

The Child Attention-Deficit Hyperactivity Disorder Teacher Telephone Interview (CHATTI): reliability and validity

JANE HOLMES, DEBORAH LAWSON, KATE LANGLEY, HELEN FITZPATRICK, ANNE TRUMPER, HELEN PAY, RICHARD HARRINGTON and ANITA THAPAR

Background The ICD–10 and DSM–IV diagnostic criteria for hyperkinetic disorder and attention-deficit hyperactivity disorder (ADHD) require symptoms or impairment in two or more settings. Thus, information on children's symptoms in school is usually required. This paper presents the Child ADHD Teacher Telephone Interview (CHATTI), an instrument aimed at systematically obtaining this information.

Aims To examine the stability, test–retest reliability and criterion validity of the CHATTI for children referred with a suspected diagnosis of ADHD.

Method Data were obtained from 79 teachers, of whom 36 were interviewed on two occasions.

Results Overall, the CHATTI shows good stability, test–retest reliability and criterion validity for symptom scores. Test–retest reliability for some individual items was low. Reliability for the operationalised criteria of 'pervasiveness' (i.e. symptoms at school and home) and 'school impairment' was excellent ($\kappa=1$).

Conclusions The CHATTI appears to be a promising tool for assessing ADHD symptoms in a school setting and could be useful in clinical as well as research settings.

Declaration of interest Work funded by the Wellcome Trust, Action Research and Sparks.

The ICD–10 (World Health Organization, 1992) and DSM–IV (American Psychiatric Association, 1994) diagnostic criteria for hyperkinetic disorder and attention-deficit hyperactivity disorder (ADHD) require symptoms or impairment in two or more settings. There is no clear consensus as to how these criteria are best assessed, although the use of multiple informants is advocated (Barkley, 1998; Taylor *et al*, 1991; Simonoff *et al*, 1998). Teacher rating scales are most commonly used but may be subject to limitations, such as rater errors and poor response rates (Conners *et al*, 1998), and children who score above an accepted cut-off do not necessarily show disorder (Taylor, 1994). Telephone interviews with teachers provide an alternative strategy. The aim of this paper is to describe the Child ADHD Teacher Telephone Interview (CHATTI), designed to assess systematically the hyperactive, inattentive and impulsive symptoms and impairment in a school setting. We present data on the stability, test–retest reliability and criterion validity of this measure in a clinical sample of children with suspected ADHD.

METHOD

Ascertainment of sample

The sample described in this paper is part of a larger sample from an ongoing genetic study consisting of children with a suspected diagnosis of ADHD referred to district child and adolescent psychiatry and paediatric clinics in South Wales, the south-west of England, Greater Manchester and Cheshire. Children with full-scale IQ test scores of below 70 (assessed using the Wechsler Intelligence Scale for Children – version III, WISC–III; Wechsler, 1992), major medical or neurological conditions, Tourette syndrome or pervasive developmental disorder were excluded. Each child was assessed comprehensively (see Holmes

et al, 2000 for a full description) with clinical information obtained using the parent version of the Child and Adolescent Psychiatric Assessment (CAPA; Angold *et al*, 1995). Diagnoses were assigned according to ICD–10, DSM–IV and DSM–III–R (American Psychiatric Association, 1987) classification systems.

Procedure and sample characteristics

The parents of each child were asked for written parental consent to contact the school. This consent form was then posted to the child's school, together with a study information sheet, teacher consent form (giving us permission to contact them by telephone to administer the CHATTI) and a short questionnaire package that contained the Abbreviated Conners Teacher Ratings Scale (ACTRS; Conners, 1973), which at the time of the study was one of the rating scales most commonly used by clinicians in the UK, and the DuPaul ADHD rating scale (DuPaul, 1981). A reminder letter was sent to teachers who failed to respond to the first mailshot. The return rate for the questionnaires was 94% (79/84). Data from the teachers of this sample of 79 children (73 males; 6 females) aged between 6 and 13 years (mean=8.76; s.d.=1.75) were used for the purposes of assessing the criterion validity of the CHATTI.

Once the consent form and questionnaire package had been returned, the researcher telephoned the teacher to carry out the CHATTI. The interview and questionnaires were always completed by the same teacher, typically the child's class teacher because the majority of children who participated in the study were in primary school. However, for those children in secondary school the teacher who had the most extensive knowledge of the child's behaviour was chosen to complete the interview and questionnaires. Phase 1 of the study involved interviews with 79 teachers, of whom 20 were reinterviewed one week later by the same experienced interviewer (a research psychologist (J.H.) and two child psychiatrists (A.Tr., H.F.), who had all been trained previously to use a research diagnostic interview). Phase 2 of the study involved a further 16 interviews (new sample) undertaken with the same teacher 1 week apart by two different interviewers (one of whom had been trained in the CAPA (D.L.) and the other who was an

assistant psychologist (H.P.) who had not received training in diagnostic interviews).

All teachers who consented to the study (94%) were interviewed. The interviews took place during regular school hours. The children in these groups were aged between 6 and 13 years (mean=9.13, s.d.=1.7; 32 males, 4 females) and fulfilled either DSM-III-R or DSM-IV criteria for ADHD or ICD-10 criteria for hyperkinetic disorder.

Measures

The CHATTI is a structured interview that takes 15–20 min to complete and should be conducted with the teacher who has the most extensive knowledge of the child's behaviour (class teacher for primary school children). The interview focuses on the occurrence of ADHD symptoms during the preceding 3 months. The CHATTI contains 18 items included in the ICD-10 criteria for hyperkinetic disorder and DSM-IV diagnostic criteria for ADHD.

The CHATTI is divided into three overall symptom areas: inattention (e.g. 'difficulty concentrating'); hyperactivity (e.g. 'fidgets or squirms in seat'); impulsiveness (e.g. 'interrupts or intrudes on others'). Each symptom is explicitly defined and teachers are asked whether the index child shows the symptom in different lessons and to a greater extent than other children of his/her developmental age. In addition to the 18 ADHD items, the CHATTI includes a question on whether the ADHD symptoms cause significant impairment in the child's social or academic functioning at school.

Symptom scores were summed to provide continuous measures, a total ADHD score and scores from the three sub-scales, namely inattention, hyperactivity and impulsiveness.

In this study the interview was developed to be used in conjunction with the parent version of CAPA (Angold *et al*, 1995) but could be used with other similar diagnostic interviews. Again in this study, diagnoses were primarily based on parent-derived interview data with the teacher information used *only* to define the ICD-10 and DSM-IV criterion of pervasiveness (i.e. symptoms at school as well as at home), but the instrument could be used differently. At the start of the study, the criterion of 'symptom pervasiveness' for ICD-10 hyperkinetic disorder was operationally defined by a consensus of

experienced child psychiatrists as the presence of at least one definite symptom from each of the symptom areas (i.e. inattention, hyperactivity, impulsiveness) reported by the teacher, with associated impairment in functioning in school in addition to meeting the diagnostic criteria at home using parental interviews. For DSM-IV ADHD, the criterion of 'some impairment from the symptoms is present in two or more settings' was rated using the response to the question on impairment of functioning.

Statistics

Symptom scores (categorical data) and scale scores (continuous data) were generated from the interview data. For the reliability analysis, Cohen's κ (Cohen, 1960) was used to assess agreement on categorical variables, whereas the scale score agreement was measured by the intraclass correlation coefficient (ICC) (Everitt, 1996). Criterion validity was assessed by investigating the association between scores on the CHATTI and ACTRS (Conners, 1973) and the DuPaul ADHD scale (DuPaul, 1981). Internal consistency also was checked using Cronbach's α coefficient. Within-subject associations of measures were investigated using Spearman's correlations, because questionnaire scores from the DuPaul rating scales and ACTRS were negatively skewed.

All statistical tests were considered significant at $P < 0.05$. Two-tailed P values are presented. Statistical analyses were carried out using the Statistical Package for the Social Sciences, Windows version 7.5 (SPSSW; SPSS Inc).

RESULTS

Criterion validity

Mean scores (and standard deviations) for all 79 teachers were as follows: total CHATTI, 10.60 (5.12); ACTRS, 19.11 (6.70); DuPaul ADHD total scale, 37.24 (11.80). Total CHATTI scores were strongly correlated with total scores on the DuPaul ADHD rating scale ($r=0.65$, $P < 0.01$) and moderately correlated with scores on the ACTRS ($r=0.46$, $P < 0.01$).

Stability of the CHATTI across time – same interviewer

Table 1 shows the agreement across time for scores on the CHATTI total scale and sub-scales. The ICCs ranged from 0.94 to

0.98. Kappa coefficients were then calculated to assess agreement between time 1 and time 2 for each individual item and for the criterion of 'pervasiveness'. Table 2 shows the stability of the criteria of 'pervasiveness' (1.0) and 'impairment from symptoms' (1.0) necessary to make a diagnosis of ICD-10 hyperkinetic disorder and DSM-IV ADHD, respectively. According to the benchmarks provided by Landis & Koch (1977), the strength of agreement for CHATTI items ranged from fair (0.35 for 'avoids tasks') to perfect (1.0 for 'can't wait turn'). Cronbach's α was 0.91 for internal consistency. Some researchers operationalise the ICD-10 criterion of pervasiveness as requiring that the full ICD-10 criteria for hyperkinetic disorder (i.e. at least six symptoms of inattention, three symptoms of hyperactivity and one symptom of impulsivity) are met at school (as well as at home). Therefore we also examined the stability of this stricter definition of 'pervasiveness'. The κ coefficient was 0.79.

Test–retest reliability across time – two different interviewers

Intraclass correlations for total and sub-scores for different raters across time are shown in Table 1. These range from 0.76 to 0.92. Table 3 shows that agreement across time for the criteria of 'impairment' (1.0) and 'pervasiveness' was perfect (1.0). However, κ coefficients for individual items were very variable (ranging from 0.16 for 'avoids tasks' to 0.87 for 'constantly on the go'). The test–retest reliability for the stricter definition of 'pervasiveness' was 0.71.

Table 1 Stability across time for the scores on the Child Attention-Deficit Hyperactivity Disorder Teacher Telephone Interview (CHATTI) (total and sub-scales) in terms of the intraclass correlation coefficient

CHATTI scales	One Interviewer ($n=20$)	Two interviewers ($n=16$)
Hyperactive	0.94***	0.92***
Inattentive	0.95***	0.76*
Impulsive	0.94***	0.82***
Total	0.98***	0.92***

* $P < 0.05$; *** $P < 0.001$.

Table 2 Stability across time for the Child Attention-Deficit Hyperactivity Disorder Teacher Telephone Interview (CHATTI) items – same interviewer ($n=20$)

CHATTI items	Cohen's κ^1 ($n=36$)	s.e.	Strength of agreement ²
Impairment	1.00	–	Very good
Symptom pervasiveness	1.00	–	Very good
Pervasiveness with full ICD–10 symptoms reported by teacher	0.79	0.14	Good
Fidgets	1.00	–	Very good
Stays seated	0.80	0.13	Good
Rushes about	0.58	0.19	Moderate
Noisy	0.41	0.20	Moderate
On the go	1.00	–	Very good
Can't concentrate	0.57	0.22	Moderate
Poor organisation	0.66	0.18	Good
Loses things	0.70	0.16	Good
Forgetful	0.40	0.20	Fair
Poor attention to details/careless mistakes	0.66	0.18	Good
Doesn't listen	0.52	0.16	Moderate
Easily distracted	0.76	0.15	Good
Following instructions	0.69	0.16	Good
Avoids tasks	0.35	0.21	Fair
Interrupts	0.89	0.11	Very good
Blurts out	0.60	0.18	Moderate
Talks excessively	0.60	0.16	Moderate
Can't wait turn	1.00	–	Very good

1. Kappa coefficient for binary data (Cohen, 1960).

2. Strength of agreement according to parameters set by Landis & Koch (1977).

Table 3 Agreement across time for the Child Attention-Deficit Hyperactivity Disorder Teacher Telephone Interview (CHATTI) items – two different interviewers

CHATTI items	Cohen's κ^1 ($n=36$)	s.e.	Strength of agreement ²
Impairment	1.00	–	Very good
Symptom pervasiveness	1.00	–	Very good
Pervasiveness with full ICD–10 symptoms reported by teacher	0.71	0.18	Good
Fidgets	0.70	0.18	Good
Stays seated	0.43	0.23	Moderate
Rushes about	0.38	0.22	Fair
Noisy	0.46	0.19	Moderate
On the go	0.87	0.13	Very good
Can't concentrate	0.47	0.16	Moderate
Poor organisation	0.33	0.23	Fair
Loses things	0.53	0.18	Moderate
Forgetful	0.42	0.25	Moderate
Poor attention to details/careless mistakes	0.60	0.24	Moderate
Doesn't listen	0.26	0.20	Fair
Easily distracted	0.28	0.24	Fair
Following instructions	0.57	0.22	Moderate
Avoids tasks	0.22	0.18	Fair
Interrupts	0.39	0.21	Fair
Blurts out	0.57	0.22	Moderate
Talks excessively	0.34	0.23	Fair
Can't wait turn	0.38	0.25	Fair

1. Kappa coefficient for binary data (Cohen, 1960).

2. Strength of agreement according to parameters set by Landis & Koch (1977).

DISCUSSION

Although telephone interviews have been used previously to assess psychopathology (Rohde *et al*, 1997), including symptoms of ADHD (Nadder *et al*, 1998), these have nearly all been designed to be used with the subject or the parent. Nadder *et al* (1998) developed a brief telephone survey for the assessment of ADHD and oppositional defiant disorder/compulsive disorder symptoms displayed at home, for twins aged 7–13 years, participating in the Virginia Twin Study of Adolescent Behavioural Development. This survey contained ten items coded as absent or present, of which six items related to ADHD symptomatology. Correlations between the summed ADHD items and maternal ratings on the Child Behavior Checklist hyperactivity sub-scale (Achenbach, 1991) were 0.67 and 0.61 for males and female twins, respectively, indicating good criterion validity for this instrument. However, this instrument was developed for the assessment of ADHD symptoms at home within a population-based sample. Furthermore, this instrument covers 6/18 items of DSM–IV and ICD–10 diagnostic criteria and was not intended to be a diagnostic instrument.

Although gathering information from teachers by telephone may occur commonly in clinical situations, to our knowledge there has been only one published report describing the use of a structured teacher telephone interview as an adjunct for making the diagnosis of ADHD for research purposes (Tannock *et al*, 2000). Although psychometric data have not been published, this instrument has been found to be a useful adjunct to parent interviews (R. Tannock, personal communication, 2003).

Although it is usual to consider interview methods as the gold standard for assessing psychopathology it is still essential to assess the reliability, validity and acceptability of a new interview-based instrument. We sought to examine criterion validity using two commonly used questionnaires. The CHATTI was found to be strongly correlated with the ADHD rating scale (DuPaul, 1981) and moderately correlated with the ACTRS (Conners, 1973). The observed strong correlations between the CHATTI and the DuPaul ADHD rating scale are not surprising, given that both measures include the DSM–III–R symptoms of ADHD.

The CHATTI was found to yield highly consistent results across a 1-week test-retest period. Specifically, the CHATTI showed excellent reliability for our operationalised definition of symptom pervasiveness ($\kappa=1.00$) and high stability for total and sub-scale scores based on symptom counts and the stricter definition of 'pervasiveness'. Test-retest reliability for total CHATTI scores (0.98) are similar or higher than those for the DuPaul ADHD scale (0.96) and ACTRS (0.7–0.90). Moreover, the CHATTI showed acceptable levels of stability even at the individual symptom level. Kappa coefficients for the majority of individual symptoms ranged from moderate to perfect agreement.

Test-retest reliability over time with two different interviewers was also examined. Here, the ICCs for symptom scores were still high and reliability for the categories of 'symptom pervasiveness' and 'impairment' was perfect and for the stricter definition of 'pervasiveness' was good. Nevertheless, reliability for individual items was highly variable and for some items it was low. We conclude that one contributory factor to this may have been the choice of interviewers. One of the two interviewers was a trained interviewer whereas the other was a psychology assistant who had not been trained in research diagnostic interviews. However, this approach was adopted not only for practical reasons (availability of researcher time) but also to consider whether it would be feasible for a clinician untrained in research diagnostic interviews to use this instrument.

Overall initial findings suggest that the CHATTI is cost- and time-efficient and acceptable to teachers. It provides a highly stable measure of symptom pervasiveness and teacher-reported total ADHD symptom scores and impairment at school. Although questionnaires are easy to use and cheap to administer, and many of them show high reliability, they can be inaccurate at identifying individuals as hyperactive and can be subject to rater biases and poor response rates (Taylor, 1994; Conners *et al*, 1998). Moreover, it is not clear how to integrate questionnaire-derived data with parent interviews to generate the criteria of 'symptom pervasiveness' or 'impairment in two or more settings' reliably. The CHATTI represents an attractive alternative to teacher questionnaires, particularly when a systematic method is required to be used in conjunction with a standard parent diagnostic

CLINICAL IMPLICATIONS

- The Child Attention-Deficit Hyperactivity Disorder Teacher Telephone Interview (CHATTI) is a teacher telephone interview that can be used to assess attention-deficit hyperactivity disorder (ADHD) symptoms in school.
- The CHATTI appears to show acceptable reliability and stability in clinical samples with suspected ADHD.
- The CHATTI is short, easy to use and acceptable to teachers.

LIMITATIONS

- The instrument was tested in a clinical sample of children with suspected ADHD.
- Patients were referred to a study of ADHD.
- Most of the interviewers were trained to use research diagnostic interviews.

JANE HOLMES, BA, DEBORAH LAWSON, BA, KATE LANGLEY, BA, Section of Child and Adolescent Psychiatry, Department of Psychological Medicine, University of Wales College of Medicine, Cardiff; HELEN FITZPATRICK, MRCPsych, ANNE TRUMPER, MRCPsych, University of Manchester Department of Child and Adolescent Psychiatry, Royal Manchester Children's Hospital, Manchester; HELEN PAY, BSc, Section of Child and Adolescent Psychiatry, Department of Psychological Medicine, University of Wales College of Medicine, Cardiff; RICHARD HARRINGTON, FRCPsych, University of Manchester Department of Child and Adolescent Psychiatry, Royal Manchester Children's Hospital, Manchester; ANITA THAPAR, MRCPsych, Section of Child and Adolescent Psychiatry, Department of Psychological Medicine, University of Wales College of Medicine, Cardiff

Correspondence: Anita Thapar, Section of Child and Adolescent Psychiatry, Department of Psychological Medicine, University of Wales College of Medicine, Heath Park, Cardiff CF14 4XN, UK

(First received 14 January 2003, final revision 28 May 2003, accepted 12 August 2003)

interview for assigning the diagnosis of ADHD or hyperkinetic disorder. The CHATTI also provides an alternative means of assessing symptoms in studies focusing on teacher-reported ADHD symptoms and in clinical settings. Indeed, it can be argued that in clinical settings, for children with suspected ADHD, early clinician contact with schools by telephone rather than by letter is highly desirable for assessment and treatment purposes.

One limitation of this study is that data were collected from a clinic sample of children with suspected ADHD, nearly all of whom fulfilled the diagnostic criteria for hyperkinetic disorder or ADHD. Diagnostic severity may influence the measurement of reliability, with reliability coefficients being higher in more severely affected groups (Jensen *et al*, 1995). Thus, it is important to examine the psychometric properties of the CHATTI

within a non-clinic sample and in children with other diagnoses before it can be recommended for widespread use in other populations. Further research also will be necessary to investigate the discriminant validity of the CHATTI to differentiate children with ADHD from other clinic groups, such as those with oppositional defiant disorder, anxiety and depression. However, we suggest that it is most useful when used as an adjunct to parent interviews to assess the presence of ADHD symptoms or impairment in more than one setting rather than as a diagnostic tool in itself.

In summary, with the advent of ICD-10 and DSM-IV, clinicians and researchers are required to assess the presence of hyperactive, impulsive and inattentive symptoms or impairment across settings, in order to determine a diagnosis of hyperkinetic disorder or ADHD. Research findings also

suggest the importance of using multiple informants for the diagnosis of hyperactivity in order to reduce rater biases and discrepancies between parent and teacher ratings of hyperactivity (Simonoff *et al*, 1998; Mitsis *et al*, 2000). The CHATTI is a new instrument designed for the assessment of ADHD symptomatology within school settings. Preliminary data suggest that the CHATTI shows acceptable reliability in clinical samples with suspected ADHD, it is easy to use and it is acceptable to teachers.

ACKNOWLEDGEMENTS

We thank the families and teachers who participated in this study.

REFERENCES

- Achenbach, T. M. (1991)** *Manual for the Child Behavior Checklist/4–18 and 1991 Profile*. Burlington, VT: Department of Psychiatry, University of Vermont.
- American Psychiatric Association (1987)** *Diagnostic and Statistical Manual of Mental Disorders* (3rd edn, revised) (DSM–III–R). Washington, DC: APA.
- American Psychiatric Association (1994)** *Diagnostic and Statistical Manual of Mental Disorders* (4th edn) (DSM–IV). Washington, DC: APA.
- Angold, A., Prendergast, M., Cox, A., et al (1995)** The Child and Adolescent Psychiatric Assessment. *Psychological Medicine*, **25**, 739–753.
- Barkley, R. A. (1998)** *Attention-Deficit Hyperactivity Disorder* (2nd edn). New York: Guilford Press.
- Cohen, J. (1960)** A coefficient of agreement for nominal scales. *Educational Psychological Measures*, **20**, 37–46.
- Conners, C. (1973)** Psychological assessment of children with minimal brain dysfunction. *Annals of the New York Academy of Sciences*, **205**, 283–302.
- Conners, C., Sitarenios, G., Parker, J. D., et al (1998)** Revision and restandardization of the Conners Teacher Ratings Scale (CTRS–R): factor structure, reliability and criterion validity. *Journal of Abnormal Child Psychology*, **26**, 279–291.
- DuPaul, G. (1981)** Parent and teacher ratings of ADHD symptoms: psychometric properties in a community based sample. *Journal of Clinical Child Psychology*, **20**, 245–253.
- Everitt, B. S. (1996)** *Making Sense of Statistics in Psychology: A Second-level Course*. New York: Oxford University Press.
- Holmes, J., Payton, A., Barrett, J. H., et al (2000)** A family-based and case–control association study of the dopamine D4 receptor gene and dopamine transporter gene in attention deficit hyperactivity disorder. *Molecular Psychiatry*, **5**, 523–530.
- Jensen, P., Roper, M., Fisher, P., et al (1995)** Test–retest reliability of the Diagnostic Interview Schedule for Children (DISC 2.1). Parent, child, and combined algorithms. *Archives of General Psychiatry*, **52**, 61–71.
- Landis, J. R. & Koch, G. G. (1977)** The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159–174.
- Mitsis, J. M., McKay, K. E., Schulz, K. P., et al (2000)** Parent–teacher concordance for DSM–IV attention-deficit/hyperactivity disorder in a clinic-referred sample. *Journal of the American Academy of Child & Adolescent Psychiatry*, **39**, 308–313.
- Nadder, N. S., Silberg, J., Eaves, L., et al (1998)** Genetic effects on ADHD symptomatology in 7-to-13-year-old twins: results from a telephone survey (1998). *Behavior Genetics*, **28**, 83–99.
- Rohde, P., Lewinsohn, P. M. & Seeley, J. R. (1997)** Comparability of telephone and face-to-face interviews in assessing axis I and II disorders. *American Journal of Psychiatry*, **154**, 1593–1598.
- Simonoff, E., Pickles, A., Hervas, A., et al (1998)** Genetic influences on childhood hyperactivity: contrast effects imply parental rating bias, not sibling interaction. *Psychological Medicine*, **28**, 825–837.
- Tannock, R., Hum, M., Maselli, M., et al (2000)** *Teacher Telephone Interview*. Toronto: The Hospital for Sick Children.
- Taylor, E. (1994)** Syndromes of attention deficit and overactivity. In *Child and Adolescent Psychiatry* (eds M. Rutter, E. Taylor & L. Hersov), vol. 3, pp. 285–307. Oxford: Blackwell Scientific Publications.
- Taylor, E., Sandberg, S., Thorley, G., et al (1991)** *The Epidemiology of Childhood Hyperactivity*. New York: Oxford University Press.
- Wechsler, D. (1992)** *Wechsler Intelligence Scale for Children: UK* (3rd edn). London: Psychological Corporation.
- World Health Organization (1992)** *The ICD–10 Classification of Mental and Behavioural Disorders*. Geneva: WHO.