Genetic analyses of maternal and teacher ratings on attention problems in 7-year-old Dutch twins

Derks, E.M.; Hudziak, J.; van Beijsterveldt, C.E.M.; Dolan, C.V.; Boomsma, D.I.

published in
Behavior Genetics
2006

DOI (link to publisher)
10.1007/s10519-006-9084-5

document version
Publisher's PDF, also known as Version of record

Behav Genet (2006) 36:833–844
DOI 10.1007/s10519-006-9084-5

ORIGINAL PAPER

Genetic Analyses of Maternal and Teacher Ratings on Attention Problems in 7-year-old Dutch Twins

E. M. Derks · J. J. Hudziak · C. E. M. Van Beijsterveldt · C. V. Dolan · D. I. Boomsma

Received: 21 September 2005 / Accepted: 12 May 2006 / Published online: 14 June 2006

Abstract The goal of the present study is to examine genetic and environmental influences on maternal and teacher ratings of Attention Problems (AP) in 7-year-old children. Teachers completed the Teacher Report Form (N = 2259 pairs), and mothers the Child Behavior Checklist (N = 2057 pairs). Higher correlations were found in twins rated by the same teacher than in twins rated by different teachers. This can be explained by rater bias or by a greater environmental sharing in twins, who are in the same classroom. We further found that 41% of the variation in maternal and teacher ratings is explained by a common factor. The heritability of this common factor is 78%. The heritabilities of the rater specific factors of mothers and teachers are 76% and 39%, respectively. Because Attention Problems that are persistent over situations may indicate more serious behavior problems than context dependent Attention Problems, we believe that gene finding strategies should focus on this common phenotype.

Edited by Richard Rose

E. M. Derks (✉) · C. E. M. Van Beijsterveldt · D. I. Boomsma
Department: Biological Psychology, Vrije Universiteit, Van der Boechorststraat 1, 1081 BT Amsterdam
The Netherlands
e-mail: em.derks@psy.vu.nl

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

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

Keywords Attention Problems · Twins · Rating scales · Multi-informant ratings

Introduction

Assessing whether a young child has Attention Problems is difficult. In order to quantify and characterize Attention Problems, researchers and clinicians often have to rely on ratings of parents and teachers. These ratings may be influenced by the rater's personal values or perspective, and by the unique settings in which the rater and child co-exist. Agreement between raters suggests that some aspects of the rated behavior can be reliably assessed regardless of rater or situation. Three different explanations exist for rater disagreement. First, different raters may assess unique aspects of the behavior, which are situation or context dependent. For example, a child's inability to concentrate or to sit still may be obvious in the classroom setting, but less evident in other settings, where sustained attention is less important (e.g., at home). Second, parents and teachers may have different perspectives to the child's behavior. The perspectives may differ for a variety of reasons; teachers are not biologically related to the children, and they are exposed to the behavior of many children of the same age. Third, raters may show rater bias, i.e., their ratings are influenced by their own personal norms and values.

In studies of the teacher and parent ratings of the same children, the agreement between these informants is modest. Van der Ende and Verhulst (2005) reported parent–teacher correlations on AP in the range of 0.29 to 0.41 in a sample of Dutch boys and girls in two different age-groups. Achenbach and

Rescorla (2000) found a correlation of 0.44 between parent and teacher ratings on Attention Problems. The FinnTwin12 study reported higher correlations for ratings on inattention, which pertained to the same setting, e.g., ratings of teachers and classmates, than the correlations for ratings which pertained to different settings, e.g., ratings of teachers and parents, or of classmates and parents (Pulkkinen et al. 1999). The latter findings imply that behavior is in part context dependent.

Previous twin studies supported the hypothesis that mothers and teachers have different perspectives on children's levels of hyperactivity and Attention Problems (Nadder and Silberg 2001), and ADHD (Martin et al. 2002; Thapar et al. 2000). To determine how much of the variation in parent and teacher ratings is due to rating similar versus situation specific components of behavior, some investigators employed bivariate model fitting analyses, which revealed that maternal and teacher ratings on hyperactivity partly reflect a common latent phenotype (Martin et al. 2002; Simonoff et al. 1998). In addition to this common phenotype, maternal ratings reflected rater contrast effects, while teacher ratings reflected aspects of the children's behavior, which did not influence maternal ratings (Simonoff et al. 1998).

It has been shown convincingly that variation in children's inattentive and hyperactive/impulsive behavior is attributable to both genetic and environmental factors. Heritability estimates of parent ratings on AP and/or Hyperactivity (HI) usually vary between 50 and 80% (Hudziak et al. 2000, 2005; Rietveld et al. 2003a; Martin et al. 2002). The heritability estimates of teacher ratings on AP and/or HI tend to be lower than those of parent ratings, and usually fall in the range of 40–70% (Vierikko et al. 2004; Kuntsi and Stevenson 2001; Thapar et al. 2000; Eaves et al. 1997; Sherman et al. 1997). The study of Martin et al. (2002), in which the number of ADHD-symptoms was established in 5–16-year-old children, is the only one, in which heritability estimates were slightly lower in parent ratings (74%) than in teacher ratings (80%).

An interesting finding is that parent and teacher ratings differ not only in the size of the heritability estimate, but also in the etiology of the sources of individual differences. Parent ratings on ADHD are often characterized by non-additive genetic effects (Martin et al. 2002), or contrast effects (Eaves et al. 1997; Kuntsi and Stevenson 2001), while teacher ratings are not. These differences are evident in the corrrelations of the parent ratings of ADHD, which are often very low in DZ twins (Simonoff et al. 1998; Eaves et al. 1997), while teacher ratings do not show these low correlations. Low DZ correlations can be explained either by the presence of non-additive genetic effects (Lynch and Walsh 1998), or by contrast effects (Eaves 1976). These two phenomena both predict low DZ correlations, but the presence of a contrast effect also predicts different variances in MZ and DZ twins. Theoretically the two can thus be distinguished, although Rietveld et al. (2003b) have shown that the statistical power of the classical twin study to do so is low. A further complication is that, given only parent ratings, one cannot distinguish between a contrast effect on the phenotypic level (sibling interaction), and a contrast effect on the observed level (rater bias). Therefore, Simonoff et al. (1998) simultaneously analyzed parent and teacher ratings on childhood hyperactivity. They found that the contrast effect in parent ratings was due to rater bias, not to sibling interaction.

With respect to teacher ratings, it is often the case that correlations are higher in children rated by the same teacher than correlations in children rated by different teachers (Saudino et al. 2005; Vierikko et al. 2004; Towers et al. 2000; Simonoff et al. 1998; but not in Sherman et al. 1997). Higher correlations in children rated by the same teacher than in children rated by different teachers, suggest that teacher rater bias plays a role. Simonoff et al. (1998) developed two different models to explore this finding. One model was based on the assumption that teachers have difficulty distinguishing the two children (''twin confusion model''). The other model was based on the assumption that ratings by the same teacher are correlated, because (a) raters have their own subjective perspective on which behaviors are (in)appropriate, or (b) raters themselves influence the behavior of the child, as a function of his/her (i.e., the rater's) own personality characteristics (''correlated errors model''). However, in their sample of 1044 twin pairs, Simonoff et al. were not able to differentiate between the twin confusion and the correlated errors model. A complicating factor in analyzing behavioral ratings of the same versus different teachers is that classroom separation may not be a random process. In Dutch twins, separation is somewhat more likely when children score high on externalizing problems at age three (Van Leeuwen et al. 2005).

In the present paper, we will examine the contribution of genetic and non-genetic factors to individual differences in Attention Problems (AP). By analyzing maternal and teacher ratings, we estimate the extent to which the agreement between maternal and teacher reports on childhood AP is caused by the same genetic and/or environmental factors being expressed in different surroundings (e.g., the classroom versus the home). Given the size and nature of our twin sample, we are also able to test the contribution of teacher rater bias, as approximately half of our sample is placed into same and half into different classrooms. Although maternal ratings may also be prone to rater bias, we can not directly test for this because twins are always rated by the same mother.

Methods

Subjects

This study is part of an ongoing longitudinal twin study in the Netherlands. The subjects were all registered at birth with the Netherlands Twin Register (Boomsma 1998; Boomsma et al. 2002). For the present study, we analyzed data of a sample of Dutch twins, whose mothers and teachers reported on their behavior, when they were 7-years-old. The twins were born between 1992 and 1996. Maternal ratings were available for 2310 complete twin-pairs and 8 incomplete twin-pairs and teacher ratings were available for 2276 complete twin-pairs and 281 incomplete twin-pairs. In 86% of the twins, ratings were available for both mothers and teachers, in 5% only from mothers, and in 9% only from teachers. Furthermore, about 53% of the twins were in the same classroom, while 36% of the twins were in different classrooms. Of the remaining 11% of the sample, it was unknown whether they were in the same or in different classrooms, mainly due to the fact that a teacher questionnaire was returned for only one of the children. Twin-pairs for whom it was unknown whether the two members of the pair were rated by the same or by different teachers were excluded from the analyses.

Zygosity diagnosis was based on DNA in 123 same-sex twin pairs. In the remaining same-sex pairs, zygosity was assessed with the use of a 10-item questionnaire. This procedure allows an accurate determination of zygosity of nearly 95%. It is described in more detail in Rietveld et al. (2000). The pairs of whom zygosity status could not be determined ($N = 31$ pairs) were excluded from the analyses. The number of twin pairs, by sex, zygosity, and informant are presented in Table 1.

Procedure

A survey, including the CBCL/4–18, was mailed to the mothers of the twins when the twins were 7-years-old. Mothers, who did not return the forms within 2 months, received a reminder. Where financially possible, persistent non-responders were contacted by phone, 7 months after the initial mailing. This procedure resulted in a 66% participation rate. Rietveld et al. (2004) showed that non-participation at age 7 is positively related to the twin's overactive behavior at age 3. However, the difference in overactive scores at age 3 between mothers who do respond (mean = 2.76), and mothers who do not respond (mean = 2.86) at age 7 is small. Once the parent's permission was procured to approach the teacher, a Teacher Report Form (TRF) was sent to the teacher. After 2 months, a reminder was sent to the non-responding teachers. The participation rate of the teachers was 78% (Van Leeuwen et al. 2005). The number of teacher ratings is greater than the number of maternal ratings due to different time schedules for the data entry.

Measures

The Child Behavior Checklist (CBCL/4–18; Achenbach 1991) contains 120 items that measure problem behavior. The items are rated on a 3-point scale ranging from ‘‘not true’’, ‘‘somewhat or sometimes true’’, to ‘‘very true or often true’’. In the present paper, we report on the Attention Problem scale (11 items). The 2-week test–retest correlation and the

Table 1 Number of twin pairs (complete/incomplete)

	Same teacher		Different teacher	
	M	T	M	T
MZM	209/0	236/0	126/2	140/1
DZM	184/1	194/0	153/0	166/2
MZF	247/1	260/1	162/0	177/2
DZF	182/0	196/1	132/1	147/0
DOS	399/2	433/4	252/1	270/3
Zygosity unknown	2/0	15/0	1/0	13/0
Total	1223/4	1334/4	826/4	913/8

M = mother; T = teacher; MZM = monozygotic male; DZM = dizygotic male, MZF = monozygotic female; DZF = dizygotic female; DOS = opposite sex twins internal consistency in this age group are 0.83 and 0.67, respectively (Verhulst et al. 1996).

The Teacher Report Form (TRF; Achenbach 1991) contains 120 items that measure problem behavior with the same three response categories as the CBCL. The Attention Problems scale contains 20 items. The 6-week test–retest correlation is 0.83. The internal consistency coeffients are 0.90 and 0.92 in boys and girls, respectively (Verhulst et al. 1997). Ten items overlap between the AP scales of the CBCL and the TRF.

Statistical Analyses

Both the TRF and CBCL data show high skewness (1.56, and 1.43, respectively), and high kurtosis (2.23, and 2.40, respectively). Derks et al. (2004a) showed that bias in parameter estimates due to non-normality of the data may be avoided by using categorical data analysis. In this approach, a liability threshold model is applied to the ordinal scores (Lynch and Walsh 1998). It is assumed that a person is ‘‘unaffected’’, if his or her liability is below a certain threshold, and that he or she is ‘‘affected’’, if his or her liability is above this threshold. In the present paper, the CBCL and TRF scores were recoded in such a way that three thresholds divide the latent liability distribution into four categories. The thresholds are chosen in such a way that the prevalences are more or less similar in each of the four categories.

In order to test whether the prevalences of Attention Problems vary by sex, or by same and different teacher, we compared the fit of a model in which the thresholds are equated with the fit of a model in which the thresholds are allowed to be different. All statistical analyses were performed in Mx (Neale 1997). The type-I error rate of all statistical tests was set at 0.01 (rather than 0.05) to accommodate multiple testing.

Genetic Modeling

Genetic analyses were performed in a multi-group design of MZM (monozygotic males), DZM (dizygotic males), MZF (monozygotic females), DZF (dizygotic females), and DOS (opposite sex twins). In addition, the twins were divided into a same teacher group and a different teacher group. This resulted in a 10-group analysis.

Univariate genetic models were fitted to maternal ratings on AP. We analyzed the data of children in the same and different classrooms separately for three reasons. First, Van Leeuwen et al. (2005) showed that in Dutch twins, separation is somewhat more likely, when children score highly on externalizing problems at age 3. If these mean differences persist to age 7, combining data from children in same versus different classrooms may give biased estimates of the correlations. Second, Simonoff et al. (1998) reported a slightly higher heritability in maternal ratings for children, who are in the same classroom than for children, who are in different classrooms. Third, it could be the case that children who have the same teacher become more similar, because of their greater environmental sharing at school.

A fully saturated model, in which all correlations and thresholds were freely estimated, was fitted to the ordinal data. Next, we examined whether the thresholds differed between MZ and DZ twins. Because contrast effects cause different variances in MZ and DZ twins, and therefore lead to different prevalences of Attention Problems among these groups, contrast effects were only included if the thresholds of MZ and DZ twins were different. Third, a model that includes additive genetic effects (A), shared environmental (C), or dominant genetic effects (D), and non-shared environmental effects (E) was fitted to the data. It should be noted that the effects of C and D cannot be modeled simultaneously, as they are not both identified. If the correlations in MZ twins were more than twice the correlations in DZ twins, D was included in the model. If the correlations in MZ twins were less than twice the DZ correlations, C was included. Finally, a series of more parsimonious models were fitted: (a) variance components in the best fitting model were constrained to be equal in boys and girls; (b) variance components A, and C or D were constrained at zero; (c) the variance components were constrained to be the same for children in the same classroom and children in different classrooms. The fit of the more parsimonious models were compared with the fit of the full model by means of the likelihood ratio test.

The univariate models that were fitted to the teacher ratings on AP were based on the models that were presented in Simonoff et al. (1998). Similarly to the model fitting of the maternal data, a fully saturated model was fitted to the data. Next, two different genetic models were fitted. In the ‘‘twin confusion’’ model (see Fig. 1), the higher twin correlations in pairs rated by same teachers are explained by the fact that teachers may not always distinguish between the two individuals in a twin pair. The confusion paths are allowed to differ according to zygosity, because we expect more confusion in MZ twins than in DZ twins. Furthermore, the confusions paths are assumed to be absent when children are rated by different teachers or when the individuals in a twin pair are of opposite sex.

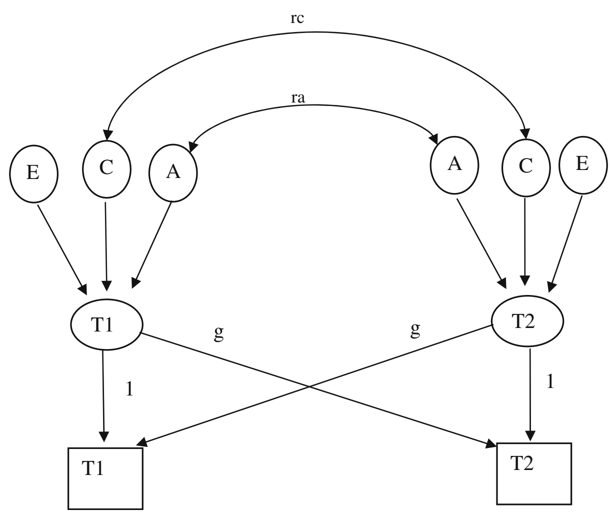

Fig. 1 Twin confusion model for attention problem scores of children rated by the same versus different teachers. Latent factors are represented as circles, observed variables are represented as squares. A = additive genetic effects; C = shared environmental effects; E = non-shared environmental effects; g = twin confusion path (the loading of twin on his/her cotwins attention problem score); T1 (circle) = latent AP score twin 1; T1 (square) = observed score twin 1; T2 (circle) = latent AP score twin 2; T2 (square) = observed score twin 2; ra = 1 (MZ) or 0.5 (DZ); rc = 1 (MZ and DZ). The loading g is allowed to vary as a function of zygosity. In opposite sex twins, and in twin-pairs in which both members of the pair are rated by different teachers, the loading g is constrained at zero. The total variance of the latent factor is constrained at 1

The second model, the ‘‘correlated errors’’ model (see Fig. 2), specifies that teachers bring their own influences into their ratings of behavior either because they have their own subjective perspective, or because they influence the behavior of the child, as a function of rater bias (i.e., the rater's own personality characteristics) (Simonoff et al. 1998). When all twin-pairs are rated by the same informant, rater bias is shared between two members of a twin-pair and is therefore confounded with shared environmental influences. Because we have access to data from twins who are rated by the same teacher and from twins who are rated by different teachers, we are able to distinguish between true shared environmental influences and rater bias. In the correlated error model, the non-shared environmental component is allowed to correlate in children that are rated by the same teacher. If this correlation is significantly greater than zero, this may be evidence of teacher rater bias. It should be noted that the term ‘‘correlated error’’ is too restrictive, because the children rated by same teachers may actually behave more alike, because of certain characteristics of the teacher and/or classmates (e.g., teaching styles, social interactions in the group), or by the fact that classroom separation depends on the level of problem behavior before separation. In these cases, the higher correlation would not be caused by error. However, to avoid confusion, we choose to retain the original name of the model.

Finally, we fitted a bivariate psychometric model to maternal and teacher data. In the psychometric model (Hewitt et al. 1992), the ratings of different informants are allowed to be influenced by a common behavioral view and shared understanding of the behavioral descriptions, and also by unique aspects of their child's behavior. In the bivariate model, we included common factors that influence both maternal and teacher ratings, specific maternal factors that influence maternal ratings only, and specific teacher factors that influence teacher ratings only. Based on the results of the univariate analyses, we identified the most appropriate bivariate model. If rater disagreement is the result of rater bias, the twin correlations of the rater specific factors would not depend on zygosity, and the rater specific variance would be explained by shared environmental influences. In contrast, when the rater disagreement is the result of each rater assessing unique aspects of the child's behavior, and given that the trait is heritable, we would expect to find genetic influences on the rater specific variance.

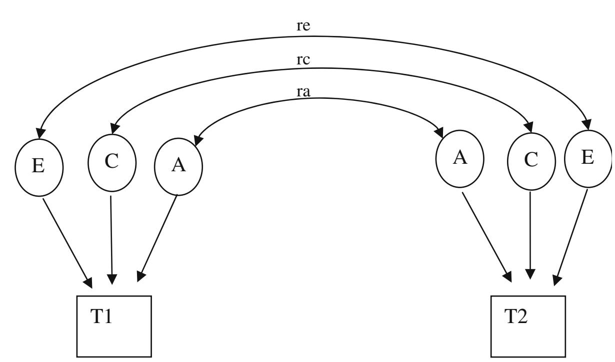

Fig. 2 Correlated error model for attention problem scores of children rated by same versus different teachers. Latent factors are represented as circles, observed variables are represented as squares. A = additive genetic effects; C = shared environmental effects; E = non-shared environmental effects; T1 (square) = observed score twin 1; T2 (square) = observed score twin 2; ra = 1 (MZ) or 0.5 (DZ); rc = 1 (MZ and DZ); re = correlated error path which is constrained at zero in twin-pairs who are rated by different teachers, and is freely estimated in twin-pairs who are rated by the same teacher

Results

Prevalence of AP

Mean scores, standard deviations, and thresholds for maternal and teacher reports on AP are summarized in Table 2. Differences in the distribution of AP were examined by equating the thresholds in Mx. Boys obtained higher AP scores than girls ($\chi^2(48) = 165.14$, $P < 0.001$; $\chi^2(48) = 223.14$, $P < 0.001$, for maternal and teacher ratings, respectively). Maternal and teacher AP scores did not differ between children in different classrooms and children in the same classroom ($\chi^2(36) = 54.31$, $P = 0.03$; $\chi^2(36) = 34.20$, $P = 0.55$, for maternal and teacher ratings, respectively).

Twin Correlations

Polychoric twin correlations were estimated for each sex-by-zygosity group in Mx. The maternal and teacher cross-twin correlations represent the agreement between the twins within each rater. The within-twin cross-rater correlations represent the agreement between the raters within the same child. Finally, the cross-twin cross-rater correlations represent the agreement between raters between the two members of a twin pair. One example of the latter is the correlation between the maternal rating of the first-born twin and the teacher rating of the second born twin.

The correlations of the maternal and teacher AP-scores are shown in Table 3. Because only 10 items overlap between the maternal and teacher AP-scales, we also calculated the correlations on the basis of the 10 overlapping items (see Table 4). The correlations of the overlapping items are no higher than the correlations of the original AP-scales. To facilitate the comparison of the results of the genetic analyses with those of other studies using the TRF, we chose to perform the statistical analyses on the original scales.

Regardless of informant, MZ twin correlations are higher than DZ twin correlations, which suggests the presence of genetic influences. The maternal cross-twin correlations are more than twice as high in MZ twins as in DZ twins, which is suggestive of genetic dominance. Therefore, we fitted an ADE model to the maternal ratings. Because the teacher cross-twin correlations are less than twice as high in MZ as in DZ twins, we fitted an ACE model to the teacher ratings. The cross-twin cross-rater correlations, which represent the common part of the maternal and teacher ratings, are much higher in MZ twins than in DZ twins. We would

Table 2 Mean and standard deviations (SD) of raw scores, and thresholds of maternal and teacher ratings on Attention Problems in 7-year-old boys and girls

	Mother, same classroom		Mother, different classrooms		Teacher, same classroom		Teacher, different classrooms	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Mean	2.86	2.23	3.37	2.44	6.17	3.73	7.37	4.31
SD	2.64	2.45	3.00	2.75	6.41	5.08	7.05	5.24
T1	−0.91	−0.55	−0.98	−0.60	0.04	0.57	−0.11	0.43
T2	0.06	0.40	−0.05	0.36	0.77	1.33	0.63	1.12
T3	1.00	1.35	0.77	1.14	1.50	1.89	1.36	1.87

T1 = threshold 1; T2 = threshold 2; T3 = threshold 3

Table 3 Polychoric Correlations of the maternal and teacher ratings on AP

		Maternal cross-twin	Teacher cross-twin	Within twin cross-rater	Cross-twin cross-rater
MZM	ST	0.77	0.81	0.48	0.40
	DT	0.79	0.56	0.51	0.38
DZM	ST	0.39	0.49	0.32	−0.04
	DT	0.25	0.22	0.51	0.02
MZF	ST	0.80	0.82	0.43	0.37
	DT	0.69	0.51	0.37	0.25
DZF	ST	0.35	0.48	0.36	0.00
	DT	0.35	0.39	0.39	0.13
DOS	ST	0.16	0.49	0.34	−0.05
	DT	0.20	0.21	0.42	0.03

ST = same teacher; DT = different teacher; MZM = monozygotic male; DZM = dizygotic male, MZF = monozygotic female; DZF = dizygotic female; DOS = opposite sex twins

Table 4 Polychoric Correlations of the 10 overlapping items of the maternal (M) and teacher (T) AP-scales

		Maternal cross-twin	Teacher cross-twin	Within twin cross-rater	Cross-twin cross-rater
MZM	ST	0.77	0.80	0.47	0.43
	DT	0.79	0.43	0.47	0.34
DZM	ST	0.39	0.44	0.29	−0.07
	DT	0.24	0.31	0.48	0.08
MZF	ST	0.80	0.83	0.38	0.32
	DT	0.69	0.44	0.32	0.17
DZF	ST	0.35	0.44	0.29	0.02
	DT	0.36	0.34	0.28	0.13
DOS	ST	0.32	0.36	0.35	−0.01
	DT	0.19	0.24	0.35	0.00

ST = same teacher; DT = different teacher; MZM = monozygotic male; DZM = dizygotic male, MZF = monozygotic female; DZF = dizygotic female; DOS = opposite sex twins

therefore expect an ADE model to provide the best fit to the common part of the bivariate model.

Statistical tests showed that maternal correlations did not differ among twin pairs in the same classroom versus different classrooms ($\chi^2(6) = 15.52$, $P = 0.02$). In contrast, teacher correlations were higher in children rated by the same teacher than in children rated by different teachers ($\chi^2(6) = 40.89$, $P < 0.001$).

Univariate Genetic Model Fitting Analyses of Teacher Ratings

In maternal ratings on AP, the thresholds did not differ between MZ and DZ twins ($\chi^2(24) = 30.10$, $P = 0.18$). Therefore, rater contrast effects were not included in the genetic model. The results of the genetic model fitting are summarized in Table 5. The best-fitting model is printed in bold. Briefly, the univariate genetic analyses showed significant influences of A, D, and E. The estimates of A, D, and E did not depend on sex, and did not differ among twins in the same classroom and twins in different classrooms. The relative influences of A, D, and E in the best fitting model were 44%, 33%, and 23%, respectively. Compared to a saturated model, the fit of this model was good ($\chi^2(10) = 18.30$, $P = 0.932$).

Univariate Genetic Model Fitting Analyses of Teacher Ratings

The results of the model fitting analyses on teacher ratings are shown in Table 6. An ACE model that allowed for different influences of A, C, and E in same and different teachers provided a good fit to the data. However, the more parsimonious ‘‘correlated errors’’ model also provided a good fit. The ‘‘twin confusion’’ model did not fit well. In the correlated errors model, the relative influences of genes and environment did not differ between boys and girls, and the influence of the shared environment was not significant. The heritability of teacher ratings on AP was 55% and the non-shared environment explained 45% of the variation. The non-shared environment correlated 0.54 when children were rated by the same teacher. This correlation was significantly greater than zero. Compared to a saturated model, the fit of the correlated errors model was good ($\chi^2(10) = 6.79$, $P = 0.745$).

Bivariate Genetic Model Fitting Analyses of Maternal and Teacher Ratings

Based on the results of the univariate genetic analyses, we fitted a bivariate model that included a common

Table 5 Univariate model fitting of maternal Attention Problem ratings in 7-year-old children

Model	Parameters	−2 LL	With model	Δ df	Δ χ^2	*P*
1. Fully saturated model	84	10201.14	–	–	–	–
2. ADE model, boys ≠ girls, same ≠ different teacher	80	10210.80	1	4	9.66	.047
3. ACE model, boys ≠ girls, same ≠ different teacher	80	10224.04	1	4	22.90	0.000
4. ADE model, boys = girls, same ≠ different teacher	76	10216.21	2	4	5.41	0.248
5. ADE model, boys = girls, same = different teacher	**74**	**10219.44**	**4**	**2**	**3.23**	**0.199**
6. AE model, boys = girls, same = different teacher	73	10226.86	5	1	7.42	0.006

A = additive genetic effects, C = shared environmental effects, D = dominant genetic effects, E = non-shared environmental effects, −2 LL = −2 log likelihood, df = degrees of freedom

Boys = girls: equating the non-standardized parameters of boys and girls

Same = different teacher: equating the non-standardized parameters of same and different teachers

Table 6 Univariate model fitting of teacher Attention Problem ratings in 7-year-old children

Model	Parameters	−2 LL	With model	Δ df	Δ χ^2	*P*
1. Fully saturated model	84	8683.69	–	–	–	–
2. ACE model, boys ≠ girls, same ≠ different teacher	80	8685.37	1	4	1.68	0.794
3. ACE, Correlated error model, boys ≠ girls	78	8688.55	1	6	4.86	0.562
4. ACE, Twin confusion model, boys ≠ girls	80	8694.12	1	4	10.43	0.034
5. ACE, Correlated error model, boys = girls	75	8690.48	3	3	1.93	0.587
6. AE, Correlated error model, boys = girls	**74**	**8690.48**	**5**	**1**	**0.00**	**–**
7. AE, boys = girls, correlated error dropped	73	8726.41	6	1	35.93	0.000

A = additive genetic effects, C = shared environmental effects, D = dominant genetic effects, E = non-shared environmental effects, −2 LL = −2 log likelihood, df = degrees of freedom

Boys = girls: equating the non-standardized parameters of boys and girls

Same = different teacher: equating the non-standardized parameters of same and different teachers

part consisting of the factors Ac, Dc, and Ec, a unique maternal part consisting of the factors Am, Dm, and Em, and a unique teacher part consisting of the factors At and Et. A correlated error was only included in the unique teacher part of the bivariate model. The bivariate model fitting results are summarized in Table 7. In the best fitting bivariate model, 41% of the variation in maternal and teacher ratings on AP was explained by a common factor. This common factor was decomposed into a dominant genetic factor, which explained 32% of the total variation, and a non-shared environmental factor, which explained 9% of the total variation. The heritability of the common factor is 78% (this can be calculated as the amount of variation explained by genetic factors divided by the total variance 0.32/0.41 = 78%). Variation in maternal ratings was further explained by Am (45%), and Em (14%). Variation in teacher ratings was explained by At (23%), and Et (36%), and a correlated error of 0.77 in same teacher ratings. Compared to a saturated model, the fit of the bivariate model was not very good ($\chi^2(66) = 103.09$, $P = 0.002$). However, it is known that in bivariate analyses, the power to detect very small differences is high. Therefore, we calculated the residuals of the expected covariance matrices of the ADE model and the expected covariance matrices under the saturated model. Expectation of these residuals showed that the misfit was mainly due to different cross-rater cross-twin correlations in the monozygotic male group rated by the same teacher (i.e., the correlation of teacher–firstborn with mother–second born is not equal to the correlation of mother–firstborn with teacher–second born). Because there is no theoretical reason why this correlation would depend on birth-order, we accepted the ADE model as the best-fitting model. This model is shown in Fig. 3, including the estimated factor loadings. As an illustration of Fig. 3, we will show how the heritability of the common factor can be derived based on the factor loadings. The total variance of the common factor is $0.30^2+0.57^2+0.00^2=0.41$. The variance explained by the genetic factor is $0.57^2=0.32$. Therefore, the proportion of the variance explained by genetic factors = 0.32/ 0.41 = 0.78, and the heritability of the common factor is 78%.

Figure 4 gives an overview of the genetic and environmental influences on the common and rater-specific parts of the model.

Discussion

The purpose of this study was to examine the genetic and environmental contributions to the variation in maternal and teacher ratings on Attention Problems in children, and to the covariation between these ratings.

Table 7 Bivariate model fitting of maternal and teacher Attention Problem ratings in 7-year-old children

Model	Parameters	−2 LL	With model	Δ df	Δ χ^2	*P*
1. Fully saturated model	216	18362.96	–	–	–	–
2. ADE model, boys ≠ girls	160	18460.51	1	56	97.55	0.000
3. ADE model, boys = girls	152	18466.05	2	8	5.54	0.699
4. ADE model, boys = girls, rater-specific D dropped	**150**	**18466.05**	**3**	**2**	**0.00**	**–**

A = additive genetic effects, C = shared environmental effects, D = dominant genetic effects, E = non-shared environmental effects, −2 LL = −2 log likelihood, df = degrees of freedom

Boys = girls: equating the non-standardized parameters of boys and girls

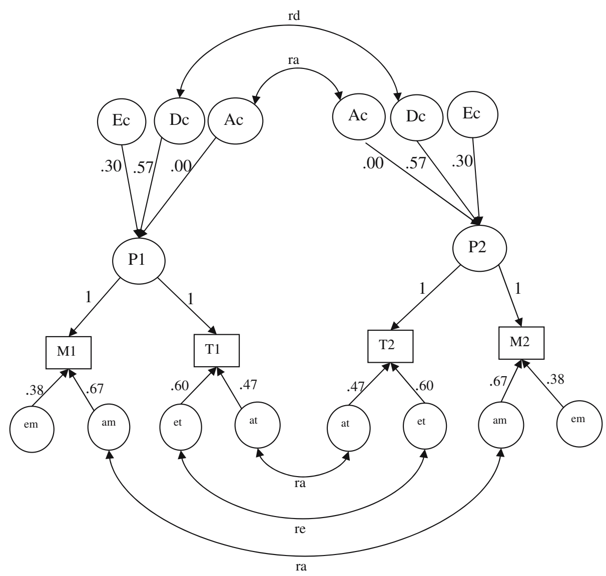

Fig. 3 Bivariate model for maternal and teacher ratings. P1 = phenotype twin 1; P2 = phenotype twin 2; M1 = Maternal rating of twin 1; M2 = Maternal rating of twin 2; T1 = Teacher rating of twin 1; T2 = Teacher rating of twin 2. Ac, Dc, and Ec are the common additive genetic, dominant genetic, and non-shared environmental effects; am and em are the unique maternal additive genetic, and non-shared environmental effects; at and et are the unique teacher additive genetic, and non-shared environmental effects. ra = 1 (MZ) or 0.5 (DZ); rd = 1 (MZ) or 0.25 (DZ); re is the correlated error path and is estimated at 0.77. It is constrained to be equal in MZ and DZ twins, and is assumed to be absent in children rated by different teachers. The paths from the latent phenotypes P1 and P2 to the maternal and teacher ratings are constrained at 1

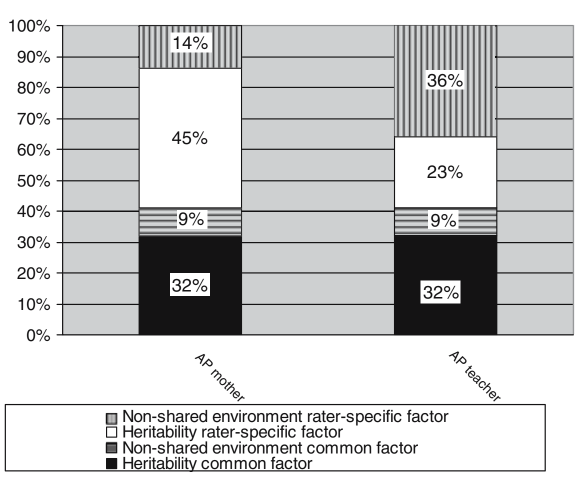

Fig. 4 Graphical representation of the influences of genes and non-shared environment on Attention Problems in 7-year-old twins

In the univariate genetic analyses, the heritability estimate was higher in maternal ratings (77%) than in teacher ratings (54%), which agrees with previous findings (Vierikko et al. 2004; Kuntsi and Stevenson 2001; Thapar et al. 2000; Simonoff et al. 1998; Eaves et al. 1997; and Sherman et al. 1997). A more thorough investigation of the correlations, however, revealed that the correlation in the maternal data was similar to the correlation in the 'same teacher' data, and that both of these were higher than the correlation in the 'different teacher' data. These data therefore support the inference that the lower heritability in teacher ratings is due to combining data from same and different teachers. This is consistent with the findings of Martin et al. (2002), who observed similar heritabilities in parent and teacher ratings in a sample consisting for 91% of children rated by same teachers. In contrast, Vierikko et al. (2004) also conducted genetic analyses on twin-pairs, in which both members were rated by the same teacher, and reported a lower heritability in teacher (49–55%) than parent (78–81%) ratings. In summary, the pattern is somewhat inconsistent, but the present results suggest that the higher heritabilities in parental ratings than teacher ratings can be explained by the fact that twins are always rated by the same parent, but in about half of the cases by different teachers.

Previously, it was shown that the higher twin correlations in children rated by the same teacher than in children rated by different teachers are associated with a higher heritability of problem behavior (Saudino et al. 2005; Simonoff et al. 1998). Simonoff et al. (1998) compared the fit of two distinct theoretical models to explain this finding, but both models fit equally well. In the present study, we were able to differentiate between these models; the correlated error model provided a better fit to the data than the twin confusion model. The correlated error may be caused by rater bias, reflecting the fact that raters have their own specific perspective on which behaviors are (in)appropriate. An alternative explanation is that the correlated error reflects true qualities of the children's behavior, which are elicited by the exposure to a particular rater (Simonoff et al. 1998). For example, different teachers may elicit different behaviors from children.

These two alternative explanations have different implications for the interpretation of the high correlations in children who are rated by the same informant. If it is true that the correlations are higher because of rater specific views, this implies that the phenotypic correlations in both maternal and same teacher ratings are overestimated, and that the influence of non-shared environmental factors is underestimated. Alternatively, if the higher correlations are the result of the fact that children behave more similarly when confronted with the same person, this suggest that the behavior of children depends on the person, with whom they interact. In this case, the lower correlation in twins rated by different informants is the result of an increase in the non-shared environmental variance, and the high correlations, when twins are rated by the same informant, reflect the true phenotypic similarity of children interacting with the same person. The second possibility may explain the higher correlation between paternal and maternal ratings than the correlation between parent and teacher ratings. The parents usually observe the children in interaction with the other parent, but not in interaction with the teacher. Consequently, correlations should be lower in parents, who are divorced, than in parents, who live together. With the available data, we cannot decide whether the high correlation in twins rated by the same informant are caused by rater specific views, or by the influence of the informant on the child's behavior.

As in previous studies on Attention Problems (Vierikko et al. 2004; Van der Ende and Verhulst 2005), the correlations between maternal and teacher ratings were moderate. We ruled out the possibility that rater disagreement is the result of non-overlapping items of the AP-scales of parents and teachers by showing that the correlations of the overlapping items are not higher than the correlations of the original scales.

The bivariate model fitting analyses showed that slightly less than half of the variation in maternal and teacher ratings is explained by common aspects of the child's behavior while the remaining variation is explained by rater or setting specific aspects. The finding of genetic influences on the rater specific variance shows that the diagreement between parents and teachers is not solely due to rater bias. Both raters assess unique aspects of the child's behavior. The common aspects, which are highly genetic, reflect the part of the phenotype that is stable across settings and raters. The genetic variation of the common factor was completely explained by dominant genetic effects. This is in agreement with the low cross-twin cross-rater correlations reported in DZ twins. The large dominant genetic influences were surprising because these were not found in the univariate analyses of the teacher ratings. How can we explain the low cross-twin cross-rater correlations in DZ twins? The fact that low cross-rater cross-twin correlations (ranging from −0.12 to −0.21) were also reported by Simonoff et al. (1998) in their study on hyperactivity suggests that these are not the result of artifacts in our data collection. The presence of sibling interaction is not likely either as these effects should also be found in the univariate analyses of maternal or teacher ratings. Rater bias does not seem to play a role, because it is hard to envisage that high teacher ratings on AP in twin 1 would lead to low maternal ratings on AP in twin 2. The only explanation that we can offer here is that variation in maternal and teacher ratings is influenced by a correlated error, which increases the correlation in MZ and DZ twins, and mimics the effect of shared environmental influences, and by dominant genetic effects. These effects might cancel each other out in the univariate analyses (which would suggest that the dominance effect reported for maternal ratings is underestimated). In the common factor of the bivariate analyses (i.e., the factor that influences both maternal and teacher ratings), correlated errors are absent, and the presence of the dominance genetic effects is evident. Some support for this explanation is provided by the different teacher correlations. In boys, the different teacher correlations show a pattern that is in agreement with the presence of genetic dominance while the same teacher correlations do not. However, in girls, the pattern of the different teacher correlations in DZ girls is not suggestive of genetic dominance. Future studies should reveal further insight regarding the low cross-twin cross-rater correlations.

The significant influence of genes on the rater-specific aspects is consistent with Martin et al. (2002), who found that variation in maternal and teacher ratings on hyperactivity is partly influenced by different genes. It implies that disagreement between parents and teachers is not merely due to rater bias. This finding is consistent with the results of Bartels et al. (2004) and Derks et al. (2004b), who found that mothers and fathers assess unique aspects of the child's behavior, although most variation in these ratings is explained by common aspects. Apparently, mothers and teachers both rate meaningful, but partly different, aspects of children's behavior.

The fact that the prevalence of ADHD was similar in MZ and DZ twins, is supportive of an absence of rater contrast or sibling interaction in maternal ratings. In the literature on AP and HI, contradictory findings are reported with respect to the presence of contrast effects in parental ratings. Significant contrast effects on AP and/or HI have been reported in some studies (Simonoff et al. 1998; Kuntsi and Stevenson 2001; Vierikko et al. 2004; Eaves et al. 1997, 2000), but not in others (Kuntsi et al. 2005; Martin et al. 2002, Thapar et al. 2000, Towers et al. 2000, and Hudziak et al. 2000). In teacher ratings, contrast effects are absent (Simonoff et al. 1998; Kuntsi and Stevenson 2001; Vierikko et al. 2004; Eaves et al. 1997). Plomin (1982) suggest that contrast effects are more likely when the items refer to global descriptions of behavior rather than to specific descriptions of behavior. This was confirmed by Saudino et al. (2004), who report a tendency for contrast effects to be more pervasive when global ratings were required. The lack of contrast effects in the current study shows that the items of the CBCL are specific enough to prevent parents from comparing the behavior of the twins.

The results should be interpreted with the following points kept in mind. First, the CBCL and TRF do not assess the presence of DSM symptoms. The CBCL-AP scale does predict the presence of DSM-IV ADHD (Hudziak et al. 2004), but whether this is so for the teacher form is unknown. There are a number of reasons that DSM interviews of teachers are rarely employed, including time burden and expense. However, perhaps the most important reason is the lack of an empirically validated DSM-IV teacher data base. Thus, although teacher reports on ADHD are commonly used, there is little known about the validity of these reports. The Netherlands Twin Register is currently collecting data on school performance, and in future studies we will address the question whether high teacher AP-scores are more predictive for problems related to school performance than high parental AP-scores. Second, the results in this study are based on analysis of Attention Problems rated by parents and teacher. It is unclear whether the results generalize to hyperactivity. However, the Attention Problem scales do include some items on hyperactivity (e.g., cannot sit still, restless, or hyperactive), and a review of epidemiological genetic studies shows that the heritabilities of Attention Problems and Hyperactivity are similar (Derks and Boomsma, in preparation).

Higher correlations are found in children rated by the same informant than in children rated by different informants. At this point, it is unclear whether the higher correlation based on ratings from the same informant overestimate the true phenotypic correlation due to rater specific views, or if the lower correlation based on ratings from different informants underestimate the true phenotypic correlation as a result of increased non-shared environmental influences.

In conclusion, we showed that a little under half of the variation in children's inattentive behavior is persistent over situations and is rater and setting independent. The heritability of this common phenotype is quite high. Todd et al. (2001) have argued that only through careful phenotype refinement will the identification of genetic and environmental influences on complex traits be realized. Because Attention Problems, which are persistent over situations, may indicate more serious behavior problems than Attention Problems that are present in only one context, we believe that gene finding strategies should focus on this common phenotype.

Acknowledgments We want to thank the participants of the twin studies. 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.), and by the Center for Neurogenomics and Cognitive Research.

References

Achenbach TM, Rescorla LA (2000) Manual for ASEBA preschool forms & profiles. University of Vermont, Department of Psychiatry, Burlington, VT

Achenbach TM (1991) Integrative guide for the 1991 CBCL 4/18, YSR, and TRF profiles. University of Vermont, Department of Psychiatry, Burlington, VT

Bartels M, Van den Oord EJCG, Hudziak JJ, Rietveld MJH, Van Beijsterveldt CEM, Boomsma DI (2004) Genetic and environmental mechanisms underlying stability and change in problem behaviors at ages 3, 7, 10 and 12. Dev Psychol 40:852–867

Boomsma DI (1998) Twin registers in Europe: an overview. Twin Res 1:34–51

Boomsma DI, Vink JM, Van Beijsterveldt CEM, Geus EJC, Beem AL, Mulder EJCM, Derks EM, Riese H, Willemsen AHM, Bartels M, Van den Berg M, Kupper HM, Polderman JC, Posthuma D, Rietveld MJH, Stubbe JH, Knol LI, Stroet T, Van Baal GCM (2002) Netherlands Twin Register: a focus on longitudinal research. Twin Research 5:401–406

Derks EM, Dolan CV, Boomsma DI (2004a) 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 (2004b) A study of genetic and environmental influences on maternal and paternal CBCL syndrome scores in a large sample of 3-year-old Dutch twins. Behav Genet 34:571–583

Eaves LJ (1976) A model for sibling effects in man. Heredity 36:205–214

Eaves LJ, Silberg JL, Meyer JM, Maes HH, Simonoff E, Pickles A, Rutter M, Neale MC, Reynolds CA, Erikson MT, Heath AC, Loeber R, Truett KR, Hewitt JK (1997) Genetics and developmental psychopathology: 2. The main effects of genes and environment on behavioral problems in the Virginia Twin Study of adolescent behavioral development. J Child Psychol Psychiatry 38:965–980

Eaves L, Rutter M, Silberg JL, Shillady L, Maes H, Pickles A (2000) Genetic and environmental causes of covariation in interview assessments of disruptive behavior in child and adolescent twins. Behav Genet 30:321–334

Hewitt JK, Silberg JL, Neale MC, Eaves LJ, Erickson M (1992) The analysis of parental ratings of children's behavior using LISREL. Behav Genet 22:293–317

Hudziak JJ, Copeland W, Stanger C, Wadsworth M (2004) Screening for DSM-IV externalizing disorders with the Child Behavior Checklist: a receiver-operating characteristic analysis. J Child Psychol Psychiatry 45:1299–1307

Hudziak JJ, Derks EM, Althoff RR, Boomsma DI (2005) The genetic and environmental contributions to Attention Deficit Hyperactivity Disorder as measured by the Conners' Rating Scales-Revised. Am J Psychiatry 162:1614–1620

Hudziak JJ, Rudiger LP, Neale MC, Heath AC, Todd RD (2000) A twin study of inattentive, aggressive, and anxious/depressed behaviors. J Am Acad Child Adolesc Psychiatry 39:469–476

Kuntsi J, Rijsdijk F, Ronald A, Asherson P, Plomin R (2005) Genetic influences on the stability of attention-deficit/hyperactivity disorder symptoms from early to middle childhood. Biol Psychiatry 57:647–654

Kuntsi J, Stevenson J (2001) Psychological mechanisms in hyperactivity: II. The role of genetic factors. J Child Psychol Psychiatry 42:211–219

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Associates, 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

Nadder TS, Silberg JL (2001) Comparison of multiple measures of ADHD symptomatology: a multivariate genetic analysis. J Child Psychol Psychiatry 42:475–486

Neale MC (1997) Mx: statistical modeling. Richmond Department of psychiatry Medical, College of Virginia, VA

Plomin R (1982) Childhood temperament. In: Lahey B, Kazdin A (eds) Advances in clinical child psychology. Plenum, New York

Pulkkinen L, Kaprio J, Rose RJ (1999) Peers, teachers and parents as assessors of the behavioural and emotional problems of twins and their adjustment: the Multidimensional Peer Nomination Inventory. Twin Res 2:274–285

Rietveld MJH, Van der Valk JC, Bongers IL, Stroet TM, Slagboom PE, Boomsma DI (2000) Zygosity diagnosis in young twins by parental report. Twin Res 3:134–141

Rietveld MJH, Hudziak JJ, Bartels M, van Beijsterveldt CEM, Boomsma DI (2003a) Heritability of attention problems in children: I. Cross-sectional results from a study of twins, age 3–12 years. Am J Med Genet B 117:102–113

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

Rietveld MJH, Posthuma D, Dolan CV, Boomsma DI (2003b) ADHD: sibling interaction or dominance: an evaluation of statistical power. Behav Genet 33:247–255

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

Saudino KJ, Wertz AE, Gagne JR, Chawla S (2004) Night and day: are siblings as different in temperament as parents say they are? J Pers Soc Psychol 87:698–706

Sherman DK, Iacono WG, McGue MK (1997) Attention-deficit hyperactivity disorder dimensions: a twin study of inattention and impulsivity–hyperactivity. J Am Acad Child Adolesc Psychiatry 36:745–753

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

Thapar A, Harrington R, Ross K, McGuffin P (2000) Does the definition of ADHD affect heritability? J Am Acad Child Adolesc Psychiatry 39:1528–1536

Todd RD, Rasmussen ER, Neuman RJ, Reich R, Hudziak JJ, Bucholz KK, Madden PAF, Heath AC (2001) Familiality and heritability of subtypes of attention deficit hyperactivity disorder in a population sample of adolescent female twins. Am J Psychiatry 158:1891–1898

Towers H, Spotts E, Neiderhiser JM, Plomin R, Hetherington EM, Reiss D (2000) Genetic and environmental influences on teacher ratings of the Child Behavior Checklist International. Int J Behav Dev 24:373–381

Van der Ende J, Verhulst FC (2005) Informant, gender and age differences in ratings of adolescent problem behaviour. Eur Child Adolesc Psychiatry 14:117–126

Van Leeuwen M, Van den Berg SM, Van Beijsterveldt CEM, Boomsma DI (2005) Effects of classroom separation on problem behavior in 7 to 12 year old twins. Twin Res 4:384–391

Verhulst FC, Van der Ende J, Koot HM (1996) Handleiding voor de CBCL/4–18. Afdeling Kinder- en jeugdpsychiatrie, Sophia Kinderziekenhuis/Academisch Ziekenhuis Rotterdam/Erasmus Universiteit Rotterdam, Rotterdam

Verhulst FC, Van der Ende J, Koot HM (1997) Handleiding voor de Teacher's Report Form (TRF). Afdeling Kinder- en jeugdpsychiatrie, Sophia Kinderziekenhuis/Academisch Ziekenhuis Rotterdam/Erasmus Universiteit Rotterdam, Rotterdam

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