

# Assessment and Etiology of Attention Deficit Hyperactivity Disorder and Oppositional Defiant Disorder in Boys and Girls

Eske M. Derks · Conor V. Dolan · Jim J. Hudziak ·  
Michael C. Neale · Dorret I. Boomsma

Published online: 19 April 2007  
© Springer Science+Business Media, LLC 2007

**Abstract** Attention deficit hyperactivity disorder (ADHD) and oppositional defiant disorder (ODD) are more common in boys than girls. In this paper, we investigated whether the prevalence differences are attributable to measurement bias. In addition, we examined sex differences in the genetic and environmental influences on variation in these behaviors. Teachers completed the Conners Teacher Rating Scale-Revised:Short version (CTRS-R:S) in a sample of 800 male and 851 female 7-year-old Dutch twins. No sex differences in the factor structure of the CTRS-R:S were found, implying the absence of measurement bias. The heritabilities for both ADHD and ODD were high and were the same in boys and girls. However, partly different genes are expressed in boys and girls.

**Keywords** ADHD · ODD · Sex differences · Conners Teacher Rating Scale · Measurement invariance

---

Edited by Stacey Cherny

---

E. M. Derks (✉) · M. C. Neale · D. I. Boomsma  
Department of Biological Psychology, Vrije Universiteit,  
Amsterdam, The Netherlands  
e-mail: em.derks@psy.vu.nl

C. V. Dolan  
Department of Psychology, University of Amsterdam,  
Amsterdam, The Netherlands

J. J. Hudziak  
Department of Psychiatry and Medicine, Division of Human  
Genetics, Center for Children, Youth and Families, College of  
Medicine, University of Vermont, Burlington, USA

M. C. Neale  
Department of Psychiatry and Human Genetics, Virginia  
Institute for Psychiatric and Behavioral Genetics, Virginia  
Commonwealth University, Richmond, USA

In both clinical and population samples, children diagnosed with attention deficit hyperactivity disorder (ADHD) and oppositional defiant disorder (ODD) are predominantly male (Gaub and Carlson 1997; Biederman et al. 2002; Loeber et al. 2000). These sex differences could either be the result of a higher liability for these disorders in boys than girls, or could be attributable to a sex effect in the actual measurement of the phenotype. For example, boys and girls with similar levels of problem behavior may receive systematically different rating scores if the items of the instrument do not reflect the problem behaviors in the same manner in boys and girls. This may be conceptualized as bias with respect to sex. In this study, we will investigate this issue in the measurement of ADHD and ODD by means of the Conners Teacher Rating Scale-Revised:Short (CTRS-R:S) version.

The presence of sex differences in the prevalence of ADHD and ODD also raises the question whether there are sex differences in the etiology of these disorders. Sex differences in etiology can only be interpreted if the measurement of a disorder is not biased with respect to sex. Lubke et al. (2004) discussed in detail the importance of establishing unbiasedness with respect to sex for the correct interpretation of any sex differences in the results of genetic modeling. The aim of this paper is, therefore, to investigate measurement bias in teacher ratings of ADHD and ODD with respect to sex. If unbiasedness can be shown to be tenable to reasonable approximation, we shall, as the second goal of this paper, estimate the genetic and environmental contributions to the phenotypic variance in ADHD and ODD in boys and girls.

Mellenbergh (1989) defined unbiasedness, or equivalently, measurement invariance (MI), with respect to group to mean that the distribution of the observed test score, conditional on the latent construct that the test measures, is

identical over groups (e.g., boys and girls). In more simple terms, this means that the instrument is measuring the same construct in boys and girls (Mellenbergh 1989; Meredith 1993). If this is the case, we expect the score of a given person to depend on that person's score on the latent construct, but not on that person's sex. If MI does not hold, a boy and a girl, who are characterized by the same degree of problem behavior, may obtain systematically (i.e., regardless of measurement error) different scores on the instrument. This is undesirable, because we wish our measurements to reflect accurate and interpretable differences between cases in different groups.

Analyses aimed explicitly at establishing MI with respect to sex, according to the approach outlined by Meredith (1993), have yet to be conducted with respect to ADHD and ODD. Although MI has not been investigated, there have been some studies, which addressed sex differences in the factor structure in teacher ratings of ADHD. Fantuzzo et al. (2001) examined the factor structure of the 28-item version of the CTRS with exploratory factor analyses. They reported a three-factor solution, which accounted for 58% of the variance. The factors admitted interpretations in terms of conduct, hyperactivity, and passivity. The invariance of the factor structure was established by comparing results from random subgroups with the results from subgroups based on sex. The results supported the similarity of the factor structure across sex. A concern in this study is that the subjects in this study were 580 children from 33 classrooms located in low-income neighborhoods. It is, therefore, unclear how representative the sample is of the general population. Furthermore, because each teacher rated more than one child, and children were clustered into classes, the assumption of independent observations, which is important in statistical inference, might not hold.

In this study, we conducted confirmatory factor analyses (CFA) of data from a large general population sample of 7-year-old twins, rated by their teachers on ADHD and ODD. Two questions are addressed. First is the measurement model that relates differences in the latent constructs of ADHD and ODD to the observed behavior problem scores identical in boys and girls, i.e., is MI tenable? Second, do the magnitudes of the genetic and environmental influences differ, or do different genes play a role in boys and girls?

## Methods

### Subjects and procedure

This study is part of an ongoing twin study of development and psychopathology in the Netherlands. The subjects were all registered at the Netherlands Twin Registry (NTR;

Boomsma et al. 2002, 2006). We assessed a sample of Dutch twins from the birth cohorts 1992–1996. These twins were assessed by their teachers when they were 7 years old.

The twins at age 3 are representative of Dutch 3-year-old children with respect to their scores on measures such as the Child Behavior Checklist (CBCL; van den Oord et al. 1995). The socioeconomic status of the parents of the twins was somewhat higher than the level in the general Dutch population (Rietveld et al. 2004). When twins reached the age of 7 years, parents were asked to provide informed consent to approach the teacher. Consent was given by 80.1% of the parents. Teachers of these pairs received a questionnaire by mail, and were asked to return it to the NTR by mail. The response rate of the teachers was 80.0%. CTRS data were available for at least one twin in 1,651 twin-pairs (1,511 complete pairs).

The maternal CBCL-AP scores at age 7 years were not significantly different between families in which parents provided permission to approach the teachers (mean = 2.95, SD = 2.93) and families in which parents did not do so (mean = 3.15, SD = 3.18;  $t(3,063) = 2.0$ ,  $p = 0.133$ ). However, mean maternal AP ratings were significantly higher in twins whose teachers did not return the questionnaire (mean = 3.34; SD = 3.13) than in twins whose teachers did return the questionnaire (mean = 2.78; SD = 2.81;  $F(1) = 16.82$ ,  $p < 0.001$ ).

To avoid biased test results due to statistical dependency of the twin data, we randomly included the score from only one of the members of a twin-pair in the CFA. The resulting sample for CFA consists of 1,651 individual twins (800 boys and 851 girls). In the genetic analyses, we included all complete twin-pairs. Data were available for both members of a twin-pair in 248 MZ male, 251 DZ male, 294 MZ female, 234 DZ female, and 484 DZ opposite sex pairs. Some twins were rated by the same teacher (877 pairs, 58%) while the remaining twins were rated by different teachers (634 pairs, 42%). The genetic analyses accounted for potential differences between same and different teacher ratings by using the model developed by Simonoff et al. (1998). Incomplete twin-pairs were excluded from the genetic analyses.

### Measure

The CTRS-R is a widely used instrument to assess behavior problems (Conners 2001; Conners et al. 1998). The CTRS-R was developed by factor analyzing a large set of items, and including items that load highly on interpretable common factors. In addition to the scales that were derived based on factor analysis, an ADHD index was included. This index comprises the best 12 items for distinguishing children with ADHD from children without

ADHD as assessed by the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; American Psychiatric Association 1994; Conners 2001). The short version of the CTRS-R, which was used in this study, contains 28 items from the following scales: oppositional (ODD, five items); cognitive problems-inattention (IN, five items); hyperactivity (HI, seven items); and the ADHD index (ADHD-I, 12 items). One of the items (item 27; excitable, impulsive) is included in both the hyperactivity scale and the ADHD index. The items are rated on a 4-point Likert scale for symptom severity (i.e., 0 = "not true at all"; 1 = "just a little true"; 2 = "pretty much true"; 3 = "very much true"). The internal consistency and stability of the CTRS-R:S version are good, as the Cronbach's alpha coefficients range from 0.88 to 0.95, and 6–8-week test–retest correlations range from 0.72 to 0.92 (Conners 2001).

#### Distribution of the items

Because of the categorical nature of the item scores, Pearson product moment correlations underestimate the correlation of the underlying latent trait, and consequently the parameter estimates obtained in factor analyses or principal component analyses based on the Pearson correlation matrices are biased (Dolan 1994). We, therefore, adopted an approach that is suitable for factor analyzing discrete item scores. Polychoric correlations between items were obtained based on the liability threshold model (Lynch and Walsh 1998). In the case of a 4-point Likert scale, three thresholds divide the latent liability distribution into four categories.

#### Criteria of MI

The criteria of MI are empirically testable in the common factor model (Meredith 1993). MI criteria are: (1) equality of factor loadings over groups; (2) equality of item intercepts over groups (i.e., differences in item means are the result of differences in factor means), and (3) equality of residual variances (i.e., variance in the observed variables, not explained by the common factor) over groups. When satisfied, these restrictions ensure that any differences in the mean and variance of the observed variables are due to differences in the mean and variance of the common factor.

For ordinal data, MI can be tested by constraining the thresholds and factor loadings, and residual variances to be equal in boys and girls. These constraints allow estimation of group differences in the means and variances of the common factor. To this end, the mean and variance of the common factor are fixed at 0 and 1, respectively, in an arbitrary reference group. We chose to estimate the mean and variance in girls, and to use boys as the reference

group. In doing so, we are modeling the observed group differences as a function of the differences in the latent liability.

#### Statistical analyses

The polychoric correlation matrices of the items of the four subscales were calculated using Prelis (Jöreskog and Sörbom 1996). All CFA were performed on raw data using Mx (Neale et al. 2003). In principle, factor analysis of  $p$  ordinal items can be carried out using full information maximum likelihood (FIML) estimation. However, this requires repeated integration of the  $p$ -variate normal distribution, which can become computationally demanding even with as few as 12 items. We, therefore, estimated the model parameters using marginal maximum likelihood estimation (MML; Bock and Aitkin 1981). MML maximizes the likelihood of the data conditional on the latent trait, in contrast to FIML, which maximizes the unconditional likelihood. The advantage of MML compared to FIML is that it is computationally much less demanding.

To test if the Conners scales are MI with respect to sex, the factor structure was constrained to be identical in boys and girls. The ODD, IN, and HI scales resulted from factor analyses, and a single factor was fit to these scales. In contrast, the ADHD-I contains items related to problems with inattention and hyperactivity, and thus a two-factor model was fitted. To detect prevalence differences in ADHD and ODD across sex, the means and variances of the latent factors were constrained to be equal in boys and girls.

The fit of the models was compared by  $\chi^2$  tests, with a type I error probability set at  $\alpha = 0.01$ . Browne et al. (2002) noted a complication of the  $\chi^2$  test. Specifically, they showed that  $\chi^2$  is influenced by the unique variances of the items. If a trait is measured reliably, the inter-correlations of the items are high, the unique variances are small, and the  $\chi^2$  test may suggest a poor fit even when the differences between the expected and observed covariance matrices are trivial. The standardized root mean square residual (SRMR; Bentler 1995) is a fit index that is not sensitive to the size of the correlations. To avoid the rejection of a simpler model due to high inter-item correlations, we only reject a model if a significant  $\chi^2$  test is accompanied by large residuals (SRMR > 0.08; Hu and Bentler 1999).

After investigating MI with respect to sex, we look at sex differences in the genetic and environmental influences on the individual differences in the sum scores of the scales, given that MI is tenable (Lubke et al. 2004). The polychoric correlations of the four scales were calculated by sex and zygosity in PRELIS (Jöreskog and Sörbom 1996). The genetic analyses were performed on the raw data using Mx (Neale et al. 2003).

With data from MZ and DZ twins, the variance in behavior can be attributed to genetic and environmental factors. In our sample, 58% of the twins were in the same classroom and 42% of the twins were in different classrooms. Correlations between twins may be higher when children are rated by the same teacher. Therefore, a correlated error model developed by Simonoff et al. (1998) was used to analyze the data. In this model, individual differences in behavior are explained by additive genetic factors (A), common environmental factors that are shared between two twins of a pair (C), and unique environmental factors (E).

The unique environmental factors are allowed to correlate in twins who are placed into the same classroom, and do not correlate in twins who are placed into different classrooms. For the genetic analyses, the items of each subscale were summed, and the sum-score was recoded so that three thresholds divide the latent liability distribution into four categories. The thresholds were chosen in such a way that the categories contain more or less equal numbers of subjects. We preferred this procedure to the analysis of the raw sum scores, because these are skewed, and therefore cannot be analyzed with maximum likelihood approaches based on the assumption of normality (Derks et al. 2004).

Sex differences in genetic and environmental influences were examined in two ways. First, we investigated if the estimates of the genetic and environmental variances are equal in boys and girls. Secondly, we investigated if the same genes influence phenotypic variation in boys and girls. These qualitative sex differences were evaluated by constraining the genetic correlation in opposite sex twins at 0.5 (similar as in same-sex DZ twins). If different genes play a role in boys and girls, the genetic correlation is expected to be lower than 0.5 in opposite sex twins.

## Results

### Measurement invariance

We tested for MI by constraining the factor loadings, thresholds, and residual covariance matrices to be equal for boys and girls while allowing the factor means and variances to be different. The factor structure of ODD was MI with respect to sex ( $\chi^2(18) = 16.66$ ,  $p > 0.10$ ; SRMR = 0.01 and 0.06 in boys and girls, respectively). MI was also tenable for the ADHD-I ( $\chi^2(55) = 70.41$ ,  $p > 0.05$ ; SRMR = .03 and 0.05 in boys and girls, respectively). Both IN and HI showed statistically significant different factor structures in boys and girls ( $\chi^2(18) = 98.45$ ,  $p < .001$ , and

$\chi^2(26) = 57.99$ ,  $p < 0.001$ , respectively). However, the residuals between expected correlation matrices under the constrained and the non-constrained model were small (SRMR = 0.01 in girls and SRMR = 0.02 in boys, for both IN and HI). Apparently, the lack of fit was the result of the high inter-correlations between the items (Browne et al. 2002), and not of large residuals between the expected covariance matrices. Therefore, we tentatively accept MI with respect to sex. Table 1 provides the factor loadings and thresholds of the best-fitting models. We also included the factor loadings as reported by Conners (2001) to facilitate the comparison of our sample with the sample that was used to create the scales. Note that the factor loadings, as reported by Conners (2001), are generally much lower, as these estimates are based on the assumption of a continuous, normal distribution of the item scores, an assumption that is obviously violated in the instance of a four-category Likert scale.

### Genetic analyses

Having established MI of the CTRS-R:S scales with respect to sex, we estimated the twin-correlations and carried out a genetic analysis of the data. Twin correlations are shown in Table 2, for same and different teachers, respectively. The genetic model fitting results of the four scales are reported in Table 3. All correlations are higher in MZ twins than in DZ twins, suggesting the presence of genetic influences on individual differences. The correlations are higher in twin pairs rated by the same teacher than for twin pairs rated by different teachers. This was taken into account by using a correlated error model (Simonoff et al. 1998). The lower correlations in opposite-sex DZ twins than in same-sex DZ twins suggest that different genes play a role in boys and girls.

Model fitting analyses showed that variation in all four scales could be explained by additive genetic and unique environmental effects. The influences of the shared environment were not statistically significant. The magnitude of the influences of genes and environment did not differ between boys and girls. The standardized estimates of genetic and environmental influences are shown in Table 4. Genetic effects explained 56–71% of the variation in the CTRS subscales. Unique environmental effects explained the remaining 29–44% of the variation. For all four scales, the genetic correlation was significantly lower than 0.5 in opposite-sex twins. This implies that different genes are expressed in males and females. The genetic correlation in opposite-sex twins was 0.16 for oppositional behavior, 0.35 for cognitive problems-inattention, 0.21 for hyperactivity-impulsivity, and 0.32 for the ADHD index.

**Table 1** Promax rotated factor loadings and thresholds (T) of the best-fitting factor model for the four subscales of the CTRS-R:S

	Items	Loadings factor 1	Loadings factor 2	Loadings as reported by Conners (2001) ( <i>N</i> = 1,897)	T1	T2	T3
Oppositional	2. Defiant	0.91	—	0.48	0.71	1.65	2.45
	6. Defies	0.88	—	0.50	1.02	2.00	2.74
	10. Spiteful	0.69	—	0.30	1.49	2.32	3.13
	15. Argues	0.84	—	0.51	0.82	1.80	2.52
	20. Explosive	0.78	—	0.41	1.20	1.92	3.12
Cognitive problems- Inattention	4. Forgets things	0.93	—	0.55	0.34	1.29	1.82
	8. Poor spelling	0.81	—	0.50	0.27	0.95	1.44
	13. Poor reading	0.71	—	0.45	0.39	0.87	1.21
	18. Lacks interest	0.70	—	0.47	1.01	1.65	2.27
Hyperactivity	22. Poor arithmetic	0.82	—	0.49	0.55	1.22	1.68
	3. Restless	0.73	—	0.62	0.10	0.94	1.51
	7. Always on the go	0.79	—	0.57	0.66	1.31	1.87
	11. Leaves seat	0.82	—	0.58	0.62	1.37	1.83
	17. Difficulty waiting	0.83	—	0.65	0.12	0.91	1.58
	21. Runs about	0.79	—	0.44	1.21	1.89	2.51
	24. Difficulty playing	0.87	—	0.52	0.34	1.17	2.03
	27. Excitable	0.86	—	0.66	0.40	1.12	1.84
ADHD index							
AP	1. Inattentive	0.99	−0.09	—	−0.36	0.76	1.47
	14. Attention span	0.95	−0.03	—	0.03	0.87	1.55
	16. Only pays attention	0.45	0.23	—	0.36	1.20	1.88
	19. Distractibility	0.94	−0.01	—	0.59	1.54	2.25
	25. Fails to finish	0.72	0.09	—	0.24	1.27	1.99
	26. Not follow instructions	0.68	0.09	—	0.38	1.30	1.86
HI	5. Disturbs	0.05	0.80	—	0.20	1.03	1.52
	9. Remain still	0.05	0.88	—	0.92	1.68	2.24
	12. Fidgets	0.16	0.67	—	0.12	0.94	1.53
	23. Interrupts	−0.05	0.85	—	0.54	1.37	2.07
	27. Excitable	−0.09	0.91	—	0.46	1.15	1.83
	28. Restless	−0.06	0.97	—	0.50	1.19	1.81

AP attention problems; HI hyperactivity

The thresholds are constrained to be equal in boys and girls. The means and variances of the latent factors in boys are constrained at 0 and 1, respectively. The means and variances of the latent factors are freely estimated in girls. The variances in girls were not significantly different from 1. The means of the latent factors are estimated at −0.60 (oppositional), −0.86 (hyperactivity), −0.42 (ADHD-I AP), and −0.52 (ADHD-I-HI). The mean of the cognitive problems-inattention factor is not significantly different between boys and girls

**Table 2** Polychoric twin-correlations of the CTRS rated by same teachers (ST) versus different teachers (DT)

	Oppositional		Cognitive problems-inattention		Hyperactivity		ADHD index	
	ST	DT	ST	DT	ST	DT	ST	DT
Monozygotic males	0.86	0.53	0.90	0.76	0.81	0.59	0.85	0.59
Dizygotic males	0.50	0.49	0.64	0.29	0.42	0.14	0.49	0.19
Monozygotic females	0.87	0.43	0.92	0.64	0.83	0.47	0.87	0.52
Dizygotic females	0.66	0.10	0.60	0.42	0.34	0.25	0.45	0.24
Opposite-sex twins	0.37	0.13	0.44	0.23	0.30	0.17	0.37	0.27



**Table 3** Genetic model fitting results on CPRS-R:S ratings

Model	Parameters	-2 LL	With model	$\Delta$ df	$\Delta\chi^2$	<i>p</i>
Oppositional behavior in 7-year-old children						
1. Fully saturated	84	4779.36	—	—	—	—
2. Equal correlations same and different teachers	78	4812.26	1	6	32.91	<0.001
3. ACE model, estimates of ACE are allowed to be different between same and different teachers	44	4817.94	1	40	38.58	0.534
4. ACE, ACE same teacher = ACE different teacher	40	4853.84	3	4	35.90	<0.001
5. Correlated errors ACE model	42	4819.80	1	42	40.44	0.540
6. Correlated errors AE model	40	4823.93	5	2	4.13	0.127
7. Correlated errors AE model, AE boys = AE girls	38	4829.76	6	2	5.83	0.054
8. Correlated errors AE model, AE boys = AE girls, opposite-sex genetic correlation-free	37	4821.86	7	1	7.90	0.005
Cognitive problems-inattention in 7-year-old children						
1. Fully saturated	84	6785.90	—	—	—	—
2. Equal correlations same and different teachers	78	6822.57	1	6	36.67	<0.001
3. ACE model, estimates of ACE are allowed to be different between same and different teachers	44	6815.94	1	40	30.04	0.874
4. ACE, ACE same teacher = ACE different teacher	40	6843.87	3	4	27.93	<0.001
5. Correlated errors ACE model	42	6815.85	1	42	29.95	0.918
6. Correlated errors AE model	40	6820.01	5	2	4.16	0.125
7. Correlated errors AE model, AE boys = AE girls	38	6821.99	6	2	1.98	0.372
8. Correlated errors AE model, AE boys = AE girls, opposite-sex genetic correlation-free	39	6817.50	7	1	4.49	0.034
Hyperactivity in 7-year-old children						
1. Fully saturated	84	6880.07	—	—	—	—
2. Equal correlations same and different teachers	78	6906.97	1	6	26.90	<0.001
3. ACE model, estimates of ACE are allowed to be different between same and different teachers	44	6914.59	1	40	34.42	0.715
4. ACE, ACE same teacher = ACE different teacher	40	6942.35	3	4	27.76	<0.001
5. Correlated errors ACE model	42	6924.12	1	42	44.05	0.385
6. Correlated errors AE model	40	6926.76	5	2	2.64	0.267
7. Correlated errors AE model, AE boys = AE girls	38	6929.06	6	2	2.30	0.317
8. Correlated errors AE model, AE boys = AE girls, opposite sex genetic correlation-free	37	6919.99	7	1	9.07	0.003
ADHD in 7-year-old children						
1. Fully saturated	84	7420.21	—	—	—	—
2. Equal correlations same and different teachers	78	7458.10	1	6	37.89	<0.001
3. ACE model, estimates of ACE are allowed to be different between same and different teachers	44	7472.59	1	40	52.38	0.091
4. ACE, ACE same teacher = ACE different teacher	40	7507.40	3	4	34.81	<0.001
5. Correlated errors ACE model	42	7477.78	1	42	57.57	0.055
6. Correlated errors AE model	40	7479.14	5	2	1.36	0.507
7. Correlated errors AE model, AE boys = AE girls	38	7480.72	6	2	1.58	0.454
8. Correlated errors AE model, AE boys = AE girls, opposite-sex genetic correlation-free	37	7475.93	7	1	4.79	0.029

## Discussion

The purpose of this study was twofold. First, we investigated if teacher ratings on ADHD and ODD are

measurement invariant with respect to sex. Secondly, genetic and environmental influences on variation in ADHD and ODD were compared between boys and girls.

**Table 4** Standardized estimates of the genetic and environmental effects on problem behavior

	% Additive genetic effects (A)	% Non-shared environmental effects (E)	Genetic correlation opposite-sex twins
Oppositional	56	44	0.16
Cognitive problems-inattention	71	29	0.35
Hyperactivity	58	42	0.21
ADHD index	61	39	0.32

### Measurement invariance

Teacher ratings on ADHD and ODD were found to be measurement invariant with respect to sex. In other words, teacher assessments of these behavior problems relate to the same latent variables in boys and girls. Sex differences in observed scores on ADHD and ODD can, therefore, be interpreted as differences with respect to the latent construct. This supports the contention that the reported sex differences in ADHD and ODD (Gaub and Carlson 1997; Loeber et al. 2000; Maughan et al. 2004) are due to a higher liability for the disorder in boys than girls and not to measurement bias.

### Quantitative and qualitative differences in the heritability among boys and girls

More than half the variance in ADHD and ODD in boys and girls is attributable to genetic influences. The remaining variance is attributable to unique environmental influences. The magnitude of the influences of genes and environment is the same in boys and girls. However, part of the variance in ADHD and ODD is attributable to different genes in boys and girls. We base this on the fact that the genetic correlation between DZ opposite-sex twins was significantly lower than 0.5, which is the theoretical value (in the absence of assortative mating), if the same genes influence behavior in boys and girls. We observed a genetic correlation lower than 0.5 in DZ opposite-sex twins for oppositional behavior, cognitive problems-inattention, hyperactivity, and the ADHD index.

Few studies have addressed quantitative and qualitative sex differences in heritability estimates from teacher ratings. Saudino et al. (2005) reported qualitative sex differences in heritability of teacher ratings of hyperactive behavior in 7-year-old twins. They did not report any quantitative sex differences, which is in agreement with the current findings. Vierikko et al. (2004) report lower

correlations in opposite-sex twins than in same-sex DZ twins of teacher ratings of hyperactivity-impulsivity in 12-year-old twins. However, both genetic and shared environmental effects were found to contribute to the phenotypic variance in these data. It was not possible to determine if the lower opposite-sex correlations were the result of sex-specific genetic influences or sex-specific shared environmental influences, although the presence of the latter appeared more likely. These findings disagree with the current results in the sense that we did not find any evidence for shared environmental influences. However, both studies suggest that teacher ratings are influenced by partly different etiological factors in boys and girls.

The finding of different genetic influences in boys and girls in teacher ratings stands in contrast with results based on parental reports. In parent ratings, qualitative sex differences are not found for attention problems (Rietveld et al. 2004) or ODD (Hudziak et al. 2005). The different findings in parent and teacher ratings may be explained by the fact that the behavior of children depends on the context in which they are observed. Apparently, inattentive, hyperactive, and oppositional behavior of boys and girls are influenced by partly non-overlapping factors at school, while this is not true for these behaviors at home.

The finding of sex-specific genetic variation has implications for gene-finding studies of ADHD and ODD. The fact that the genes which influence the behavior of boys and girls do not completely overlap indicates that some quantitative trait loci may explain variation in boys but not in girls and vice versa. Therefore, the data from boys and girls cannot be collapsed when studying genetic effects in teacher ratings.

In the NTR, teacher data are collected at the ages 7, 10, and 12 years. The sample sizes at the ages 10 and 12 are currently relatively small. In the future, we plan to address the issue of qualitative sex differences in teacher ratings in a longitudinal framework. The results of such a study will reveal if the finding of sex differences in the specific genes that play a role is also present in older children. Another important issue that may be addressed is the MI of ADHD with respect to age.

The results of this study should be interpreted in the light of the following limitations. First, we did not replicate the factor structure of the CTRS-R:S by means of exploratory factor analyses of the 28 items. To take the ordinal nature of the data into account, we used the liability threshold model (Lynch and Walsh 1998). We limited the number of common factors to keep the computational burden manageable. Therefore, we performed CFA, in which we assumed that the items are correctly assigned to the four scales and that cross-loadings are absent. Second, teacher ratings were shown to be measurement invariant with respect to sex, but this finding may not generalize to

parent ratings. The correlations between Conners parent and teacher ratings are small to moderate with a range 0.18–0.52 (Conners 2001). It has been shown that parents and teachers rate partly different aspects of the child's behavior (Derks et al. in press; Martin et al. 2002). Future studies will reveal if MI is also tenable in parent ratings.

Assessment of ADHD and ODD symptoms, through teacher reports on the CTRS-R:S, provides a solid starting point for measuring sex differences in mean scores or in heritabilities. Variation in teacher ratings of children's problem behavior is mainly influenced by genetic factors. The size of the genetic influences does not depend on the child's sex, but partly different genes are expressed in boys and girls. Future studies should reveal if these findings generalize to children from different age groups.

**Acknowledgments** This work was supported by NWO Grant numbers 575-25-006, 575-25-012, and 904-57-94 (Boomsma, P.I.), NIMH Grant number MH58799 (Hudziak, P.I.), NIH grant MH-65322 (Neale, P.I.), and by the Centre for Neurogenomics and Cognition Research (CNCR) of the Vrije Universiteit, Amsterdam.

## References

- American Psychiatric Association (1994) Task force on DSM-IV. Diagnostic and statistical manual of mental disorder: DSM-IV, 4th edn. American Psychiatric Association, Washington, DC
- Bentler PM (1995) EQS structural equations program manual. Multivariate Software, Encino, CA
- Biederman J, Mick E, Faraone SV, Braaten E, Doyle A, Spencer T et al (2002) Influence of gender on attention deficit hyperactivity disorder in children referred to a psychiatric clinic. *Am J Psychiatry* 159:36–42
- Bock RD, Aitkin M (1981) Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika* 46:443–459
- Boomsma DI, Geus de EJC, Vink JM, Stubbe JH, Distel MA, Hottenga JJ, Posthuma D, van Beijsterveldt CEM, Hudziak JJ, Bartels M, Willemsen G (2006) Netherlands Twin Register: from twins to twin families. *Twin Res Hum Genet* 9(6):849–857
- Boomsma DI, Vink JM, van Beijsterveldt CEM, Geus EJC, Beem AL, Mulder EJCM et al (2002) Netherlands Twin Register: a focus on longitudinal research. *Twin Res* 5:1–6
- Browne MW, MacCallum RC, Kim C, Andersen BL, Glaser R (2002) When fit indices and residuals are incompatible. *Psychol Methods* 7:403–421
- Conners CK (2001) Conners' rating scales-revised. Multi-Health Systems, Inc., New York
- Conners CK, Sitarenios G, Parker JDA, Epstein JN (1998) Revision and restandardization of the Conners teacher rating scale (CTRS-R): factor structure, reliability, and criterion validity. *J Abnormal Child Psychol* 26:279–291
- Derks EM, Dolan CV, Boomsma DI (2004) Effects of censoring on parameter estimates and power in genetic modeling. *Twin Res* 7:659–669
- Derks EM, Hudziak JJ, van Beijsterveldt CEM, Dolan CV, Boomsma DI (in press) Genetic analyses of maternal and teacher ratings on attention problems in 7-year-old Dutch twins. *Behav Genet*
- Dolan CV (1994) Factor analysis of variables with 2, 3, 5 and 7 response categories: a comparison of categorical variable estimators using simulated data. *Br J Math Stat Psychol* 47:309–326
- Fantuzzo J, Grim S, Mordell M, McDermott P, Miller L, Coolahan K (2001) A multivariate analysis of the revised Conners teacher rating scale with low-income, urban preschool children. *J Abnormal Child Psychol* 29:141–152
- Gaub M, Carlson CL (1997) Gender differences in ADHD: a meta-analysis and critical review. *J Am Acad Child Adolescent Psychiatry* 36:1036–1045
- Hu L, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Eq Model* 6:1–55
- Hudziak JJ, Derks EM, Althoff RR, Copeland WM, Boomsma DI (2005) The genetic and environmental contributions to oppositional defiant behavior: a multi-informant twin study. *J Am Acad Child Adolescent Psychiatry* 44:907–914
- Jöreskog KG, Sörbom D (1996) PRELIS 2 user's reference guide. Scientific Software International, Chicago
- Levy F, Hay DA, Bennett KS, McStephen M (2005) Gender differences in ADHD subtype comorbidity. *J Am Acad Child Adolescent Psychiatry* 44:368–376
- Loeber R, Burke JD, Lahey BB, Winters A, Zera M (2000) Oppositional defiant and conduct disorder: a review of the past 10 years, part I. *J Am Acad Child Adolescent Psychiatry* 39:1468–1484
- Lubke GH, Dolan CV, Neale MC (2004) Implications of absence of measurement invariance for detecting sex limitation and genotype by environment interaction. *Twin Res* 7:292–298
- Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer, Sunderland, MA
- Martin N, Scourfield J, McGuffin P (2002) Observer effects and heritability of childhood attention-deficit hyperactivity disorder symptoms. *Br J Psychiatry* 180:260–265
- Maughan B, Rowe R, Messer J, Goodman R, Meltzer H (2004) Conduct disorder and oppositional defiant disorder in a national sample: developmental epidemiology. *J Child Psychol Psychiatry* 45:609–621
- Mellenbergh GJ (1989) Item bias and item response theory. *Int J Educ Res* 13:127–143
- Meredith W (1993) Measurement invariance, factor-analysis and factorial invariance. *Psychometrika* 58:525–543
- Neale MC, Boker SM, Xie G, Maes HH (2003) Mx: statistical modeling, 6th edn. Department of Psychiatry, Richmond, VA
- van den Oord EJCG, Koot HM, Boomsma DI, Verhulst FC, Orlebeke JF (1995) A twin-singleton comparison of problem behaviour in 2–3-year-olds. *J Child Psychol Psychiatry* 36:449–458
- Rietveld MJH, Hudziak JJ, Bartels M, van Beijsterveldt CEM, Boomsma DI (2004) Heritability of attention problems in children: longitudinal results from a study of twins, age 3 to 12. *J Child Psychol Psychiatry* 45:577–588
- Saudino KJ, Ronald A, Plomin R (2005) The etiology of behavior problems in 7-year-old twins: substantial genetic influence and negligible shared environmental influence for parent ratings and ratings by same and different teachers. *J Abnormal Child Psychol* 33:113–130
- Simonoff E, Pickles A, Hervas A, Silberg JL, Rutter M, Eaves L (1998) Genetic influences on childhood hyperactivity: contrast effects imply parental rating bias, not sibling interaction. *Psychol Med* 28:825–837
- Vierikko E, Pulkkinen L, Kaprio J, Rose RJ (2004) Genetic and environmental influences on the relationship between aggression and hyperactivity-impulsivity as rated by teachers and parents. *Twin Res* 7:261–274