Expressed Emotion about children: reliability and validity of a Camberwell Family Interview for Childhood (CFI-C)

STEPHEN SCOTT, Institute of Psychiatry, Camberwell, London, UK
CAROLINE CAMPBELL, Maudsley Hospital, Camberwell, London

ABSTRACT A Camberwell Family Interview for Childhood (CFI-C) was developed by adding questions about the family impact of the child’s problems to a semi-structured interview on child psychiatric symptoms. The whole CFI-C took under an hour to administer; the questions about family impact added 15–20 minutes. The inter-rater reliability was good (kappa 0.64–1.0). Mothers of 25 boys aged four to nine years referred with disruptive behaviour, and 25 matched controls were interviewed twice in five months. Test-retest stability was fair to good (kappa 0.36–1.0). Discriminant validity between referred and control samples was strong for critical comments, positive comments and warmth, but not significant for emotional overinvolvement or hostility. The same three scales showed strong discriminant validity between child symptom domains, being strongly correlated with conduct symptoms (kappa = 0.49–0.71) but not emotional symptoms (kappa = 0.10–0.17). Sensitivity to change with treatment was shown by a reduction in the mean number of critical comments from 4.7 to 2.9, an increase in positive comments from 2.3 to 3.9, and an increased score on the warmth scale from 2.1 to 2.6. The CFI-C is a useful instrument for the study of the relationship between parenting style and child psychiatric symptoms.

Key words: Expressed emotion, childhood, conduct problems, emotional problems

Introduction
Expressed emotion (EE) was a concept invented by George Brown and Michael Rutter 35 years ago, while working at the Institute of Psychiatry in Camberwell, London (Brown and Rutter 1965; Rutter and Brown 1965). They devised the Camberwell Family Interview (CFI) to measure emotions expressed by relatives about adult psychiatric patients. The original CFI often lasted over three hours, so Vaughn and Leff (1976) developed the abbreviated version that typically took little more than an hour. The shortened CFI has been used to study a wide range of adult psychiatric disorders. The most widely replicated finding is that high EE predicts relapse in patients with schizophrenia (Bebbington and Kuipers, 1994). High EE also predicts prognosis in several other disorders, such as anorexia nervosa and depression (Butzlaff and Hooley, 1998).

The psychometric properties of the abbreviated CFI are good (Bentsen, Boye, Munkvold, and Uren, 1996), but it still takes over an hour to administer. This has led to the development of the Five Minute Speech Sample (FMSS) (Magana, Goldstein, Karno, Miklowitz, Jenkins and Falloon, 1986), in which relatives are instructed ‘I’d like you to tell me what kind of person X is, and how you get along together’ and their responses tape recorded for five minutes. Coding is along dimensions with similar names to the original CFI. However, because the material elicited is necessarily so limited, the designation of high EE often depends on the presence of a single comment. Moreover, whilst the FMSS purports to measure EE, the definition of terms is substantially different in the FMSS. When the two methods were directly compared, the FMSS misclassified 43% of high EE cases in one study (Magana et al., 1986) and 52% in another (Leeb, Hahlweg, Goldstein, Feinstein, Mueller, Dose, and Magana-Amato, 1991). As the FMSS uses such a small sample of human communication, codes it using different concepts from the CFI, and gives substantially different results regarding the presence of EE, there may be considerable value in using the CFI, which has proved so useful in predicting the course of adult disorders. Using the CFI would become practicable for larger scale surveys if it could be administered concurrently with another interview, and so took only an extra 20 minutes or so to give.
In child psychiatry there have been relatively few studies of EE using the CFI, and none that we are aware of specifically examining its psychometric properties. This is perhaps surprising given that many theories of child psychopathology are concerned with the destructive effects on children of parental criticism and negative attention. For example, Patterson, Chamberlain and Reid (1982) in longitudinal and laboratory studies have put forward convincing evidence that parental negativity is a major cause of child conduct disorder. Richman, Stevenson and Graham (1982) showed criticism was strongly related to antisocial behaviour but not emotional problems in three year olds, but it was not predictive of persistence of disorder at eight years. This study reported measuring criticism using global ratings rather than the CFI. Vostanis, Nicholls and Harrington (1994) used a form of the CFI and found more critical comments, fewer positive comments and less warmth expressed by parents of children with conduct problems than controls. Parents of children with emotional problems were less critical than even the controls, but were also less positive and less warm. At a nine-month follow up of the same sample, EE levels in both child diagnostic groups were approaching the level of the controls (Vostanis and Nicholls 1995). Schwartz, Dorer, Beardslee, Lavori and Keller (1990) used the CFI but only reported criticism as present or absent, and showed it was associated with increased levels of child psychopathology but did not subdivide by type. Szumukler, Eisler, Russell and Dure (1985) studied parents of adolescents with eating disorders. Although the amount of criticism was low, typically one or two comments, the level nonetheless predicted dropping out of treatment.

The limited nature of the information from the relatives used to score EE in the FMSS suggests that it would be valuable to develop a CFI for childhood that did not add more than say 15–20 minutes to a clinical interview about a child, but that offered the possibility of the richness of information that the CFI can give. We therefore set out to add the CFI to an established semi-structured psychiatric interview for children and to see whether it had good reliability and test–retest stability. If so, we wished to discover whether the CFI for Childhood (CFI-C) showed good validity in discriminating groups, had different strengths of association with disruptive as opposed to emotional symptoms, and was sensitive to change.

Method

Overview
The instrument was developed and field tested, and its inter-rater reliability was established. Then it was administered twice, with an interval of five months, to parents of 50 children. Twenty-five had been referred to local services and went on to receive treatment, and 25 were matched controls from local schools.

Instrument development

Rationale
The overriding principle was to keep as closely as possible to the standard Camberwell Family Interview as used with adults (Vaughn and Leff, 1976). This meant first asking questions about the main symptoms of childhood behavioural and psychiatric disorders, then specifically probing for their impact on family life. The symptom questions were taken from the Parent Account of Child Symptoms (PACS), a relatively short semi-structured interview of demonstrated reliability and validity developed for use with prepubertal children by Eric Taylor (Taylor et al., 1986; Taylor et al., 1991). The PACS is similar in style to the Isle of Wight Interview (Graham and Rutter, 1968) but takes only 30–40 minutes. The PACS covers three symptom domains:

- Emotional problems rating frequency and severity of misery, worries, health worries, fears, eating problems, and sleeping problems.
- Hyperactivity rating attention span, motor restlessness, and fidgeting while watching TV, reading
alone, playing alone, and playing with other children; rating motor restlessness during mealtimes, out shopping, and on family outings.

- **Conduct problems** rating frequency and severity of lying, stealing, temper tantrums, rudeness, disobedience, refusal to go to bed, aggressiveness, and destructiveness.

For the main problem in each symptom domain, parents are asked about:

- **Effects on family life.** Probes include:

  - How much of a problem do you consider this to be?
  - What effect has this problem had on the family?
  - What effect has it had on your life?
  - How has it affected your routine?
  - How do you feel when X behaves like that?
  - What goes through your mind?

- **Parents’ way of coping.** Probes include:

  - How do you cope when X behaves like that?
  - What did you do the last time?
  - Did it work?
  - How effective were you in dealing with it?
  - Does your 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?

The final section of the interview contains probes to elicit the parent’s feelings about the child as a person and their relationship in general, rather than specific behaviours, and ends on a positive note, asking what are the best things about him and what he is really good at.

**The scales**

Critical comments and positive comments are rated individually throughout the interview, so scores could vary from zero to 10 or more. Warmth is rated at the end of the interview on a global scale: 0 = no warmth, through to 5 = high warmth. Similarly, emotional overinvolvement is rated globally on a 0–5 scale. Hostility is rated as absent (score 0), generalized (score 1), rejection (score 2), or both generalized hostility plus rejection (score 3). However, hostility of any type is uncommon, and hostility of the specific subtypes is rare; previous studies have failed to find that subtypes of hostility have any validity. Therefore, in line with other expressed emotion research, hostility was recorded as present or absent.

**Field testing**

The interview was tried out on five clinically referred children and the form of wording slightly modified to improve the flow and precision. A copy is available from the first author on request.

**Inter-rater reliability**

In order to ensure that the criteria for this interview would be the same as those in the original Camberwell Family Interview, each of the authors was trained by Dr Christine Vaughn, the originator and official trainer of the modern shorter form of the CFI (Vaughn and Leff, 1976). This involved taking an intensive two-week course and rating subsequent interviews until acceptable agreement with standardized tapes was reached.

Next, the mothers of ten boys aged four to eight years, referred to the Maudsley Hospital for disruptive behaviour, were interviewed. The mean duration of the interview was 54 minutes (sd = 17, range = 32–78). The interviews were tape recorded and then rated independently by each author.

**Samples for main field trial**

Two samples were recruited for testing – one clinical and one consisting of matched controls. The clinical sample was of 25 boys aged four to nine years who had been consecutively referred to The Maudsley Hospital and three local clinics because of disruptive behaviour. The other criteria were no apparent general learning disability (mental retardation), and willingness by the parents to engage in treatment. The control sample was chosen by approaching two schools in the same area. For each boy in the clinical group, teachers were asked to identify a control boy who had no behaviour problems as far as they knew, who matched the clinical one on four criteria:

- age (to within three months);
- family configuration (two birth parents, lone parent, or reconstituted family);
- parental employment (professional, manual, unemployed); and
- ethnic group.

The researchers checked that selected control
pupils met the matching criteria. Only mothers were interviewed.

Procedures
The study was approved by the local research ethics committee. Parents who consented to take part were visited at home where the interviews took place and were tape recorded. The clinical group was then offered treatment, and attended a mean of 5.8 sessions (sd = 4.7, range = 1–11). Repeat interviews were carried out after a mean of 5.2 months (sd = 2.3, range = 3.9–7.8); the control group were followed up after a mean of 5.1 months (sd = 1.7, range = 4.2–7.1). The difference in the interval between assessments was 0.1 month. All statistical calculations were performed using SPSS 9.0 package.

Results

Inter-rater reliability
Intraclass correlations were 0.77 for critical comments and 0.71 for positive comments, calculated by single estimations using a one-way random effects model, which takes the component of the variance arising from the subjects as the numerator and the variance components of the subjects and an error term as the denominator. An unweighted kappa statistic was used as this is more conservative. Although the weighted version has the advantage of giving credit for close ratings, rather than treating them as total disagreements, there is a range of possible weights that could be assigned, and the interpretation of resulting kappa values is harder. The unweighted kappa statistic was 0.66 for warmth, 0.64 for emotional overinvolvement, and 1.0 for hostility, which was present in two out of 10 cases, for which agreement was total. Crosstabulations for each rater's scores are given in Table 1. These values indicate good levels of agreement for all scales.

Main study sample characteristics
These are given in Table 2. None of the group differences approached statistical significance. As might be expected for an inner-city sample suffering conduct problems, lone parenthood and poverty were common. Child age, family configuration, socioeconomic status and race showed no association with any EE scale on initial testing by Spearman correlation, so are not considered further as possible covariates in analyses.

Table 1: Inter-rater reliability of expressed emotion scales

(a) Crosstabulation of warmth scores

<table>
<thead>
<tr>
<th>Warmth score</th>
<th>Rater 2</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater 1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>10</td>
</tr>
</tbody>
</table>

Kappa 0.66, p = 0.000

(b) Crosstabulation of emotional overinvolvement scores

<table>
<thead>
<tr>
<th>Emotional overinvolvement</th>
<th>Rater 2</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater 1</td>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>10</td>
</tr>
</tbody>
</table>

Kappa 0.64, p = 0.002

(c) Crosstabulation of hostility scores

<table>
<thead>
<tr>
<th>Hostility</th>
<th>Rater 2</th>
<th>0</th>
<th>1</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater 1</td>
<td></td>
<td>8</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>8</td>
<td>2</td>
<td>10</td>
</tr>
</tbody>
</table>

Kappa 1.0, p = 0.002

Test–retest stability
This was calculated for the control sample only as the clinical sample had received treatment designed to change levels of expressed emotion. The measurement interval was 5.1 months. The intraclass correlation for critical comments was 0.68, for positive comments 0.66. The kappa value for warmth was 0.42, for emotional overinvolvement 0.36, and for hostility 1.0 – there was only one case of hostility in this sample, and the rating did not change. Crosstabulations of each rater's scores are given in Table 3. The values suggest moderate to strong test-retest stability.
Discriminant validity between groups

The distribution of critical comments was irregular in the clinical sample but approximately normal in the controls, so the paired difference scores were not normally distributed, rendering parametric testing invalid. The distribution of positive comments also varied between the samples, with difference scores also not being normally distributed. Warmth and emotional overinvolvement were rated on five-point scales so could not be normally distributed. Therefore the clinical sample and the controls were compared using Wilcoxon's ranked pairs test (two-tailed) for all scales except hostility, which was compared using McNemar's test for a dichotomous variable. The results are given in Table 4. Critical and positive comment counts, and the warmth scale discriminated well between groups, but emotional overinvolvement and hostility did not.

Discriminant validity across symptom domains

The clinical and control samples were combined to ensure a range of symptom scores. Spearman correlations are given in Table 5. Conduct problems were strongly associated with critical comments, few positive comments and a lack of warmth. They were weakly associated with hostility and not at all with emotional scores. There was a risk that these associations were made stronger by combining the samples, so correlations were also calculated for each sample separately. The same pattern was seen in each sample, with conduct symptoms being most strongly associated with measures of EE, then hyperactivity, with only weak associations with emotional scores. It should be pointed out that emotional scores were not high in either sample, as subjects were not chosen for being high on emotional problems, so higher correlation values might have been obtained with a sample with emotional disorders; this has not, however, been reported in the literature.

Sensitivity to change after treatment in clinical group

To test whether the CFI-C was sensitive to change, we tested the difference in scores before and after treatment in the clinically referred group, using the Wilcoxon signed ranks test (McNemar for hostility).
Table 4: Discrimination between clinical and control samples

<table>
<thead>
<tr>
<th></th>
<th>Clinical sample n = 25</th>
<th>Control sample n = 25</th>
<th>Significance of difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (sd)</td>
<td>Range</td>
<td>Mean (sd)</td>
</tr>
<tr>
<td>Critical comments</td>
<td>4.6 (3.1)</td>
<td>1–12</td>
<td>1.6 (1.3)</td>
</tr>
<tr>
<td>Positive comments</td>
<td>2.3 (1.7)</td>
<td>0–7</td>
<td>4.4 (2.9)</td>
</tr>
<tr>
<td>Warmth</td>
<td>2.1 (1.1)</td>
<td>0–4</td>
<td>3.0 (0.9)</td>
</tr>
<tr>
<td>Emotional overinvolvement</td>
<td>0.64 (1.1)</td>
<td>0–3</td>
<td>0.24 (0.6)</td>
</tr>
<tr>
<td>Hostility</td>
<td>20%</td>
<td></td>
<td>4%</td>
</tr>
</tbody>
</table>

1 McNemar test

Table 5: Correlation between measures of parental expressed emotion and child symptom domains (samples combined, n = 50)

<table>
<thead>
<tr>
<th></th>
<th>Conduct problems</th>
<th>Hyperactivity</th>
<th>Emotional problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical comments</td>
<td>0.71**</td>
<td>0.58**</td>
<td>0.16</td>
</tr>
<tr>
<td>Positive comments</td>
<td>−0.61**</td>
<td>−0.43**</td>
<td>−0.17</td>
</tr>
<tr>
<td>Warmth</td>
<td>−0.49**</td>
<td>−0.35*</td>
<td>−0.17</td>
</tr>
<tr>
<td>Emotional overinvolvement</td>
<td>−0.08</td>
<td>0.13</td>
<td>0.10</td>
</tr>
<tr>
<td>Hostility</td>
<td>0.28*</td>
<td>0.16</td>
<td>−0.12</td>
</tr>
</tbody>
</table>

** correlation is significant at the 0.01 level (two-tailed); * at the 0.05 level

Table 6: Sensitivity to change in clinical sample before and after treatment

<table>
<thead>
<tr>
<th></th>
<th>Before n = 25</th>
<th>After n = 25</th>
<th>Significance of difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (sd)</td>
<td>Range</td>
<td>Mean (sd)</td>
</tr>
<tr>
<td>Critical comments</td>
<td>4.6 (3.1)</td>
<td>1–12</td>
<td>2.2 (2.7)</td>
</tr>
<tr>
<td>Positive comments</td>
<td>2.3 (1.7)</td>
<td>0–7</td>
<td>3.9 (2.1)</td>
</tr>
<tr>
<td>Warmth</td>
<td>2.1 (1.1)</td>
<td>0–4</td>
<td>2.6 (1.1)</td>
</tr>
<tr>
<td>Emotional overinvolvement</td>
<td>0.64 (1.1)</td>
<td>0–3</td>
<td>0.24 (0.7)</td>
</tr>
<tr>
<td>Hostility</td>
<td>20%</td>
<td></td>
<td>0%</td>
</tr>
</tbody>
</table>

1 McNemar test
Results are given in Table 6. These show highly significant changes in critical comments, positive comments, and warmth, a moderately significant reduction in emotional over-involvement, and a trend towards reduction of hostility that did not reach significance.

Discussion
This study has shown that a Camberwell Family Interview for Childhood (CFI-C) can be completed in under an hour as part of a semi-structured interview to assess child symptomatology. The extra time taken for the CFI-C questions was around 15–20 minutes. The format was successful in eliciting expressed emotion, and the scores on each of the scales were similar to those found in the study of comparable children by Vostanis et al. (1994). Inter-rater reliability was good, test–retest stability over five months was moderate to good. The ability of the critical comment, positive comment and warmth scales to discriminate between clinical and control groups was very strong, but emotional over-involvement and hostility scales did not distinguish the groups. This may be because both emotional over-involvement and hostility were seldom detected, so a floor effect operated. Also emotional overinvolvement tends to be associated with more severe psychosomatic and emotional disorders, which were uncommon in this sample.

Good discrimination was found between child symptom domains on the same three scales. The disruptive symptoms (conduct problems and hyperactivity) were strongly associated, whereas emotional symptoms were not. This is in keeping with epidemiological studies that used global ratings such as the Isle of Wight study (Rutter, Tizard and Whitmore, 1970). The CFI-C showed good sensitivity to change associated with treatment. Many measures that discriminate well between groups are less sensitive to change, but the CFI-C is very useful for this task. Our findings are similar to those of Vostanis and Nicholls (1995) in their nine-month follow up of treated patients using the CFI.

The sample used here was relatively small, so findings should be interpreted bearing this in mind. However, the results are in line with those found by others.

Conclusion
The CFI-C provides a reliable and valid way to measure EE that is sensitive to change and only adds 15–20 minutes to an appropriate semi-structured interview of child symptomatology. It should be a useful tool to help disentangle cause and effect in childhood disorders. This is particularly pertinent at present when older theories about the prime importance of parental behaviour on child development are being challenged by findings from the new behavioural genetics. These have been interpreted by some to suggest that, for the majority of children, shared family influences such as parenting style are insignificant in determining many personality characteristics and disorders (Scarr, 1992; Rowe, 1994). The truth is likely to involve a more complex interaction of genetic and environmental influences (Rutter, Silberg, O’Connor and Simonoff, 1999), and to unravel it, sound methods of specifying emotional environment such as the CFI-C will be needed.

Acknowledgements
Dr Scott was supported by a Wellcome Research Training Fellowship while the initial part of this study was carried out, and the research was supported by a grant from the UK Medical Research Council. We are grateful to the parents and schools involved, and staff at Brixton, Belgrave, and Camberwell Child and Adolescent Mental Health Services for enabling the children to be studied.

References
Hibbs E D, Hamburger SD, Lenane M, Rapoport JL, Krueger MJP, Keyser CS, Goldstein MJ. Determinants of

Correspondence to Stephen Scott, Senior Lecturer in Child and Adolescent Psychiatry, Institute of Psychiatry, De Crespigny Park, Camberwell, London SE5 8AF. Telephone (+44) 0207-919-3467 Fax (+44) 0207-708-5800 Email: Stephen.Scott@iop.kcl.ac.uk