

# Genetic Analyses of Teacher Ratings of Problem Behavior in 5-Year-Old Twins

Tinca J. C. Polderman,<sup>1,4</sup> Daniëlle Posthuma,<sup>1</sup> Leo M. J. De Sonneville,<sup>2,3</sup> Frank C. Verhulst,<sup>4</sup> and Dorret I. Boomsma<sup>1</sup>

<sup>1</sup> Department of Biological Psychology, Vrije Universiteit of Amsterdam, Amsterdam, the Netherlands

<sup>2</sup> Department of Pediatrics, Vrije Universiteit Medical Center, Amsterdam, the Netherlands

<sup>3</sup> Department of Clinical Child and Adolescent Studies, Faculty of Social Sciences, Leiden University, Leiden, the Netherlands

<sup>4</sup> Department of Child and Adolescent Psychiatry, Erasmus MC — Sophia, Rotterdam, the Netherlands

Behavioral problems in young children can be assessed by asking their parents or teachers to rate their behaviors. Genetic analyses of parental ratings show relatively large heritabilities for emotional and behavioral problems in young children, but data from teachers for this age group are scarce. Sources of variation in the Teacher's Report Form (TRF) problem scales were examined. The TRF was completed for 211 Dutch 5-year-old twin pairs and 4 single twins. Twins rated by different teachers had higher means and variances than twins rated by the same teacher, in addition twin correlations were lower in this group. In both groups monozygotic (MZ) correlations were generally higher than dizygotic (DZ) correlations. A model for twin resemblance was tested that allowed for these effects. For 5 problem scales (Withdrawn, Social Problems, Aggressive Behavior, Rule Breaking Behavior and Attention Problems) a model with genetic and unique environmental sources of variation fitted best to the data. For 3 problem scales (Anxious/Depressed, Thought Problems and Somatic Complaints) there were familial influences but it was not possible to distinguish between common environmental influences or genetic influences. Heritability was 63% for Attention problems, around 45% for Withdrawn, Social Problems, Aggressive Behavior and Rule Breaking Behavior, and around 30% for Anxious/Depressed, Thought Problems and Somatic Complaints.

Behavioral and emotional problems such as overactivity, anxiety, lack of attention or aggression are relatively common in young children (Bartels et al., 2004; Campbell, 1995; Koot et al., 1997; van den Oord et al., 1996). High levels of these problem behaviors are stable over time and not limited to clinical groups. Several studies have reported stability from childhood into adulthood (Campbell & Ewing, 1990; Lavigne et al., 1998; Roza et al., 2003; Verhulst & van der Ende, 1992). For this reason, early detection and knowledge about the etiology of childhood behavior problems is crucial. Keenan and Wakschlag (2000, 2002) assume that in preschool

children, problem behavior like aggression is less entrenched than later in life, and that behavioral control emerges during this developmental period. Therefore, intervention at this young age is more effective than later in childhood.

Individual differences in parent-reported problem behavior in children aged 2 to 7 years are, to a large degree, genetically influenced. Studies using Child Behavior Checklist (CBCL; Achenbach, 1991a) data from the Netherlands Twin Registry (NTR) found genetic influences for both internalizing and externalizing problems in preschool children. Rietveld et al. (2003) and van den Oord et al. (1996) studied overactive behavior and attention problems in 3-year-old twin pairs. Their results showed heritability estimates around 68%. Derks et al. (2004) investigated aggressive, oppositional, overactive, withdrawn and anxious behavior in 9689 3-year-old twin pairs. They found high genetic contributions for all problem scales (40% to 70%), except for oppositional behavior where 20% (as rated by the father) to 32% (as rated by the mother) of the variance was explained by genetic influences.

Using the Devereux Child Behavior Rating Scale (DCB; Spivack & Spotts, 1966) in a sample of 8041 5-year-old twin pairs, Van Beijsterveldt et al. (2004) reported heritabilities of 50 to 80% for maternal ratings of aggression, anxiety, dependency, emotional lability and attention problems.

Research from other twin registries also found genetic influences on parental reports of behavioral problems in young children. In a sample of 199 Norwegian 5- and 6-year-old twin pairs, Gjone et al. (1996) reported a heritability of 73% for internalizing problems. Schmitz et al. (1995) reported for 260 2- and 3-year-old twin pairs of the Colorado Twin Registry, heritabilities of between 29 and 52% for

Received 3 May, 2005; accepted 11 July, 2005.

Address for correspondence: Tinca Polderman, Department of Biological Psychology, Vrije Universiteit of Amsterdam, Van der Boechorststraat 1, 1081 BT Amsterdam, the Netherlands. E-mail: [jc.polderman@psy.vu.nl](mailto:jc.polderman@psy.vu.nl)

anxious behavior, withdrawn, sleep problems, somatic complaints, aggression and destructive behavior.

Genetic research on young children's problems is often based on parental reports while behavioral data from teachers about preschool children are scarce. Parents are an important source for the assessment of their children's problems (Arseneault et al., 2003; Bartels et al., 2003; van der Valk et al., 2001). However, situational variation in children's behaviors at home and at school makes teachers another important source of information. For example, teachers may have a unique view on problems that are specific to the classroom or other school situations, such as problems in the social interactions with other children, or task-oriented situations. Teachers also have an advantage over parents in their wide exposure to children of the same age, which makes them able to compare the child's behavior with that of many same-aged peers (Verhulst et al., 1997).

Zahn-Waxler et al. (1996) studied internalizing and externalizing problems and related problems of attention and hyperactivity in 5-year-old twins. They used parental data collected with the CBCL, and teacher data collected with the Preschool Behavior Questionnaire (PBQ; Behar & Stringfield, 1974). For teacher and mother reports they found significant genetic influences for all three problem scales. The father reports showed significant genetic influences for externalizing and attention/hyperactivity problems but not for internalizing problems. Genetic variation in antisocial behavior was investigated in a representative-plus-high-risk sample by Arseneault et al. (2003). For 1116 5-year-old twins they had data from mothers, teachers, examiner-observers and children themselves. They reported high heritabilities ranging from 42% based on children's self-report to 76% for the teacher reports.

From genetic studies using teacher data in older twin samples, we know that twin correlations can be much higher for twin pairs rated by the same teacher than for twin pairs rated by different teachers (Derks et al., 2005; Simonoff et al., 1998; Towers et al., 2000; Vierikko et al., 2004). According to Simonoff et al. (1998) and Derks et al. (2005), these high twin correlations are not due to twin confusion (i.e., the teacher confuses the members of a twin pair) but are likely to be associated with teacher and classroom characteristics. Teacher ratings of a particular child's behavior may be influenced by the teacher's expectations about normal and abnormal behavior, by aspects of the relationship between a teacher and a particular child and by influences the teacher imposes on the classroom as a whole, such as a specific educational approach.

The age of 5 corresponds with the developmental transition from preschool to elementary school and marks an important change in daily occupation, and social and cognitive functioning. In the Dutch system, school is obligatory from the age of 5 years onward, but most children start school when they are 4 years

old. In 2 years, from the age of 4 to 6 years, children are expected to adapt to the social, emotional and cognitive demands which involve their school participation. Genetic studies covering this important developmental time span are rare. Of the four studies that investigated the behavior of 5-year-old twin pairs, only Zahn-Waxler et al. (1996) and Arseneault et al. (2003) used teacher data. The other studies (i.e., Gjone et al., 1996; Van Beijsterveldt et al., 2004) used parental data.

In the present study Teacher's Report Form data (TRF; Achenbach, 1991b) of 211 Dutch 5-year-old twin pairs and four single twins were analyzed. The first aim of this study was to examine the sources of variation of eight specific problem scales of the TRF (Withdrawn, Anxious/Depressed, Social Problems, Aggressive Behavior, Rule Breaking Behavior, Attention Problems, Thought Problems and Somatic Complaints).

A second aim of the study was to explore whether teacher-specific styles influence the teacher's ratings. In the sample, 126 twin pairs shared the same teacher and 89 twin pairs had a different teacher. A model was evaluated in which the nonshared environmental component was allowed to correlate in children who were rated by the same teacher.

Power analyses showed that the power to detect sex differences in heritabilities was low (see Appendix). Therefore, male and female twin data for both zygositys were combined. A four group design was used for the analyses: monozygotic (MZ) twins rated by the same teacher, MZ twins rated by different teachers, dizygotic (DZ) twins rated by the same teacher and DZ twins rated by a different teacher.

## Materials and Methods

### Subjects

The twins were registered at birth with the NTR kept by the department of Biological Psychology of the Vrije Universiteit in Amsterdam. Of all multiple births in the Netherlands, 40 to 50% are registered by the NTR (Boomsma, 1998; Boomsma et al., 2002). Parents of the twins received a behavior questionnaire every 2 years (CBCL; Achenbach, 1991a). A subsample of all twin pairs participated in a study on neuropsychological development and attention (Groot et al., 2004; Stins et al., 2005). This sample consisted of 237 twin pairs (mean age = 5.8 years,  $SD = 0.1$ ). The sample was selected on the basis of age, city of residence, and zygosity which was determined on the basis of DNA polymorphisms. For practical reasons, children had to live within 100 kilometer radius of the Vrije Universiteit. None of the children suffered from severe physical or mental handicaps. Parents signed an informed consent form.

The TRF was returned for 215 first-born twins and for 213 second-born twins. IQ data and behavioral ratings by the parents were available for all 237 twin pairs. TRF responders and TRF nonresponders were compared on these variables. There were no significant

differences between responders and nonresponders in total IQ score, socioeconomic status or the problem scales of the DCB, as reported by the parents.

Most twin pairs ( $N = 126$ ) shared the same classroom and the same teacher, 89 twin pairs were in parallel classrooms and were assessed by different teachers. There is no official policy about separating twin pairs when they go to school. However, Dutch twin organisations advise parents and schools to do so, which is reflected by the fact that large schools often enforce such policies. The decision to separate a twin pair thus may be based on parental choice, may be enforced by the school or, in the case of smaller schools, is not possible because there is simply only one classroom for a particular age group.

Only one TRF was returned for four of the twin pairs, (i.e., for the first-born twin or for the second-born twin) of which three twin pairs had different teachers and one twin pair had the same teacher. Two children were excluded because of missing items. The final sample of twin pairs consisted of 45 MZ twin pairs rated by different teachers (MZ-DT), 44 DZ twin pairs rated by different teachers (DZ-DT), 67 MZ twin pairs rated by the same teacher (MZ-ST) and 59 DZ twin pairs rated by the same teacher (DZ-ST).

#### Procedure

The children were visited at home to perform a neuropsychological test battery consisting of an intelligence test and several tasks of the Amsterdam Neuropsychological Tasks (ANT; De Sonneville, 1999). Behavioral data were collected with the DCB (Spivack & Spotts, 1966) filled in by the parents, and after permission from the parents, the TRF (Achenbach, 1991b; Verhulst et al., 1997) filled in by the teachers.

The TRF consists of two parts. The first part contains questions about daily functioning and school results of the child, the second part consists of 120 problem items. Teachers are instructed to rate the child's behavior over the last 2 months with 0 if the behavior is not true, 1 if the behavior is sometimes or somewhat true, and 2 if the behavior is very or often true. Items can be scored on eight specific problem scales, two broad band scales (internalizing and externalizing) and a total problem scale. The eight specific problem scales are Withdrawn, Anxious/Depressed, Social Problems, Aggressive Behavior, Rule Breaking Behavior, Attention Problems, Thought Problems and Somatic Complaints. As the data were not normally distributed, problem scores were square-root transformed.

#### Analyses

A MANOVA was performed in SPSS 11.5 (2002) to test for possible differences in problem behavior between twins who shared the same teacher and twins who had different teachers, as reported by the mother with the CBCL at age 3, and the DCB at age 5. Internalizing and externalizing broad band scales were used with the CBCL data, and seven specific problem scales were used for the DCB data. The CBCL age 2/3

is described by Derks et al. (2004) and the derived problem scales of the DCB are described by Van Beijsterveldt et al. (2004).

Genetic analyses were carried out using the statistical software package Mx (Neale, 1999). The total variation in problem scores was decomposed into sources of additive genetic variance (A), common environmental variance (C) and unique environmental variance (E). The A variance is due to additive effects of different alleles, C is due to environmental influences shared by members of a twin pair, and E is due to environmental influences not shared by members of a twin pair. E also includes measurement error and is therefore always included in the models.

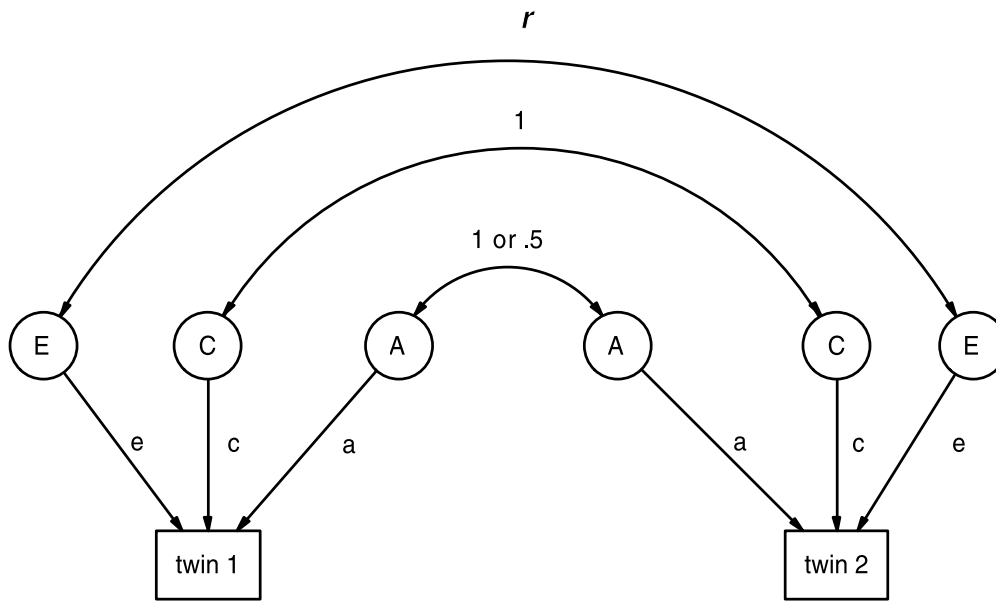
Full ACE models were fitted to the data of each problem scale. To obtain the most parsimonious model for each problem scale, AE and CE models were compared to the full ACE model using the likelihood ratio test which is computed by taking twice the difference between the log-likelihood of the full model and the log-likelihood of a reduced model. The associated degrees of freedom are computed as the difference in degrees of freedom between the two nested models.

The possibility that teachers bring in their own influences when rating the children was tested with the model as proposed by Simonoff et al. (1998), where the increase in correlation for twins rated by the same teachers is reflected as a correlation different from zero between the latent unique environmental influences (E) of the two members of the twin pairs who share the same teacher (see Figure 1a).

## Results

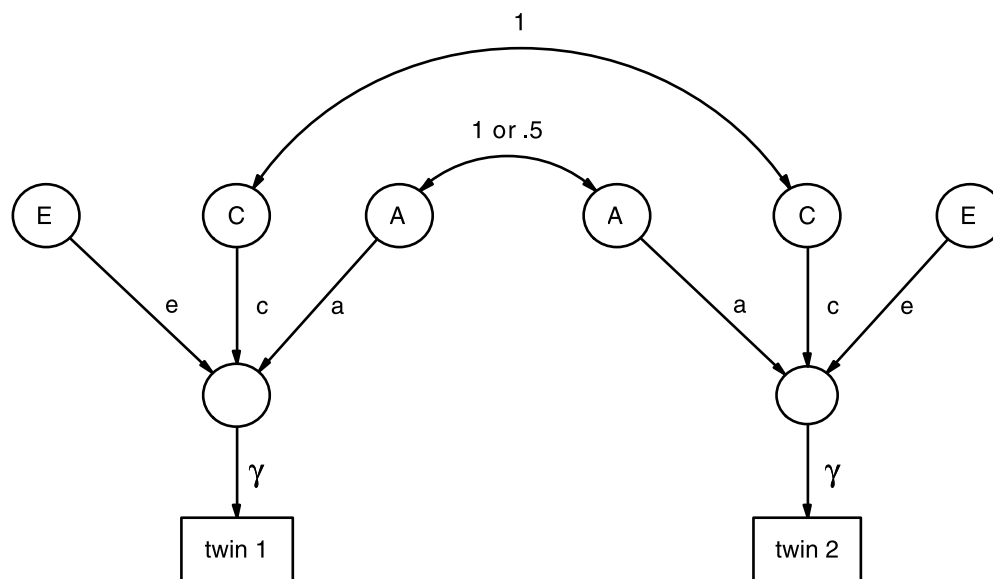
The results of the MANOVA showed that there were no significant differences in problem behavior between twins who shared the same teacher and children who were in separate classrooms, as rated by the mother at age 3 and age 5 ( $p = .24$ ). When examining the univariate results, only the aggression scale of the DCB showed a significant difference ( $p = .087$  for the first-born twin,  $p = .009$  for the second-born twin). No significant differences were found at the age of 3 years ( $p > .7$ ).

Using Mx, it was established in a saturated model that means and variances of the TRF problem scales were equal for both members of a twin pair, and for MZ and DZ twins. Teacher differences (i.e., if the ratings by different teachers differed from the ratings by the same teacher) in means and variances were then tested. The means and variances of twins who were rated by the same teacher were significantly lower than the means and variances of twins rated by different teachers for all problem scales, except for aggressive behavior. We then tested if twin correlations were equal for both groups. Twin correlations of twin pairs who were rated by the same teacher were significantly higher than twin correlations of children who were rated by different teachers. Table 1 shows means and standard deviations (based on

**Figure 1a**

Path diagram for twins rated by the same teacher.

A is the additive genetic variance, correlated 1 for MZ twins and .5 for DZ twins, C is the shared environmental variance, correlated 1 for MZ and DZ twins. E is the unique environmental variance, correlated  $r$  for MZ and DZ twins who share the same teacher.

**Figure 1b**

Path diagram for twins rated by different teachers.

The variances of twins who were rated by different teachers were modeled as scalar ( $\gamma$ ) times the total variance of twins rated by the same teacher. Path coefficients  $a$ ,  $c$  and  $e$  are equal for both groups.

untransformed data), and twin correlations (based on transformed data) according to zygosity and same versus different teacher.

As the variances of twin pairs who were rated by different teachers were higher than the variances of twin pairs rated by the same teacher, a scalar model was fitted in which the variances of twins who were rated by different teachers were modeled as scalar

times the total variance of twins rated by the same teacher (see Figure 1b). Table 2 shows the genetic modeling analyses for each problem scale. The correlation ( $r$ ) of the unique environmental influences in MZ and DZ twin pairs who shared the same teacher was substantial (.35 to .66) and could not be left out of the model without significantly reducing the fit of the model to the data.

**Table 1**

Means and Standard Deviations (Based on Untransformed Data), and Twin Correlations With Confidence Intervals (Based on Transformed Data) for Each Problem Scale, for MZ and DZ Twin Pairs Rated by the Same Teacher (ST) and for MZ and DZ Twin Pairs Rated by Different Teachers (DT)

	Means DT (means girls/ boys)	SD	Twin correlations MZ-DT ( <i>n</i> = 45)	Twin correlations DZ-DT ( <i>n</i> = 44)	Means ST (means girls/ boys)	SD	Twin correlations MZ-ST ( <i>n</i> = 67)	Twin correlations DZ-ST ( <i>n</i> = 59)
Withdrawn	2.14/2.53	2.83	.48 (.25-.66)	.05 (-0.29-0.36)	1.29/1.42	1.80	.74 (.62-0.82)	.42 (.15-.61)
Anxious/depressed	3.07/3.72	4.01	.19 (-.10-.44)	.41 (.11-.62)	1.57/2.72	2.88	.71 (0.57-0.80)	.45 (.22-.62)
Social problems	1.70/2.35	2.72	.39 (.14-.59)	.31 (-.03-.56)	0.77/1.78	1.95	.71 (0.58-0.80)	.49 (.26-.646)
Aggressive behavior	3.33/7.18	6.75	.40 (.16-.59)	.21 (-.16-.50)	2.97/4.90	6.07	.84 (.76-.89)	.43 (.16-.62)
Rule breaking behavior	0.43/0.93	1.22	.34 (.09-.54)	.41 (.09-.63)	0.28/0.47	0.84	.82 (.73-.88)	.48 (.24-.65)
Attention problems	5.22/7.77	6.69	.67 (.48-.78)	.31 (-.03-.56)	2.79/5.43	5.35	.81 (.71-.87)	.58 (.39-.71)
Thought problems	0.27/0.60	0.97	.28 (.04-.49)	.28 (-.17-.57)	0.20/0.31	0.69	.62 (.47-.73)	.40 (.09-.61)
Somatic complaints	0.61/0.92	1.02	.23 (-.003-.44)	.39 (-.22-.66)	0.26/0.30	0.66	.55 (.37-.68)	.34 (.05-.56)

Reduced models were compared to the full ACER model. Analyses showed that for all problem scales, C could be dropped from the full model. However, for the problem scales Anxious/Depressed, Thought Problems and Somatic Complaints it was possible to drop A or C from the full model, indicating that the variance was explained by familial influences, however it was not possible to distinguish between genetic or common environmental influences. E contributed most to the total variance (around .60) except for Attention Problems (.39).

## Discussion

Five years of age is a vital period of progress in which children change from dependent, needy toddlers to self-assured, independent school children. The development of social competence and emotion regulation in preschool children is crucial as it can predict both emotional disorders and academic outcomes such as school readiness, and positive attitudes toward school (Blair et al., 2004). Therefore, it is important to examine possible underlying sources of variation in this stage of development. However, genetic research involving the behavior of 5-year-old children is scarce. Genetic studies containing teacher ratings about the behavior of children at this young age are especially limited. In the current study teacher reports were assessed in 211 5-year-old twin pairs and four single twins. The relative contribution of genetic and environmental influences were estimated for individual differences in eight specific problem scales of the TRF (Achenbach, 1991b). Furthermore whether or not teachers bring in specific influences to their ratings was tested.

It was found that genetic influences explained around 40% of the total variance, and unique envi-

ronmental influences around 60% for Withdrawn, Social Problems, Aggressive Behavior and Rule Breaking Behavior. For Attention Problems the genetic component explained around 60% of the variation and the unique environmental component about 40%. For Anxious/Depressed, Thought Problems and Somatic Complaints, familial influences were found but no distinction could be made between common environmental influences and genetic influences.

It is known that in addition to information from parents about a child's behavior, teacher information is valuable as well (van der Ende & Verhulst, 2005). The school situation enables teachers to compare the behavior of one child with the behavior of many other same-aged, older or younger children, and to judge whether the behavior of the child is appropriate for his or her age. However, as parents may have a possible rater bias by lacking internal standards to determine 'normal' levels of behavior, teachers may too have their own kinds of 'bias'.

In our results the means and variances for all problem scales were higher in twins rated by different teachers compared to twins rated by the same teacher. This was neither due to the fact that more boys were in parallel classrooms than girls (as boys generally tend to have higher means than girls), nor to the fact that children were in different schools, for example, specific schools for learning or behavioral problems. The question arises then why twins who were rated by different teachers appear to have more behavioral and emotional problems? A first explanation may be that twins who were rated by different teachers were probably separated from each other for the first time in their lives. Their increased levels of problems may be a reaction to this separation. The reverse may also be true. It may be that problematic twins were separated



**Table 2**  
Model Fitting Results

	Model	$\Delta\chi^2$	$\Delta df$	$p$	AIC	A or C	E	$r$	$\Upsilon$
Withdrawn	ACE $r$	1.88	1	.17	-0.12				
	ACE	6.67	1	.01	4.66				
	<b>AEr</b>	<b>0.00</b>	<b>1</b>	—	<b>-2.00</b>	<b>.49</b>	<b>.51</b>	<b>.47</b>	<b>1.63</b>
	CE $r$	13.26	1	.00	11.26				
	Er	21.49	2	.00	17.49				
Anxious/depressed	ACE $r$	4.48	1	.03	2.48				
	ACE	9.56	1	.00	7.57				
	<b>AEr</b>	<b>0.02</b>	<b>1</b>	<b>.90</b>	<b>-1.98</b>	<b>.34</b>	<b>.66</b>	<b>.52</b>	<b>1.36</b>
	CE $r$	3.16	1	.08	1.16	.29	.71	.42	1.35
	Er	10.47	2	.01	6.47				
Social problems	ACE $r$	0.36	1	.55	-1.64				
	ACE	7.32	1	.01	5.32				
	<b>AEr</b>	<b>0.08</b>	<b>1</b>	<b>.77</b>	<b>-1.92</b>	<b>.42</b>	<b>.58</b>	<b>.49</b>	<b>1.42</b>
	CE $r$	4.97	1	.03	2.97				
	Er	16.98	2	.00	12.98				
Aggressive behavior	ACE $r$	3.19	1	.07	1.19				
	ACE	18.84	1	.00	16.84				
	<b>AEr</b>	<b>0.00</b>	<b>1</b>	—	<b>-2.00</b>	<b>.49</b>	<b>.51</b>	<b>.66</b>	<b>1.01</b>
	CE $r$	20.88	1	.00	18.88				
	Er	30.96	2	.00	26.96				
Rule breaking behavior	ACE $r$	4.11	1	.04	2.11				
	ACE	17.91	1	.00	15.91				
	<b>AEr</b>	<b>0.00</b>	<b>1</b>	—	<b>-2.00</b>	<b>.45</b>	<b>.55</b>	<b>.65</b>	<b>1.55</b>
	CE $r$	13.04	1	.00	11.04				
	Er	25.21	2	.00	21.21				
Attention problems	ACE $r$	0.51	1	.48	-1.85				
	ACE	4.62	1	.03	2.62				
	<b>AEr</b>	<b>0.67</b>	<b>1</b>	<b>.41</b>	<b>-1.33</b>	<b>.63</b>	<b>.37</b>	<b>.49</b>	<b>1.56</b>
	CE $r$	14.76	1	.00	12.76				
	Er	40.50	2	.00	36.50				
Thought problems	ACE $r$	0.67	1	.41	-1.33				
	ACE	6.64	1	.01	4.64				
	<b>AEr</b>	<b>0.00</b>	<b>1</b>	—	<b>-2.00</b>	<b>.32</b>	<b>.68</b>	<b>.43</b>	<b>1.49</b>
	CE $r$	2.56	1	.11	0.56	.28	.72	.37	1.47
	Er	9.75	2	.00	5.75				
Somatic complaints	ACE $r$	1.33	1	.25	-0.67				
	ACE	4.03	1	.05	2.03				
	<b>AEr</b>	<b>0.00</b>	<b>1</b>	—	<b>-2.00</b>	<b>.27</b>	<b>.73</b>	<b>.35</b>	<b>1.52</b>
	CE $r$	1.27	1	.26	-0.73	.26	.74	.29	1.52
	Er	6.74	2	.03	2.74				

Note: Submodels ACE, AEr, CE $r$  and Er are compared with ACE $r$  model, which in turn is compared with the fully saturated model.

Bold indicates the best fitting model based on Akaike's Information Criterion.

in school because parents or teachers expected that these twins would be easier to handle apart from each other, in separate classrooms. The differences in teacher ratings may also reflect differences related to location and size of the school: it is likely that children from larger towns and villages have more opportunity to go to separate classrooms than children from smaller villages in which only one school for elementary

education is available and in which no parallel classes exist. Tully et al. (2004) investigated the effects of classroom separation on the behavior of twins when separated at the age of 5 or 7 years. Compared to nonseparated twins, separated twins showed significantly more internalizing problems at both ages, as rated by the teacher. As some twins were not separated at age 5 but only at age 7, the investigators

could test if the internalizing problems were already present at age 5. This was not the case, suggesting that the separation caused these children's emotional problems. No differences between separated and nonseparated twins were found for externalizing problems or ADHD. Our TRF results partly confirmed these findings. We found differences in mean problem scores for Withdrawn and Anxious/Depressed, and Social Problems (see Table 1). Tully et al. (2004) did not observe differences for Attention Problems, however, whereas we did.

Van Leeuwen et al. (2005) analyzed mother ratings (CBCL at ages 3, 7 and 12 years) and teacher ratings (TRF at ages 7 and 12 years) in a much larger group of twins from the NTR. No TRF data at age 5 were available as the NTR routinely collects TRF data only in children of 7 years and older. Van Leeuwen et al. (2005) found that separating the twins when going to school led to more internalizing problems at age 7, but that these problems were no longer present when the children were 12 years old. Twins who were separated at age 5 already had more externalizing problems at age 3. These externalizing problems persisted in time but when the data were corrected for these problems at age 3, there were no additional effects at older ages. This suggested that the decision to separate the twins when starting school was based in part on existing externalizing problems. However, in the current, much smaller sample, there were no differences in problem behavior between twins sharing a classroom and twins who were in separate classrooms according to maternal ratings of these twins at age 3, and at age 5.

Twin correlations in twins who were rated by different teachers were lower than twin correlations in twin pairs rated by the same teacher. Twin studies using teacher data in older children (age 12 to 14 years) showed the same pattern (Simonoff et al., 1998; Towers et al., 2000; Vierikko et al., 2004). An explanation for the high resemblance in twins rated by the same teacher is that these ratings include teacher-specific influences. The results of the current study confirmed that specific teacher styles influenced their ratings. With the unique correlation ( $r$ ) estimated around 50%, it was not possible to drop  $r$  from the model without worsening the fit significantly.

Specific teacher styles can cover a whole range of domains, including personal values and pedagogic qualities to school systems and educational approaches. Some teachers, for example, prefer strict rules in the classroom, whereas others have a more lenient style. Some children prosper better under free conditions whereas others need a structured environment. An accurate way to explore this phenomenon is by obtaining ratings of the twins from multiple teachers. This is perhaps not an impossible job, as children change teachers almost every school year.

A final question is how to describe the teacher-specific influences as reflected in our model. 'Correlated error' as proposed by Simonoff et al. (1998), suggests

a real bias in the teacher ratings, but this is not the case. 'Uniquely correlated' might be a better description in this context.

### Acknowledgments

This work was supported by NWO grant 904-57-94 and NWO 575-25-006, and Zorgonderzoek Nederland (grant 28-2523), The Hague, the Netherlands.

### References

- Achenbach, T. M. (1991a). *Manual for the Child Behavior Checklist/4-18*. Burlington, VT: University of Vermont, Department of Psychiatry.
- Achenbach, T. M. (1991b). *Manual for the Teacher's Report Form*. Burlington, VT: University of Vermont, Department of Psychiatry.
- Arseneault, L., Moffitt, T. E., Caspi, A., Taylor, A., Rijdsdijk, F. V., Jaffee, S. R., Ablow, J. C., & Measelle, J. R. (2003). Strong genetic effects on cross-situational antisocial behavior among 5-year-old children according to mother, teachers, examiner-observers, and twins' self-reports. *Journal of Child Psychology and Psychiatry*, *44*, 832-848.
- Bartels, M., van den Oord, E. J. C. G., Hudziak, J. J., Rietveld, M. 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.
- Bartels, M., Hudziak, J. J., Boomsma, D. I., Rietveld, M. J. H., Van Beijsterveldt, C. E. M., & van den Oord, E. J. C. G. (2003). A study of parent ratings of internalizing and externalizing problem behavior in 12-year-old twins. *Journal of the American Academy of Child and Adolescent Psychiatry*, *42*, 1351-1359.
- Behar, L., & Stringfield, S. (1974). A behavior rating scale for the preschool child. *Developmental Psychology*, *10*, 601-610.
- Blair, K. A., Denham, S. A., Kocanoff, A., & Whipple, B. (2004). Playing it cool: Temperament, emotion regulation and social behavior in preschoolers. *Journal of School Psychology*, *42*, 419-443.
- Boomsma, D. I. (1998). Twin registers in Europe: An overview. *Twin Research*, *1*, 34-51.
- Boomsma, D. I., Vink, J. V., Van Beijsterveldt, C. E. M., de Geus, E. J. C., Beem, A. L., Mulder, E. J. C. M., Derks, E. M., Riese, H., Willemsen, A. H. M., Bartels, M., van den Berg, M., Kupper, H. M., Polderman, J. C., Posthuma, D., Rietveld, M. J. H., Stubbe, J. H., Knol, L. I., Stroet, T., & van Baal, G. C. M. (2002). Netherlands Twin Register: A focus on longitudinal research. *Twin Research*, *5*, 401-406.
- Campbell, S. B. (1995). Behavior problems in pre-school children: A review of recent research. *Journal of Child Psychology and Psychiatry*, *36*, 113-149.

- Campbell, S. B., & Ewing, L. J. (1990). Follow-up of hard-to-manage preschoolers: Adjustment at age 9 and predictors of continuing symptoms. *Journal of Child Psychology and Psychiatry*, *31*, 871–889.
- Derks, E. M., Hudziak, J. J., Van Beijsterveldt, C. E. M., Dolan, C. V., & Boomsma, D. I. (2004). A study of genetic and environmental influences on maternal and paternal CBCL syndrome scores in a large sample of 3-year-old Dutch twins. *Behavior Genetics*, *34*, 571–583.
- Derks, E. M., Hudziak, J. J., Van Beijsterveldt, C. E. M., Dolan, C. V., & Boomsma, D. I. (2005). *Genetic analyses of maternal ratings and teacher ratings on attention problems in 7-year-old twins*. Manuscript submitted for publication.
- De Sonneville, L. M. J. (1999). Amsterdam Neuro-psychological Tasks: A computer-aided assessment program. In B. P. L. M. Den Brinker, P. J. Beek, A. N. Brand, S. J. Maarse, & L. J. M. Mulder (Eds.), *Cognitive ergonomics, clinical assessment and computer-assisted learning* (Vol. 6; pp. 187–203). Lisse, the Netherlands: Swets and Zeitlinger.
- Gjone, H., Stevenson, J., Sundet, J. M., & Eilertsen, D. E. (1996). Changes in heritability across increasing levels of behavior problems in young twins. *Behavior Genetics*, *26*, 419–426.
- Groot, A. S., De Sonneville, L. M. J., Stins, J. F., & Boomsma, D. I. (2004). Familial influences on sustained attention and inhibition in preschoolers. *Journal of Child Psychology and Psychiatry*, *45*, 306–314.
- Keenan, K., & Wakschlag, L. S. (2000). More than the terrible twos: The nature and severity of behavior problems in clinic-referred preschool children. *Journal of Abnormal Child Psychology*, *28*, 33–46.
- Keenan, K., & Wakschlag, L. S. (2002). Can a valid diagnosis of disruptive behavior disorder be made in preschool children? *American Journal of Psychiatry*, *59*, 351–358.
- Koot, H. M., van den Oord, E. J. C. G., Verhulst, F. C., & Boomsma, D. I. (1997). Behavioral and emotional problems in young pre-schoolers: Cross-cultural testing of the validity of the Child Behavior Checklist 2/3. *Journal of Abnormal Child Psychology*, *25*, 183–196.
- Lavigne, J. V., Arend, R., Rosenbaum, D., Binns, H. J., Christoffel, K. K., & Gibbons, R. D. (1998). Psychiatric disorders with onset in the preschool years. I: Stability of diagnoses. *Journal of the American Academy of Child and Adolescent Psychiatry*, *37*, 1246–1254.
- Neale, M. C. (1999). *Mx: Statistical modeling* (5th ed.) [Computer software]. Richmond, VA: Department of Psychiatry, Medical College of Virginia.
- Rietveld, M. J. H., Hudziak, J. J., Bartels, M., van Beijsterveldt, C. E. M., & Boomsma, D. I. (2003). Heritability of attention problems in children: Longitudinal results from a study of twins, age 3 to 12. *Journal of Child Psychology and Psychiatry*, *45*, 577–588.
- Roza, S. J., Hofstra, M. B., van der Ende, J., & Verhulst, F. C. (2003). Stable prediction of mood and anxiety disorders based on behavioral and emotional problems in childhood: A 14-year follow-up during childhood, adolescence, and young adulthood. *American Journal of Psychiatry*, *160*, 2116–2121.
- Schmitz, S., Fulker, D. W., & Mrazek, D. A. (1995). Problem behavior in early and middle childhood: An initial behavior genetic analysis. *Journal of Child Psychology and Psychiatry*, *36*, 1443–1458.
- 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.
- Spivack, G., & Spotts, J. (1966). *The Devereux Child Behavior (DCB) Rating Scale*. Devon, PA: The Devereux Foundation.
- SPSS Inc. (2002). *SPSS for Windows* (Release 11.5) [Computer software]. Chicago, IL: SPSS, Inc.
- Stins, J. F., De Sonneville, L. M. J., Groot, A. S., Polderman, J. C., Van Baal, G. C. M., & Boomsma, D. I. (2005). Heritability of selective attention and working memory in preschoolers. *Behavior Genetics*, *35*, 407–416.
- Towers, H., Spotts, E., Neiderhiser, J. M., Hetherington, E. M., Reiss, D., & Plomin, R. (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.
- van Beijsterveldt, C. E. M., Verhulst, F. C., Molenaar, P. C. M., & Boomsma, D. I. (2004). The genetic basis of problem behavior in 5-year-old Dutch twin pairs. *Behavior Genetics*, *34*, 229–242.
- van der Ende, J., & Verhulst, F. C. (2005). Informant, gender and age differences in ratings of adolescent problem behavior. *European Child and Adolescent Psychiatry*, *14*, 117–126.
- van Leeuwen, M., van den Berg, S. M., Van Beijsterveldt, C. E. M., & Boomsma, D. I. (2005). Effects of twin separation in primary school. *Twin Research and Human Genetics*, *8*, 384–391.
- van den Oord, E. J. C. G., Koot, H. M., & Boomsma, D. I. (1996). A genetic study of maternal and paternal ratings of problem behaviors in 3-year-old twins. *Journal of Abnormal Psychology*, *105*, 349–357.
- van der Valk, J. C., van den Oord, E. J. C. G., Verhulst, F. C., & Boomsma, D. I. (2001). Using parental ratings to study the etiology of 3-year-old twins' problem behaviors: Different views or rater bias? *Journal of Child Psychology and Psychiatry*, *42*, 921–931.
- Verhulst, F. C., van der Ende, J., & Koot, H. M. (1997). *Manual for the Teacher's Report Form* (Dutch



translation). Rotterdam, the Netherlands: Department of Child and Adolescent Psychiatry.

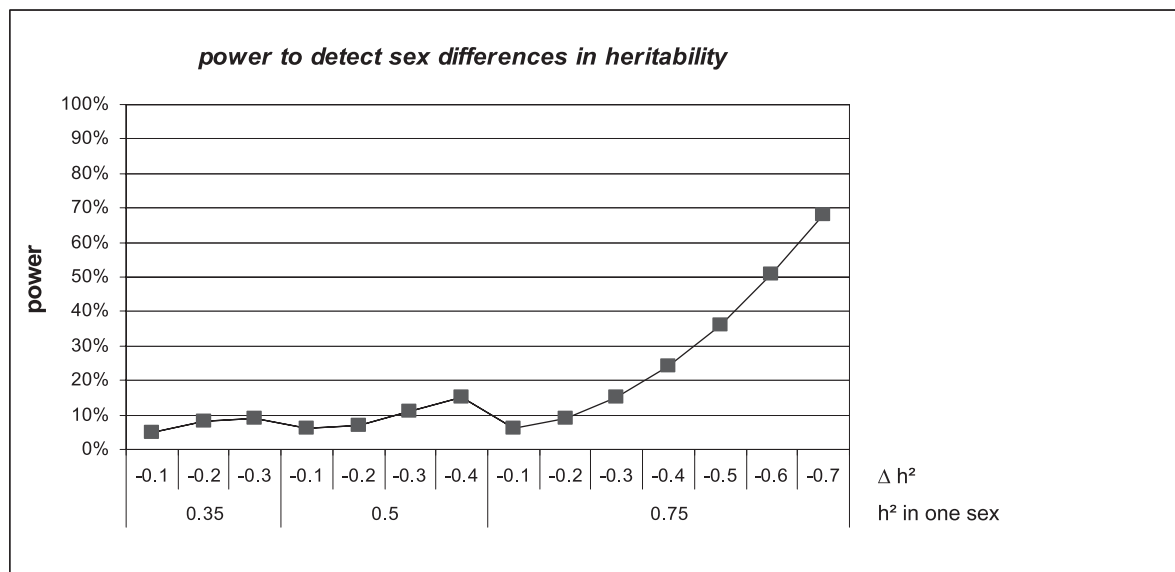
Verhulst, F. C., & van der Ende, J. (1992). Six-year stability of parent-reported problem behavior in an epidemiological sample. *Journal of Abnormal Child Psychology*, 20, 595–610.

Vierikko, E., Pulkkinen, L., Kaprio, J., & Rose, R. J. (2004). Genetic and environmental influences on the

relationship between aggression and hyperactivity-impulsivity as rated by teachers and parents. *Twin Research*, 7, 261–274.

Zahn-Waxler, C., Schmitz, S., Fulker, D. W., Robinson, J., & Emde, R. (1996). Behavior problems in 5-year-old monozygotic and dizygotic twins: Genetic and environmental influences, patterns of regulation, and internalization of control. *Development and Psychopathology*, 8, 103–122.

### Appendix



Appendix shows the power (y-axis) to detect differences in heritability (x-axis) between boys and girls given a sample size of 215 twin pairs. For a fixed heritability ( $h^2$ ) of .35, .50 and .75 in one sex, the power to detect a difference of heritability ( $\Delta h^2$ ) between boys and girls of respectively, .10, .20 and .30; .10, .20, .30 and .40; and .10, .20, .30, .40, .50, .60 and .70 is shown.

