Assessment of doctors’ consultation skills in the paediatric setting: the Paediatric Consultation Assessment Tool

R J Howells,1 H A Davies,2 J D Silverman,3 J C Archer,4 A F Mellon5,6

ABSTRACT
Objective: To determine the utility of a novel Paediatric Consultation Assessment Tool (PCAT).

Design: Developed to measure clinicians’ communication behaviour with children and their parents/guardian, PCAT was designed according to consensus guidelines and refined at a number of stages. Volunteer clinicians provided videotaped real consultations. Assessors were trained to score communication skills using PCAT, a novel rating scale.

Setting: Eight UK paediatric units.

Participants: 19 paediatricians collected video-recorded material; a second cohort of 17 clinicians rated the videos.

Main outcome measures: Itemised and aggregated scores were analysed (means and 95% confidence intervals) to determine measurement characteristics and relationship to patient, consultation, clinician and assessor attributes; generalisability coefficient of aggregate score; factor analysis of items; comparison of scores between groups of patients, consultations, clinicians and assessors.

Results: 188 complete consultations were analysed (median per doctor = 10). 3 videos marked by any trained assessor are needed to reliably (r>0.8) assess a doctor’s triadic consultation skills using PCAT, 4 to assess communication with just children or parents. Performance maps to two factors – “clinical skills” and “communication behaviour”; clinicians score more highly on the former (mean (SD) 95% CI 0.52 (0.075)). There were significant differences in scores for the same skills applied to parent and child, especially between the ages of 2 and 10 years, and for information-sharing rather than relationship-building skills (2-tailed significance <0.001).

Conclusions: The PCAT appears to be reliable, valid and feasible for the assessment of triadic consultation skills by direct observation.

Embedded in good medical practice and other models of professional activity, doctor–patient interactions are central to clinical practice.1,2 The General Medical Council’s 0–18 years: guidance for all doctors emphasises the essential role of effective communication for the good care of children and young people.3

Common among the methods used to assess doctor–patient communication are patient and peer ratings. International research has shown peer ratings (multisource feedback) to be reliable, feasible and versatile for assessment of many attributes.4–6 Parent and adult patient rating tools have been developed for communication assessment.7–9

Some licensing bodies make use of direct observation, either live or via video/audio recordings, to measure doctor–patient interactions. Of these, video appears to have the greatest effect on behavioural response to feedback perhaps because

What is already known on this topic
- Video observation is suitable for in-training assessment.
- Validated tools for assessment of consultation skills have focused on two-way consultations with adult patients.

What this study adds
- Reliable assessment of paediatric triadic consultation skills can be achieved by applying the Paediatric Consultation Assessment Tool to 2–3 cases.
- Reliable, individual assessment of child-oriented and parent-oriented communication is also possible within 3–4 cases.
- Training in paediatric communication skills should emphasise consulting with children themselves, and on information sharing rather than rapport building.

METHOD
PCAT design: assuring content validity
We designed PCAT as an itemised rating scale to simultaneously but separately rate doctor–parent and doctor–child communication. PCAT’s content was configured according to consensus guidelines, a model of competencies for paediatric consultations, and the Calgary–Cambridge Referenced Observation Guide’s scheme.12–14 We included items relating to:
- Relationship building
- Structuring the consultation
- Initiating the consultation
- Information-gathering behaviour
- Information-giving behaviour and shared decision-making
Closure of the consultation

Following a nominal group exercise conducted using Royal College of Paediatrics and Child Health (RCPCH) college tutors and regional advisors in 2004, we included items assessing clinicians’ clinical skills and judgement, as these were deemed ‘essential skills’.

PCAT’s format (figs 1–4) was configured according to the best practice of developing health measurement scales, comprising:

- Assessment of skills (rather than consultation outcomes) and use of an itemised rating scale, for maximising educational potential through feedback.\(^{15}\)
- Sixteen scores related to themed groups of individual skills and two “global” scores.
- A 7-point scale: an optimal number of response categories, feasible for use without substantial loss of information, marked using behaviourally anchored ratings.\(^{16–18}\)
- Space for comments/observations, for feedback.

PCAT was tested and refined through an iterative process at four time points. Four authors initially piloted the PCAT for face validity and scope using videotaped consultations. At this stage the number of items were determined. In a second

---

**Figure 1**

Paediatric Consultation Assessment Tool

Use one copy per consultation.

On this page, give details of the consultation, the clinician and yourself. Then score the clinician’s diagnostic and management skills.

Overleaf, make judgments about the skills the clinician uses to communicate with the child / young person and his/her carers. Use the behaviourally anchored ratings (pages 3,4) to guide you. Fill in the boxes for each individual skills, then use these judgements to guide your overall scores for each section.

Where indicated, and whenever the patient is able to contribute to the conversation, evaluate the skills used with child and parent/s separately.

Use the free text boxes to record what you saw / heard. Record actual phrases if you can.

### Consultation

| Patient ID |  |
| Age of child / young person |  |
| Diagnosis / Diagnoses |  |
| Length of consultation (minutes) |  |
| Other people present |  |

### Clinician

| Name | Cons | SpR | SHO | Other |
| Seniority of clinician |  |

### Observer

| Name | Cons | SpR | SHO | Other |
| Seniority of observer |  |

### Content Skills

<table>
<thead>
<tr>
<th>1 = very poor, 7 = very good</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gathers relevant essential information</td>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>Undertakes pertinent examination</td>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>Generates appropriate diagnosis</td>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>Formulates relevant management plan</td>
<td>1 2 3 4 5 6 7</td>
</tr>
</tbody>
</table>

©Howells, Silverman, Kurtz, Davies, Archer 2002
Exercise, RCPCH college tutors informally tested the potential for educational feedback and reliability; here, the number of response categories were optimised. At two further stages, paediatricians from centres involved in this study refined the PCAT’s format into themed domains informed by discrete skills.

**PCAT evaluation**

To evaluate the utility of PCAT, we undertook a study to determine:

- The feasibility of collecting video material
- The reliability of the tool to assess paediatricians’ consultation behaviour
- Construct validity of the tool, by testing whether scores were:
  - Higher for items relating to clinical skills and judgement than communication skills. We hypothesised that a valid assessment tool would demonstrate such differences given that postgraduate paediatric training usually focuses on clinical rather than communication skills.
  - Higher for items relating to doctor–parent interaction compared with those relating to doctor–child interaction. We hypothesised that a valid assessment tool would demonstrate these differences given that none of our clinician samples had received specific child-oriented communication training.
  - Higher for adult-oriented items than respective child items, particularly information-sharing items. We hypothesised that a valid assessment tool would demonstrate these differences given that none of our clinician samples had received specific child-oriented communication training.

Paediatric consultants and specialist registrars recorded videotaped consultations with patients and their families, in accordance with General Medical Council guidelines.20 With...
consent from parents and children, we recorded consultations from each clinician’s out-patient practice, across a wide age (newborn–16 years) and case range.

A second cohort of paediatricians rated the video recordings, having been trained to use PCAT during standardised training sessions lasting 90–120 minutes. Sessions involved familiarisation with PCAT followed by benchmarking between markers using videotapes of consultations. Markers independently rated each videotape in “real time”. They were asked to judge for themselves whether to score items for communication with the child, based on the actual age of the child (supplied) and their assessment of the developmental stage of the child seen on videotape.

Multicentre ethical approval was obtained for this study.

Statistical analysis
For quantitative analysis, scores from PCAT’s 16 items were combined to produce one aggregate score per consultation per assessor (AggregateO). Scores from five items (relationship building, initiating the session, gathering information, explanation/planning and closure with the adult) were combined to produce one “adult-aggregate” score (AggregateA) and six other items (relationship building, initiating the session, gathering information, physical examination, explanation/planning and closure with the child) combined to produce one “child-aggregate” score (AggregateC).

Reliability – the reliability coefficient (R) – the expression of reproducibility of “true” differences in performance between doctors given any consultation when assessed by any assessor – was determined using a generalisability (“G-study”) analysis in SPSS V.13.0. A fully nested design was used: “assessors nested within cases, nested within clinicians”. A “D-study” predicting the “number of consultations required for satisfactory reliability when assessed by any trained assessor” was determined using Microsoft Excel (2000).

Determination of construct validity by comparison of mean scores for groups of items
Principle components analysis was used to determine the relationships between item scores. Two factors accounted for 68% of variance of score: factor 1 (“communication behaviour”,

Figure 3

<table>
<thead>
<tr>
<th>Building the relationship</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-verbal skills: eye contact, open posture, avoids writing / reading notes</td>
<td>Minimal eye contact, obstructive posture, excessive reading or writing notes with no explanation</td>
<td>Minimal eye contact, awkward posture, refers to / writes in notes with no explanation</td>
<td>Quite good eye contact and posture, occasionally inappropriate non-verbal behaviour</td>
<td>Very good non-verbal skills – explains need to use notes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is empathetic and supportive – shows concern, responds to family’s predicament</td>
<td>Ignores patient’s predicament completely, totally unsupportive or rude</td>
<td>Minimal response to parent/child’s predicament, limited support</td>
<td>Warmth towards family but occasionally poor response to predicament</td>
<td>Much warmth and natural empathy throughout consultation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriately engages child from the early stages of consultation</td>
<td>Completely ignores child</td>
<td>Almost completely ignores child or is awkward / inappropriate</td>
<td>Engages child but occasionally inappropriate, insensitive or overpowering</td>
<td>Slightly tries to engage child, adjusting approach to child’s response</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Initiating the session

| Introduces self, clarifies role, determines who is present | Does not establish who is present | Minimal information clarified | Most but not all is clarified eg self and role, but not family members | Explains role and identity, sensitively determines who is present |
| Identifies reasons for the consultation – the doctor’s and family’s | Neither explains purpose for consultation nor checks family’s reasons | Limited coverage of own or family’s reasons for consultation | Explains own reasons for consultation; responds poorly to family’s reasons | Explains own reasons for consultations and acknowledges those of family |
| Screens for other problems and negotiates the agenda for the consultation | No check for other problems and no agenda identified | Screens for but ignores other problems | Screens but no agenda set / only doctors needs are discussed | Screens for and responds to other problems, contracts a clear agenda with family |

Gathering information

| Listens attentively, facilitating verbally and non-verbally | Interrupts all the time or invalidates family’s story or excessive use of notes | Interrupts moderately often or poor facilitation or a little too much use of notes | Moderate facilitation with occasional interruption | Shows interest by verbal and non-verbal facilitation, good use of silence |
| Picks up and responds to verbal and non-verbal cues | No response to obvious verbal non-verbal cues or is rude in response | Minimal response to verbal / non-verbal cues (despite being present) | Picked up and responds to cues but occasionally misses cues | Repeatedly picks up verbal and non-verbal cues, with insensitive response |
| Uses appropriate questioning techniques (eg open–closed questions) | Obstructs patient narrative, exclusive use of closed questions | Too little space for narrative or too many closed too early | Appropriate questioning, a few too many closed questions | Natural sequence of questions, sensitive questioning style for all family members |
| Explores parent/child’s ideas, concerns, feelings, expectations | Does not explore parent/child’s perspective at all, obstinately discussion | Limited or awkward discussion of parent/child’s perspective | Moderate amount of exploration of parent/child’s perspective | Slightly explores and values parent/child’s perspective |
Influence of covariates on score – we collected data about the following attributes:

**Patient attributes**
- Age
- Diagnoses
- Co-attendants

**Consultation attributes**
- Length
- New or follow-up
- Difficulty of consultation (as rated by observing assessor)
- Within first or last three consultations on VHS tape

We used paired and independent t tests to analyse the effect of the attributes on aggregate score, and linear regression analysis to determine the interdependency of covariates affecting score.

---

58% of variance) and factor 2 (“clinical skills”, 10% of variance). Highly correlated items which could be accounted for by either one of the two factors were aggregated.

Paired t tests were used to analyse the difference between the mean factor aggregate scores and between AggregateA and AggregateC.

---

**Table 1**

<table>
<thead>
<tr>
<th>No cases</th>
<th>Overall Reliability</th>
<th>Adult Reliability</th>
<th>Child Reliability</th>
<th>Key</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.62</td>
<td>0.54</td>
<td>0.49</td>
<td>&lt;0.6</td>
</tr>
<tr>
<td>2</td>
<td>0.77</td>
<td>0.70</td>
<td>0.66</td>
<td>&lt;0.6</td>
</tr>
<tr>
<td>3</td>
<td>0.83</td>
<td>0.78</td>
<td>0.74</td>
<td>&lt;0.7</td>
</tr>
<tr>
<td>4</td>
<td>0.87</td>
<td>0.82</td>
<td>0.80</td>
<td>&lt;0.8</td>
</tr>
<tr>
<td>5</td>
<td>0.89</td>
<td>0.85</td>
<td>0.83</td>
<td>&lt;0.9</td>
</tr>
<tr>
<td>6</td>
<td>0.91</td>
<td>0.88</td>
<td>0.85</td>
<td>&lt;0.9</td>
</tr>
<tr>
<td>7</td>
<td>0.92</td>
<td>0.89</td>
<td>0.87</td>
<td></td>
</tr>
</tbody>
</table>

---

**Figure 4**

**Paediatric Consultation Assessment Tool – Marking Key**

<table>
<thead>
<tr>
<th>Process Skills</th>
<th>1</th>
<th>3</th>
<th>5</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Physical examination</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prepares for examination eg by attending to privacy, toys</td>
<td>Does not prepare for examination at all</td>
<td>Minimal preparation for examination</td>
<td>Some preparation for examination</td>
<td>Carefully attends to accessories, seating, others in room, etc</td>
</tr>
<tr>
<td>Maintains rapport with the child throughout the examination</td>
<td>No rapport: is rude or abrupt, or ignores child</td>
<td>Some effort to develop rapport but still awkward</td>
<td>Good rapport but occasional inappropriate behaviour</td>
<td>Sensitively maintains rapport throughout examination process</td>
</tr>
</tbody>
</table>

| **Explanation and Planning** | | | | |
| Tailors amount and type of information for parents and child | No attempt at all to adjust information to parent/child’s needs | Determines but does not respond to starting point, little chunking and checking | Some tailoring/ chunking of info but could still be better | Checks starting point, tailors information well |
| Uses skills which aid recall and understanding | Muddled information, lots of jargon, no checking for understanding | Disorganised information, a little jargon, limited or no checking for understanding | Organised information, a little jargon, some checking for understanding | Well-organised, jargon-free info, repeated checks for understanding asks for restate |
| Incorporates parent/child’s perspective into explanation | Completely ignores or belligerent perspective | Limited incorporation of parent/child’s perspective into explanation | Some incorporation of parent/child’s perspective into explanation | Very good incorporation of parent/child’s perspective into explanation |
| Involves parent/child in decision making | Prescriptive management plan, no consideration of family’s wishes | Describes options but still offers little choice to parent/child | Some involvement of parent/child in decision making | Parent/child fully participant in decision making |

| **Closure** | | | | |
| Establishes and clarifies next steps with parents and child | Does not establish or clarify next steps at all | Some plans re next steps made, but no cross-checking with family | Clear plans made re next steps but not fully cross-checked with family | Carefully explains next steps and cross-checks with family if acceptable and understood |
| Makes contingency plans | No contingency plans or discrepancies requests for contingencies | Limited or transparent contingency plans | Contingency plans made but not clarified with family | Clear contingency plans clarified with family |

| **Structuring** | | | | |
| Uses skills which provide structure (eg summarising and signposting) | Neither of skills used at all, despite opportunity OR chaotic consultation without being patient-centred | Minimal use of summarising and signposting, despite opportunity | Some use of summarising and signposting, could have beneficially used more | Very good (abundant) and appropriate use of summarising and signposting |

---

RESULTS

Descriptive results

Ninety-three point eight per cent of families approached gave consent for video recording within their consultations, and of these 96.1% of recordings were of satisfactory quality (sound, image, completeness). The age range of patients was 5 weeks–15 years + 10 months. Nineteen clinicians recorded a median of 10 (range 8–14) satisfactory quality consultations onto videotape. The median number (range) of clinics needed to acquire the videos was two (1–7) per clinician.

Seventeen clinicians rated a total of 188 video recordings. One hundred and sixty-two consultations were rated in triplicate, 26 twice.

Figure 5 shows the distribution of 538 overall aggregate scores ("AggregateO"), which approximate to a normal distribution. The slight skew towards higher scores was not significantly greater than would be expected by chance. Mean (SE) overall aggregate score was 4.78 (0.05). Individual clinician’s mean (SE) overall aggregate scores ranged from 3.01 (0.12) to 6.65 (0.05).

Reliability analysis

Table 1 illustrates how many cases are needed for a reliable assessment of any clinician, when marked by any one trained assessor. For R>0.7, two cases are needed. For R>0.8, three cases are needed.

Analysis of “adult”-related items and “child”-related items separately gave reliability coefficients of 0.70 and 0.66, respectively. The numbers of cases needed to reliably (R>0.7 or R>0.8, respectively) assess doctors’ behaviour with children and parents separately are three or four, respectively, in each instance.

Tests of construct validity

Clinical and communication skills

The mean (SD) aggregate score (95% CI) for “clinical skills” was 0.520 (0.445 to 0.595) higher than that for “communication behaviour” (p<0.001). There was high correlation between pairs of scores (coefficient 0.73, p<0.001).

Comparison of communication with parents and children

Communication with parents scored more highly than that with children, although the two correlated highly (coefficient 0.76, p<0.001). The mean (SD; 95% CI) difference between AggregateA and AggregateC scores was 0.64 (0.08), p<0.001. When segregated by age, the difference between AggregateA and AggregateC scores was only significant when the patient was aged between 2 and 10 years (post hoc analysis of variance, p<0.001).

Scores were higher for all five specifically adult-oriented items (relationship building, initiating the session, gathering information, explanation and planning, closure) compared with the same child-oriented items. Differences between parent-oriented and child-oriented scores were significantly greater for three information-sharing items (gathering information, explanation and planning, closure), p<0.001 in all cases.

Influence of covariates on score

Patient and consultation attributes

Consultations judged to be of “average” or “difficult” complexity yielded higher aggregate scores (mean (SD; 95% CI) 0.25 (0.19) than consultations considered to be “easy” (p<0.05). This association was accounted for by consultations attended by a non-parent (eg, grandparent, social worker) as well as a parent or guardian (p<0.001). Length of consultation and new/ follow-up status did not significantly influence score. Scores were no higher for the first or last three consultations recorded.

DISCUSSION

Reliability

The PCAT offers the opportunity to assess doctors’ consultation skills in the triadic setting by direct observation. Despite the complexity of triadic interactions, the tool appears highly reliable, requiring a very small number of cases for summative assessment of a doctor’s consultation skills. This is true not only for assessment of overall communication performance (where three cases are needed for summative assessment), but also for that of communication with the child or parent alone (where four cases are needed). As the PCAT has been able to discriminate between performances of volunteer clinicians, it is likely to be able to do so among a random sample of doctors, to identify those who communicate poorly.

Feasibility

The high degree of reliability makes the assessment tool more feasible to administer than originally anticipated. Assuming a mean length of consultation of 15 minutes, assessors would need to observe 30–45 minutes of video-recorded material per trainee, or slightly more where focused evaluation of communication with just parent or child was required. Given high levels of consent and satisfactory recordings, as with our experience, it should be feasible to acquire sufficient video material from one or two outpatient clinics. Training time of 90–120 minutes adds to the practicability of implementing PCAT using clinicians as assessors. The use of web-based video streaming and online forms accessible to assessors would also improve feasibility and is already being utilised in other video-based assessments.

Validity

PCAT’s content validity has been assured by its design using evidence-based communication theory. The tool has also demonstrated construct validity, scores being higher for clinical than communication skills, which is not surprising given that only two clinicians had received specific training in doctor-patient communication and no training in paediatric communication. The finding that scores were highest for communication with “parents” rather than children themselves, particularly for information-sharing items, also adds to construct validity. Observational research has shown children’s
professionals in primary care. 

Further evaluation

PCAT’s design, with space for free-text comments and documentation of what has been seen and heard, makes it ideal for formative assessment where feedback can be used to guide future training and learning. We have not explored the educational impact of PCAT in this study, but this should be carried out when the assessment tool is incorporated into a wider assessment programme for paediatric trainees such as that organised by the RCPCH. PCAT’s value within such a programme is likely to be one of exploration of specific communication difficulties, where these have been suspected by other methods such as multisource feedback.

Implications for assessment

What sort of cases should be recorded?

To assess the breadth of a clinician’s performance, assessment should include recordings which capture children of a range of ages including those able to contribute to “medical” information-sharing parts of the encounter, and extended family members/carers.

Should clinicians themselves be allowed to select which videotaped consultations are assessed?

Our data suggest that performance does not vary hugely between individual consultations. Similarly, multisource feedback scores appear to be unaffected by case/respondent selection by the assessor. Thus clinicians should be allowed to choose their own consultations to submit for assessment.

Does it matter whether the first videotaped consultations are assessed relative to a later sample?

Corroborating with other researchers’ findings that doctors’ performance is not affected by the presence of video cameras for more than two or three consultations, this study found no significant difference between scores for the first and last three consultations recorded. Therefore assessors should be allowed to record consultations that represent cases from their usual practice, and be allowed to submit any of these, including the first 2–3 recorded, for evaluation.

Implications for training

It is not surprising, given other observational research, that this research shows clinicians’ behaviour with children to be different from that with their parents. Children resent being left out of discussion about their illness but respond well to information which is specifically tailored to them. Therefore, although training in paediatric consultation skills should encompass rapport building with children and communication with their parents, more emphasis should be given to information sharing with children themselves. Training needs to be focused on interacting with pre-adolescent children, as well as teenagers.

Acknowledgements

Grateful thanks to families involved in the study and clinicians who were involved in collecting and assessing video material.

Funding

RH was supported by a grant donated to the Royal College of Paediatrics and Child Health by WellChild, and by the Department of Paediatrics, University of Cambridge. RH sought multicentre ethical approval, collected video material, trained assessors and undertook data analysis. HD, JS and JA supported development of the PCAT and study methodology. AM recruited clinicians/assessors.

Competing interests

None.

Ethics approval

Multicentre ethical approval was obtained for this study.

Patient consent

Obtained.

Provenance and peer review

Not commissioned; externally peer reviewed.

REFERENCES