

Parent and Teacher (Dis)Agreement on the Conners Rating Scale: Revised-Long Format

Karen E. Waldie,¹ Christine M. Cornforth,² Edwin A. Mitchell,¹ Rebecca F. Slykerman,¹
Claire Wall,¹ and John M. D. Thompson¹

¹The University of Auckland, New Zealand

²The University of Liverpool, United Kingdom

Abstract

The Conners Rating Scale: Revised-Long version (CRS: R-L) has been used widely in clinic and research settings to measure child behavior and symptoms of attention deficit hyperactivity disorder (ADHD). Despite evidence of poor agreement between parent and teacher reports of child internalizing and externalizing problems, no study to date has examined the levels of agreement between parent and teacher reports of child behavior using the CRS: R-L. Our aim was to measure the level of agreement between parent and teacher perception of child emotion and behavior problems using the CRS: R-L. A total of 591 New Zealand European mothers and their 7-year-old children participated in Phase 4 of the Auckland Birthweight Collaborative (ABC) study. Child behavior was measured when the children were 7 years old using the parent and teacher versions of the CRS: R-L. Parent and teacher agreement was analyzed using weighted kappa coefficient, Pearson correlation coefficient, and Bland-Altman plots. Agreement between parent and teacher reports of child behavior was found to range between poor and low. Agreement decreased with reports of increasing behavioral problems and was generally higher for externalizing subscales and lower for internalizing subscales.

Keywords: Child behavior, inter-rater reliability, parent, teacher, longitudinal, small for gestational age (SGA), Conners Rating Scale: Revised-Long version (CRS: R-L)

It is well understood that behavior problems in children can vary from one situational context to another (Achenbach & McConaughy, 1997; Nass, 2005). Therefore, proving the “pervasiveness” of disorders such as attention deficit hyperactivity disorder (ADHD) requires the involvement of multiple informants in the assessment of children’s behavior (World Health Organization, 2004).

When assessing a child’s behavior, it is important that the chosen informants are familiar with the child across a range of different settings (Miller et al., 2001). As well as standardized clinical interviews, researchers and clinicians often use the parent and teacher report of a particular behavior rating scale as part of a diagnostic and/or treatment assessment. For example, in an Australian study of children with ADHD, parent and teacher ratings of behavior were used in

the diagnostic process in 85%-88% of cases (Efron et al., 2016). While it is well understood that the correspondence between direct observation and behavior rating scale diagnoses supports the validity of rating scales for making a diagnosis, behavior rating scales should not be used as a diagnostic tool in isolation (Kazdin, 1995; Nass, 2005).

Parents and teachers are usually the two groups of respondents who are asked to provide information about a particular child’s behaviour. Parents provide valuable information as they are considered to be most familiar with their child across time and in different environments (Cohen-Kettenis & Van-Goozen, 2002). For school-aged children, teachers can also be important second informants. They observe children learning, playing, and interacting with their peers in the school setting, which allows them a unique per-

spective of a child's social and cognitive functioning (Nottelmann, 2002). Furthermore, teachers also have other children of similar ages to compare with, making them better placed at judging whether a child's behaviours are outside the normal range.

It has been widely observed that the joint use of multiple informants has led to problems of cross-informant effects (e.g., Lohaus et al., 2019). Not surprisingly, studies examining agreement levels between different raters of a child's behaviour have found that agreement is generally higher between the parents of a child than between those who observe the child in different environments; for example, teacher vs. football coach (Achenbach et al., 1987) and preschool teacher vs. parents (Penninx et al., 2008).

Two meta-analyses have examined the cross-informant correlations and situational specificity of child behaviour problems (Achenbach et al., 1987; Lee et al., 1994). Both studies reported low levels of parent and teacher agreement. Specifically, Achenbach and colleagues (1987) included 41 studies in their analysis, of which 4 reported a correlation of 0.5 and above; the remaining studies reported correlations ranging from 0.2 to 0.5. Other studies have revealed similar findings (Gagnon et al., 1992; Lee et al., 1994; Mitsis, 2000; Youngstrom et al., 2000). For example, Youngstrom and colleagues (2000) reported a correlation of 0.3 between parent and teacher reports of child externalising problems and 0.2 for internalising problems.

Further, in a meta-analysis of 49 studies on the agreement between informants on externalising, internalising, and social problems in children with autism spectrum disorder (ASD) or intellectual disability (ID), Stratis and Lecavalier (201) found that, consistent with previous studies, the correlation between informants was higher for externalising problems ($r = .42$) than for internalising problems ($r = .35$) or social problems ($r = 0.30$). Level of agreement for each of these domains differed by the child's age, diagnosis, and IQ.

Mitsis and colleagues (2000) examined parent and teacher agreement for 74 children who were clinically referred with ADHD. The results of their study revealed correlations of 0.3 for ADHD Inattentive Type, 0.4 for ADHD Hyperactive-Impulsive Type, and 0.4 for ADHD Combined Type. The authors also reported that agreement was lower for categorical measures using intra-correlations (Mitsis, 2000). In another study of agreement between parent and teacher reports of disruptive behaviours of children diagnosed with ADHD, Angtrop et al. (2002) found no significant associations between parent and teacher report of ADHD symptoms ($r = 0.13$ for inattention

and $r = 0.09$ for hyperactivity/impulsivity), moderate agreement for problems relating to conduct disorder ($r = 0.36$), and good agreement for problems relating to oppositional defiant disorder ($r = 0.56$).

Correlation between paediatrician- and parent-reported internalising and externalising problems in children with ADHD has also been shown to be poor in an Australian study despite most parents agreeing with the diagnosis of ADHD (Efron et al., 2016). For example, in a study of parent and teacher ratings of preschool children, Orylska and colleagues (2016) found that the agreement between parents and teachers on the Conners Early Childhood Behaviour Scale was low to moderate. Furthermore, they noted that parent and teacher ratings of working memory ability mediated the relationship between inattentive/hyperactive behavior on the Conners Early Childhood Behaviour Scale CEC BEH [S] and fluid intelligence (Orylska et al., 2016).

For adolescents, self-reported symptoms of ADHD can be obtained in addition to observer reports. A study of 80 adolescent survivors of childhood cancer found moderate agreement between parent and teacher reports, with lower agreement between self-reports and observer reports of problems on the Conners 3 Rating Scale. Across all scales, parents reported more problems than either teachers or the young persons themselves (Willard et al., 2016). Further, comparing agreement on the Conners 3 Rating Scale between teachers, parents, and youth between 6 to 18 years of age, Izzo et al. (2019) found that correspondence between the three groups of informants was moderate only. Finally, in adults, the concordance between self-reported symptoms of ADHD and informant report (partner or parents) has been shown to range between small to moderate, with the authors concluding that multi-informant information is important in the diagnosis of adult ADHD (Abt-Mörstedt et al., 2015).

The Conners Rating Scale: Revised (CRS-R)

The Conners Rating Scale: Revised (CRS-R) is among the most widely used behavior rating scales in the history of research on children with ADHD (Barkley, 1998). The CRS-R Long Format (CRS: R-L) includes parallel parent and teacher questionnaires, with the aim of measuring the various aspects of problems (including those associated with ADHD) in different situational contexts. Cohen and colleagues (1990) compared parent and teacher ratings on the earlier versions of the Conners and found correlations between parent and teacher ratings to be 0.40 for the ADHD Index scale and 0.42 for the Oppositional scale.

Purpose of the Present Study

The aim of the current study was to examine the level of agreement between parent and teacher ratings of behaviour in a community-based sample of 7-year-old New Zealand European children using the CRS: R-L. Some lack of agreement between the different measures is inevitable, but a focus of this study was the extent by which methods disagree (Bland & Altman, 1986). Despite its widespread use both clinically and in research, to our knowledge no studies to date have reported the levels of agreement between parent and teacher reports of child behaviour problems using the CRS: R-L.

It is noteworthy that the CRS: R-L (Conners, 1997) is no longer published or sold and was replaced by the Conners 3 (Conners, 2008). Despite using an earlier version of the Conners, we believe that the current study makes an important contribution to the wider discussion about agreement between teachers and parents when assessing phenomena as complex as behavior and attention.

Symptoms of ADHD need to be present in at least two different settings to indicate a diagnosis according to the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5; American Psychiatric Association [APA], 2013), so multi-informant assessment approaches are typically adopted. Such information and decisions about the child by specialists (psychologists, social workers, special educators, etc.) are largely based on the information provided by parents and teachers. As such, the current research about rater-agreement remains relevant for practice and has implications for the accurate identification of children with clinically significant attention problems.

Method

Participants and General Procedure

The Auckland Birthweight Collaborative (ABC) Study includes a longitudinal cohort with disproportionate sampling. The methodology of previous phases of the ABC study has been described in detail previously (Thompson et al., 2001). In brief, participants are mothers and their singleton infants born full term (≥ 37 completed weeks of gestation) in the Auckland and Waitamata Healthcare regions between 1995 and 1997. Approximately half of the infants were small for gestational age (SGA) with birthweights equal to or below the sex-specific tenth percentile for gestation; the remainder were a random sample of infants born appropriate for gestational age (AGA) (Thompson et

al., 1994). Infants were excluded from participating in the study if they did not reside in the study region or if they had any congenital abnormalities likely to affect subsequent growth or development.

Perinatal information was collected on 1,714 participants at birth (Phase 1), of which 863 (50.4%) were male, 870 (50.8%) were born AGA, and 871 (50.8%) had mothers of European ethnicity. The cohort was followed up at one year of age (via a postal questionnaire), and face-to-face assessments were undertaken at 3½ years and beyond. At the 3½-year follow-up, the response rates for other ethnicities was substantially lower than for European participants, so in consultation with our M ori co-investigators and advisors, it was concluded that results for these groups were unlikely to be generalizable, and further follow-up attempts with non-European participants ceased.

Data collection for Phase 4 of the study took place when the children were 7 years old. As noted, due to the differential response rates amongst ethnic groups at previous phases of the ABC study, the present study was restricted to New Zealand European mothers and their children. A total of 871 New Zealand European mothers and infants enrolled at birth were eligible for follow-up at 7 years of age; of these, 591 (67.7%) participated.

Of the 591 New Zealand European mothers and their 7-year-old children, we collected complete sets of parent and teacher reports of child behavior for 553 children. The sample consisted of 292 (52.8%) females and 261 (47.2%) males. Compared with all mothers who were initially recruited into the study, respondents at Phase 4 were more likely to be older at the birth of their child ($p < 0.0001$), married ($p < 0.0001$), of high socioeconomic status ($p = 0.0002$), have tertiary education ($p < 0.0001$), be non-smokers during pregnancy ($p < 0.0001$) and not use alcohol during this time ($p = 0.02$), have good social support in pregnancy ($p = 0.03$), and have AGA infants ($p < 0.0001$). Respondents and non-respondents did not differ on obstetric factors, including gestational age, infant sex, parity, and levels of maternal stress ($p > 0.05$).

Measures

Child behavior was measured when study children were age 7 years old using the parent and teacher versions of the Conners Rating Scale: Revised-Long Format (CRS: R-L; Conners, 1997). Parents and teachers indicate on a 4-point Likert scale how much each item applies to a particular child (0 = "not at all," 1 = "just a little," 2 = "pretty much," and 3 = "very much"). Relevant items are summed to provide total

scores for the following behavioural factors that are not specific to ADHD: Oppositional Problems, Cognition, Hyperactivity, Emotion, Perfectionism, and Anxiety. In addition, several subscales specific to ADHD are also provided, including (a) *Diagnostic and Statistical Manual of Mental Disorders – Fourth Edition* (DSM-R-IV; APA, 1994) symptom subscales (Inattentive, Hyperactive/Impulsive, and Total); (b) Conners Global Indices (GGI; Conners, 1997; Restless-Impulsive, Emotional Liability, and Total); and (c) an ADHD Index scale.

The DSM-IV symptom subscales include the ADHD symptom criteria from DSM-IV (APA, 1994). The Conners Global Index includes 10 items (7 relating to Restless-Impulsive symptoms and 3 relating to Emotional Liability). The CGI is recommended for treatment monitoring and characterization of patterns of behavior (Conners, 1997). The ADHD Index includes 12 items relating to behaviors typical of children with ADHD.

The CRS: R-L parent version was given to one parent (usually the mother) of participating children at the 7-year follow-up, with standardized instructions for completion. The CRS: R-L teacher form was posted to each child's current teacher together with standardized instructions and a prepaid envelope. Raw scores were converted to T-scores, which have a mean of 50 and a standard deviation of 10. All subscales were investigated except Psychosomatic, as this is only available for parent report.

Ethical approval for this research was obtained from the North Health Research Ethics Committee (AKX/02/00/319).

Statistical Analyses

Agreement between parent and teacher report of behavior problems was calculated using several statistical methods. Specifically, the relationships between parent and teacher scores were examined categorically using weighted kappa correlation coefficient (K) and continuously using Pearson correlation coefficient (r). The kappa statistic was interpreted according to Landis and Koch (1977): almost perfect 0.81-1.00, substantial agreement 0.61-0.80, moderate agreement 0.41-0.60, fair agreement 0.21-0.40, slight agreement 0.00-0.20, and poor agreement <0. The strength of the correlation coefficients was interpreted according to published guidelines: excellent >0.90; good 0.80-0.90; moderate 0.50-0.70; low 0.30-0.50; poor <0.30 (Andrews et al., 1993).

Agreement was also examined using the statistical procedure outlined by Bland and Altman (Bland, 1986). Bland-Altman plots were used to examine the agreement between parent and teacher report of behavior problems for all CRS: R-L subscales. The Bland-Alt-

man plot includes the within-pair mean (in this study, this is the mean of the parent and the teacher report of behavior problems) to estimate the severity of a child's behavior problems on the abscissa. The difference between parent and teacher reports was used to represent the inter-rater agreement level on the ordinate.

A large and positive value of the difference indicated that parents reported more problems than teachers, whereas a large and negative value indicated that teachers reported more problems than parents. The smaller the absolute value, the higher the agreement between the parents and teachers report of behavioral symptoms.

Statistical analyses were calculated in SAS version 9.4 (SAS Institute).

Results

Means and standard deviations are presented for parent and teacher report of child behavior for all scales (see Table 1), including the Conners Core Factors, Conners ADHD Index, and CGI subscales. All parental report scores were found to be slightly above the mean (T-score >50). Parents' scores were consistently higher than teacher scores except for the Anxious/Shy subscale; however, these differences were not statistically significant ($T > 0.05$).

Atypical behavior problems were defined by scores that were equal to or above 1 SD (\geq T-score of 60). Scores that were below this threshold were considered normal. Parents identified a higher prevalence of all problems than teachers, except for the subscale relating to symptoms of anxiety (teacher prevalence = 21.3%, parent prevalence = 17.0%). Kappa statistics for parent and teacher report of behavior for all CRS: R-L subscales ranged from poor to low. The kappa coefficients were comparable to the Pearson correlation coefficients (see Table 2). However, it is possible that the categorization of subscales resulted in a loss of power, which is reflected in the kappa coefficients.

Pearson and kappa correlation statistics were used to assess the relationship between parent and teacher reports of child behavior for all subscales. As shown in Table 2, all Pearson correlations were positive and significant ($p < 0.0001$), ranging from $r = 0.15$ for Perfectionism to $r = 0.50$ for Cognitive problems, indicating a significant but poor-to-moderate relationship between parent and teacher reports. Agreement was higher for the externalizing subscales; for example, Conduct problems ($r = 0.28$) and Hyperactivity ($r = 0.43$). Subscales related to internalizing problems were found to have lower levels of agreement (e.g., Anxious/Shy problems; $r = 0.19$). Levels of parent and teacher agreement for the ADHD subscales ranged from $r = 0.24$ for CGI:

Emotional Liability to $r = 0.47$ for the Conners ADHD Index. A similar pattern was observed for parent and teacher agreement on the DSM-IV ADHD subscales (Inattentive $r = 0.45$, Hyperactive/Impulsive $r = 0.39$, and Total $r = 0.46$).

Bland-Altman plots were produced for all subscales. Figures 1a-2f show the plots for the Conners Core Factors, and Figures 2a-2g show the Conners

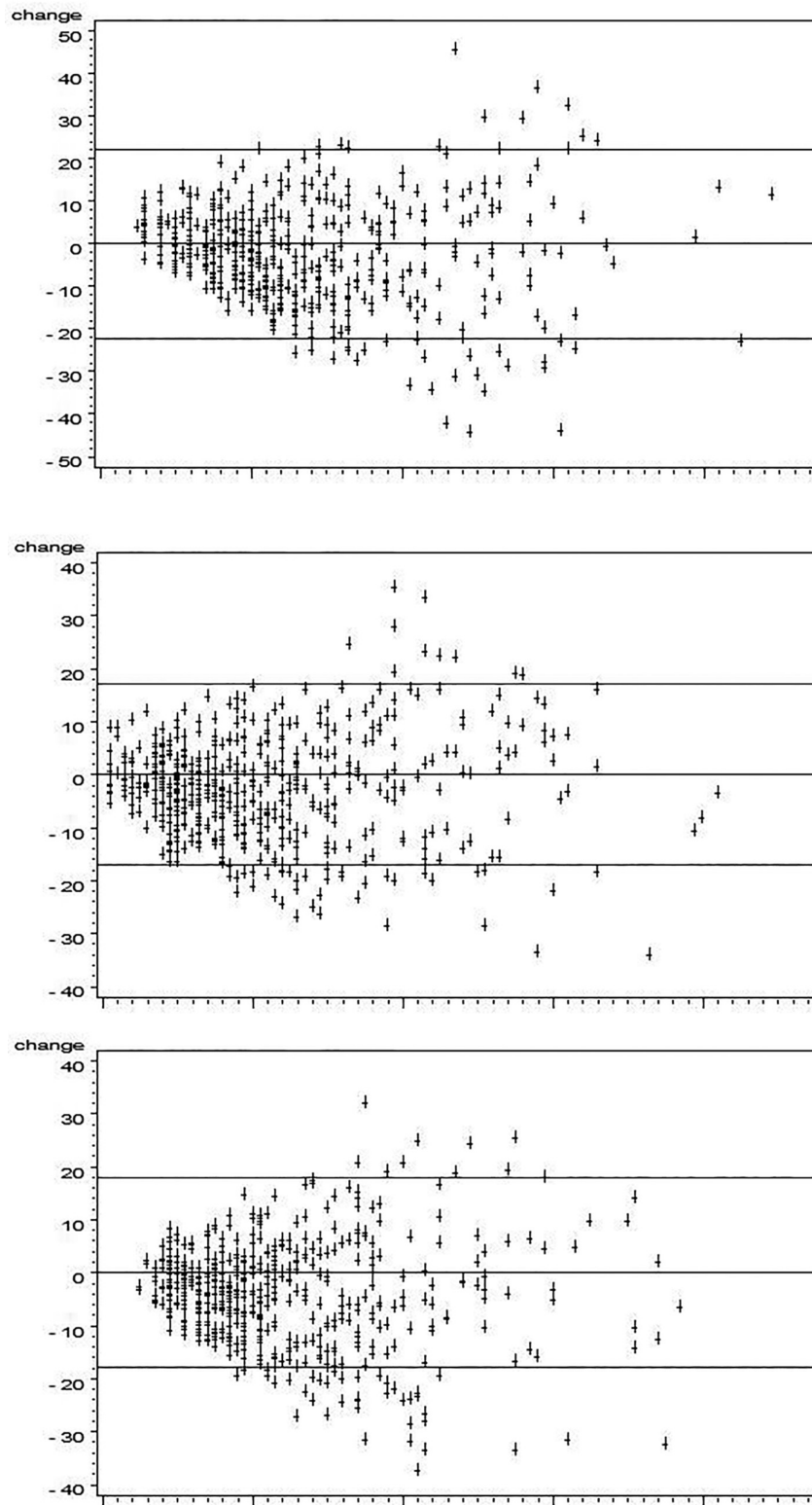
DSM-IV Hyperactive-Impulsive, Inattentive, and Combined Type ADHD subscales. All plots reveal very similar patterns with great variation in the differences between parent and teacher reports of child behavior. Furthermore, they reveal that the differences between parent and teacher reports for all subscales increased as the mean of the subscale score increased, as indicated by the funnel shape formed by the points.

Table 1
Mean T-Scores and Standard Deviation (SD) for Parent and Teacher Report of Child Behavior by Sex for All CRS: R-L Subscales

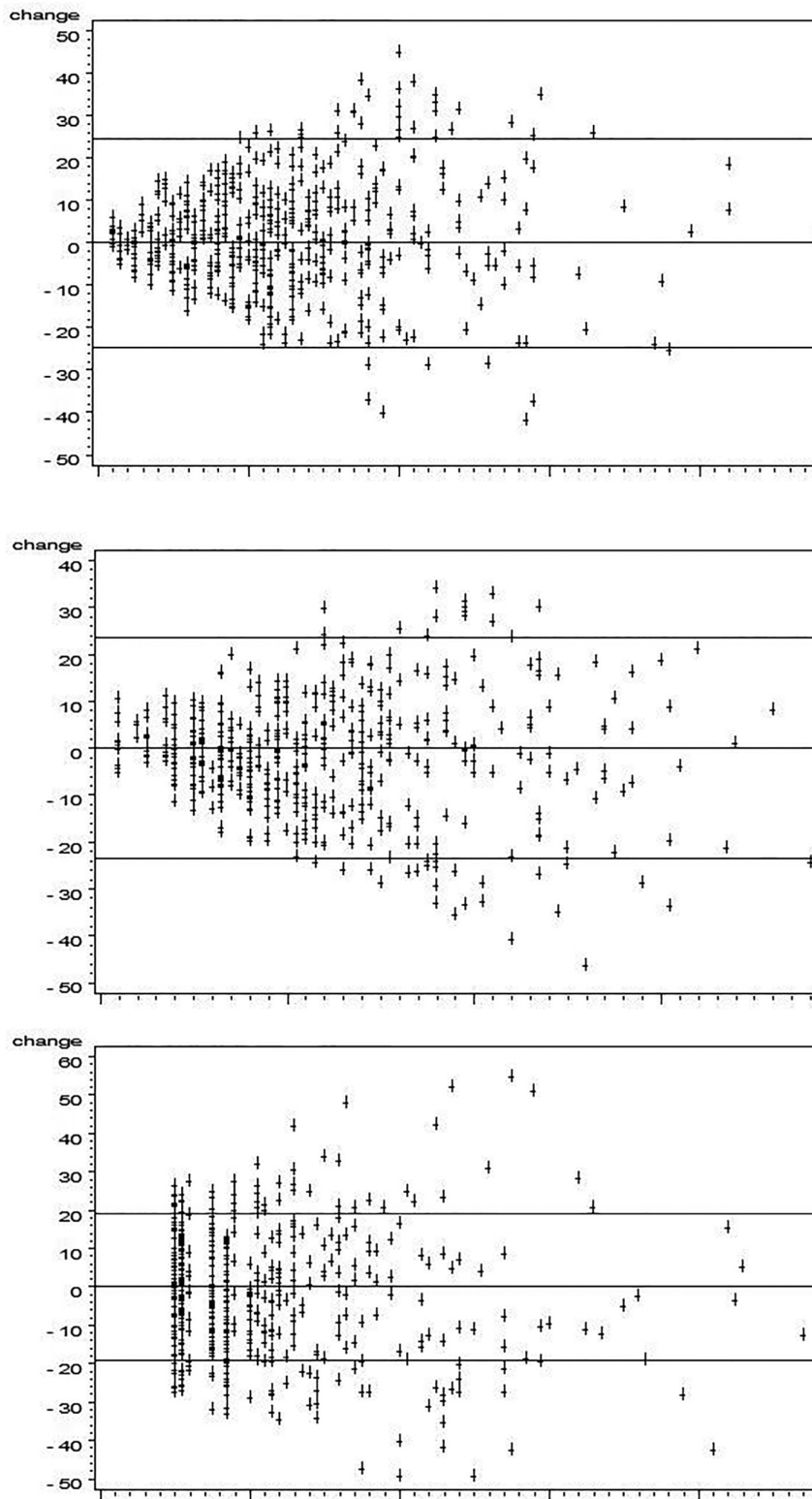
	Parent				Teacher			
	Male		Female		Male		Female	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Core Factors								
Conduct	54.62	9.64	54.54	9.52	52.15	8.12	52.56	9.42
Cognitive	51.64	9.34	52.02	7.82	49.32	9.60	49.72	7.47
Hyperactivity	54.14	9.11	53.24	8.24	49.70	8.08	50.68	8.21
Anxious/Shy	50.81	8.98	51.88	9.86	53.63	10.18	52.97	9.53
Perfectionism	52.55	9.17	51.97	9.53	52.30	9.00	50.22	8.36
Social Problems	51.06	8.73	51.04	9.34	49.91	7.80	50.48	8.05
ADHD Subscales								
Conners ADHD Index	51.77	8.73	51.73	7.37	49.87	8.00	49.99	8.49
CGI: Restless Impulsive	54.03	9.25	53.11	8.41	50.91	9.21	51.01	9.20
CGI: Emotional Liability	54.03	9.25	53.11	8.40	50.91	9.20	51.01	9.20
CGI: Total	53.33	8.97	52.53	8.41	50.26	9.22	50.55	9.40
DSM-IV: Inattentive	51.36	8.92	51.00	7.07	48.29	8.41	49.54	7.63
DSM-IV: Hyperactive Impulsive	54.14	9.06	54.12	8.47	49.78	8.40	50.98	9.97
DSM-IV: Total	52.90	9.01	52.52	7.50	48.63	8.02	49.91	7.87

Table 2
Weighted Kappa Coefficient and 95% Confidence Interval (CI) for Parent and Teacher Report of Child Behavior According to All CRS: R-L Subscales

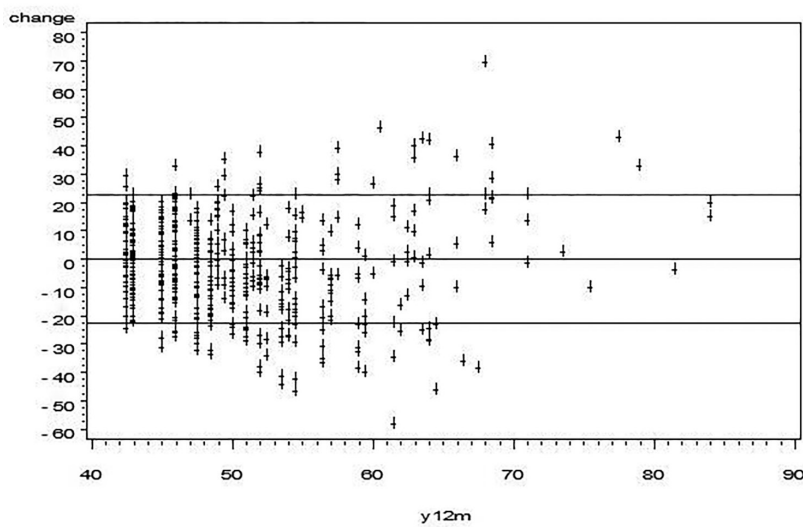
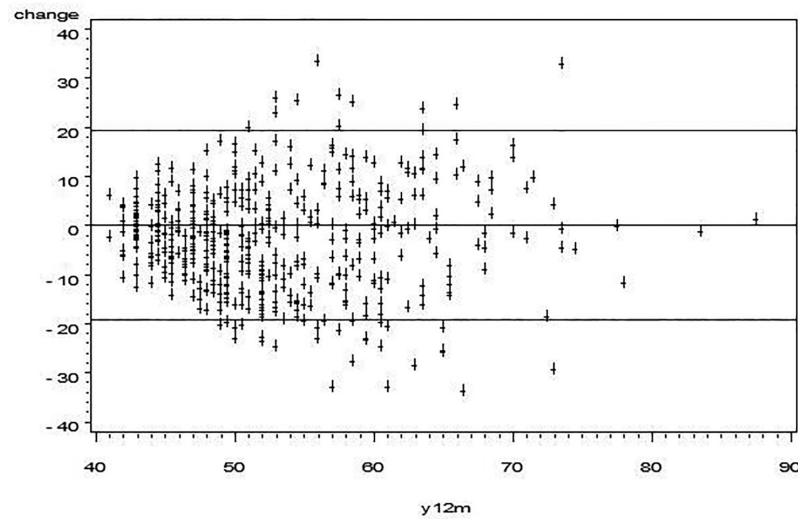
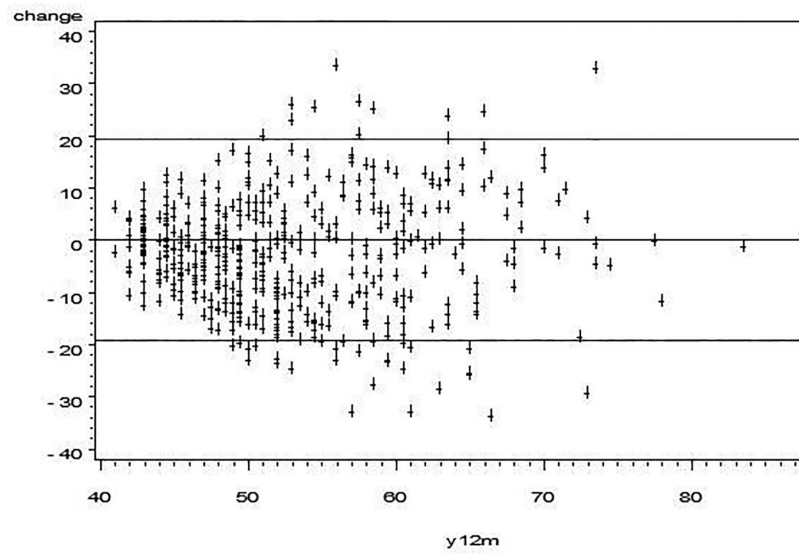
	Weighted Kappa Coefficient		Pearson Correlation Coefficient	
	K	95% CI	r	p-value
Core Factors				
Conduct	0.15	0.07 – 0.24	0.28	<0.0001
Cognitive	0.30	0.20 – 0.39	0.50	<0.0001
Hyperactivity	0.28	0.18 – 0.38	0.43	<0.0001
Anxious/Shy	0.15	0.06 – 0.23	0.19	<0.0001
Perfectionism	0.13	0.04 – 0.21	0.15	<0.0001
Social Problems	0.27	0.16 – 0.38	0.37	<0.0001
ADHD Subscales				
Conners ADHD Index	0.27	0.18 – 0.36	0.47	<0.0001
CGI: Restless Impulsive	0.26	0.17 – 0.34	0.43	<0.0001
CGI: Emotional Liability	0.14	0.05 – 0.24	0.24	<0.0001
CGI: Total	0.22	0.13 – 0.31	0.40	<0.0001
DSM-IV: Inattentive	0.26	0.16 – 0.37	0.45	<0.0001
DSM-IV: Hyperactive Impulsive	0.27	0.18 – 0.36	0.39	<0.0001
DSM-IV: Total	0.24	0.14 – 0.33	0.46	<0.0001



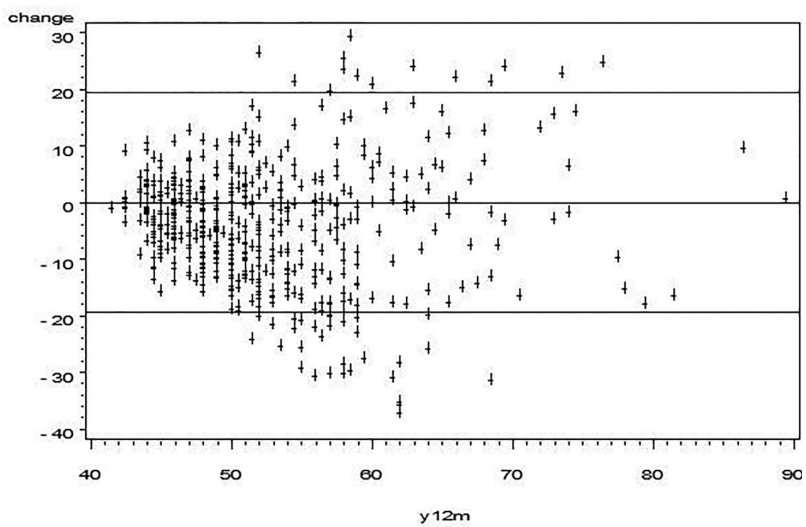
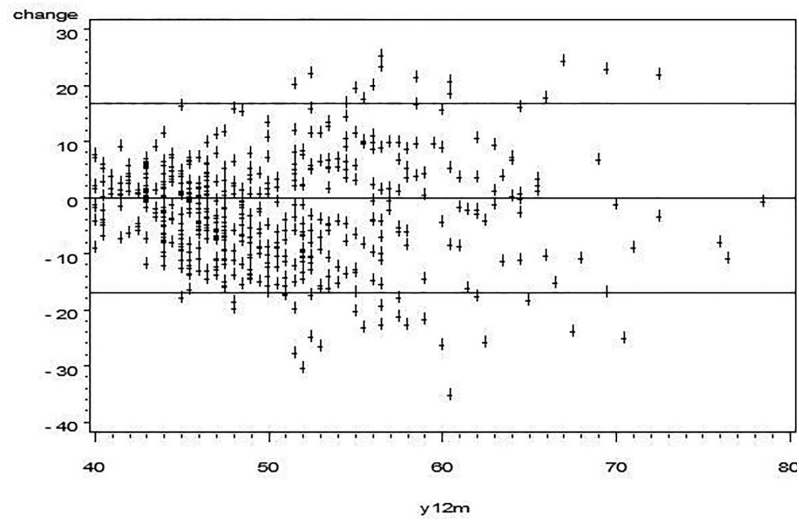
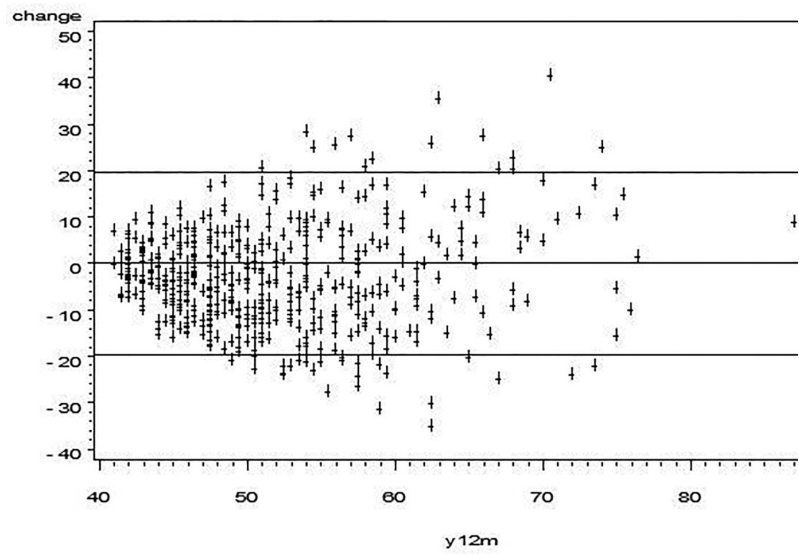
Figures 1a-1c
Bland-Altman plots for CRS: R-L Core Factors



Figures 1d-1f
Bland-Altman plots for CRS: R-L Core Factors



Figures 2a-2c
Bland-Altman plots for CRS: R-L ADHD Subscales



Figures 2d-2f
Bland-Altman plots for CRS: R-L ADHD Subscales

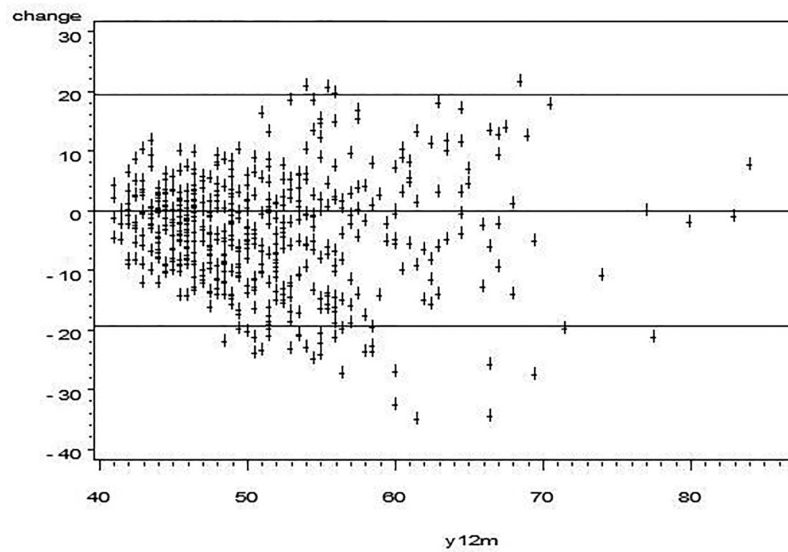


Figure 2g
Bland-Altman plot for CRS: R-L ADHD Subscales

Discussion

To our knowledge, this study is the first to examine parent and teacher agreement using the CRS: R-L. Our main findings were as follows: (a) parent and teacher agreement for all Conners behaviour subscales was poor to low; (b) agreement was highest for the Cognitive problems and DSM-IV Total subscales, and the lowest for the Perfectionism subscale; and (c) parent and teacher agreement declined as behavioural problem scores increased.

The finding that parent ratings were generally higher than teacher ratings is not surprising and is consistent with other studies (Angtrop et al., 2002; Deng et al., 2004; Mitsis, 2000). In a recent study using the latest Conners version (Colomer et al., 2020), compared to adolescents without ADHD, adolescents with ADHD underreported their symptoms and impairment when parent reports, but not teacher reports, were the indicator of performance.

Our study confirms previous findings that parent and teacher reports of child behaviour are not as consistent as one would expect. There are several possible reasons for this. First, children arguably behave differently at school than at home, so the differences in parent and teacher perceptions of child behaviour are due to the different observational contexts. Second, it is also possible that parents and teachers under- or overreport symptoms of behaviour problems depending on the family and/or school climate. For example, parents who are depressed, anxious, or stressed may overreport their children's behaviour

problems, and teachers who fear the repercussions of classifying a child as having behaviour problems may also underreport a child's symptoms.

Further, the present study found that agreement between parent and teacher report of ADHD symptoms is low. This is consistent with earlier research (Angtrop et al., 2002; Biederman et al., 1993; Cohen et al., 1990; Loeber et al., 1990; Mitsis 2000). Parent and teacher ratings of oppositional defiant disorder and conduct disorder are generally lower for a normal sample than for a clinically referred group of children (Angtrop et al., 2002).

The current study supports the idea that low levels of agreement do not mean that one informant is providing invalid or unreliable information (Achenbach et al., 1987) and that teachers and parents differ in their perception of problem behaviour (Reid & Maag, 1994). Low levels of parent and teacher agreement may be explained in terms of the situation specificity of the child's behaviour. Thus, it is well known that symptoms of ADHD fluctuate across environments (Barkley, 1998). For example, a child with ADHD may benefit from structured activities, resulting in lower levels of problem behaviour at school. The low levels of agreement nonetheless complicate assessment of ADHD. Some authors (e.g., Collett et al., 2003) have suggested that the term *pervasiveness* does not mean that ratings of both informants need to reach a certain cut-off score in order to meet the diagnostic criteria for ADHD. The DSM-V (APA, 2013) defines pervasiveness as "some impairment from symptoms required in two or more settings" but

does not indicate the degree of impairment required in each setting (p. 62). Findings from this and other studies suggest that assessment protocols are needed in order to clearly define what is meant by pervasiveness in order to overcome the problems of contradictory parent and teacher reports of symptoms.

Similar to other studies in this area, the present study found that agreement was higher for externalising than internalising problems (Achenbach et al., 1987; Loeber et al., 1990). Deng and colleagues (2004) suggested that internalising problems are difficult for teachers to detect in school. If this were true, this study would not find that teachers reported significantly more Anxious/Shy problem behaviours than parents (teacher = 21.3%, parent = 17.0%). One possible explanation for the finding that teachers reported significantly more Anxious/Shy problem behaviours than parents is that children are exposed to new experiences, challenges, or even confrontations in the classroom, which in turn provide more opportunities for anxious, withdrawn behaviours. In contrast, the child's home environment may provide fewer opportunities for such behaviours to be observed.

Consistent with other studies, the present study also found that parent and teacher agreement was highest for the Conners Cognitive Problems subscale (Deng et al., 2004). Willard et al. (2016) reported highest levels of agreement on the Conners 3 Learning Problems subscale in adolescents. As this subscale consists of items that relate to inattention, a possible reason for this finding is that cognitive problems (particularly inattention) are expected to be more stable across environments (Kazdin, 1995).

The present study found that parent and teacher agreement declined as mean scores increased. Although this finding is consistent with two other population-based studies in this area (Kolko & Kazdin, 1993; Deng et al., 2004), it is still surprising. Anecdotally, one would expect that agreement would increase as behaviour problems became more severe. It is important to take into consideration the context specificity of a child's behaviour and how this can impact the level of agreement between different raters. Others have highlighted the importance of this for both clinical and research practice (Achenbach et al., 1987; Deng et al., 2004; Kolko & Kazdin, 1993). It firstly stresses the importance of using multiple informants in measuring children's behavioural problems, and secondly it provides guidance for the clinical evaluation of children's behavioural problems (Deng et al., 2004; Kolko & Kazdin, 1993). If both a child's parent and teacher report relatively few behavioural problems, the child is unlikely to have a disorder. If both

a child's parent and teacher report higher levels of behaviour problems, further investigation is required. Moreover, if the reports for parent and teacher differ, further investigation is needed to determine the origin of the difference.

Limitations and Conclusions

The present study has several limitations. First, the CRS: R-L (Conners, 1997) is no longer published or sold and was replaced by the Conners 3 (Conners, 2008) Rating Scales. However, even though our study used an earlier version of the Conners questionnaire, our results reinforce the clinical importance of considering multi-informant perspectives when assessing children. As such, they provide a foundation for future studies on agreement between teacher and parent ratings using later versions of the Conners Rating Scales. It is notable that there are few psychometric studies using the more recent version of the Conners. Only one study has investigated the psychometric properties of the Conners 3 Parent and Teacher scales (Christiansen et al., 2016), and these researchers confirmed the original factor structure (Conners, 2008).

Second, although there are six major ethnic groups in New Zealand (European, Māori, Pacific Peoples, Asian, Middle Eastern/Latin American/African, Other), the sample was restricted to New Zealand European mothers and their children. Because of a poor response rate from children of non-European mothers at ages 1 (Phase 2) and 3.5 years (Phase 3), only participants born to European mothers were included in subsequent phases due to concerns regarding sample representativeness. The response rate for these participants was 67.9% for data used in the present study (Phase 4). Replication with a non-European sample may be useful in establishing the generalizability of our results.

This study has a number of strengths. It was empirically based, using information from multiple informants to assess each child's behaviour. It also adopted empirically based assessment procedures when using psychometrics. These include standardised procedures, the use of multiple items to measure particular aspects of behaviour, aggregated items to provide a quantitative scale for measuring each aspect of behaviour, normative scales to help comparisons of children with relevant reference groups; besides, the measure (Conners) has been tested for reliability and validity to ensure psychometric soundness (Achenbach et al., 1987).

In conclusion, the results of the current study revealed that parent and teacher agreement about behaviour problems using the Conners were poor to low. Agreement was higher for the ADHD subscales,

particularly the DSM-IV subscales, and the Cognitive problems subscale (high scores reflect children who are likely to be inattentive, have organizational problems, have difficulty completing tasks and concentration problems). The study also found that agreement decreases as mean scores increase. It is likely that multiple informants contribute additional variation, which may limit the agreement found between reports by different informants. Children with behavior disorders such as ADHD often exhibit behaviors that are situation-specific; therefore,

measures such as the CRS: R-L are likely to provide results that are inconsistent across situations. The American Academy of Paediatrics guidelines (2004) suggest that these discrepancies may be due to differences across settings in terms of expectations, levels of structure, behavioural management strategies, and/or environmental circumstances. In conclusion, the findings from our study indicate that assessment protocols are needed to overcome or identify the reasons for the contradictory parent and teacher reports of symptoms.

References

- Abt-Mörstedt, B., Corbisiero, S., Bitto, H., & Stieglitz, R.-D. (2015). Attention-deficit/hyperactivity disorder (ADHD) in adulthood: Concordance and differences between self- and informant perspectives on symptoms and functional impairment. *PLoS ONE*, *10*(11). e0141342. <https://doi.org/10.1371/journal.pone.0141342>
- Achenbach, S., & McConaughy, S. M. (1997). *Empirically based assessment of child and adolescent psychopathology – Practical applications* (2nd ed.). Sage.
- Achenbach, S., McConaughy, S. M., & Howell, C. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, *101*, 213-232. <https://doi.org/10.1037/0033-2909.101.2.213>
- American Academy of Pediatrics, Steering Committee on Quality Improvement and Management. (2004). Classifying recommendations for clinical practice guidelines. *Pediatrics*, *114*(3), 874-877. doi:10.1542/peds.2004-1260
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders, fourth edition* (DSM-IV).
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders, 5th edition* (DSM-V).
- Andrews, J., Lewinson, P., Hops, H., & Roberts, R. (1993). Psychometric properties of scales for the measurement of psychosocial variables associated with depression in adolescence. *Psychological Reports*, *73*, 1019-1046. doi:10.1177/00332941930733pt146
- Angtrop, I., Roeyers, H., Oosterlaan, J., & Van Oost, P. (2002). Agreement between parent and teacher ratings of disruptive behavior disorders in children with clinically diagnosed ADHD. *Journal of Psychopathology and Behavioral Assessment*, *24*(1), 67-73. <https://doi.org/10.1023/A:1014057325752>
- Barkley, R. A. (1998). *Attention deficit hyperactivity disorder: A handbook for diagnosis and treatment*. Guilford.
- Biederman, J., Faraone, S. V., Milberger, S., & Doyle, A. (1993). Diagnoses of attention-deficit hyperactivity disorder from parent reports predict diagnoses based on teacher reports. *Journal of the American Academy of Child & Adolescent Psychiatry*, *32*(2), 315-317. <https://doi.org/10.1097/00004583-199303000-00011>
- Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, *1*(8476), 307-310. PMID: 2868172
- Christiansen, H., Hirsch, O., Drechsler, R., Wanderer, S., Knospe, E.-L., Gunther, T., et al. (2016). German validation of the Conners 3 Rating Scales for parents, teachers, and children. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, *44*, 139-147. <https://doi.org/10.1024/1422-4917/a000408>
- Cohen-Kettenis, P., & Van-Goozen, S. (2002). Adolescents who are eligible for sex reassignment surgery: Parental reports of emotional and behavioural problems. *Clinical Child Psychology and Psychiatry*, *7*, 412-422. doi:10.1177/1359104502007003008
- Cohen, M., Becker, M., & Campbell, R. (1990). Relationships among four methods of assessment of children with attention deficit hyperactivity disorder. *Journal of School Psychology*, *28*(3), 189-202. [https://doi.org/10.1016/0022-4405\(90\)90011-U](https://doi.org/10.1016/0022-4405(90)90011-U)
- Collett, B. R., Ohan, J. L., & Myers, K. M. (2003). Ten-year review of rating scales.V: Scales assessing attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *42*(9), 1015-1037. <https://doi.org/10.1097/00004583-200310000-00006>
- Conners, K. (1997). *Conners Rating Scales – Revised technical manual*. Multi-Health Systems.
- Conners, C. K. (2008). *Conners 3rd edition: Manual*. Multi-Health Systems.
- Colomer, C., Wiener, J., & Varma, A. (2020). Do adolescents with ADHD have a self-perception bias for their ADHD symptoms and impairment. *Canadian Journal of School Psychology*, *35*(4), 238-251. <https://doi.org/10.1177/0829573520936457>
- Deng, S., Liu, X., & Roosa, M. W. (2004). Agreement between parent and teacher reports on behavioral

- problems among chinese children. *Journal of Developmental & Behavioral Pediatrics*, 25(6), 407-414. doi:10.1097/00004703-200412000-00004
- Efron, D., Bryson, H., Lycett, K., & Sciberras, E. (2016). Children referred for evaluation for ADHD: Comorbidity profiles and characteristics associated with a positive diagnosis. *Child: Care, Health and Development*, 42(5), 718-724. doi:10.1111/cch.12364
- Gagnon, C., Vitaro, F., & Tremblay, R. E. (1992). Parent-teacher agreement on kindergartners' behavior problems. *Journal of Child Psychology and Psychiatry*, 33(7), 1255-1261. <https://doi.org/10.1111/j.1469-7610.1992.tb00944.x>
- Hardy, K. K., Willard, V. W., Gioia, A., Sharkey, C., & Walsh, K. S. (2018). Attention-mediated neurocognitive profiles in survivors of pediatric brain tumors: Comparison to children with neurodevelopmental ADHD. *Neuro Oncology*, 20(5), 705-715. doi:10.1093/neuonc/nox174
- Izzo, V. A., Donati, M. A., & Primi, C. (2019). Assessing ADHD through the multi-informant approach: The contribution of the Conners' 3 Scales. *Journal of Attention Disorders*, 23(6), 641-650. doi:10.1177/1087054718815581
- Kazdin, A. (1995). *Conduct disorders in childhood and adolescence*. Sage.
- Kolko, D., & Kazdin, A. (1993). Emotional/behavioural problems in clinic and non-clinic children: Correspondence among child parent and teacher reports. *Journal of Child Psychology and Psychiatry*, 41, 253-260. doi:10.1111/j.1469-7610.1993.tb01103.x
- Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174. <https://doi.org/10.2307/2529310>
- Lee, S. W., Elliott, J., & Barbour, B. (1994). A comparison of cross-informant behavior ratings in school-based diagnosis. *Behavioral Disorders*, 19(2), 87-97. doi:10.1177/019874299401900202
- Lohaus, A., Rueth, J.-E., & Vierhaus, M. (2019). Cross-informant discrepancies and their association with maternal depression, maternal parenting stress, and mother-child relationship. *Journal of Child and Family Studies*, 29. <https://doi.org/10.1007/s10826-019-01625-z>
- Loeber, R., Green, S. M., & Lahey, B. B. (1990). Mental health professionals' perception of the utility of children, mothers, and teachers as informants on childhood psychopathology. *Journal of Clinical Child Psychology*, 19(2), 136-143. https://doi.org/10.1207/s15374424jccp1902_5
- Miller, M., Bowen, J. R., Gibson, F. L., Hand, P. J., & Ungerer, J. A. (2001). Behaviour problems in extremely low birthweight children at 5 and 8 years of age. *Child: Care, Health & Development*, 27(6), 569-581. doi:10.1046/j.1365-2214.2001.00223.x
- Mitsis, E. M., McKay, K. E., Schulz, K. P., Newcorn, J. H., & Halperin, J. M. (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(3), 308-313. doi:10.1097/00004583-200003000-00012
- Nass, R. D. (2005). Evaluation and assessment issues in the diagnosis of attention deficit hyperactivity disorder. *Seminars in Pediatric Neurology*, 12(4), 200-216. doi:10.1016/j.spn.2005.12.002
- Nottelmann, E. (2002). Competence and self-esteem during transition from childhood to adolescence. *Developmental Psychology*, 23, 441-450. <https://doi.org/10.1037/0012-1649.23.3.441>
- Orylska, A., Brzezicka, A., Racicka-Pawlukiewicz, E., Albinski, R., & Sedek, G. (2016). Parent-teacher concordance in rating preschooler difficulties in behavioural and cognitive functioning and their dyadic predicting of fluid intelligence. *Polish Psychological Bulletin*, 47(1), 81-91. doi:10.1515/ppb-2016-0009
- Penninx, B. W. J. H., Beekman, A. T. F., Smit, J. H., Zitman, F. G., Nolen, W. A., Spinhoven, P., et al. (2008). The Netherlands study of depression and anxiety (NESDA): Rationale, objectives and methods. *International Journal of Methods in Psychiatric Research*, 17(3), 121-140. <https://doi.org/10.1002/mpr.256>
- Reid, R., & Maag, J. W. (1994). How many fidgets in a pretty much: A critique of behavior rating scales for identifying students with ADHD. *Journal of School Psychology*, 32(4), 339-354. [https://doi.org/10.1016/0022-4405\(94\)90032-9](https://doi.org/10.1016/0022-4405(94)90032-9)
- Stratis, E. A., & Lecavalier, L. (2017). Predictors of parent-teacher agreement in youth with autism spectrum disorder and their typically developing siblings. *Journal of Autism and Developmental Disorders*, 47(8), 2575-2585. doi:10.1007/s10803-017-3173-6
- Thompson, J. M., Clark, P. M., Robinson, E., Becroft, D. M., Pattison, N. S., Glavish, N., Pryor, J. E., Wild, C. J., Rees, K., & Mitchell, E. A. (2001). Risk factors for small-for-gestational-age babies: The Auckland Birthweight Collaborative Study. *Journal of Paediatrics & Child Health*, 37(4), 369-375. doi:10.1046/j.1440-1754.2001.00684.x
- Thompson, J. M., Mitchell, E. A., & Borman, B. (1994). Sex specific birthweight percentiles by gestational age for New Zealand. *New Zealand Medical Journal*, 107(970), 1-3. PMID:8295745
- World Health Organization. (2004). *World health report 2004: Changing history*.
- Youngstrom, E., Loeber, R., & Stouthamer-Loeber, M. (2000). Patterns and correlates of agreement between parent, teacher, and male adolescent ratings of externalizing and internalizing problems. *Journal of Consulting & Clinical Psychology*, 68(6), 1038-1050. doi:10.1037/0022-006x.68.6.103