

Gene–environment interaction in teacher-rated internalizing and externalizing problem behavior in 7- to 12-year-old twins

Diane J. Lamb,¹ Christel M. Middeldorp,^{1,2,3} Catarina E.M. Van Beijsterveldt,¹ and Dorret I. Boomsma¹

¹Department of Biological Psychology, Vrije Universiteit, Amsterdam, The Netherlands; ²Department of Child and Adolescent Psychiatry, Academic Medical Center, Amsterdam, The Netherlands; ³Department of Child and Adolescent Psychiatry, GGZ inGeest/VU Medical Center, The Netherlands

Background: Internalizing and externalizing problem behavior at school can have major consequences for a child and is predictive for disorders later in life. Teacher ratings are important to assess internalizing and externalizing problems at school. Genetic epidemiological studies on teacher-rated problem behavior are relatively scarce and the reported heritability estimates differ widely. A unique feature of teacher ratings of twins is that some pairs are rated by different and others are rated by the same teacher. This offers the opportunity to assess gene–environment interaction. **Methods:** Teacher ratings of 3,502 7-year-old, 3,134 10-year-old and 2,193 12-year-old twin pairs were analyzed with structural equation modeling. About 60% of the twin pairs were rated by the same teacher. Twin correlations and the heritability of internalizing and externalizing behavior were estimated, separately for pairs rated by the same and different teachers. Socioeconomic status and externalizing behavior at age 3 were included as covariates. **Results:** Twin correlations and heritability estimates were higher when twin pairs were in the same class and rated by the same teacher than when pairs were rated by different teachers. These differences could not be explained by twin confusion or rater bias. When twins were rated by the same teacher, heritability estimates were about 70% for internalizing problems and around 80% in boys and 70% in girls for externalizing problems. When twins were rated by different teachers, heritability estimates for internalizing problems were around 30% and for externalizing problems around 50%. **Conclusions:** Exposure to different teachers during childhood may affect the heritability of internalizing and externalizing behavior at school. This finding points to gene–environment interaction and is important for the understanding of childhood problem behavior. In addition, it could imply an opportunity for interventions at school. **Keywords:** Internalizing problem behavior, externalizing problem behavior, teacher ratings, gene–environment interaction.

Introduction

Internalizing (INT) and externalizing (EXT) problem behavior at school can have major consequences for a child. It can seriously affect learning achievements, lead to bullying by other children or negative interactions with the teachers (Deater-Deckard, 2001; Ladd & Burgess, 1999). Moreover, a DSM-IV diagnosis requires that problem behavior during childhood is not restricted to one environment (American Psychiatric Association, 2000). It is therefore common practice to take into account INT and EXT behavior at school as rated by the teacher when assessing a child for psychiatric problems in a child and adolescent psychiatry clinic.

Behavior genetic studies that investigate the relative influence of genetic and environmental risk factors on teacher-rated behavior are relatively scarce and sample sizes are small. An often used paradigm within behavior genetics is the twins design, which makes use of the different genetic relatedness of monozygotic (MZ) and dizygotic (DZ) twins. As MZ

twins are (nearly) genetically identical whereas DZ twins share on average 50% of their genes a higher correlation in MZ twin pairs than in DZ twin pairs suggest that part of the individual differences can be explained by genetic risk factors (Boomsma, Busjahn, & Peltonen, 2002a). A correlation in MZ twin pairs which is less than twice the DZ twin correlation indicates the influence of common environmental factors shared by children in the same family. Differences in problem behavior within MZ twin pairs are due to unique environmental influences, which also include measurement error.

Heritability estimates based on teacher ratings of twins behavior vary widely, e.g. from 31% to 81% for EXT and conduct problems, respectively (Saudino, Ronald, & Plomin, 2005; Towers et al., 2000). One of the reasons for the varying estimates is that twin correlations tend to be higher when the twins' behavior is rated by the same teacher than when it is rated by different teachers (e.g. Saudino et al., 2005; Simonoff et al., 1998; Towers et al., 2000). Assuming that same teacher ratings are inflated or that using multiple teachers to rate a twin pair increases the

Conflict of interest statement: No conflicts declared.

error in the measurement, these ratings were sometimes excluded, resulting in a wide range of heritability estimates for problem behavior (Hartman, Rhee, Willcutt, & Pennington, 2007; Towers et al., 2000). Heritability is estimated around 72% for different types of problem behavior when twin pairs are rated by the same teacher, in contrast to estimates of about 54% when twin pairs are rated by different teachers (Saudino et al., 2005).

However, it is an important question whether teacher ratings are inflated or biased and if this bias explains the differences in correlations between same and different teacher ratings. In addition to teacher bias leading to correlated errors, the higher correlations in MZ twins may be explained by confusion of twins rated by the same teacher, or by gene–environment interaction (Baker, Jacobson, Raine, Lozano, & Bezdjian, 2007; Derks, Dolan, Hudziak, Neale, & Boomsma, 2007; Derks, Hudziak, van Beijsterveldt, Dolan, & Boomsma, 2006; Polderman, Posthuma, De Sonneville, Verhulst, & Boomsma, 2006; Saudino et al., 2005; Simonoff et al., 1998). Rater bias refers to the teacher's own bias which affects the similarity of ratings of twin behavior if the same teacher rates more than one child (correlated errors). Twin confusion is present when a teacher does not always distinguish the two individuals in a twin pair and therefore rates their behavior as more similar. A third explanation is the presence of gene–environment interaction (G × E). G × E implies that the influence of genetic factors depends on the environment (Boomsma & Martin, 2002; Eaves, 1984). Thus, heritability estimates can differ between groups that are rated in different environments.

In this report, we focus on the etiology of teacher-rated INT and EXT problem behavior in a large sample of male and female twins at age 7, 10 and 12 years. The sample includes twin pairs who are rated by the same teacher and pairs who are rated by different teachers. First, we investigate the influence of social-economic status (SES) and earlier problem behavior on the chance of twins being separated or not. Earlier research has shown that SES and earlier problem behavior are associated with problem behavior at later ages (Bartels et al., 2004; Bradley & Corwyn, 2002). If these factors are also related to twins being separated, this can lead to confounding. Consequently, significant risk factors are included in the following analyses as covariates. Next, we tested which of the three models, that is, twin confusion, correlated errors or G × E, described the data best and estimated to what extent individual differences in teacher-rated INT and EXT behavior are explained by genetic and environmental risk factors.

Methods

Subjects

All twins participated in the ongoing longitudinal study of the Netherlands Twin Register (NTR; Boomsma et al.,

2002b). Most twins were registered with the NTR at birth. Parents of twins completed questionnaires assessing the twin's behavior from birth to age 12. If parents provided informed consent, teachers were asked to report on the twin's behavior at age 7, 10 and 12. Assessment took place starting from 1999 up to the year 2010. The current study included twins with maternal ratings on externalizing problems at age 3, in addition to teacher ratings at age 7, 10 or 12. Teacher ratings were available for 3,700 complete and 701 incomplete (i.e. only one teacher returned the questionnaire for one twin from a pair) 7-year-old twin pairs, 3,310 complete and 776 incomplete 10-year-old twin pairs, 2,315 complete and 618 incomplete 12-year-old twin pairs. Twins were excluded from the analyses when teachers reported that they did not know the student well (at age 7, 10 and 12: 0.9%, 1.5%, 0.8%), or had only known the student for a short time (at age 7, 10 and 12: 1.6%, 2.0%, 1.7%). Twin pairs were excluded when one or both twins attended special education (at age 7, 10 and 12: 0.8%, 1.7% and 1.9%). Special education included schools for physical and cognitive disabled children and children with learning and extreme behavioral problems. Classes in special education are different from regular primary school classes in that they are smaller. Zygosity determination took place in different projects in which subgroups of the twins participated. For 2% of the twins, zygosity was determined based on blood group polymorphisms, for 12% based on DNA polymorphisms, for 55% based on parental ratings of zygosity and for 31% based on the fact that the twins were of opposite sex. Twins were only included if information on zygosity and SES was available leading to missing data around 1%.

Measures

Teachers completed the Teacher Report Form (TRF; Achenbach & Rescorla, 2001; Achenbach, 1991b), which contains 120 items measuring problem behavior. The TRF includes the broad band INT scale which consists of the subscales Anxious/Depressed, Withdrawn behavior and Somatic complaints, and the EXT scale which consists of subscales Aggressive Behavior and Rule Breaking Behavior.

Parents reported on the twins' behavior by means of the Child Behavior Check List (CBCL; Achenbach, 1991a). The CBCL and the TRF are both part of the Achenbach system of empirically based assessment (ASEBA) and are comparable. The broad band INT scale for the CBCL 2/3 consists of the Anxious/Depressed and Withdrawn Behavior subscales (and not Somatic Complaints), EXT consists of the Rule Breaking Behavior and Aggressive Behavior subscales.

Social-economic status of the twin families was measured at age 3, 7 and 10 years. SES was obtained from a full description of the occupation of the parents for about two-third of the twin sample. Subsequently, SES was coded according to the standard classification of occupations (2001). For the remaining families, SES was obtained by the EPG-classification scheme (Erikson, Goldthorpe, & Portocarero, 1979), combined with information on parental education. All twins were classified into three SES levels (1 = low; 2 = middle; 3 = high).

Statistical analyses

Means model and covariates. INT and EXT problem behavior showed an L-shaped distribution. Therefore, prior to the analyses, the data were normalized in PRELIS. PRELIS makes use of a normalizing transformation that fits the inverse normal density function to the ranked data (Jöreskog & Sörbom, 2002). This did not influence the estimates of the variances.

The effect of SES and problem behavior on twin separation at age 7, 10 and 12, was analyzed performing binary logistic regression analyses in SPSS (Chicago: SPSS Inc, 2008). INT and EXT as rated by the mother and the father at age 3 and SES were used as predictors and twin pair separation at age 7, 10 and 12, was used as the dependent variable.

Structural equation modeling as implemented in the software package Mx was used to estimate means, variances and twin correlations as a function of zygosity for twin pairs rated by the same and by different teachers (Neale, Boker, & Maes, 2006). First, a so-called saturated model was fitted to the INT and EXT data at age 7, 10 and 12 in which all these parameters were freely estimated. Means were estimated separately for twins in the same class and in a different classes. Sex, SES and maternal ratings of EXT at age 3 were included as covariates on the INT and EXT scores and were tested for significance in the saturated model. Differences in mean problem behavior between the same and different teacher groups were also tested in the saturated model.

Genetic epidemiological analyses. The classical twin design (see Figure 1a) was used to estimate the influence of genetic, shared environmental and unique environmental factors on INT and EXT data of twins rated by the same teacher and on data of twins rated by different teachers. If estimates for the genetic variance differ between same teacher ratings and different teacher ratings then this constitutes evidence for $G \times E$. $G \times E$ can be confounded with gene–environment correlation, that is, the phenomenon that environmental exposure is associated with individual's genetic make-up (Purcell, 2002). To correct for this effect, means were estimated separately for twins in the same class and in different classes.

Next, differences between parameter estimates for twins rated by the same and by different teachers were tested by constraining them to be equal. Further, sex differences in parameter estimates were tested as well as the significance of the effect of shared environmental factors.

Correlated errors model. Teachers can bring a bias into their ratings which can result in children being rated as more similar when rated by one teacher. As a result, the correlations of twins rated by the same teacher are inflated. The teacher's bias can be accounted for by estimating a correlation between nonshared environment (which also includes measurement error) of twin pairs rated by the same teacher (Simonoff et al., 1998; Figure 1b).

Twin confusion model. Twin confusion is present if a teacher does not always discriminate between the two individuals in a twin pair (Simonoff et al., 1998;

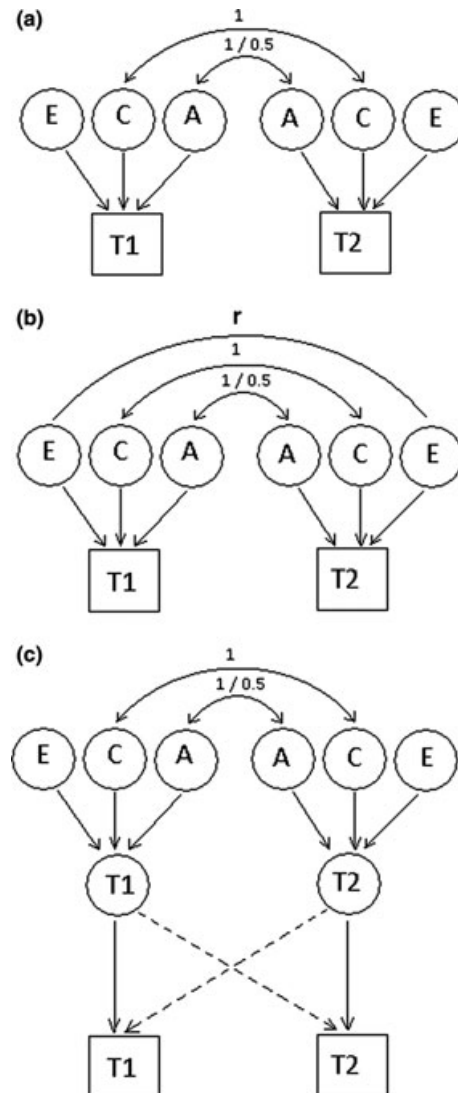


Figure 1 a. Basic genetic model to estimate the effect of additive genetic (A), common environment shared by children in the same family (C) and unique environmental (E) effects on behavioral traits. Latent variables are given in circles, observed variables in squares. The underlying assumption is that E is uncorrelated between twin 1 (T1) and twin 2 (T2), the correlation of C is 1, and A correlates in MZ twins with 1 and in DZ twins with 0.5. b. Correlated errors model; E is correlated (r) when pairs are rated by the same teacher. c. Twin confusion model; behavior of one twin is sometimes ascribed to the other twin when both are rated by the same teacher

Figure 1c). Behavior of one twin (in the figure depicted with a circle) is sometimes ascribed to his or her co-twin (depicted with the broken line pointing to the teacher's ratings of the co-twin).

In all models sex, SES and EXT at age 3 were included as covariates on the means.

Results

Twin pairs rated by the same teacher and by different teachers

Table 1 shows the INT and EXT scores for the same teacher group and the different teacher group at age 3 (rated by the mother) and at age 7, 10 and 12 (rated

Table 1 Raw scores on mean (*SD*) internalizing and externalizing behavior at age 3, 7, 10 and 12

Questionnaire	Age	Internalizing				Externalizing			
		ST	DT	<i>t(df)</i>	<i>p</i>	ST	DT	<i>t(df)</i>	<i>p</i>
CBCL 2/3	3	4.6 (3.8)	4.9 (4.1)	2.2 (3716)	.03	15.6 (9.5)	17.1 (10.1)	4.8 (3490)	.00
TRF	7	4.1 (4.7)	5.1 (5.5)	6.4 (3905)	.00	4.1 (6.8)	4.7 (7.1)	2.7 (4177)	.01
TRF	10	4.4 (5.3)	5.7 (6.1)	7.3 (3214)	.00	4.4 (7.1)	5.3 (8.0)	3.6 (3346)	.00
TRF	12	3.8 (5.0)	5.2 (5.8)	6.3 (1791)	.00	4.1 (7.0)	5.2 (8.1)	3.7 (1838)	.00

ST, same teacher; DT, different teacher; CBCL, Child Behavior Check List; TRF, Teacher Report Form.

by the teacher). The presented statistics are of the oldest twin to avoid paired observations when including both twins of a pair. EXT scores at age 3 were higher in twins who were separated at school than twins who were not separated ($p < .01$). INT scores at age 3 were only marginally higher for the separated twins ($p = .03$). A binary logistic regression showed that EXT at age 3 was positively predictive of being separated at age 7, 10 and 12 (all tests showed $p < .01$). INT at age 3 had no power to predict twin separation at later ages (all tests showed $p > .05$). Binary logistic regression also showed that SES was positively predictive of being in the different teacher group at age 7 and 10, but not at age 12 (for age 7 and 10 $p < .01$, for age 12 $p > .05$).

At age 7, same teacher ratings on INT and EXT were lower than the different teacher ratings. The same pattern was seen at age 10 and 12 (see Table 1). After correcting for the effects of sex, SES and EXT at age 3, scores on EXT at age 7 could be constrained to be equal in the same and different teacher group ($p > .05$). This was not the case for scores on EXT at age 10 and 12, and for scores on INT at all ages.

The effects of sex, SES and EXT at age 3 on INT and EXT were also tested in the means model. Sex had no significant effect on INT scores at age 7, 10 or 12 (all tests showed $p > .05$). There was a significant effect of sex on EXT behavior at age 7, 10 and 12 (all tests showed $p < .01$), with boys scoring higher on EXT than girls. There was an effect of SES on mean INT and EXT scores at each age, except for INT at age 7 ($p > .05$). Mean scores on problem behavior decreased with increasing SES. EXT behavior at age 3 was a significant predictor of mean INT and EXT scores at each age (all tests showed $p < .01$).

Genetic analyses

Twin variances and correlations as a function of zygosity, age, sex and rating by the same or by different teachers are presented in Table 2. Twin pair correlations for MZ twins rated on EXT by the same teacher were higher than MZ twins rated by different teachers, the same holds for the DZ twins. Roughly the same pattern was seen for ratings on INT, with a few exceptions. MZ correlations were higher than DZ correlations, indicating that genetic factors influence problem behavior. In general, the MZ correlations

were about twice as high as the DZ correlations, which indicates little or no influence of shared environmental factors. This was the case for both the same teacher and different teacher group. No influence of shared environment in the same teacher group indicates that the correlated errors model probably does not fit the data, as in this model the presence of shared environment is expected because MZ as well as DZ correlations are inflated due to correlated errors. Furthermore, the variances were about equal for the same and different teacher group, indicating that the twin confusion model is also not likely to describe the data well, as variances of the same teacher ratings are inflated when teachers confuse twins.

The correlated errors, twin confusion and the G × E interaction models were compared to the saturated model (see online appendix Table S1 for fit statistics). At all ages and for both phenotypes the G × E model gave a good fit to the data, except for INT at age 7 which showed a marginally significant deterioration of fit compared to the saturated model ($p = .03$). Both the twin confusion model and the correlated errors model did not fit the data, except for the correlated errors model on INT at age 12. Because the overall trend (a good fit of the G × E model at all ages) we adopted this model for INT and EXT.

The raw estimates and the proportions of variance explained by additive genetic, common and unique environmental factors are presented in Table 3. At all ages, heritability estimates were higher in the same teacher group than in the different teacher group. There was no effect of shared environment at any age for any type of problem behavior.

For INT there was no difference between boys and girls in the amount of variance explained by genetic or unique environmental factors. This was the case for the same teacher group and for the different teacher group. At all ages heritability estimates were about 70% in the same teacher group. Heritability in the different teacher group was nearly 30% at age 7 and 10, and 39% at age 12.

There was a significant effect of sex on the heritability estimates of EXT, although the effect was small. For boys, estimates of the effect of additive genetic factors were between 79% and 82% in the same teacher group. In the different teacher group heritability estimates were between 47% and 53%.

Table 2 Variances and correlations for internalizing (INT) and externalizing (EXT) as a function of twin pairs rated by the same or by different teachers and zygosity

Zyg	Age 7						Age 10						Age 12					
	Same teacher			Different teacher			Same teacher			Different teacher			Same teacher			Different teacher		
	Var	r	N pairs	Var	r	N pairs	Var	r	N pairs	Var	r	N pairs	Var	r	N pairs	Var	r	N pairs
Internalizing																		
MZM	23	.65	315 + 7	25	.27	242 + 5	30	.72	320 + 4	31	.28	182 + 5	27	.68	242 + 2	28	.37	123 + 6
DZM	26	.33	275 + 5	27	.21	267 + 16	33	.47	273 + 4	32	.19	219 + 7	24	.39	214 + 7	26	.18	120 + 4
MZF	25	.71	380 + 4	26	.33	273 + 8	29	.66	408 + 4	29	.22	210 + 16	26	.67	349 + 9	29	.43	142 + 5
DZF	24	.20	302 + 9	22	.10	235 + 9	30	.37	274 + 7	32	.22	174 + 17	25	.37	211 + 6	27	.18	106 + 2
DOS	24/23	.36	622 + 17	24/28	.08	465 + 18	29/28	.34	550 + 18	28/29	.07	410 + 16	26/25	.41	403 + 11	29/26	.20	229 + 9
Externalizing																		
MZM	50	.81	324 + 1	52	.49	246 + 1	61	.78	324 + 2	54	.39	188 + 0	54	.81	244 + 0	54	.50	127 + 2
DZM	50	.37	278 + 3	49	.25	279 + 4	65	.37	279 + 1	69	.37	224 + 2	58	.44	219 + 2	67	.37	121 + 3
MZF	37	.67	388 + 1	39	.55	297 + 2	40	.75	418 + 3	45	.50	220 + 0	38	.74	355 + 3	40	.56	146 + 1
DZF	37	.34	310 + 3	38	.27	233 + 3	40	.40	280 + 2	44	.11	190 + 2	38	.33	216 + 3	46	.35	108 + 0
DOS	54/36	.35	641 + 4	54/41	.20	480 + 4	63/41	.31	565 + 7	64/48	.24	423 + 4	57/40	.31	412 + 2	53/39	.20	235 + 3

N pairs gives first the number of complete pairs and second the number of incomplete pairs. For the DOS pair the variance of the male twin is given first, and the variance of the female twin second. Zyg, zygosity; var, variance; r, correlation; MZM, monozygotic male; DZM, dizygotic male; MZF, monozygotic female; DZF, dizygotic female; DOS, dizygotic opposite sex.

Table 3 Raw estimate and proportion of variance explained by A, C and E for internalizing (INT) and externalizing (EXT) as a function of twin pairs rated by the same or by different teachers, sex and age

Teacher	Age 7						Age 10						Age 12					
	Same teacher			Different teacher			Same teacher			Different teacher			Same teacher			Different teacher		
	A	C	E	A	C	E	A	C	E	A	C	E	A	C	E	A	C	E
Internalizing																		
Boys & girls	16.56	-	7.84	7.34	-	18.15	20.52	-	9.06	7.78	-	22.37	17.06	-	8.18	10.82	-	16.73
% (CI)	68 (64-71)	-	32 (29-36)	29 (22-35)	-	71 (65-78)	69 (66-72)	-	31 (28-34)	26 (18-33)	-	74 (67-82)	68 (64-71)	-	32 (29-36)	39 (30-48)	-	61 (52-70)
Externalizing																		
Boys	42.64	-	9.36	24.01	-	27.25	50.27	-	13.62	29.92	-	32.95	46.51	-	10.24	30.47	-	27.56
% (CI)	82 (79-85)	-	18 (15-21)	47 (38-55)	-	53 (45-62)	79 (75-82)	-	21 (18-25)	48 (38-56)	-	52 (44-62)	82 (78-85)	-	18 (15-22)	53 (41-62)	-	47 (38-59)
Girls	24.40	-	12.25	21.62	-	17.89	30.69	-	10.18	20.80	-	24.11	28.62	-	10.37	23.52	-	18.15
% (CI)	67 (61-71)	-	33 (29-39)	55 (47-62)	-	45 (38-53)	75 (71-78)	-	25 (22-29)	48 (38-56)	-	52 (44-62)	73 (69-77)	-	27 (23-31)	56 (46-65)	-	44 (35-54)

CI = 95% confidence interval.

For girls, the heritability estimates were lower in the same teacher group (67%–75%) and marginally higher (48%–56%) in the different teacher group.

Discussion

The focus of this report was on the etiology of teacher-rated problem behavior at age 7, 10 and 12. Results in twin pairs rated by different teachers were compared to twin pairs rated by the same teacher. Correlations and heritability estimates were higher when one teacher-rated both twins of a pair than when different teachers rated the twins. We tested if twin confusion or correlated errors could explain the difference in the correlations between the two groups. Both models gave a significant worsening of the fit of the model, thus did not provide a good explanation. This indicates an effect of G × E for boys and girls at each age. In other words, different classes with different teachers, peers and possibly classroom climate trigger different behavior depending on the child's genetic make-up. In the same teacher group, shared environment interacts with genetic risk factors, causing the twins to behave more similar in accordance with their genotype (and thus showing a higher heritability). A lower heritability is found in the different teacher group with the nonshared class environment, triggering different behavior in twins.

Our estimates of the effect of additive genetic, shared and unique environmental factors are roughly comparable with studies in which same or different teacher ratings were excluded from the analyses. Hartman et al. (2007) found a heritability of 74% for Attention deficit hyperactivity disorder (ADHD) in 106 twin pairs rated by the same teacher. Towers et al. (2000) estimated the heritability for EXT at 31% and for INT at 15% in a sample of 88 twin pairs rated by different teachers. Given the large differences between the heritabilities in the two groups, it is not surprising that our results differ from studies in which twins rated by the same and by different teachers were analyzed simultaneously (Haberstick, Schmitz, Young, & Hewitt, 2005; Vierikko, Pulkkinen, Kaprio, & Rose, 2006). When taking the same and different teacher ratings' ratio into account, our results are roughly comparable to the findings of Haberstick et al. (2005) in a total sample of 382 twin pairs. Possibly due their smaller sample size, they did not report a difference in heritability between boys and girls in EXT. The estimates that Vierikko et al. (2006) reported from a sample of 1,041 twin pairs were low in comparison to the present findings. This might be explained by differences in phenotypes. This study focused on EXT, which consists of rule breaking and aggressive behavior, while Vierikko et al. (2006) only considered aggression. In contrast to a meta-analysis of the environmental contributions to child and adolescent INT and EXT, no effect of shared environmental factors was detected (Burt, 2009). In the meta-analysis

the effect of shared environment on EXT was estimated around 20% for both sexes, and on INT for boys at 24% and for girls at 10%. These estimates were based on parent, child and teacher ratings. Based on teacher ratings the meta-analysis reported a large dispersion in the effect of shared environmental factors. This is in line with the large dispersion in the effects of genetic factors as reported in the introduction.

To our knowledge, this is the first study investigating the twin confusion and the correlated errors model and G × E interaction in INT and EXT behavior in twins rated by same and different teachers. The twin confusion and correlated errors models have been previously applied to teacher ratings of problem behavior. For attention problems and related traits, Simonoff et al. (1998) found evidence for the twin confusion model, whereas others favored the correlated errors model (Baker et al., 2007; Derks, 2006; Derks et al., 2007; Polderman et al., 2006). It is possible that teacher bias plays a role in behavioral problems in the domain of hyperactivity and attention deficit hyperactivity disorder, but not in INT and EXT problem behavior. However, these measures, especially ADHD and EXT, are significantly correlated.

Based on the current study, it is not possible to distinguish which differences between classes play a role in the G × E interaction. Earlier studies suggest that peer behavior can induce G × E in teacher-rated problem behavior. The genetic component to teacher-rated aggression is moderated by peer victimization (Brendgen et al., 2008). In addition, the estimated heritability of teacher-rated depressive behavior is moderated by peer rejection (Brendgen et al., 2009). Teacher quality can also moderate genetic effects. Taylor, Roehrig, Hensler, Connor, and Schatschneider (2010) reported for early reading, that when teacher quality is low, genetic variance is restricted, whereas, when teacher quality is high, children's reading ability has room to be expressed according to their genetic potential. As these examples show, both teacher and classroom (peer) factors can act as environmental moderators.

It appeared that EXT problem behavior scores of twins rated by different teachers remained higher than of twins rated by the same teacher. Moreover, twins rated by different teachers also scored higher on INT problem behavior than twin pairs that stayed together, while there was only a marginal difference between these groups at age 3. This difference seemed to increase, while the difference in EXT between the two groups remains about the same. This indicates that twin separation had a negative effect on INT and no effect on EXT behavior. Studies on the effect of twin separation at school on problem behavior are relatively scarce. Two studies reported that twin pairs at age 7, in one study rated by the mother and in the other study rated by the teacher, who were separated displayed more problems than

pairs who attended the same classroom. (DiLalla & Mullineaux, 2008; van Leeuwen, van den Berg, van Beijsterveldt, & Boomsma, 2005). A third study only found an effect in teacher ratings of INT problems, and not in other types of problem behavior, at age 7 (Tully et al., 2004). In this third study, the authors concluded that some twins (but not all) will experience significant and long-lasting INT problems due to twin separation. Our findings support this conclusion.

Conclusions

Our study demonstrated $G \times E$ in teacher-rated INT and EXT problem behavior. This implies possibilities for interventions to buffer against the genetic risk of developing INT or EXT behavior during childhood. It is very well possible that the classroom environment consisting of teacher and peers is a very important factor in the expression of problem behavior and can sometimes make the difference between a child that is hard to handle at school versus a child that is not to handle at all and needs to go to a special school. Therapeutic interventions should therefore also involve the school. Specifically applicable to twins, our study showed that separating twins at school has a negative effect on anxious depressed and withdrawn behavior based on ratings of teachers, who observe children in the school environment. It might therefore be advisable for twins not to separate them unless there is an urgent reason.

Acknowledgements

This research was supported by ZonMw, grant number 920-03-268 and NWO 400-03-330 and 451-04-034, and by 'Spinozapremie' (NWO/SPI 56-464-14192); 'Twin-family database for behavior genetics and genomics studies' (NWO 480-04-004); Genetics of Mental Illness: European Research Council (ERC-230374); Genetic influences on stability and change in psychopathology from childhood to young adulthood (NWO/ZonMw 91210020). C.M.M. is supported by NWO (VENI 916-76-125). We are grateful to the twins (and their families) for their participation.

Correspondence to

Diane J. Lamb, MSc, VU University Amsterdam, Department of Biological Psychology, Van der Boechorststraat 1, 1081 BT, Amsterdam, The Netherlands; Email: dj.lamb@vu.nl

Supporting information

Additional supporting information is provided along with the online version of this article.

Table S1 Results of model fitting for internalizing (INT) and externalizing (EXT) at age 7, 10 and 12 (Word document)

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Key points

- Studies that focus on the etiology of teacher ratings are relatively scarce and report a wide range of heritability estimates for problem behavior.
- Twin correlations for pairs rated by the same teacher are higher than for pairs rated by different teachers, some studies argue due to teacher bias.
- Using structural equation modeling, this study finds that teachers are not biased and that gene–environment interaction explains the difference in correlations between twin pairs rated by the same teacher and pairs rated by different teachers.
- The findings are clinically relevant because it implies the opportunity for teacher based interventions.

References

- Achenbach, T.M. (1991a). *Manual of the Child Behavior Checklist/4-18 and 1991 profile*. Burlington, VT: University of Vermont, Department of Psychiatry.
- Achenbach, T.M. (1991b). *Integrative guide for the 1991 CBCL/4-18, YSR, and TRF profiles*. Burlington, VT: Dept. of Psychiatry, University of Vermont.
- Achenbach, T.M., & Rescorla, L.A. (2001). *Manual for ASEBA school-age forms & profiles*. Burlington, VT: University of Vermont Research Center for Children, Youth, & Families.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (4th ed., text rev.)*. Washington, DC: Author.
- Baker, L.A., Jacobson, K.C., Raine, A., Lozano, D.I., & Bezdjian, S. (2007). Genetic and environmental bases of childhood antisocial behavior: A multi-informant twin study. *Journal of Abnormal Psychology, 116*, 219–235.
- Bartels, A., van den Oord, E.J.C.G., Hudziak, J.J., Rietveld, A.J.H., van Beijsterveldt, C.E.M., & Boomsma, D.I. (2004). Genetic and environmental mechanisms underlying stability and change in problem behaviors at ages 3, 7, 10, and 12. *Developmental Psychology, 40*, 852–867.
- Boomsma, D., Busjahn, A., & Peltonen, L. (2002a). Classical twin studies and beyond. *Nature Reviews Genetics, 3*, 872–882.
- Boomsma, D., & Martin, N.G. (2002). gene–environment interactions. In H. D'haenen, J.A. den Boer, & P. Willner (Eds.), *Biological psychiatry* (pp. 181–187). New York: John Wiley & Sons Ltd.
- Boomsma, D.I., Vink, J.M., van Beijsterveldt, T.C., de Geus, E.J.C., Beem, A.L., Mulder, E.J.C., ... & van Baal, G.C. (2002b). Netherlands Twin Register: A focus on longitudinal research. *Twin Research and Human Genetics, 5*, 401–406.

- Bradley, R.H., & Corwyn, R.F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, 53, 371–399.
- Brendgen, M., Boivin, M., Vitaro, F., Girard, A., Dionne, G., & Perusse, D. (2008). gene–environment interaction between peer victimization and child aggression. *Development and Psychopathology*, 20, 455–471.
- Brendgen, M., Vitaro, F., Boivin, M., Girard, A., Bukowski, W.M., Dionne, G., ... & Perusse, D. (2009). gene–environment interplay between peer rejection and depressive behavior in children. *Journal of Child Psychology and Psychiatry*, 50, 1009–1017.
- Burt, S.A. (2009). Rethinking environmental contributions to child and adolescent psychopathology: A meta-analysis of shared environmental influences. *Psychological Bulletin*, 135, 608–637.
- Chicago: SPSS Inc. (2008). PAWS Statistics 17.0 [Computer software].
- Deater-Deckard, K. (2001). Annotation: Recent research examining the role of peer relationships in the development of psychopathology. *Journal of Child Psychology and Psychiatry*, 42, 565–579.
- Derks, E.M. (2006). *Assessment and genetic aetiology of attention problems, hyperactivity and related disorders*. Unpublished doctoral dissertation, UU University, Amsterdam.
- Derks, E.M., Dolan, C.V., Hudziak, J.J., Neale, M.C., & Boomsma, D.I. (2007). Assessment and etiology of attention deficit hyperactivity disorder and oppositional defiant disorder in boys and girls. *Behavior Genetics*, 37, 559–566.
- Derks, E.M., Hudziak, J.J., van Beijsterveldt, C.E.M., Dolan, C.V., & Boomsma, D.I. (2006). Genetic analyses of maternal and teacher ratings on attention problems in 7-year-old Dutch twins. *Behavior Genetics*, 36, 833–844.
- DiLalla, L.F., & Mullineaux, P.Y. (2008). The effect of classroom environment on problem behaviors: A twin study. *Journal of School Psychology*, 46, 107–128.
- Eaves, L.J. (1984). The resolution of genotype × environment interaction in segregation analysis of nuclear families. *Genetic Epidemiology*, 1, 215–228.
- Erikson, R., Goldthorpe, J.H., & Portocarero, L. (1979). Inter-generational class mobility in 3 western European societies – England, France and Sweden. *British Journal of Sociology*, 30, 415–441.
- Haberstick, B.C., Schmitz, S., Young, S.E., & Hewitt, J.K. (2005). Contributions of genes and environments to stability and change in externalizing and internalizing problems during elementary and middle school. *Behavior Genetics*, 35, 381–396.
- Hartman, C.A., Rhee, S.H., Willcutt, E.G., & Pennington, B.F. (2007). Modeling rater disagreement for ADHD: Are parents or teachers biased? *Journal of Abnormal Child Psychology*, 35, 536–542.
- Jöreskog, K.G., & Sörbom, D. (2002). *PRELIS 2 user's reference guide: A program for multivariate data screening and data summarization: A preprocessor for LISREL*. Lincolnwood: Scientific Software International.
- Ladd, G.W., & Burgess, K.B. (1999). Charting the relationship trajectories of aggressive, withdrawn, and aggressive/withdrawn children during early grade school. *Child Development*, 70, 910–929.
- van Leeuwen, M., van den Berg, S.M., van Beijsterveldt, T.C.E.M., & Boomsma, D.I. (2005). Effects of twin separation in primary school. *Twin Research and Human Genetics*, 8, 384–391.
- Neale, M.C., Boker, S.M., & Maes, H. (2006). *Mx statistical modeling* (7th edn). VCU Box 900126, Richmond, VA 23298: Dept of Psychiatry.
- Polderman, T.J.C., Posthuma, D., De Sonneville, L.M.J., Verhulst, F.C., & Boomsma, D.I. (2006). Genetic analyses of teacher ratings of problem behavior in 5-year-old twins. *Twin Research and Human Genetics*, 9, 122–130.
- Purcell, S. (2002). Variance components models for gene–environment interaction in twin analysis. *Twin Research*, 5, 554–571.
- Saudino, K.J., 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. *Journal of Abnormal Child Psychology*, 33, 113–130.
- Simonoff, E., Pickles, A., Hervas, A., Silberg, J.L., Rutter, M., & Eaves, L. (1998). Genetic influences on childhood hyperactivity: Contrast effects imply parental rating bias, not sibling interaction. *Psychological Medicine*, 28, 825–837.
- Standard Classification of Occupations. (2001). *Central Bureau of Statistics*. Available from: <http://www.cbs.nl/nl-L/menu/methoden/classificaties/overzicht/sbc/1992/default.htm> [last accessed 20 September 2011].
- Taylor, J., Roehrig, A.D., Hensler, B.S., Connor, C.M., & Schatschneider, C. (2010). Teacher quality moderates the genetic effects on early reading. *Science*, 328, 512–514.
- Towers, H., Spotts, E., Neiderhiser, J.M., Hetherington, E.M., Plomin, R., & Reiss, D. (2000). Genetic and environmental influences on teacher ratings of the Child Behavior Checklist. *International Journal of Behavioral Development*, 24, 373–381.
- Tully, L.A., Moffitt, T.E., Caspi, A., Taylor, A., Kiernan, H., & Andreou, P. (2004). What effect does classroom separation have on twins' behavior, progress at school, and reading abilities? *Twin Research*, 7, 115–124.
- Vierikko, E., Pulkkinen, L., Kaprio, J., & Rose, R.J. (2006). Genetic and environmental sources of continuity and change in teacher-rated aggression during early adolescence. *Aggressive Behavior*, 32, 308–320.

Accepted for publication: 17 October 2011

Published online: 23 April 2012