A Study of Parent Ratings of Internalizing and Externalizing Problem Behavior in 12-Year-Old Twins

MEIKE BARTELS, PH.D., JAMES J. HUDZIAK, M.D., DORRET I. BOOMSMA, PH.D., MARJOLEIN J.H. RIETVELD, PH.D., TOOS C.E.M. VAN BEIJSTERVELDT, PH.D., AND EDWIN J.C.G. VAN DEN OORD, PH.D.

ABSTRACT

Objective: Studies on 3-, 7-, and 10-year-old twins' internalizing and externalizing problems have emphasized the importance of understanding sources of agreement and disagreement between maternal and paternal ratings. A psychometric model that assumes that each parent assesses rater-specific aspects of the child's behavior provided the best explanation for parental disagreement. This study investigates two models that have been used to explain the agreement and disagreement between mothers and fathers in the ratings of their children. **Method:** Child Behavior Checklists filled in by mothers and fathers were collected for a sample of 1,481 twelve-year-old twin pairs. Genetic and environmental influences on internalizing and externalizing problems were estimated using models that corrected for rater bias, rater-specific effects, and unreliability. **Results:** The psychometric model fitted the data significantly better than a rater bias model. Significant influences of genetic, shared, and nonshared environmental factors were found for internalizing and externalizing behavior. Parent-specific views, rater bias, and unreliability were significant. **Conclusions:** The best-fitting model implies that disagreement between parents is due to the fact that mothers and fathers provide information from their own perspective. This information should be seen as important and adding to the diagnostic formulation rather than as a point of disagreement. The finding that internalizing and externalizing problems are influenced by genetic and environmental factors fosters the understanding that it is the interaction of nature and nurture that puts children at risk for common behavioral disorders. *J. Am. Acad. Child Adolesc. Psychiatry,* 2003, 42(11):1351–1359.
Key Words: parental rating, rater bias, problem behavior, twins.

There is little debate about the prevalence, morbidity, and mortality associated with emotional and behavioral disorders. Multiple studies have demonstrated these disorders to be highly prevalent. Two studies in representative samples of Dutch children reported that 7% to 8% of preschool-age children show problem behavior (Verhulst et al., 1985, 1997). These studies indicate that internalizing problems (anxious/depressed behavior, withdrawn behavior) are more prevalent in girls and externalizing problems (aggressive behavior, rule-breaking) are more prevalent in boys. Although experts agree that these behaviors are common, there remains considerable debate about how best to conceptualize these disorders (categorical vs. quantitative), who best to ask about these problems (mothers, fathers, or the children themselves), and to what degree these problems are due to genetic and/or environmental factors (the nature vs. nurture debate). Behavioral genetic approaches allow us to answer many of these questions. Prior work indicates that these behaviors are influenced by genetic and environmental factors. Furthermore, behaviors such as internalizing and externalizing behavior generally do not fall in distinct categories (present or absent) but are better conceptualized as quantitative variations of behaviors that most children display to

Accepted June 13, 2003.

Drs. Bartels, Rietveld, Van Beijsterveldt, and Boomsma are with the Department of Biological Psychology, Vrije Universiteit, Amsterdam, the Netherlands. Dr. Hudziak is with the Departments of Psychiatry and Medicine (Division of Human Genetics), Center for Children, Youth and Families, University of Vermont, College of Medicine Burlington, Burlington; Dr. Van den Oord is with the Virginia Institute of Psychiatric and Behavioral Genetics, Medical College of Virginia/Virginia Commonwealth University, Richmond.

This research was financially supported by The Netherlands Organization for Scientific Research (575-25-012). The Netherlands Organization for Scientific Research (NWO: R 56-467) and the Stichting Simonsfonds (SF053-iz) provided travel grants to facilitate collaboration with Dr. Van den Oord at the Department of Psychiatry, Medical College of Virginia/Virginia Commonwealth University, Richmond.

Correspondence to Dr. Bartels, Department of Biological Psychology, Vrije Universiteit, Room 1F 66, van der Boechorststraat 1, 1081 BT, Amsterdam, the Netherlands; e-mail: m.bartels@psy.vu.nl.

DOI: 10.1097/01.CHI.0000085755.71002.5d

some degree. Debate remains on who is the best informant.

One approach to quantify children's problem behavior is by asking the parents to score behavioral problems on questionnaires. A meta-analysis by Achenbach and colleagues (1987) showed a mean correlation of 0.60 between maternal and paternal ratings of the same child. This underscores that parents can assess their child's behavior, for if parental ratings would reflect nothing but error, the correlation between their ratings would probably be low. The less-than-perfect correlation implies parental disagreement. From a clinical point of view, it remains a struggle to determine what to do with that disagreement. Is it best to assume that there is one best informant, that one parent is more reliable than the other, or that each parent presents a rater-specific viewpoint on his or her child, thus providing unique and valuable information to be used in assessment?

To study parental disagreement, Hewitt and colleagues (1992) proposed rater bias and psychometric models that combine data of mother and father ratings. The rater bias model assumes that parents assess exactly the same behaviors in the child (common behavioral view) and that they share a common understanding of the behavioral descriptions. Disagreement between the raters is regarded as error, resulting from rater bias and/or unreliability. Sources of rater bias are stereotyping, employing different normative standards, or having certain response styles (i.e., judging problem behaviors more or less severely). Unreliability can become an important source of disagreement when raters cannot give an accurate description about relevant behaviors. The psychometric model assumes that, besides a common view, each parent assesses a rater-specific aspect of the child's behavior. This will occur when the parent also observes the child in situations where he or she is exposed to distinct samples of the behavior. For instance, the parent who usually brings the child to school may be more familiar with the child's behavior outside the home. Moreover, each parent may interact differently with the child (Achenbach et al., 1987). These unique interactions between a parent and a child may allow parents to provide additional information about the child's behavior. Therefore, disagreement in this model does not merely arise due to unreliability and/or rater bias, but also because each parent contributes, from his or her own perspective, different but valid information on the child's functioning.

Several quantitative genetic studies have used the Child Behavior Checklist (CBCL) (Achenbach, 1991) to examine the etiology of children's problem behaviors (Edelbrock et al., 1995; Gjone and Stevenson, 1997; Hudziak et al., 2000; Leve et al., 1998; Schmitz et al., 1995; Silberg et al., 1994; Van den Oord et al., 1995; van der Valk et al., 1998a,b; Zahn-Waxler et al., 1996). However, only a few studies have employed models that incorporated rater differences. Rowe and Kandel (1997) collected the CBCL completed by mothers and fathers for their oldest two offspring (aged 9–17) in 76 families. They did not fit either psychometric or rater bias models. Still, their results, based on an "individual view-shared view" model, showed that the parental ratings contained a substantial shared behavioral view. Simonoff and colleagues (1995), in a study of 282 twin pairs aged 8 to 16, also found evidence in favor of a shared behavioral view for antisocial behaviors. Hewitt and colleagues (1992) applied both the rater bias and psychometric models on parental ratings of the internalizing scale of the CBCL for 983 twin pairs. They found that both for their prepubertal cohort (8–11 years) and for their pubertal cohort (12–16 years), the psychometric model fitted the data better than the rater bias model. Rater-specific genetic influences were found, implying that the rater differences reflected the existence of real rater-specific behavioral views and not just error and bias. Further insight into issues of rater bias is presented by van der Valk and colleagues (2001, in press). Rater bias models and psychometric models were tested on large groups of 3- and 7-year-old Dutch twins. As in the previous studies, the psychometric model fitted the data significantly better at both ages. In a twin sample of 10-year-old twins, the same results are found (Bartels et al., unpublished, 2003).

In the present study rater bias and psychometric models were fitted to the internalizing and externalizing scale of 1,481 Dutch 12-year-old twin pairs to examine the processes underlying parental disagreement and to estimate genetic and environmental influences on both behaviors.

METHOD

Subjects

All participants were registered by the Netherlands Twin Registry (Boomsma et al., 2002). Data from twins from the birth cohorts 1986 to 1990 were used. Questionnaires were mailed to families within 3 months of the twins' 12th birthday. After 3 months reminders were sent, and after 4 months nonresponders were contacted by phone. This procedure resulted in an 80% participation rate. From the remaining sample 80 twin pairs were excluded because either one or both of the children had a disease or handicap that interfered severely with daily functioning. The final sample for analysis consisted of 1,481 mother ratings and 1,156 father ratings. The data could be divided into twin pairs for which both mother and father had replied (225 MZM, 180 DZM, 240 MZF, 187 DZF, and 324 DOS) and twin pairs for which only mothers had replied (58 MZM, 51 DZM, 75 MZF, 41 DZF, and 100 DOS). Zygosity was determined for 472 same-sex twin pairs by DNA analysis or blood group polymorphisms. For all other same-sex twin pairs zygosity was determined by discriminant analysis, using questionnaire items. The questionnaire items allow accurate determination of zygosity of nearly 95% (Rietveld et al., 2000).

Measures

The Child Behavior Checklist (CBCL 4–18) (Achenbach, 1991) was developed for parents to score the behavioral problems of their 4- to 18-year-old children. It consists of 120 problem items that are scored on a 3-point scale based on the occurrence of the behavior during the preceding 6 months: 0 if the problem item was not true, 1 if the item was somewhat or sometimes true, and 2 if it was very true or often true. The syndrome scales were composed according to the 1991 profile (Achenbach, 1991). Dutch syndrome scales and comparability with the syndrome scales, as developed by Achenbach, are reported in the Dutch manual (Verhulst et al., 1996). The internalizing scale consists of the anxious/depressed, somatic complaints, and withdrawn subscales. The externalizing scale consists of the aggressive and rule-breaking behavior subscales. The data were square root-transformed to approximate normal distributions that are required for maximum likelihood estimation. After transformation, all skewness and kurtosis indices were between −1.0 and 1.0 (Muthén and Kaplan, 1985).

Statistical Analysis

Descriptive statistics and prevalence of internalizing and externalizing behaviors were calculated using SPSS/Windows 10. Pearson correlations were used to calculate twin correlation within and across raters. Twin correlations for the five zygosity groups give a first impression of the genetic and environmental influences on both behaviors. The cross-twin/cross-rater correlations give information on the underlying causes of parental disagreement.

Classic Twin Design and Multiple Rater Models

Genetic model fitting of twin data allows for separation of the observed phenotypic variance into genetic and environmental components. Additive genetic variance (A) is the variance that results from the additive effects of alleles at each contributing genetic locus. Shared environmental variance (C) is the variance that results from environmental events common to both members of a twin pair (style of parenting, socioeconomic level, or religion). Nonshared environmental variance (E) is the variance that results from environmental effects that are not shared by members of a twin pair (illness, relationships with peers, or measurement errors). Estimates of the nonshared environmental effects also include measurement error.

The different degree of genetic relatedness between monozygotic (MZ) and dizygotic (DZ) twin pairs can be used to estimate the contribution of genetic and environmental factors to the trait under study. Similarities for MZ twins are assumed to be due to additive genetic influences plus environmental influences that are shared by both members of a twin pair. Experiences that make MZ twins different from one another are nonshared environmental influences. Because DZ twins share 50% of their genetic material on average, like other siblings, genetic factors contribute only half to their resemblance. As for MZ twins, the shared environment contributes fully. Model fitting to twin data is based on the comparison of the variance–covariance matrices in MZ and DZ twins. (For a summary of the twin method, the various assumptions, and the plausibility of these assumptions, see, for example, Neale and Cardon, 1992.)

If data of multiple raters are available it is possible, besides estimating genetic and environmental influences, to gain insight into the underlying causes of disagreement. Two different models are available to study this disagreement.

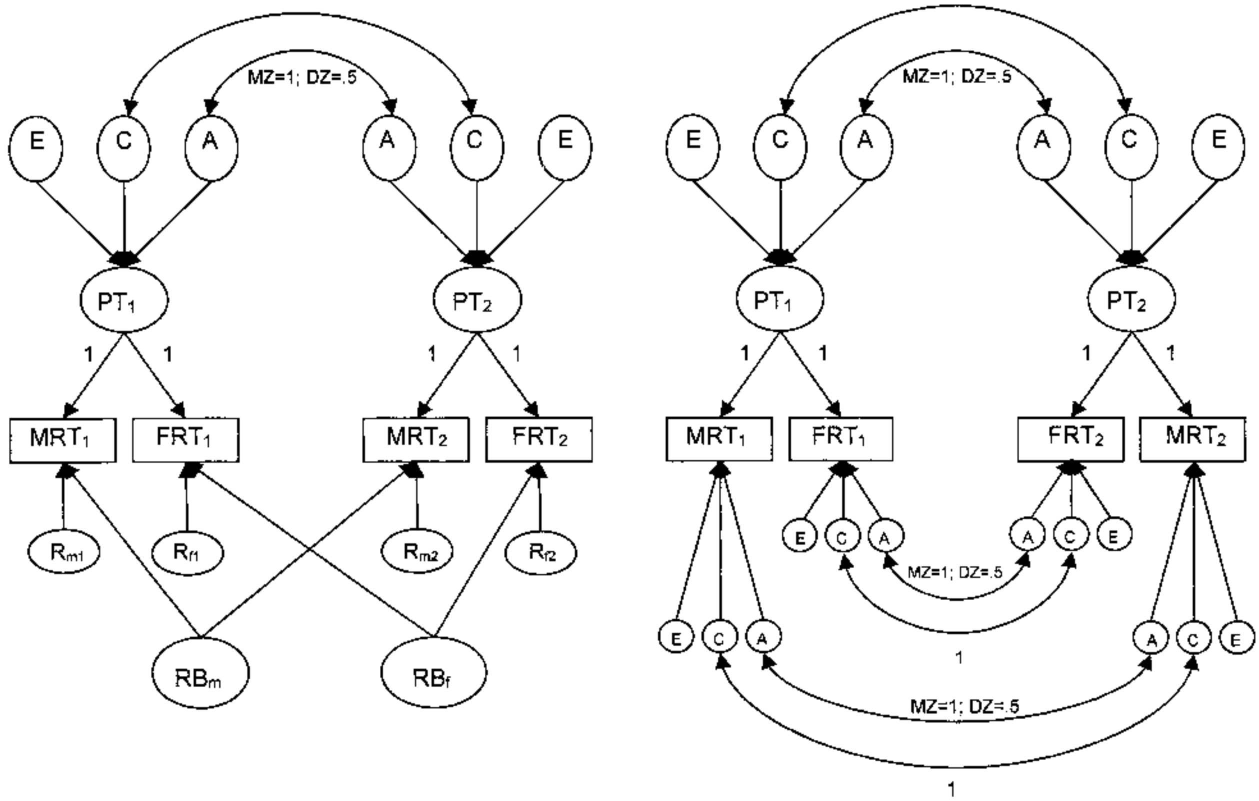
In the rater bias model (Hewitt et al., 1992) (Fig. 1, left) the variance in behavior similarly assessed by both parents (PT_1 and PT_2) can be decomposed into additive genetic variance (A), shared environmental variance (C), and variance due to nonshared environmental influences (E). In addition to these three common factors, rater-specific factors representing parental disagreement and measurement error are modeled: a maternal rater bias factor, a paternal rater bias factor, and residual (unreliability) factors affecting each rating. The influence of the common factors is assumed to be independent of the maternal and paternal rater bias and unreliability factors.

The psychometric model (Hewitt et al., 1992) (see Fig. 1, right) also estimates the influence of a genetic (A), a shared environmental (C), and a nonshared environmental factor (E) on the behavior similarly assessed by both parents (PT_1 and PT_2). In addition, six rater-specific factors are estimated for the disagreement in ratings of mother and fathers. Disagreement between parents in this model can be caused by rater-specific behavioral views, leading to different but valid information of each rater. The psychometric model tests this by examining whether there are significant child genetic effects on the rater-specific behavioral views. If the behaviors uniquely rated by the parents are shown to be influenced by genetic factors of the child, the parent must have been assessing "real" rater-specific behavioral views, for error and/or unreliability cannot cause the systematic effects necessary for the model to estimate genetic influences. Disagreement can also be caused by rater bias, which will confound the rater-specific shared environmental effects and thus influences both children of a twin pair, or by unreliability or measurement error, which will confound the rater-specific nonshared environmental effects, which will not systematically influence both children of a twin pair but only one.

Model Fitting

The structural equation modeling package Mx (Neale et al., 1999) was used to analyze the data through a simultaneous analysis of the 4 × 4 variance–covariance matrices in the five zygosity by sex twin groups (MZM, DZM, MZF, DZF, DOS) where both mother and father ratings were available, and the 2 × 2 variance-covariance matrices in the five zygosity by sex twin groups with only mother ratings. A good model is indicated by a low nonsignificant χ^2 test statistic ($p > .05$). Apart from the χ^2 test statistic, Akaike's Information Criterion (AIC = $\chi^2 - 2 \times$ degrees of freedom) was computed. The lower the AIC, the better the fit of the model to the observed data.

Fitting the rater bias and psychometric model to the data showed which model best described the processes involved in disagreement between the parental ratings. The strength of the common and rater-specific genetic and environmental influences was allowed to differ for boys and girls. This model was further examined for possible simplifications. It was tested whether the common and/or rater-specific factors could be removed from the model, whether



The Rater Bias Model

Rater bias model for ratings of a pair of twins (oldest and youngest twin) by their parents. Mother's and father's observed ratings (in squares: MRT_1, FRT_1, MRT_2, FRT_2) are linear functions of the latent phenotypes of the twins (PT_1 and PT_2), mother's and father's bias (RB_m, RB_f), and residual error (R_{m1}, R_{f1}, R_{m2}, R_{f2}). Latent phenotypes of the twins (PT_1 and PT_2) are influenced by A (additive genetic factors), C (shared environmental factors), and E (nonshared environmental factors).

The Psychometric Model

Psychometric model for ratings of a pair of twins (oldest and youngest twin) by their parents. Mother's and father's observed ratings (in squares: MRT_1, FRT_1, MRT_2, FRT_2) are linear functions of the latent phenotypes of the twins (PT_1 and PT_2) and rater specific variance. Latent phenotypes of the twins are influenced by common (i.e. across both parents) A (common genetic factor), C (common shared environmental factor), and E (common nonshared environmental factor). Rater specific variance is made up of rater specific (i.e. unique to each parent) A (rater specific genetic factor), C (rater specific shared environmental factor and/or rater bias), E (rater specific nonshared environmental factor and/or measurement error/ unreliability).

Fig. 1 Models for multiple rater data.

estimates for boys and girls could be constrained to be the same, and if the rater-specific factors for mothers and fathers could be constrained to be equal. The only factor that was never dropped from the model was the rater-specific nonshared environmental factor, because apart from the influences of idiosyncratic experiences, measurement errors are also estimated in this factor.

RESULTS

Description of the Data

The untransformed mean problem scores, standard deviations, and prevalence of internalizing and externalizing behavior (T ≥ 65) in the 12-year-old Dutch twins are presented in Table 1. The ratings given to the twins were similar to the ratings given to a Dutch community sample (Verhulst et al., 1996). Within the twin group, one-way ANOVA showed no significant mean differences between MZ and DZ twin pairs. Both mothers and fathers gave significantly higher ratings to boys for the externalizing scale (MZ: mothers: $F_{1,1196} = 34.421$, $p = .000$; fathers: $F_{1,924} = 26.960$, $p = .000$; DZ: mothers: $F_{1,925} = 22.505$, $p = .000$; fathers: $F_{1,738} = 17.853$, $p = .000$). A paired t test showed

TABLE 1
Mean Problem Scores, Standard Deviations, and Prevalence of Behaviors

	Internalizing					Externalizing				
				Prevalence					Prevalence	
	Mother	Father	*n* (M/F)	M	F	Mother	Father	*n* (M/F)	M	F
Male										
MZM	4.11 (4.63)	3.20 (4.40)	561/447	5.6%	4.7%	7.25 (6.91)	6.07 (6.84)	569/447	11.2%	12.5%
DZM	4.09 (4.67)	3.49 (4.50)	453/362			6.71 (6.52)	5.94 (6.52)	468/363		
DOS	3.79 (4.68)	2.75 (3.49)	410/317			6.70 (6.76)	5.53 (5.62)	416/318		
Female										
MZF	4.79 (4.82)	3.08 (3.26)	617/469	9.0%	9.2%	5.18 (5.28)	4.10 (4.58)	629/479	4.3%	5.2%
DZF	4.63 (4.55)	4.05 (4.65)	448/377			4.89 (5.04)	4.19 (4.62)	459/377		
DOS	4.31 (4.43)	2.98 (3.35)	427/334			4.79 (5.30)	3.62 (4.09)	440/333		

Note: MZM/DZM = monozygotic/dizygotic males; MZF/DZF = monozygotic/dizygotic females; DOS = dizygotic opposite sex; *n* children M/F = number of children for mothers (M) and fathers (F).

that the ratings for the internalizing and externalizing scales given by mothers were significantly higher than ratings given by fathers (internalizing: boys: $t = 7.566$, $df = 1{,}093$, $p = .000$; girls: $t = 9.744$, $df = 1{,}139$, $p = .000$; externalizing: boys: $t = 6.729$, $df = 1{,}113$, $p = .000$; girls: $t = 7.339$, $df = 1{,}174$, $p = .000$). The homogeneity of the variance was tested with Mx (Neale et al., 1999). No differences could be found in the variances of MZM, DZM, MZF, DZF, and DOS, for both scales.

Correlations

Table 2 shows twin correlations for both raters separately. For mother and father ratings, the MZ twin correlations were higher than the DZ twin correlation, indicating influences of genetic factors. The MZ correlations, though, were less than twice as high as the DZ correlations, which is an indication of the influence of shared environmental factors or rater bias. No differences between same-sex DZ twins and opposite-sex DZ twins were observed, so no heterogeneity is expected. The cross-twin/cross-parent correlations in Table 2 indicate, by higher cross-twin/cross-parent correlations for MZ twins compared to DZ twins, that rater-specific parental view and rater bias are possible underlying causes for parental disagreement. On average, the interparent correlations for the internalizing scale were 0.63; for the externalizing scale they were 0.72. This resembled the interparent correlations obtained in the Dutch norm group (Verhulst et al., 1996).

Model Fitting Results

As indicated by the lower χ^2 test statistic and the lower AIC in Table 3, the psychometric model fitted

TABLE 2
Twin Correlations for Raters

	Internalizing						Externalizing					
	Twins		Different Rater				Same Rater		Different Rater			
	Twins		Twins		Interparent		Twins		Twins		Interparent	
	M/M	F/F	M/F	F/M	O	Y	M/M	F/F	M/F	F/M	O	Y
MZM	.73	.74	.48	.47	.66	.60	.87	.90	.71	.68	.78	.75
DZM	.46	.47	.31	.36	.67	.68	.54	.61	.39	.39	.71	.71
MZF	.73	.69	.41	.43	.55	.61	.84	.83	.66	.62	.70	.72
DZF	.54	.61	.38	.40	.65	.64	.56	.58	.44	.35	.74	.69
DOS	.59	.49	.41	.32	.63	.61	.61	.57	.42	.48	.74	.69

Note: MZM/DZM = monozygotic/dizygotic males; MZF/DZF = monozygotic/dizygotic females; DOS = dizygotic opposite sex twins; Same Rater Twins = correlation between the oldest and the youngest twin, rated by M/M (mothers) or F/F (fathers); Different Raters Twins = cross-correlation: either oldest twin rated by mothers and youngest by fathers (M/F) or the other way around (F/M); Different Raters Interparent: O = correlation between mother and father ratings for the oldest child; Y = idem for the youngest child.

TABLE 3
Psychometric Versus Rater Bias Model

	Internalizing						
	χ^2	*df*	*p*	AIC	$\Delta\chi^2$	Δdf	*p*
Overall model							
Psychometric model	51.557	47	.30	–42.44			
Rater bias model	78.467	49	.01	–19.53			
Simplification overall model							
Factor estimates							
No common genetic effects	73.595	49	.01	–24.60	22.038	2	.000
No rater-specific genetic effects	86.08	51	.00	–15.92	34.523	4	.000
No common shared env.	80.224	49	.00	–17.78	28.667	2	.000
No rater-specific shared env.	75.905	51	.01	–26.10	24.348	4	.000
No common nonshared env.	257.69	49	.11	159.69	206.133	2	.000
Sex differences							
No sex dif. common effects	56.771	50	.24	–43.23	**5.214**	**3**	**.157**
No sex dif. rater-specific effects	54.122	53	.43	–51.88	**2.565**	**6**	**.861**
No sex dif. common + rater-specific	60.47	56	.32	–51.53	**8.931**	**9**	**.445**
Rater differences							
Rater-specific effects: M-F identical	55.654	53	.38	–50.35	**4.097**	**6**	**.665**
Simplified model	**63.981**	**59**	**.31**	**–54.02**			

	Externalizing						
	χ^2	*df*	*p*	AIC	$\Delta\chi^2$	Δdf	*p*
Overall model							
Psychometric model	38.532	47	.81	–55.47			
Rater bias model	74.875	49	.01	.23–13			
Simplification overall model							
Factor estimates							
No common genetic effects	199.621	49	.99	101.62	161.089	2	.000
No rater-specific genetic effects	80.345	51	.00	–21.66	41.813	4	.000
No common shared env.	59.763	49	.14	–38.24	21.231	2	.000
No rater-specific shared env.	91.164	51	.00	–1.084	52.632	4	.000
No common nonshared env.	186.068	49	.00	88.07	147.536	2	.000
Sex differences							
No sex dif. common effects	57.103	50	.23	–42.90	18.571	3	.000
No sex dif. rater-specific effects	48.232	53	.66	–57.77	**9.700**	**6**	**.138**
No sex dif. common + rater-specific	71.068	56	.09	–40.93	32.536	9	.000
Rater differences							
Rater-specific effects: M-F identical	50.524	53	.57	–55.48	**11.992**	**6**	**.062**
Simplified model	**55.070**	**56**	**.051**	**–56.93**			

Note: Model reductions that did not give a significant change in the fit of the model are in boldface type. AIC = Akaike's Information Criterion.

the data better than the rater bias model for both scales. This finding signifies that besides a common view, each parent assesses a rater-specific aspect of the child's behavior. For the sake of comparison we performed a Cholesky decomposition (also called a biometric model). This model can be viewed as a psychologically less informative rotation of the psychometric model (Hewitt et al., 1992). Neither for the internalizing scale nor for the externalizing scale did this saturated model fit the data better than the psychometric model.

The psychometric model was further examined for possible simplifications. None of the common and rater-specific parameters could be dropped from the model. Between boys and girls, the estimates of the common and the rater-specific factors could be constrained to be equal for the internalizing scale. For the externalizing scale, only the rater-specific effects could be set equal for boys and girls. Mother and father ratings could be constrained to be equal for both scales. The fit of the best model is given in Table 3.

The percentages of variance explained by common and rater-specific genetic, shared, and nonshared environmental factors are given in Table 4. A major part of the variance was explained by common factors. For the externalizing scale, the largest part of the variance was explained by the common genetic factor, and sex differences in the strength of this common genetic influence are found. For the internalizing scale, both common genetic and common shared environmental factors explained significant parts of the variance. Common nonshared environmental influences were small but significant. The rater-specific factors explained a relatively smaller part of the variance, ranging from 6% to 15%.

DISCUSSION

In a sample of 12-year-old twin pairs, we studied the genetic and environmental contribution to internalizing and externalizing behavior by considering the role of informant on these estimates. By studying the underlying agreement and disagreement between maternal and paternal ratings, we were able to test whether parental ratings of internalizing and externalizing behavior in 12-year-old twins conformed to rating patterns seen in younger children. In addition, we tested if there were differences in the way mothers versus fathers rated internalizing and externalizing behavior, and if parents rated their daughters differently than their sons.

Parent Rating Styles

For externalizing behavior, both mothers and fathers rated their daughters as having fewer problems than their sons. Mothers rated their daughters as having higher mean scores on internalizing problems than their sons, whereas fathers reported similar rates of internalizing behavior for daughters and sons. The ratings for the internalizing and externalizing scales given by mothers were significantly higher than ratings given by fathers for both boys and girls. These reporting patterns conform to Achenbach and colleagues' finding (1987) and our findings on large samples of younger children (Bartels et al., unpublished, 2003; van der Valk et al., 2001, in press).

TABLE 4
Variation Explained by Genetic, Shared Environmental, and Nonshared Environmental Factors

	Internalizing (%)	Externalizing (%)	
		Male	Female
Genetic factor			
Common	22	48	40
Rater-specific	15	9	11
Shared environmental factor			
Common	25	20	20
Rater-specific	11	10	13
Nonshared environmental factor			
Common	16	7	8
Rater-specific	11	6	8

Agreement and Disagreement

The psychometric model fit the data best for both scales. These results indicate that differences between mother and father reports are not due to measurement errors, but that parents assess different aspects of the child's behavior. As suggested by Achenbach and colleagues (1987), unique interactions might allow each parent to provide additional information about the child's behavior. Because no single rater may be able to provide a complete picture of the child's behavior, the implication is that it is important to collect data from multiple informants. These findings are of specific interest to the study of developmental psychopathology, because as children enter, endure, and exit puberty, it will be important for parents to realize that they see and respond to different aspects of behavior.

More complete information on the child's behavior could also be obtained by adding data on teacher ratings and self-report. In future studies we will extend these findings by including reports from teachers (Teacher Report Form) and the children themselves (Youth Self-Report Form). By including data from multiple informants, we will be able to estimate how mothers', fathers', and teachers' ratings of children compare to self-reports. By analyzing these reports on a large sample of twin data, we hope to come up with phenotyping strategies that will improve our ability to select subjects for genotyping and treatment studies.

Variance Decomposition

Genetic factors were most important for the reliable trait variance (behavior similarly assessed by both parents) of the externalizing scale, explaining over 50% of the variance in boys and girls (boys, 63%; girls, 60%). Heritabilities of around 50% were found for 3- and 7-year-old twin pairs (van der Valk et al., 2001, in press). Previous studies also report that genetic influences explained about half of the variance of the externalizing scale (Edelbrock et al., 1995; Gjone and Stevenson, 1997; Zahn-Waxler et al., 1996). At age 10 heritabilities explain around 65% and 50% of the variance in boys and girls, respectively (Bartels et al., unpublished, 2003).

Shared environmental influences explained 27% of the reliable trait variance of the externalizing scale in boys and 29% in girls. This again was in accordance with the shared environmental influences observed for the 3- and 7-year-old twin pairs (van der Valk et al., 2001, in press) and the results found in the studies of Edelbrock and colleagues (1995), Gjone and Stevenson (1997), and Zahn-Waxler and colleagues (1996). Apart from quantitative genetic studies, various epidemiologic studies have also demonstrated the importance of shared environmental factors in the etiology of externalizing behaviors. Family discord and disruption, lack of affection, and poor supervision all predispose to conduct problems and antisocial behavior (Rutter, 1985).

Genetic influences for the internalizing scale accounted for 35% of the reliable trait variance at age 12. In contrast, for younger twin pairs the internalizing scale was predominantly influenced by the child's genotype (van der Valk et al., 2001; Zahn-Waxler et al., 1996). Shared environmental influences showed a complementary increase in influences over time, having almost no influence on the internalizing scale of 3-year-old twin pairs (van der Valk et al., 2001) and explaining around 35% of the variance of the internalizing scale for 7-year-old twin pairs (van der Valk et al., in press), 10-year-old twin pairs (Bartels et al., unpublished, 2003), and 12-year-old twin pairs. A differential genetic influence for internalizing problems of older versus younger children was also found in other studies (Gjone and Stevenson,1997; O'Connor et al., 1998).

The most plausible explanation for the differences in genetic influences on internalizing behavior is provided by the developmental perspective, which allows that different genetic and environmental influences can contribute to the expression of psychopathology at different ages. At early ages, genetic influences predominate; as the child ages, and his or her environment allows for different types of behavior, shared environmental influences become more important. Evidence of these developmental findings is rooted in cognitive psychology, developmental biology, and parenting. As brain maturation proceeds, children accrue added skills in communication, attention, and motor behavior. These behaviors are influenced in the child's cognitive level. As Piaget (1954) argued, children pass through successive developmental stages during which different cognitive styles are evidenced. Furthermore, as the child's neurodevelopment proceeds, in concert with his or her cognitive maturation, the parenting style will play an increasingly important role. It may be important to realize that shared environment is not necessarily confined to the home environment: there are indications that environmental effects are shared not merely by siblings but also by cousins (Van den Oord and Rowe, 1998, 1999). This suggests that shared environment reflects the wider community in which families are embedded as well (Bronfenbrenner, 1979; Parke and Kellam, 1994).

Fitting models to the observed data that explicitly incorporate rater bias and unreliability ensured that these effects could not distort estimates of the shared and nonshared environmental factors. Pure nonshared environmental influences (undistorted by error or unreliability) are found for the 3-, 7-, 10-, and 12-year-old twin pairs (Bartels et al., unpublished, 2003; van der Valk et al., 2001, in press). Thus idiosyncratic experiences are important to explain preschool and school-age children's problem behaviors. Measurement error and unreliability were estimated in the rater-specific nonshared environmental factor. Neither for the externalizing scale nor for the internalizing scale did this factor account for more than 11% of the variance. Rater bias was included in the estimate of the rater-specific shared environmental factor, accounting for at most 13% of the variance for the internalizing and externalizing scale.

Limitations

Our assessment instrument, the CBCL, does not measure *DSM* or *ICD* psychiatric diagnoses. Although the relations between internalizing and externalizing behavior and *DSM* disorders are well known, the purpose of this report was to explore the genetic and environmental influences on internalizing and externalizing behavior and not on *DSM* diagnosis. We are collecting *DSM* data on a selected group of these twins to comment on diagnoses as well.

Estimates found using quantitative genetic studies do not pertain to the individual but involve average differences between individuals in the population. For other populations, or for specific individuals, other estimates may apply. This study used a nonclinical sample of Dutch twin pairs showing problem behaviors in the normal range. Whether similar results will be obtained in clinical populations showing more extreme problem behaviors remains to be explored.

Clinical Implications

Our results confirm and extend earlier findings on the study of internalizing and externalizing behavior in twins. Based on both mother and father reports, estimates of genetic influences on externalizing behavior account for approximately 40% to 50% of the total variance. These results are consistent with the genetic influences reported by parents for earlier ages. Internalizing behavior in 12-year olds was influenced to a much smaller degree by additive genetic influences than those reported at earlier ages. These differences are striking and must be taken into account when conceptualizing diagnostic, therapeutic, and prognostic approaches. With externalizing behavior, genetic contributions are best described as robust and stable across development. In fact, Hudziak and colleagues (in press) report that although raw scores of aggressive behavior (a component of externalizing behavior) decrease for boys and girls from age 3 to 12, the relative genetic contribution to this syndrome remains constant. On the other hand, our data support the contention that genetic contributions to internalizing behavior decrease as a child ages. As a result, a one-size-fits-all approach to developmental psychopathology does not apply.

REFERENCES

Achenbach TM (1991), *Manual for the Child Behavior Checklist/4-18* and *1991 Profile*. Burlington: University of Vermont Department of Psychiatry

Achenbach TM, McConaughy SH, Howell CT (1987), Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. *Psychol Bull* 101:213–232

Boomsma DI, Vink JM, van Beijsterveldt CEM et al. (2002), Netherlands Twin Register. *Twin Res* 5:401–406

Bronfenbrenner U (1979), *The Ecology of Human Development*. Cambridge, MA: Harvard University Press

Edelbrock C, Rende R, Plomin R, Thompson L (1995), A twin study of competence and problem behavior in childhood and early adolescence. *J Child Psychol Psychiatry* 36:775–785

Gjone H, Stevenson J (1997), The association between internalizing and externalizing behavior in childhood and early adolescence: genetic and environmental common influences? *J Abnorm Child Psychol* 25:277–286

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, 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

Hudziak JJ, van Beijsterveldt CEM, Bartels M, Derks E, Boomsma DI (in press), Individual differences in aggression in young children: cross-sectional analyses in Dutch twins. *Behav Genet*

Leve LD, Winebarger AA, Fagot BI, Reid JB, Goldsmith HH (1998), Environmental and genetic variance in children's observed and reported maladaptive behavior. *Child Dev* 69:1286–1298

Muthén B, Kaplan D (1985), A comparison of some methodologies for the factor analysis of nonnormal Likert variables. *Br J Math Stat Psychol* 38:171–189

Neale MC, Boker SM, Xie G, Maes HH (1999), *Mx: Statistical Modeling*, 5th ed. VCU Department of Psychiatry, Box 900126, Richmond, VA 23298

Neale MC, Cardon LR (1992), *Methodology for Genetic Studies of Twins and Families.* Dordrecht: Kluwer Academic

O'Connor TG, Neiderhiser JM, Reiss D, Hetherington EM, Plomin R (1998), Genetic contributions to continuity, change, and co-occurrence of antisocial and depressive symptoms in adolescence. *J Child Psychol Psychiatry* 39:323–336

Parke RD, Kellam SG, eds. (1994), *Exploring Family Relationships With Other Social Contexts.* Hillsdale, NJ: Erlbaum

Piaget J (1954), *The Construction of Reality in the Child*. New York: Basic Books

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

Rowe DC, Kandel D (1997), In the eye of the beholder? Parental ratings of externalizing and internalizing symptoms. *J Abnormal Child Psychol* 25:265–275

Rutter M (1985), Family and school influence on behavioural development. *J Child Psychol Psychiatry* 26:349–368

Schmitz S, Fulker DW, Mrazek DA (1995), Problem behavior in early and middle childhood: an initial behavior genetic analysis. *J Child Psychol Psychiatry* 36:1443–1458

Silberg JL, Erickson MT, Meyer JM, Eaves LJ, Rutter ML, Hewitt JK (1994), The application of structural equation modeling to maternal ratings of twins' behavioral and emotional problems. *J Consult Clin Psychol* 62:510–521

Simonoff E, Pickles A, Hewitt JK et al. (1995), Multiple raters of disruptive child behavior: using a genetic strategy to examine shared views and bias. *Behav Genet* 25:311–326

Van den Oord EJCG, Koot HM, Boomsma DI, Verhulst FC, Orlebeke JF (1995), A twin-singleton comparison of problem behaviour in 2-3-year-olds. *J Child Psychol Psychiatry* 36:449–458

Van den Oord EJCG, Rowe DC (1998), An examination of genotype-environment interactions for academic achievement in a U.S. National Longitudinal Survey. *Intelligence* 25:205–228

Van den Oord EJCG, Rowe DC (1999), A cousin study of associations between family demographic characteristics and children's intellectual ability. *Intelligence* 27:251–266

Van der Valk JC, van den Oord EJCG, Verhulst FC, Boomsma DI (2001), Using parental ratings to study the etiology of 3-year-old twins' problem behaviors: different views or rater bias? *J Child Adolesc Psychiatry* 42:921–931

Van der Valk JC, van den Oord EJCG, Verhulst FC, Boomsma DI (in press), Using common and unique parental views to study the etiology of 7-year-old twins' internalizing and externalizing problems. *Behav Genet*

Van der Valk JC, Verhulst FC, Neale MC, Boomsma DI (1998a), Longitudinal genetic analysis of problem behaviors in biologically related and unrelated adoptees. *Behav Genet* 28:365–380

Van der Valk JC, Verhulst FC, Stroet TM, Boomsma DI (1998b), Quantitative genetic analysis of internalizing and externalizing problems in a large sample of 3-year-old twins. *Twin Res* 1:25–33

Verhulst FC, Berden G, Sanders-Woudstra JAR (1985), Mental health in Dutch children II: prevalence of psychiatric disorder and relationship between measures. *Acta Psychiatr Scand* 72(suppl):322

Verhulst FC, Van der Ende J, Ferdinand RF, Kasius MC (1997), The prevalence of *DSM-III-R* diagnosis in a national sample of Dutch adolescents. *Arch Gen Psychiatry* 54:329–336

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

Zahn-Waxler C, Schmitz S, Fulker DW, 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. *Dev Psychopathol* 8:103–122