Treatment fidelity as a predictor of behaviour change in parents attending group-based parent training

C. Eames, D. Daley, J. Hutchings, C. J. Whitaker, K. Jones, J. C. Hughes and T. Bywater

School of Psychology, College of Health and Behavioural Sciences, Bangor University, Gwynedd, UK
Accepted for publication 8 February 2009

Abstract

Background Change in parenting skills, particularly increased positive parenting, has been identified as the key component of successful evidence-based parent training (PT), playing a causal role in subsequent child behaviour change for both prevention and treatment of Conduct Disorder. The amount of change in parenting skills observed after PT varies and may be accounted for by both the content of the programme and by the level of PT implementer process skills. Such variation in implementer skills is an important component in the assessment of treatment fidelity, itself an essential factor in successful intervention outcome.

Aims To establish whether the Leader Observation Tool, a reliable and valid process skills fidelity measure, can predict change in parenting skills after attendance on the Incredible Years PT programme.

Results Positive leader skills categories of the Leader Observation Tool significantly predicted change in both parent-reported and independently observed parenting skills behaviour, which in turn, predicted change in child behaviour outcome.

Conclusions Delivering an intervention with a high level of treatment fidelity not only preserves the behaviour change mechanisms of the intervention, but can also predict parental behaviour change, which itself predicts child behaviour change as a result of treatment.

Introduction

Conduct Disorder in the early years

One in 10 of 5- to 16-year-olds in the UK has a mental health disorder meeting clinical criteria, with Conduct Disorder (CD) the most prevalent of these, accounting for 60% of children with mental health disorders (Office for National Statistics 2004). Disruptive behaviour problems among children aged five and over are the most common reason for mental health service referral in both the UK and the USA (Belsky et al. 2006). Externalizing behaviour patterns such as aggression towards peers, non-compliance and poor impulse regulation can be identified in early childhood and are precursors to clinically significant behaviour problems that become evident during school age and which, without intervention, can remain stable from toddler-hood throughout childhood, adolescence and adulthood (Richman et al. 1982; Campbell 1995; Pierce et al. 1999). Despite formal diagnosis predominantly occurring between middle childhood and middle adolescence, it may also occur in the preschool years [American Psychiatric Association (APA) 2000], sometimes as early as 2 years of age (Shaw et al. 2000; Wakschlag and colleagues 2005).
Depending on the population surveyed, percentage rates of preschool and young school age children who meet criteria for CD range from 7% to 35% (Campbell & Ewing 1990; Webster-Stratton & Lindsay 1999), with higher rates found in single-parent families (Bloomquist & Schnell 2002), low-income families (Attride-Stirling et al. 2000; Kazdin & Wassell 2000) and in families with a parental history of psychopathology (Alpern & Lyons-Ruth 1993; Kazdin 1995), parental substance misuse or criminal activity (Patterson et al. 1989; Frick et al. 1991) and marital problems (Webster-Stratton & Herbert 1994). Without intervention, early aggressive tendencies become established patterns of responding at around 8 years of age (Eron 1990). Campbell (1995) found that 50% of children who exhibited moderate to severe externalizing behavioural difficulties in the preschool years continued to show such behaviour during the school years, meeting full diagnostic criteria by 9 years of age.

In light of the high prevalence of such behaviour problems among young children, the need for early intervention services to both promote child mental health and prevent the development of mental illness has been recognized by both the UK and US Governments, with the Surgeon General in the US urging early intervention, before problems become established and hard to treat (Department of Health and Human Services 2004). The DSM-IV-R identifies parenting behaviours that predispose a child to the development of CD, to include parental rejection and neglect, inconsistent child-rearing practices with harsh discipline and lack of supervision (APA 2000). Because of the significant role parenting plays in the development and maintenance of CD (Patterson 1997; Gardner et al. 1999), parent training (PT) is considered to be the most effective intervention for both prevention and treatment of CD (Kazdin 1997; Brestan and Eyberg 1998; Beauchaine et al. 2005) and has been recommended by the UK National Institute of Health and Clinical Excellence as a clinical treatment for referred children with CD (NICE 2006). PT aims to decrease child antisocial behaviour and increase child prosocial behaviour by strengthening parent management skills, with particular focus on promoting positive relationships between parent and child. Increased positive parenting after PT intervention has been identified as a partial mediator of child outcome, a key component of successful evidence-based PT interventions (F. Gardner, J. Hutchings & T. Bywater, unpublished).

Parent training community replications and treatment fidelity

Parent training programmes were originally delivered and evaluated within clinical and research settings, but the increasing recognition of the need to intervene in the early years, and the efficacy of PT in preventing CD, led to a rise in the use of such programmes in community services (Bond et al. 2000; Gardner et al. 2004). Both clinical and community replications however, do not always achieve the same outcomes as the original research evaluation (Durlak & Wells 1998; Forgatch et al. 2005), and this can be accounted for by lack of treatment fidelity (Mills & Ragan 2000; Mowbray et al. 2003). Treatment fidelity refers to the extent to which intervention delivery adheres to the original intervention protocol (Dane & Schneider 1998; Durlak 1998; Institute of Medicine 2001; Mihalic et al. 2002). Developers of evidence-based programmes who address treatment fidelity do so by a number of strategies, which can include the use of manuals, supervision and training in their programmes (Schnie et al. 1991; Harchik et al. 1992), providing checklist measures of programme content and qualitative forms or Likert-scale measures of process. Replications not only require that the programme content is monitored, but also that the process skills involved in delivering an intervention successfully are utilized. Such fidelity tools are necessary in order to preserve the behaviour change mechanisms for interventions that stipulate process skills as an essential factor in implementation success, especially for interventions delivered in service settings by group leaders from various agencies and backgrounds (Webster-Stratton & Herbert 1994; Arthur & Blitz 2000; Hutchings et al. 2004).

Treatment fidelity of the Incredible Years parenting programme and the Leader Observation Tool

The Incredible Years (IY) BASIC Parenting Programme (Webster-Stratton 1989) has achieved a high scientific standard of programme effectiveness and is identified as a ‘model’ programme for both treatment and prevention of CD, having both content and process fidelity checklists in place (Mihalic et al. 2002). The Leader Observation Tool (LOT; Eames et al. 2008) is a reliable and valid observational measure of process skills fidelity originally developed as an objective quantitative measure for the IY BASIC PT Programme, but has versatility as a both.
clinical and research tool for other PT programmes that utilize the same core components as the IY PT programme. The LOT has 18 standard behaviour categories, forming four skills subgroups: listening, empathy, physical encouragement and positive behaviour, as well as other/negative behaviours. Coding is continuous and records total frequency of leader behaviours within the 2 h parenting group session.

Rationale for study

The current study aims to investigate if LOT skills subgroups predict change in parenting behaviour, for parents with preschool children at risk of developing CD who attended IY PT groups within Sure Start areas in North and Mid Wales and the borders (see Hutchings et al. 2007 for full details of the trial). The hypotheses are twofold. First, it is predicted that greater exposure will produce greater change in observed positive parenting behaviours and change in self-reported parenting style. Second, it is hypothesized that observed change in positive parenting will predict change in observed child outcome scores, and change in parent-reported parenting style will predict change in parent-reported child behaviour problems.

Methods

Participants

Twelve IY BASIC Parenting intervention groups were delivered within 12 Sure Start areas in North and Mid Wales and the borders. Interview and observational home visits were conducted with the recruited families at baseline prior to the intervention starting and 6 months post baseline, once the intervention was complete (follow-up 1) (see Hutchings et al. 2007 for details of the full trial). A Health Visitor seconded to the research team recruited families, with children aged between 36 and 59 months and scoring above the clinical cut-off on the Child Behaviour Check List (CBCL; Achenbach & Edelbrock 1986: r = 0.35 to 0.53; Eyberg & Robinson 1981) and the Beck Depression Inventory (BDI; Beck et al. 1988: r = 0.30 for Overreactivity). Initial data demonstrated the Parenting Scale’s ability to distinguish between clinical and non-clinical groups, although the results were based upon small sample sizes (Arnold et al. 1993). The Parenting Scale demonstrated reasonable internal consistency, with coefficient alpha for the factor and total scores ranging from 0.63 to 0.84. Test–retest reliability ranged from 0.79 to 0.84. Moderate construct validity was obtained from correlations with the Child Behaviour Check List (CBCL; Achenbach & Edelbrock 1986: r = 0.35 to −0.53) and the Beck Depression Inventory (BDI; Beck et al. 1988: r = 0.30 for Overreactivity).

The intervention is a 12-week group-based PT programme. Parents who attended one or more parenting session and had completed both baseline and follow-up 1 assessment were included in the analysis (n = 86), giving a mean number of seven parents per group (range 5–11), with an average attendance of nine sessions (range 1–12). Of the 86 parents, with a mean age of 30 years (range 19–54 years), only one was male and all were the biological parent of the index child. Thirty-five were lone parents, 23 cohabiting, 23 married and five had a non-resident partner. Of the 86 index children, 37 were female, and 49 were male, with a mean age of 46 months. Baseline measures were typically collected up to 2 months prior to the intervention starting with 6-month follow-up conducted within a similar time frame post intervention.

Twenty-two leaders delivered the intervention, with two leaders per group (two leaders co-led two different groups). Twenty-one were female, with leaders’ mean age of 44 years (range 26–59 years).

Measures

Leader measure

The Leader Observation Tool (LOT; Eames et al. 2008) This observational measure was designed to record the leader process skills involved in delivering the IY group parenting intervention (Eames et al. 2008). The LOT has demonstrated high code–recode and inter-rater reliability (mean of 0.92 and 0.95 respectively) and moderate concurrent validity with parent- and leader-reported measures of group process (mean of 0.49 and 0.51 respectively).

Parent and child measures

The Parenting Scale (Arnold et al. 1993) This is a 30-item inventory completed by the parent, designed to measure parenting practices in parents of children aged between 18 and 48 months. The measure comprises three sub-scales: Laxness, Overreactivity and Verbosity, aimed at identifying parental discipline practices, without being tied to the frequency of a child’s misbehaviour. Responses are recorded on a 7-point scale between two alternative responses to a parent/child situation, where a score of 1 represents the highest score of effectiveness, and 7 the highest score of ineffectiveness.

Initial data demonstrated the Parenting Scale’s ability to distinguish between clinical and non-clinical groups, although the results were based upon small sample sizes (Arnold et al. 1993). The Parenting Scale demonstrated reasonable internal consistency, with coefficient alpha for the factor and total scores ranging from 0.63 to 0.84. Test–retest reliability ranged from 0.79 to 0.84. Moderate construct validity was obtained from correlations with the Child Behaviour Check List (CBCL; Achenbach & Edelbrock 1986: r = −0.35 to −0.53) and the Beck Depression Inventory (BDI; Beck et al. 1988: r = 0.30 for Overreactivity).

The Dyadic Parent-Child Interaction Coding System (DPICS; Eyberg & Robinson 1981) The DPICS is an observational...
system designed to aid the assessment of the quality of parent–child interaction. The measure consists of 44 parent and child behaviour categories that are summarized in terms of parent behaviours, child deviance, child responses to commands and parent and child affect. Coding is continuous and records the total frequency of behaviours per specified time interval. The DPICS has demonstrated high inter-rater reliability for parent and child behaviours ($r = 0.67$ to $1.0$ and $r = 0.76$ to $1.0$ respectively; Eyberg and Robinson 1981) and discriminant validity by identifying and classifying $100\%$ of normal families and $85\%$ of treatment families (Eyberg & Robinson 1981). The DPICS was utilized as an observational measure of both parent and child behaviour. For the parent measure of behaviour, observed positive parenting was included in the analysis (a composite of labelled praise, unlabelled praise, positive affect, physical positive and problem solving). In terms of observed child behaviour, child compliance to direct commands and child positives (to include positive affect verbal, positive affect non-verbal and physical warmth) were utilized.

Eyberg Child Behaviour Inventory (ECBI; Eyberg 1980) This is a 36-item inventory, completed by the parent, to assess problem behaviour in children aged 2–16 years. Each item is rated on two scales: a 7-point Intensity Scale and a Yes/No response Problem Scale. The Intensity Scale aims to establish how often the behaviour occurs, with Intensity response ranging from never (1) to always (7). The Problem Scale identifies whether the parent currently perceives particular behaviours to be a problem. Test–retest reliability and internal consistency yielded high correlation coefficients for the measure ($r = 0.86$ and $r = 0.98$ respectively; Robinson et al. 1980). The ability of the ECBI to discriminate between groups of children with and without CD has been well documented (see Eyberg & Ross 1978; Baden & Howe 1992; Webster-Stratton & Hammond 1997).

Design
The IY BASIC PT programme comprises 12 weekly 2-h sessions, divided into four three-session sections: (1) play and relationship building; (2) praise and reward; (3) effective limit setting; (4) handling misbehaviour. All 12 sessions for each intervention group had been videotaped as a requirement of the main trial. One session from each section, sessions 2, 5, 8 and 11 were coded for both leaders in each intervention group by using the LOT, giving a total of 8-h of observation per group, 96-h of leader observations in total. Of these, 30% were selected for second coding. Secondary observations were conducted by trained LOT coders who had achieved an inter-rater agreement of $70\%$ or above during training (see Eames et al. 2008)

Procedure
Leader observations were conducted on selected videotapes for each intervention group by the primary coder and second coded for both inter-rater and code–recode reliability on $30\%$ of observations. Participants were allocated a leader exposure score, based on the number of sessions they had attended. Each leader skills subgroup was totalled per intervention group, and then divided by the total number of sessions delivered ($n = 12$). This number was then multiplied by the number of sessions each participant attended, and thus a level of exposure was calculated for each participant. This allowed variability in the leader scores to complement the variability in the parent and child outcome data. Parent and child outcomes were obtained from the main trial (Hutchings et al. 2007; Bywater et al. 2009). Parent and child data were collected on separate home visits by researchers blind to condition. Interviews were firstly conducted in the home environment to collect the parent-reported data, secondary home visits conducted by an observer blind to interview data. Twenty per cent of home observations were inter-rater reliability visits, achieving $70\%$ agreement or above.

Analysis strategy/data preparation
Change scores were calculated from baseline to follow-up 1 for parenting and child behaviours. First, the relationship between LOT skills subgroups and parenting behaviours was explored, to identify any significant relationships. Once these were identified, stepwise regressions were conducted. Because of the nature of the intervention, whereby the leaders only have contact with the parents, and it is the parents who practise the learned techniques with the children to instigate change, regressions were firstly conducted to establish which leader behaviours could predict change in parenting behaviours. It was then explored whether this predicted change in parenting behaviour was associated with change in child behavioural outcome. Results of a Kolmogorov–Smirnov test indicated that the data leant itself to non-parametric correlations, although graphical analysis confirmed that the assumptions of normality pertinent to regression analysis were held (Field 2005).

Results
Internal reliability was initially calculated on total frequencies of leader behaviours (see Eames et al. 2008). To ensure the LOT retained internal reliability once the exposure score was
calculated, Spearman’s correlation coefficient was conducted. An examination of Table 1 demonstrated that LOT exposure scores retained similar and consistent levels to LOT scores. Specifically, Table 1 indicates the significant correlations between the positive categories of the LOT, reflecting the correlations originally observed with total leader scores.

Mean parent and child baseline and follow-up data are recorded in Table 2. The estimate of clinically significant change was conducted by using Cohen’s (1988) criteria for magnitude of change, where 0.3 SD indicates moderate change, 0.8 SD indicates large change and 1.5 SD indicates very large change. Table 2 presents the clinical improvement in both parent-reported and independently observed measures after attending the intervention. For all the measures analysed, a high percentage of clinically significant change was observed.

The relationship between LOT exposure scores and parent outcome was explored to establish any significant relationships. Given the causal role of positive parenting in changing child behaviour, the relationship between LOT skills subgroups and observed positive parenting was explored and is summarized in Table 3. The relationship between LOT skills subgroups and parent-reported parenting style change scores, as measured by the Parenting Scale was explored and is also summarized in Table 3. One-way Analysis of Variance was applied in both cases, summarizing LOT scores into low, medium and high frequencies. These were calculated as +1/−1 standard deviation from the mean for the high and low groups respectively, with medium frequencies consisting of those scores clustered around the mean for each LOT skills subgroup.

Significant relationships were obtained between observed leader behaviours and change in observed positive parenting, as summarized in Table 3. Moderate effect sizes were calculated for each of the LOT exposure score subscales and change in observed positive parenting. These results demonstrate that the higher the groups’ LOT exposure score, the larger the change in observed positive parenting as measured by independent coders using the DPICS. The relationship between the LOT and parenting is further explored and summarized in Table 3 with LOT exposure score subscales and change in parent-reported parenting style as measured by the Parenting Scale. Leader positive behaviour yields the only significant difference with parent-reported parenting style, obtaining a weak to moderate effect size, although a similar trend was observed with leader empathic skills. Again, a relationship is observed between LOT exposure score subscales and parenting behaviour, although it is weaker with change in parent-reported behaviours than with independently observed parenting behaviours. Having confirmed relationships between the LOT and changes in parenting
outcome, stepwise regression was then conducted to see whether leader behaviours could predict the change in parent outcome scores. Parent outcome change was then entered into stepwise regression to see which parent change scores best predicted change in child outcome scores.

Results indicated that higher levels of positive leader behaviour predicted a greater change in observed positive parenting \((\beta = 0.28, P < 0.01; R^2 = 0.08)\) and the more physical encouragement from the leader predicted greater change in the parent-reported Parenting Scale \((\beta = 0.26, P < 0.01; R^2 = 0.07)\). Change in observed positive parenting predicted a greater change in observed child positive behaviours \((\beta = 0.60, P < 0.01; R^2 = 0.36)\) and to a lesser extent, observed change in child compliance to direct commands \((\beta = 0.23, P < 0.05; R^2 = 0.05)\). Parent-reported change in Parenting Scale predicted greater change in both parent-reported ECBI-Intensity and ECBI-Problem \((\beta = 0.23, P < 0.05; R^2 = 0.05 \text{ and } \beta = 0.34, P < 0.01; R^2 = 0.12 \text{ respectively})\).

### Discussion

Skills subgroups of the LOT exposure score predict change in observed parenting, and, to a lesser extent, some aspects of self-reported parenting. The ability to predict factors that influence change in positive parenting is especially pertinent, considering it has been identified as a key component of successful PT intervention, and a partial mediator of child outcome (F. Gardner, J. Hutchings & T. Bywater, unpublished; K. Jones, D. Daley, J. Hutchings, T. Bywater, C. Eames & C. Whitaker, unpublished). This is further emphasized by the current findings that change in positive parenting significantly predicts change in child behaviour.

Results follow the expected trend of observational measures being significantly related to each other, both leader to parent and parent to child, as were parent-reported measures of parent and child behaviours. The quality of information about child behavioural problems varies depending on the methods used, with parent-reported measures at times both under- or over-reporting the extent of the antisocial behaviour (Loeber et al. 1991; Turner et al. 1998; Handwerk et al. 1999), whereas direct observational methods provide a more precise account of behaviour (Aspland & Gardner 2003). The current results support such findings with observed parenting change significantly predicting observed change in child behaviour, and self-reported parenting change significantly predicting change in parent-reported child behaviour. Because of the nature of the

| Table 2. Parenting and child pre- and post-intervention scores and indices of clinically significant change
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Magnitude of change (%)</td>
</tr>
<tr>
<td></td>
<td>Baseline score (pre-intervention)</td>
<td>Follow-up score (post intervention)</td>
</tr>
<tr>
<td>ECBI-I*</td>
<td>145.92 (25.19)</td>
<td>116.19 (32.09)</td>
</tr>
<tr>
<td>ECBI-P*</td>
<td>16.56 (6.92)</td>
<td>9.55 (7.94)</td>
</tr>
<tr>
<td>The Parenting Scale†</td>
<td>3.47 (0.81)</td>
<td>2.65 (0.81)</td>
</tr>
<tr>
<td>DPICS: positive parenting†</td>
<td>23.05 (19.49)</td>
<td>32.44 (19.27)</td>
</tr>
<tr>
<td>DPICS: child positive†</td>
<td>6.59 (6.68)</td>
<td>6.53 (5.81)</td>
</tr>
<tr>
<td>DPICS: child compliance to direct commands†</td>
<td>8.92 (7.78)</td>
<td>7.99 (6.36)</td>
</tr>
</tbody>
</table>

| Table 3. Change in observed positive parenting and parent-reported parenting style for different levels of Leader Observation Tool exposure
<table>
<thead>
<tr>
<th>Mean leader behaviour scores (SD)</th>
<th>Change in observed positive parenting</th>
<th>Change in parent-reported parenting style</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>P</td>
</tr>
<tr>
<td>Listening</td>
<td>969.45</td>
<td>579.36 (390.09)</td>
</tr>
<tr>
<td>Empathy</td>
<td>72.23</td>
<td>42.83 (29.40)</td>
</tr>
<tr>
<td>Physical encouragement</td>
<td>961.77</td>
<td>622.26 (339.51)</td>
</tr>
<tr>
<td>Positive behaviour</td>
<td>1112.32</td>
<td>773.68 (338.64)</td>
</tr>
<tr>
<td>Other/negative</td>
<td>62.97</td>
<td>37.47 (25.50)</td>
</tr>
</tbody>
</table>

High and low groups were calculated as +1/-1 standard deviation from the mean; medium group consists of scores clustered around the mean.

*P < 0.01; **P < 0.05. d.f. = 2.
intervention, leaders’ interaction was with the parents, and not the children, and therefore the relationships between leader and parent were explored. Significant results were demonstrated for both independently observed parenting change and parent-reported change, a significant finding considering the discrepancies often reported between such measures. The findings also confirm the importance of incorporating process measures into the assessment of treatment fidelity (Hohmann & Shear 2002; Paulson et al. 2002) where they are deemed to impact on the intervention, and demonstrate the important role that these process skills play in determining treatment outcome in such programmes. The findings are further strengthened given the stringent design of the original trial whereby fidelity was addressed by content and process checklist measures being completed by leaders, as well as receiving supervision by an IY mentor throughout the period of intervention delivery (Hutchings et al. 2007).

Despite the impressive findings, the study is not without its limitations. The study was somewhat limited because of the relatively small sample size of the leaders, although obtaining data from 12 community-led IV intervention groups led by 22 different staff lends strength to the findings. The calculation of exposure scores as opposed to having true data for each session of the intervention limits the study to some extent; however, representative sessions were chosen because of the time constraints, with observational coding being a time-consuming process. Despite these limitations, significant results were found within a ‘high risk’ community group, with experienced leaders delivering an evidence-based programme and receiving continuous supervision.

Treatment fidelity is an important consideration in all programme replications, but especially so within community-led interventions. This is because such replications are often more difficult to implement, delivered by staff of varying backgrounds and skills level, and the resources and time available to deliver an intervention with fidelity are usually limited. The ability to monitor and preserve the behaviour change mechanisms of an intervention is key, especially considering the increasing numbers of interventions delivered in such settings. This is of even greater importance when recognizing the increasing UK Government funds being allocated to provide PT to tackle the increasing numbers of disruptive behaviour problems among young children, as reflected by Sure Start and Flying Start initiatives in England and Wales to support families of preschool children living in identified high-risk areas (Department for Education and Skills 2005; Welsh Assembly Government 2005). By delivering interventions as early preventative support to families, and monitoring both content and process skills fidelity, high levels of programme effectiveness may be maintained, ensuring evidence-based interventions are delivered to the same standard as the original research evaluation of the programme, and therefore likely to achieve the same outcomes. In light of such initiatives, further investigation into properties of the LOT in determining outcome within routine clinical practice and community services would lend greater strength to the current findings. Moreover, its ability as a both clinical and research tool within such services, for other parenting programmes that employ the same core components of the IY PT programme, would broaden the LOT’s utility and aid comparisons of process fidelity skills in predicting and maintaining change in the longer term.

Key messages

- Poor parenting skills are the most common factor associated with the development and maintenance of CD.
- Change in positive parenting skills after PT acts as a partial mediator of child behaviour change.
- The amount of change in parenting skills varies and may be attributed to the process skills of the PT implementer.
- The LOT provides a quantitative measure of implementer process skills in delivering PT.
- Levels of LOT process skills significantly predict change in both observed positive parenting and parent-reported change in parenting style, which in turn, predict positive child outcome change.

References

American Psychiatric Association (2000) Diagnostic and Statistical Manual of Mental Disorders, 4th edn, Text Revision. Authors, Washington, DC, USA.


Institute of Medicine (2001) *Improving the Quality of Long-Term Care*. National Academy Press, Washington, DC, USA.


