J. Child Psychol. Psychiat. Vol. 42, No. 7, pp. 921–931, 2001
Cambridge University Press

0021–9630/01 $15.00+0.00

Using Parental Ratings to Study the Etiology of 3-year-old Twins' Problem Behaviors: Different Views or Rater Bias?

J. C. van der Valk
Vrije Universiteit, Amsterdam, and Erasmus University, Rotterdam, The Netherlands

E. J. C. G. van den Oord
University of Utrecht, The Netherlands

F. C. Verhulst
Erasmus University, Rotterdam, The Netherlands

D. I. Boomsma
Vrije Universiteit, Amsterdam, The Netherlands

Child Behavior Checklist questionnaires (Achenbach, 1992), filled in separately by mothers and fathers, were collected for an effective sample of 3501 Dutch 3-year-old twin pairs. To disentangle the child's phenotype from that of the rater, two contrasting models were fitted to the data. One model, called a Rater Bias model, is based on the assumption that both parents assess exactly *the same* behaviors in the child. A weaker alternative of this model, called a Psychometric model, assumes that apart from these common behavioral views, each parent also assesses a *unique aspect* of the child's behavior. A Psychometric model fitted the data of both Internalizing and Externalizing scales significantly better than a Rater Bias model. This implied that each parent provided unique information from his or her own perspective, apart from the common behavioral view. Using this best fitting model, the etiology of both the Internalizing and Externalizing scales was studied. Common factors (influencing behaviors similarly assessed by both parents) were more important than unique factors (influencing behaviors uniquely assessed by one parent). Common genetic factors explained about 50% of the variance of both scales, indicating a possible inborn vulnerability to childhood psychopathology. Common environmental factors not shared between twins (free of unreliability and error) explained around 14% of both scales, suggesting the importance of pure idiosyncratic experiences even for children as young as 3 years. Common environmental factors shared between twins (unconfounded by rater bias) were only found for the Externalizing scale, explaining 18% of the variance. Rater bias and unreliability, if present in the data, were included in the estimates of the unique factors. Unique genetic, shared, and nonshared environmental factors each explained around 8% of the variance for both scales. These small effects could be detected because of the large sample of twin pairs used.

Keywords: Behavior genetics, children, Child Behavior Checklist, problem behavior, rater bias, twins.

Abbreviations: CBCL: Child Behavior Checklist; DOS: dizygotic opposite-sex; DZF: dizygotic females; DZM: dizygotic males; MZF: monozygotic females; MZM: monozygotic males; NTR: Netherlands Twin Registry.

Introduction

To study children's behaviors parental descriptions are often used. Parents observe the child in natural situations at home and in the playground and so are a useful source of information. However, parents do not generally agree in detail about a given child's behavior (Achenbach, McConaughy, & Howell, 1987). There are very good reasons why this should be so. Ratings obtained via the assessment of children by their parents are a function of both parent and child. As noticed by Neale and Cardon (1992) each parent has a different situational exposure, a different degree of insight, and a different perception, evaluation, and normative standard that may create rater differences of various kinds in reporting behaviors. Therefore, when using parental ratings, disentangling the child's phenotype from that of the parent becomes an important methodological problem. For the analysis of genetic and environmental contributions to children's behavior, solutions to this are available when multiple raters, e.g. two parents, rate multiple children, e.g. twins (Neale & Cardon, 1992). To disentangle the child's phenotype from that of the rater two contrasting models have been developed. One model, called the Rater Bias model (Hewitt, Silberg, Neale, Eaves, & Erickson, 1992; Neale & Stevenson, 1989), is based on the assumption that both parents are rating *the same* behaviors in their children. A weaker alternative of this model, called the Psychometric model (Hewitt et al., 1992) assumes that parents are rating *correlated* behaviors in their children.

A Rater Bias model may apply when both parents are equally confronted with the behaviors shown by the child (for instance at home). In this case the parents may have

Requests for reprints to: J. C. van der Valk, MA, The Department of Biological Psychology, Vrije Universiteit, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands (E-mail: Jolande@bhp.be).

a common behavioral view (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/or unreliability. Rater bias in this context is considered to be the tendency of an individual rater to overestimate or underestimate scores consistently. Sources of rater bias are stereotyping, employing different normative standards, or having certain response styles, i.e. judging problem behaviors more or less severely. Because these types of bias may differ between raters, they may also lead to disagreement between raters. Unreliability can become an important source of disagreement when raters cannot give an accurate description of relevant behaviors. For instance, evidence is found that parents may be relatively insensitive to affective disturbances in children (Angold et al., 1987). Using the Rater Bias model it becomes possible to partition the variance in the parental ratings into their components due to reliable trait variance, due to parental bias, and due to unreliability or error in the particular rating of a particular child. Only the reliable trait variance will then be decomposed into its components due to genetic and environmental influences (Neale & Cardon, 1992).

A Psychometric model may apply when, in addition to the common behavioral view and shared understanding of the behavioral descriptions, parents also assess a unique aspect of their child's behavior. Unique behavioral views will occur when the parent also observes the child in distinct situations where they are exposed to different samples of the behavior. For instance, the parent who usually brings the child to a day-care center may also 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 each parent to provide additional information about the child's behavior, apart from the information on which they both agree. Disagreement in this model does not merely arise due to rater bias and/or unreliability, but also because each parent contributes, from his or her own perspective, different but valid information on the child's functioning. Using the Psychometric model it becomes possible to partition the variance in the parental ratings into their components due to trait variance shared between parents and due to trait variance unique to one parent. Genetic and environmental influences can then be estimated apart for the trait variance shared between parents and the trait variance unique to one parent. For the trait variance shared between parents genetic and environmental influences contain only reliable variance. Possible rater bias and/or unreliability can, in this model, only confound the environmental influences estimated for the trait variance unique to one parent. When genetic factors are estimated to influence the behaviors uniquely rated by one parent, the parent must have been assessing "real" unique behavioral views, for error and/or unreliability cannot cause the systematic effects necessary for the model to estimate genetic influences.

Several quantitative genetic studies have collected parental ratings using the Child Behavior Checklist (CBCL; Achenbach, 1991, 1992) to examine the etiology of children's problem behaviors (Edelbrock, Rende, Plomin, & Thompson, 1995; Gjone & Stevenson, 1997; Leve, Winebarger, Fagot, Reid, & Goldsmith, 1998; Schmitz, Fulker, & Mrazek, 1995; Silberg et al., 1994; Van den Oord, Verhulst, & Boomsma, 1996; Van der Valk, Verhulst, Neale, & Boomsma, 1998; Van der Valk, Verhulst, Stroet, & Boomsma, 1998; Zahn-Waxler, Schmitz, Fulker, Robinson, & Emde, 1996). Yet, only a few studies 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 to 17) in 76 families. They did not fit either Psychometric or Rater Bias models. Still, their results showed that the parental ratings contained a substantial shared behavioral view. Simonoff et al. (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. However, from their analyses they could not determine what underlay the shared parental view and described it as due to a shared set of expectations of the parents against which both twins were rated. Hewitt et al. (1992) applied both the Rater Bias and Psychometric model on parental ratings of the Internalizing scale (CBCL) for 983 twin pairs. They found that for both their prepubertal cohort (8 to 11 years) and their pubertal cohort (12 to 16 years) the Psychometric model fitted the data better than the Rater Bias model. Hewitt et al. concluded that for the Internalizing scale, mothers and fathers rate the same phenotype in their children (i.e., have a shared behavioral view). However, unique genetic influences were also found, implying that the rater differences reflected the existence of real unique behavioral views and not just error and/or rater bias.

In the present study we fitted Rater Bias and Psychometric models to the Internalizing and Externalizing scale of 3501 Dutch 3-year-old twin pairs to examine whether disagreement was caused by rater bias and unreliability, or whether it also involved the fact that parents provide unique and complementary information about their children's functioning. A correct representation is not only important from a substantive point of view, but also to obtain accurate estimates of genetic and environmental effects. If a quantitative genetic model does not take rater bias into account, its presence will cause environmental influences shared between twins to be overestimated. Similarly, possible measurement errors will magnify the estimates of idiosyncratic environmental influences. Moreover, it may be incorrect to assume similar heritabilities when parents are actually exposed to different samples of behavior. Thus, using a model that takes possible rater bias and/or unreliability into account allows us to estimate accurate genetic and environmental influences on the behaviors studied. The large sample of twin pairs used in this study provided the power necessary to be able to detect possible small effects.

In short, the processes underlying parental disagreement were examined in a sample of 3-year-old twin pairs and, using a model that best fitted the data, the etiology of Internalizing and Externalizing Problems was studied.

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, 1998; Boomsma, Orlebeke, & Van Baal, 1992). For this study, data from all twins from the birth cohorts 1987–1991 were used. Questionnaires were mailed to 5103 families within 3 months of the twins' third birthday. After 2 to 3 months reminders were sent and 4 months after the initial mailing persistent nonresponders were contacted by phone. Families whose address was not available were included in the non-response group. A response rate of 78.7% was obtained ($N = 4016$ families). Sixty twin pairs were excluded from the analyses because either one or both of the children had a disease or handicap that interfered severely with daily functioning. Another 303 twin pairs were excluded because questionnaire items of either one or both of the children were missing.

Zygosity was determined for 880 same-sex twin pairs by DNA analyses or blood group polymorphisms (tests were administered for 719 twin pairs by the NTR, and for 161 twin pairs, of whom the NTR had no zygosity data available, by their parents). For all other same-sex twin pairs zygosity was determined by discriminant analysis, using questionnaire items. The discriminant function was created using 784 same-sex twin pairs, for which both DNA/blood results and questionnaire items were available. Around the twins' fifth birthday mothers, and around their seventh birthday both mothers and fathers, completed a zygosity questionnaire. Parents 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 a 93.5% correct classification, suggesting that at most 3% of the twins' zygosity was wrongly classified {(6.5% × [4016 − 880 − 1284]) (dizygotic opposite-sex twins not included in group with DNA/blood data or in discriminant analysis)/4016}. Zygosity could not be determined for 152 twin pairs because neither the results from DNA/blood analyses nor the zygosity questionnaires were available. These twin pairs were excluded from the study.

This left a sample of 567 monozygotic males (MZM), 596 dizygotic males (DZM), 654 monozygotic females (MZF), 521 dizygotic females (DZF), and 1163 dizygotic opposite-sex (DOS) twin pairs. For half of the sample both mothers and fathers had been asked to complete a CBCL, and for the other half of the sample only mothers had been asked to reply. Therefore data could be further divided into twin pairs for which both mothers and fathers had replied (293 MZM, 303 DZM, 333 MZF, 261 DZF, 547 DOS) and twin pairs for which only mothers had replied (274 MZM, 293 DZM, 321 MZF, 260 DZF, 616 DOS).

Measures

The CBCL 2/3 (Achenbach, 1992) was developed for parents to score the behavioral and emotional problems of their 2- and 3-year-old children. It consists of 100 problem items that are scored by the parents on a 3-point scale based on the occurrence of the behavior during the preceding 2 months: 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. Dutch syndrome scales and comparability with the syndrome scales as developed by Achenbach (1992) are reported by Koot, Van den Oord, Verhulst, and Boomsma (1997). In this paper the two broad-band scales Internalizing and Externalizing are analyzed. The Internalizing scale consists of the Anxious and Withdrawn/Depressed subscales. The Externalizing scale consists of the Aggressive, Oppositional, and Overactive subscales. For the Internalizing scale subjects were only included if not more than one item was missing for the Anxious, and not more than two items were missing for the Withdrawn/Depressed scale. For the Externalizing scale the inclusion criterion was not more than one item missing for the Aggressive and the Overactive and not more than three items for the Oppositional scale. This ensured that the two syndrome scales were always composed of all problem behaviors 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 kurtosis indices were between −1.0 and 1.0, implying that not much distortion is to be expected (Muthén & Kaplan, 1985).

The Twin Method

Data from monozygotic and dizygotic twins were used to decompose the scores on the Internalizing and Externalizing scales into a contribution of the additive effects of many genes, environmental influences that are shared by twins (like style of parenting, socioeconomic level, or religion), and environmental influences that are not shared by twins (such as an illness, relationships with peers, or measurement errors). For a summary of the twin method, the various assumptions, and the plausibility of these assumptions see Eaves (1982); Falconer (1989); Kendler and Eaves (1986); Martin and Eaves (1977); Neale and Cardon (1992); Plomin, DeFries, and McClearn (1990); for a short explanation in relation to children's problem behaviors see Van der Valk, Verhulst, and Boomsma (1999).

The relative importance of the additive genetic, shared environmental, and nonshared environmental variance components can be derived from the resemblance between MZ twins who are genetically identical and DZ twins who share on average half of their genes. Genetic effects are indicated when the MZ twin correlation r_{mz} is higher than the DZ twin correlation r_{dz}. Shared environmental effects are indicated if the twin correlations are larger than 0 after the genetic effects are partialled out, and nonshared environmental effects are indicated if the correlation between MZ twins is smaller than 1.0. Assuming additive genetic variance so that the genotypic correlation is .5 for DZ twins, the proportion of variance explained by each component can be calculated as follows: genetic variance $= 2 \times (r_{mz} - r_{dz})$, shared environmental variance $= 2 \times r_{dz} - r_{mz}$, and nonshared environmental variance $= 1 - r_{mz}$.

To decompose the variance shared by both parents, the correlation between the twins rated by different raters (cross-correlation) has to be used. This way, the variance is decomposed into additive genetic, shared environmental, and nonshared environmental contributions for which both parents agree. The decomposition can again be made by comparing the resemblance of MZ twins versus DZ twins. Genetic effects are indicated when the cross-correlation is higher for MZ twins compared to DZ twins. Shared environmental effects are indicated if the cross-correlations are larger than 0 after the genetic effects have been partialled out, and a nonshared environmental contribution is indicated when the cross-correlations for MZ twins is smaller than the interparent correlation. Similar formulae to the ones discussed above for the variances can again be used to compute the contributions of each component: genetic contribution $= 2 \times (r_{mz\text{-}cross} - r_{dz\text{-}cross})$, shared environmental contribution $= 2 \times r_{dz\text{-}cross} - r_{mz\text{-}cross}$, and nonshared environmental contribution = interparent correlation $- r_{mz\text{-}cross}$.

These formulae indicate that the whole variance-covariance matrix can be decomposed into a matrix of genetic variances and covariances, a matrix of shared environmental variances and covariances, and a matrix of nonshared environmental variances and covariances. Instead of decomposing each variance and covariance separately, it is preferable to make such a decomposition by fitting multivariate genetic models. For this purpose Hewitt et al. (1992) proposed a Rater Bias and Psychometric model.

Structural Equation Modeling of Twin Data Rated by More than One Rater

In the Rater Bias model (see Fig. 1) (Hewitt et al., 1992) the phenotypes of the twins are a function of three common factors underlying the ratings of both mothers and fathers: a genetic factor (A), a shared environmental factor (C), and a nonshared environmental factor (E). In addition to these three common factors unique factors are modeled: a maternal rater bias factor, a paternal rater bias factor, and residual (unreliability) factors

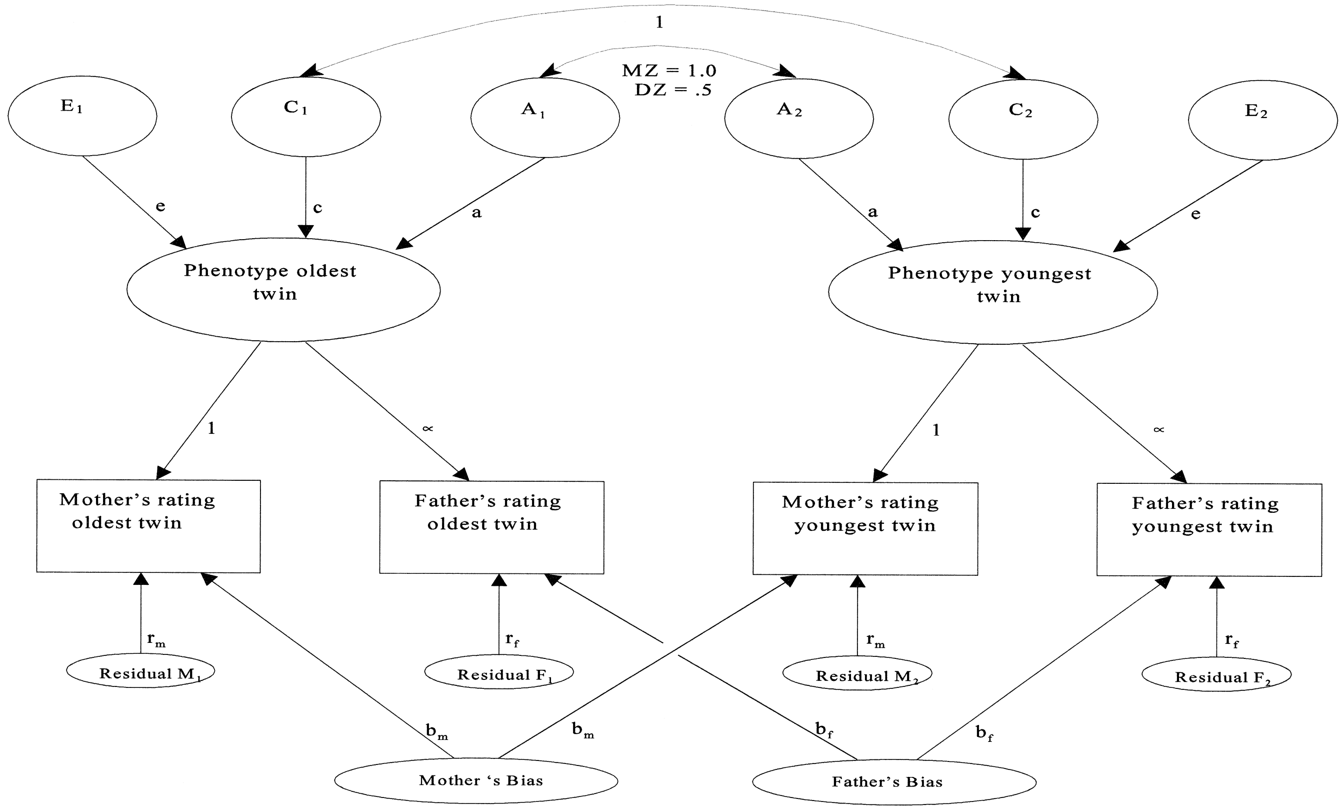

Figure 1. Rater Bias model.

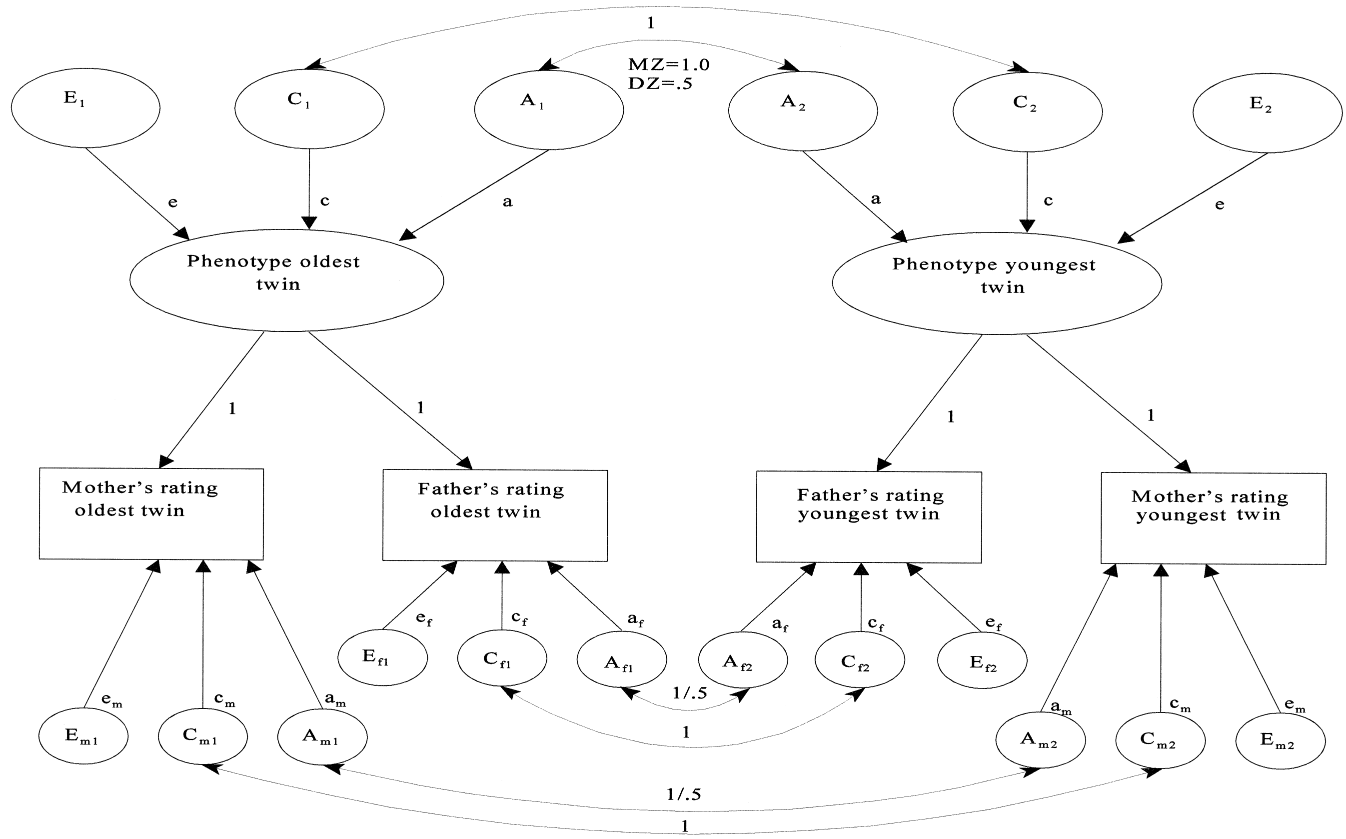

Figure 2. Psychometric model.

affecting each rating. The influence of the common factors (A, C, and E) is assumed to be independent of the maternal and paternal rater bias and unreliability factors.

The Psychometric model (see Fig. 2) (Hewitt et al., 1992) also estimates for the behavioral view common to both parents the influence of a common genetic (A), a common shared environmental (C), and a common nonshared environmental factor (E). These three common factors loading on the twins' phenotypes contain only reliable trait variance, causing the common nonshared environmental factor to contain only pure idiosyncratic environmental effects (McArdle & Goldsmith, 1990) and the common shared environmental factor to contain only pure shared familial environmental effects. In addition the model estimates for the behavioral view unique to one parent three unique factors, a unique genetic ($A_{m/f}$), a unique shared environmental ($C_{m/f}$), and a unique nonshared environmental factor ($E_{m/f}$). In this model disagreement between parents can either be caused by parent's unique behavioral views, leading to different but valid information of each rater, or by rater bias and/or unreliability. Rater bias will confound the unique shared environmental effects, whereas unreliability will confound the unique nonshared environmental effects.

Model Fitting

The program Mx (Neale, 1997) was used to analyze the data through a simultaneous analysis of the 4×4 variance-covariance matrices of 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 of the five zygosity by sex twin groups with only mother ratings. The model describes the observed variance-covariance matrices adequately when the residual variance-covariance matrices are trivially small. 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. Although the Rater Bias model and the Psychometric model do not form a nested pair, they may be compared in terms of parsimony and goodness of fit because they represent alternative sets of constraints on a more general model (Neale & Cardon, 1992).

Fitting the Rater Bias and Psychometric model of Hewitt et al. (1992) to the data showed which model described the processes involved in either agreement or disagreement between the parental ratings best. Monozygotic twin covariances and dizygotic twin covariances were modeled, assuming a correlation between the twins' shared environmental factors of 1.0, regardless of twin type, and a genotypic correlation of 1.0 for monozygotic twins and .5 for dizygotic twins. Estimates for male and female twins were allowed to differ. This model was further examined for possible simplifications. It was tested whether the common and/or unique factors could be removed from the model, whether estimates for boys and girls could be constrained to be the same, and if the unique factors for mothers and fathers could be constrained to be equal. The only factor that was never dropped from the model was the unique nonshared environmental factor, because apart from the influences of idiosyncratic experiences, measurement errors are also estimated in this factor.

Results

Description of the Data

For half of the sample both mothers and fathers, and for the other half of the sample only mothers were asked to complete a CBCL. One-way ANOVA indicated that the ratings for the "mothers only" group did not differ from the mothers in the "mothers and fathers" group. Thus in the analyses, no differences had to be made between mothers who were asked to complete a CBCL alone and mothers who were asked to complete a CBCL while the fathers also filled in a questionnaire. When calculating the means, standard deviations, and correlations both types of mothers were taken as one group. During model fitting, estimates of the "mothers only" group were constrained to be equal to the estimates of the mothers in the "mothers and fathers" group.

The untransformed mean problem scores and standard deviations of the twin sample and those of a Dutch community sample of 2- and 3-year-old children (Koot, 1993) are given in Table 1. For both the Internalizing and Externalizing scale, the ratings given to the twins were quite similar to the ratings given to the Dutch community sample. In a previous study, a comparable level of problem behaviors between 2- and 3-year-old twins and singletons was also found for the subscales of the CBCL and for the Total Problem score (Van den Oord, Koot, Boomsma, Verhulst, & Orlebeke, 1995). Within the twin group, one-way ANOVA showed no significant mean differences between MZ and DZ twin pairs for boys (MZM vs. DZM) or for girls (MZF vs. DZF), neither for maternal nor for paternal ratings. Comparing boys and girls (MZM vs. MZF, and DZM vs. DZF), both mothers and fathers gave significantly higher ratings to the boys for the Externalizing scale; MZ: Mothers: $F(1, 2512) = 30.383, p = .000$; Fathers: $F(1, 1399) = 19.413, p = .000$; DZ: Mothers: $F(1, 2281) = 16.618$, $p = .000$; Fathers: $F(1, 1259) = 7.867$, $p = .005$. For the Internalizing scale ratings for boys and girls did not differ. Comparing mother and father ratings, a paired T-test showed that the ratings for the Externalizing scale given by mothers were significantly higher than ratings given by fathers for both boys and girls; boys: $T(1823) = 4.997$, $p = .000$; girls: $T(1817) = 4.848, p = .000$. For the Internalizing scale no differences were found. Thus, MZ and DZ twin pairs were not rated differently, allowing the use of twin data

Table 1
Means (SDs) *and Sample Sizes for the Internalizing and Externalizing Scale, in a 3-year-old Twin (Per Zygosity) and a 2- and 3-year-old Dutch Community Sample*

	Males				Females			
	Twins				Twins			
	MZM	DZM	DOS	Commun.	MZF	DZF	DOS	Commun.
Internalizing								
Mothers	4.66 (4.05)	4.51 (4.00)	4.59 (4.02)	4.5 (4.4)	4.85 (4.22)	4.74 (4.00)	3.97 (3.88)	4.3 (3.6)
Fathers	4.38 (3.69)	4.53 (4.09)	4.50 (3.93)		4.55 (3.94)	4.81 (4.08)	3.74 (3.75)	
N children M/F	1168/657	1212/672	1193/628	215	1347/744	1072/591	1196/617	205
Externalizing								
Mothers	17.82 (10.50)	16.69 (9.79)	16.00 (10.05)	17.5 (9.5)	15.55 (10.16)	15.02 (9.73)	13.93 (9.49)	16.5 (8.8)
Fathers	16.95 (10.24)	15.94 (9.54)	15.05 (9.79)		14.65 (9.36)	14.44 (9.43)	13.61 (9.19)	
N children M/F	1167/657	1211/669	1195/628	215	1347/744	1072/592	1198/617	205

Commun. = Dutch community sample; *N* children M/F = number of children for mothers (M) and fathers (F).

Table 2
Correlations (Ratings Given by the Same Rater), and Cross-correlations (Ratings Given by Different Raters) between the Twins and the Interparent Correlations, Per Zygosity, for 3-year-olds and Sample Sizes

	Internalizing scale						Externalizing scale							
	Same rater[a]		Different raters[b,c]				Same rater[a]		Different raters[b,c]					
	Twins		Twins[b]		Interparent[c]		Twins		Twins[b]		Interparent[c]		Sample sizes[d]	
	M/M	F/F	M/F	F/M	O	Y	M/M	F/F	M/F	F/M	O	Y	M	M+F
MZM	.65	.65	.49	.49	.63	.61	.79	.75	.61	.59	.71	.66	274	293
DZM	.37	.39	.27	.24	.69	.65	.58	.49	.38	.38	.72	.74	293	303
MZF	.73	.71	.55	.52	.66	.67	.81	.78	.61	.63	.71	.72	321	333
DZF	.35	.43	.27	.22	.70	.64	.53	.41	.26	.35	.67	.71	260	261
DOS	.36	.39	.29	.26	.63	.64	.51	.49	.38	.33	.67	.69	616	547

[a] Same rater: Twins = correlation between the oldest and youngest twin, rated by M/M = mothers or F/F = fathers.
[b] Different raters: Twins = cross-correlation: either oldest twin rated by mothers and youngest by fathers (M/F) or the other way around (F/M).
[c] Different raters: Interparent = correlation between mother and father ratings for (O) oldest child and (Y) youngest child.
[d] Sample sizes: M = number of twin pairs rated by mothers only, M + F = number of twin pairs rated by both mothers and fathers.

for genetic analyses. Boys did receive higher ratings than girls for the Externalizing scale. For this same scale, mothers gave higher ratings to their twin children than fathers did, implying possible rater differences. For the Internalizing scale no differences between boys and girls or between mothers and fathers were found.

The homogeneity of the variance was tested with Mx (Neale, 1997). No differences could be found in the variances and covariances of MZM, DZM, MZF, DZF, and DOS, neither for the Externalizing scale nor for the Internalizing scale. Because the variances did not differ depending on zygosity, siblings were not expected to influence each others' behaviors (sibling interactions).

Twin Correlations

Table 2 shows, for both the Internalizing and Externalizing scale, in the first and second column the correlations between the twins rated by the same rater (mother *or* father rated both children), and in the third and fourth column the cross-correlations between the twins each rated by a different rater (mother *and* father each rated one child). In the fifth and sixth column the interparent correlations between mothers and fathers is given, both for oldest and youngest twin. The interparent correlations were comparable for both oldest and youngest twin for all zygosity by sex groups.

The correlations between the oldest and youngest twins both rated by mothers (M/M; first column) and those both rated by fathers (F/F; second column) can be used to obtain a first estimate of the genetic influences (h^2), the shared environmental influences (c^2), and the nonshared environmental influences (e^2) on the total variance. For instance, if we take for the Internalizing scale the first column "M/M": the genetic influences for boys can be estimated as $(r_{MZM} - r_{DZM}) \times 2 = (.65 - .37) \times 2 = .56$. Nonshared environmental influences for boys can be estimated as $(1 - r_{MZM}) = (1 - .65) = .35$. Following, the shared environmental influences for boys can be estimated as $(2 \times r_{DZM}) - r_{MZM} = (2 \times .37) - .65 = .09$. For girls, father ratings of the Internalizing scale, and mother and father ratings of the Externalizing scale, the correlations between the MZ and DZ twin pairs can be compared in similar ways to obtain a first impression of the genetic and environmental influences.

Table 3
Univariate Estimates of Genetic and Environmental Influences on Internalizing and Externalizing Problems Rated for 3-year-old Twins

	Internalizing problems		Externalizing Problems	
	Mothers	Fathers	Mothers	Fathers
Genetic	69%	59%	52%	56%
Shared		10%	27%	19%
Nonshared	31%	31%	21%	25%

Fitting univariate models (one for mother ratings of Internalizing, one for father ratings of Internalizing, one for mother ratings of Externalizing, and one for father ratings of Externalizing) that estimated three factors—A, C, and E and possible sex differences—the obtained results were comparable to those calculated by comparing the MZ and DZ correlations. Take, for example, the Internalizing scale rated by mothers. As shown in Table 3, no differences between boys and girls were found. The genetic factor explained 69% of the variance and the nonshared environmental factor explained 31%. Using a model fitting approach, no significant shared environmental influences were found.

Univariate analyses make a decomposition of the total variance in genetic, shared environmental, and nonshared environmental factors. To take rater differences into account, the information from the twin's cross-correlations has to be used. By calculating cross-correlations between mother ratings of oldest twins with father ratings of youngest twins (M/F; third column of Table 2) or the other way around (F/M; fourth column), one can make a decomposition of the variance on which both kinds of raters agree. The difference between the decomposition of the variance shared between raters (i.e. common view) and the decomposition of the total variance can be used to estimate the genetic, shared environmental, and nonshared environmental influences on the variance uniquely rated by one particular rater (i.e., unique view). Take, for instance, for the Internalizing scale, the cross-correlations between mother ratings of oldest twins and father ratings of youngest twins (M/F) for boys. The same comparisons between the r_{MZ} and r_{DZ} can be

Table 4
Model Fitting Statistics for Psychometric and Rater Bias Model and Simplification of Best Fitting (Psychometric) Model, for 3-year-old Twins' Internalizing and Externalizing Problems

	Internalizing problems							Externalizing problems						
	χ^2	*df*	*p*	AIC	χ^2diff.	*df*	*p*	χ^2	*df*	*p*	AIC	χ^2diff.	*df*	*p*
Overall model														
Psychometric model	*58.295*	*47*	*.125*	*−35.71*				*56.616*	*47*	*.159*	*−37.38*			
Rater Bias model	81.761	49	.002	−16.24				85.607	49	.001	−12.39			
Simplification of overall model														
Factor estimates														
No common genetic effects	235.911	49	.000	137.91	177.616	2	.000	284.837	49	.000	186.84	228.221	2	.000
No unique genetic effects	84.26	51	.002	−17.74	25.965	4	.000	87.722	51	.001	−14.28	31.106	4	.000
No common shared environment	*58.845*	*49*	*.158*	*−39.16*	*0.55*	*2*	*.760*	89.651	49	.000	−8.35	33.035	2	.000
No unique shared environment	72.67	51	.025	−29.33	14.375	4	.006	108.344	51	.000	6.34	51.728	4	.000
No common nonshared environment	378.837	49	.000	280.84	320.542	2	.000	471.444	49	.000	373.44	414.828	2	.000
Sex differences														
No sex differences common effects	59.928	50	.159	−40.07	1.633	3	.652	59.751	50	.163	−40.25	3.135	3	.371
No sex differences unique effects	65.825	53	.111	−40.18	7.53	6	.275	63.253	53	.158	−42.75	6.637	6	.356
No sex differences common + unique	*71.032*	*56*	*.085*	*−40.97*	*12.737*	*9*	*.175*	*69.166*	*56*	*.111*	*−42.83*	*12.55*	*9*	*.184*
Rater differences														
Unique rater effect: mother-father identical	*66.587*	*53*	*.099*	*−39.41*	*8.292*	*6*	*.217*	*67.635*	*53*	*.085*	*−38.37*	*11.019*	*6*	*.088*
Simplified model	*78.852*	*60*	*.052*	*−41.15*				*78.766*	*59*	*.044*	*−39.23*			

made to estimate the genetic influences on the variance shared by raters, namely $2 \times (r_{MZM\text{-cross}} - r_{DZM\text{-cross}}) = (.49 - .27) \times 2 = .44$. Thus we can conclude that the total genetic variance of 56% can be divided into a genetic influence for behaviors that are similarly rated by the parents of 44% and a genetic influence for behaviors that are uniquely rated by mothers of 12%. This shows that genes of the child affect the unique part of the maternal ratings, implying that the parental disagreement is not merely caused by measurement errors but that mothers, in addition to the common view, also assess a valid unique part of the child's behavior. Finding genetic influences for behaviors that are differently rated by mothers and fathers does not seem to be a chance finding, but arises systematically in the data. Also for the father ratings of boys and for the mother and father ratings of girls, both for the Internalizing and Externalizing scale, similar unique genetic effects are found.

To estimate the environmental influences on the variance shared by raters the interparent correlations (fifth and sixth columns for oldest and youngest twin, respectively) have to be used. Table 2 shows that for the Internalizing scale the interparent correlation (between mothers and fathers of the same child) in the MZM group was .63 for the oldest twin. The cross-correlation (between mothers and fathers of different children) was .49, indicating a nonshared environmental contribution on the variance shared by raters of: interparent correlation $- r_{mzm\text{-cross}} = .63 - .49 = .14$. Thus the nonshared environmental influences can be divided into an influence for behaviors that are similarly rated by both parents of 14% and an influence for behaviors that are uniquely rated by mothers of 21% (i.e., 35%–14%). Shared environmental influences on the variance shared by raters can be estimated as $(2 \times r_{DZM}) - r_{MZM} = (2 \times .27) - .49 = .05$. Taking rater differences into account the shared environmental influences can be divided into an influence for behaviors that are similarly rated by the parents of 5% and an influence for behaviors that are differently rated by mothers of 4% (i.e., 9%–5%). For the cross-correlations of father ratings for boys, mother and father ratings for girls, and all ratings of the Externalizing scale, similar comparisons can be made.

Rater Models

As indicated by the lower χ^2 test statistic and the lower AIC in Table 4, the Psychometric model fitted the data better than the Rater Bias model both for the Internalizing and the Externalizing scale. This signified that although both parents partially assessed the same behaviors, there also was a component that was unique to each rater. For sake of comparison we also performed a Cholesky or triangular 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). It assumes that each parent only assesses on unique aspects of the child's behavior. Parental ratings may be correlated but for unspecified reasons. This view may be appropriate if mothers and fathers only report on behaviors observed in distinct situations, or if they do not share a common understanding of the behavioral descriptions. Neither for the Internalizing scale nor for the Externalizing scale did this saturated model fit the data any better than the Psychometric model. The high *p*-values obtained for the Psychometric model of both problem scales were remarkable, especially when considering the large sample size used (Neale, 1997). This indicated a very good fit of the model to the data.

Table 5
Genetic and Environmental Influences, Estimated Using Best Fitting Psychometric Model, for Internalizing and Externalizing Problems Rated for 3-year-old Twins

	Internalizing	Externalizing
Genetic factor		
Common genetic factor	57%	47%
Unique genetic factor	9%	7%
Shared environmental factor		
Common shared environment		18%
Unique shared environment	5%	8%
Nonshared environmental factor		
Common nonshared environment	16%	12%
Unique nonshared environment	13%	8%

The Psychometric model was further examined for possible simplifications. Only the common shared environmental factor could be omitted from the model for the Internalizing scale. For the Externalizing scale none of the common or unique factors could be omitted. Other model simplifications worked for both scales. Between boys and girls, the estimates of the common and the unique factors could be constrained to be equal. Between mother and father ratings of a sibling only the estimates of the unique factors could be constrained to be equal. The fit of the most simplified model is given in Table 4.

The percentages of variance explained by the common and unique genetic, shared, and nonshared environmental factors are given in Table 5. A major part of the variance was explained by common factors. For both the Internalizing and the Externalizing scale the largest part of the variance was explained by the common genetic factor, explaining 57% and 47% respectively. The common nonshared environmental factor explained 16% of the variance for the Internalizing scale and 12% for the Externalizing scale. The common shared environmental factor only had an influence on the Externalizing scale, explaining 18% of the variance. The unique factors explained a relatively small part of the variance. For the Internalizing scale unique genetic factors explained 9%, unique shared environmental factors explained 5%, and unique nonshared environmental explained 13% of the variance. For the Externalizing scale unique factors also explained relatively small parts of the variance, of 7% genetic influence, 8% shared, and 8% nonshared environmental influences.

Discussion

We examined the processes underlying agreement and disagreement between maternal and paternal ratings and, using a model that best fitted the data, studied the etiology of the Internalizing and Externalizing scale, employing a sample of 3501 Dutch 3-year-old twin pairs. The Psychometric model (Hewitt et al., 1992) fitted the data significantly better than the Rater Bias model, implying that although both parents partially assessed the same behaviors in their children, there also was a component that was unique to each rater. These results are in agreement with the results of Hewitt et al. (1992), who also found a good fitting of the Psychometric model for both their prepubertal (8 to 11 years) and their pubertal (12 to 16 years) cohort of twin pairs. Also Rowe and Kandel (1997), although not fitting Psychometric and Rater Bias models, found that mother and father ratings contained a component that was unique to one rater in addition to a shared behavioral view.

When a Psychometric model fits genetically informative data better than a Rater Bias model, unique genetic factors can be estimated for behaviors that are differently assessed by the diverse raters. If unique genetic factors are estimated in a model, systematic effects must exist in the data that are not expected when differences between parental ratings are only caused by rater bias and unreliability. Thus, because unique genetic factors were estimated in the present study the conclusion must be that the observed rater differences are the result of the fact that raters really do assess different aspects of the child's behavior. As was already suggested by Achenbach et al. (1987), unique interactions between a certain parent and the child might allow each parent to provide additional information about the child's behavior, apart from the information on which they both agree. It thus seems important to collect data from multiple informants. As outlined by Achenbach (1992), "because any reports by any informants may be affected by characteristics of the informants, as well as by their own particular knowledge of the child's behavior, no single informant's reports can provide a complete picture".

The genetic and environmental influences were estimated while the underlying processes causing agreement or disagreement between the individual raters were taken into account. By taking these effects along in the model, more accurate estimates of genetic and environmental influences were obtained. The common genetic, shared, and nonshared environmental factors, influencing behaviors similarly assessed by both parents, explained the largest part of the variance (around 75%). Thus, although each parent does assess unique aspects of the children's behaviors, most of the behaviors are similarly assessed by both parents. Common genetic factors explained about 50% of the variance of both the Internalizing and Externalizing scale, implying a possible inborn vulnerability.

Decomposing the genetic, shared, and nonshared environmental influences into common and unique factors allowed us to estimate the common shared and nonshared environmental factors apart from rater bias and unreliability. The common nonshared environmental factor explained 14% of the variance, indicating a pure independent environmental effect on the Internalizing and Externalizing scales that cannot be explained by measurement error or unreliability. Thus, for children as young as 3 years of age, idiosyncratic experiences already seem to be influencing their behaviors. The common shared environmental factor explained 18% of the variance, suggesting that for the Externalizing scale there is a pure shared environmental influence that is not confounded by rater bias. The importance of shared environmental influences for externalizing behaviors have been demonstrated by various epidemiological studies. Family discord and disruption, lack of affection, and poor supervision all predispose to conduct disturbance and antisocial behavior (Rutter, 1985). However, often it is not family adversity as such but its persistence that predicts chronic problems (Campbell, 1995). To detect shared environmental effects for 3-year-old children thus seems a remarkable finding. An alternative explanation might be that the siblings have been imitating each others' behaviors. Even though the variances and covariances were found to be the same for all five twin groups (MZM, DZM, MZF, DZF, DOS), in a previous study we did find a small influence of sibling interactions (Van der Valk et al., 1998a). Sibling interactions for externalizing behaviors have also been found by Hewitt et al. (Neale & Cardon, 1992) for a sample of 8- to 16-year-old twins. If siblings imitate each other's externalizing behaviors, the estimates of the common shared environmental factor for the Externalizing scale might be inflated. Another explanation might be correlated rater bias, for instance parents copying each others' answers. However, this explanation does not seem very likely. For if this would have been the case, the same common shared environmental influence should also have been found for the Internalizing scale, because the items of the Internalizing and Externalizing scales were mixed on the CBCL.

Unique genetic, shared, and nonshared environmental factors, influencing behaviors differently rated by the parents, each explained around 8% of the variance for both scales. Rater bias and unreliability, if present, were included in the estimates of these unique factors. Probably these small unique effects were only detected because this study used a large sample of 3501 twin pairs. Rater bias may have confounded the estimates of the unique shared environmental factors. Nevertheless, considering the modest influence of the unique shared environmental factors of 5% and 8% for the Internalizing and Externalizing scale, respectively, these possible effects of rater bias were small. This result contrasts with the findings of Neale and Stevenson (1989) and Simonoff et al. (1998). These two studies tested for the possible influence of rater bias and found this influence to be significant. Maybe this difference emerged because of the subject studied. Neale and Stevenson investigated temperament in $3\frac{1}{2}$-year-old twins and Simonoff et al. examined hyperactivity in 8- to 16-year-old twins. For temperament and activity measures, it is common to find DZ twin correlations that are too low. Simonoff et al. examined this phenomenon and found that these too low DZ correlations were not caused by siblings influencing each other, but by parental rater bias (parents contrasting the twins when rating their hyperactivity). Possibly parents do not contrast their children's behaviors for the Internalizing and Externalizing scale. However, the difference may also have emerged because the various questionnaires used may have differed in their sensitivity for rater bias. The current study used the CBCL (Achenbach, 1992), while Neale and Stevenson used EASI temperament scales (Buss & Plomin, 1975), and Simonoff et al. used three hyperactivity items from the Rutter A questionnaire (Rutter, Tizard, & Whitmore, 1970).

Unreliability and measurement error may have confounded the estimates of the unique nonshared environmental factors. Nevertheless, considering the small size of these estimates, of 13% and 8% for the Internalizing and Externalizing scale, respectively, these effects cannot have been strong. Possibly measurement error and unreliability were low because of the high internal consistency shown by the Dutch factor solution of the CBCL (Koot et al., 1997). Cronbach's alphas for the Externalizing scales are: Aggressive .82, Oppositional .91, Overactive .78; and for the Internalizing scales: Withdrawn/Depressed .64, Anxious .83.

Neither sex differences, nor distinct estimates for mothers and fathers for the unique factors, were needed. The behaviors of 3-year-old children are predominantly influenced by the child's genotype. Parental guidance in this case may not be so dependent on the parents' own values and ideas, but may be more directed by the child's genotype. Maybe parents can only guide the child's behavior when the child is somewhat older, able to understand other peoples' values, and can thus direct his or her behavior accordingly. This could mean that at such a young age, the genotype of the child determines what kind of environmental influences the child experiences. In the literature there is cumulating evidence that genotype-environment correlations are important for children's development. For example, a number of studies have shown that when environmental measures (such as parenting behaviors) are used as the dependent variable in a behavior genetic analysis, the correlations between environmental measures of relatives increases with the degree of genetic relatedness (Braungart, Fulker, & Plomin, 1992; Goodman & Stevenson, 1991; Plomin, Reiss, Hetherington, & Howe, 1994; Rende, Slomkowski, Stocker, Fulker, & Plomin, 1992). This suggests that environmental measures tend to reflect the differential genetic resemblance of relatives and that they are dependent on the genetic propensities of individuals. A correlation between genotype of the child and environmental influences was not incorporated in the model and thus could have inflated the genetic estimates.

If at a young age the genotype of the child determines the environmental influences the child experiences, a relatively high genetic estimate with smaller shared and nonshared environmental estimates would be expected. Subsequently, when the child matures, parental guidance may become less directed by the child's genotype and more by the parent's own values and ideas. If this is correct, estimates of environmental influences will then increase for school-aged children compared with preschool children.

This paper used a nonclinical sample of twin pairs, showing problem behaviors in the normal range. Whether similar results apply to clinical populations, showing problem behaviors in an extreme range, remains to be explored. Also, estimates found using quantitative genetic techniques do not pertain to the individual but involve average differences between individuals in the population. For other populations, or for specific individuals, different estimates might be applicable. Even though large genetic influences were found for both problem scales, implying a possible inborn vulnerability for children with problem behaviors, this does not mean that those behaviors are unchangeable. The finding of genetic effects implies hereditary propensities, not predestination (Plomin & Daniels, 1986).

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

References

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

Achenbach, T. M. (1992). *Manual for the Child Behavior Checklist/2–3 and 1992 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

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

Angold, A., Weissman, M. M., John, K., Merikangas, K. R., Prusoff, B. A., Wickramaratne, P., Gammon, G. D., & 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., & Van Baal, G. C. M. (1992). The Dutch twin register: growth data on weight and height. *Behavior Genetics*, *22*, 247–251.

Braungart, J., Fulker, D. W., & Plomin, R. (1992). Genetic mediation of the home environment during infancy. A sibling adoption study of the HOME. *Developmental Psychology*, *28*, 1048–1055.

Buss, A. H., & Plomin, R. (1975). *A temperament theory of personality development*. New York: Wiley.

Campbell, S. B. (1995). Behavior problems in preschool children: A review of recent research. *Journal of Child Psychology and Psychiatry*, *36*, 113–149.

Eaves, L. J. (1982). The utility of twins. In V. E. Anderson, W. A. Hauser, J. K. Penry, & C. F. Sing (Eds.), *Genetic basis of the epilepsies* (pp. 249–276). New York: Raven Press.

Edelbrock, C., Rende, R., Plomin, R., & 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.

Falconer, D. S. (1989). *Introduction to quantitative genetics*. Harlow, UK: Longman Scientific and Technical.

Gjone, H., & 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.

Goodman, R., & Stevenson, J. (1991). Parental criticism and warmth towards unrecognized monozygotic twins. *Behavior and Brain Sciences*, *14*, 394–395.

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

Kendler, K. S., & Eaves, L. J. (1986). Models for the joint effect of genotype and environment on liability to psychiatric illness. *American Journal of Psychiatry*, *143*, 279–289.

Koot H. M. (1993). *Problem behaviour in Dutch pre-schoolers*. Doctoral dissertation, Erasmus University, Rotterdam.

Koot, H. M., Van den Oord, E. J. C. G., Verhulst, F. C., & Boomsma, D. I. (1997). Behavioural and emotional problems in young pre-schoolers: Cross-cultural testing of the validity of the Child Behavior Checklist 2/3. *Journal of Abnormal Child Psychology*, *25*, 183–196.

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

Martin, N. G., & Eaves, L. J. (1977). The genetical analysis of covariance structure. *Heredity*, *38*, 79–95.

McArdle, J. J., & Goldsmith, H. H. (1990). Alternative common factor models for multivariate biometric analyses. *Behavior Genetics*, *20*, 569–608.

Muthén, B., & 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 modelling* (4th ed). Richmond, VA: Medical College of Virginia, Department of Psychiatry.

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

Neale, M. C., & Stevenson, J. (1989). Rater bias in the EASI Temperament Scales: A twin study. *Journal of Personality and Social Psychology*, *56*, 446–455.

Plomin, R., & Daniels, D. (1986). Genetics and shyness. In W. H. Jones, J. M. Cheek, & S. R. Briggs (Eds.), *Shyness: Perspectives on research and treatment* (pp. 63–80). New York: Plenum Press.

Plomin, R., DeFries, J. C., & McClearn, G. E. (1990). *Behavioral genetics: A primer*. San Francisco: Freeman.

Plomin, R., Reiss, D., Hetherington, E. M., & Howe, G. (1994). Nature and nurture: Genetic influence on measures of the family environment. *Developmental Psychology*, *30*, 32–43.

Rende, R. D., Slomkowski, C. L., Stocker, C., Fulker, D. W., & Plomin, R. (1992). Genetic and environmental influences on maternal and sibling interaction in middle childhood: A sibling adoption study. *Developmental Psychology*, *28*, 484–490.

Rowe, D. C., & Kandel, D. (1997). In the eye of the beholder? Parental ratings of Externalizing and Internalizing symptoms. *Journal of Abnormal Child Psychology*, *25*, 265–275.

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

Rutter, M., Tizard, J., & Whitmore, K. (1970). *Education, health and behaviour*. Longman: London (reprinted 1981, Melbourne Krieger).

Schmitz, S., Fulker, D. W., & Mrazek, D. A. (1995). Problem behavior in early and middle childhood: An initial behavior genetic analysis. *Journal of Child Psychology and Psychiatry*, *36*, 1443–1458.

Silberg, J. L., Erickson, M. T., Meyer, J. M., Eaves, L. J., Rutter, M. L., & 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., Hervas, A., Silberg, J. L., Rutter, M., & Eaves, L. (1998). Genetic influences on childhood hyperactivity: Contrast effects imply parental rating bias, not sibling interaction. *Psychological Medicine*, *28*, 825–837.

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

Van den Oord, E. J. C. G., Koot, H. M., Boomsma, D. I., Verhulst, F. C., & Orlebeke, J. F. (1995). A twin-singleton comparison of problem behaviour in 2–3-year-olds. *Journal of Child Psychology and Psychiatry*, *36*, 449–458.

Van den Oord, E. J. C. G., Verhulst, F. C., & Boomsma, D. I. (1996). A genetic study of maternal and paternal ratings of problem behaviors in 3-year-old twins. *Journal of Abnormal Psychology*, *105*, 349–357.

Van der Valk, J. C., Verhulst, F. C., & Boomsma, D. I. (1999). Studying the development of children's problem behaviors using quantitative genetic techniques. In H. M. Koot, A. A. M. Crijnen, & R. F. Ferdinand (Eds.), *Child psychiatric epidemiology: Accomplishments and future directions* (pp. 116–141). Assen, The Netherlands: Van Gorcum.

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

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

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

Manuscript accepted 14 March 2001