that those therapists who adhered best to the treatment manual were those who tended to be more controlling, hostile and prone to negative interactions. Given the widely accepted view that the therapeutic alliance is an important factor in the outcome of treatment (Svartberg, Seltzer & Stiles, 1998), these results are worrying. However Addis et al. (1999) report that the development of a therapeutic relationship is possible when using manualised treatments, finding superior relationships with the use of manuals compared to treatment as usual. This finding is supported by Krupnick et al. (1996), who found uniformly high levels of therapeutic alliance in all four manual based treatments used in the NIMH study of depression. It is therefore important that manuals incorporate strategies for the development of an effective therapeutic alliance so that this important, relational aspect of therapy is not sidelined in preference to specific technical skills.

1.2.2.2. The Needs of Clients

Clinicians have raised several concerns relating to the idea that manualised treatments do not meet the needs of their clients. Firstly, manualised treatments have been criticised for not taking into consideration individual client differences, for example personality, life situation or ethnicity, but relying on diagnostic criteria to guide treatment choice. Wolpe (1997), for example, objected to the use of specific treatments for specific phobias, emphasising the point that phobias can have different

causes that require different treatment interventions. However, the use of manuals does not diminish the necessity to tailor treatment to individual clients needs, and there is no data to suggest that clients needs are unmet with manual based treatments. Secondly, it has been suggested that manualised treatment are unable to treat clients with a co-morbid diagnosis. Hickling and Blanchard (1997), in their study of manualised therapy for Post Traumatic Stress Disorder, found their clients still reported symptoms of depression and generalised anxiety that required further treatment, concluding that "manualised treatments do not allow the clinician to complete the whole treatment of problems presented by the patient" (Hickling & Blanchard, p.199). However, a number of studies have shown that manualised treatments are equally efficacious in the treatment of clients with single or multiple disorders when outcome is measured in terms of the targeted problem (Brown, Anthony & Barlow, 1995; Wade, Treat & Stuart, 1998). These studies also demonstrate that manualised therapies targeted at a specific disorder are effective in reducing or eliminating co-morbid psychopathology. Wade et al. (1998), for example, found a significant reduction in symptoms of depression, social phobia and generalised anxiety with patients treated specifically for panic disorder. It would appear, therefore, that manualised treatments do not inherently prevent 'whole treatment' and further empirically supported treatments should be employed in a sequential manor to treat any remaining symptoms. Finally, clinicians are concerned that manual-based treatments ignore clients' emotions. However the identification and experiencing of emotions are critical components of any manual-based treatment.

1.2.2.3 Clinical Innovation

Another common concern of clinicians is that manualised treatments may restrict clinical innovation by replacing the therapist with the therapy manual and by reducing the development of new theories and alternative interventions. Whilst there may be a role for the use of para-professionals and computer programmes to deliver some forms of behavioural health care (Strosahl, 1998), manualised treatments do not assume rigidity in their application and often require clinical skill and innovation to administer the treatment in an individualised fashion. Treatment manuals may also help stimulate future research by providing a clearer picture of treatment failures and necessary future modifications (Addis et al., 1999).

1.2.2.4. Empirical Validation

Many clinicians believe that they already offer effective treatment to their clients using individual case formulation and are reluctant to learn new standardised manualised therapies. Persons (1991) argues that the fundamental difference between idiographic case formulation and manual-based treatment is the separation of assessment and treatment. She claims that, using case formulation, therapists assess each client's individual problems and design a treatment plan based on this assessment; progress is then assessed and this information is used to further revise the treatment plan, thus providing a continual interplay between assessment and treatment. This contrasts with her view of manualised treatments where clients' problems are identified and then treated with a standardised protocol, undermining the central role of assessment and case formulation. However, the evidence suggests that clinical judgement can be all too fallible, with therapists struggling to reliably integrate all the information they receive from assessment without incorporating their

own clinical and cognitive biases. Whilst therapists assume that clinical experience enhances expertise and effectiveness, the evidence to support this claim is lacking. To the contrary, research suggest that neither clinical experience nor training are related to clinical judgement (American Psychological Association, 1982) and that experienced clinicians are no more successful than non-professionals or less experienced therapists (Christenson & Jacobson, 1994; Smith, Glass & Miller, 1980). It would therefore appear that whist clinical judgement seeks to achieve a perfectly accurate individualised treatment plan, this approach is risky as errors and biases occur in human inference and judgement. By contrast, manual-based treatment, consistent with an actuarial approach, assumes a probabilistic model that, on average, seeks to achieve optimal but not perfect accuracy. As Einhorn (1984) commented, accepting error to make less error is likely to be a safer and more accurate strategy over a wide range of practical situations.

1.2.2.5. Feasibility

A common set of concerns centre around the feasibility of implementing manualised therapies in clinical practice. Garfield (1996) points out that whist psychotherapists in research settings are trained and monitored to conform to a specific manual, this is not the case in the average clinic. However, there are a small number of studies that show manualised treatments can be implemented in a range of clinical settings, when adequate support is provided (Addis et al., 1999). There is also a question over the number of highly technical, disorder specific, manualised treatments clinicians will be required to learn in order to best serve a diverse group of clients. Addis et al. (1999), however, suggest that clinicians only require an initial solid grounding in the common elements of manual-based interventions, and proficiency in one manualised treatment

for each broad classification of problems. Finally, concerns over client acceptability arise, questioning whether client preference should play a role in the decision making process. To date there is very little research on clients' preferences regarding manual based treatments.

1.2.2.6. Competence and Job Satisfaction

Finally there are concerns regarding therapists' ability to learn and successfully implement manualised treatment as well as fears that such treatments are uncreative, constraining, boring and unfulfilling. For example, Davidson and Lazarus (1995) suggest that manuals "obfuscate clinical artistry" (p.110). Whilst manualised therapies are designed to take precedence over clinical judgement, as previously discussed, this does not mean that therapists do not require considerable skill in implementing the specific techniques, building a therapeutic relationship and engaging clients in the change process. Wade et al. (1998) trained and supervised staff in providing manualised treatments for anxiety and major depression, finding anecdotal reports from staff that this work was the most satisfying work they had done, with every case providing new challenges and observing measurable client improvement.

It would therefore appear that whilst clinicians have a number of concerns over the use of manual-based treatments, many of these concerns are not supported by research, suggesting that manualised treatments may have something useful to offer clinical practice. Other obvious advantages are that they have been shown to be effective in controlled outcome studies, they provide structured, time limited, goal

orientated therapies, they make training and supervision easier, increase dissemination to practitioners and have the potential to facilitate clinical audit (Wilson, 1996, 1997).

Future work must therefore consider how manual based therapies can be integrated into routine clinical practice, whilst maintaining the positive effects observed in efficacy research. One specific area that needs to be considered is clinicians' adherence to the manualised protocol in a setting that no longer provides the same level of constraints and routine monitoring of therapy delivery.

1.2.3. Treatment adherence

As previously discussed, the use of manuals may provide a useful tool for disseminating treatments found to be efficacious, under conditions of high internal validity, into the less controlled arena of the clinic. It remains important, however, despite this lack of control that treatment is still implemented as intended with the therapist adhering to the manual. This is what is commonly referred to as treatment adherence, treatment fidelity or treatment integrity.

Whilst these terms are often used interchangeable within the literature there is some confusion as to their specific meaning. For example, according to Moncher and Prinz (1991), treatment fidelity, specifically, refers to two related but distinct issues. The first involves the degree to which a treatment is implemented as intended – treatment integrity. The second refers to whether treatment conditions differ from one another as intended – treatment differentiation.

Dane and Schneider (1998) on the other hand break the term treatment fidelity into four primary components: (1) Adherence, whether the treatment is delivered as intended; (2) Exposure, whether the number of sessions, length of sessions or frequency of session were implemented as intended; (3) Quality of treatment, whether the treatment was delivered with the appropriate level of skill required; and (4) Participant responsiveness, the extent to which the participants were engaged with the treatment.

For the purpose of the present literature review, the terms will be used interchangeable as used by each individual author. For the purpose of the study, however, the term treatment adherence will be used to refer to how well the delivery of a program adheres to the intended model.

The identification and development of efficacious and effective treatments has aided the development of manualised treatments that can be used in the clinic. However, we do not know, as yet, what specific components of a treatment are responsible for therapeutic change. It is therefore important that therapies are implemented in their entirety, maximising the likelihood of producing the effect sizes found in research settings. As Arthur and Blitz (2000) have commented, programs must be implemented with fidelity to the original model to preserve the behaviour change mechanisms that made the original model effective. Indeed as effective therapies are more widely disseminated in naturalistic settings, under less favourable conditions, the chances that key treatment components are modified or active ingredients are dropped becomes more likely. Therefore, whilst it is important to identify and develop efficacious and effective treatments, their widespread implementation is

unlikely to improve outcome within the clinic unless careful attention is given to treatment adherence. Boruch and Gomez (1977) in fact attribute the frequent lack of effectiveness of interventions evaluated in the clinic to the decreased integrity of laboratory proven, effective treatments.

Moncher and Prinz (1991) also point out the importance of maintaining fidelity to the treatment model within treatment outcome research as it has an effect on internal validity, external validity, construct validity and statistical power. Treatment fidelity is necessary to maintain internal validity and ensure a fair comparison of treatments. Without a measure of fidelity it would be impossible to know whether positive results are due to an effective treatment or to unknown contaminants added to the treatment; conversely it could not be concluded that non-significant results are due to an ineffective treatment or an inadequately administered treatment.

Treatment fidelity relates to external validity by making replication and comparison across studies possible, thus enabling clinicians to see whether a treatment, when administered with fidelity, is able to generalise to the specific environment in which they work. Fidelity also impacts on statistical power as by increasing internal validity the chances of observing a positive effect is increased (Smith, Glass & Miller, 1980). In addition, the monitoring and documenting of treatment fidelity is important on practical grounds, allowing for the early detection and correction of errors, thus reducing costs and improving treatments.

Despite the importance of maintaining and monitoring treatment fidelity, both in research and within the clinical setting, very few studies attend to this issue. In 1980,

Billingsley, White and Munson reviewed 108 psychotherapy treatment outcome studies and found only 5.6% assessed treatment implementation. Ten years later Moncher and Prinz (1991) evaluated 356 outcome studies and found that 45% had attended to the issues of treatment fidelity. The issue has been similarly neglected within the prevention research literature. A review by Durlack (1997) found only 5% of 1,200 published prevention studies provided data on implementation and Dane and Schneider (1998) found 39 of 162 prevention interventions contained information on fidelity, with only 13 of these studies considering the impact of fidelity on outcome.

There have been a number of strategies suggested to help clinicians maintain adherence to the treatment they are implementing. Moncher and Prinz (1991) identify four precursors to facilitate adherence: (1) a clear conceptualisation and operationalisation of the treatment, specifying the content and procedures expected, enabling clinicians to ensure that the active ingredients are delivered; (2) implementor training, providing guidance on boundaries for treatment delivery; (3) treatment manuals, providing explicit guidelines for techniques and strategies that comprise acceptable implementation; (4) supervision, enabling treatment to be assessed and, if necessary, adjusted. Whilst these four strategies may provide guidance for clinicians, increasing the likelihood that the treatment model will be adhered to, there is also a need for systematic verification of implementation, assessing whether the clinicians performed the major operations prescribed by the intervention and did not engage in non-prescribed procedures (Moncher & Prinz, 1991).

Currently, there is no widely accepted technique for conducting treatment adherence research, with great heterogeneity in the methods being used. These methods,

however, generally fall into two broad categories: (1) retrospective cataloguing of in-session behaviour, for example by reviewing client contact logs, filling out therapists' self-report rating scales, or client-report scales; and (2) observational review of therapists' conduct during sessions, for example using trained raters to code videotapes according to their adherence to the prescribed treatment (Hogue, Liddel & Rowe, 1996).

There are, however, a number of difficulties with all these approaches. The use of self-report questionnaires are likely to be affected by biases in recall, with therapists wanting to report good treatment adherence, or simply being biased in their perception of the session. Client reports are likewise limited, as they may not have sufficient knowledge or training to describe the session content at the level required. Use of observational review, however, poses its own methodological difficulties. This approach is highly resource intensive, requiring the training of raters to reliably rate adherence and the subsequent rating of multiple therapy sessions, providing a representative view that could not be gained from a rating of only one session. Typically researchers use independent raters who are blind to therapist identity, context of the session and the specific goals of the study. This, however, limits the extent the raters can take into account the stage of therapy, the clients' presenting problems, and the severity of these problems, all of which will determine the strategies the therapist should employ according to the manual. Including more contextual factors in the measurement of adherence, however, makes it more difficult to obtain a reliable measure with good inter-rater reliability. Raters can also be misled due to halo effects, where perceived-to-be competent therapists are given high ratings on all constructs, independently of whether they meet the necessary criteria or

when raters allow one impressive interaction to generate positive spill-over into ratings of similar items (Hogue et al., 1996; Waltz et al., 1993).

Another major problem in the assessment of treatment fidelity is the differentiation between adherence and competence. Measures of fidelity often focus on whether a therapist used the interventions and approaches prescribed by the treatment manual and avoided the use of interventions and procedures proscribed by the manual. However many of these measures fail to take into consideration the level of skill, or competence, shown by the therapist in delivering the treatment. According to Waltz et al. (1993), competence refers to the extent to which the therapist takes into account relevant aspects of the therapeutic context. These include the clients' degree of impairment; the clients' specific problems; the clients' life situation and stresses; and factors such as stage in therapy, degree of improvement and sensitivity to the timing of interventions. It seems obvious that, despite good adherence in the administration of a specific therapy, therapeutic change is unlikely if these techniques and procedures are delivered incompetently. Garfield (1996) asks the question, is the therapeutic approach more important than the overall therapeutic skill of specific therapists? A study by Luborsky et al. (1985) found important differences between therapists participating in a research project that used a manualised treatment, indicating that despite the use of manuals, individual differences still exist and therefore therapist competence must also be accounted for.

Finally, there has been a great deal of evidence within psychotherapy research that demonstrates the importance of the therapeutic relationship with regard to therapeutic change. Hogue et al. (1996) in fact suggest a framework for intervention strategies

that consists of three pillars: adherence, competence and the therapeutic alliance. This will be discussed in more detail in the following section.

Whilst there are only a few studies that have looked into the effects of treatment adherence on outcome, those that have, have found positive results. A study by Henggeler, Melton, Brondino, Scherer and Hanley (1997) looked at the effectiveness of multisystemic therapy (MST) for violent and chronic juvenile offenders in community mental health settings which lacked the clinical oversight of a multisystemic therapist expert. They found that whilst the level of decreased criminal activity was not as favourable as had been previously observed on other trials of MST, outcomes were substantially better in cases where treatment adherence ratings were high. These results were replicated in a subsequent study examining the effects of MST with substance abusing delinquents (Henggeler, Pickrel & Brondino, 1999), with high adherence scores being associated with decreased criminal activity and out of home placements. A review paper by Mihalic (2002), on violence prevention initiatives in the US, consistently found that the closer an intervention adheres to the original design, the greater the degree of behavioural change. Similarly, Kam, Greenberg and Walls (2003) found that a high degree of classroom implementation by teachers contributed to the success of a school based violence prevention program.

Adherence to the treatment protocol therefore seems important if the findings of efficacy research are to be replicated in the clinic. However, as Waltz, Addis, Koerner and Jacobson (1993) have pointed out, this should not be at the expense of therapeutic competence, skill and other common factors such as the therapeutic alliance.

1.2.4. Common factors in therapy

Traditionally, change factors within psychotherapy have been divided into 'specific factors', those that derive from an identified theory or model of therapy, and 'common factors', those that are shared by all forms of therapy. Over the past few decades researchers have sought to discover which of these two factors has the greatest impact on therapeutic change, with many investigators concluding that the common factors carry a substantial amount of the treatment outcome variance (Nathon, Stuart & Dolan, 2000).

One of the most widely researched common factors is that of the therapeutic alliance. Horvath and Greenberg (1994) conclude that client ratings of the therapeutic alliance are the best predictors of clinical outcome. Lambert and Bergin (1992) completed a review of outcome studies and suggested that about 30% of psychotherapy outcome variance is attributable to therapist variables such as empathy, warmth and acceptance, and Swatberg, Seltzer and Stiles (1998) conclude that the therapeutic alliance is the most important factor in determining positive therapeutic outcomes. Bachelor (1991) similarly found that client rated provision of help, warmth, caring and emotional involvement on behalf of the therapist were key factors in determining therapeutic outcome.

A review paper by Ackerman and Hilsenroth (2003) found a number of therapist personal attributes had a positive influence on the therapeutic alliance; these included flexibility, honesty, respectfulness, trustworthiness, confidence and warmth. Such factors may therefore be important mechanisms of therapeutic change.

It is therefore important that manualised therapies incorporate common factors into their manuals and encourage the development of a warm, empathic therapeutic relationship that has been shown to promote therapeutic change. The monitoring of these common factors will also be important during treatment adherence research in order to determine the degree to which therapeutic outcome is a result of the specific manualised therapy and/or non-specific common factors that are present across all therapeutic domains.

1.3. Rationale and aims of the present study

1.3.1 Summary and rationale

As already mentioned, the Webster-Stratton BASIC parenting program has been found in numerous randomised trials to reduce childhood conduct problems, improve positive child-parent interactions and to reduce parents' use of violent forms of discipline. However, these studies have been conducted under conditions of high internal validity with multiple checks of treatment adherence to ensure the program is delivered as intended (Webster-Stratton, 1998). From these studies, therefore, we can only conclude that this parenting program works if it is implemented in its entirety as set out in the leaders' manual.

Within everyday clinical practice, however, there are fewer resources and opportunities to provide the same level of adherence monitoring found in these research trials. This therefore questions the extent to which parenting programs conducted in the clinic are delivered as they are intended and to the level required to produce the effects that have been found in the efficacy research. The study by Scott

et al. (2001), conducted in regular NHS settings with standard referrals to child mental health service and using regular clinic staff to carry out the intervention, went some way to answering this question. Their results showed that the Webster-Stratton BASIC parenting program effectively reduced antisocial behaviour in children in regular NHS settings. The study, however, still incorporated exclusion criteria and built in regular checks of treatment adherence, leaving the question as to whether parenting groups are effective in regular clinical practice largely unanswered.

The role of treatment adherence within the parenting program literature is currently unclear as there has been no study that has investigated the relationship between treatment adherence and outcome. In the Scott et al. (2001) study, variability in outcome between the treatment groups was observed and researchers speculated that treatment adherence might be playing a role (personal correspondence), but as yet this has not been examined.

Within the UK there has been a recent increase in government funding to provide parenting groups, largely due to the beneficial effects found in research trials. Therefore, as these groups become more widely disseminated, it can be hypothesised that the number and quality of checks to treatment adherence will diminish. This could reduce the likelihood that the program will be delivered as it was intended and potentially reduce the effectiveness of the program in changing parenting behaviours and subsequently reducing child conduct problems. There is, therefore, a need to monitor adherence to the program within everyday clinical practice and tools need to be developed to enable this to occur.

1.3.2. Aims of the present study

The research had two aims:

1) The central aim was to design an instrument that could be used by trained raters to rate treatment adherence to the Webster-Stratton BASIC parenting program.

2) A secondary aim was to explore the relationship between treatment adherence and outcome, as measured by a reduction in child conduct difficulties, using the developed instrument to measure adherence.

2. Method

The research reported in this thesis builds on a research trial that commenced in 1996, studying the effectiveness of parenting groups for childhood antisocial behaviour in regular clinical practice (Scott et al., 2001). The first part of the present study involves the development of an instrument for coding adherence to the Webster-Stratton BASIC parenting program, and the second part involves its subsequent use in an exploratory analysis, looking at the relationship between treatment adherence and outcome. This chapter will therefore firstly describe the method for developing the coding instrument and secondly outline the method for the exploratory part of the study, including a brief overview of the methodology used in the Scott et al. research trial.

2.1. Ethics

Ethical approval for the project was granted by the Bethlem and Maudsley NHS Trust/Institute of Psychiatry ethical committee (see Appendix 1).

2.2. Part 1: Development of a measure of treatment adherence

The instrument for measuring treatment adherence to the Webster-Stratton BASIC parent-training program was developed by a co-worker (AC) and myself. Both of us had been involved in previous research using the BASIC parent-training program and were therefore familiar with this model of parent training. Initially, time was taken to re-familiarise ourselves with the leaders' manual and with the material used in the groups.

The instrument was designed as an observational measure, using global rating scales of videotaped sessions. The videotapes used in the initial stages of measure development were taken from the SPOKES research trial (Scott & Sylva, 2002). The tapes used in the measure of inter-rater reliability were taken from both the SPOKES trial and the Scott et al. (2001) trial.

2.2.1. Variables to be rated

The leadership materials, supplied with the program, include a peer and self-evaluation form provided to help maintain leadership adherence to the program. This formed the basis for our coding instrument as it provided a comprehensive overview of the program's expectation of group leaders. The form comprises 53 individual criteria grouped into six categories: group process skills, leadership skills, relationship building skills, knowledge, methods and parent responses.

The first step was to adapt the form for use with videotapes of parent training sessions. Criteria were excluded if 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. Further adaptations were made with criteria being 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 labelled "encourage participation". Table 1 shows the content of the original self-evaluation form and the resulting list of 27 variables, after all the adaptations had been made.

Table 1. Adaptations made to peer and self-evaluation form

Peer and self-evaluation form	Adaptations made
1. Builds rapport with each member of the group	1. Encourage participation
2. Encourages everyone to participate	
3. Views every member of the group as equally important	
	2. Practical arrangements
4. Models open ended questions to facilitate decision	3. Open ended questions
5. Reinforces parents' ideas and fosters self learning	4. Reinforces parents' ideas and fosters self learning
6. Fosters idea that parents will learn from each other	
7. Helps parents learn to support and reinforce each other	
8. Creates an atmosphere where parents feel they are decision-makers and discussion and debate are paramount	
9. Encourage parents to problem solve where possible	
10. Creates a feeling of safety amongst the group	5. Feeling of safety amongst the group
11. Establishes ground rules for group	6. Ground rules
12. Started and ended meeting on time	Not included: Could not be code from videotape
13. Explained Agenda for session	7. Agenda
14. Reviews homework from previous session	8. Reviews homework and gives feedback and praise
15. Review homework and gives feedback	
16. Summarises and restates important points	9. Summarise and restates main points
17. Focuses group on key points	10. Leads the discussion
18. Imposes sufficient structure to facilitate group process	
19. Prevents side tracking by participants	
20. Knows when to be flexible and allow a digression for an important issue and knows how to tie it into sessions content	
21. Maintains leadership of group	
22. Anticipates potential difficulties	11. Predicts relapse
23. Predicts relapses	
24. Predicts behaviours and feelings	
25. Encourages generalisation of concepts to different settings and	12. Generalisation of concepts

situations	
26. Encourages parents to work for long term goals as apposed to quick fixes	Not included: unclear what each parents' long term goals are
27. Helps group focus on positive	13. Focus and highlight the positive
28. Uses humour and fosters optimism	
29. Identifies each families strength	
30. Balances group discussion on affective and cognitive domain	14. Validates/supports parents' feelings
31. Validates and supports' parents feelings (reflective statements)	
32. Reviews handouts and homework for next week	15. Issue homework
33. Emphasizes the importance of homework	
34. Evaluates session	Not included: could not be coded from videotape
35. Shares personal experiences when appropriate	16. Collaborative vs. expert model
36. Fosters a partnership or collaborative model (as opposed to an expert model)	
37. Fosters a coping model as opposed to a mastery model of learning	17. Fosters a coping vs. mastery model of learning
38. Normalises problems when appropriate	
39. Reframes experiences from the child's viewpoint and modifies negative attributions	18. Reframes experiences from the child's viewpoint and modifies parents' negative attributions
40. Strategically confronts, challenges and teachers parents when necessary	19. Confronts, challenges, teaches parents when necessary
41. Identifies and discusses resistance	Not included: occurred too infrequently
42. Advocates for parents	Not included: could not be coded from videotape
43. Demonstrates knowledge of content covered in session	20. Knowledge of content covered in session
44. Explains rational for principles covered in clear, convincing manner	
45. Prepares materials in advance of session and is prepared for group	21. Material prepared in advance
46. Integrates parents' ideas and problems with important content and child development principles	22. Integrate parents' ideas with important content and child development principles
47. Uses analogies and metaphors to explain theories or concepts	23. Use of analogies and metaphors to explain theories
48. Uses videotaped examples efficiently and strategically to trigger group discussion	24. Vignettes used to trigger discussion
49. Uses role play and rehearsal to reinforce learning	25. Use role play to reinforce learning

50. Uses modelling by self or other group members when appropriate	Not included: occurred too infrequently
51. Parents appear comfortable and involved in the session	26. Parents appear comfortable and involved
52. Parents complete homework, ask question and are active participants	27. Parents refer to using the skills learnt during the week
53. Parents complete positive evaluations of sessions	Not included: could not be coded from videotape

After these initial adaptations were made, AC and I carried out preliminary testing of the 27 variables to establish the feasibility of rating these from videotapes: this involved videotapes of sessions that would not be used later to formally establish reliability. This initial testing stage was carried out by both raters together, allowing for discussion of each variable. 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.

2.2.2. Initial operationalisation

After deciding on which criteria to include in the instrument, the next step was to operationalise each variable, defining exactly what was being rated, thus enabling other people to use the instrument in the future. This process took into consideration whether the measure was one of quantity, e.g. how many vignettes were shown during the session, or quality, e.g. the efficient use of vignettes to discuss key parenting issues. It also enabled us to take into account the difference in leadership style expected at different stages of the program; for example, during early sessions group discussion might need to be facilitated by using open ended questions, whereas in later sessions discussion might occur spontaneously as parents become familiar with the program format. Similarly it would be expected for ground rules to be established

and mentioned in early sessions, or when new members join, but in later sessions they might only be referred to if they are being broken.

A five point rating scale was devised for each variable, with each point on the scale defined in order to clarify exactly what was required for each specific code. During this process, practice videotapes that would not be included in the measure of inter-rater reliability, were continually being coded, providing the fuel for discussion over definitions and enabling us to informally monitor our inter-rater agreement.

2.2.3. Finalising the coding system

A number of decisions had to be made at this point. The first was whether to code each individual leader's adherence, or the overall combined leadership's adherence to the model. It was decided that what was important was whether the overall delivery of the program was delivered as intended, independently of how each leader faired individually. Secondly, we discussed whether it was important to only code the leader's behaviour, as recommended by Hogue et al. (1996), or to include what was happening in the group due to the behaviour of the parents. For example, in early groups leaders may have to work at encouraging parents to participate, creating a feeling of safety in the group. In later groups, however, this may occur more spontaneously due to the work that has already been put in to create a safe atmosphere. It was felt that both these observations should be included in the measure for "feeling of safety amongst the group", therefore including some parental behaviours as contributing to the measurement of adherence. A third issue that was highlighted at this stage was the need to buffer against halo effects, whereby a rating for a given item is biased by a rating awarded to another item or by a global rating of

the session as a whole (Hogue et al., 1996). For example, high ratings might be made on all items because of a general opinion that the therapist is competent, or conversely low ratings being given across items due to a negative impression of the leader. It was therefore emphasised that each variable should be coded separately as an independent entity, with an effort to avoid being influenced by ratings given to any other items. Writing notes, and recording examples of each variable whilst watching the videos helped this process, as did referring to the individually operationalised five point scoring system when deciding on the code for each variable.

The final coding system is shown in Appendix 2 as it is too long to include in this chapter. Appendix 3 also shows the form that was used during the rating of each videotape.

2.2.4. Coding of the video tapes

A total of 25 videotapes were rated independently by both AC and myself, seven taken from the SPOKES trial and 18 from the Scott et al. (2001) trial, including at least one tape from each of the 15 parenting groups. AC and I met regularly to compare our scores and make minor adjustments or clarifications to the coding criteria where necessary.

2.2.5. Method of analysis

Pearson's correlations were used to measure the inter-rater reliability between the ratings of the two researchers. This is reported in chapter 3.

2.3. Part 2: Relationship between treatment adherence and outcome

The exploratory part of the study, using the adherence measure described above to examine the relationship between treatment adherence and outcome, builds on a research trial that commenced in 1996. I will therefore give an overview of the methodology used in this trial before outlining the methods used in the current study. (For a more detailed presentation of Scott et al's methods see Scott et al., 2001).

2.3.1. Overview of Scott et al's (2001) study

Scott et al's (2001) research trial set out to examine whether the Webster-Stratton behaviourally based BASIC parenting program was an effective treatment for antisocial behaviour in children, when carried out in regular clinical practice. The study took place from 1995-1999 in four NHS child and adolescent mental health services in London and the South East of England.

2.3.1.1. Design

The study was a randomised controlled trial allocating the parents from each NHS centre to the intervention or control group using a permuted block design. Parents of 141 children were allocated, 90 to the parenting group and 51 to the waiting list control.

2.3.1.2. Participants

Eligible children were all those aged between 3-8 years referred to their local NHS service with antisocial behaviour. Exclusion criteria were clinically apparent major depression, hyperkinetic syndrome, or any other condition requiring separate

treatment. Parents had to be able to understand English and attend a weekly parent-training group.

Of the 90 children allocated to the parenting groups, 84 completed the trial. Of these, 61 were male (72.6%) and 23 were female (27.4%). The mean age of the children was 5.4 years.

2.3.1.3. Intervention

The Webster-Stratton BASIC parenting program was administered to groups comprising 6-8 parents for two hours each week over 13-16 weeks. During the period of the research trial 15 parenting groups were delivered across the five different sites. Thirteen group leaders took part in the study working in different pairing combinations. Intervention sessions were videotaped and weekly supervision meetings were held to promote adherence to the manual.

2.3.1.4. Measures

Measures were taken from the mothers on entry to the trial and after completion of the intervention five to seven months later. These included demographic details, six measures of child behaviour and one of parent behaviour. The 'Parent Accounts of Child Symptoms Interview' (PACS) (Taylor, Chadwick, Heptinstall & Danckacets, 1996) was used as the primary outcome measure for antisocial behaviour. This is a well validated semi-structured interview that uses investigator based criteria to assess the frequency and severity of behaviour on three different sub-scales: disruptive behaviour, hyperactivity and emotional problems. (Due to its length only the disruptive behaviour subscale is shown in Appendix 4).

2.3.1.5. Findings from Scott el al. (2001)

When compared with population norms the participants' mean scores at time 1 were in the 97th centile for conduct problems, above the 90th centile for hyperactivity and above the 78th centile for emotional problems. There were no significant differences found between the intervention and control groups. Family demographics showed that the majority of families in the trial were poor and disadvantaged with no differences being found between the groups.

Results from the PACS showed that, for antisocial behaviour, control children showed no change but intervention children showed a large improvement, falling from the 97th centile to the 82nd centile on population means, which is within the normal range (effect size between groups 1.06). A similar picture was observed on all the other outcome measures. Differences in outcome were observed between the 15 interventions groups.

Regression analyses found that the presence of ADHD and the age of the child were significant predictors of outcome, whereas gender, social class, emotional symptoms in the child, family structure and ethnicity did not significantly predict outcome.

2.3.2. Procedure of the current study

The current study aimed to look at the relationship between treatment adherence to the Webster-Stratton BASIC program and outcome, using data from Scott et al.'s study. In order to do this videotapes of selected sessions were rated for adherence using the measure described in section 2.2.

46

A total of 60 tapes were rated, 4 from each of the 15 parenting groups. The tapes were not viewed prior to selection but were chosen on the basis of where they were from in the program, i.e. one early session (weeks 2-4), two middle sessions (weeks 5-8) and one late session (weeks 9+). The two researchers who developed the coding instrument, who were blind to treatment outcome, rated the tapes. Eighteen tapes were taken form the earlier reliability study (see section 2.2.4); these were rated by both researchers and any discrepancies in scores were discussed and a final score to be included in the analysis was agreed upon. A further 11 tapes were coded by AC and the remaining 31 tapes were coded by myself.

Details of the final set of adherence variables that were included in the statistical analyses will be given in Chapter 3.

3. <u>Results</u>

The first section of this chapter presents the data concerning the measure development part of the study. The second section reports the results of the analysis examining the relationship between treatment adherence and outcome.

3.1. Part 1: Development of a measure of treatment adherence

3.1.1. Inter-rater reliability

Pearson's correlations were used to measure inter-rater reliability for each of the adherence variables. Twenty-five videotapes were rated by both researchers for each of the 27 variables. As can be seen in Table 2, nine of the 27 variables were found to have an inter-rater correlation greater than 0.6, which is an acceptable standard for inter-rater reliability (Barker, Pistrang & Elliott, 2002; Cone, 1999).

Of the nine variables that reached an acceptable level of reliability, five ("Arrangements", "Vignettes", "Role play", "Review homework", and "Issue homework") were more concrete in nature making them easier to define and observe. The four remaining variables that also reached an acceptable level of reliability, "Encourage", "Reinforce ideas", "Summarise" and "Confronts", tapped somewhat less concrete behaviours but had very specific definitions which perhaps made them easier to score.

Of the remaining 18 variables that did not obtain an acceptable level of inter-rater reliability only two related to fairly specific tasks, "Ground rules" and "Agenda". Whilst is it is slightly surprising that these two variables were not reliably rated, observations show that the inter-rater reliability for agenda appeared to improve over

Table 2. Inter-rater reliability

Variable	Abbreviation	Pearson's Correlation
Encourage participation	Encourage	.67
Practical arrangements	Arrangements	.88
Open ended questions	Open ended	.38
Reinforce parents ideas/fosters self learning	Reinforce ideas	.70
Feeling of safety amongst group	Safety	-.35
Ground rules	Ground rules	.35
Agenda	Agenda	.57
Summarize/restate main point	Summarize	.66
Leads the discussion	Discussion	.39
Generalisation of concepts	Generalisation	.33
Predicts relapse	Predicts relapse	.47
Focus/highlights positive	Focus on positive	.46
Validates/supports parents feelings	Validates feelings	.47
Collaborative vs. expert model	Collaborative	.30
Fosters coping vs. mastery model	Coping	.51
Reframes from child's perspective & modifies negative attributions	Reframes	.16
Confronts/challenge /teach when necessary	Confronts	.71
Knowledge of content covered	Knowledge	.46
Integrates child developmental principles	Developmental	.56
Use of analogies and metaphors	Analogies	.46
Materials prepared in advance	Preparation	.48
Vignettes used to trigger discussion	Vignettes	.77
Role play used to reinforce learning	Role play	.87
Review homework and give feedback/praise	Review homework	.71
Issue homework	Issue homework	.85
Parents appear comfortable and involved	Comfortable	-.27
Parents refer to using skills learnt during week	Use skills	.56

Note. abbreviated variable names only will be used for the remainder of the thesis

time, suggesting that the raters were using the definitions slightly differently early on. Observing the scores of the "Ground rules" variable suggests that most of the time the raters used similar scores, but occasionally were two to three points out. This seemed to occur as it only took a rater to miss one example of a broken rule that was not addressed by the group leaders for scores to differ considerably on this variable.

Two variables are notable in that they received negative inter-rater correlations: "Comfortable" and "Safety". It was noticed during coding that these variables appeared to be measuring the same construct, with the raters scoring them very similarly for each tape. Both variables also tended to be rated within a restricted range, rarely receiving less than a four (on the five point scale), which may explain the poor inter-rater reliability for these variables.

Inter-rater reliability was also particularly low (.30 or less) on two additional variables: "Collaborative" and "Reframes". During measure development and coding it was noticed that it was particularly difficult to define these variables. This probably explains the poor reliability scores and suggests that the raters were measuring different things.

In the remaining sections of the chapter, only those variables obtaining an inter-rater reliability coefficient of 0.6 or greater will be included. One of those variables "Confronts", occurred infrequently, i.e. it was rated in only nine of the 25 videotapes analysed. It will therefore be excluded from the following analyses, leaving 8 variables for inclusion. The analyses included in the remaining sections also include data from all 60 videotapes rated by the two researchers.

Table 3. Mean scores and standard deviations for each adherence variable across group

Variable	Group															Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
Encourage	4.50 (1.00)	4.25 (.96)	5.00 (.00)	3.50 (.58)	4.25 (.50)	4.25 (.96)	4.00 (.82)	3.25 (.50)	4.50 (5.78)	4.50 (1.00)	4.50 (1.00)	4.00 (.82)	3.75 (.96)	4.75 (.50)	4.25 (.96)	4.25
Arrangements	4.50 (1.00)	4.50 (.58)	4.50 (1.00)	3.50 (1.00)	3.00 (.00)	3.00 (.00)	4.00 (1.15)	3.50 (.58)	4.25 (.96)	3.25 (.50)	4.00 (.82)	5.00 (.00)	3.25 (.50)	4.25 (.96)	5.00 (.00)	3.96
Reinforce ideas	4.00 (.82)	4.50 (1.00)	4.25 (.96)	2.00 (.82)	3.50 (.58)	4.75 (.50)	3.00 (.82)	3.00 (.00)	4.75 (.50)	4.00 (1.15)	5.00 (.00)	4.75 (.50)	3.25 (5.00)	5.00 (.00)	3.50 (1.00)	3.95
Summarize	4.75 (.50)	4.75 (.50)	5.00 (.00)	2.50 (.58)	3.00 (1.41)	4.75 (.50)	3.25 (.50)	3.25 (.50)	4.50 (.58)	5.00 (.00)	5.00 (.00)	4.25 (.50)	3.50 (1.00)	5.00 (.00)	4.50 (.58)	4.20
Vignettes	3.5 (1.91)	4.75 (.50)	4.00 (.00)	3.25 (.50)	3.75 (.96)	4.00 (1.15)	3.50 (.58)	3.25 (.50)	4.50 (1.00)	5.00 (.00)	4.50 (1.00)	4.50 (.58)	4.25 (.50)	3.75 (1.89)	3.50 (.58)	4.00
Role Play	3.75 (.50)	4.25 (.96)	2.50 (1.91)	2.25 (1.5)	2.00 (1.41)	3.75 (.96)	2.25 (1.50)	2.00 (2.00)	4.00 (.82)	3.50 (1.91)	3.50 (1.29)	3.50 (.58)	2.25 (1.5)	4.25 (.50)	2.25 (1.89)	3.07
Review homework	3.75 (1.89)	4.25 (.50)	5.00 (.00)	2.25 (.50)	4.25 (.96)	4.00 (1.15)	3.00 (1.41)	2.75 (.96)	5.00 (.00)	4.50 (.58)	4.75 (.50)	3.50 (.58)	4.25 (.50)	4.75 (.50)	3.75 (.96)	3.85
Issue homework	3.50 (1.00)	3.75 (.50)	3.50 (.58)	2.50 (.58)	3.75 (.50)	3.25 (.96)	3.25 (.50)	2.75 (.96)	2.50 (1.00)	3.50 (.58)	3.75 (.96)	3.75 (.50)	2.00 (.82)	3.00 (.82)	3.50 (.58)	3.22

Note.   All variables were rated on a scale of 1 to 5, where high scores indicate greater adherence to the model.
Numbers in brackets refer to standard deviations.

3.1.2. **Adherence scores**

Table 3 shows the mean scores for each of the eight adherence variables across the 15 groups. The overall picture shows that adherence is generally good, with the majority of scores reaching above three. Two groups, four and eight, appear to be scoring consistently lower on most variables, with a number of variables rated below 3. Observing the ratings across groups suggests that two variables, "Role-play" and "Issue homework", received lower ratings. That is, adherence seemed to be generally lower on these two aspects of leaders' behaviour.

3.1.3. Inter-correlations of variables

Table 4 shows the correlations between each of the variables. Whilst "Issue homework" and "Arrangements" do not correlate highly with any of the other variables, the remaining six variables are generally highly correlated with each other.

3.1.4. Factor analysis

Following on from the inter-correlations a factor analysis was conducted to see how the variables loaded onto different factors. A principal component factor analysis with varimax rotation was used. As can be seen in Table 5, the eight variables load onto two separate factors, supporting what was observed in the above correlations.

The variables that load onto factor 1 ("Encourage", "Reinforce ideas, "Summarize", "Vignettes", "Role play" and "Review homework") seem to be those that tap aspects of leaders' behaviour concerned with group facilitation. Whilst "Vignettes" and "Role play" might appear to be related to more practical aspects of the program, these variables included a substantial qualitative component concerned with group

Table 4. Inter-correlations of adherence variables

	1	2	3	4	5	6	7
1. Encourage	-						
2. Arrangements	.34	-					
3. Reinforce ideas	.70**	.34	-				
4. Summarize	.80**	.43	.84**	-			
5. Vignettes	.40	.39	.62*	.56*	-		
6. Role play	.50	.29	.81**	.74**	.56*	-	
7. Review homework	.83**	.11	.76**	.73**	.64*	.48	-
8. Issue homework	.47	36	.43	.44	.24	.23	.23

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

Table 5. Varimax rotated factor loadings

Variables	Factor	
	1	2
Encourage	.73	.44
Arrangements	.47	.85
Reinforce	.87	.35
Summarize	.81	.46
Vignettes	.81	-.08
Role-play	.76	.24
Review homework	.90	.06
Issue homework	.22	.72

discussion; for example, the "Vignettes" and "Role play" variables were concerned with how the vignettes and role plays were discussed within the group and not just simply whether they occurred. This first factor will therefore be labeled "Group facilitation".

The variables that loaded onto factor 2, "Arrangements" and "Issue homework", tap more practical aspects of leader behaviour. This factor will therefore be labeled "Practicalities".

3.2. Part 2: Relationship between treatment adherence and outcome

Linear regression analysis was used to see whether adherence predicted outcome, as measured by post-treatment disruptive behavior. The two factors from the factor analysis, "Group facilitation" and "Practicalities" were used in these analyses.

One option here would be to use a change score, i.e. the difference between pre- and post-treatment disruptive behaviour, as the dependant variable. However a statistically preferred option is to control for pre-treatment disruptive behavior by entering it into the regression analysis as one of the independent variables. Therefore, post-treatment disruptive behaviour was used as the dependant variable and pre-treatment disruptive behaviour as an independent variable. The two adherence factors, "Group facilitation" and "Practicalities", were then entered in as independent variables, in parallel rather than hierarchical equations.

Table 6. Regression analysis of treatment outcome on adherence

	Variable	B	SE B	Beta	Sig.
Model 1	Pre-treatment disruptive behaviour	.51	.18	.44	.00
Model 2a	Pre-treatment disruptive behaviour	.51	.11	.45	.00
	Group facilitation	-.19	.08	-.24	.01
Model 2b	Pre-treatment disruptive behaviour	.52	.11	.46	.00
	Practicalities	-.08	.10	-.09	.40

Note.   There is no hierarchical regression from Model 2a to Model 2b.

The regression analysis (see Table 6) found that 20% of the variance in post-treatment

disruptive behavior was accounted for by pre-treatment disruptive behavior $(R^2 = .20,$

$F(2,82) = 20.28$, p = .00). When "Group facilitation" was also included in the model

the variance explained increased by 6% (increase in $R^2 = .06$, $F(1,82) = 6.36$, p =

0.01). "Practicalities" did not add to the variance explained by pre-treatment

disruptive behaviour (increase in $R^2 = .01$, $F(1,82) = .72$, p = .40).

The results from the Scott et al. study (2001) found that age of child and hyperactivity

before treatment influenced outcome. These variables were therefore included in a

regression model as independent variables. The regression analysis (see Table 7)

showed that these two variables together increased the predictive value of the model,

over and above pre-treatment disruptive behaviour, accounting for 28% of the

variance (increase in $R^2 = .08$, $F(2,81) = 4.41$, p = .02). When "Group facilitation"

was entered into the model, the variance explained increased to 30%, but this was not

significant (increase in $R^2 = .03$, $F(1,80) = 2.94$, p = 0.09).

Table 7. Regression analysis of treatment outcome including age of child and hyperactivity

	Variable	B	SE B	Beta	Sig.
Model 1	Pre treatment disruptive behaviour	.51	.11	.44	.00
Model 2	Pre treatment disruptive behaviour	.46	.11	.40	.00
	Age of child	-.05	.32	-.17	.10
	Hyperactivity	.14	.08	.19	.07
Model 3a	Pre treatment disruptive behaviour	.47	.11	41	.00
	Age of child	-.04	.03	-.12	.25
	Hyperactivity	.12	.07	.17	.10
	Group facilitation	-.14	.08	.17	.09
Model 3b	Pre treatment disruptive behaviour	.47	.11	.41	.00
	Age of child	-.05	.03	-.16	.11
	Hyperactivity	.14	.08	.19	.07
	Practicalities	-.07	.10	.07	.45

Note.   There is no hierarchical regression form Model 3a to Model 3b.

4. Discussion

The primary aim of this study was to design an instrument that could be used by trained raters to measure adherence to the Webster-Stratton BASIC parenting program. A secondary aim was to use this instrument to explore the relationship between treatment adherence and outcome, as measured by a reduction in child conduct difficulties. This chapter therefore, will firstly report the main findings from both arms of the study before discussing each one in more detail.

4.1. Summary of findings

Results from the measure development arm of the study found that of the 27 variables included in the instrument used to measure treatment adherence, nine achieved adequate inter-rater reliability. These nine variables showed that treatment adherence, across both variables and groups, appeared generally high. Of the nine variables one was later dropped due to its low frequency in occurrence, leaving eight variables for inclusion in the further analyses. A factor analysis found that these variables loaded onto two separate factors, the first including variables measuring leader behaviours concerned with group facilitation (labelled "Group facilitation") and the second concerned with practical aspects of the program (labelled "Practicalities"). Regression analyses showed that the group facilitation factor was a significant predictor of treatment outcome, but this finding was eradicated when age of child and hyperactivity were taken into account.

4.2. Measure development

Of the 27 variables included in the measure of treatment adherence, nine achieved an acceptable level of inter-rater reliability (>.6). Of these nine variables, seven obtained correlations greater than .7 and three of these were above .85, suggesting a good level of reliability between the two raters. These variables tended to be those that were more concrete in nature, were more easily defined on the five-point scale and tended to occur more frequently, such that a measure of frequency could be included in the definitions of these variables.

4.2.1. Difficulties in rating adherence

Eighteen of the variables included in the measure failed to obtain an adequate level of inter-rater reliability. These results are disappointing and point to the difficulty in designing an observational global rating scale to measure leaders behaviour in parent training groups. This section will therefore discuss the main difficulties encountered.

4.2.1.1. Defining the variables

One of the key strategies for maximising inter-rater reliability is to provide raters with well defined variables and good examples of what constitutes a specific rating (Barker et al., 2002; Cone, 1999). During the early stages of measure development it was apparent that some of the variables included in the measure were more difficult to define than others, particularly those that were more abstract in nature and required the observer to quantify an overall impression rather than record specific concrete examples. For example, the variable "Collaborative" required the raters to rate the leaders on their overall approach to facilitating the group. Whilst there were a number of specific examples for the raters to look out for, for example, the leaders

sharing their own personal parenting difficulties, the variable tended to be rated from an overall impression of the leaders' ability, increasing the difficulty in obtaining a reliable measure.

It was also noted that some of the variable definitions were rating factors more related to outcome rather than adherence. For example, the definition of "Safety" incorporated aspects of parental behaviour synonymous with the feeling of safety, rather than defining the behaviours of the leaders that would promote safety within the group.

It was also difficult, at times, to define each point on the five-point rating scale, enabling raters to differentiate between the different ratings. Whilst some variables seemed to be easily differentiated (e.g. "Arrangements"), others fitted less easily on to the scale, with no obvious factors that differentiated each rating (e.g. "Reframes"). This made defining these variables more difficult, again reducing the likelihood that they would achieve adequate inter-rater reliability.

4.2.1.2. Coding complex variables: the use of global rating scales

The instrument made use of global rating scales in order to make an overall judgement of the behaviours that were observed during the course of the parenting group. This allowed for the coding of complex events, particularly around issues of therapeutic process and moved the ratings away from simply recording the frequency of specific behaviours. Global rating scales are, however, generally less precise than the more behavioural methods, reducing inter-rater reliability (Barker et al, 2002; Cone, 1999).

The use of global rating scales, however, allowed each variable to include both quantitative and qualitative aspects. For example, the rating scale for the variable "Vignettes" included both the number of vignettes that were shown as well as the quality of the discussion that followed, making it a useful learning experience for the parents. Therefore, in order for leaders to obtain a five, the highest rating, on this variable, they must have shown a sufficient number of vignettes and used discussion efficiently and effectively to have drawn out the main teaching points of each vignette.

This is an important aspect of treatment adherence, as mentioned in Chapter 1. A number of measures of adherence focus solely on whether the prescribed interventions were administered, and fail to account for the skill and level of competence demonstrated by the therapist (Hogue et al., 1996; Waltz et al., 1993). The Webster-Stratton BASIC parenting program includes in its leaders' manual information on what interventions should be delivered, as well as how it should be delivered, in order to maximise the program's effectiveness. It was therefore important to incorporate both of these aspects into our measure of treatment adherence, necessitating the use of global rating scales.

Whilst the inclusion of both quantitative and qualitative aspects of program delivery are important within measures of adherence, it does increase the complexity of defining each variable and reduce the likelihood of achieving adequate inter-rater reliability. One possible solution would be to separate ratings for these two aspects of program delivery onto two separate subscales.

4.2.1.3. The number of variables and their frequency of occurrence

Another difficulty was the inclusion in the measure of so many variables used to rate adherence. Keeping all 27 variables in mind whilst observing the videos was difficult, increasing the likelihood that some examples of variables were missed and therefore not included in the rating of adherence. This also made it difficult to use the global rating scales, as the raters had to keep in mind their impressions of each variable over the 3 hour-long videos.

Many of the variables that did not reach an adequate level of reliability were those that were not frequently observed during the groups, for example, generalisation of concepts. Their low frequency may have reduced the raters' attention to their occurrence, increasing the likelihood that important examples were missed and therefore not included in the adherence rating. This may explain the poor inter-rater reliability scores that some of these variables achieved, as missing recording one example of these infrequently observed variables would have a greater impact on the overall rating as opposed to missing recording an example of a more frequently occurring variable. The low frequency variables may also have reduced the raters' ability to recall their overall impression of that specific variable as required on many of the global rating scales.

The raters attempted to aid this process by making written notes throughout the coding process. They also made continuous reference to the coding manual whilst scoring each variable. However, it was found to be easier to record examples of variables that occurred more frequently or examples that very obviously fitted a

specific variable definition. Again those variables that were less concretely defined were more difficult to record during the rating process, reducing the chances that they would be rated reliably.

Making written notes and referring to the manual was also aimed at reducing the likelihood of halo effects influencing the rating process where by an overall positive impression of a group spills over into scoring highly across all variables, despite not satisfying the specific criteria. Waltz et al. (1993) suggest that raters should be cautioned about these and should be reminded that each item is to be rated as a separate, independent entity, not being influenced by ratings given to other items. Again, the use of written notes is more likely to aid the independent coding of the more frequent, concrete variables where specific examples are easier to record than the less frequent, more abstract variables making halo effects more likely with these latter variables.

4.2.1.4. Combined leadership and group dynamics

The measure was designed to rate the overall combined leaders' adherence to the model, rather than each individual leader's adherence, as it was decided that it was the overall delivery of the program that was important with regards to adherence. However, this produced some difficulties, particularly when there was a considerable difference in quality and style between the two group leaders. For example with one particular group, one leader was very good at encouraging participation, bringing everyone into the discussion and using inclusive language and non-verbal behaviour, where as the second leader conversely used very technical language and a closed posture that had the effect of discouraging group participation. It was therefore very

difficult to know how to code for such discrepancies, as such specific eventualities were not included in the scoring criteria.

Scoring was also complicated by the group dynamics within the parenting groups. Some groups, for example, required very little facilitation to discuss the vignettes or initiated praise and specific questioning themselves during homework feedback. It was therefore difficult to know whether these discussions should be included in the scoring of adherence for these variables or whether leader behaviour alone should be the basis of the rating. Hogue et al. (1996) suggest that the rating of treatment adherence should be geared exclusively towards the therapist's behaviour and should not include the behaviour of the clients. However, whilst this would reduce the complexity of the coding instrument, thereby increasing the reliability of the adherence ratings, it could mean that the leadership of certain parenting groups would be rated low on treatment adherence despite the majority of the program being delivered by the parents themselves. This would have the effect of penalising those group leaders whose group contained competent, psychologically minded parents. It therefore seems important, with some variables, to include within their definitions specific aspects of client behaviour that will be included in the rating. Whilst this is included in some of the variable definitions in this study, there are a number of variables where it was not anticipated that this would be a problem, creating confusion for the raters as to whether they should be including client behaviour or not.

As was discussed in Chapter 1, one of the common concerns regarding the use of treatment manuals is that they do not take into consideration individual client differences (Addis et al, 1999). However, it was observed that parents, very

obviously, brought very different issues into each group, requiring flexibility on behalf of the leader to tailor the session to the groups needs. The Webster–Stratton BASIC program does provide guidance for flexibility within group facilitation, allowing for the discussion of important issues. However allowing for such flexibility within the rating of adherence again complicates the process, bringing in issues of subjectivity and making reliable rating more difficult.

4.2.1.5. Contextual factors

The instrument also allowed the raters to take some aspects of the session context into account, rating tapes slightly differently depending on whether the session was early, middle or late. This is an important aspect of adherence as changes in leader facilitation over the course of the sessions are included in the program manual. However, again this increases the complexity of defining each variable and introduces difference and a degree of subjectivity into the rating process, making it harder to achieve good inter-rater reliability. Hogue et al. (1993) talk about the importance of including contextual factors whilst measuring treatment adherence, suggesting that competency is always linked to context. They state that factors such as client difficulties, client characteristics and stage of therapy should have an effect on the therapist choice of intervention and therefore should be incorporated within a measure of adherence. Despite the importance of these factors, they observe that the majority of researchers fail to take them into account when rating treatment adherence or therapists' competency, questioning the usefulness of these measures. Therefore whilst incorporating contextual factors into the measure of treatment adherence increases the complexity of the measure, this is an important aspect of treatment delivery which needs to be addressed within any adherence instrument.

The inclusion of contextual factors into the measure of adherence, however, increases the amount of knowledge the raters need to know about the parenting program in order for them to reliably take context into account. This will also have the disadvantage of increasing the amount of training raters will need before they will be able to use the instrument. This became apparent within this study as one of the raters had more experience of running the parenting groups than the other and therefore was initially more able to understand the context of each session and take this into account during coding.

4.2.1.6. Common Factors

Throughout the rating process it was observed that a number of the variables were rating skills that the majority of the leaders had a reasonable level of ability in, for example leading a discussion or creating a feeling of safety amongst the group. These skills tended to be those that are common to many forms of psychotherapy, rather than being specific to the BASIC parenting program. It therefore became very difficult to tease out what makes some group leaders better than others in these areas with very subtle differences differentiating between the ratings, again increasing the difficulty in obtaining reliable ratings. This also meant that some variables only received scores at the high end of the spectrum, reducing the range of scores used and making inter-rater reliability more difficult to obtain.

The importance of common factors in determining therapeutic outcome was addressed in Chapter 1, with Lambert and Bergin (1992) have suggested that around 30% of outcome variance is attributable to variables such as empathy, warmth and acceptance. It is therefore important that the use of manualised therapies does not

preclude therapists from developing an effective therapeutic relationship with their clients, as both Castonquay et al. (1996) and Henry et al. (1993) found. Manualised therapies therefore need to include, within their manuals, strategies for developing and maintaining an effective therapeutic relationship alongside the specific therapeutic skills. This will ensure that these important factors in determining therapeutic change are not lost with the use of manuals. The BASIC parenting program does in fact incorporate a number of common factors within its manual aimed at developing an effective relationship with the parents in the groups. As noted in Chapter 1 (Table 1), of the 53 criteria included in the original peer and self evaluation form, 12 are specifically aimed at relational aspects of group facilitation generic to all forms of psychotherapy. Whilst this is an important area for manualised therapies it does increase the complexity for rating treatment adherence. In the present study the raters found the more common, relational aspects of therapy more difficult to rate reliably, particularly within the context of the group. These variables tended to be very complex in nature, being more subjective and more affected by the group dynamics and the context of the group. Therefore whilst it is important to maintain a measure of adherence to these more relational aspects of group facilitation, future work could attempt to differentiate this aspect of program delivery onto a third separate subscale, dealing specially with the treatment alliance, rather than attempting to incorporate this into the measure as a whole.

4.2.2. Future developments

4.2.2.1. Improving inter-rater reliability

The role of adherence within parent training programs is an important area of research. This study has highlighted some of the complex issues involved with rating

adherence using an observational method of whole videotaped sessions. Whilst the overall inter-rater reliability for the instrument was disappointing, a number of variables were coded reliably. These tended to be those that were more concrete in nature and therefore more easily defined on the 1-5 scale and occurred frequently throughout the course of the session. This reduced the subjectivity involved in the rating process and the chance that important examples were missed during coding.

Future work in this area will benefit from addressing some of the difficulties that have been discussed above. Firstly it would be important to clarify some of the definitions of variables in order to reduce the subjective nature of some of these variables. This could be achieved by providing more examples of what constitutes a specific score, providing less room for rater intuition. The variables that would particularly benefit from this are: "Safety", "Ground rules", "Discussion", "Generalisation", "Predicts relapse", "Focus on positive", "Validates feelings", "Collaborative" and "Reframes". Variable definitions could also be improved for those variables that are currently measuring outcome rather than adherence, for example "Safety" needs to include aspects of leadership behaviour that promote a feeling of safety within the group.

Another area of rater difficulty was related to the complex nature of some of the variables, which included both qualitative and quantitative features, as well as relational aspects of adherence. Future work could attempt to be more precise about each of these aspects in the definition of each variable, again reducing subjectivity. However, another idea, would be to separate the different aspects of a variable into different subscales, in order to reduce the complexity of the rating system. For example, each variable could be independently rated as to the frequency of its

occurrence, the competency with which it was carried out by the leader and the degree to which it facilitated a good therapeutic alliance.

The rating of treatment adherence may also benefit from reducing the number of variables that are included in the instrument, particularly those that appear to be measuring similar constructs and those that provide little differentiation between leaders. For example, the variables "Safety" and "Comfortable" appeared to be tapping similar concepts and therefore could be combined to form one variable. The variable "Open ended" on the other hand appeared to show very little differentiation between groups as the majority of leaders made effective use of open-ended questions. This variable could therefore be dropped or added to the "Leads discussion" variable, further reducing the size of the instrument. Finally variables that do not occur very frequently could be excluded from the instrument as they are easily missed during the rating process and are therefore difficult to rate reliably. This was observed specifically with the variables "Confronts" and " Generalisation".

Another area that could be improved upon in future research is the amount of training raters are given in order for them to use the instrument reliably. It is important that raters have a competent grasp of the BASIC parenting program and the therapeutic methods involved. It is also important that raters are given sufficient time to practice using the instrument with videotaped sessions so that they know what specific areas of the program relate to each individual variable, making the process of coding easier. It would also be beneficial for the researchers to meet regularly throughout the period of rating, to discuss examples from the videotaped sessions, helping to maintain inter-rater reliability and prevent "observer drift" (Barker et al., 2002). Whilst such

meetings did occur frequently at the beginning of this study, time constraints limited their inclusion towards the latter stages of the rating process, increasing the likelihood that the ratings drifted apart.

4.2.2.2. Assessment of the measure's validity

Finally, future work in this area also needs to take into account the issue of validity, to see if the instrument is actually measuring treatment adherence. There are a number of ways of capturing issues of validity (Barker et al., 2002; Cone, 1999). Firstly, content validity addresses whether the items adequately sample the different aspects of adherence to the program specified in the program manual. Related to this, face validity addresses whether the items look like they are measuring what they are intending to measure – in this case, adherence to the program. The use of Webster-Stratton's own peer and self evaluation form as the starting point to developing our instrument suggests that it will include ratings of the aspects of adherence that are fundamental to the BASIC parenting program. Future work could attempt to assess this further by getting accredited group leaders to appraise the instrument on how well they think it adequately rates aspects of treatment adherence. If we refer to Dane and Schneider's (1998) work in this area a measure of adherence should be sampling aspects of program exposure, quality of treatment, participant responsiveness as well as whether the specific treatments were administered. Our measure currently includes three of these aspects within its ratings: quality of treatment, participant responsiveness and whether specific treatments were administered, allowing us to assume some degree of content validity.

Thirdly, criterion validity asks how well the measure correlates with other relevant criteria or indicators. This could be examined in the future, for example, by seeing whether the ratings of treatment adherence attained using our instrument, correlate with the leaders' self-ratings of adherence, with parents' ratings of what occurred in the sessions, or with the feedback leaders have received from Webster-Stratton's own accreditation system. Criterion validity also brings in the issue of sensitivity, whether the instrument is sensitive enough to pick up leaders who do not adhere to the program. Whilst this is a very important aspect of validity the restricted range in adherence scores observed in this study limit the extent to which this can be assessed.

Finally, construct validity addresses a more complex issue, examining the validity of the construct being measured rather than the individual methods of measuring that construct. Is treatment adherence, therefore, a valid construct to be measuring? Future research could attempt to assess construct validity by ascertaining whether the pattern of relationships between measures of adherence and measures of other constructs are consistent with theoretical expectations. For example, one might expect to find moderate correlations between treatment adherence and the therapeutic alliance given that adherence includes aspects of the therapeutic alliance. However, if strong correlations were found this might indicate that they were measuring the same construct.

4.3. Relationship between treatment adherence and outcome

Overall, treatment adherence was not found to play a major role in predicting clinical outcome. The two adherence factors, "Practicalities" and "Group facilitation", were entered into regression analyses, the results of which were generally non-significant.

The "Group facilitation" factor did appear to account for a small proportion of the reduction in conduct difficulties, but this finding was eradicated when the effect of age of child and hyperactivity were taken into account. There are a number of possible explanations that may account for the lack of significant findings, which are discussed below.

4.3.1. Design of the study

The initial study by Scott et al. (2001) was designed to see whether parenting groups can be effective at reducing conduct problems in everyday clinical practice. It was therefore not designed specifically to look at the effects of treatment adherence on outcome. There are, therefore, a number of design factors that reduce the likelihood of finding a significant association between adherence and outcome in this study. For example, the study built in regular checks of treatment adherence by providing weekly supervision meetings for the therapists, videotaping all intervention sessions, and ensuring leaders kept detailed weekly checklists of group process, thus increasing the likelihood that the groups were delivered as they were intended in the manual. Whilst this is important in controlled studies as it reduces variability within the groups, increases internal validity and maximises the likelihood of detecting treatment effects, it will have reduced the amount these groups resemble everyday clinical practice and reduced the variability in adherence we may have expected to find.

Whilst researchers in the Scott et al. study speculated that variability in treatment outcome may have been due to differences in adherence, our results show that the general level of treatment adherence across the 15 groups was relatively high, with the majority of variables across groups scoring consistently above a three on the five-

point scale. Only two of the 15 groups appeared to be scoring slightly lower on the measure of adherence. We can therefore hypothesise that this lack of variability in our measure of treatment adherence may have reduced the likelihood of finding any significant associations between adherence and treatment outcome.

There is also an issue over the power of the study. Whilst the number of participants in the study was sufficiently large to detect an association between adherence and outcome (n = 84) the data set is complicated by the fact that there were only 15 groups that received a measure of adherence, reducing the power of the analysis. The lack of significant findings therefore, could be due to a type II error.

4.3.2. The role of treatment adherence

Given the limitations of the study, discussed above, it would be premature to rule out the possibility that adherence to the manual plays a role in treatment outcome. Despite the lack of significant results the study did suggest that certain aspects of the parenting program are more important with regards to outcome than others. The rating of a large number of parenting sessions also provided the opportunity to observe the role certain aspects of the program appeared to have in challenging and changing parent behaviours. This allows for speculation about the importance of adherence, particularly regarding specific aspects of the BASIC parenting program.

The regression analyses showed that the practical aspects of the program, such as arranging the room appropriately, sitting amongst the group and issuing homework did not have any association with outcome. The group facilitation factor, however, which included "Encourage", "Reinforce", "Summarise", "Vignettes" "Review

homework" and "Role play", did appear to be having more of an influence on treatment outcome. These variables include the more active ingredients of the program and would therefore be expected to be associated with treatment outcome. This may give some indication to group leaders as to the areas of the program that they should be focusing on in order to improve outcome. It appears that it may not be so important to spend time arranging the room and worrying about the practical details, but rather focussing on those factors that facilitate group discussion, reinforce parental self-learning, increase parental knowledge through the use of vignettes and provide parents with opportunities to practice their skills through the effective use of role-play.

During the coding of the videotapes it was apparent that some group leaders were reluctant to incorporate role play into their sessions: they appeared embarrassed at suggesting their use and would often be discouraged from following through with them by the parents in the group. They also made no attempts to encourage parents in this task, reduce the parents' anxiety about participation, or highlight the important learning potential that arises from their use. In contrast, some leaders were much more willing and able to use role plays within their sessions. Although in early groups this may have caused some embarrassment, over time parents seemed to become used to this aspect of the group, and in fact appeared to enjoy their inclusion and find them a helpful tool to initiate discussion, solve specific problems and practice skills that they would be using over the week. It is my hypothesis that this may be one of the areas of the program that would be less likely to be included once it is more widely disseminated within everyday practice when there are fewer built in checks to monitor treatment adherence. It is also my hypothesis that this is one of the

areas of the program that is important in producing changes in parental behaviour. It provides parents with an opportunity to empathise with the experience of their child, practice implementing new skills and increases the parents feeling of self-efficacy so that they will feel more able to implement behavioural change outside of the parenting groups.

The use of manualised treatments reduces the role for clinical judgement in deciding what interventions should be prescribed for each individual client. Whilst many clinicians are critical of this, believing that it undermines the role of individual assessment and case formulation (Persons, 1991), it does reduce the possibility that the clinician will incorporate their own clinical and cognitive biases into the decision making process. This seems particularly relevant with the use of role plays, where the clinicians' own reluctance and embarrassment to use role plays within their groups may increase the likelihood that they would not be included in a treatment package that was not closely guided by the use of a manual. Therefore, in this case one can see the advantage of using a more actuarial approach to treatment planning that is less likely to incorporate errors of human inference and judgement.

It was also observed, during the rating process, that there was a noticeable difference in the use of vignettes amongst the different leaders. Whilst in nearly all the sessions some vignettes were shown (therefore generally scored a minimum of three on the rating scale), some leaders failed to use the vignettes as an effective trigger for discussion and a useful learning opportunity. Often in these examples the discussion would be very general and fail to draw out the main teaching areas for that vignette. Again this highlights the importance of including an aspect of the competency of

clinicians in their administration of the specific techniques and not just a measure of whether the techniques are included per se. However, it also calls into question the claim by Davidson and Lazarus (1995) that the use of manuals "obfuscates clinical artistry", as there appeared to be an obvious difference in the leaders' ability to make use of the techniques included in the manual so that they provided a useful learning opportunity for the parents in the group. Incorporating the use of vignettes, role plays and homework assignments, for example, so that they are used to their utmost effect, appeared to require considerable skill on behalf of the leader. This questions the assumption that manualised therapies are by definition uncreative, constraining, boring and unfulfilling (Addis et al., 1999).

There were also a number of adherence criteria, included in the Webster-Stratton's BASIC parenting program that could not be rated form videotape and were therefore not included in the measure of treatment adherence. For example, the manual suggests that group leaders should make weekly phone calls to each parent, monitoring their progress, help troubleshoot or provide encouragement and should act as advocates for the parents outside of the group. The manual also advises service providers to offer the parents help with transport to and from the group and provide crèche facilities whilst the group is running. Whilst these are all criteria that are impossible to rate using videotapes of sessions, adherence to their implementation may be important with regard to treatment outcome. For example, lack of such practical assistance could affect parent's ability to attend sessions regularly, reducing the opportunities they have to learn from the group and feel supported in this learning process. Again it can be hypothesised that these aspects of the program may be less likely to be provided outside of the research setting, where group leaders have larger

caseloads, more constraints on their time and less financial resources. Future research may therefore need to include a measure of adherence to these aspects of the program.

4.3.3 Implications for future research

This study has clearly highlighted some important areas for future research. Firstly the issue of statistical power could be addressed by including more parenting groups from different research trials, therefore increasing the number of groups included in the analysis and reducing the chance of making a type II error.

Secondly, this research suggests that it would be beneficial to conduct a study looking at the role of treatment adherence outside of the research setting where there is less provision for regular checks of treatment adherence. Within such a setting it can be hypothesised that there will be more variability in adherence to the program manual as leaders start to incorporate their own clinical and cognitive biases into the decision making process, missing out areas of the program they feel less comfortable including. This potentially increased variability in adherence may therefore provide a better opportunity to look at the relationship between adherence and treatment outcome.

Conducting research within ordinary clinical settings will also allow the opportunity to ascertain whether certain areas of the program are more likely to be left out once the checks to adherence are no longer in place. As mentioned earlier one of these areas may be the effective use of role- play. It would be interesting to see if, without regular supervision, this is an area that group leaders fail to include and what the effects of this might be on treatment outcome.

This leads on to another area for future research: examining the differential importance of specific aspects of the parenting program. This study suggested that the group facilitation aspects of what the leaders did were more influential in their effect on treatment outcome than the practical aspects. There may also be specific components of how the leaders facilitate group discussion and maximise learning that have a greater influence on treatment outcome than others. I have already speculated about the importance of the use of role play within the groups but there may be other factors that are important in promoting therapeutic change and are essential to effective program delivery. Future research could aim to investigate this more fully by systematically leaving out specific aspects of the program and observing the effect this has on outcome.

Currently, the leadership materials supplied with the BASIC parenting program include a peer and self-evaluation form to help leaders maintain adherence to the program through self monitoring. This form comprises 53 individual criteria that leaders are expected to rate themselves on, after every group. Due to the length of the form, this takes a considerable amount of time and it is unlikely that it will be completed within everyday clinical practice due to time constraints; again this will reduce the likelihood that treatment adherence is monitored and hence maintained. The suggestion that some of these criteria may be more important in relation to treatment outcome than others, as hinted at in this study, suggests that future work could attempt to reduce the size of this evaluation form, giving more emphasis to those variables that affect therapeutic change. This may increase the likelihood that

group leaders would complete the form, increasing the monitoring and maintenance of important areas of the program.

Given the concerns raised over the use of manualised therapy, highlighted by Addis et al. (1999), future research might also examine whether the manualisation of the BASIC parenting program does in fact have any negative consequences for leaders. Does it, for example, inhibit the development of an effective therapeutic relationship, prevent leaders from meeting the individual needs of clients or reduce the leaders' job satisfaction by curbing their clinical creativity? Both parent and group leader questionnaires could be used to assess these areas of concern.

4.3.4. Clinical Implications

The research undertaken in this thesis aimed to design an instrument that could measure treatment adherence to the Webster-Stratton BASIC parenting program and, secondly, to use this instrument to explore the role of treatment adherence on treatment outcome.

This research has relevance to the literature on interventions for child conduct problems, as parent training is an area of health care that is currently receiving much government interest and financial support. This has enabled parent training programs, such as the Webster-Stratton BASIC program, to be more widely disseminated around the UK, providing more parents with the opportunity to learn more effective parenting behaviour. As these programs, however, move further away from the constraints of the research setting there is more opportunity for group leaders to depart form the

model. It is currently unclear how this will affect the potency of the program in promoting therapeutic change.

This study tentatively found that treatment adherence might play a role in affecting treatment outcome. If this result is replicated in future research, one needs to consider ways of ensuring adherence to the manual once these programs are disseminated into clinical settings. One of the most effective methods used to increase adherence is the provision of supervision. This provides group leaders with support in administering the program and an opportunity to think about how program delivery could be improved. Whilst the provision of good supervision is fundamental to the delivery of effective parenting groups, its occurrence outside of research settings is less frequent and there are a number of reasons for this. Firstly, clinicians in their everyday practice settings typically have large caseloads and therefore have less time to spend in supervision. They tend to work with a heterogeneous client group, administering multiple treatment programs and thus reducing the amount of time they can spend thinking about, and receiving supervision for, each individual treatment. Within the clinical setting there are also fewer people available who have the expertise required to provide adequate supervision and clinicians may find themselves being the sole provider of parent training in their trust, making peer supervision impossible. Finally, the provision of supervision is costly, and this cost is often not included in the financial budget provided for administering the parenting group.

The importance of supervision in maintaining the positive treatment outcomes observed in research settings has been highlighted in a small number of studies. Both Henggeler et al. (1995) and Scott et al. (2002) found positive results in clinical

practice, similar to those found in research settings. These studies appear to maintain the level of supervision provided in efficacy research, enabling clinicians to attend closely to treatment process and monitor adherence to the specific therapeutic models that are being delivered. These findings therefore suggest that researchers should include supervisory aspects of program delivery in the development of their models of treatment, providing clinicians with methods to maintain sufficient levels of supervision within the more difficult conditions of the clinic.

Poor treatment adherence, however, is not the only possible reason for the poorer outcomes observed in clinical settings. The client group in these settings often have a more heterogeneous presentation making group facilitation more complicated. Their presentation is also usually more complex, including issues around physical and mental health, drug and alcohol dependency, relationship difficulties, housing problems, and financial disadvantages. These can all have a major impact on treatment outcome, by reducing parents motivation to change, or practically preventing them from attending the groups. These issues, therefore, must also be addressed during the development of specific treatments, designing and providing treatments with high ecological validity.

4.4. Conclusions

With the increasing prevalence of childhood antisocial behaviour (Rutter et al., 1998) and the known deleterious consequences it has to individuals, families and society (Farrington, 1995; Kazdin, 1995; Silberg et al., 1996; Webster–Stratton & Spitzer, 1994), the need to find efficacious and cost effective treatments for this client group is imperative. The results from laboratory based studies of parent training interventions

have been encouraging, showing an increase in parental skill, more positive parental attitudes and a significant reduction in child conduct problems (Kazdin & Kendall, 1998). These results have been found to remain at long term follow up (Long et al., 1994). They have also been shown to have indirect beneficial effects, reducing family conflicts, decreasing parental depression and reducing the chance of similar behaviours being displayed in siblings (Dishion & Andrews, 1995; Kazdin 1985; Webster-Stratton & Hammond, 1995).

Whilst these results are encouraging and have led to the recommendation that parent training programs are one of the treatments of choice for children with conduct problems (Roth & Fonagy, 1998), the results have yet to be replicated under natural clinic conditions, questioning the extent to which they can be transferred from the laboratory into the clinic. Future research in this area must therefore attempt to bridge this lab-clinic divide, enabling clinicians to provide treatments that are as effective as they are efficacious.

Researchers therefore must start to design treatments that take into account the conditions of the clinic: the heterogeneous nature of the client group, the demands made on clinicians due to large case loads and the financial constraints of the service providers. This would increase the likelihood that clinicians would be able and willing to adhere to manualised treatments and hence replicate the positive findings that are observed under highly controlled conditions. Clinicians, however, also have a role in this process, working along side researchers to help design treatments that can be reliable implemented in the clinic so that treatments can be more widely disseminated.

References

Addis, M. E., Wade, W. A. & Hatgis, C. (1999). Barriers to dissemination of evidence-based practices: Addressing practitioners concerns about manual-based psychotherapy. *Clinical Psychology: Science and Practice*, 6, 430-441.

American Psychology Association, (1982). Report of the task force on the evaluation of education training and service in psychology. Washington DC: American Psychology Association.

Andrade, A., Lambert, W. & Bickman, L. (2000). Dose effect in child psychotherapy: outcomes associated with negligible treatment. *Journal of American Academic Child and Adolescent Psychiatry*, 39, 161-168.

Bachelor, A. (1991). Comparison and relationship to outcome of diverse dimensions of the helping alliance a seen by client and therapist. *Psychotherapy*, 28, 534-549.

Barker, C. Pistrang, N. & Elliott, R. (2002). Research Methods in Clinical and Counselling Psychology. Chichester, John Wiley.

Barlow, J. (1999). Systematic review of the effectiveness of parent-training programmes in improving behaviour problems in children age 3-10 years. Oxford: University of Oxford Health Services Research Unit.

Billingsley, F., White, O. R. & Munson, R. (1980). Procedural reliability: A rational and an example. *Behavioural Assessment*, 2, 229-241.

Boruch, R.F & Gomez, H. (1977). Sensitivity, bias and theory in impact evaluations. *Professional Psychology*, 8, 411-434.

Brown, T. A., Antony, M. M. & Barlow, D. D. (1995). Diagnostic co-morbidity in panic disorder: Effect on treatment outcome and course of comorbid diagnoses following treatment. *Journal of Consulting and Clinical Psychology*, 63, 408-418.

Campbell, S. (1990). *Behaviour problems in preschool children: clinical and developmental issues.* New York, NY: Guildford Press.

Castonguay, L. G., Goldfried, M. R., Wiser, S., Raue, P. J. & Hayes, A. M. (1996). Predicting the effect of cognitive therapy for depression: A study of unique and common factors. *Journal of Consulting and Clinical Psychology,* 64, 497-504.

Chambless, D., Babich, K., Crits-Christoph, P., Frank, E., Gilson, M., Montgomery, R., Rich, R., Steinberg, J. & Weinberger, J. (1993). Task force on promotions and dissemination of psychological procedures. Report adopted by the APA Division 12 Board, 1-17.

Christenson, A. & Jacobson, N. S. (1994). Who (or what) can do psychotherapy: the status and challenge of non-professional therapies. *Psychological Science,* 5, 8-14.

Coie, J. D. (1990). Toward a theory of peer rejection. In S. R. Asher& J. D. Coie (Eds), *Peer rejection in childhood* (pp365-398). Cambridge, England: Cambridge University Press.

Coie, J. D., Watt, N. F., West, S. G., Hawkins, D., Asarnow, J. R., Markman, H. J., Ramey, S. L., Shure, M. B., Long, B. (1993). The science of prevention: A conceptual framework and some directions for a national research program. *American Psychologist,* 48, 1013-1022.

Cone, J., D. (1999). Observational assessment: Measure development and research issues. In Kendall, P., C., Butcher, N., B. & Holmbeck, G., R. (Eds), *Handbook of research methods in clinical psychology* (pp. 183-223). New York: Wiley.

Conner-Smith, K. K. & Weisz, J. R. (2003). Applying treatment outcome research in clinical practice: Techniques for adapting interventions to the real world. *Child and Adolescent Mental Health,* 8, 3-10.

Dane, A. V. & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology review*, 18, 23-45.

Dishion, T.J. & Andrews, D. W. (1995). Preventing escalation in problems behaviour with high-risk young adolescents: Immediate and 1-year outcomes. *Journal of Consulting and Clinical Psychology*, 63, 538-548.

Durlack, J. A. (1997). *Successful prevention programs for children and adolescents*. New York: Plenum.

Einhorn, (1984), Accepting error to make less error. *Journal of Personality Assessment*, 50, 387-395.

Farrington, D. P. (1995). The development of offending and antisocial behaviour from childhood: key findings from the Cambridge study in delinquent development. *Journal of Child Psychology and Psychiatry*, 36, 929-964.

Francis, A., First, M. B. & Pincus, H. A. (1995). *DSM-IV Guidebook*, American Psychiatric Press, Inc.

Garfield, S. L. (1996) Some problems associated with 'validated' forms of psychotherapy. *Clinical Psychology: Science and Practice*, 3, 218-441.

Henggeler, S. W., Schoenwald, S. W. & Pickrel, S. G. (1995). Multisystemic therapy: Bridging the gap between university-and community-based treatment. *Journal of Consulting and Clinical Psychology*, 63, 709-717.

Henggeler, S. W., Melton, G. B., Brondino, M. J., Scherer, D. G. & Hanley, J. H. (1997). Multisystemic therapy with violent juvenile offenders and their families: The role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology*, 65, 821-833.

Henggeler, S. W., Pickrel, S. G. & Brondino, M. J. (1999). Multisystemic treatment of substance abusing and dependant delinquents: Outcomes, treatment fidelity, and transportability. *Mental Health Services Research*, 1, 171-184.

Henry, W. P., Strupp, H. H., Butler, S. F., Schacht, T. E. & Binder, J. L. (1993) Effects of training in time limited dynamic psychotherapy: Changes in therapist behaviour. *Journal of Consulting and Clinical Psychology*, 61, 434-440.

Hickling, E. J. & Blanchard, E. B. (1997). The private practice psychologist and manual-based treatments: A case study in the treatment of post-traumatic stress disorder secondary to motor vehicle accidents. *Behaviour Research and Therapy*, 35, 191-203.

Hogue, A., Liddle, H. A. & Rowe, C. (1996). Treatment adherence process research in family therapy: A rationale and some practical guidelines. *Psychotherapy*, 33, 332-345.

Horvath, A. O. & Greenberg, L. S. (1994*). The working alliance: theory research and practice.* New York, John Wiley.

Kam, C., Greenberg, M. T. & Walls, C. T. (2003). Examining the role of implementation quality in school-based prevention using paths curriculum. *Prevention Science*, 4, 55-63.

Kazdin, A. E (1985). *Treatment of antisocial behaviour in children and adolescence.* Homewood, IL: Dorsey press.

Kazdin, A. E. (1995). *Conduct disorders in childhood and adolescence.* Thousand Oaks, CA: Sage Publications.

Kazdin, A. E. & Kendall, P. C. (1998). Current progress and future plans ofr developing effective treatments: Comments and perspectives. *Journal of Clinical Child Psychology*, 27, 217-226.

Krupnick, J. L., Sotsky, S. M., Simmens, S., Moyer, J., Elkin, I., Watkins, J. & Pilkonis, P. A. (1996). The role of the therapeutic alliance in psychotherapy and pharmacotherpy outcome: Findings in the national institute of mental health treatment of depression collaborative research program. *Journal of Consulting and Clinical Psychology*, 64, 532-539.

Lambert, M. J. & Bergin, A. E. (1994). The effectiveness of psychotherapy. In S. L. Garfield & A. E. Bergin (Eds), *Handbook of psychotherapy and behaviour change* (4th ed., pp 143-189). New York, Wiley.

Loeber, R. (1985). Patterns and developments of antisocial behaviour. *In Annals of child development*, vol. 2, edited by G, Whitehurst. New York, NY: JAI Press (pp. 77-116).

Long, P. Forehand, R. Wierson, M. & Morgan A. (1994). Moving into adulthood: Does parent training with young noncompliant children have long-term effects? *Behaviour Research and Therapy, 32, 101-107.*

Luborsky, L., Crits-Christoph, P., Mclellan, A. T., Woody, G., Piper, W., Liberman, B., Imber, S. & Pilkonis, P. (1986). Do therapists vary much in their success? Findings from four outcome studies. *American Journal of Orthopsychiatry*, 56, 501-512.

Meltzer, H., Gatward, R., Goodman, R. & Ford. T. (1990). The mental health of children and adolescence in Great Britain. London: office of national statistics.

Mihalic, S. (2002), The importance if implementation fidelity. In Mihalic, S., Fagon, A., Irwin, K., Ballard, D. and Elliott, D. (2002) *Blueprints for violence prevention replications: factors for implementation success.*

Miller, G. E. & Prinz, R. J. (1990). The enhancement of social learning family interventions for childhood conduct disorders. *Psychological Bulletin*, 108, 291-307.

Moffitt, T. (1993) Adolescence-limited and life-course-persistent antisocial behaviour: a developmental taxonomy. *Psychological Review*, 100, 674-710.

Moncher, F. J. & Prinz, R. J. (1991) treatment fidelity in outcome studies. *Clinical Psychology Review*, 11, 247-266.

Nathan, P. E., Stuart, S. P. & Dolan, S. L. (2000). Research on psychotherapy efficacy and effectiveness: Between Scylla and Charybdis? *Psychological Bulletin*, 6, 964-981.

Norquist, G., Lebowitz, B., & Hyman, S. (1999). Expanding the frontier of treatment research. *Prevention and Treatment*, 2, 1-8.

Patterson, G. R. (1982). *Coercive family process*. Eugene, OR: Castalia Press.

Patterson, G. R. & Strouthamer-Loeber, M. (1984). The correlation of family management practices and delinquency. *Child development*, 55, 1299-1307.

Patterson, G. R. (1986). Performance models for antisocial boys. *American Psychologist*, 41, 432-444.

Persons, J. B. (1991). Psychotherapy outcome studies do not accurately represent current models : A proposed remedy. *American Psychologist*, 46, 99-106.

Reid, J., Taplin, P. & Loeber, R. (1981) A social interactional approach to the treatment of abusive families. In R. B. Stuart (Eds.), *Violent behaviour: Social learning approaches to prediction management and treatment* (pp. 83-101). New York: Brunner/Mazel.

Roth, A. & Fonagy, P. (1998). *What works for whom: a critical review of psychotherapy research*. Guildford Press.

Rutter, M. (1980). *Changing youth in a changing society*. Cambridge, MA: Harvard University Press.

Rutter, M., Hagell, A. & Giller, H. (1998). *Antisocial behaviour by young people*. Cambridge: Cambridge University Press.

Scott, S., Spender, Q,. Doolan, M, Jacobs, B. & Aspland, H. (2001). Multicentre controlled trial of parenting groups for childhood antisocial behaviour in clinical practice. *British Medical Journal*, 323, 1-7.

Scott, S. & Sylva, K. (2002). The "spokes" project: supporting parents on kids education. Report to the Department of Health.

Silberg, J., Meyer, J., Pickles, A, Simonoff, E., Eaves, L., Hewett J., et al. (1996) Heterogeneity among juvenile antisocial behaviours: findings from the Virginia twin study of adolescent behaviour. In G. Bock (Eds.), *Genetics of criminal and social behaviour*, London, Wiley.

Smith, M. L., Glass, D. & Miller, (1980). *The benefits of psychotherapy*. Baltimore, MD: Johns Hopkins University Press.

Strosahl, 1998, K. H. (1998). The dissemination of manual based psychotherapies in managed care: Promises, problems and prospects. *Clinical Psychology, Science and Practice*, 5, 382-386.

Svartberg, M. D., Seltzer, M. H. & Stiles, T. C. (1998). The effects of common and specific factors in short-term anxiety provoking psychotherapy. *The Journal of Nervous and Mental Disea*se, 186, 691-696.

Taylor, T. K., Schmidt, F., Pepler, D. & Hodgins, C. (1998). A comparison of eclectic treatment with Webster-Stratton's parents and children series in a children's mental health clinic: A randomised control trial. *Behavior Therapy*, 29, 221-240.

Taylor, E., Chadwick, O., Heptinstal, E. & Danckacets M. (1996). Hyperactivity and conduct problems as risk factors for adolescent development. *Journal of American Academic Child and Adolescent Psychiatry*, 35, 1213-1226.

Wade, W.A., Treat, T. A. & Stuart, G. L. (1998) Transporting an empirically supported treatment for panic disorder to a service clinic setting: A benchmarking strategy. *Journal of Consulting and Clinical Psychology*, 66, 231-239.

Waltz, J., Addis, M. E., Koerner, K. & Jacobson, S. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology*, 61, 620-630.

Webster-Stratton, C. (1982). The long term effects of a videotape modelling parent education program; Comparison of immediate and one year follow up results. *Behavior therapy*, 13, 702-714.

Webster-Stratton, C. (1984). Randomised trial of two parent-training programs for families with conduct disordered children. *Journal of Consulting and Clinical Psychology*, 52, 666-678.

Webster-Stratton, C. (1985). Predictors of treatment outcome in parent training for conduct-disordered children. *Behavior Therapy*, 16, 223-243.

Webster-Stratton, C. (1990). Long term follow up of families with young conduct problem children: From preschool to grade school. *Journal of Clinical and Child Psychology*, 19, 144-149.

Webster-Stratton, C. (1990). A potential disruptor of parent perceptions and family interactions. *Journal of Clinical Child Psychology*, 19, 302-312.

Webster-Stratton, C. (1994). Advancing videotape parent training: a comparison study. *Journal of consulting and clinical psychology*, 62, 583-593.

Webster-Stratton, C. (1998). Preventing conduct problems in head start children: Strengthening parenting competencies. *Journal of consulting and clinical psychology*, 66, 715-730.

Webster-Stratton, C. (2000). The incredible years training series. *Juvenile Justice Bulletin*, June 2000, 1-23.

Webster-Stratton, C. & Hancock, L. (1998). Training for parents of young children with conduct problems: Content, methods and therapeutic processes. In C. E. Schaefer & J. M. Briesmeister (Eds.), *Handbook of parent training* (pp. 98-152). NY: John Wiley & Sons.

Webster-Stratton, C. & Hammond, M. (1990). Predictors of treatment outcome in parent training for families with conduct problem children. *Behavior Therapy*, 21, 319-337.

Webster-Stratton, C., Hollinsworth, T. & Kolpacoff, M. (1989). The long term effectiveness of and clinical significance of three cost-effective training programs for families with conduct problem children. *Journal of Consulting and Clinical Psychology*, 57, 550-553.

Webster-Stratton, C. & Spitzer, A. (1996). Parenting a young child with conduct problems: new insights using qualitative methods. In T. H. Ollendick & R. J. Prinz (Eds.), *Advances in clinical psychology*, volume 18, (pp. 1-62).

Weisz, J. R., Weiss, B. & Donenberg, G. R. (1992). The lab verses the clinic: Effects of child and adolescent psychotherapy. *American Psychologist*, 47, 1578-1585.

Weisz, J. R., Weiss, B., 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.

Wilson, G. T. (1996). Manual-based treatments: The clinical application of research

findings. *Behaviour Research and therapy*, 34, 295-314.

Wilson, G. T. (1997) Treatment manuals in clinical practice. *Behaviour Research and*

Therapy, 35, 205-210.

THE MAUDSLEY

Advancing mental health care

Appendix 1: Ethics approval

Maudsley Hospital

ETHICAL COMMITTEE (RESEARCH)

Denmark Hill
London
SE5 8AZ

Telephone: 0171 703 6333
Fax: 0171 919 2171

Tel: (0171 919) 2892

24 October, 1997
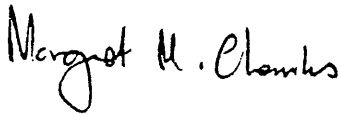
Dr S Scott
Department of Child & Adolescent Psychiatry
Institute of Psychiatry

Dear Dr Scott

Re: Parent management training for childhood conduct disorder (39/93)

At its meeting on 17 October 1997, the Ethical Committee (Research) considered and confirmed Chair's action to approve the protocol amendments requested in your letter of 15 September 1997.

Yours sincerely

Margaret M. Chambers

Margaret M Chambers
Research Ethics Coordinator

THE
BETHLEM ROYAL HOSPITAL AND THE MAUDSLEY HOSPITAL
SPECIAL HEALTH AUTHORITY

ETHICAL COMMITTEE (RESEARCH)

WTD/PDD/DISC.EC7/SCOTT.LTR

25 January 1993

· Please reply to

Bethlem Royal Hospital
Monks Orchard Road
Beckenham
Kent BR3 3BX

Telephone: 081-777 6611
Extension: ~~773~~ 2892
Fax: 081-777 1668

Dr S Scott
Department of Child and Adolescent Psychiatry
Institute of Psychiatry

Dear Dr Scott

PARENT MANAGEMENT TRAINING FOR CHILDHOOD CONDUCT DISORDER (39/93)

The Ethical Committee (Research) has approved, from the ethical standpoint, the research project involving investigations on human subjects as set out above.

If patients of the Joint Hospital are involved, you may find it helpful to obtain consultant approval by means of the attached proforma letter.

In any further correspondence about this project would you please quote the project number given above.

Yours sincerely

Mr W Tautz-Davis
ADMINISTRATIVE OFFICER

Appendix 2: Final coding instrument

<u>Encourage Participation</u>
> o Give every parent a chance to feedback, asking what people thought, brings everyone into the discussion, eye contact with all, inclusive language and non-verbal behaviour, builds rapport with each member of the group. Manage dominant or reticent parents.

1 = No encouragement to participate in discussion or to feedback; non verbal behaviour lacks warmth; no attempt made to build rapport with each member of the group; failure to manage dominant participants.
2 = Very few skills demonstrated and executed poorly.
3 = Some demonstration of these skills, may be inconsistently applied, or executed less rigorously.
4 = Most things demonstrated well.
5 = Everything demonstrated very well.

<u>Practical Arrangements</u>
> o Seating in a semi circle in view of TV, leader and co-leader sat amongst the group.

1 = No preparation for the session; parents' seating arrangement disqualifies parents from entering into discussion and being able to see the television; leaders sat together.
2 = Some attempt at preparation but chairs not ideally positioned <u>and</u> leaders sitting together.
3 = Some attention to seating arrangement, but leaders sitting together <u>or</u> chairs not ideally positioned.
4 = attempts made to arrange seating appropriately, but leaders seated near one another due to small numbers or empty seats.
5 = Perfect seating arrangement, leaders sat amongst the group, television in full view.

<u>Open-ended questions</u>
> o To facilitate the start of the discussion (later groups – session 5 onwards - might discuss spontaneously without need for as many open ended question*). Scored poorly if uses too many closed questions where could have used open.

1 = No use of open-ended questions; inappropriate use of closed questions.
2 = Limited use of open questions, inappropriate closed questions.
3 = Some demonstration of open-ended questions applied appropriately with few unnecessary closed questions.
4 = Good use of open-ended questions allowing open discussion; rare use of inappropriate closed questions*.
5 = Excellent use of open-ended questions to facilitate open and free flowing discussion; no inappropriate closed questions*.

Reinforce parent's ideas and fostering self-learning
    o   Asks parents what they would do, and give credit for their solutions, praise their own ideas, encourage problem solving. Learn from each others experiences and support each other, drawing out ideas from parents.

1 = No attempts made to draw out, acknowledge or praise parents' own ideas. No encouragement for parents to learn from each other.
2 = One attempt to draw out or acknowledge parents own ideas.
3 = Some demonstration of reinforcement of parents' ideas, through praise and acknowledgement. Attempts to draw out or encourage ideas and solutions so that parents can learn from each other.
4 = Good demonstration of reinforcement of parents' ideas through praise and encouragement, however some opportunities are missed.
5 = Excellent reinforcement of parents' ideas, use of praise and encouragement of problem solving and learning from each other.

Feeling of safety amongst group members
    o   In early groups evidence of this being created by the leaders, later groups that parents appear safe, able to express their own ideas, disagree, discuss, laugh. Evidence early on would be saying, its ok to disagree, give own opinions, that its difficult to do etc.

1 = No evidence of safety amongst group or safety being created in earlier sessions. Parents appear uncomfortable when expressing their own ideas. Unwilling to debate or discuss ideas. No humour to put parents at ease.
2 = Very little demonstration of safety, one mother may appear at ease in the group but leaders make no attempt to create a feeling of safety.
3 = Some demonstration of safety amongst group members or that it is being created by the leaders. Some parents appear reasonably comfortable presenting their own ideas but for others the interaction is not spontaneous and leaders fail to pick up on this. There may be reluctance to challenge/offer opposing views. May appear self-conscious about role-plays and need persuading.
4 = Most parents seem at ease and contribute spontaneously to the discussion. Parents offer encouragement and solutions to one another and use humour as appropriate. Members may disclose sensitive information about themselves, or do not appear anxious about participating in role plays.
5 = Excellent demonstration of safety amongst group or being created by leader. All parents appear very relaxed, free to present their own ideas, challenge others and debate issues. Use of humour, willingness to participate in role-plays with no signs of self-consciousness.

Ground rules
    o   Evidence that ground rules have been established. Verbal reference to a rule or signs that the rules are pinned up in the room are scored positively. Obvious breaking of rules are picked up on by the leader. Failure to do so scores negatively. Basic rules – confidentiality, time-keeping, mobile phone/bleeper interruptions, giving everyone a chance to be heard.

1 = No evidence of ground rules having been established or stuck to throughout the session.

2 = Evidence of vague guidance, but not clearly conveyed or understood by the group, shown through inconsistent behaviour by parents and/or leader?

3 = Ground rules established. Rules are generally stuck to, but not revisited at appropriate times.

4 = Evidence of ground rules adherence, instances of broken rule(s) addressed, or the importance of rules are stressed.

5 = Ground rules established. Parents adhere to them or breaking of rules is addressed. New members are informed of the rules. Rules are added to, if necessary, throughout the sessions.

Agenda

- o Evidence that there is an agenda, up on wall, discussed or clearly divided into different areas. Ideally, the outline for the session is mentioned within first 15 minutes of the meeting. Sticking to timings, not left at the end having not done half the material.

1 = No evidence of a written or spoken outline for the session.

2 = Some evidence of an agenda, with session following a clear structure, but this is not made explicit so that parents are unaware of the session outline

3 = Some evidence of either a verbal or written agenda stating session objectives so that parents have an idea of what is being covered in the session. Most items covered during the session.

4 = Agenda specifically referred to **early on** in the meeting, so that parents know the outline for the session, a timekeeper maybe assigned. However due to poor time management one agenda item is not addressed

5 = Agenda posted on the wall and referred to at the beginning of the session so that parents are clear of the objectives for the session. The session is well structured, addressing all the intended items with good time management.

Summarises restate main points

- o Highlight key points from homework examples, paraphrase and highlights points made by parents Main point of session frequently reminded throughout the session.

1 = No summary of key points, paraphrasing or highlighting of important points raised by parents.

2 = Few attempts made

3 = Some attempt to restate main points through homework examples, and/or paraphrasing parents' views. On occasions opportunities to highlight points are missed, or technical terms are used.

4 = Parent-friendly language used to summarise and restate main points. Many opportunities to highlight points and paraphrase comments utilised.

5 = Frequent and effective summaries of main points. Parent's views often highlighted and successfully paraphrased to clarify understanding. Examples given by parents are integrated into programme principles and aims of session.

Leads discussion
- o Allow for some discussion outside the immediate topic but bring it back to specific topic after a short period. Begins topic for the day with open-ended question to get the parents thinking. Focus on key points, sufficient structure prevents sidetracking, knows when to be flexible and digress for an important issue. Maintains leadership.

1 = No control over topic of discussion. Considerable side tracking evident.

2 = Attempts to lead discussion unsuccessful, digression frequent with little focus on key points or, important issues ignored.

3 = Some attempt to lead discussion and focus on key points. Fair use of open-ended questions to trigger discussion but difficulty in controlling digressions or fails to allow for an important issue to be discussed. Leader's questions may be too broad/less focused on the topic

4 = Good use of open ended questions, limited side tracking evident, reasonable time allowed for important digressions, key points reiterated and discussion brought back on course. Most parents are asked specific questions about their week in relation to the homework task.

5 = Excellent use of open-ended questions, allows for discussion of important issues outside the immediate topic, but after a short period brings the discussion back on track, focusing on the main issues. During homework feedback, leader questions are specific to the topic and the discussion is contained. Maintains leadership and moves the discussion along once parents have understood.

Generalisation of concepts
- o Generalisation of concepts to different settings, situations and behaviours. Skills are readdressed regarding new, or predicted problem behaviours. Parents receive guidance for how new skills can be applied to current or future misbehaviours. Statements like "so, perhaps xx would be useful when/if (s)he does x as well". Behaviour goals revisited.

1 = No generalisation of concepts.

2 = Attempt to generalise is unsuccessful.

3 = Generalisation of concepts somewhat evident, very little evidence of anticipating future misbehaviour.

4 = Frequent generalisation of skills to different settings, situations and behaviours. May refer to future predicted behaviours <u>or</u> apply skills to behaviours that parents want to see changed.

5 = Excellent generalisation of concepts throughout the session. Leader generalises skills for use with future behaviours <u>and</u> to different current behaviour the parents want to see changed.

Predicts relapse
- o Leaders predicts future relapses and difficulties, helping parents to understand concept of the tool box. Positive but realistic levels of expectation. Anticipates potential difficulties. Doesn't have to be frequent but used appropriately.

1 = No prediction of relapses or anticipation of future difficulties.

2 = Vague reference to future relapse or difficulties, or predictions exaggerated or understated.

3 = Some evidence of realistic prediction of relapse, but parents not helped to identify the tools/skills they could use to manage these.

4 = Some realistic predictions made and parents are encouraged to think about toolbox. May be inconsistently applied, or leaders immediately offer the skills.

5 = Realistic expectations discussed with relapse and difficulties anticipated. Parents encouraged to think about toolbox analogy and how they could apply the appropriate tool for different behaviours. Parents encouraged to be positive and to persevere, i.e. relapse does not mean failure.

Focus on or highlights the positive

- o Reminds parents of their achievements, what they have managed not what they haven't, noting their strengths.

1 = No attempt to focus on parents achievement always focuses on problems and weaknesses.

2 = rare attempts made to focus on the positives.

3 = Some attempt to highlight parents strengths and achievements.

4 = Takes the majority of opportunity to highlights the parents achievements and strengths.

5 = Takes every opportunity to highlight the parents achievements and strengths providing support and encouragement to parents.

Validates/supports parents feelings

- o Use of reflective statements and supportive comments in response to parents affect, may include some self-disclosure. Parents' get a sense that the leader understands their situation or feelings. Leader is warm and empathic.

1 = No attempt to acknowledge affect or make reflective statements.

2 = At least 1 attempt to acknowledge affect or use a reflective statement but delivered with little warmth.

3 = Some attempt to acknowledge affect and make reflective statements.

4 = Good use of reflective statements and affect acknowledged, helping parents to feel understood. However, some instances to do so may be missed.

5 = Excellent use of reflective statements and affect acknowledged. All key instances addressed. Leaders appear warm and empathetic and parents feel supported and understood.

Collaborative vs. expert model

- o Shares own personal experiences when appropriate, work in partnership. Shows that they are not perfect and don't know everything, as apposed to always telling parents how to do it, that they know all the answers. The overall approach should empower the parents. Ask for other members' opinion.

1 = Leader takes an expert approach, telling parents what they should do. Makes no attempt to empower the parents.

2 = Few attempts made to work in partnership with parents but these are poorly executed.
3 = Leader attempts to work in partnership and asks the parents for their opinions. Still may take more of the expert role and doesn't show their imperfections.
4 = Works in partnership, taking less of an expert role or shows their imperfections.
5 = Works in partnership, without giving parents all the answers and taking on the expert role. Shares own experiences and difficulties.

<u>Fosters coping vs. mastery model of learning</u>
- o Noting that parents are not going to be perfect but good enough, normalises problems when appropriate.

1 = Emphasises perfect parenting and doesn't attempt to normalise child difficulties.
2 = One attempt to normalise child difficulties or suggest that parents are not going to get it perfectly correct all the time.
3 = Occasional attempts to suggest that parents are not going to get it right all the time or that some child difficulties should be expected.
4 = Several attempts are made to normalise problems and suggest that parents are not going to get it right all the time, but some instances may be missed.
5 = Takes a good enough approach, accepts that no parent is going to be perfect and children will have difficulties at times, normalises these problems.

<u>Reframes experiences from child viewpoint and modify parent's negative attributions</u>
- o May not always be apparent especially in later groups where less negativity. Putting the child's perspective across. Tries to pull out parents positive view of child rather than all bad.

1 = Makes no attempt to modify negative attributions of child and no reframing.
2 = One attempt to reframe experiences from child's point of view or modify negative attributions but many opportunities missed.
3 = attempts to reframe experiences from child's point of view or modifies negative attributions.
4 = Good attempts to put the child's perspective across and draws pout negative attributions, however some opportunities are missed.
5 = Enables parents to see the child's perspective using reframes and draws out parents positive views of their children if they fail to see any.

<u>Confront, challenge or teach parents when necessary</u>
- o When necessary, quality of how achieved, if missed an opportunity scores poorly. Occurrences may be rare.

1 = If missed an obvious opportunity to challenge or confront a parent on what could be a dangerous issue.
2 = If missed a less obvious opportunity or made a poor attempt to challenge, confront or teach.
3 = Attempted to challenge a parent on an issue but without much success
4 = good attempt made to challenge, confront or teach a parent, but may miss another opportunity to do this.
5 = Challenged or confronted a parent on an important issues with care, consideration and skill.

Knowledge of content covered in session

      o   Explains the rationale for principles in clear convincing manner. Knows what they are talking about and what the main points of the session are.

1 = Shows little understanding of the principles or main teaching points of the session. Unable to draw from other, earlier or later areas of the program.
2 = Understands one area of the session.
3 = Shows a reasonable understanding of the program and able to explain the rationale but does not attempt to draw from other topics earlier or later in the program.
4 = Good understanding of the program seen through explanation of the rationale. At least one attempt to link current session to previous or later topic areas.
5 = Demonstrates a very clear explanation of the rationale. Demonstrates an excellent understanding of the program, and draws from other topic areas, keeping the bigger picture in mind.

Integrates parents ideas and problems with content and child development principles

      o   Notes difference at different stages of development, knows what should expect from a child and what they need.

1 = Makes no attempt to integrate child development principles or shows little understanding of developmental stages.
2 = Vague references to developmental stages, but its relevance is unclear.?
3 = Some attempt(s) to integrate developmental principles, but further elaboration needed.
4 = Integrated developmental principles with good elaboration, but inconsistently demonstrated.
5 = Frequently integrates child development principles and demonstrates good understanding of developmental stages.

Use of analogies and metaphors to explain theories

      o   Uses parent friendly terms, that increase the parents understanding of theories and ideas. Good use of analogies and metaphors.

1 = Difficulty explaining concepts in a way that increases understanding, no use of metaphors.
2 = One attempt made to explain a concept.
3 = Attempts to explain concepts in a way that parents will understand.
4 = Concepts explained well and attempts at using effective analogies/metaphors.
5 = Excellent use of metaphors or analogies to increases parents understanding. Presents material in a very clear parent friendly manor.

Leadership methods
Material prepared in advance

      o   Homework given back, leaders prepared for current session tasks including flipchart, toys and vignettes. Evidence of lack of preparation is negative.

1 = No evidence of homework given back, or available for this session. Necessary materials not available and vignettes not cued.
2 = One aspect of the session prepared.

3 = Some preparations made for current session homework. Evidence of preparation of most of the main materials needed, or vignettes cued at the right place.

4 = Good preparation made for the current session. All but one area prepared well.

5 = Excellent preparations for the session, homework, all materials and vignettes prepared well.

Vignettes used to trigger discussion

o   Start with open ended question (may not be needed in later groups and parents discuss spontaneously), acknowledge parents responses, paraphrase and highlights main points, pulls out the main issues and "typical questions" for the vignettes and moves onto the next one. Used efficiently and strategically. Quality and quantity, if got it move on.

1 = no evidence of vignettes used.

2 = shown one or two vignettes but not used to facilitate learning.

3 = Some evidence of vignettes used to trigger discussion. Only a few vignettes shown with some issues discussed, but not all main issues drawn out, or discussion continues unnecessarily.

4 = Vignettes used effectively, but lacks range and/or depth of discussion.

5 = Sufficient use of vignettes to draw out main issues and typical questions. Good and efficient discussion, acknowledging parents' responses, paraphrasing and highlighting main points before moving on.

Use role-play to reinforce leaning

o   At least one role play or "show me" detailed descriptive praise of the role play, asking parents how it felt, what was good what could be different leader shouldn't take part if possible, and never play the expert.

1 = No use of role plays.

2 = Attempt at a role play but fails to see it through, or no attempt to discuss it.

3 = One attempt at a role play or "show me", but not effectively used to reinforce learning, or, leader takes the expert part, or the role play is not directly related to the session topic.

4 = At least one role play focused on the session topic, but only some parents asked for their comments about affect and improvements.

5 = At least one role play used to facilitate discussion. 'Actors' asked to comment on their role, how it felt and observers included in how it could be improved. Leader coaches if necessary.

Review homework and give feedback and praise

o   Praise any efforts made over the week, highlight key principles that the example illustrates, explore with those parents who didn't complete homework what made it difficult and how they can adapt. Except responsibility for any misunderstandings of homework.

1 = No acknowledgement of parents' effort made over the week.

2 = Attempts to praise parents' effort, but failure to accept responsibility for any homework misunderstanding, or no attempt to problem solve homework difficulties.

3 = Some feedback and praise for each parents effort over the week. Parent experiences used to highlight key principles, or an attempt to problem solve homework difficulties.

4 = Positive comments and praise offered to each parent. Some linking of experiences and program principles, though may be inconsistently demonstrated.

5 = Each parent's week reviewed, feedback and praise offered. Clear links between parent experiences and key principles of the programme are made and alternative strategies for completing homework developed.

Issue homework

o Review the refrigerator notes and homework, clear objective and methods, why its important and how they are going to do it. Comprehensive guidance. Given adequate time and not just tagged on at the end, supported in it and understand it.

1 = No homework issued (unless the observed session is the last one).

2 = very little time allotted to explain the aim of the week's homework, or parents uncertain about how they should conduct the task.

3 = Some guidance for homework completion offered, or, importance emphasised and some attempt to convey the principles to which the task relates, though somewhat hurried.

4 = Clear instructions ensure understanding of homework task, links between programme principles and task objectives discussed and importance emphasised.

5 = Excellent coverage of refrigerator notes and homework task. Homework objective clearly set out and methods to aid completion discussed. Sufficient time allotted to discuss homework task, its value and importance.

Parent Responses
Parents appear comfortable and involved in the session

o No obvious signs of self-consciousness, are active participants.

1 = High levels of self-consciousness / anxiety amongst the group. Frequently looking at the camera or constant shuffling or fidgeting.

2 = Distinct signs of self-consciousness observable in some members. Attempts to participate seem strained, very little eye contact with each other.

3 = Most members seem comfortable, attempting to involve themselves in the session, or respond appropriately to invitation to participate. Rare glances at the camera

4 = Majority of parents actively engage in discussion, seem comfortable with each other and leaders, seem not to notice camera.

5 = Excellent interaction by all parents throughout the session body language appears relaxed, lots of humour or appropriate responses observed, good eye contact, keen to contribute view. Presence of camera does not inhibit nature of responses.

Parents refer to using the skills learnt during the past weeks

o Complete homework and make reference to homework and using the skills they are learning at home. Not that just sitting in session and not making any changes at home. Implemented during the week.

1 = No reference made to using the skills at home by any member of the group.

2 = One parent completes homework or talks about using a skill at home.
3 = Some members complete homework and refer to using the skills they are learning.
4 = Nearly all members complete homework and refer to using the skills at home.
5 = All members complete homework and refer to using the skills they are learning at home. May bring in examples of charts they are using and show them to the group.

Appendix 3: Treatment adherence rating form

Variable	Comments & rating	Variable	Comments & rating
• Encourage participation		• Fosters coping vs. mastery model	
• Practical arrangements		• Reframes & modifies attributions	
• Open-ended Questions		• Confronts/ challenges / teachers	
• Reinforce parents ideas foster self learning		• Knowledge of content	
• Feeling of safety amongst group		• Integrates child dev principles	
• Group rules		• Use of analogies & metaphors	
• Agenda		• Material prepared in advance	
• Summarise/restate main points		• Vignettes used to trigger discussion	
• Leads the discussion		• Use of role play to reinforce learning	
• Generalisation of concepts		• Review homework and praise	
• Predicts relapse		• Issues homework	
• Focus/highlights positive		• Parents appear comfortable and involved	
• Validates/supports parents feelings		• Parents refer to using the skills	
• Collaborative vs. expert model			

Appendix 4. Disruptive Behaviour sections of the PACS

SECTION III: DISRUPTIVE BEHAVIOUR

1. TELLING LIES

Now I would like to ask about some of the things most children do to some extent. For example, would X exaggerate, make up stories that are not true or tell lies?
If the answer is NO: **In the past week for example, did he exaggerate, make up stories or tell lies?**
If the answer is YES: **Could you give me an example?**
Does he admit to lying when you confront him with it?

HIGHEST LEVEL OF SEVERITY (in last month)

No exaggeration, making up of stories or lies	0
Exaggeration or making up stories .	1
Lies in order to get out of trouble (white lies) or lies to make trouble for others (siblings) but does not persist when challenged	2
Serious lies, e.g. lies about where he goes or what he does, never admits to lies or lies to obtain goods or favours or to avoid obligations	3
Not applicable or situation not arisen	8
No information, don't know or unreliable information	9

How many days in a week would he usually exaggerate, make up stories or tell lies ?
If the answer is vague: **Would it be more or less than 3 days a week?**

FREQUENCY OF HIGHEST LEVEL CODED (In the last month.)

Never or less than weekly	0
On 1 or 2 days a week	1
On 3 to 6 days a week	2
Daily	3
Not applicable or situation not arisen	8
No information, don't know or unreliable information	9

FREQUENCY OF LEVEL ONE (In the last month.)
Exaggeration or making up stories:

Never or less than weekly	0
On 1 or 2 days a week	1
On 3 to 6 days a week	2
Daily	3
Not applicable or situation not arisen	8
No information, don't know or unreliable information	9

II. STEALING

Would X take things that don't belong to him ?
If the answer is NO: **In the past week for example, did he take things?**
If the answer is YES: **Has it happened in the past month?**

What kinds of things did he take?

NOTES: Include stealing at school, either from other children or items belonging to the school (e.g. pens, pencils, rubbers etc.), or from relatives' or friends' homes. Do not include taking food or swapping toys

HIGHEST LEVEL OF SEVERITY (in last month)

Did not steal in last month	0
Small items or small amounts of money	1
Valuable family possessions	2
Large sums of money or steals from shops	3
Not applicable or situation not arisen	8
No information, don't know or unreliable information	9

FREQUENCY OF HIGHEST LEVEL CODED (In the last month.)

Never or less than once a month	0
More than once a month but less than weekly	1
Once or twice a week	2
More than twice a week	3
Not applicable or situation not arisen	8
No information, don't know or unreliable information	9

FREQUENCY OF LEVEL ONE (In the last month.)
Small items or amounts of money:

Never or less than weekly	0
On 1 or 2 days a week	1
On 3 to 6 days a week	2
Daily	3
Not applicable or situation not arisen	8
No information, don't know or unreliable information	9

III. TEMPER TANTRUMS
Does X sometimes lose his temper? Does he start shouting or screaming or stamping his feet?
If the answer is NO: **In the past week for example, has he done anything like that?**
If the answer is YES: **Would he also throw things, or kick something like a wall or table?   Would he break things or hit someone?**

HIGHEST LEVEL OF SEVERITY (in last month)
No tantrums in last month   0

Mild, shouting, waving arms, stamping feet   1

Marked, throws things, kicks objects   2

Severe, breaks things, kicks or hits people   3

Not applicable or situation not arisen   8
No information, don't know or unreliable information   9

NOTE: In case of differing severities, rate the most severe behaviour

How many days in a week would he usually do that?
If the answer is vague: **Would it be more or less than 3 days a week?**

FREQUENCY OF HIGHEST LEVEL CODED (In the last month.)
Never or less than weekly   0

On 1 or 2 days a week   1

On 3 to 6 days a week   2

Daily   3

Not applicable or situation not arisen   8
No information, don't know or unreliable information   9

FREQUENCY OF LEVEL ONE (In the last month.)
Shouting, waving arms, stamping feet:
Never or less than weekly   0

On 1 or 2 days a week   1

On 3 to 6 days a week   2

Daily   3

Not applicable or situation not arisen   8
No information, don't know or unreliable information   9

IV. RUDENESS

Is X cheeky, would he answer you back?
If the answer is NO: **In the past week, for example, has s/he been cheeky?**
If the answer is YES: **Would s/he ever speak to you in a disrepectful, rude way?**
And has s/he ever sworn at you or spoken to you in an abusive way?
Does he swear at all ? (Not necessarily at you).

HIGHEST LEVEL OF SEVERITY (in last month)
No rudeness or less than weekly   0

Cheeky, answering back, pulling faces without intensely
negative attitude   1

Rude, more disrespectful than being cheeky   2

Swearing or abusive to one or two parents   3

Not applicable or situation not arisen   8
No information, don't know or unreliable information   9

NOTES: Rate the most severe behaviour. Do not include cheekiness which by the parents' account is playful and not intended to be disrespectful.

How many days in a week would he usually do that?
If the answer is vague: **Would it be more or less than 3 days a week?**

FREQUENCY OF HIGHEST LEVEL CODED (In the last month.)
Never or less than weekly   0

On 1 or 2 days a week   1

On 3 to 6 days a week   2

Daily   3

Not applicable or situation not arisen   8
No information, don't know or unreliable information   9

FREQUENCY OF LEVEL ONE (In the last month.)
Cheeky, answering back, or pulling faces but not particularly rude:
Never or less than weekly   0

On 1 or 2 days a week   1

On 3 to 6 days a week   2

Daily   3

Not applicable or situation not arisen   8
No information, don't know or unreliable information   9

V. DISOBEDIENCE
I'd like to ask about times s/he refuses do what you ask, like putting toys away or laying the table. Does s/he ever resist doing things you ask?
If the answer is NO: **In the past week for example, did s/he disobey you?**
If the answer is YES: **How strongly does s/he resist? Does s/he grumble but eventually do it? Or does s/he sometimes flatly refuse to do things you ask? And does s/he then sometimes get rude or have a tantrum?**

HIGHEST LEVEL OF SEVERITY (in last month)

No refusing	0
Mild resistance, answers back but not rude	1
Marked resistance, refuses to comply, negative or rude replies	2
Severe, refuses, leading to tantrums or aggressive behaviour	3
Not applicable or situation not arisen	8
No information, don't know or unreliable information	9

NOTE: Include any disobedience, even if parents say the child's disobedience is related to only particular activities

How many days a week would he be disobedient?
If the answer is vague: **Would it usually be more or less than 3 days a week?**

FREQUENCY OF HIGHEST LEVEL CODED (In the last month.)

No disobedience or less than weekly	0
On 1 or 2 days a week	1
On 3 to 6 days a week	2
Daily	3
Not applicable or situation not arisen	8
No information, don't know or unreliable information	9

FREQUENCY OF LEVEL ONE (In the last month.)
Mild resistance: Answers back but not rude:

Never or less than weekly	0
On 1 or 2 days a week	1
On 3 to 6 days a week	2
Daily	3
Not applicable or situation not arisen	8
No information, don't know or unreliable information	9

VI. REFUSAL TO GO TO BED

Do you have trouble getting him off to bed at night, does he refuse to go to bed?

If the answer is NO: **In the past week, for example, was he difficult about that?**

If the answer is YES : **Can you describe what usually happens?**

HIGHEST LEVEL OF SEVERITY (in last month)
No difficulties   0

Mild, grumbling or stalling but not intense or prolonged   1

Marked, child refuses or has to be coerced into going   2

Strong, refusal leading to tantrum   3

Not applicable or situation not arisen   8
No information, don't know or unreliable information   9

NOTE: Rate the most severe behaviour

How many days in a week does he usually behave like that?
If the answer is vague: **Would it usually be more or less than 3 days a week?**

FREQUENCY OF HIGHEST LEVEL CODED (In the last month.)
Never or less than weekly   0

On 1 or 2 days a week   1

On 3 to 6 days a week   2

Daily   3

Not applicable or situation not arisen   8
No information, don't know or unreliable information   9

FREQUENCY OF LEVEL ONE (In the last month.)
Mild grumbling or stalling but not intense or prolonged:
Never or less than weekly   0

On 1 or 2 days a week   1

On 3 to 6 days a week   2

Daily   3

Not applicable or situation not arisen   8
No information, don't know or unreliable information   9

VII. DESTRUCTIVENESS

Have there been any times recently that he has deliberately broken, torn or spoiled something, like his toys or things belonging to another child? or things in the home?

If the answer is NO: **In the past week, did he ever do anything like that?**

If the answer is YES: **What did he do?**

HIGHEST LEVEL OF SEVERITY (in last month)

No destructiveness	0
Destroyed own property only	1
Destroyed siblings possession or caused mild damage outside the home	2
Caused serious damage in the home (e.g. fire setting) or outside the home	3
Not applicable or situation not arisen	8
No information, don't know or unreliable information	9

NOTE: Rate the most serious behaviour

How many days a week would he usually do that?
If the answer is vague: **Would it be more or less than 3 days a week?**

FREQUENCY OF HIGHEST LEVEL CODED (in the last month)

Never or less than weekly	0
On 1 or 2 days a week	1
On 3 to 6 days a week	2
Daily	3
Not applicable or situation not arisen	8
No information, don't know or unreliable information	9

FREQUENCY OF LEVEL ONE (In the last month.)
Destroyed own property only:

Never or less than weekly	0
On 1 or 2 days a week	1
On 3 to 6 days a week	2
Daily	3
Not applicable or situation not arisen	8
No information, don't know or unreliable information	9

VIII. AGGRESSIVENESS

Does X sometimes get aggressive to other people ? For example, would s/he hit his brothers or sisters (or friends) if provoked?

If the answer is YES: **How aggressive does s/he get? Does s/he hurt anyone? Do you have to restrain him/her ?**

HIGHEST LEVEL OF SEVERITY ((in the last month)

No aggressiveness	0
Mild, threatens only or lashes out when provoked	1
Marked, is physically aggressive, but only transiently and not intensely	2
Severe, attacks people, hurts them, has to be restrained	3
Not applicable or situation not arisen	8
No information, don't know or unreliable information	9

NOTE: Rate the most severe behaviour

How many days in a week would he usually do that?
If the answer is vague: Would it be less or more than 3 days a week?

FREQUENCY OF HIGHEST LEVEL CODED (in the last month)

Never or less than weekly	0
On 1 or 2 days	1
On 3 to 6 days	2
Daily	3
Not applicable or situation not arisen	8
No information, don't know or unreliable information	9

FREQUENCY OF LEVEL ONE (In the last month.)
Mild, threatens only or lashes out when provoked

Never or less than weekly	0
On 1 or 2 days a week	1
On 3 to 6 days a week	2
Daily	3
Not applicable or situation not arisen	8
No information, don't know or unreliable information	9

PARENTAL PERCEPTIONS OF OPPOSITIONAL BEHAVIOUR

So the problems that you have highlighted in this section are........./ there don't seem to be any particular problems in this section, is that right?
(Choose most severe problem)

When did it begin ?
When did you first notice that it was a problem ?

RATE AGE OF ONSET IN MONTHS: _____ (2 DIGITS)

How much do you think that it is within his control?

Do you think he could do more to stop it happening?

How much is it his fault?

PARENTAL PERCEPTION OF LOCUS OF CONTROL
Almost always outside child's control/can't help it	1
Between 1 and 3	2
Partly in child's control, partly beyond; can stop it on some occasions and not on others	3
Between 3 and 5	4
Almost always in child's control; could stop it on most occasions if he wanted to	5
Not applicable or situation not arisen	8
No information, don't know or unreliable information	9

How much of a problem do you consider this to be?

PARENTAL PERCEPTION AND EMOTIONAL REACTION
No problem for them, unconcerned	0
Minor problem, parent slightly worried about child's disruptive behaviour	1
Major problem, parent very concerned about child's disruptive behaviour	2
Severe problem, parent constantly worrying; very upset/ close to breaking point at times	3
Not applicable or situation not arisen	8
No information, don't know or unreliable information	9

MOTHER'S COPING WITH DIFFICULT BEHAVIOUR

How do you react when X behaves like that?

What did you do the last time?

Did it work? How effective were you in dealing with it?

Have you found any other ways of dealing with the behaviour?

Who usually "wins"? When you have differences, who usually makes it up?

RATE MOTHER'S COPING WITH OPPOSITIONAL OR
DIFFICULT BEHAVIOUR ★   ___ (1 DIGIT)

(Responses to this section will also contribute to global parenting ratings at end of interview)

Does you partner agree with the way you handle it?

Do you cope in different ways, or over rule each other?

Do you argue in front of him?

RATE INTER-PARENTAL CONSISTENCY ★   ___ (1 DIGIT)

(★*See the manual for ratings*)

NOTE: For one-parent families, rate inter-parental consistency 8.

PACS: HAND SCORING GUIDES

SECTION III: ANTISOCIAL BEHAVIOUR

1. **LIES**   SEVERITY =   TOTAL / 2 = *MEAN*
                                 FREQUENCY OF LEVEL 1 =

2. **STEALING**   SEVERITY =   TOTAL / 2 = *MEAN*
                                 FREQUENCY OF LEVEL 1 =

3. **TANTRUM**   SEVERITY =   TOTAL / 2 = *MEAN*
                                 FREQUENCY OF LEVEL 1 =

4. **RUDENESS**   SEVERITY =   TOTAL / 2 = *MEAN*
                                 FREQUENCY OF LEVEL 1 =

5. **DISOBEDIENCE**   SEVERITY =   TOTAL / 2 = *MEAN*
                                 FREQUENCY OF LEVEL 1 =

6. **REFUSAL BED**   SEVERITY =   TOTAL / 2 = *MEAN*
                                 FREQUENCY OF LEVEL 1 =

7. **DESTRUCTIVENESS**   SEVERITY =   TOTAL / 2 = *MEAN*
                                 FREQUENCY OF LEVEL 1 =

8. **AGGRESSIVENESS**   SEVERITY =   TOTAL / 2 = *MEAN*
                                 FREQUENCY OF LEVEL 1 =

CONDUCT SCORE: FINAL TOTAL OF *MEANS* DIVIDED BY NO. OF ITEMS RATED   (MAX. 8) =

Norms for 6 - 7 year olds: no disorder 0.8, +1SD (top 16%)1.2, +2SD (top2%) 1.6
Maudsley referred cases with conduct disorder 1.5

117