Behavior Genetics, Vol. 33, No. 4, July 2003 (© 2003)

Using Shared and Unique Parental Views to Study the Etiology of 7-Year-Old Twins' Internalizing and Externalizing Problems

J. C. van der Valk,[1,2,4] E. J. C. G. van den Oord,[3] F. C. Verhulst,[2] and D. I. Boomsma[1]

Received 17 July 2000—Final 5 Mar. 2002

In a sample of 1,940 Dutch 7-year-old twin pairs we studied the etiology of individual differences in Internalizing and Externalizing behavioral problems. For the majority of twins in the sample, both maternal and paternal ratings of behavioral problems were obtained from the Child Behavior Checklist. This made it possible to take into account processes underlying agreement and disagreement between maternal and paternal ratings. For both problem behaviors, a Psychometric model fitted the data better than a Rater Bias model, implying that parents, in addition to the behaviors they similarly observed, also assessed unique aspects of their children's behaviors. Relatively large genetic influences were found for Externalizing problems, explaining over 50% of the variance in both boys and girls. For internalizing problems, the heritability was over 30% in both sexes. Shared environmental factors were nearly as important as genetic influences in explaining the variation in behavioral problems. For both Externalizing and Internalizing problems, around 30% of the variance was accounted for by the shared environmental factors.

KEY WORDS: Behavior genetics; children, Child Behavior Checklist; problem behavior; multiple informants; twins.

INTRODUCTION

Epidemiological studies conducted during the last three decades have suggested a median prevalence as high as 13% for problem behaviors in children (Verhulst and Koot, 1992). Furthermore, these studies indicate that girls display more internalizing problems, whereas boys show more externalizing problems. Quantitative genetic studies can shed more light on the genetic and environmental factors underlying children's problem behaviors. This knowledge may not only be of scientific value but could ultimately further improve clinical interventions.

The use of multiple informants such as parents, teachers, clinicians, and children themselves has become the standard practice to assess problem behaviors in childhood. However, it has long been realized that agreement between different raters may be low (e.g., Peterson, 1961). Achenbach *et al.* (1987) reviewed 119 studies and found an average correlation of .6 between similar raters (e.g., parents) dropping to about .2 between different kinds of raters (e.g., children's self-reports and reports from parents or teachers). Despite the common use of multiple raters, there is still no profound understanding of the processes underlying this disagreement and consequently no consensus on how to combine information from multiple sources. One

[1] Department of Biological Psychology, Vrije Universiteit, Amsterdam, The Netherlands.
[2] Department of Child and Adolescent Psychiatry, Erasmus University, Rotterdam, The Netherlands.
[3] Virginia Institute for Psychiatric and Behavioral Genetics, Medical College of Virginia, Richmond, Virginia, USA.
[4] To whom correspondence should be addressed at The Department of Biological Psychology, Vrije Universiteit, Van der Boechorststraat 1, 1081 BT, Amsterdam, The Netherlands. Tel: (32) 36.33.38.44. Fax: (32) 36.33.10.47. e-mail: Jolande@bhp.be

0001-8244/03/0700-0409/0

approach has been to attribute the low agreement to "error" in one or both informants and to select an optimal informant (Fergusson, 1987). Indeed, there is evidence that, for instance, parents or teachers may be relatively insensitive to affective problems in children (Angold, 1987). Self-reports may therefore be preferred to assess affective problems. The alternative view has been that informants contribute, each from their own perspective, different but valid information (Achenbach, 1987). Loeber *et al.* (1989) found, for instance, that children's reports on their conduct problems tended to complement the reports by adults. From this point of view, combining information from all informants would be necessary to obtain a comprehensive picture of children's functioning.

To assess problem behaviors in young children, parent reports are most often used (Simonoff *et al.,* 1995). Parents observe their young child in a variety of situations and across day and night and thus are a valuable source of information. Although ratings of mothers and fathers are correlated, there is never perfect agreement (Achenbach *et al.,* 1987). One explanation is that mothers and fathers may observe the child's behavior in different situations and interact with their child in different ways. For instance, the parent who brings the child to school may be more familiar with the quality of the child's relations with classmates, which is an important correlate of depression and aggression (Newcomb *et al.,* 1993; Parker and Asher, 1987). The disagreement may therefore reflect actual behavior of the child. On the other hand, parents may differ in their perception of the child's behavior and apply different normative standards when evaluating their child's behavior. The parents *themselves* may therefore also be a source of disagreement.

Disentangling true effects of the child's behavior from the parental factors underlying disagreement is an important step to understand rater differences. For this purpose Hewitt *et al.* (1992) proposed so-called Rater Bias and Psychometric models. The Rater Bias model assumes that parents assess exactly the same behaviors in the child and share a common understanding of the behavioral descriptions. Disagreement between the raters is regarded as error, resulting from rater bias and unreliability. The Psychometric model assumes that, in addition to the common (shared) view, parents assess a unique aspect of their child's behavior. Therefore, in the Psychometric model disagreement between the parents arises not only because of error but also because each informant provides from his or her own perspective different but valid information on the child's functioning. It is only with genetically informative data that it becomes possible to discriminate between these two models for rater (dis)agreement. In the Psychometric model the unique aspect of each parent's assessment reflects actual behavior of the child. If genetic influences on the unique aspects of parental assessments are found in a child-based genetic design, that rater-specific genetic variance must reflect the behaviors of the child, because it is unlikely for biases or measurement errors to cause the systematic effects necessary for a model to estimate genetic influences. In contrast, such a finding would exclude the Rater Bias model where this unique aspect of each parent's rating is a function of rater effects only.

Several studies have investigated the etiology of problem behaviors in children using the Child Behavior Checklist (CBCL; Achenbach, 1991; Silberg *et al.,* 1994; Edelbrock *et al.,* 1995; Schmitz *et al.,* 1995; Van den Oord *et al.,* 1996; Zahn-Waxler *et al.,* 1996; Gjone and Stevenson, 1997; Leve *et al.,* 1998; Van der Valk *et al.,* 1998a; Van der Valk *et al.,* 1998b). Only few have taken the processes underlying parental disagreement into account. Hewitt *et al.* (1992) fitted Rater Bias and Psychometric models to the ratings of the Internalizing scale of the CBCL using 983 twin pairs. The Psychometric model fitted better, both for the prepubertal (8 to 11 years) and pubertal cohort (12 to 16 years). In a previous study (Van der Valk *et al.,* 2001) both rater models were fitted to the CBCL ratings of Internalizing and Externalizing problems using 4,016 Dutch 3-year-old twin pairs. The Psychometric model fitted again better than the Rater Bias model. Thus, results from both studies suggested that in addition to a shared view, each parent may provide a unique piece of information about the child's functioning.

In this article, Rater Bias and Psychometric models were fitted using mother and father CBCL ratings of Internalizing and Externalizing behaviors (Achenbach, 1991) in a sample of 1,940 Dutch 7-year-old twin pairs. In addition to studying sources of (dis)agreement between raters, the best-fitting model was used to estimate the genetic and environmental influences. An advantageous byproduct of estimating genetic and environmental effects in the context of a rater model is that more refined estimates may be obtained. For instance, rater bias inflates estimates of shared environment, and measurement error inflates estimates of nonshared environment. However, these rater effects only confound the unique part of the parental ratings. The part of the child's behavior assessed by both parents now solely reflects the behavior of the child so that better estimates are obtained of the relative influences of genetic and environmental factors on children's problem behaviors.

METHOD

Subjects

All participants were members of the Netherlands Twin Registry (NTR), kept by the Department of Biological Psychology at the Vrije Universiteit in Amsterdam. Of all multiple births in the Netherlands, 40–50% are registered by the NTR (Boomsma *et al.,* 1992; Boomsma, 1998). For this study, data from all twins from the birth cohorts 1987, 1988, and 1989 were used. Questionnaires were mailed to 2,855 families within three months of the twins' seventh birthday. After two to three months, reminders were sent and four months after the initial mailing persistent nonresponders were contacted by phone. Families whose address was not available were included in the nonresponse group. A response rate of 68% was obtained ($N = 1{,}940$ families). For 27 twin pairs either one or both of the children had a disease or handicap that interfered severely with daily functioning. These twins were excluded from the analyses. Another 28 twin pairs were omitted because questionnaire items of either one or both of the children were missing.

Zygosity was determined for 639 same-sex twin pairs by DNA or blood group polymorphisms. For the remaining 680 same-sex twin pairs, zygosity was determined by discriminant analysis, using questionnaire items filled out by the mothers. The discriminant function was created using data from 595 twin pairs for which both DNA/blood results and questionnaire items were available. Mothers were asked how much the twins resembled each other in facial structure, hair color, facial color, eye color, and whether they were ever mistaken for each other by the parents themselves, by family, or by strangers. They were also asked if the twins were as much alike as two peas in a pod, whether it was difficult for the parents to separate the twins on a recent picture, and how similar the twins' hair structure was. The discriminant analysis resulted in 6% misclassifications. This implied that merely 2% of the total number of twin pairs was wrongly classified: $(6\% \times 680)/1{,}940 = 2\%$. One twin pair had to be excluded from the study because both the DNA/blood results and the questionnaire on zygosity information were missing.

This procedure left a sample of 342 monozygotic males (MZM), 316 dizygotic males (DZM), 360 monozygotic females (MZF), 300 dizygotic females (DZF), and 566 dizygotic opposite sex (DOS) twin pairs. Data were further divided into twin pairs for which both mothers and fathers had completed the CBCL (267 MZM, 233 DZM, 280 MZF, 230 DZF, 421 DOS) and twin pairs for which only mothers had responded (75 MZM, 83 DZM, 80 MZF, 70 DZF, 145 DOS).

Measures

The Child Behavior Checklist (CBCL, Achenbach, 1991) is a parental rating scale for assessing behavioral and emotional problems in 4- to 18-year-old children. It consists of 20 competence items and 120 problem items. Only the problem items are used in this study. The items pertain to the occurrence of the behavior during the preceding 6 months. Response categories are: 0 if the problem item was not true of the child, 1 if the item was somewhat or sometimes true, and 2 if it was very true or often true.

In this paper the two broad-band scales Internalizing problems and Externalizing problems are analyzed. The scales were composed according to the 1991 profile (Achenbach, 1991). The Internalizing scale consists of the Withdrawn, Somatic Complaints, and Anxious/Depressed syndrome scales. The Externalizing scale consists of the Rule-Breaking Behavior (previously labeled Delinquent Behavior) and Aggressive Behavior syndrome scales. Internalizing problems were only scored when not more than 2 items were missing for the Withdrawn and Somatic Complaints syndrome scale and not more than 3 items were missing for the Anxious/Depressed syndrome scale. For the Externalizing scale the criterion was not more than 3 items missing for both the Rule-Breaking Behavior and Aggressive Behavior syndrome scales. If this inclusion criterion was not reached, the broad-band scale for that subject was regarded as missing. This ensured that the two broad-band scales were always composed of all syndrome scales loading on that scale.

The data were square-root transformed to approximate normal distributions that are required for maximum likelihood estimation. After transformation, all skewness and kurtoses indices were between -1.0 and 1.0, implying that not much distortion is to be expected (Muthén and Kaplan, 1985).

Statistical Analyses

One-way ANOVAs were used to test for mean differences between monozygotic (MZ) and dizygotic

(DZ) twins, between boys and girls, and between mother and father ratings. The homogeneity of the variance in the five zygosity-by-sex groups (MZM, DZM, MZF, DZF, DOS) was tested using Mx (Neale, 1997). The Rater Bias model (Fig. 1) and Psychometric model (Fig. 2) proposed by Hewitt *et al.* (1992) were fitted to the observed variance–covariance matrices. Ten variance–covariance matrices were analyzed simultaneously, namely, those in the five zygosity-by-sex twin groups for which ratings from both parents were available and those in the five zygosity-by-sex twin groups for which only maternal ratings were available. Goodness of fit was assessed by the χ^2 test statistic and Akaike's Information Criterion (AIC). After selecting the best-fitting model, additional tests were performed by fixing genetic and environmental effects to zero, constraining parameters to be equal for boys and girls, and equating rater-specific effects to be the same in mothers and fathers.

RESULTS

Table I shows the mean and standard deviation of the untransformed scales in the twin sample (maternal and paternal ratings) and the Dutch norm group (maternal ratings) discussed by Verhulst *et al.* (1996). For both the Internalizing and Externalizing scale, the means in the twins were quite similar to those in the Dutch norm group. One-way ANOVAs showed no significant differences in means between same-sex MZ and DZ twin pairs. Boys scored significantly higher than girls on the Externalizing scale. Girls scored consistently higher scores than boys on the Internalizing scale, but this was only significant for the maternal

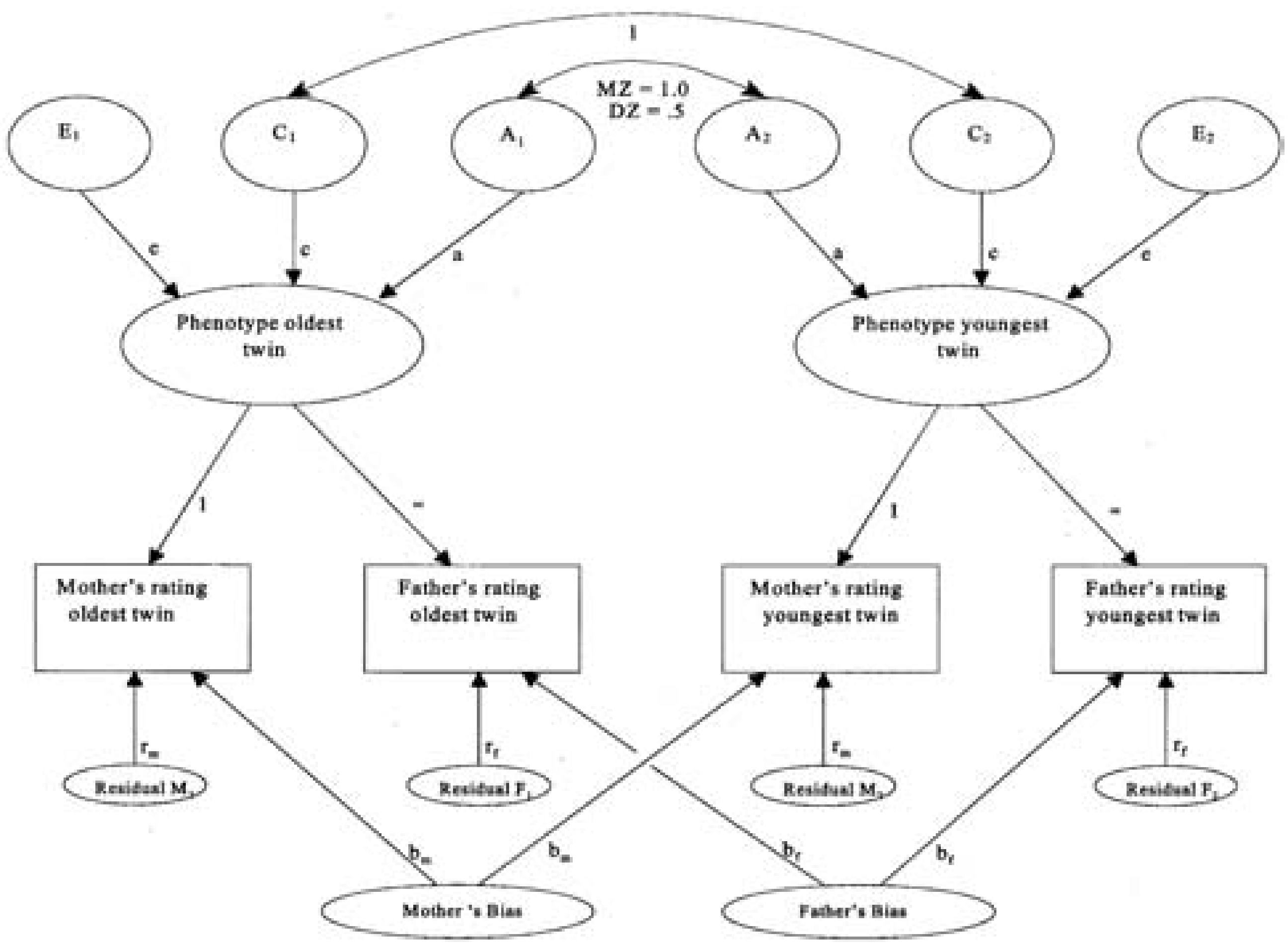

Fig. 1. Rater Bias model for ratings of a pair of twins (oldest and youngest twin) by their parents. Mother's and father's observed ratings (in squares) are linear functions of the latent phenotypes of the twins, mother's and father's bias, and residual errors (M = mother, F = father). Latent phenotypes of the twins are influenced by A, C, and E, representing genetic, shared environmental, and nonshared environmental factors.

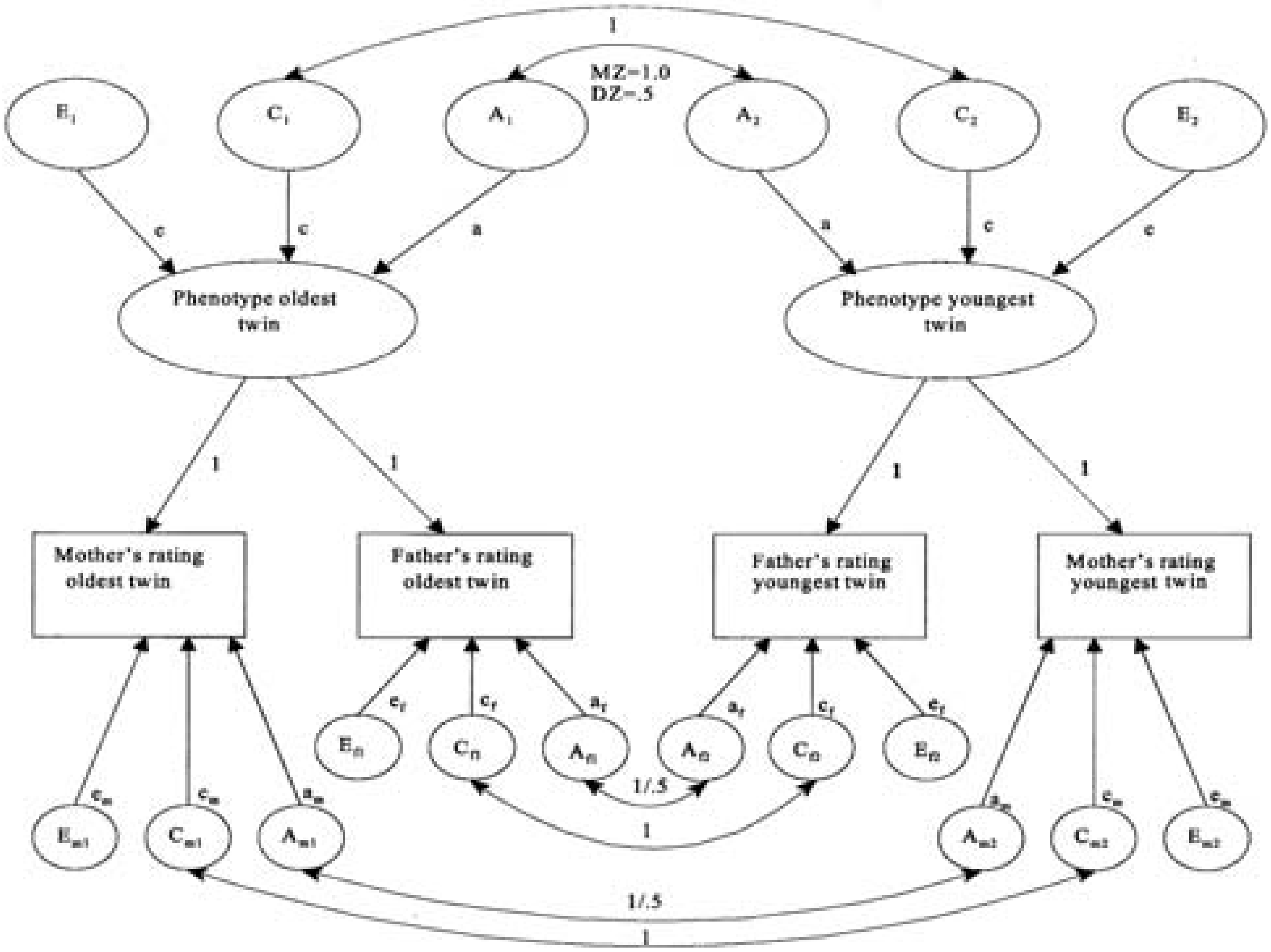

Fig. 2. Psychometric model for ratings of a pair of twins (oldest and youngest twin) by their parents. Mother's and father's observed ratings (in squares) are linear functions of the latent phenotypes of the twins and rater-specific variance. Latent phenotypes of the twins are influenced by common (i.e., across both parents) A, C, and E, representing common genetic, common shared environmental, and common nonshared environmental factors. Rater-specific variance is made up of unique (i.e., to each parent) A, C, and E, representing unique genetic, unique shared environmental, and unique nonshared environmental factors.

ratings of MZ twins (MZM vs. MZF). Maternal and paternal ratings both indicated that MZ and DZ females differed significantly from DOS females with respect to Internalizing and Externalizing problems. Both maternal and paternal ratings also indicated significant differences between MZ and DOS males for Externalizing problems. DZ males differed from DOS males with respect to Internalizing problems when tested both with maternal and paternal ratings. Comparing maternal and paternal ratings, a paired *t*-test showed that maternal ratings were significantly higher than paternal ratings. In summary, no differences were found between same sex MZ and DZ twin pairs. Boys scored higher than girls on Externalizing problems, and girls tended to score higher on the Internalizing scale. Both internalizing and externalizing problem scores were higher for same-sex female twin pairs than opposite-sex female twin pairs. For boys this pattern of differences was not observed. It may be that for girls, being a member of an opposite-sex twin pair buffers against problem behaviors. Alternatively, this finding could be a Type I error. Mothers scored higher levels of problem behavior than fathers, implying a rater effect.

For maternal and paternal ratings the variances of MZM, DZM, MZF, and DZF but not the DOS twins could be constrained to be equal for the Internalizing and Externalizing scale. This suggested that phenomena like sibling interactions that cause these variances to differ (Eaves, 1976; Neale and Cardon, 1992) may not be important. The variances of the ratings of

Table I. Means (Standard Deviations) and Sample Sizes for the Internalizing and Externalizing Scale in a 7-Year-Old Twin Group (per Zygosity; for Both Mother's and Father's Ratings) and a 4- to 11-Year-Old Dutch Norm Group (Mother's Ratings)

	Males				Females			
	Twins			Norm group	Twins			Norm group
	MZM	DZM	DOS		MZF	DZF	DOS	
Internalizing								
Mothers	4.39 (4.24)	4.88 (4.86)	4.04 (3.91)	4.52 (4.27)	5.46 (4.92)	5.42 (4.83)	4.25 (4.38)	5.16 (5.02)
Fathers	3.54 (3.76)	3.95 (4.22)	3.18 (3.29)		3.83 (3.77)	4.23 (4.37)	3.40 (3.71)	
N Moth/Fath	686/571	625/503	567/448	579	723/584	585/483	569/446	593
Externalizing								
Mothers	9.42 (7.07)	8.72 (7.05)	8.66 (7.05)	8.26 (6.26)	7.35 (6.21)	6.76 (5.81)	6.06 (5.72)	6.04 (5.57)
Fathers	8.37 (6.80)	8.15 (6.71)	7.53 (6.52)		6.34 (5.50)	5.96 (5.58)	5.00 (5.07)	
N Moth/Fath	694/580	643/510	575/448	579	732/590	602/492	576/446	593

Note: MZM/DZM = monozygotic/dizygotic males; MZF/DZF = monozygotic/dizygotic females; DOS = dizygotic opposite sex; *N* Moth/Fath = number of children rated by mothers (Moth) and fathers (Fath).

Internalizing and Externalizing problems in the "mothers-only" group could be equated with those of the mothers in the "mothers-plus-fathers" group. Therefore, in the further analyses parameter estimates in the "mothers-only" group were constrained to be equal to the estimates for the mothers in the "mothers-and-fathers" group.

Twin Correlations

Table II shows for the Internalizing and Externalizing scale in the first and second columns the intrarater twin correlations where *one* parent rated both children, in the third and fourth columns the cross-rater twin correlations where either mothers rated the oldest twin and fathers the youngest (Moth/Fath) or the other way around (Fath/Moth), and in the fifth and sixth columns the interparent correlations where mothers and fathers either both rated the oldest or both rated the youngest child. The interparent correlations were comparable for the oldest and youngest twin in all zygosity-by-sex groups. On average, the interparent correlations for the Internalizing scale were .66, and for the Externalizing scale .75. This resembled the interparent correlations obtained in the Dutch norm group (Verhulst *et al.,* 1996). Correlations of opposite-sex and same-sex DZ twin pairs were similar, suggesting that there were no large differences between genetic and environmental effects in boys and girls.

The twin correlations were higher for MZ than for DZ twins, suggesting genetic effects. This difference was especially large for the Externalizing scale. This implied a higher heritability for the Externalizing than the Internalizing scale. Especially for Internalizing problems, the MZ twin correlations were smaller than 1. This indicated nonshared environmental influences. Shared environmental influences were implied by the fact that the DZ twin correlations were larger than half the MZ twin correlations.

Table III shows the genetic, shared, and nonshared environmental influences estimated by fitting univariate models with possible sex differences. For both the Internalizing (first two columns) and Externalizing (last two columns) scale, parameter estimates for mothers' and fathers' ratings were comparable. Although the differences were small, sex differences were significant for the Externalizing scale. The sex differences were neither scalar sex differences nor could they be pinpointed to a specific factor. Most likely, the sex differences were a multivariate effect, caused by small effects on multiple factors. Genetic influences (first row) were largest for the Externalizing scale, explaining more than half of the total variance. For the Internalizing scale, genetic factors ex-plained around 36% of the variance in problem scores. Shared environmental influences (second row) were similar for both the Internalizing and the Externalizing scale, explaining around 32% of the variance. Nonshared environmental influences (third row) explained around 31% of the variance of Internalizing problems and 15% of the variance of Externalizing problems.

The univariate analyses presented above yield a decomposition of the total phenotypic variance and no

Table II. Phenotypic and Twin Correlations for Parental Ratings of Problem Behaviors

	Internalizing scale						Externalizing scale							
	Same rater		Different raters				Same rater		Different raters					
	Intrarater		Crossrater		Interparent		Intrarater		Crossrater		Interparent			
	Twins		Twins		Child		Twins		Twins		Child		Sample sizes[a]	
	Moth/Moth	Fath/Fath	Moth/Fath	Fath/Moth	O	Y	Moth/Moth	Fath/Fath	Moth/Fath	Fath/Moth	O	Y	Moth	Moth + Fath
MZM	.69	.68	.42	.45	.60	.65	.85	.87	.66	.66	.75	.73	75	267
DZM	.49	.44	.27	.39	.73	.68	.54	.55	.41	.38	.76	.67	83	233
MZF	.71	.67	.42	.45	.65	.65	.84	.86	.72	.65	.78	.75	80	280
DZF	.54	.55	.30	.35	.65	.64	.55	.61	.43	.44	.77	.74	70	230
DOS	.51	.54	.26	.29	.61	.60	.62	.57	.42	.41	.76	.64	145	421

Note: MZM/DZM = monozygotic/dizygotic males; MZF/DZF = monozygotic/dizygotic females; DOS = dizygotic opposite sex twins. Same rater: intrarater twins = correlation between the oldest and youngest twin, rated by Moth/Moth = mothers or Fath/Fath = fathers. Different raters: crossrater twins = cross-correlation: either oldest twin rated by mothers and youngest by fathers (Moth/Fath) or the other way around (Fath/Moth). Different raters interparent child: O = correlation between mother and father ratings for the oldest child; Y = correlation between mother and father ratings for the youngest child.

[a] Sample sizes Moth = number of twin pairs rated by mothers only, sample sizes Moth + Fath = number of twin pairs rated by both mothers and fathers.

distinction is made between the variance that is shared and unique to mothers and fathers. To make a separate decomposition of these two parts, the correlations between raters are required. To compute the contributions of the genetic and environmental influences to the variance shared by both raters, MZ and DZ correlations can be compared as in the univariate case. However, now the cross-rater twin correlations (see, for each scale, the third and fourth columns in Table II) have to be used. For both scales the cross-rater twin correlations (third and fourth columns) were higher for MZ than for DZ twins, suggesting that genetic factors affect shared variance. Genetic influences seemed again larger for the Externalizing than for the Internalizing scale. The genetic contribution to the shared variance can be subtracted from the genetic contribution to the total variance, to estimate the genetic effects on the variance that is unique for mothers or fathers. For instance, for the mother ratings of Internalizing problems in boys, the genetic influence on the variance shared by both raters $[2 \times (r_{\text{MZM-cross}} - r_{\text{DZM-cross}}) = 2 \times (.42 - .27) = .30]$ can be subtracted from the genetic contribution to the total variance $[2 \times (r_{\text{MZM}} - r_{\text{DZM}}) = 2 \times (.69 - .49) = .40]$ to obtain an estimate of the genetic contribution to the unique variance $(.40 - .30 = .10)$. Similar calculations performed for girls and for both sexes on the other scale suggested a genetic con-tribution to the unique variance too. These findings are consistent with the Psychometric model, assuming that parental disagreement is not merely caused by "error" but may also reflect that parents assess partly different behaviors.

To estimate the nonshared environmental influences on the variance shared by raters, the interparent correlations have to be used. Table II shows that for Internalizing problems in the MZM group, the interparent correlation was .60 for the oldest twin. However, the MZM cross-rater twin correlation (Moth/Fath) of .42 was lower. This suggested a nonshared environmental contribution of $.60 - .42 = .18$. Shared environmental influences on the variance shared by raters can be estimated as $(2 \times r_{\text{DZM-cross}}) - r_{\text{MZM-cross}} = (2 \times .27) - .42 = .12$. Subtraction from the total variance will again yield estimates of the shared or nonshared contribution to the variance unique to each rater.

Rater Models

Fit indices of the Rater Bias and Psychometric model are presented in Table IV. As indicated by

Table III. Univariate Estimates of Genetic and Environmental Influences on the Internalizing and Externalizing Scale Rated for 7-Year-Old Twins

	Internalizing scale		Externalizing scale	
	Mothers	Fathers	Mothers	Fathers
Genetic	38%	35%	52%	56%
Shared	32%	33%	32%	30%
Nonshared	30%	32%	16%	14%

the lower χ^2 and AIC, the Psychometric model fitted the data better than the Rater Bias model for both the Internalizing and Externalizing scale. This implied that although parents partially assessed the same behaviors (estimated by the common factors in the model), there also was a component that was unique to each rater (estimated by the unique factors in the model). The Cholesky decomposition can be viewed as a psychologically less informative rotation of the Psychometric model (Hewitt *et al.,* 1992). Neither for Internalizing nor for Externalizing did the Cholesky fit the data clearly better than the Psychometric model. Considering the large sample size used, the *p*-values obtained for the Psychometric model were high. Also, the AICs were low for both scales. Taken together, these findings indicated that the Psychometric model fitted reasonably well.

The Psychometric model was further examined for possible simplifications. We did not fit a model that fixed the unique nonshared environmental effect to zero. The reason is that this component includes the measurement error which is always larger than zero for behavior problem scales. Neither for the Internalizing nor for the Externalizing scale could any of the common or unique effects be removed from the model. For the Internalizing scale, the estimates for boys and girls could be constrained to be equal. However, the estimates for the unique factors of mothers and fathers differed. For the Externalizing scale, the estimates for the unique factors of mothers and fathers could be constrained to be equal, but sex differences were found for the common effects. The fit of the simplified model is given in Table IV.

The parameter estimates (expressed as percentages of the variance) for the Internalizing scale, calculated using the best-fitting Psychometric model, are given in Table V. No sex differences were found. Genetic factors explained for mothers 24% and for fathers 28% of the variance in the behavioral ratings that the parents shared. Unique genetic factors explained for mothers 14% and for fathers 4% of the variance in the behavioral ratings uniquely rated by each parent. Estimating unique genetic factors implied that parental disagreement was not merely caused by measurement errors but that each rater also assessed, from his or her own perspective, different but valid aspects of the child's behavior.

Shared environmental factors explained for mothers 19% and for fathers 23% of the variance shared by both raters. These estimates pointed to a pure shared environmental effect unlikely to be affected by possible rater bias. Unique shared environmental factors explained about 13% of the variance unique to mothers and father. Nonshared environmental factors explained for mothers 19% and for fathers 22% of the variance that the parents shared. Unique nonshared environmental factors explained for mothers 11% and for fathers 10% of the variance.

The parameter estimates for the Externalizing scale, calculated using the best-fitting Psychometric model, are summarized in Table VI. Estimates for the ratings of mothers and fathers could be constrained to be equal. Genetic factors explained about half of the observed variance in the behavioral ratings of the Externalizing scale for boys and girls. For boys, genetic factors explained 44% of the variance in the behavioral ratings shared by the parents. For girls, 41% of the variance in the behavioral ratings shared by parents was explained by genetic factors. Of the variance in the behavioral ratings uniquely rated by mothers and fathers, 9% for boys and 10% for girls was accounted for by unique genetic factors.

Shared environmental factors explained for boys 22% and for girls 21% of the variance shared by the parents. This implied that there were pure shared environmental effects on the Externalizing scale. Unique shared environmental factors explained for boys 10% and for girls 12% of the variance. Nonshared environmental factors explained both for boys and girls 10% of the variance shared by the parents. Unique nonshared environmental factors explained for boys 5% and for girls 6% of the variance in the behavioral ratings uniquely rated by mothers and fathers, suggesting small effects for possible unreliability and measurement errors.

DISCUSSION

We studied processes underlying (dis)agreement between maternal and paternal ratings and used the best-fitting model to estimate genetic and environmental influences on Internalizing and Externalizing problems in a sample of 1,940 Dutch 7-year-old twin

Table IV. Model-Fitting Statistics for Psychometric and Rater Bias Model and Simplification of Best-Fitting (Psychometric) Model, for the Internalizing and Externalizing Scale of 7-Year-Old Twin Pairs

	Internalizing scale							Externalizing scale						
	χ^2	*df*	*p*	AIC	χ^2 diff.	*df*	*p*	χ^2	*df*	*p*	AIC	χ^2 diff.	*df*	*p*
Overall model:														
Psychometric model	**53.6**	**47**	**.235**	**−40.4**				**68.8**	**47**	**.021**	**−25.18**			
Rater Bias model	73.0	49	.015	−25.0				129.1	49	.000	31.13			
Simplification of overall model:														
Factor estimates:														
No common genetic effects	87.3	49	.001	−10.7	33.7	2	.000	238.7	49	.000	140.68	169.9	2	.000
No unique genetic effects	79.8	51	.006	−22.3	26.1	4	.000	129.7	51	.000	27.71	60.9	4	.000
No common shared environment	80.0	49	.003	−18.0	26.4	2	.000	100.2	49	.000	2.20	31.4	2	.000
No unique shared environment	114.7	51	.000	12.7	61.1	4	.000	148.0	51	.000	46.03	79.2	4	.000
No common nonshared environment	509.9	49	.000	411.9	456.2	2	.000	358.4	49	.000	260.45	289.6	2	.000
Sex differences:														
No sex differences common effects	55.9	50	.263	−44.1	2.3	3	.514	81.2	50	.003	−18.76	12.4	3	.006
No sex differences unique effects	60.4	53	.226	−45.6	6.8	6	.341	**77.7**	**53**	**.015**	**−28.34**	**8.8**	**6**	**.183**
No sex differences common + unique	**63.1**	**56**	**.241**	**−49.0**	**9.4**	**9**	**.399**	92.4	56	.002	−19.58	23.6	9	.005
Rater differences:														
Unique rater effect: mother-father identical	86.6	53	.002	−19.4	33.0	6	.000	**78.4**	**53**	**.013**	**−27.59**	**9.6**	**6**	**.143**
Simplified model:	**63.1**	**56**	**.241**	**−49.0**				**86.3**	**56**	**.006**	**−25.74**			

pairs. The Psychometric model fitted the data significantly better than the Rater Bias model for both problem scales. This implied that in addition to a common view, parents assess a unique aspect of their child's behavior. These results are in agreement with a previous study (Van der Valk *et al.,* 2001) in which we fitted rater models to the CBCL ratings of Internalizing and Externalizing Problems in 4,016 Dutch 3-year-old twin pairs. They also replicate the findings of Hewitt *et al.* (1992), who reported a better fit of the Psychometric model versus the Bias model in a prepubertal (8 to 11 years) and pubertal (12 to 16 years) cohort of twin pairs.

In the present study, rater-specific genetic variance significantly explained between 4% and 14% of the variance in Externalizing and Internalizing behaviors. In our design that detects effects of the genes of the children, rater effects cannot account for this unique genetic variance. The conclusion must therefore be that part of the rater differences were the result of mothers

Table V. Standardized Genetic and Environmental Influences, Estimated Using Best-Fitting Psychometric Model, for the Internalizing Scale of 7-Year-Old Twins

	Internalizing scale	
	Age 7	
	Mothers	Fathers
Genetic factor:		
Common genetic factor	24%	28%
Unique genetic factor	14%	4%
Shared environmental factor:		
Common shared environment	19%	23%
Unique shared environment	13%	13%
Nonshared environmental factor:		
Common nonshared environment	19%	22%
Unique nonshared environment	11%	10%

Table VI. Standardized Genetic and Environmental Influences, Estimated Using Best-Fitting Psychometric Model, for the Externalizing Scale of 7-Year-Old Twins

	Externalizing scale	
	Age 7	
	Boys	Girls
Genetic factor:		
Common genetic factor	44%	41%
Unique genetic factor	9%	10%
Shared environmental factor:		
Common shared environment	22%	21%
Unique shared environment	10%	12%
Nonshared environmental factor:		
Common nonshared environment	10%	10%
Unique nonshared environment	5%	6%

and fathers assessing different aspects of the child's behavior. Although the rater-specific variance must reflect the behavior of the child, it is likely that this unique variance is confounded by rater effects too. First, unique shared environmental influences explained 10–13% of the total variance. Rater bias may account for at least part of this shared environmental effect. This view would be consistent with the psychometric literature demonstrating the existence of individual differences in response tendencies and normative standards. It is also consistent with epidemiological studies showing that levels of psychopathology in parents are positively associated with parent-reported problem behavior in their children (Breslau *et al.,* 1988; Fergusson and Horwood, 1987; Jensen *et al.,* 1988; Phares *et al.,* 1989). Second, because of the less-than-perfect reliabilities of the assessment scale, the rater-specific nonshared environmental component will be confounded with measurement error.

Collecting information from both parents, in particular the father, may be difficult. Therefore, both in clinical situations and in research, usually mothers are the sole source of parental information about their children's behavioral and emotional problems. Based on findings from the present study, one may conclude that it is important to obtain information from both the father and the mother. However, the rater-specific variance was rather small (about 30% of the total variance) and likely to be confounded with rater bias and measurement error. Thus, the extra information that can be obtained by collecting information from both parents seems modest and, because it is difficult to separate child from rater effects, it is not immediately clear how this rater-specific information should be used. An explanation for the relatively small specific rater variance is that parents mainly interact with their children in the home environment. To obtain a complete picture of the child's functioning, it may therefore be more useful to complement parental information with information from teachers or peers who interact with the child outside the home situation.

Another issue involves the nature of the child effects in the rater-specific component. Parents may agree about the child's behavior that is stable across time and situations. In contrast, rater-specific effects could involve temporary behavioral fluctuations or unique parent–child interactions. From a clinical point of view, the part assessed by both parents, predicting cross-situational consistency and long-term adjustment problems, would then be much more important. Further studies including teacher ratings and assessment made at different ages will be helpful to elucidate the nature of the rater-specific variance and the extent it contributes clinically meaningful information.

Because of the overlap between parental ratings, it could be argued that to obtain a comprehensive picture of the child's functioning other combinations of raters may be more useful. However, this substantial shared variance between parental ratings makes it possible to correct for rater bias and measurement error in the individual ratings and enable researchers and clinicians to assess problem behavior more reliably. If two raters would be assessing very different aspects of the child's behavior, it would not make sense to make such a correction. In this study we used this unique advantage of parental ratings to obtain better estimates of genetic and environmental effects. Results showed that genetic factors were quite important for the Externalizing scale, explaining about 50% of the total variance, and 57% of the reliable trait variance in the behavioral ratings for 7-year-old twin pairs. Heritabilities of the same size were found for 3-year-old twin pairs (Van der Valk *et al.,* 2001), 5-year-old twin pairs (Zahn-Waxler *et al.,* 1996), 5- to 15-year-old twin pairs (Gjone *et al.,* 1996), and 7- to 15-year-old twin pairs (Edelbrock *et al.,* 1995). These results suggest that genetic influences for the Externalizing scale are strong throughout childhood. Sizeable shared environmental effects were also found. This was in accordance with the shared environmental influences observed in 3-year-old twin pairs (Van der Valk *et al.*, 2001), 5-year-old twin pairs (Zahn-Waxler *et al.,* 1996), 5- to 15-year-old twin pairs (Gjone *et al.,* 1996), and 7- to 15-year-old twin pairs (Edelbrock *et al.,* 1995). The shared environmental effects on the reliable trait variance indicated that rater bias was not the sole source. This seems to confirm studies showing that factors like family discord and disruption, lack of affection, and poor supervision may predispose to conduct problems and antisocial behaviors (Rutter, 1985). However, shared environmental influences reflect not only the environment children share within the family but also the environment they share in the wider community such as peers and social norms (e.g., Harris, 1995).

For the Internalizing scale, genetic factors explained 35% of the total variance and nearly 40% of the reliable trait variance. Gjone *et al.* (1996) found similar results for a sample of 5- to 15-year-old twin pairs. However, for 3-year-old twin pairs (Van der Valk *et al.,* 2001) and 5-year-old twin pairs (Zahn-Waxler *et al.,* 1996), genetic influences were much larger. In addition, in the present study shared environment accounted for a considerable part of the reliable trait variance, while for 3-year-old twin pairs no pure shared environmental effect was found (Van der Valk *et al.,* 2001). Gjone *et al.* (1996), examining a sample of twin pairs aged 5–9 and 12–15 years, found a near-significant effect of age on the genetic influence for Internalizing behaviors in terms of a decreasing genetic influence with increasing age. O'Connor *et al.* (1998), studying a sample of 720 siblings initially aged 10 to 18 years, also found a decrease in heritability and a complementary increase in environmental influences over a three-year interval for a composite score of depressive symptoms. Possibly these changes in genetic and environmental effects are caused by developmental differences between older and younger children. Another explanation could be that the CBCL Internalizing assessment for 2- and 3-year-old children taps somewhat different behaviors than the CBCL Internalizing assessment for 4- to 18-year-old children. For instance, at very young ages, Internalizing problems could more strongly reflect temperamental factors (like shyness) than at older ages when these behavior problems might be more closely related to affective symptoms.

For both Internalizing and Externalizing problems the nonshared environmental factor explained a considerable proportion of the reliable trait variance. Because these estimates are unlikely to be affected by measurement error, this study shows that pure idiosyncratic experiences, including those related to peer relationships, diseases, or accidents, are also of importance in the etiology of problem behaviors of 7-year-old twins. No sex differences emerged in genetic and environmental estimates.

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.

ACKNOWLEDGMENTS

This research was supported by a grant from the Sophia Foundation for Medical Research.

REFERENCES

Achenbach, T. M., McConaughy, S. H., and Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin* **101**(2):213–232.

Achenbach, T. M. (1991). *Manual for the child behavior checklist/4-18 and 1991 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Angold, A., Weissman, M. M., John, K., Merikangas, K. R., Prusoff, B. A., Wickramaratne, P., Gammon, G. D., and Warner, V. (1987). Parent and Child Reports of depressive symptoms in children at low and high risk of depression. *Journal of Child Psychology and Psychiatry* **28**:901–915.

Boomsma, D. I. (1998). Twin registers in Europe: An overview. *Twin Research* **1**:34–51.

Boomsma, D. I., Orlebeke, J. F., and Van Baal, G. C. M. (1992). The Dutch twin register: Growth data on weight and height. *Behavior Genetics* **22**:247–251.

Breslau, N., Davis, G. C., and Prabucki, K. (1988). Depressed mothers as informants in family history research. *Psychiatry Research* **24**:345–349.

Eaves, L. J. (1976). A model for sibling effects in man. *Heredity* **36**: 205–214.

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

Fergusson, D. M., and Horwood, L. J. (1987). The trait and method components of ratings of conduct disorder-Part II. Factors related to the trait component of conduct disorder scores. *Journal of Child Psychology and Psychiatry* **28**:261–272.

Gjone, H., Stevenson, J., Sundet, J. M., and Eilertsen, D. E. (1996). Changes in heritability across increasing levels of behavior problems in young twins. *Behavior Genetics* **26**:419–426.

Gjone, H., and Stevenson, J. (1997). The association between Internalizing and Externalizing behavior in childhood and early adolescence: Genetic and environmental common influences? *Journal of Abnormal Child Psychology* **25**:277–286.

Harris, J. R. (1995). Where is the child's environment? A group socialization theory of development. *Psychological Review* **102**: 458–489.

Hewitt, J. K., Silberg, J. L., Neale, M. C., Eaves, L. J., and Erickson, M. (1992). The analysis of parental ratings of children's behavior using LISREL. *Behavior Genetics* **22**(3):293–317.

Jensen, P. S., Traylor, J., Xenakis, S. N., and Davis, H. (1988). Child psychopathology rating scales and interrater agreement I: Parents' gender and psychiatric symptoms. *Journal of the American Academy of Child and Adolescent Psychiatry* **27**:442–450.

Leve, L. D., Winebarger, A. A., Fagot, B. I., Reid, J. B., and Goldsmith, H. H. (1998). Environmental and genetic variance in children's observed and reported maladaptive behavior. *Child Development* **69**:1286–1298.

Loeber, R., Green, S. M., Lahey, B. B., and Stoutham-Loeber, M. (1989). Optimal informants on childhood disruptive behaviors. *Development and Psychopathology* **1**:317–337.

Muthén, B., and Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of nonnormal Likert variables. *British Journal of Mathematical and Statistical Psychology* **38**:171–189.

Neale, M. C. (1997). *Mx: Statistical modeling* (4th ed.), Richmond, VA: Department of Psychiatry, Medical College of Virginia.

Neale, M. C., and Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Dordrecht, The Netherlands: Kluwer Academic.

Newcomb, A. F., Bukowski, W. M., and Pattee, L. (1993). Children's peer relations: A meta-analytic review of popular, rejected, neglected, controversial, and average sociometric status. *Psychological Bulletin* **113**:99–128.

O'Connor, T. G., Neiderhiser, J. M., Reiss, D., Hetherington, E. M., and Plomin, R. (1998). Genetic contributions to continuity, change, and co-occurrence of antisocial and depressive symptoms in adolescence. *Journal of Child Psychology and Psychiatry* **39**: 323–336.

Parker, G. P., and Asher, S. R. (1987). Peer relations and later personal adjustment: Are low-accepted children at risk? *Psychological Bulletin* **102**:357–389.

Peterson, D. R., Becker, W. C., Schoemaker, D. J., Luria, Z., and Hellmer, L. A. (1961). Child behavior problems and parental attitudes. *Child Development* **32**:151–162.

Phares, V., Compas, B. E., and Howell, D. C. (1989). Perspectives on child behavior problems: Comparisons of children's self-reports with parent and teacher reports. *Psychological Assessment: A Journal of Consulting and Clinical Psychology* **1**:68–71.

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

Schmitz, S., Fulker, D. W., and 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.

Silberg, J. L., Erickson, M. T., Meyer, J. M., Eaves, L. J., Rutter, M. L., and Hewitt, J. K. (1994). The application of structural equation modeling to maternal ratings of twins' behavioral and emotional problems. *Journal of Consulting and Clinical Psychology* **62**: 510–521.

Simonoff, E., Pickles, A., Hewitt, J., Silberg, J., Rutter, M., Loeber, R., Meyer, J., Neale, M., and Eaves, L. (1995). Multiple raters of disruptive child behavior: Using a genetic strategy to examine shared views and bias. *Behavior Genetics* **25**(4):311–326.

Van den Oord, E. J. C. G., Verhulst, F. C., Stroet, T. M., and 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**(3):349–357.

Van der Valk, J. C., Verhulst, F. C., and Boomsma, D. I. (1998a). Quantitative genetic analysis of Internalising and Externalising problems in a large sample of 3-year-old twins. *Twin Research* **1**:25–33.

Van der Valk, J. C., Verhulst, F. C., Neale, M. C., and Boomsma, D. I. (1998b). Longitudinal genetic analysis of problem behaviors in biologically related and unrelated adoptees. *Behavior Genetics* **28**(5):365–380.

Van der Valk, J. C., Van den Oord, E. J. C. G., Verhulst, F. C., and Boomsma, D. I. (2001). Using parental ratings to study the etiology of 3-year-old twins' problem behaviors: Different views or rater bias? *The Journal of Child Psychology and Psychiatry and Allied Disciplines* **42**(7):921–931.

Verhulst, F. C., and Koot, H. M. (1992). *Child psychiatric epidemiology: Concepts, methods, and findings.* London: Sage.

Verhulst, F. C., Van der Ende, J., and Koot, H. M. (1996). *Handleiding voor de CBCL/4-18.* Rotterdam: Afdeling Kinder-en Jeugdpsychiatrie, Sophia Kinderziekenhuis/Academisch Ziekenhuis Rotterdam/Erasmus Universiteit Rotterdam.

Zahn-Waxler, C., Schmitz, S., Fulker, D., Robinson, J., and 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.

Edited by Carol Prescott