Assessing Emotional and Behavioural Problems in Children: Factors Associated with Multiple Informant Consistency in New Zealand

Heather Clay, *Caterbury District Health Board*

Lois J. Surgenor & Christopher M. Frampton, *University of Otago*

Behaviour checklists using multiple informants are commonly used in the assessment of childhood behavioural and emotional problems. Characteristics of both the informants and children may effect such ratings, and in turn, this may effect inter-rater consistency and the confidence by which accurate conclusions can be drawn. This study investigates the extent to which selected demographic and clinical factors affect the consistency of ratings between informants on the commonly used Child Behaviour Checklist and the Teacher Report Form. Data were collated from a cohort of 200 cases aged 6-10 years referred to a New Zealand specialist outpatient clinic. Significant effects were found in relation to two problem areas. First, in regard to anxiety/depression problems, informants reported significantly greater consistency of agreement when the identified case was male, or when both informants were female. Second, in regard to the rating of thought problems, informants reported significantly greater consistency of agreement when initial assessment suggested the presence of a psychiatric diagnosis. Both the observation and expression of anxiety/depressive symptoms appear particularly affected by gender issues. These results reinforce the need for multiple sources of information when assessing emotional and behavioural problems in children.

Clinicians assessing the presence and extent of childhood emotional and behavioural disorders often employ a two-stage procedure to identify problematic conditions. In conjunction with detailed initial interviews with the child and relevant adults, behaviour surveys/checklists are frequently used as an economical way to survey multiple domains of functioning and help evolve clinical formulations. Such checklists are also used to obtain 'cut off' scores enabling children to be categorised as potential 'cases' for research or clinical purposes, and to assess meaningful changes in psychopathology arising from treatment (Visser et al., 2003). Surveys or checklists are typically completed by parents, teachers, or children, or any combination of such informants. Indeed, capturing information on children's specific behaviour across and within situations is considered important (American Psychiatric Association, 2000).

An important aspect of the assessment, treatment and research of children's behavioural or emotional problems is deciding who should rate such behaviour. However different informants differ with respect to their sensitivity, perception, and relationship to aspects of a child's behaviour. As such, gathering information from a number of sources poses some unique problems regarding the extent to which accurate conclusions can be drawn for diagnostic or research purposes. This is an issue for both 'clinical' and 'non-clinical' populations, and for this reason, the issue of inter-rater consistency has long been discussed, especially in regard to one of the most common rating scales – the Child Behaviour Checklist (CBCL) and it's parallel informant supplements (Achenbach, 1991a, , 1991b; Achenbach & Edelbrock, 1991). For example, Rutter, Tizard, Yule, Graham and Whitmore (1977) found that mothers and teachers agreed on the presence of behaviour disorders in only 7% of the disturbed children. Other studies echo concerns about poor correlations (Fergusson & Horwood, 1987a, 1987b, 1989), leading some to conclude that clinicians simply should expect that significant disagreement between informants will be the norm (Lee et al., 1994). The implications of this extend beyond clinical situations where assessment and treatment may be

the focus. Sawyer, Baghurst and Clark (1992) noted that differences between informant reports could seriously distort prevalence estimates in epidemiological studies.

Low agreement between different informants for any measures may be interpreted negatively and thereby cast doubt on one or both informants. Another interpretation is that, rather than reports being invalid, different informants contribute different information simply on the basis that symptoms or behaviours vary from one situation to another (Achenbach, McConaughy et al., 1987). That is, different inter-informant correlations may reflect disagreement and the influence of informant characteristics, but also real situational variability of children's behaviour. Likewise, different types of informants may systematically and habitually report different rates of concern, irrespective of the problem at hand. For example, Sawyer et al. (1992) noted that children reported the most problems, teachers reported the fewest, and parents reported a frequency of problems between the two. This was regardless of whether the focus was on externalising (e.g. aggressive and delinquent behaviours) or internalising problems (e.g. withdrawn, somatic complaints and anxious/depressed behaviours). Similarly, Verhulst, Koot and Van der Ende (1994) found that, while there was significant stability of ratings obtained from the same individual (e.g., the same parent on different occasions), parents emphasised aggressive behaviour while teachers highlighted problems relating to cognition and withdrawal, and the children themselves emphasised socialisation problems.

Several other demographic and clinical variables have been investigated with respect to inter-rater consistency on the CBCL and its supplements. Gender stereotyping may also affect the way informants perceive behaviour, or even how a child behaves. In one study teacher informants scored boys significantly higher than girls on four syndromes (Weine et al., 1995). While this suggests that both the child gender and observable clinical features can affect rates of reported problematic behaviour, no comment was made in respect to caregiver informant gender. Jensen, Traylor, Xenakis and Davis (1988) found that fathers and mothers differed in ratings of behavioural problems in sons verses daughters. This study also noted that agreement was enhanced between raters particularly in regard to external and observable behaviours that violated social role expectations. It was suggested internalising symptoms in girls might be more difficult to register. Likewise, aggression may be less tolerated in girls based on expectations of appropriate gender-specific behaviours. However, these findings have not been reported consistently. Stanger and Lewis (1993) found that gender of the informant or child had no significant effect on inter-informant reliability, whereas Seiffge-Krenke and Kollmar (1998) found that both mothers and fathers showed greater agreement with their daughters than with their sons. Taken together, these findings raise the possibility that interrater reliability could be influenced by both the gender of the adult informant and the child, irrespective of whether this adult was a parent or teacher.

Age and ethnicity have also been considered, albeit in a limited way. While the expression of psychopathology varies considerably over the period of childhood, much less has been noted specifically about the relationship between the age of a child and the level of agreement between informants. Achenbach, Verhulst, Baron and Akkerhuis (1987) noted that informant disagreement increased with the age of the child, but this age effect has not been consistently found. Indeed, Seiffge-Krenke and Kollmar (1998) found increasing inter-parental agreement over a period of four years. Measures such as the CBCL and its supplements are valid and reliable cross-culturally (Davies & McKelvey, 1998; Yang et al., 2000), but whether ethnicity itself systematically influences inter-rater consistency is unclear. There appears to be considerable scope for further investigations to include both age and ethnicity as factors likely to influence multiple informant agreement.

The type of symptoms or behaviours being rated has been associated with issues of inter-rater agreement. Stanger and Lewis (1993) and Verhulst et al. (1994) both found that adults are more likely to agree when the problem is one where the child was in overt conflict with others. Whereas when defining problematic behaviours is more reliant on judgement of functioning (internalising behaviours), there is more likelihood of disagreement. In understanding this, it has been suggested that externalising behaviours may be more consistently exhibited across settings, whereas children "may display internalising behaviours in the presence of their parents, but not their teachers. Externalising behaviours may also be more readily observable and more likely to cause management problems, which may make them more salient to external observers" (Stanger & Lewis, 1993, p/113). Kazdin, Esveldt-Dawson and Loar (1983) examined the correspondence of teacher ratings and direct observations of classroom behaviour of children in a psychiatric inpatient service. Rater-observer correlations between the CBCL and the teacher informant parallel measure (Teacher Informant Form, TRF) were consistently higher (r = .67) than teacher-observer correlations (r = .23). Danforth and DuPaul (1996) also highlighted the importance of the frequency and type of interactions informants have with clients in determining the level of agreement between different observers/raters. Clearly teachers report different behaviours to parents although both show greatest agreement regarding externalising behaviours than internalising behaviours. It is suggested that this is due in part to externalising behaviour being more likely to evoke a classroom management problem. Little has been reported about the effect of the teacher's gender on the rating of children's problem behaviour and whether the gender of the child has any effect on this. Given that the gender of the child is significant in the level of agreement between other caregiver informants, teacher gender as a variable warrants further investigation.

In summary, rating scales such as the CBCL and its supplements have been popular for both research and clinical purposes and have multiple applications including epidemiological comparisons of populations, research into specific demographic or diagnostic groups, and for use in

clinical assessment for diagnostic and treatment purposes. Research using these tools has consistently highlighted the problem of inter-rater reliability. Factors hypothesised to affect inter-rater agreement can generally be divided into the child variables (age, gender, and clinical features) and informant variables (parent/teacher/case). Very little work has included ethnicity, and clinical features of the child have commonly been grouped into the two broad groups of internalising and externalising problems, with externalising problems generally showing greater inter-rater consistency between parent/teacher informants. In the context of few local studies, and the failure of wider studies to adequately explain informant discrepancies to date (De Los Reyes & Kazdin, 2005), there appears to be considerable scope for detailed investigations into what factors may be involved in inter-rater consistency in the New Zealand setting.

In the context of a typical clinical setting, the aim of this study is to investigate the extent to which routinely known demographic and clinical variables contribute to identified discrepancies between informants on the commonly used CBCL and TRF rating scales.

Method

Participants and Setting

The Christchurch Child and Family Speciality Service (CFSS) is a psychiatric outpatient service in the community for children from 0-13 years servicing a primarily urban population of approximately 350,000 people. It is funded by central government and clients are usually referred by general practitioners, paediatricians or education services. Services provided include assessment and treatment of moderate to severe emotional, behavioural or psychiatric problems. All patients undergo an initial comprehensive screening interview during which DSM-IV based diagnoses (American Psychiatric Association, 2000) are considered. The education institution, (e.g., kindergarten or school) of every client is contacted immediately following the initial assessment for information regarding the child's learning, socialisation and behaviour in school. This multi-axial assessment routinely includes a CBCL by at least one parent/caregiver and a TRF by the key classroom teacher.

The sample was drawn from a 12-month cohort of clients ending in December 2001. This incorporated an initial sample of 227 consecutive cases aged between 6-10 years for whom the CBCL (maternal caregiver informants) and the TRF (teacher or teacher aid informants) were completed in the weeks following the initial assessment. The study was restricted to this age band because including older cases would have required the use of informant versions other than the TRF. Twenty-seven of the 227 cases were subsequently excluded because a significant proportion of the CBCL or TRF items had not been completed, leaving a final sample of 200. It is far more common in this clinic population to have a maternal rather than paternal figure attend assessments, and thus only CBCLs completed by maternal figures were used.

Diagnostic and demographic information regarding the sample was obtained from a separate patient information system and case records completed by clinicians at the time of initial assessment. Informants did not know the diagnostic status of the cases.

Measures

The CBCL and the TRF are standardised parent and teacher rating scales of problematic emotional and behavioural problems in children over the preceding six months. Updated in 1991 (Achenbach, 1991a, , 1991b), these scales have been validated internationally in clinical and general populations (Achenbach, Verhulst et al., 1987; Bird et al., 1991; Sawyer et al., 1992; Weine et al., 1995; Weisz et al., 1987). The main differences with the teacher-report companion measure to the CBCL are the inclusion of items relating to academic performance/school behaviour and slight wording changes, for example, referring to "pupils" rather than "children". By omitting instrument-specific items, eighty-nine core items common to the CBCL and TRF make up eight empirically based cross-informant syndromes. Scored in a 0 (not true), 1 (somewhat or sometimes true) or 2 (very or often true) format, the number of common items in each scale varies (Withdrawn n = 8; Somatic Complaints n = 9 Anxious/Depressed n = 13; Social Problems n = 11; Thought Problems n = 10; Attention Problems n = 10; Delinquent Behaviour n = 10; Aggressive Behaviour n = 18). The eight scales based on these syndromes correlate well with clinical diagnoses of behavioural problems. Scales can be further categorised as internalising problems reflecting possible internal distress (Withdrawn, Somatic Complaints, Anxious/Depressed), or externalising problems reflecting conflicts with other people and their expectations of the child (Delinquent Behaviour, Aggressive Behaviour).

Statistical Analysis

Descriptive statistics (mean, standard deviation) were used to summarise the demographic, clinical and psychometric features of the population. Raw scores rather than scaled scores were used to reflect actual distributions. Reliability between the eight CBCL and TRF cross-informant syndrome scores was investigated by calculating Pearson correlation coefficients. This approach to the assessment of consistency assumes that parents and teachers should be providing coherent assessments of the cases. It is clear that consistency between parent and teacher assessments may depend on individual case or informant characteristics. We sought to explore the strength of the associations between parent and teacher consistency and a number of demographic and clinical features of the cases. The association between selected informant and case characteristics and reliability between corresponding CBCL/TRF scale raw scores was investigated as follows. The correlation coefficients were calculated for each subgroup defined by baseline demographic and clinical features and compared between the subgroups using a randomisation test procedure. This process selects subgroups (without replacement) of the same size as the groups of informants being compared. The correlation was then calculated for each group and the difference calculated. This random selection was repeated 10,000 times and an empirical distribution (null hypothesis) of the differences

Table 1

Primary presenting problem at initial assessment (N=200)

Presenting problem	*n*	%
Attention Deficit Hyperactivity Disorder	34	17
Elimination Disorder	22	11
Anxiety Disorder	12	6
Autistic Spectrum Disorder	9	4.5
Major Depressive Episode	9	4.5
Oppositional Defiant Disorder	7	3.5
Parent-Child Relationship Problem	6	3
Adjustment Disorder	6	3
Sexualised Behaviour	4	2
Attachment Disorder	3	1.5
Bipolar Disorder	1	0.5
Dyspraxia (Motor Coordination Disorder)	1	0.5
Intellectual Disability	1	0.5
Post Traumatic Stress Disorder	1	0.5
Tourettes Disorder	1	0.5
No psychiatric diagnosis	83	41.5

between correlation coefficients was then generated. The observed difference between the correlation coefficients was then compared to this distribution and an empirical two-tailed p-value calculated. The significance level accepted was $p < .05$.

Demographic and clinical features of the cases and informants

Just over two thirds of the cases were boys (69.5%), and most (86%) identified their ethnicity as New Zealand European. This predominance of boys was in keeping with the usual case gender mix of the service. As expected based on inclusion criteria, the mean age was 8.1 years old (SD = 1.3 years). The range and frequency of the primary presenting problem identified by clinicians at the point of initial assessment is described in Table 1. Fifty-nine percent of the cases were considered to have a preliminary DSM-IV based diagnosis (American Psychiatric Association, 2000), and of these, 29% had more than one diagnosis. Given that 15 different problem areas were recorded, for the purposes of data analysis, cases were categorised as having, or not having, a provisional DSM-IV diagnosis.

Almost all (91.5%) maternal informants were the biological mother of the case, with a small number being completed by various others including foster mothers (3%), grandmothers (2.5%), foster mothers (3%), stepmothers (0.5%), or another maternal care giver (2.5%). Consistent with the gender mix of teachers in New Zealand primary schools, most of the teacher informants were female (79.5%).

Results

Inter-informant scores and reliability

Table 2 summarises each scale score obtained from the different informants, along with the correlation between informants. CBCL mean scores ranged from 2.52 (Thought Problems) to 18.32 (Aggressive Behaviour) whereas TRF mean scores ranged from a low of 0.93 (Somatic complaints) to 9.58 (Aggressive behaviour).

Case or informant variables were significantly associated with three inter-rater correlation coefficients on two scales (Table 3). First, case gender was significantly associated with the level of agreement between informants on the Anxious/Depressed scale ($p = .00$). That is, when assessing anxiety or depressive symptoms, there was a significantly greater level of agreement between maternal caregivers and teachers when the gender of the case was male ($r = .24$) rather than female ($r = -.12$). On this same scale, informant gender was significantly associated with the level of agreement between informants ($p = .01$). That is, maternal caregivers had a significantly greater level of agreement with female teachers ($r = .20$) compared to male teachers ($r = -.14$). Finally, in regard to the Thought Problems scale, there was a significant difference between informants in regard to the presence ($r = .22$) or not ($r = .01$) of an initial diagnosis ($p = .02$). That is, there was a

Table 2

Mean CBCL (caregiver) and TRF (teacher) scale raw scores and associated Pearson correlation coefficients

	Informant Scale				
	CBCL		TRF		
Subscale	*M*	*SD*	*M*	*SD*	*r*
Withdrawn	3.98	2.74	3.12	3.10	.23
Somatic complaints	3.00	3.06	0.93	1.61	.25
Anxious/ Depressed	8.68	6.12	4.28	4.12	.14
Social problems	5.67	3.26	3.80	2.84	.38
Thought problems	2.52	2.46	1.14	1.81	.14
Attention problems	8.58	4.16	7.16	4.26	.35
Delinquent behaviour	4.68	2.97	2.64	2.83	.44
Aggressive behaviour	18.32	9.51	9.58	9.04	.33

Table 3
Analysis of consistency (Pearson correlation coefficients)

Subscales	Withdrawn	Somatic complaints	Anxious depressed	Social problems	Thought problems	Attention problems	Delinquent behaviour	Aggressive behaviour
Gender of case								
Male (*n* = 139)	.32	.19	**.24**	.44	.17	.38	.41	.37
Female (*n* = 61)	.04	.23	**-.12**	.23	.08	.27	.48	.25
p	.06	.79	**.00**	.29	.35	.64	.58	.20
Diagnosis of case								
None (*n* = 83)	.31	.32	.17	.45	**.01**	.37	.44	.38
Established (*n* = 117)	.18	.15	.11	.32	**.22**	.33	.44	.31
p	.35	.29	.50	.53	**.02**	.90	.99	.47
Informant gender								
Male (*n* = 41)	.23	.18	**-.14**	.22	-.01	.24	.39	.30
Female (*n* = 159)	.23	.21	**.20**	.42	.20	.38	.45	.34
p	.98	.83	**.01**	.45	.09	.59	.67	.64
Age of case (years)								
6 (*n* = 27)	.36	-.01	.27	.38	.40	.35	.45	.43
7 (*n* = 43)	.36	.10	.12	.45	.14	.39	.44	.46
8 (*n* = 47)	.09	.17	.13	.22	-.03	.03	.49	.23
9 (*n* = 42)	.23	.17	-.01	.43	.14	.51	.32	.25
10 (*n* = 41)	.16	.53	.18	.33	.12	.33	.46	.32
p	.72	.89	.30	.08	.30	.94	.11	.98
Ethnicity of case								
New Zealand European (*n* = 172)	.23	.22	.13	.36	.14	.32	.41	.34
Polynesian (*n* = 28)	.26	-.08	.14	.44	.12	.52	.53	.30
p	.60	.23	.45	.87	.31	.43	.46	.96

significantly greater level of agreement between informants regarding thought problems in the children thought to have an initial psychiatric diagnosis.

Discussion

The purpose of this study was to investigate the extent to which routinely known demographic and clinical variables contributed to consistency between informants on the CBCL and its companion measure, the TRF. The advantage of this study is that it involves a large routine clinical population of New Zealand children referred for the assessment of moderate to severe emotional, behavioural and psychiatric problems.

In this study, gender was shown to be significant only in respect to the consistency of the assessment of anxiety/depressive symptoms, and this could be explained by a number of possible factors. The greater agreement with respect to male children could reflect the fact that any anxious or depressive behaviour in boys is more easily detected because of gender-based social norms (Jensen et al., 1988). That is, on the basis that depression and anxiety are diagnosed more frequently in females across most age ranges (American Psychiatric Association, 2000; Piccinelli & Wilkinson, 2000; Schniering et al., 2000), such symptoms in boys may be perceived as more serious departures from the social norm. Whatever the phenomenon is that is being observed, deviations from expectations are likely draw greater consensus. A corresponding explanation may also account for the similar observed trend with respect to withdrawn behaviour. This too may be more aberrant in boys. The other consequence of this argument is that, in respect to girls, clinicians should routinely feel less confident about achieving inter-rater consensus, and may need to seek further information to assist in what will routinely be differences of opinion. As with other forms of evaluation, gender role stereotyping may be involved in what is endorsed, or not, on standardised rating scales. Our findings of a significant effect for male children stands in contrast to the one other study that found greater agreement for female children (Kolko & Kazdin, 1993).

There appears to be limited research demonstrating that teacher informant gender could be significant despite the wider literature regarding parents to suggest that this might occur. Our findings suggest that informant gender factors are also important but only in regard to the appraisal of child anxiety/depression problems. A potential reason for the difference between female-female informant dyads and female-male informant dyads is their life experiences. As both anxiety and depression are diagnosed more frequently in women, mothers and female teachers may have a greater familiarity and experience with these problems, and thus be in a better position to identify and agree about these in children under their care. Alternatively, as has been suggested by Seiffge-Krenke and Kollmar (1998), this may be further evidence that boys may be less inclined to discuss their private worries with their male caregivers and male teachers (in the case of this study), and this contributes to why female-female informant dyads have a higher level of agreement. Whatever the case, the influence of informant gender on inter-rater reliability clearly has implications for clinicians in that the utility of information may vary depending on the gender of the informant. This also points to the need for further investigation into the effect of teacher gender, and not just caregiver gender. It cannot be assumed that there is simply a maternal bias at work, which is the more often discussed clinical and research issue (De Los Reyes & Kazdin, 2005). A range of teacher-informant variables are likely to be important, and further research is required to stand this alongside the more developed caregiver-informant literature.

The greater level of agreement

between maternal and teacher informants on the Thought Problem scale in the presence of a provisional psychiatric diagnosis could be explained by the fact that this scale refers to explicit symptoms such as thought disorder and other psychotic phenomenon. Such problems are more likely to be noted by informants, irrespective of informant type or whether the concern arises from an initial clinical interview or a subsequent pencil-and-paper rating scale. In this respect, when appraising severe psychopathology, rating scales may have less utility since they may not add additional information.

Child age was not a significant contributor to informant discrepancy, and this stands in contrast to some studies (Achenbach, McConaughy et al., 1987; Grills & Ollendick, 2003) but not others (Choudhury et al., 2003; Kolko & Kazdin, 1993). The absence of effect in other studies has been attributed to methodological problems such as small sample sizes or categorising age into bands. These criticisms do not apply in the present study. That said, the study is limited by a number of issues including the early collection of the provisional diagnosis, thereby reducing the opportunity to detect diagnoses that take time to declare themselves. This means that the recorded diagnoses were probably under-inclusive of actual clinical difficulties. Furthermore, the ability to consider the effect of case ethnicity on inter-informant reliability was limited by the relatively small number of non-European cases (14%). This rate was in keeping with the known ethnicity of the Christchurch population, and thus further investigation of this issue should consider either using larger samples or conducting studies in other sites.

Collectively these findings suggest the need for considerable caution in expecting adult informants to agree about what emotional and behavioural problems require attention when children present to clinics. While the full suite of factors contributing to such differences of opinion may not be fully understood, it is important to know at least that the routinely observable attributes of the observer and observed appear to systematically influence such opinions. In short, clinicians should consider the results of the questionnaires alongside information about who filled them in. Future research is recommended to investigate the consistency between male-male informant dyads, and whether the significant gender findings in this study also occur in male caregiver-female teacher dyads.

References

Achenbach, T. M. (1991a). *Manual for the Child Behaviour Checklist/4-18 and 1991 Profile*. Burlington, VT: Department of Psychiatry, University of Vermont.

Achenbach, T. M. (1991b). *Manual for the Teacher's Report Form and 1991 Profile*. Burlington, VT: Department of Psychiatry, University of Vermont.

Achenbach, T. M., and Edelbrock, C. (1991). *Integrative Guide for the 1991 CBCL/4-18, YSR, and TRF profiles*. Burlington, VT: Department of Psychiatry, University of Vermont.

Achenbach, T. M., McConaughy, S. H., and Howell, C. T. (1987). Child/adolescent behavioural and emotional problems: implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*, 213-232.

Achenbach, T. M., Verhulst, F. C., Baron, G. D., and Akkerhuis, G. W. (1987). Epidemiological comparisons of American and Dutch children: I. Behavioral/emotional problems and competencies reported by parents for ages 4 to 16. *Journal of the American Academy of Child & Adolescent Psychiatry, 26*, 317-325.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (4th ed.). Text revision*. Washington, D.C.: APA.

Bird, H. R., Gould, M. S., Rubio-Stipec, M., Staghezza, B. M., and Canino, G. (1991). Screening for childhood psychopathology in the community using the Child Behavior Checklist. *Journal of the American Academy of Child & Adolescent Psychiatry, 30*, 116-123.

Choudhury, M. S., Pimentel, S. S., and Kendall, P. C. (2003). Childhood anxiety disorders: parent-child (dis)agreement using a structured interview for the DSM-IV. *Journal of the American Academy of Child & Adolescent Psychiatry, 42*, 957-964.

Danforth, J. S., and DuPaul, G. J. (1996). Interrater reliability of teacher rating scales for children with attention-deficit hyperactivity disorder. *Journal of Psychopathology and Behavioural Assessment, 18*, 227-237.

Davies, L. C., and McKelvey, R. S. (1998). Emotional and behavioural problems and competencies among immigrant and non-immigrant adolescents. *Australian & New Zealand Journal of Psychiatry, 32*, 658-665.

De Los Reyes, A., and Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: a critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131*, 483-509.

Fergusson, D. M., and Horwood, L. J. (1987a). The trait and method components of ratings of conduct disorder--Part I. Maternal and teacher evaluations of conduct disorder in young children. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 28*, 249-260.

Fergusson, D. M., and Horwood, L. J. (1987b). The trait and method components of ratings of conduct disorder--Part II. Factors related to the trait component of conduct disorder scores. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 28*, 261-272.

Fergusson, D. M., and Horwood, L. J. (1989). Estimation of method and trait variance in ratings of conduct disorder. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 30*, 365-378.

Grills, A. E., and Ollendick, T. H. (2003). Multiple informant agreement and the anxiety disorders interview schedule for parents and children. *Journal of the American Academy of Child & Adolescent Psychiatry, 42*, 30-40.

Jensen, P. S., Traylor, J., Xenakis, S. N., and Davis, H. (1988). Child psychopathology rating scales and interrater agreement: I. Parents' gender and psychiatric symptoms. *Journal of the American Academy of Child & Adolescent Psychiatry, 27*, 442-450.

Kazdin, A. E., Esveldt-Dawson, K., and Loar, L. L. (1983). Correspondence of teacher ratings and direct observations of classroom behavior of psychiatric inpatient children. *Journal of Abnormal Child Psychology, 11*, 549-564.

Kolko, D. J., and Kazdin, A. E. (1993). Emotional/behavioral problems in clinic and nonclinic children: correspondence among child, parent and teacher reports. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 34*, 991-1006.

Lee, S. W., Elliott, J., and Barbour, J. D. (1994). A comparison of Cross-Informant Behaviour Ratings in school-based diagnosis. *Behavioural Disorders, 19*, 87-97.

Piccinelli, M., and Wilkinson, G. (2000). Gender differences in depression. Critical review. *British Journal of Psychiatry, 177*, 486-492.

Rutter, M., Tizard, J., Yule, W., Graham, P.,

and Whitmore, K. (1977). Isle of Wight Studies. In S. Chess & A. Thomas (Eds.), *Annual progress in clinical psychology and child development* (pp. 359-392). New Your: Brunner/Mazel.

Sawyer, M. G., Baghurst, P., and Clark, J. (1992). Differences between reports from children, parents and teachers: implications for epidemiological studies. *Australian & New Zealand Journal of Psychiatry, 26*, 652-660.

Schniering, C. A., Hudson, J. L., and Rapee, R. M. (2000). Issues in the diagnosis and assessment of anxiety disorders in children and adolescents. *Clinical Psychology Review, 20*, 453-478.

Seiffge-Krenke, I., and Kollmar, F. (1998). Discrepancies between mothers' and fathers' perceptions of sons' and daughters' problem behaviour: a longitudinal analysis of parent-adolescent agreement on internalising and externalising problem behaviour. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 39*, 687-697.

Stanger, C., and Lewis, M. (1993). Agreement among parents, teachers and children on internalising and externalising behaviour problems. *Journal of Clinical Child Psychology, 2*, 107-115.

Verhulst, F. C., Koot, H. M., and Van der Ende, J. (1994). Differential predictive value of parents' and teachers' reports of children's problem behaviours: a longitudinal study. *Journal of Abnormal Child Psychology, 22*, 531-546.

Visser, J. H., van der Ende, J., Koot, H. M., and Verhulst, F. C. (2003). Predicting change in psychopathology in youth referred to mental health services in childhood or adolescence. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 44*, 509-519.

Weine, A. M., Phillips, J. S., and Achenbach, T. M. (1995). Behavioral and emotional problems among Chinese and American children: parent and teacher reports for ages 6 to 13. *Journal of Abnormal Child Psychology, 23*, 619-639.

Weisz, J. R., Suwanlert, S., Chaiyasit, W., Weiss, B., Achenbach, T. M., and Walter, B. R. (1987). Epidemiology of behavioural and emotional problems among Thai and American children: parent reports for ages 6 to 11. *Journal of the American Academy of Child & Adolescent Psychiatry, 26*, 890-897.

Yang, H. J., Soong, W. T., Chiang, C. N., and Chen, W. J. (2000). Competence and behavioral/emotional problems among Taiwanese adolescents as reported by parents and teachers. *Journal of the American Academy of Child & Adolescent Psychiatry, 39*, 232-239.

Address for correspondence:

Lois Surgenor PhD
Department of Psychological Medicine
University of Otago, Christchurch
PO Box 4345
Christchurch
New Zealand

Tel: 64 3 3720 400

Fax: 64 3 3720 407

Email: lois.surgenor@otago.ac.nz