*Behavior Genetics, Vol. 25, No. 4, 1995*

# Multiple Raters of Disruptive Child Behavior: Using a Genetic Strategy to Examine Shared Views and Bias

**Emily Simonoff,[1,5] Andrew Pickles,[1] John Hewitt,[2] Judy Silberg,[3] Michael Rutter,[1] Rolf Loeber,[4] Joanne Meyer,[3] Michael Neale,[3] and Lindon Eaves[3]**

*Received 31 May 1994—Final 28 Nov. 1994*

Most research on child behavior incorporates information from different individuals. While agreement between informants is generally only modest, there is little understanding of the processes underlying disagreement. In twin studies, differential agreement among raters for MZ and DZ twins is of particular concern. The processes underlying differences among mother, father, and child ratings of oppositional and conduct disorder symptoms are explored. Evidence in favor of a shared parental view of behavior is presented. Parental ratings give higher intrapair correlations, which could be due to either parents rating their twins more similarly or twins contrasting themselves. Rater bias and situational specificity are among the possible explanations of differential ratings. The effects of incorporating multiple raters of behavior on estimates of genetic and environmental effects are explored. These suggest that genetic influences are greater for the shared (multiple-rater) phenotype than for individual ratings; reduction in measurement error is only a partial explanation.

**KEY WORDS:** Twins; rater bias; child behavior problems; parental ratings; conduct disorder; heritability.

## INTRODUCTION

Multiple raters of an individual's behavior are frequently used in the behavioral sciences. The practice arises out of an awareness not only of the error associated with a single measure of behavior but also of the different points of view and interpretations of behavior that various raters may make. The use of multiple raters of behaviors, traits, and disorders in childhood has now become the standard practice. Young children, e.g., under the age of 8, are often considered unable to give reliable and valid assessments of their own behavior and there is no general consensus regarding the age at which their self-ratings should receive the priority given to adult self-ratings. In middle childhood and adolescence, there is less concern regarding children's understanding of the concepts they are asked to rate, but their perception of their own behavior may be different from that of others. Most frequently, parents have been used as raters of children's behavior and emotions, along with teachers and the children themselves. In general, there is only moderate agreement between the various raters even in general population samples. The finding includes both questionnaire and interview measures. In questionnaire studies, raters have included children, parents and teachers. In general the greatest agreement has been between mothers and fathers, followed by parents (usually mothers) and teachers

[1] MRC Child Psychiatry Unit, Institute of Psychiatry, De Crespigny Park, London SE5 8AZ, England. Fax: 0171 252 5106. e-mail: SPJWEAS@UCL.AC.UK.
[2] Institute of Behavior Genetics, University of Colorado, Boulder, Colorado 80302.
[3] Department of Human Genetics, Medical College of Virginia, Richmond, Virginia.
[4] Western Psychiatric Institute and Clinic, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania.
[5] To whom correspondence should be addressed.



0001-8244/95/0700-0311$07.50/0 © 1995 Plenum Publishing Corporation

(e.g., Kazdin *et al.*, 1983; Phares *et al.*, 1989). Usually the poorest agreement is between children and adults (generally parents, as teacher-child agreement is less frequently examined). Interview studies have compared the ratings of parents and children. While some studies, e.g., Herjanic *et al.* (1975), found a reasonably high agreement when using the percentage of questions in which parents and children give the same answer, those which have used the kappa statistic to account for chance agreement have suggested that the consensus is in fact quite low (Herjanic and Reich, 1982; Kashani *et al.*, 1985; Orvaschel *et al.*, 1982). The reasons for higher mother-father than parent-teacher agreement may include the fact that mothers and fathers are evaluating their children's behavior in the same setting, while teachers are seeing children in a different situation with often very different expectations.

There is at present little real understanding of what underlies the lack of agreement. For both questionnaires and interviews, the lack of agreement is considerably below the test-retest reliability (Achenbach *et al.*, 1987). One possible explanation is that raters vary in their conceptualization of the behaviors and emotions being tapped. This might be due in part to a *lack* of understanding in younger children. In support of this, several studies have found that correlations between parent and child ratings increased with age (Edelbrock *et al.*, 1986; Herjanic *et al.*, 1975; Silverman, 1992). However, in their meta-analysis, Achenbach *et al.* (1987) found that parent-child agreement was generally better in studies of 6 to 11 year olds than for adolescents. Another explanation could be that either parents or children are more likely to describe pathological behaviors. There is considerable evidence that parents are more likely to report symptoms of overactivity, inattention, and oppositional behavior than their children (Edelbrock *et al.*, 1986; Herjanic and Reich, 1982; Kashani *et al.*, 1985; Loeber *et al.*, 1991). On the other hand, children more frequently endorse emotional symptoms, including phobias and obsessional behavior (Herjanic and Reich, 1982) and depression (Angold *et al.*, 1987; Kashani *et al.*, 1985; Verhulst and van der Ende, 1992). This may reflect that the effects of children's behavior are more apparent to others than to themselves, while children's feelings are more accessible to the children. A third possibility is that the same emotions or behaviors are interpreted differently by parents and children. Thus, in administering a standardized research interview to both parents and children, Kashani *et al.* (1985) highlighted several cases in which children's self-reports of anxiety and depression symptoms were not supported by parents, who endorsed a range of oppositional symptoms instead.

While low agreement between informants is now well recognized, there is no consensus on how to deal with it. One way is to select the best informant, as judged in a number of ways, such as agreement with other raters and predictive validity. Thus Loeber and colleagues (1989, 1991) have argued that parents and teachers are better informants for hyperactivity and oppositional behavior, while children and parents should be used to elicit conduct disorder symptomatology. Another possibility is to combine the information from all available informants. This can be done in a number of ways. Individual scores from different raters can be added to create one grand total. Alternatively, individual behaviors or symptoms can be coded as present if any rater endorses the item. Finally, for categorical data, such as the presence or absence of a psychiatric diagnosis, information can be combined at the diagnostic level. Reich and Earls (1987) suggest a more complex, diagnosis-specific method which requires certain numbers of symptoms from different informants before a diagnostic threshold is reached. The major drawback to these methods of combining data across raters is that it ignores the processes that play a role in developing agreement or disagreement. Such an understanding is important in determining not only which rating is "correct," but also why the raters disagree. Much research treats lack of agreement between raters as a nuisance, a marker of error in ratings, rather than a systematic effect. Thus, the work by Fergusson and Horwood (1987a, b) uses a latent variable approach to defining a pervasive conduct disorder trait as rated by mothers and teachers and showed that this predicted later pervasive conduct problems better than either rating alone. However, conduct problems as perceived by only mothers or only teachers may have *different* correlates, of equal importance, than those that are pervasive.

Rater differences may have other effects on genetic designs and, in particular, twin studies. Systematic effects in ratings may lead twins to be judged as more similar to or more different from each other, thereby affecting twin correlations.

Such rater effects could apply differentially to monozygotic (MZ) and dizygotic (DZ) twin pairs, influencing heritability estimates.

This paper examines ratings made by twins, their mothers, and their fathers and explores some of the processes that may be influencing agreement between raters. Data on oppositional and conduct behavior are used, as it is an area of only fair agreement across raters. We also look across both raters and types of information (questionnaires versus interviews) in a variant of the multimethod, multitrait analysis (Campbell and Fiske, 1959; Fergusson and Horwood, 1987a, b). Unlike other multimethod, multitrait analyses, however, the use of twin data allows examination of the effect of combining raters on genetic-environmental parameter estimates.

## METHODS

### Design

The Virginia Twin Study of Adolescent Behavioral Development (VTSABD) is a cohort-sequential study of twins aged 8-16 years (Hewitt *et al.*, 1995). The study aims to examine the role of genetic and environmental influences on the development and maintenance of behavioral and emotional problems. For this reason, measures of psychopathology and the normal variation associated with psychopathology are assessed, along with potential risk factors for child psychiatric disorders. Two types of behavioral measures were used in the study. The Child and Adolescent Psychiatric Assessment (CAPA; Angold *et al.*, 1995; Angold and Costello, 1995) was used to determine the presence of child psychiatric disorders. CAPA is a semistructured, investigator-based instrument which is designed to make DSM-III-R [American Psychiatric Association (APA), 1987] and ICD-10 research diagnoses [World Health Organization (WHO), 1993], as well as to collect information on impairment. For any potential symptoms that are reported, CAPA requires a behavioral description from which the interviewer then decides whether the symptom meets criteria as defined in the interview. CAPA has similar versions for child and parent report. Lay interviewers were trained in its use by a child psychiatrist (E.S.) or a clinical psychologist (J.S.) during an intensive 3-week training course followed by regular follow-up training sessions. In addition, all interviews were reviewed by a team of monitors, headed by a clinically trained psychologist, to ensure that they had been correctly administered and completed. The CAPA was administered to each twin and to both parents, regarding both twins.

Questionnaire measures were also used both to tap normal variation in the areas of psychopathology covered in the CAPA and to obtain information regarding related factors, such as life events and temperament. As with the CAPA, the questionnaire measures have parallel parent and child versions which each child completed on himself and parents completed on both twins. In children aged 11 or less, the questionnaires were read aloud by the interviewers. In children aged 12 or above, the Slosson (1963) reading test was administered to ensure that they were reading at age-appropriate level. If they were reading at an age level of 12 years or above, they were allowed to complete the fuller battery of questionnaires on their own; otherwise, the questionnaires were completed in the same manner as for the younger children. Parents completed questionnaires on their own, unless interviewers were concerned about their literacy, in which case the interviewers offered to administer them.

Two interviewers went into each home. Information was generally obtained from children first. The interview was administered, followed by the questionnaires. While the interviewers were seeing the children, the parents completed their questionnaires and were interviewed after the twins. Each child was assessed by a different interviewer and subsequently each parent was interviewed by one of the same two interviewers. Parents were instructed to complete the entire set of questionnaires on one twin before moving on to consider the second twin. Similarly, the CAPA was completed fully first on one twin and then the other. The determination as to which interviewer assessed each child and parent was made randomly, as was the order in which the parents were asked about the two twins.

### Subjects

The VTSABD sample comprises an epidemiological population of Caucasian twins aged 8 to 16, ascertained through the Virginia public and private school systems during the school year 1986/1987 and 1987/1988, respectively. From this sample, 1900 families were contacted and partial or complete information was obtained from 1412, for a participation rate of 74%.

## Zygosity

Zygosity of same-sex twin pairs was determined by blood typing, when available, or by a combination of parental questionnaire and resemblance in photographs. Parents were asked separately three questions regarding zygosity: whether their twins were mistaken by strangers frequently, occasionally or never, whether the twins were "as alike as two peas in a pod" or "only of family likeness," and whether they believed the twin to be identical or fraternal. These measures were validated against the 105 pairs in whom blood typing had been performed. From this, a tentative questionnaire zygosity was determined from each rater. In addition, photographs were judged independently by two raters, who assigned a zygosity from these. In cases where the judges agreed on zygosity, a picture zygosity was also given (Maes *et al.*, 1995). For these analyses, only twin pairs in which parent ratings agreed with the photograph zygosity and/or where blood typing was available were used.

## Selection of Subjects

In the analyses presented here, same-sex male twins only were used. As the purpose of the current analyses is to examine the effects of combining different raters, only cases where nearly complete data were available from all three sources (child, mother, and father) and on both twins were used. Among the same-sex male twins, there were 312 cases in which the data from fathers were incomplete, compared to 110 cases for mothers and 55 for children, leaving 629 individuals. The majority of missing father data reflected instances where the father was no longer living with the family. Such selection has implications for the nature of the cases analyzed. Comparison of families in which the father did and did not participate revealed significantly higher mean scores on mother CAPA measures in those families with missing fathers but no significant differences for mother questionnaire or child CAPA or questionnaire scores. Similar comparisons for cases in which mothers or children had not participated showed no significant differences. For genetic analyses, a total of 169 MZ and 113 DZ twin pairs was used.

## Measures

Aggressive and antisocial behavior was examined. The dimensional, questionnaire measure used to tap this area is that designed by Olweus (1980), with minor modifications. Each of the 35 items in the questionnaire was coded 0 if the respondent had endorsed "no, never," 1 if the response was "yes, but not in the last 3 months," and 2 when the response was "yes, in the last 3 months." Items were summed to provide a total score, with prorating of missing items when five or fewer items were missing. The scores were transformed using a logarithmic function, prior to obtaining covariances. Pearson correlations for twin 1 $\times$ twin 2 are given in Table I, along with untransformed means and standard deviations.

The interview measures used the CAPA sections on oppositional and conduct disorder. The section consists of 20 behaviors, which are used in the diagnosis of oppositional or conduct disorder according to DSM-III-R or ICD-10 criteria. All information pertained to behavior over the last 3 months. For each symptom, an algorithm, incorporating the severity and frequency of the behavior, determined whether a threshold rendering the behavior symptomatic had been passed. Each item was coded 0 if below threshold and 1 if above. A score of 1 indicates that the behavior has reached a sufficient severity to be considered a symptom potentially related to disorder. The items were summed to give each individual a score. As with the questionnaire data, logarithmic transformations were used prior to calculating covariances. Pearson correlations, means, and standard deviations are shown in Table II.

## Statistics

SAS (1988) was used to obtain the descriptive statistics and the correlations and covariances. All structural equation modeling was performed using the statistical package EQS (Bentler, 1989). Modeling employed covariance structures or raw data, using the maximum-likelihood method.

**Table I.** Twin Correlations for Questionaire Scores[a]

| | Twin 1 | | | Twin 2 | | | | |
|---|---|---|---|---|---|---|---|---|
| MZ twins | Mother | Father | Child | Mother | Father | Child | Mean | SD |
| Twin 1 | | | | | | | | |
| Mother | 1.00 | | | | | | 3.68 | 4.04 |
| Father | 0.62 | 1.00 | | | | | 3.13 | 3.70 |
| Child | 0.36 | 0.39 | 1.00 | | | | 4.01 | 4.84 |
| Twin 2 | | | | | | | | |
| Mother | 0.81 | 0.57 | 0.25 | 1.00 | | | 3.64 | 3.73 |
| Father | 0.54 | 0.77 | 0.30 | 0.56 | 1.00 | | 2.91 | 3.23 |
| Child | 0.45 | 0.49 | 0.60 | 0.46 | 0.45 | 1.00 | 5.20 | 6.44 |
| Twin 1 | | | | | | | | |
| Mother | 1.00 | | | | | | 3.13 | 3.72 |
| Father | 0.61 | 1.00 | | | | | 3.47 | 3.36 |
| Child | 0.54 | 0.45 | 1.00 | | | | 4.01 | 4.92 |
| Twin 2 | | | | | | | | |
| Mother | 0.68 | 0.51 | 0.44 | 1.00 | | | 3.11 | 3.61 |
| Father | 0.32 | 0.58 | 0.29 | 0.59 | 1.00 | | 3.14 | 3.49 |
| Child | 0.17 | 0.25 | 0.41 | 0.28 | 0.44 | 1.00 | 4.66 | 4.88 |

[a] Pearson correlations are calculated from log-transformed scores; means and standard deviations are not transformed.

**Table II.** Twin Correlations for CAPA Scores[a]

| | Twin 1 | | | Twin 2 | | | | |
|---|---|---|---|---|---|---|---|---|
| | Mother | Father | Child | Mother | Father | Child | Mean | SD |
| MZ twins | | | | | | | | |
| Twin 1 | | | | | | | | |
| Mother | 1.00 | | | | | | 0.68 | 1.21 |
| Father | 0.35 | 1.00 | | | | | 0.66 | 1.58 |
| Child | 0.31 | 0.27 | 1.00 | | | | 1.17 | 1.73 |
| Twin 2 | | | | | | | | |
| Mother | 0.65 | 0.21 | 0.31 | 1.00 | | | 0.69 | 1.29 |
| Father | 0.33 | 0.75 | 0.26 | 0.25 | 1.00 | | 0.68 | 1.41 |
| Child | 0.25 | 0.29 | 0.42 | 0.17 | 0.30 | 1.00 | 1.08 | 1.78 |
| DZ twins | | | | | | | | |
| Twin 1 | | | | | | | | |
| Mother | 1.00 | | | | | | 0.85 | 1.48 |
| Father | 0.41 | 1.00 | | | | | 0.53 | 1.20 |
| Child | 0.31 | 0.12 | 1.00 | | | | 0.99 | 1.99 |
| Twin 2 | | | | | | | | |
| Mother | 0.37 | 0.26 | 0.13 | 1.00 | | | 0.78 | 1.30 |
| Father | 0.08 | 0.28 | −0.02 | 0.35 | 1.00 | | 0.53 | 1.03 |
| Child | 0.01 | 0.11 | 0.11 | 0.33 | 0.11 | 1.00 | 1.07 | 1.65 |

[a] Pearson correlations are calculated from log-transformed scores; means and SDs have not been transformed.

## MODELS AND RESULTS

### Questionnaires

The data were initially treated as though only a single rater was available for each twin and estimates of genetic, shared environmental, and unique or specific environmental influences were calculated using the standard ACE model, as depicted in Fig. 1. In this model, and subsequent ones, the latent variables are $A$, additive genetic

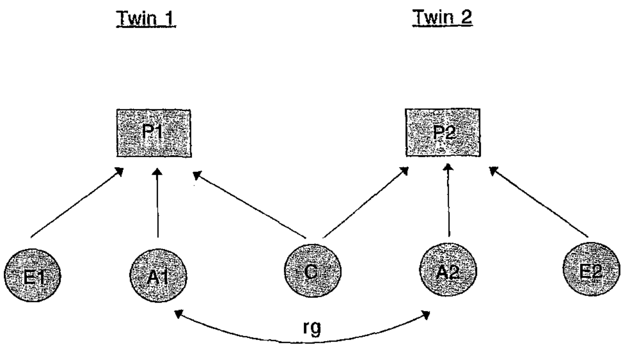


**Fig. 1.** Path diagram for the basic *ACE* model for single raters. P1 and P2 represent the observed phenotype (score) for twin 1 and twin 2, respectively; A1 and A2, the additive genetic component; C, the common environment; E1 and E1, the nonshared environment. A1 and A2 are correlated ($r_g$; rg) 1.0 for MZ and 0.5 for DZ twins.

**Table III.** Single-Rater Estimates for the *ACE* Model for Questionnaire and CAPA Data

| | Parameter estimate[a] | | | Fit | | |
|---|---|---|---|---|---|---|
| Rater | $h^2$ | $c^2$ | $e^2$ | $\chi^2$ | df | *p* |
| Questionnaires | | | | | | |
| Mothers | .23 | .58 | .19 | .43 | 3 | .93 |
| Fathers | .34 | .42 | .24 | 2.90 | 3 | .41 |
| Boys | .34 | .25 | .41 | 1.32 | 3 | .72 |
| CAPA | | | | | | |
| Mothers | .64 | .03 | .33 | 1.55 | 3 | .67 |
| Fathers | .73 | .00 | .27 | 4.16 | 3 | .24 |
| Boys | .40 | .00 | .60 | .99 | 3 | .80 |

[a]$h^2$, additive genetic variance; $c^2$, common environmental variance; $e^2$, nonshared environmental variance.

influence; *C*, common, or shared environment; and *E*, nonshared environment. *P*, the phenotype, is the observed measure. The results for the ratings of questionnaire data are given in Table III. The fit of all models is good. Because the aim of the single rater models was to obtain comparative genetic-environmental estimates, further, more parsimonious models were not fitted. In each cases, about one-quarter of the variance is additive genetic in origin, between one-quarter and one-half common environmental, and between one-fifth and two-fifths nonshared environmental. Mothers' and fathers' ratings show more shared environment, while boys' ratings have relatively more nonshared environment.

A *multiple-rater model* (Fig. 2) was then constructed in which the three observed measures (mother, father, and child ratings) for each twin were indicators of the latent phenotype (LP1 and LP2 for twin 1 and 2, respectively). The genetic and environmental influences loaded onto the latent phenotype. This model allows variation in the contribution of the different raters to the latent phenotype and the degree of error with which each of the three ratings is made. The model is similar to the common pathway model (Kendler *et al.*, 1987; Neale and Cardon, 1992) but with different raters replacing different traits or symptoms.

The fit and parameter estimates of this model are given in the first rows of Tables IV and V, respectively. While little attention should be paid to the parameter estimates because of the poor fit ($\chi^2_{34}$ = 186.50, $p < .001$), it can be seen that relatively little weight is given to the child's account and that this rating has a correspondingly high error variance.

The simple multiple-rater model treats all six ratings of each twin as though made by separate individuals. However, while the twins each rate only his own behavior, each parent rates both twins. To the extent that these ratings are influenced by qualities of the parent as well as by the twins' behavior, these ratings will be influenced in a way that affects scores of both twins. For example, if a parent has high expectations of children's behavior, he may rate both his twins' behavior as more deviant than a parent with low expectations. This would lead to a shift in the level of rating on both twins and a higher correlation between ratings. This is the situation depicted in the correlated errors model (Fig. 3). The model differs from the simple latent phenotype model by introducing two further latent variables, $V_m$ and $V_f$, which represent the view that mothers and fathers bring to their ratings. The paths from the observed to the latent variables are allowed to differ for mothers and fathers but are constrained to be the same for parents of MZ and DZ twins. A similar model, described as the psychometric model, has been described previously by Hewitt *et al.* (1992). The fit of this model is better than that of the simple model as demonstrated by the significance of the likelihood-ratio chi-square (LR $\chi^2_2$ = 116.46, $p < .0001$). However, the overall fit of this model is still poor, and once again, caution should be exercised interpreting parameter estimates. Compared

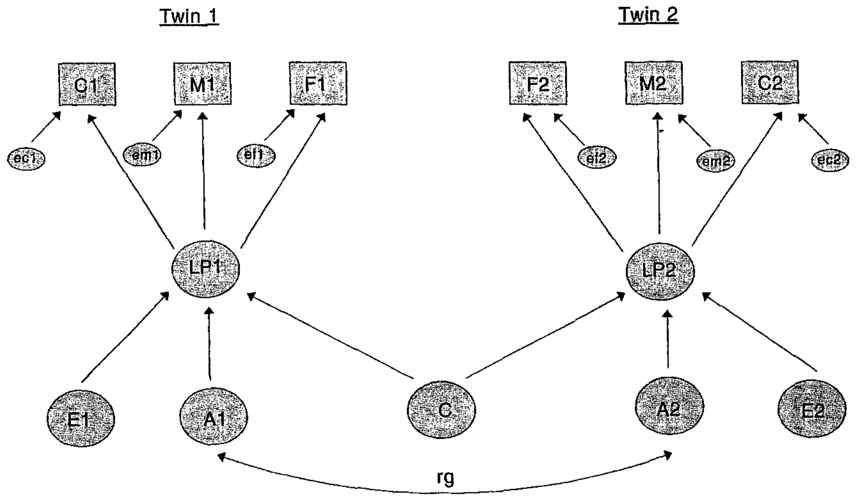


**Fig. 2.** Multiple-rater model. Notation as for Fig. 1. C1 and C2 represent the child ratings, M1 and M2 the mother ratings, and F1 and F2 the father ratings for twin 1 and twin 2, respectively. Each rating has its own error, which is constrained equal across twin pairs.

**Table IV.** Fit of Joint Rating Models of Questionnaire

| Model | $\chi^2$ | df | $p$ | LR $\chi^2$ | df | $p$ |
|---|---|---|---|---|---|---|
| 1. Simple multiple rater | 186.50 | 34 | <.0001 | — | — | — |
| 2. Correlated errors | 70.04 | 32 | <.0001 | 116.46 | 2 | <.0001 |
| 3. Shared parental view, $r_v$ MZ = DZ[a] | 39.19 | 31 | .148 | 30.85 | 1 | <.0001 |
| 4. Shared parental view, $r_v$ MZ≠DZ[a] | 38.41 | 30 | .140 | .78 | 1 | .377 |
| 5. Shared parental view, $r_v$ MZ≠DZ, $V_m$ MZ≠DZ[a] | 37.49 | 29 | .134 | .32 | 1 | .572 |
| 6. Shared parental view, $r_v$ MZ≠DZ, $V_m$, $V_f$ MZ≠DZ[a] | 37.44 | 28 | .109 | .02 | 1 | .888 |

[a] See Fig. 3 for parameters.

**Table V.** Parameter Estimates of Questionnaire Joint Ratings[a]

| | Genetic estimated | | | Latent phenotype | | | Error | | | Parental view | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | $h^2$ | $c^2$ | $e^2$ | M | F | C | M | F | C | $V_m$ | $V_f$ | $r_v$ |
| 1. Simple multiple rater | .30 | .70 | 0 | .69 | .57 | .27 | .31 | .43 | .73 | — | — | — |
| 2. Correlated errors | .50 | .44 | .06 | .52 | .57 | .38 | .16 | .20 | .62 | .31 | .23 | — |
| 3. Shared parental view | .63 | .29 | .08 | .30 | .34 | .61 | .17 | .21 | .39 | .53 | .45 | .56 |

[a] $h^2$, $c^2$, and $e^2$ refer to additive genetic, common environmental, and nonshared environmental variances, respectively. Latent phenotype refers to squared standardized path coefficients from the latent phenotype to the observed measures; error, to the error on the observed measures; and parental view, to the squared standardized path coefficients from the parental view to the observed measures. M, F, and C refer to mother, father, and child parameters, respectively. Number of model refers to models as numbered in Table IV. Genetic estimates sum to unity, as do the other estimates for an individual rater, e.g., latent phenotype, error, and parental view.

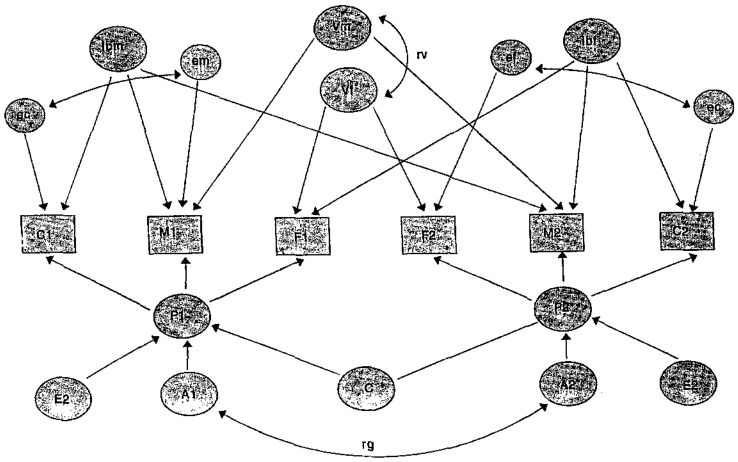


**Fig. 3.** Correlated errors/shared parental view model. Notation is the same as for previous models. Vm and Vf refer to the mother's and father's view of child behavior, respectively; rv, to the correlation between the parental views (present only for models incorporating a shared parental view). For CAPA models only, Ibm and Ibf refer to the interviewer bias for mothers and fathers, respectively.

to the simple multiple-rater model, a greater proportion of the variance relating to the phenotype is due to additive genetic factors and less to shared environment. This can be understood in terms of a portion of the parental intertwin correlations now being explained by the parental view method factor. Additionally, relatively more emphasis is placed on the child's rating in this model, as demonstrated by the rater parameters, which indicate the weight given to each of the three raters. While the correlated errors model takes into account the fact that the parental ratings for each twin are not independent, it assumes that the ratings made by mothers are correlated with those made by fathers only because they are measuring the same latent phenotype. As highlighted earlier, there is substantial evidence to suggest that parents may have shared expectations and/or ratings, although the mechanisms underlying this are unclear. The correlations in Table II reveal greater agreement between mothers and fathers than between either parent and the child. To examine the extent to which the parental views are associated, a *shared parental view model*, in which $r_v$ represents the correlation ($r_v$) between the parental views, $V_m$ and $V_f$, was tested (see Fig. 3). The model (3 in Tables IV and V) follows directly from the correlated errors model; the likelihood-ratio chi-square test indicates a significant improvement in fit ($LR\chi^2_1 = 30.85$, $p < .0001$), and the overall chi-square indicates a good fit to the data ($\chi^2_{31} = 39.19$, $p = .148$). Inspection of the parameter estimates show a further increase in the relative weighting of the three raters toward the child rating, along with a corresponding reduction in the error associated with the child report. The loadings of both parental ratings on the latent phenotype are further reduced, but with the addition of the correlation, $r_v$, between parental views, the loadings on both parental views are increased.

Further elaborations of the shared parental view model were fitted, to test the extent to which parental correlated errors and shared views are the same across zygosity. If the assumptions of the twin method are correct, there should be no zygosity difference in either the extent to which parents are employing a set of expectations in the rating of their twins or the extent to which they agree in this view. This was tested in models 4 to 6 (Table IV), where first the correlation, $r_v$, between parental views was relaxed across zygosity (model 4). The mother view was then allowed to differ (model 5), and then the father view (model 6). In no case did the relaxation of parameters improve the fit significantly.

## CAPA

A similar set of analyses was carried out with CAPA interview data. While questionnaire and interview measures are broadly aimed at examining the same behaviors, there are three important dif-

**Table VI.** Fit of Joint Rating Models of CAPA

| Model | $\chi^2$ | df | *p* | LR $\chi^2$ | df | *p* |
|---|---|---|---|---|---|---|
| 1. Simple multiple rater | 150.85 | 34 | <.0001 | | | |
| 2. Correlated errors | 48.67 | 32 | .030 | 102.18 | 2 | <.0001 |
| 3. Shared parental view, $r_v$ MZ=DZ[a] | 47.99 | 31 | .026 | .68 | 1 | .410 |
| 4. Shared parental view, $r_v$ MZ≠DZ[a] | 46.80 | 30 | .026 | 1.19 | 1 | .273 |
| 5. Shared parental view, $r_v$ MZ≠, $V_m$ MZ≠DZ[a] | 45.92 | 29 | .024 | .88 | 1 | .348 |
| 6. Shared parental view, $r_v$ MZ≠DZ, $V_m$, $V_f$ MZ≠DZ[a] | 37.94 | 28 | .100 | 7.08 | 1 | .008 |
| 7. Interviewer bias per model 6 | 37.49 | 27 | .086 | .45 | 1 | .502 |
| 8. Contamination per model 7 | 37.49 | 26 | .068 | .00 | 1 | 1.00 |

[a] See Fig. 3 for parameters.

ferences. CAPA obtains information in person, with the interviewer ensuring through follow-up questions that the behavior described by the respondent is, in fact, what was meant by the symptom. CAPA places relatively greater weight on oppositional symptoms, which represent a style of behavior, as opposed to conduct symptoms, which are generally tied to discrete acts of antisocial and aggressive behavior and may therefore be less judgmental. Finally, CAPA symptoms were derived by applying thresholds of frequency as well as severity to individual behaviors; endorsement of CAPA symptoms represents a more extreme degree of deviance, with each endorsement being at a severity level seen in frank disorder. Examination of the twin correlations (Table II) shows lower inter-rater agreement for CAPA, which is due in part to the more discrete scores. All these factors make it likely that there will be some differences not only in the genetic-environmental estimates but also in the fit of the different latent phenotype models.

The results of the individual rater models are presented in Table III. In comparison to the questionnaire estimates, there are two striking differences. The first is the lack of agreement between parents and children on the role of genetic influences. Parental ratings give considerably more genetic variance, as do boys (64-73 compared to 40%). Furthermore, for all three raters, the remaining variance is almost completely explained by nonshared environment, with common environment accounting for no more than 3% of the variance.

The simple multiple-rater model was fitted in the same manner as for the questionnaire data. As with the questionnaire data, it also gave a poor fit, as shown in Table VI. The correlated errors model, while still fitting poorly, led to a significant improvement in the fit (LR $\chi^2_2 = 102.18, p < .0001$). Inspection of the parameter estimates as given in Table VII shows that the shift in rater loadings from the simple multiple-rater to the correlated errors model includes both a reduction in the mother rater loading and an increase in the child rater loading, as seen with the questionnaire data. The shared parental view model was then fitted; this, by itself, did not lead to a significant improvement in fit (LR $\chi^2_1 = .68, p = .410$). However, elaboration of the model allowing the parental views to differ for MZ and DZ pairs (while requiring that the total phenotype variance remained the same across zygosity) did improve the fit. The improvement in fit upon releasing the constraint on the maternal view was not significant (LR $\chi^2_1$ .88, $p = .348$) but substantial for the father view (LR $\chi^2_1 = 7.08, p = .008$). The results indicate that the strength of the father view is greater for MZ than DZ twins (see model 6 in Table VII). This pattern can be seen in the correlations (Table II), where the cross-rater correlations for MZ twins are of the same magnitude for ratings of the cotwin as for the same twin when father ratings are considered. For example, looking at the MZ cross-rater correlations between father on twin 1, we see that agreement is slightly higher with child on twin 2 (0.29) than with child on twin 1 (0.27). In the mother-father cross-twin ratings, the correlation between mother on twin 1 and father on twin 1 (0.35) is of the same magnitude as that for mother on twin 1 and father on

**Table VII.** Parameter Estimates of CPA Joint Ratings[a]

| | Genetic estimate | | | Latent phenotype | | | Error | | | Parental view | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | $h^2$ | $c^2$ | $e^2$ | M | F | C | M | F | C | $V_m$ | $V_f$ | $r_v$ |
| 1. Simple multiple rater | 1.00 | 0 | 0 | .19 | .69 | .11 | .81 | .32 | .90 | — | — | — |
| 2. Correlated errors | 1.00 | 0 | 0 | .36 | .31 | .26 | .34 | .30 | .74 | .30 | .39 | — |
| 6. Shared parental view | 1.00 | 0 | 0 | .26 | .18 | .37 | .32 | .27 | .63 | .41 | .55 | .14 (MZ) |
| | | | | | | | .44 | .52 | .63 | .30 | .30 | .44 (DZ) |

[a] $h^2$, $c^2$, and $e^2$ are additive genetic, common environmental, and nonshared environmental loadings, respectively. Latent phenotype refers to squared standardized path coefficients from the latent phenotype to the observed measures; error, to the error on the observed measures; and parental view, to the squared standardized path coefficients from the parental view to the observed measures. M, F, and C refer to mother, father, and child parameters, respectively. Model numbers associated with models as given in Table VI. Genetic estimates sum to unity, as do other estimates for an individual rater, e.g., latent phenotype, error, and parental view.

twin 2 (0.33). This is different from the mother-child cross-twin correlations, where mother on twin 1 and child on twin 1 (0.31) is larger than the correlation between mother on twin 1 and child on twin 1 (0.25). This is also quite different from the pattern in DZ twins, where all three pairs of cross-rater correlations go in the predicted direction, i.e., with the correct twin.

For CAPA models, it was also important to look at the role of the interviewer in determining the ratings. As described under Methods, a pair of interviewers went into the home, each interviewing one twin and then one parent on *both* twins. There are two possible effects of such a protocol. The first is a straightforward bias on the part of the interviewer in determining whether or not symptoms are rated. This could act in a manner similar to the maternal and paternal view, with the interviewers applying their own expectations of normal and abnormal behavior to the ratings. To the extent that interviewers differ in these views, both the cross-twin child and the cross-parent within-twin correlation would be reduced, because the twins themselves were interviewed by different individuals, as was each parent. However, within-parent, cross-twin correlations would be inflated, as these ratings were made by the same interviewer. While CAPA makes every attempt to reduce such bias by thorough training of interviewers and the use of clearly defined coding rules, the presence of such an effect was still possible. Such as *interviewer bias model* was tested by inserting parameters ($I_{bm}$ and $I_{bf}$ in Fig. 3) connecting the three ratings made by the same interviewer. The data were sorted in advance so that twin 1 was assessed by the same interviewer as the mother, and twin 2 by the same interviewer as the father. In the correlations (Table II), this would be revealed as a greater correlation between boys' and mothers' ratings of twin 1 compared to boys' and mothers' ratings of twin 2, and vice versa for fathers. This pattern can be observed in the correlations presented in Table II. Hence, for MZ twins, the correlation between the mothers' and the boys' ratings is 0.31 for twin 1 and 0.17 for twin 2, while the correlation between fathers' and boys' ratings is higher for twin 2 (0.30) than for twin 1 (0.27). The pattern is not seen for DZ pairs, however. Model-fitting was used to test whether the effect was significant. Initially the interviewer effects were constrained to be the same for fathers and mothers (as the interviewers were allocated randomly) and for MZ and DZ twins. The additional parameter did not lead to a significant improvement in the fit (model 7, Table VI; LR $\chi^2_1$ = .45, $p$ = .502). Subsequent release of the constraints on $I_{bm}$ and $I_{bf}$ across parents and zygosity (not shown) also failed to significantly improve the fit.

The other way in which interviewer bias could act is by influencing the parental ratings of the twin who has already been assessed by the same interviewer. While the child and parent interviews are considered confidential and the interviewers are instructed to use only information given by the respondent in their ratings, the interview is semistructured and does leave to the interviewer's discretion the decision on when to probe further. Thus, there was a possibility that there would be greater concordance in the parent-child ratings when assessed by the same interviewer. This *con-*

*tamination model* is represented in Fig. 3 by the correlation between the error terms of the individuals who were assessed by the same interviewer, as well as the presence of the interviewer bias parameters. Compared to the interviewer bias model, the contamination model predicts an even greater correlation between the child 1 and the mother 1 and between the child 2 and the father 2 correlations. The contamination model was fitted with constraints across parents and zygosity for the reasons given above. The results (model 8 in Table VI) also did not lead to a significant improvement in fit, nor did subsequent models releasing the constraints for the correlated errors across parents and zygosity (not shown).

## Models Combining Questionnaire and CAPA Ratings

So far, only models examining questionnaire and CAPA ratings separately have been considered. Although there are issues specific to each measure and the results of model-fitting are somewhat different, both are designed to tap the same underlying trait. Hence we examined two models combining both multiple raters and different instruments, which can be seen as variants of the multimethod, multitrait models (Campbell and Fiske, 1959). The *combined latent phenotype model*, depicted in Fig. 4a, has a general latent phenotype (LP1 and LP2), upon which the questionnaire (Q1, Q2) and CAPA (CA1, CA2) latent phenotypes load. The questionnaire and CAPA latent phenotypes are specified according to the previous model providing best fit. Hence the questionnaire latent phenotype included a shared parental view constrained across zygosity, while the CAPA latent phenotype allowed the shared parental view to differ for MZ and DZ pairs. Interviewer bias and contamination were not included in the model, as they had not significantly improved the fit of the CAPA models. In addition to the correlations across parents for the same rating ($r_{vq}$ and $r_{vc}$), there were also correlations within parent across ratings ($r_{vm}$ and $r_{vf}$). The genetic and environmental parameters were estimated on the general latent phenotype. The model fit is acceptable ($\chi^2$ = 219.01, df 126, CFI = .934). Table VIII shows the parameter estimates. The questionnaire and CAPA ratings load roughly equally on the combined latent phenotype and the genetic-environmental estimates are intermediate between the two traits. The correlations between shared parental views were greater for questionnaires (constrained across zygosity) and higher for DZ than MZ twins on CAPA, consistent with previous models. The correlations across measures were slightly smaller, although this difference should not be overinterpreted.

The second joint model, the *combined multivariate model* depicted in Fig. 4b, links the questionnaire (Q1 and Q2) and the CAPA (CA1 and CA2) latent phenotypes through shared genes, common environment, and unique environment. In addition, each measure has genetic, common environmental, and unique environmental components that are specific, i.e., not shared with the latent phenotype that is tapped by all the measures. As with the combined latent phenotype model, shared parental views were measured in the manner that best fitted the questionnaire and CAPA models. The overall fit of the model was acceptable ($\chi^3$ = 191.82, df 125, CF1 = .954). The parameter estimates revealed that the questionnaire and CAPA measures were linked almost exclusively by additive genetic variance and that all of the genetic variance for questionnaires was shared with CAPA. CAPA had further additive genetic variance through its specific genetic factor but no additional common or nonshared environmental loading. The questionnaire-specific loadings, on the other hand, showed some common environment and a small amount of nonshared environment. The overall decomposition of genetic-environmental variance for questionnaire and CAPA measures was consistent with that reported for the individual measures; this model adds further information by suggesting that the link between the two measures arose through shared genes. The combined latent phenotype and combined multivariate model cannot be compared directly as they are not nested. While the multivariate model gave a slightly better fit, the differences are small and it is not clear that one conceptualization of the link between the two measures is clearly better than the others.

## DISCUSSION

The models described explore different ways of combining data from different raters. Despite the generality of the issue in social sciences measurement, there is no consensus on which ratings to use or how to combine multiple ratings. While some

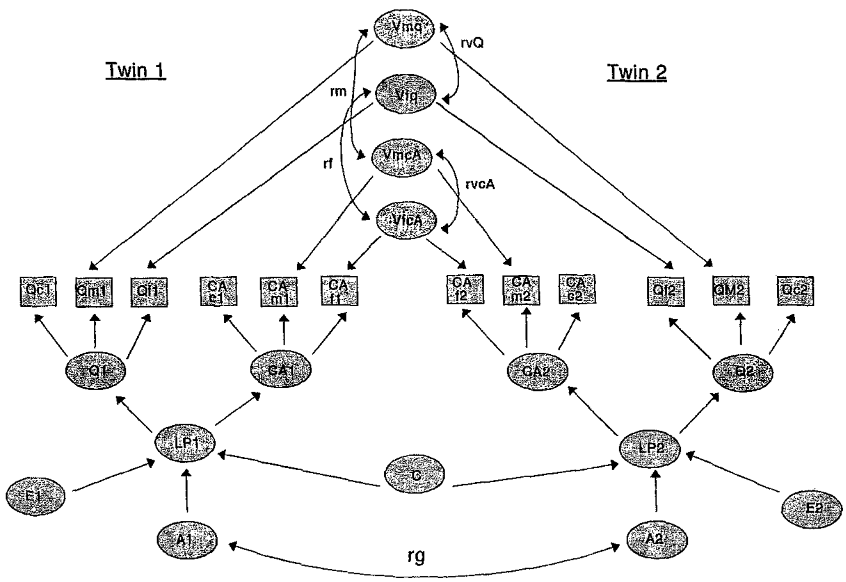


**Fig. 4a.** Combined questionnaire/CAPA models: combined latent phenotype. Notation as for previous models. Q1 and Q2 represent the latent questionnaire measures, CA1 and CA2 the CAPA latent measures, and LP1 and LP2 the combined latent phenotypes for twins 1 and 2, respectively. Correlations rvQ and rvcA are shared parental views for questionnaire and CAPA measures; rm and rf are maternal and paternal correlations in correlated errors across questionnaire and CAPA measures. Errors directly loading on the observed measure were included but are not depicted in the path diagram.

**Table VIII.** Parameter Estimates of Combined Questionnaire/CAPA Models[a]

| | | | | Latent phenotype loading | | Latent phenotype error | | Parental view correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | $h^2$ | $c^2$ | $e^2$ | Q | CA | Q | CA | rvq | rvcA | rvm | rvf | |
| (a) Latent phenotype | .81 | .02 | .17 | .96 | .79 | .04 | .21 | .51 | .23 | .20 | .17 | (MZ) |
| | | | | | | | | .51 | .46 | .15 | .23 | (DZ) |

| | Shared loading | | | Questionnaire specific loading | | | CAPA specific loading | | | Parental view correlation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $h^2$ | $c^2$ | $e^2$ | $h^2$ | $c^2$ | $e^2$ | $h^2$ | $c^2$ | $e^2$ | rvq | rvca | rvm | rvf | |
| (b) Multivariate | .70 | 0 | .02 | 0 | .22 | .06 | .28 | 0 | 0 | .48 | .09 | .21 | .21 | (MZ) |
| | | | | | | | | | | .48 | .16 | .23 | .26 | (DZ) |

[a] $h^2$, $c^2$, and $e^2$ refer to additive genetic, common environmental, and nonshared environmental loadings, respectively. Q refers to questionnaire; CA, to CAPA. rvq refers to the mother–father correlation on questionnaire measures; rvca, to the mother–father agreement on CAPA measures; rvm, to the questionnaire–CAPA correlation on maternal ratings; and rvf, to the questionnaire-CAPA correlation on paternal ratings.

researchers have advocated rules for deciding which rater receives precedence or a priori methods for combining scores, such techniques leave unexamined the reasons for and effects of agreement and disagreement. One of the strengths of combining the data through structural equation modeling is that it allows examination of the processes involved in agreement between raters. While for some analyses, determination of the partitioning of genetic and environmental variance components may be the primary aim of the analyses, genetic data can also be used as they have been here, to

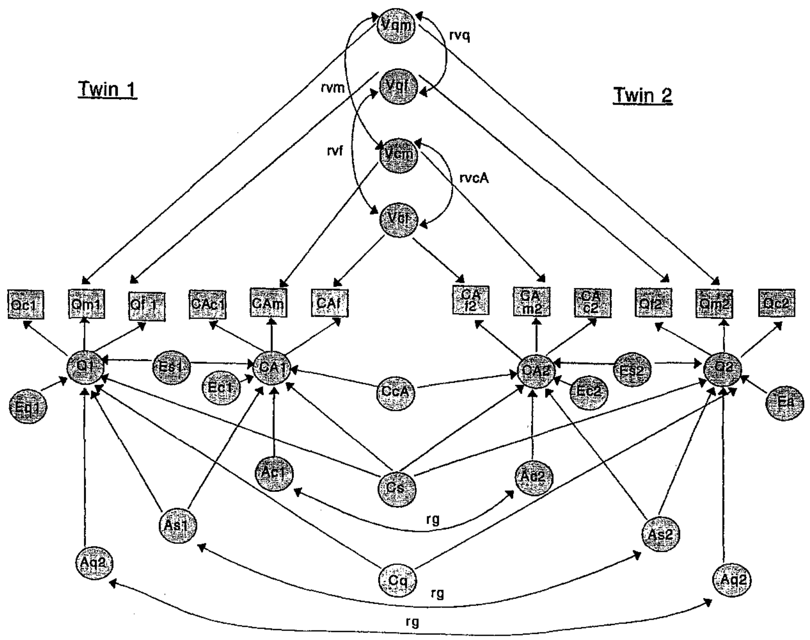


**Fig. 4b.** Combined questionnaire/CAPA models: combined multivariate model. Notation as for previous models. The latent phenotypes for questionnaire (Q1 and Q2) and CAPA (CA1 and CA2) measures are comprised of specific additive genetic (Aq, Ac), common environmental (Co, CcA), and nonshared (Eo, Ec) components and ones shared across questionnaire and CAPA phenotypes (Am, Cs, Ex). Errors on observed measures were estimated but are not depicted in the diagram.

allow exploration of the processes involved in the differences between raters.

Consistent across both questionnaire and CAPA measures is the finding of a shared parental view. This is in accord with most other data showing higher mother-father than parent-child agreement for behavioral ratings (Achenbach *et al.*, 1987). What is less clear is what underlies the higher parental agreement. While it is described here in general terms as due to a shared set of expectations against which both twins are rated, at least three other possible explanations are possible. The first is that the greater parental agreement represents situational specificity in the child's behavior. Children may behave in a different manner when they are with their parents. Parents can rate only those aspects of their children's behavior of which they are aware; particularly with regard to conduct disorder, children may be engaged in a variety of behaviors about which they do not tell their parents. Without such information, parents will be rating primarily behavior that occurs in the home and that may be more likely to involve both children than behavior in other settings, where each child may be on his own or where different adults for each twin may be involved. As parents are likely to share their information regarding their children's behavior, this situational specificity could lead to greater interparental agreement.

Another possibility is that correlated parental views represent *bias* (Hewitt *et al.*, 1992; Neale and Stevenson, 1989), that is, consistent distortion in the rating of their children's behavior for some reason. While Neale and Stevenson (1989) used parental self-ratings compared to ratings of their children, as a way of assessing bias, the current analyses are based on differences between parent and child ratings. Bias could occur for a variety of reasons, as discussed previously. The fact that the model-fitting for both questionnaires and CAPA indicates a significant component is related to a shared parental view suggests that individual parental factors are not the only cause. The third possibility is that assortative mating for views and expectations of child behavior has occurred in the parents. It should be borne in mind that whatever processes are operating apply within the general population.

For the questionnaire ratings, it is difficult to distinguish between bias in the increased agreement of parental ratings and other reasons for a parental view. In the CAPA analyses, however, the fact that the father view is significantly stronger for MZ than DZ twins suggests that bias may play a role in the higher twin correlations, at least from fathers' rating. As discussed above, the cross-rater cross-twin correlations for MZ twins provide little discrimination between the two twins in the fathers' ratings. One possible interpretation is that fathers are rating their MZ twins according to a particular view of their behavior and do not distinguish the differences reported by the children and, to a lesser extent, by the mothers. This interpretation is strengthened by the fact that the father MZ twin correlation (0.75) is the largest of the three raters. The fact that this occurs only for MZ twins suggests that the process is at least partly one of bias. It is not clear whether the phenomenon relates to expectations of MZ twins being very similar or whether it reflects difficulty in discrimination. One way to examine the effect would be to see whether it occurred in twins whose parents had mistaken their zygosity. If the effect is due to altered expectations of MZ twins, the phenomenon will go with perceived rather than true zygosity.

It is unclear why the effect should be seen only in CAPA. Both questionnaires and CAPA were completed in their entirety for one twin first and then the other. However, in answering questions about one twin, many parents compare and contrast to the other twin. While the interviewers would not proceed with coding the cotwin during the assessment of the first twin, the information remembered from the first twin might be used in questioning during assessment of the second twin.

Another possible explanation of the parental view is that the intra- and interparental correlations are a true reflection of the similarity of the twins and that there is a contrast effect in the twins' self-ratings. This seems less plausible for CAPA than questionnaire measures, because of the MZ-DZ difference in paternal view. Without another, independent rater of the twins' behavior, it is not possible to determine whether twin contrast or parental view (or both) is operating. While teacher reports might be helpful, they are rating children's behavior in a different situation and are therefore unlikely to resolve the situation. The CAPA models, in showing no significant effect of either interviewer bias or contamination, support the research design and the robustness of CAPA to possible confounding effects.

In comparing single to multiple-rater models, there is a shift in the genetic-environmental parameter estimates for both questionnaire and CAPA measures. In both, nonshared environment accounts for a substantial part of the variance (one-fifth to one-half) in all single-rater measures. In going to a simple multiple-rater model, nonshared environment accounts for none of the variance. Shared parental view models give a very small amount of nonshared environment for questionnaires and none for CAPA. At the same time, genetic estimates increase in multiple-rater models. The reduction in nonshared environment alone could be accounted for by the use of a latent phenotype. With an observed phenotype, nonshared environment includes all measurement error (unless this is dealt with separately), but with a latent phenotype much of this error is taken up elsewhere. However, the reduction in shared environment was less expected. For both questionnaire and CAPA, genetic influences become the most important. It should be borne in mind that the high genetic loadings on multiple-rater models indicate the importance of genetic influences on those aspects of conduct disorders that are common to these different raters. Other types of conduct disorder, possibly specific behaviors and ones that occur in other settings, such as school, may be less genetically mediated.

There have been relatively few studies with large enough numbers of individuals to draw firm conclusions regarding the role of genetic influences on conduct disorder. McGuffin and Gottesman (1985) showed high twin concordance rates for both MZ and DZ twins for juvenile delinquency and argued for the importance of shared environmental effects. However, environmental effects could influence being apprehended and prosecuted as much as committing antisocial behavior. In a moderate-sized general population sample of 13-year-old twins, rated on the Rutter questionnaires (Rutter *et al.*, 1970), Stevenson and Graham (1988) found genetic influences ($h^2$ = .42) for boys only. Previous analyses on a subsample from VTSABD, using the externalizing scale of the Achenbach CBC-L (Achenbach and Edelbrock, 1981), allowed for individual parental bias and unreliability and found 31-40% of the variance for boys' scores to be additive genetic in origin. Hence, the results from individual raters presented here are consistent with previous reports. The findings from the joint rater latent phenotype models suggest that there may be types of conduct disorder that are highly genetic. In this connection, it should be recalled that these analyses use a total score, rather than subdividing conduct disorder, as has been advocated by other researchers (Lahey and Loeber, 1994). Genetic analyses examining subgroups of conduct disorder, classified either on an a priori basis or according to latent classes, used the antisocial Rutter A scale items in the VTSABD sample (Eaves *et al.*, 1993). The results suggest classes of conduct disorder that are highly genetic.

In summary, the combination of multiple raters of behavior gives importantly different results to genetic analyses compared with single raters. Because such models can take into account potential sources of bias of individual raters, they may give more accurate results. In the case of conduct disorder, the findings indicated that aggressive and antisocial behavior may be more genetic than previously appreciated.

## ACKNOWLEDGMENTS

This work was supported by NIH Grant MH 45268 and the Medical Research Council. Thanks go to Andrew Heath for comments on the manuscript and to Linda Wilkinson for secretarial assistance.

## REFERENCES

Achenbach, T. M., and Edelbrock, C. S. (1981). Behavior problems and competencies reported by parents of normal and disturbed children age four through sixteen. *Monogr. Soc. Res. Child Dev.* **46**:No. 188.

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. *Psychol. Bull.* **101**:213–232.

American Psychiatric Association (1987). *The Diagnostic and Statistical Manual of Mental Disorders (Third Edition, Revised)*, American Psychiatric Association, Washington, DC.

Angold, A., and Costello, E. J. (1995). A test-retest reliability study of child-reported psychiatric symptoms and diagnosis using the Child and Adolescent Psychiatric Assessment (CAPA). *Psychol. Med.* (in press).

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. *J. Child Psychol. Psychiat.* **28**:901–915.

Angold, A., Prendergast, M., Cox, A., Harrington, R., Simonoff, E., and Rutter, M. (1995). The Child and Adolescent Psychiatric Assessment (CAPA) Interview. *Psychol. Med.* (in press).

Bentler, P. M. (1989). *EQS: Structural Equations Program Manual*, BMDP Statistical Software, Los Angeles.

Campbell, D. T., and Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait multimethod matrix. *Psychol. Bull.* **105**:317–327.

Eaves, L. J., Silberg, J. L., Hewitt, J. K., Rutter, M., Meyer, J. M., Neale, M. C., and Pickles, A. (1993). Analyzing twin resemblance in multisymptom data: Genetic applications of a latent class models in symptoms of conduct disorder in juvenile boys. *Behav. Genet.* **22**:5–19.

Edelbrock, C., Costello, A. M., Dulcan, M. K., Conover, N. C., and Kala, R. (1986). Parent-child agreement on child psychiatric symptoms assessed via structured interview. *J. Child Psychol. Psychiat.* **27**:181–190.

Fergusson, D. M., and Horwood, L. J. (1987a). The trait and method components of ratings of conduct disorder. I. Maternal and teacher evaluations of conduct disorder in young children. *J. Child Psychol. Psychiat.* **28**:249–260.

Fergusson, D. M., and Horwood, L. J. (1987b). The trait and method components of ratings of conduct disorder. II. Factors related to the trait component of conduct disorder scores. *J. Child Psychol. Psychiat.* **28**:261–271.

Herjanic, B., and Reich, W. (1982). Development of a structured psychiatric interview for children: Agreement between child and parent on individual symptoms. *J. Abnorm. Child Psychol.* **10**:307–324.

Herjanic, B., Herjanic, M., Brown, F., and Wheatt, T. (1975). Are children reliable reporters? *J. Abnorm. Child Psychol.* **3**:41–48.

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. *Behav. Genet.* **22**:293–318.

Hewitt, J. K., Eaves, L. J., Silberg, J. L., Rutter, M. E. S., Meyer, J. M., Loeber, R., Neale, M. C., Erickson, M., Kendler, K. S., Heath, A. C., Pickles, A., Truett, K. R., and Maes, H. (1995). Genetics and developmental psychopathology: The Virginia Twin Study of Adolescent Behavioral Development (submitted for publication).

Kashini, J. H., Orvaschel, H., Burk, J. P., and Reid, J. C. (1985). Informant variance: The issue of parent-child disagreement. *J. Am. Acad. Child Psychiat.* **24**:437–441.

Kazdin, A. E., Esveldt-Dawson, K., Unis, A. S., and Rancurello, M. D. (1983). Child and parent evaluations of depression and aggression in psychiatric inpatient children. *J. Abnorm. Child Psychol.* **11**:401–413.

Kendler, K. S., Heath, A. C., Martin, N. G., and Eaves, L. J. (1987). Symptoms of anxiety and symptoms of depression. Same genes, different environments? *Arch. Gen. Psychiat.* **44**:451–457.

Lahey, B. B., and Loeber, R. (1994). Framework for a developmental model of oppositional defiant disorder and conduct disorder. In Routh, D. K. (ed.), *Disruptive Behavior Disorders in Childhood*, Plenum, New York, pp. 139–180.

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

Loeber, R., Green, S. M., Lahey, B. B., and Stouthamer-Loeber, M. (1991). Differences and similarities between children, mothers, and teachers as informants on disruptive child behavior. *J. Abnorm. Child Psychol.* **19**:75–95.

Maes, H., Meyer, J., Silberg, J., Hewitt, J., Simonoff, E., Pickles, A., Rutter, M., and Eaves, L. J. (1995). Zygosity determination in the Virginia Twin Study of Adolescent Behavioral Development (in preparation).

McGuffin, P., and Gottesman, I. I. (1985). Genetic influences on normal and abnormal development. In Rutter, M., and

Hersov, L. A. (eds.), *Child and Adolescent Psychiatry: Modern Approaches*, Blackwell Scientific, Oxford, pp. 17–33.

Neale, M. C., and Cardon, L. R. (1992). *Methodology for Genetic Studies of Twins and Families*, Kluver Academic, Dordrecht.

Neale, M. C., and Stevenson, J. (1989). Rater bias in the EASI temperament scales: A twin study. *J. Personal. Soc. Psychol.* **56**:446–455.

Olweus, D. (1980). Familial and temperamental determinants of aggressive behavior in adolescent boys: A causal analysis. *Dev. Psychol.* **16**:644–660.

Orvaschel, H., Puig-Antich, J., Chambers, W., Tabrizi, M. A., and Johnson, R. (1982). Retrospective assessment of prepubertal major depression with Kiddie-SADS-E. *J. Am. Acad. Child Psychiat.* **21**:392–397.

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. *Psychol. Assess.* **1**:68–71.

Reich, W., and Earls, F. (1987). Rules for making psychiatric diagnoses in children on the basis of multiple sources of information: Preliminary strategies. *J. Abnorm. Child Psychol.* **15**:609–616.

Rutter, M., Tizard, J., and Whitmore, K. (1970). *Education, Health and Behavior,* Longman, London.

Silverman, W. K. (1992). Age differences in the reliability of parent and child reports of child anxious symptomatology using a structured interview. *J. Am. Acad. Child Adolesc. Psychiat.* **31**:117–124.

Simonoff, E., Pickles, A., Silberg, J. L., Hewitt, J. K., Loeber, R., Eaves, L. J., Meyer, J. M., and Rutter, M. (1995). Age effects in genetic influences on oppositional and conduct World Health Organization (1993). *The ICD-10 Classification of Mental and Behavioural Disorders. Diagnostic Criteria for Research*, World Health Organization, Geneva.

Stevenson, J., and Graham, P. (1988). Behavioral deviance in 13-year old twins: An item analysis. *J. Am. Acad. Child Adolesc. Psychiat.* **27**(5):791–797.

Edited by N. G. Martin