Child: care, health and development

Original Article

doi:10.1111/j.1365-2214.2008.00828.x

The Leader Observation Tool: a process skills treatment fidelity measure for the Incredible Years parenting programme

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

College of Health and Behavioural Science, School of Psychology, University of Wales, Bangor, Gwynedd, UK

Accepted for publication 23 November 2007

Abstract

Background Despite recognition of the need to deliver evidence-based programmes in the field of mental health, there is little emphasis on implementing such programmes with fidelity. Attempts by programme developers to ensure adherence to their programmes include the development of training, manuals and content scales, but these alone may be insufficient to ensure fidelity in replication. Observational measures lend themselves as a potentially useful assessment of intervention outcomes, providing accurate and objective accounts of the intervention process.

Aim To develop a reliable and valid observational treatment fidelity tool of process skills required to deliver the Incredible Years (IY) BASIC parenting programme effectively.

Methods An objective observational fidelity measure was developed to assess adherence to the IY BASIC parenting programme protocol. Observations were conducted on 12 IY BASIC parenting programme groups, attended by parents of pre-school children displaying signs of early onset conduct disorder.

Results The Leader Observation Tool (LOT) achieved high internal reliability and good code–recode and inter-rater reliability. Evidence of concurrent validity was also obtained.

Conclusions Having demonstrated that the LOT is a reliable and valid measure of implementation fidelity, further research is necessary to examine the relationship between LOT scores and intervention outcome.

Keywords
behaviour, parenting, pre-school children, treatment

Correspondence:
Catrin Eames, College of Health and Behavioural Science, School of Psychology, University of Wales, Bangor, Gwynedd LL57 2AS, UK
E-mail:
c.eames@bangor.ac.uk

Introduction

Parenting intervention

Parenting that involves ineffectual commands, inconsistent discipline and punishment plays a significant role in the development and maintenance of conduct disorder (CD) (Patterson 1997; Gardner *et al.* 1999). Parent training (PT) aims to strengthen parent management skills, increase child prosocial behaviour and decrease child antisocial behaviour (Kazdin 1997). As a result, PT has been demonstrated to be the most effective intervention for both the prevention and treatment of CD in the early years (Kazdin 1997; Brestan & Eyberg 1998; Beauchaine *et al.* 2005). Studies indicate that fewer than 30% of children who need treatment for CD receive them (Brestan & Eyberg 1998). For those children that do receive a service, there is a wide variation in the type and level of services provided (Webster-Stratton 2003), with few evidence-based practices delivered by specialist practitioners (Kurtz *et al.* 1994). These needs are being addressed in the UK with guidance on the delivery of evidence-based parenting interventions for the treatment of CD, published by the National Institute for Health and

Clinical Excellence (NICE 2006). Moreover, Governments are now demonstrating a commitment to deliver evidence-based practices effectively, by training professionals working with parents, as funded by the Welsh Assembly Government (WAG) as part of their Parenting Action Plan for Wales (DfTE 2005) and the Respect Action Plan by the UK Government (Home Office 2006).

Treatment fidelity in programme replication

Moncher and Prinz (1991) maintain that in order for an intervention to be delivered effectively by therapists, in addition to having an adequate evidence base, it needs thorough documentation of the methods used and a means of assessing fidelity to ensure effective replication. Treatment fidelity may be referred to as the degree to which intervention delivery adheres to the original intervention protocol (Institute of Medicine 2001). With growing numbers of agencies adopting evidence-based programmes, delivered by a range of professionals, treatment fidelity becomes increasingly important. Programme drift is common when mainstream services deliver programmes (Bond *et al.* 2000) and fidelity criteria that document implementation during intervention can correct this (Mowbray *et al.* 2003). Without adequate fidelity measurement, there is no way of knowing what occurred during the intervention (Vermilyea *et al.* 1984; Chen 1990; Domitrovich & Greenberg 2000). Valid fidelity measures are particularly necessary components in multi-centred research in order to ensure that the intervention provided across centres is the same (Mihalic *et al.* 2002). If differences in outcome do emerge, these can be understood in terms of variations in delivery (Paulson *et al.* 2002). In randomized controlled trials, fidelity measures can establish the extent of treatment differentiation between groups (Epstein *et al.* 2005).

Treatment fidelity measures

Treatment fidelity is addressed when programme developers provide manuals, checklists, training and methods for assessing the quality of delivery, such as supervision and certification in their programmes (Schinke *et al.* 1991; Harchik *et al.* 1992). These tools enable community/service-based staff to deliver the programme to the same standard as the original evidence-based intervention (Epstein *et al.* 2005). Typically, such tools provide a measure of programme content, adopting checklist or Likert scale formats, to rate the degree of adherence to the theoretical dimensions or components of the intervention (Domitrovich & Greenberg 2000). Whereby these measures document the content of the programme protocol, process scales are concerned with the skills associated with delivering the content of the programme, primarily adopting a qualitative format, which is often not generalizable and requires complex analysis. The validity of these measures is, however, questionable; for example, Likert scales can provide confounding results owing to the nature of the mid-range attributions of the scale (Clark & Watson 1995). The validity of these measures is further threatened when used by service providers as self-monitoring tools, owing to the potential for subjective bias (Bentler 1969; Green *et al.* 1993).

Benefits of observation

Observation provides a rich source of information, enabling a precise account of interaction as it unfolds (Taplin & Reid 1977; Aspland & Gardner 2003). Direct observational methods provide accurate and objective accounts by directly measuring the behaviours of interest as they occur (Patterson *et al.* 1989; Webster-Stratton *et al.* 1989; Webster-Stratton & Herbert 1994; Webster-Stratton & Hancock 1998). Such accounts of behaviour could not be extracted as effectively from other measures (Margolin *et al.* 1998; Gardner 2000) and bias owing to treatment expectancy and an overestimation of the phenomenon under investigation is reduced with independent observations (Patterson 1982; Aspland & Gardner 2003).

Observational research uses both continuous and interval time-sampling methods. With continuous sampling, all behaviour that occurs within a specified time period is coded, whereas interval sampling consists of the coder typically having an 'observe' period followed by a 'record' period during which the behaviour in the 'observe' period is recorded on a score sheet. Continuous time-sampling systems provide a rigorous means of processing the 'stream' of reciprocal exchanges between individuals into meaningful discrete codes. This allows a continuous recording of the interaction and contributes to validity and utility by providing a complete account of all behaviour, permitting data to be collected in less time than typically required by interval sampling methods (Altmann 1974; Gardner 2000). Continuous sampling can be demanding on the coder, especially when interaction is dense, and therefore requires intensive training. Interval sampling, while somewhat easier on the coder, has the potential to lose crucial information and it becomes difficult to obtain truly interactive data (Powell 1984). Consequently, despite its demands, continuous sampling methods yield more precise interactive data, and are the preferred sampling method wherever possible (Gardner 2000).

Rationale for study

The IY BASIC Parenting Programme (Webster-Stratton 1989) is one of two programmes identified by NICE (2006) as a suggested intervention of best practice, as well as being funded by the WAG (DfTE 2005) and by the Department for Education and Skills (DfES) for the English Pathfinder Project (for more information, see http://www.respect.gov.uk/members/article.aspx?id=8846). There has been over 10 000 professionals and practitioners trained in the BASIC programme, including 3000 in the UK. The programmes are being delivered and evaluated internationally, including Australia, Canada, Denmark, England, Finland, Holland, Germany, Jamaica, Norway, New Zealand, Portugal, Wales and the USA (for more information, see http://www.incredibleyears.com).

With the numbers of independent replications and people being trained to deliver the programme continuously increasing, a new observational measure was developed to provide a quantitative measure of process skills associated with treatment fidelity of the IY BASIC parenting programme. The primary objective of this paper is to present the psychometric properties of the new measure.

Method

Participants

Twenty-two trained IY BASIC parenting group leaders from 12 groups delivered in 11 Sure Start areas (two groups were delivered in one setting), with two leaders per group (two leaders co-lead two different groups). Group leaders had a mean age of 44.3 years (ranging from 26 to 59 years old) with an average of nine IY groups run prior to the research and a combined average of five and a half years IY experience. Sure Start funding was drawn from the WAG by a variety of different organizations with six of the groups delivered by services managed by Barnardos children's charity, three by local health trusts, and the remaining three by the NCH children's charity, a local education authority and a local community group respectively.

Data set

The IY BASIC parenting programme is divided into four, three-session, sections: (1) play and relationship building; (2) praise and reward; (3) effective limit setting; and (4) handling misbehaviour. All 12 sessions had been videotaped as part of the IY BASIC Parenting evaluation (Hutchings *et al.* 2007). Four sessions, one from each section, were selected to be coded for each intervention group, providing a total data set of 48 2-h parenting sessions. Of these, 30% were randomly selected for second coding to establish reliability ($n = 14$). Secondary observations were conducted by trained coders, who had achieved an inter-rater overall agreement of 70% or above during training.

Measures

The Leader Observation Tool

Code development Components of the Leader Observation Tool (LOT) were theoretically derived from the work of Patterson and colleagues (1969), and Eyberg and Robinson (1981). Additional codes were created to depict the process skills used in delivering the content of the programme as outlined in the IY parenting manual (Webster-Stratton 1989).

There are 18 standard behaviour categories (see Table 1). These behaviour categories form four skill subgroups: listening, empathy, physical encouragement and positive behaviour. Table 1 provides an example of each behaviour category and lists their respective skill subgroup.

Each category is given a definition, a number of illustrative examples, specific guidelines to aid discrimination between categories, and decision rules to help the observer when there is uncertainty as to which category to code. Systematic descriptions in the LOT manual are designed to provide an accurate description of leader behaviour during a 2-h group session. Codes are exhaustive and mutually exclusive to the extent that only one unit of verbal behaviour by leaders can occur at any one time, although one unit of verbal, and one unit of non-verbal behaviour can occur simultaneously. The LOT allows for both leaders to be coded simultaneously in order to capture overall parental experience, and coding is continuous, with each coding sheet used to record the total frequency of each behaviour category per 10-minute interval. The coding sheet provides a space to record every leader verbalization and physical behaviour. Behaviours are coded by making a tally mark in the appropriate space on the recording sheet each time the behaviour occurs. Observations using the LOT may be conducted either by live coding at the group session or retrospectively from videotaped recordings, allowing for flexibility of both practical and research design considerations.

Additional measures

In addition to the LOT, a subsample of content and process skills measures from the IY BASIC PT manual (Webster-Stratton 1989) collected by leaders during the study were

Skill	Behaviour category	Example
Listening	Acknowledgement	Yes, no, hmm
	Clarifying question	You went shopping yesterday?
	Reflective	You've done your homework
Empathy	Feelings acknowledgement	That must have been hard
	Self-reflection	I've done the same
Physical encouragement	Positive body language	Thumbs up, nodding
	Positive effect	Smiles, laughs
	Physical positive	Pat on the back
Positive behaviour	Engagement	Could you help me please?
	Role play	Lets have a practice
	Praise	That's great, well done
	Principle reflection	How about Julie's principle.
	Thought provoking	What do you think will happen?
	Reframing	You realize and count to 10.
Other LOT categories	Negative body language	Frowning
	Critical	No, that's wrong
	Closed question	Was that good?
	Off agenda	When discussions veer off agenda
	On agenda	When discussions return on agenda
	Time off agenda	Off-agenda discussion duration

Table 1. Behaviour categories, with exemplars and summaries of the type of skills demonstrated, as defined by the Leader Observation Tool (LOT)

selected to explore the validity of the LOT. These consisted of parent ratings of leader skill and group discussion, scored on a Likert scale (process), and leader self-reports of content delivered in each session. The latter involved the degree to which they adhered to the protocol for the session, the total number of vignettes shown (video-clips that are used as a learning tool) and the total number of checklist items implemented during the session. These measures were routinely collected by the group leaders and were completed as part of the main trial by both leaders and parents involved in the programme.

Procedure

Coding team

The coding team consisted of the first author who developed the measure (primary coder) and two others (secondary coders). All coders had extensive knowledge of observational measures, with a high level of experience in conducting both live and videotaped parent–child and teacher–child observations. Once behaviour category codes had been finalized, the primary coder observed a video-recorded parenting group not included in the main data set, until code–recode reliability reached 70% agreement or above. Secondary coders then undertook an intensive training course on the LOT, led by the primary coder.

Initial training involved thorough reading of the manual, repeated testing of category understanding using a variety of learning resources and building familiarity of the coding sheet to develop speed. Initial training adopts a cumulative approach to learning codes so that consolidation of the categories learned is achieved and a good comprehension of each category is maintained throughout the training course. Training sessions make use of a variety of learning processes and include: tutorial and instruction with discussion, reviewing previously learned codes, assignment feedback, reading sections of the manual, a battery of learning materials (such as quizzes, dialogues, transcripts and practice coding on video sessions) to encourage long-term retention of codes, with the aim of inter-rater reliability between the primary coder and each secondary coder reaching an overall agreement of 70% or above. Subsequent weekly supervision ensured that a high level of overall percentage agreement reliability was maintained throughout the video analysis period.

Typically, initial training consists of 40 taught hours, with each subsequent supervision session running for a minimum of 1.5 h, depending on the number of coders in the team. Each member of the coding team, in this instance, had a good understanding of behaviour categories and their definitions owing to their high level of experience so that initial training, having achieved an overall inter-rater agreement above 70%, was completed in 30 h.

Videotapes

As a requirement of the main trial (Hutchings *et al.* 2007), each group recorded every parenting group session. These tapes were made available for the current study. The primary coder undertook the coding of all the parenting group sessions selected for

the data set. A subsample of these was second coded for reliability purposes.

Analysis strategy/data preparation

First, the internal reliability of the LOT was explored, together with the mean and standard deviations per group. Second, intraclass correlations were conducted on code–recode and inter-rater observations to establish reliability. Third, concurrent validity was explored by examining associations between LOT scores and parent and leader reported data. Finally, frequencies, consistencies and the range of the LOT behaviour categories were explored. Results of a Kolmogorov-Smirnov test showed that some variables violated the assumptions of normality, i.e. some variables were significant, therefore non-normal, indicating that the data would lend itself to non-parametric analysis.

Results

Internal reliability

Spearman's correlation coefficients were conducted to establish the internal reliability of LOT categories and are presented in Table 2. Despite the small data set, significant correlations at both the $P < 0.05$ level and the $P < 0.01$ level were obtained. Each behaviour category code demonstrated a significant positive correlation with at least one other behaviour category code. Categories have been logically grouped together into skills subgroups and these groupings demonstrate considerable internal reliability with moderate to large positive correlations.

Observer reliability

There is no agreed criterion for reliable observer agreement, although 70% and above is generally considered acceptable (Aspland & Gardner 2003). The first author was the primary coder and an average code–recode agreement of 87% was achieved over four 2-h tapes. Intraclass correlation coefficients (see Table 3) yielded high code–recode reliability results for each category of the measure. Average inter-rater agreement of 84% was achieved between three coders. Ten of the 48 tapes coded in this study were coded by two coders, the first author and one other (20% of all tapes), and highly reliable intraclass correlation coefficients between codes for each category of the LOT were achieved, with a mean of 0.9158 and 0.9479 for code–recode and inter-rater reliability correlations respectively (see Table 3). These demonstrate that coders achieved a high standard of coding reliability on each category of the LOT.

Validity

Concurrent validity was examined by correlating the LOT with both parent and leader reports about the process of the group intervention (see Table 4). Parental ratings of group discussion were positively correlated with observed physical positive behaviour and negatively correlated with observed leader negative body language indicating that parents rated group discussions more positively when leaders were positive in their behaviour towards them. Parent-reported ratings of group discussion yielded a significant positive correlation with clarifying questions, suggesting that parents rated discussions as more helpful when leaders actively demonstrated listening skills.

The number of vignettes shown during the session correlated significantly with principle reflection, indicating that the more vignettes shown during the group, the more the leaders drew/reflected upon principles from the parents. Leaders self-report checklists of the degree to which they adhered to the protocol for the session were positively correlated with physical positive, negatively with negative body language, and positively with engagement, reframing and reflective behaviours. Together, these yield some evidence for the concurrent validity of the LOT as a fidelity measure.

Variability in leader behaviour between groups

Mean and standard deviations of the LOT across the 12 groups indicate large variation between the groups (see Table 5). These indicate the existence of differences in leader styles in delivering the programme. Group 11 was identified as consistently demonstrating low frequencies of positive behaviours, listening skills, empathy and physical encouragement. Group 6 also scored low frequencies for listening skills and physical encouragement, together with a low frequency and high variance of positive behaviour. The group observed as demonstrating the highest frequency of positive behaviours, physical encouragement and listening skills was group 7. Group 8 demonstrated some consistency in their use of positive behaviour, listening skills and empathy with comparatively low standard deviations. While purely descriptive, an examination of Table 5 demonstrated that the leaders of group 7 engaged in almost four times as much listening behaviours as leaders in group 1. Leaders in group 9 displayed six times more empathic behaviours as leaders in group 11, leaders in group 5 displayed four times more encouragement as in group 6 and leaders in group 7

Table 2. Internal correlations of Leader Observation Tool behaviour categories ($n = 24$)

	1 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Acknowledgement (1)	**0.308**	**0.518****	0.378	0.038	0.838**	0.636**	0.324	0.365	0.668**	0.644**	0.545**	0.641**	0.614**	−0.243	−0.336	−0.004	0.251	0.199	0.346
Clarifying question (2)		**0.533****	0.475*	0.475*	0.371	0.220	0.325	0.359	0.260	0.448*	−0.084	0.433*	0.287	−0.205	0.317	0.319	0.307	0.227	0.109
Reflective (3)			0.832**	0.583**	0.482*	0.555**	0.551**	0.555**	0.266	0.660**	0.188	0.559**	0.670**	−0.230	0.140	−0.123	0.323	0.2880	0.225
Feelings acknowledgement (4)				**0.705****	0.281	0.495*	0.418*	0.658**	0.218	0.493*	0.155	0.662**	0.649**	−0.096	0.283	−0.122	0.263	0.224	0.083
Self-reflection (5)					−0.006	0.088	0.101	0.434*	−0.150	0.249	−0.101	0.406*	0.279	−0.360	0.295	−0.098	0.210	0.265	0.114
Positive body language (6)						**0.610****	**0.427***	0.235	0.552**	0.662**	0.498*	0.568**	0.627**	−0.197	−0.373	0.159	0.235	0.161	0.272
Positive effect (7)							**0.478***	0.501*	0.344	0.471*	0.429*	0.637**	0.743**	−0.217	−0.179	−0.031	0.173	0.143	0.111
Physical positive (8)								0.462*	0.310	0.636**	0.084	0.317	0.617**	−0.103	0.062	0.023	0.404	0.293	0.302
Engagement (9)									**0.373**	**0.554****	**0.305**	**0.374**	**0.634****	−0.213	−0.072	−0.137	0.099	0.218	0.076
Role play (10)										**0.339**	**0.578****	**0.550****	**0.466***	0.111	−0.030	0.082	0.146	0.174	0.176
Praise (11)											**0.359**	**0.405***	**0.715****	−0.145	−0.224	−0.057	0.416*	0.300	0.409*
Principle reflection (12)												**0.431***	**0.571****	−0.005	−0.463*	−0.133	−0.142	−0.092	−0.019
Thought provoking (13)													**0.672****	−0.129	0.165	0.047	0.342	0.285	0.260
Reframing (14)														−0.162	−0.238	−0.173	0.421*	0.358	0.339
Negative body language (15)															0.311	−0.258	−0.308	−0.415*	−0.481*
Critical (16)																0.010	0.130	0.106	−0.095
Closed question (17)																	0.083	−0.008	0.108
Off agenda (18)																		0.901**	0.878**
On agenda (19)																			0.818**
Time off agenda (20)																			

See parenthesis in the left-hand column for definitions of corresponding numbers in the title row. Bolded items indicate respective groups.
*$P < 0.05$; **$P < 0.01$.

Table 3. Intraclass correlation coefficients per category of the Leader Observation Tool for code–recode and inter-rater observations

Code–recode intraclass correlations		Inter-rater intraclass correlations	
Acknowledgements	0.9914	Acknowledgement	0.9697
Feelings acknowledgement	0.9930	Feelings acknowledgement	0.9712
Self-reflection	0.9720	Self-reflection	0.9792
Positive body language	0.9718	Positive body language	0.9587
Positive effect	0.9169	Positive affect	0.8383
Physical positive	0.9914	Physical positive	0.9846
Negative body language	0.6316	Negative body language	1.0000
Critical	1.0000	Critical	0.9480
Engagement	0.9325	Engagement	0.9866
Role play	1.0000	Role play	0.9410
Praise	0.9496	Praise	0.9881
Principle reflection	0.9811	Principle reflection	0.9198
Clarifying question	0.8828	Clarifying question	0.8550
Closed question	0.3810	Closed question	0.9405
Thought provoking	0.9662	Thought provoking	0.9851
Reframing	0.9874	Reframing	0.9744
Reflective	0.9753	Reflective	0.9105
Off agenda	1.0000	Off agenda	0.9640
On agenda	0.8571	On agenda	0.8959

Table 4. Spearman's correlation coefficients of the Leader Observation Tool (LOT) with parent-reported and leader-reported measures

	Parent-reported measure		Leader-reported measure	
LOT behaviour category	**Leader skills**	**Group discussion**	**Vignettes shown**	**Checklist items delivered**
Acknowledgements	–0.108	0.213	0.217	–0.061
Clarifying question	0.244	0.478*	–0.113	–0.013
Reflective	0.154	0.211	–0.032	0.424*
Feelings acknowledgement	–0.048	–0.019	0.051	0.368
Self-reflection	0.301	0.083	–0.194	0.341
Positive body language	0.029	0.338	0.082	–0.101
Positive effect	0.019	0.327	0.019	0.195
Physical positive	0.391	0.470*	–0.209	0.458*
Engagement	0.057	0.122	–0.018	0.413*
Role play	–0.200	–0.015	0.233	–0.142
Praise	0.255	0.406	–0.213	0.238
Principle reflection	–0.281	–0.025	0.456*	–0.127
Thought provoking	0.051	0.225	0.165	0.058
Reframing	0.131	0.359	0.100	0.430*
On agenda	0.440*	0.458*	0.034	0.671*
Off agenda	0.502*	0.563*	–0.082	0.628*
Time off agenda	0.462*	0.534*	0.51	0.559**
Negative body language	–0.235	–0.440*	–0.164	–0.437*
Critical	0.082	–0.188	–0.252	0.076
Closed question	0.092	0.193	–0.017	–0.214

*$P < 0.05$; **$P < 0.01$.

engaged in almost three times more positive behaviours as leaders in group 6. These figures highlight the heterogeneity in implementation fidelity between the groups.

Discussion

The results suggest that the LOT has strengths in terms of reliability and validity. All coders maintained an overall percentage agreement for both inter-rater and code–recode reliability above the required minimum. The reliability of the LOT was further confirmed by high intraclass correlations on each item. The LOT has achieved some evidence of concurrent validity. Significant correlations were achieved with parental ratings of group discussion, leader self-reported protocol checklists and the number of vignettes shown during the session, further strengthening its utility as an objective fidelity tool.

Table 5. Mean and standard deviations per group for each skills subgroup as depicted by the Leader Observation Tool (LOT)

Group (Mean attendance)	LOT skills subgroup Mean (SD)			
	Listening	Empathy	Physical encouragement	Positive behaviour
1 (5.2)	346.5 (194.5)	13.0 (17.0)	506.5 (103.9)	452.5 (378.3)
2 (7.8)	458.0 (268.7)	28.5 (14.8)	438.5 (67.2)	422.5 (99.7)
3 (4.3)	515.0 (67.9)	42.5 (7.8)	478.5 (101.1)	475.0 (58.0)
4 (5.6)	484.0 (125.9)	17.0 (14.1)	624.0 (7.1)	549.5 (221.3)
5 (4.7)	435.5 (38.9)	11.0 (1.4)	543.5 (33.2)	506.0 (131.5)
6 (4.6)	165.5 (188.8)	33.0 (41.0)	137.5 (139.3)	234.0 (250.3)
7 (6.3)	915.0 (485.1)	36.5 (3.5)	672.5 (156.3)	652.0 (83.4)
8 (6.4)	325.0 (35.4)	24.0 (4.2)	205.0 (104.7)	386.0 (34.0)
9 (5.8)	409.5 (94.0)	69.5 (20.5)	442.0 (73.5)	600.5 (75.7)
10 (8.8)	366.0 (442.6)	29.0 (41.0)	321.5 (359.9)	272.0 (258.8)
11 (5.7)	192.5 (48.8)	10.0 (1.4)	143.0 (9.9)	237.5 (34.6)
12 (4.4)	453.9 (323.7)	13.5 (0.7)	403.5 (362.7)	397.5 (96.9)

The frequencies of the LOT behaviour categories followed an expected trend, with positive behaviours and listening skills reported consistently across groups but with considerable variation. The utility of the LOT is further strengthened by its ability to identify a diverse range of group leader styles.

The primary objective was to develop a quantitative measure of process skills pertaining to treatment fidelity. The LOT utilizes the benefits of observation and continuous time sampling, enabling the documentation of specific leader behaviours in delivering an evidence-based intervention as they unfold (Patterson *et al.* 1989; Gardner 2000). It addresses limitations of current fidelity measures, such as ungeneralizable qualitative accounts of process, and subjective checklists or Likert content scales (Clark & Watson 1995; Aspland & Gardner 2003). By adopting quantitative objective observations, the threats to validity are reduced, such as overestimation and subjective biases (Taplin & Reid 1977; Aspland & Gardner 2003).

Limitations and future studies

Two limitations of the study are worthy of comment. First, owing to the focus of the LOT on process skills, the prospect of confirming its concurrent validity against the weekly session content checklists was limited; however, some evidence of concurrent validity was obtained. Second, the testing of the measure was reliant upon leaders running groups for parents with pre-school children. To confirm the reliability and validity of the measure further, it would be beneficial to apply the measure to sessions delivered to the programmes' identified age range. The next step in terms of research is to establish whether the components of treatment fidelity as measured by the LOT can predict behaviour change. To this end, an investigation into both the discriminant validity and the predictive validity of the LOT will aid this exploration. Examining the observation ratings of a session by trained IY leaders and/or mentors against LOT observations would further establish its validity.

Clinical implications

The LOT was developed primarily as a research tool, but could be used effectively as an audit tool to help clinical services to ensure that they are delivering the programme with fidelity. It has clinical implications for services in enabling them to maintain a high level of delivery skill, ensuring that families consistently receive a high standard of treatment. Moreover, families who attend an intervention group delivered with high levels of process skill and fidelity could potentially require fewer subsequent services, therefore reducing service costs and burden (Mihalic *et al.* 2002).

The LOT was developed to explore treatment fidelity of the IY BASIC parenting programme and has the potential to be used to evaluate what is effective in other programmes in the IY series because it focuses on the collaborative leader skills (for a full review of these programmes, see http://www.incredibleyears.com). These are also core components in other effective parent programmes. Its versatility as both a clinical and research tool in live and/or videotape-recorded sessions allows ongoing assessment of treatment fidelity by supervisors/mentors, and potentially as a self-supervision tool for leaders to identify the frequencies of their behaviours and level of adherence.

Key messages

- Parent training is the most effective intervention for both prevention and treatment of early onset CD.
- Only when an intervention has a sound evidence base and has thorough means of assessing implementation fidelity are successful results achieved.
- The IY BASIC Parenting Programme is an internationally evaluated intervention for CD, addressing fidelity through training and qualitative materials.
- Measuring implementation fidelity is complex, owing to the collaborative process involved in delivering the BASIC Parenting Programme effectively.
- The LOT depicts the process skills used in delivering the programme effectively, allowing for the measurement of fidelity in a quantitative manner.

References

Altmann, J. (1974) Observational study of behaviour: sampling methods. *Behaviour*, **49**, 227–267.

Aspland, H. & Gardner, F. (2003) Observational measures of parent–child interaction: an introductory review. *Child and Adolescent Mental Health*, **8**, 136–143.

Beauchaine, T. P., Webster-Stratton, C. & Reid, J. (2005) Mediators, moderators, and predictors of 1-year outcomes among children treated for early-onset conduct problems: a latent growth curve analysis. *Journal of Consulting and Clinical Psychology*, **73**, 371–388.

Bentler, P. M. (1969) Semantic space is (approximately) bipolar. *Journal of Psychology*, **71**, 33–40.

Bond, G. R., Evans, L., Salyers, M., Williams, J. & Hea-Won, K. (2000) Measurement of fidelity in psychiatric rehabilitation. *Mental Health Services Research*, **2**, 75–87.

Brestan, E. V. & Eyberg, S. M. (1998) Effective psychosocial treatments of conduct disordered children and adolescents: 29 years, 82 studies and 5272 kids. *Journal of Clinical Child Psychology*, **27**, 180–189.

Chen, H. (1990) *Theory-Driven Evaluations*. Sage, Thousand Oaks, CA, USA.

Clark, L. A. & Watson, D. (1995) Constructing validity: basic issues in objective scale development. *Psychological Assessment*, **7**, 309–319.

DfTE (2005) *Parenting Action Plan: Supporting Mothers, Fathers and Carers with Raising Children in Wales*. DfTE, Welsh Assembly Government Information Document, Cardiff, UK. No: 054-05, December 2005.

Domitrovich, C. E. & Greenberg, M. T. (2000) The study of implementation: current findings from effective programs that prevent mental disorders in school-aged children. *Journal of Educational and Psychological Consultation*, **11**, 193–221.

Epstein, M. H., Kutash, K. & Duchnowski, A. J. (2005) *Outcomes for Children and Youth with Emotional and Behavioural Disorders and Their Families: Programs and Evaluation Best Practices*, 2nd edn. PRO-ED Inc., Austin, TX, USA.

Eyberg, S. M. & Robinson, E. A. (1981) *Dyadic Parent–Child Interaction Coding System*. University of Washington, The Parenting Clinic, Washington, DC, USA.

Gardner, F. (2000) Methodological issues in the direct observation of parent–child interaction: do observational findings reflect the natural behaviour of participants? *Clinical Child and Family Psychology Review*, **3**, 185–198.

Gardner, F., Sonuga-Barke, E. & Sayal, K. (1999) Parents anticipating misbehaviour. An observational study of strategies parents use to prevent conflict with behaviour problem children. *Journal of Child Psychology and Psychiatry*, **40**, 1185–1196.

Green, D. P., Goldman, S. L. & Salovey, P. (1993) Measurement error masks bipolarity in affect ratings. *Journal of Personality and Social Psychology*, **64**, 1029–1041.

Harchik, A. E., Sherman, J. A., Sheldon, J. B. & Strouse, M. C. (1992) Ongoing consultation as a method of improving performance of staff members in a group home. *Journal of Applied Behavioural Analysis*, **25**, 599–610.

Home Office (2006) Respect Action Plan. Availale at: http://www.homeoffice.gov.uk/documents/respect-action-plan?view=Binary (accessed from 3 April 2006).

Hutchings, J., Bywater, T., Daley, D., Gardner, F., Whitaker, C., Jones, K., Eames, C. & Edwards, R. T. (2007) A pragmatic randomised controlled trial of a parenting intervention in Sure Start services for children at risk of developing conduct disorder. *British Medical Journal*, **334**, 679–682.

Incredible Years (IY) Series (no date) *Parents, Teachers and Children Training Series*. Available at: http://www.incredibleyears.com, 30/1/07 (accessed from 12 June 2007).

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

Kazdin, A. E. (1997) Practitioner review: psychosocial treatments for conduct disorder in children. *Journal of Clinical Psychology and Psychiatry*, **38**, 161–178.

Kurtz, Z., Thornes, R. & Wolkind, S. (1994) *Services for the Mental Health of Children and Young People in England*. South Thames Regional Health Authority, London, UK.

Margolin, G., Oliver, P. H., Gordis, E. B., O'Hearn, H. G., Medina, A. M., Ghosh, C. M. & Morland, L. (1998) The nuts and bolts of behavioural observation of marital and family interaction. *Clinical Child and Family Psychology Review*, **1**, 195–203.

Mihalic, S., Fagan, A., Irwin, K., Ballard, D. & Elliott, D. (2002) *Blueprints for Violence Prevention Replications: Factors for Implementation Success*. Center for the Study and Prevention of Violence, Boulder, CO, USA.

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

Mowbray, C. T., Holter, M. C., Teague, G. B. & Bybee, D. (2003) Fidelity criteria: development, measurement, and validation. *American Journal of Evaluation*, **24**, 315–340.

National Institute for Health and Clinical Excellence (NICE) (2006) *Parent Training/Education Programmes in the Management of Children with Conduct Disorders.* NICE technology appraisal. Guidance 102. SCIE, NHS, London, UK. Available at: http://www.nice.org.uk/TA102 (accessed from 8 September 2006).

Pathfinder & FIP Projects (no date) Family Intervention. Information available at: http://www.respect.gov.uk/members/article.aspx?id=8846 (accessed from 20 July 2007).

Patterson, G. R. (1982) *Coercive Family Process*. Castalia, Eugene, OR, USA.

Patterson, G. R. (1997) Performance models for parenting: a social interactional perspective. In: *Parenting and the Socialization of Values: A Handbook of Contemporary Theory* (eds J. Grusec & L. Kuczynski), pp. 193–235. Wiley, New York, NY, USA.

Patterson, G. R., DeBaryshe, B. D. & Ramsey, E. (1989) A developmental perspective on anti-social behaviour. *American Psychologist*, **44**, 329–335.

Patterson, G. R., Ray, R. S., Shaw, D. A. & Cobb, J. A. (1969) *A Manual for Coding Family Interactions*. ASIS National Auxiliary Publications Service, c/o CMM Information Service Inc., 909 Third Avenue, New York, NY 10002. Document No. D1234.

Paulson, R. I., Post, R. L., Henrickx, H. A. & Risser, P. (2002) Beyond components: using fidelity scales to measure and assure choice ion program implementation and quality assurance. *Community Mental Health Journal*, **38**, 119–128.

Powell, J. (1984) Some empirical justification for a modest proposal regarding data acquisition via intermittent direct observation. *Journal of Behavioural Assessment*, **6**, 71–80.

Schinke, S. P., Botvin, G. J. & Orlandi, M. A. (1991) *Substance Abuse in Children and Adolescents: Evaluation and Intervention (22)*. Sage, Thousand Oaks, CA, USA.

Taplin, P. S. & Reid, J. B. (1977) Changes in parent consequences as a function of family intervention. *Journal of Consulting and Clinical Psychology*, **45**, 973–981.

Vermilyea, B. B., Barlow, D. H. & O'Brien, G. T. (1984) The importance of assessing treatment integrity: an example in the anxiety disorders. *Journal of Behavioural Assessment*, **6**, 1–11.

Webster-Stratton, C. (1989) *The Incredible Years: The Parents and Children Series*. IY, Seattle, WA, USA.

Webster-Stratton, C. (2003) Aggression in Young Children Perspective: Services Proven to be Effective in Reducing Aggression. Available at: http://www.incredibleyears.com/research/article-aggression-in-young-children-perspective.pdf (accessed from 8 January 2005).

Webster-Stratton, C. & Hancock, L. (1998) Training for parents of young children with conduct problems: content, methods, and therapeutic processes. In: *Handbook of Parent Training* (eds C. E. Schaefer & J. M. Briesmeister), pp. 98–152. John Wiley, New York, NY, USA.

Webster-Stratton, C. & Herbert, M. (1994) *Troubled Families-Problem Children*. John Wiley, Chichester, UK.

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

