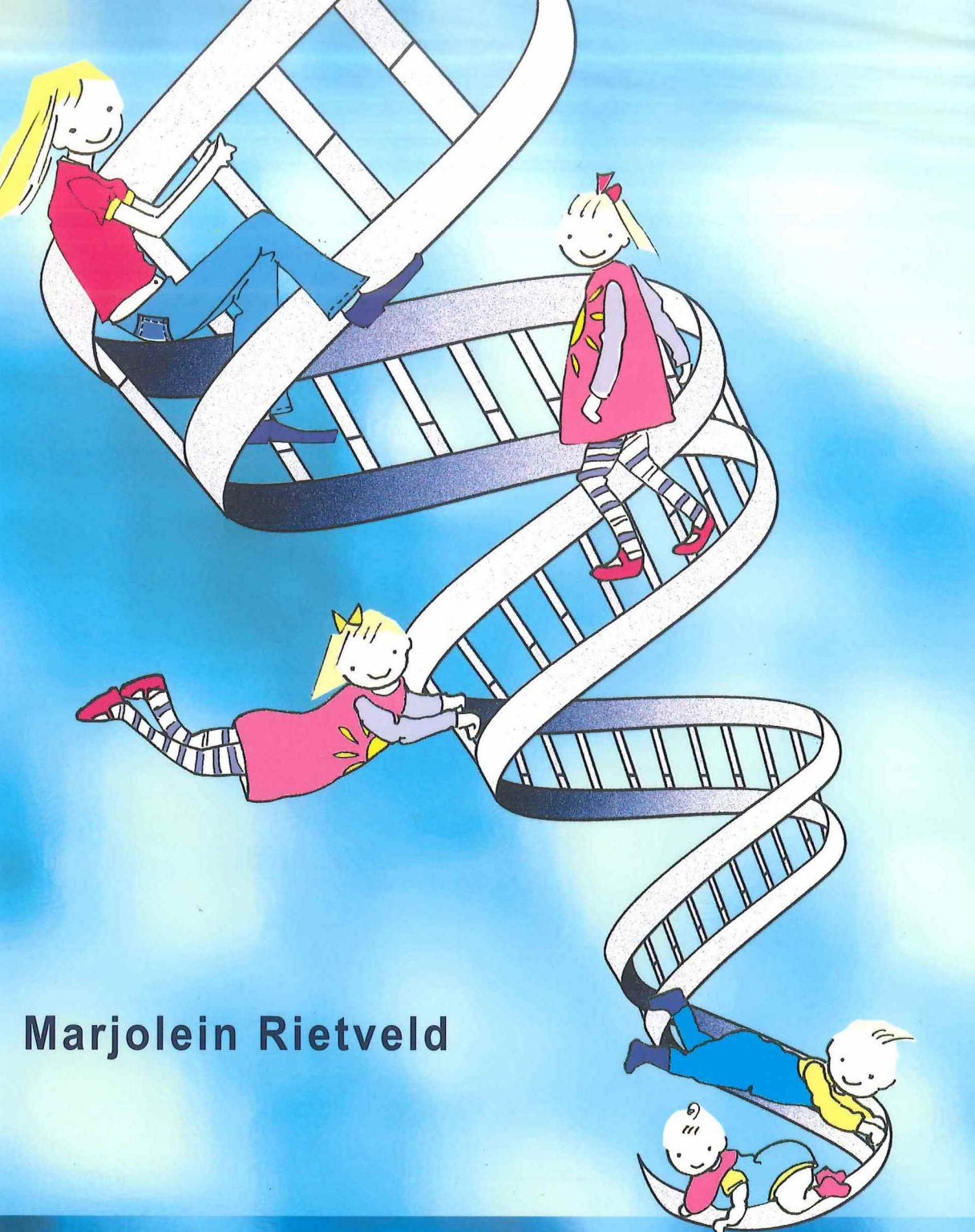


Marjolein Rietveld

Heritability of Cognitive Abilities and of Attention Problems

a longitudinal twin study in childhood

Kids Who Are Different

Here's to the kids who are different,
The kids who don't always get A's,
The kids who have ears twice the size of their peers,
And noses that go on for days...

Here's to the kids who are different,
The kids they call crazy or dumb,
The kids who don't fit, with the guts and the grit,
Who dance to a different drum...

Here's to the kids who are different,
The kids with a mischievous streak,
For when they have grown, as history's shown,
It's their difference that makes them unique.

Digby Wolfe, 1982
UNM/ Theatre & Dance
Albuquerque, NM

This research was financially supported by a grant from the Universitair Stimulerings Fonds (USF #96/22) awarded to prof. dr. Boomsma.

ISBN 90-9016914-8
Printed by [OPTIMA] grafische communicatie, Rotterdam
Cover Aline van Veelen, Amsterdam
Lay-out Marjolein Rietveld

VRIJE UNIVERSITEIT

Heritability of cognitive abilities and of attention problems

A longitudinal twin study in childhood

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. T. Sminia,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Psychologie en Pedagogiek
op dinsdag 17 juni 2003 om 13.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Maria Johanna Helena Rietveld

geboren te Vlissingen

promotor : prof.dr. D.I. Boomsma
copromotor : dr. C.V. Dolan

Contents

Chapter 1

Introduction

Intelligence and childhood psychopathology have both been the subject of intense research within the field of behavior genetics. The substantial heritability of psychometric measures of general intelligence is now widely acknowledged. In contrast, much less is known about the genetic architecture of specific cognitive abilities. Heritability of childhood psychopathology varies across the various dimensions of psychopathology. Heritability of both cognitive abilities and psychopathology may vary with age. The present thesis covers two studies; the structure and stability of Intelligence, and the development of Overactive / Attention Problem behavior. These variables are measured in large non-clinical samples of twins. The study of twins makes it possible to evaluate the contribution of genes and environment to phenotypic individual differences. The present studies are particularly important as they make use of a longitudinal design whereas most previous studies have taken an age-specific or cross-sectional approach. Repeated assessments of the same children provide the means to explore the causes of stability in Intelligence, and Overactive / Attention Problem behavior. In addition, a unique feature of both studies is the young age of the participating twins. Assessments of cognitive abilities took place at ages 5, 7, and 10 years, and assessments of Overactivity / Attention Problems took place at ages 3, 7, 10, and 12 years. This chapter provides an introduction of the literature concerning Cognitive Abilities and Overactivity / Attention Problems, aims of the two studies and applied methods of analysis.

The structure and stability of Intelligence

In this thesis, intelligence is viewed as a multi-dimensional construct comprising distinct but correlated specific abilities. This definition is closely related to Thurstone's theory of Primary Mental Abilities (1938). Thurstone proposed that human intelligence was too diverse to be expressed as a single, general ability (Spearman, 1904), and that intelligence could best be defined by several uncorrelated abilities. In the years following the introduction of this theory, the concept that specific intellectual abilities were uncorrelated was found untenable in the light of empirical results (Thurstone & Thurstone, 1941). The concept of intelligence as a general ability or as multiple correlated abilities continues to be a topic of intense debate (e.g., Sternberg & Grigorenko, 2002).

In this thesis a longitudinal study of intelligence is described. At the start of the longitudinal study, a group of 209 twin pairs were administered the shortened version of the Revisie Amsterdamse Kinder Intelligentie Test (RAKIT; Bleichrodt et al., 1984) at the age of 5 years. One and a half years later, 192 pairs of the original sample provided complete data on all subtests (Boomsma & Van Baal, 1998; Van Baal, 1997). The number of participating twin pairs increased to 197 when the children were tested around their 10th birthday. The shortened version of the RAKIT consists of six subtests that each require specific cognitive skills to complete. The composite score of these six subtests scores represent general IQ. The study of intelligence is unique in the following ways; the children were very young at the start of the study, the sample sizes at each age are considerable, boys and girls are equally represented, and the study is longitudinal. Because data were collected in twins, behavior genetic methods can be used to investigate the importance of genes and environment in explaining individual variation in intelligence and stability of intelligence during childhood. The study of intelligence has three goals. First, the clustering of specific abilities at the ages 5, 7, and 10 is explored by means of factor analyses of the multivariate age-specific data. Second, taking the age-specific findings as a starting point, the stability of this clustering from age to age is studied. The clustering and stability is explored at the phenotypic, environmental, and genetic level. Third, throughout these analyses, differences between boys and girls in mean performance and in genetic and environmental variances are evaluated.

The factor structure of intelligence

The phenotypic structure of psychometric intelligence has been explored extensively in factor analytic studies (Carroll, 1993; McCall et al., 1973; Reinert, 1970; Schaie et al., 1998). In the factor analytic approach, the structure of intelligence is assessed by modeling the covariation between the subtests (or items) of the intelligence test. Studies of intelligence

have arrived at factor patterns that reflect both general and specific intellectual abilities. An example of a general ability is the well-known concept of *g* (Jensen, 1998; Spearman, 1927). The Primary Mental Abilities theory of Thurstone (1938) postulates that intelligence is manifested by multiple specific abilities. The concepts of Spearman and Thurstone can both be integrated in the hierarchical factor model (Vernon, 1965; Carroll, 1993). This model presents the general intelligence factor (*g*) as a second-order general factor common to first-order factors. These first-order factors represent the more specific abilities of intelligence like verbal and spatial skills. The hierarchical model has become a popular method to analyze the structure of intelligence (Gustafsson, 1984).

A phenotypic factor structure, however, does not necessarily reveal the underlying causes of clusters of abilities. Clearly, the covariation between abilities arises because they share a common source of influence. Application of multivariate behavior genetic methods indicates whether this common source is either genetic, environmental, or both. In this thesis, the factor structure of intelligence is first explored at the phenotypic level, then at the genetic and environmental level. With these analyses, the causes of observed covariation among cognitive abilities are identified. There have been numerous behavior genetic studies that addressed the factor structure of intelligence at the genetic and environmental level (Alarcón et al., 1998, 1999; Cardon et al., 1992; Casto et al., 1995; Hay & O'Brien, 1983; LaBuda et al., 1987; Luo et al., 1994; Pedersen et al., 1994; Petrill et al., 1996, 1998; Price et al., 2000; Rice et al., 1989; Rijsdijk et al., 2002; Tambs et al., 1986). Predominantly, these studies have adopted an age-specific or cross-sectional design. That is, data were either collected at one specific age or pooled across a wide range of ages. These studies differ in measurements, factor analytic approaches, and models, by which the genetic and environmental (co)variances were analyzed. This variation has resulted in a variety of genetic and environmental factor models to the observed clustering of cognitive abilities. In general, the genetic factor structure is complex and dependent on the age composition of the sample. Genetic factors account both for variance that is common to abilities and for variance that is specific to abilities. The environment that is unique to the individual explains variance that is specific to abilities. This source of variance thus makes no contribution to the observed clustering of intellectual abilities. The family environment that makes members of the same household more alike contributes to the covariance of abilities.

In this thesis, the phenotypic factor structure of the RAKIT at age 5 is explored by means of exploratory factor analyses. The best fitting model obtained at age 5 is applied to the data at ages 7 and 10 by confirmatory factor analyses. The exploratory and confirmatory approaches differ in the degree of hypothesized clustering of subtests; the explorative approach emphasizes the identification of clusters whereas the confirmative approach seeks to confirm an explicit hypothesis (e.g., Hartman et al., 1999). Subsequently, the ge-

netic and environmental contributions to the observed variance and covariance of the subtest scores is investigated.

Multivariate genetic analyses

The different degree of genetic relatedness between monozygotic (MZ) and dizygotic (DZ) twin pairs is used to estimate the genetic and environmental contributions to (co)variance of the phenotypic trait scores. Similarity between MZ twins is due to additive genetic and genetic dominance influences plus environmental influences that are common to both twins. Variation within MZ pairs arises only due to environmental influences that are unique to the individual. Members of a DZ pair share on average 50% of their additive genetic effects and 25% of their genetic dominance effects. Similarity between DZ twins is further increased by shared environmental influences. Estimation of the relative importance of the different sources of variance is obtained by analyzing the variance and covariance in MZ and DZ twins (Martin & Eaves, 1977). A first impression of the relative importance of each variance component is obtained by inspecting the standardized covariances, which are the twin correlations (Plomin et al., 2001). In the univariate case, a twin correlation is calculated between a subtest score of one member of a pair and a score of the same subtest of the other member of the pair. The importance of additive genetic influences is assessed by doubling the difference between MZ and DZ correlations. An indication of the importance of common environmental influences is obtained by subtracting the MZ correlation from twice the DZ correlations (Falconer & MacKay, 1996). MZ twins differences are due to experiences unique to the individual[1]. The extent to which MZ correlations differ from unity is thus indicative of the importance of the unique environment. The estimate of the unique environmental effect thus obtained includes measurement error. The presence of genetic dominance is indicated when the MZ correlations are larger than twice the DZ correlations.

In Chapters 2 and 3, the causes of the observed covariance among the six subtests of the RAKIT were investigated by means of multivariate analyses. To this end, cross-correlations were also considered. This is the correlation between the scores of twins on one subtest and the scores of the co-twins on another subtest. Cross-correlations may be interpreted in the same manner as in the univariate case. That is, if MZ cross-correlations are larger than DZ cross-correlations, this suggests that genetic effects contribute to the phenotypic covariation between subtest scores. The observed covariation between subtests is both a function of the subtest-specific genetic and environmental contributions, and a function of the covariation between these subtest-specific contributions. A large

[1]Epigenetic dissimilarity in MZ twins is not considered in this thesis (Petronis, 2001).

covariation between two sets of genes that influence two subtests does not imply a large observed covariation. Analogously, a large contribution of both sets of genes to the subtests does not imply a large observed covariation.

Initially, an unconstrained decomposition of the phenotypic covariance structure of the six subtests into genetic and environmental covariances matrices was considered by means of a Cholesky decomposition. A Cholesky or triangular decomposition model decomposes the phenotypic (co)variances into additive genetic, shared and unique environmental (co)variances that are common across subtests and into additive genetic, shared and unique environmental variances that are unique to each subtest. The common variance is indicative of the extent to which additive genetic, shared, and unique environmental variance is shared across subtests. The unique variance is indicative of the extent to which variance is subtest-specific and not shared across subtests. A Cholesky decomposition provides the most general approach to estimating the multivariate genetic and environmental components of variance and covariance. The genetic and environmental covariance matrices thus obtained are subjected to no other constraint than that they be positive (semi-)definite.

The results of the Cholesky decomposition provided a starting point for fitting first-order factor models to the genetic and environmental covariance structures. Three different factor models were considered. Firstly, the genetic and environmental structures were modeled by two correlated group factors with subtest-specifics. This model is plausible when the subtests may be ordered in two categories that reflect distinct, but correlated intellectual abilities. An example is the distinction between verbal and spatial abilities. Second, genetic and environmental covariance matrices were modeled by one general factor with subtest-specifics. This model represents the concept of general intelligence. Third, it was hypothesized that there were no sources of variation common to the subtests, that is, the model included subtest-specifics only. This model, which was considered mainly in comparison to the other models, implies that each subtest taps on a specific intellectual ability. Each of these models was applied to the genetic and environmental covariance matrices.

Modeling of the data

Modeling of the data, that is fitting theoretical models to the data, involves both the estimation of parameters and the evaluation of goodness of fit. The theoretical models concern the sources of phenotypic individual differences. Theoretical models give rise to specific hypotheses concerning the phenotypic covariance structure. By the solution of a set of simultaneous non-linear equations, that is by structural equation modeling, parameters are estimated by minimizing some measure of distance between the observed covari-

ance structure and the theoretical structure. The evaluation of goodness-of-fit involves comparing the discrepancy between the observed and expected covariance structures of a number of alternative models. A well fitting model indicates that the hypothesis is supported by the data. A poor fit indicates a discrepancy between the model and the data, and, once this discrepancy is located, may suggest possible revisions of the model. Throughout this thesis, structural equation modeling was done using (normal theory) Maximum Likelihood (ML) estimation in Mx (Neale et al., 1999). An important aspect of model fitting is the principle of parsimony. The addition of parameters to model each possible effect that appears to be present in the data may provide a useful baseline, but one should seek a parsimonious solution. The economic use of parameters enhances the understanding of the trait of interest. To decide whether a more parsimonious model (i.e., incorporating fewer parameters) is to be preferred to a less parsimonious model, hierarchical (likelihood ratio) tests are used. When covariance matrices of the observations in twin pairs are used as input in the model fitting procedure, a chi-square measure of fit is obtained. The chi-square and associated degrees of freedom are used as a measure of goodness-of-fit of the tested model. If the difference in chi-square and associated degrees in freedom between the initial model and more parsimonious model is not significant, the excluded parameter(s) account(s) for an insignificant part of the observed (co)variance. Provisional acceptation of the restricted model is then justified, in the understanding that further reduction of the model may be possible. If the difference in chi-square and degrees in freedom is significant, the excluded parameter(s) make(s) a significant contribution to the observed data and the restricted model is not tenable. The use of raw data as input in the model fitting procedure provides a log-likelihood function value as an index of the goodness-of-fit. Minus twice the difference in log-likelihood between the models with and without the restriction is asymptotically chi-square distributed, given that the more restricted model is true. The degrees of freedom equal the difference in the number of parameters of the two models. So, as in the case of covariance matrices, hierarchical chi-square tests are used to evaluate the plausibility of the restrictions that are imposed (e.g., Loehlin, 1992).

The factor structure of intelligence across development

It has been suggested that over the course of normal development, a change occurs in the structure of intelligence. Garrett (1946) hypothesized that intellectual abilities measured after childhood are much more independent of one another than abilities in earlier years. Although changes in the structure of intellectual abilities have great intuitive appeal from a developmental viewpoint, evidence from factor analytic research is not convincing. Reinert (1970) reviewed nearly 60 factor analytic studies. On the basis of those studies that

reported a change, it was concluded that a trend towards increased differentiation appeared to be present. However, more recent factor analytic studies of intelligence have provided no evidence in favor of the differentiation hypothesis (Bickley et al., 1995; Carroll, 1993; Deary et al., 1996; Juan-Espinosa et al., 2000; Werdelin & Stjernberg, 1995). Changes in the structure of intellectual abilities as a function of age were not observed. In this thesis, the issue of differentiation of abilities over development is addressed by analyzing the longitudinal data. Differentiation of cognitive abilities can be considered both at the phenotypic level and at the genetic and environmental level. Because the pattern of abilities at the phenotypic level is the result of a combination of genetic factors and environmental factors, a variety of patterns across phenotypic, genetic and environmental levels may emerge. This also implies that a lack of differentiation of abilities at the phenotypic level does not exclude differentiation of abilities at the genetic and environmental level. From a behavior genetic perspective, differentiation is evaluated by the covariance among genetic factors and environmental factors specified at multiple ages. A high degree of covariation between factors over time suggests stability, or a lack of differentiation. However, if these factors make a minimal contribution at each time point, then differentiation is not an issue. Differentiation thus both depends on the degree of covariation among factors over time and the magnitude of the contribution of the factors. Only a few twin studies have addressed the longitudinal aspects of intellectual abilities in infancy and childhood (Cardon & Fulker, 1993; Price et al., 2000; Reznick et al., 1997). As with the age-specific studies, the interpretation of the results is complicated by differences in applied factor models. Genes contribute both to stability and change in intellectual performance. Environmental experiences that are unique to the individual do not explain the covariance among intellectual abilities and contribute only minimally to stability. Where detected, the shared environment contributes both to the covariance among abilities and across time. These few longitudinal studies of specific intellectual abilities suggest that the relative effect of the shared environment declines with age. Convincing evidence is presented by a much larger number of longitudinal studies of general IQ (McCartney et al., 1990; McGue et al., 1993; Patrick, 2000; Posthuma et al., 2003). With increasing age, the relative effects of genes on variance in intelligence increase, whereas the relative effects of the shared environment, which make individuals from the same family more alike, decrease. Studies performed at the Netherlands Twin Registry indicate that from age 5 to age 16, genetic influences on general IQ increase steadily from 26% to 60% (Bartels et al., 2002; Boomsma & Van Baal, 1998; Rijsdijk et al., 2002), whereas shared environmental influences decrease to a negligible contribution to the observed variation.

One aim of present thesis is to establish whether the clustering of abilities is due mostly to shared environmental influences in early childhood and mostly due to genetic effects

in late childhood. This issue is investigated by the analyses of the longitudinal, multivariate dataset.

Longitudinal genetic analyses

The longitudinal genetic analyses were based on the autoregressive, or simplex model (Boomsma & Molenaar, 1987; Guttman, 1954). This particular model was chosen because it provides a straightforward account of the stability of individual differences over time. In a simplex model individual differences at a given measurement occasion t are regressed onto individual differences at the immediately preceding occasion t-1, but not on individual differences at any earlier occasions (e.g., t-2). This gives rise to the autoregressive property that the correlations between measurement occasions are higher, when the occasions are closer in time. Covariances among the three ages of measurements are modeled by genetic and environmental factors specific to each age, and by 'carry-over' or transmission effects of these factors to subsequent ages. The model specifies the genetic and environmental variance unique to each measurement occasion by an innovation term that comes into play at each time point. The total variance is a product of the age-specific effects and age-to-age transmission effects. Bartels et al. (2002) applied the simplex model to a general measure of intelligence, measured at ages 5, 7, 10, and 12 years. In this thesis, the simplex model is applied to the RAKIT subtests, measured at ages 5, 7, and 10 years. The best fitting factor models that were derived by age-specific multivariate genetic analyses were retained in the longitudinal genetic analyses. That is, transmission parameters were introduced between the first-order genetic and environmental factors that were specified at ages 5, 7, and 10. To accommodate possible correlations between the residuals of a given subtest measured at different ages, correlations between these subtest specifics were introduced. In Chapter 3, a path diagram of the combined simplex-factor model is presented for both the genetic and environmental covariance matrices.

Sex differences

Reviews and meta-analytic studies suggest that gender differences in the means of cognitive ability test scores during childhood are absent (Helgeson, 2002; Sadker et al., 1991). It should be noted that a sex effect may only be detected if the test scores are not standardized by sex. The RAKIT employs identical norms for boys and girls.

In behavior genetic studies sex differences may be addressed in two ways. First, intelligence in boys and girls may be influenced by different sets of genes. Second, the relative magnitude of the genes may vary by sex. Similar issue can be raised concerning environmental effects. Gender effects have not been addressed systematically in behavior genetic studies of intelligence (Plomin et al., 2000; Thompson, 1993). The limited evidence sug-

gests the absence of sex differences in heritabilities of specific intellectual abilities during childhood and adolescence (Baker et al., 1994; Rijsdijk et al., 2002). Further, it appears that with respect to intelligence both sexes share the same environment and set of genes. These conclusions are confirmed by a study of general IQ (Bartels et al., 2002). The sample in that study is identical to the sample in this thesis. However, because general IQ is a composite of the individual subtests, differences between boys and girls in the subtests data remain to be investigated.

The analyses of sex differences

The assessment of a sex effect takes place initially by inspection of the twin correlations. If a distinct pattern in twin correlations between boys and girls is observed, it suggests that the magnitudes of the genetic and environmental influences differ by sex. Formal testing of the sex effect takes place by comparing the fit of a model in which the genetic and environmental parameter estimates are equal in boys and girls to the fit of a model in which these estimates are allowed to vary by sex. The question whether boys and girls share the same environment and set of genes can be addressed only if data from pairs of opposite-sex are available. If the correlation of opposite-sex pairs is distinctly different from the correlation of dizygotic same-sex pairs, variation in intelligence in boys may be influenced by different genes and (or) different environments than girls. Following inspection of the twin correlations, formal testing in Mx may be performed. Under the assumption that there is no sex effect, the genetic correlation of DZ twins is fixed at .5 and the shared environmental correlation is fixed at 1.00. One of these correlations may be freely estimated. If a genetic correlation of smaller than .5 is estimated it suggests that boys and girls do not share the exact same set of genes. If a shared environmental correlation of smaller than 1.00 is estimated it suggests that boys and girls do not share those environmental influences that are common to all members of the same family.

The development of Overactive and Attention Problem behavior

The constellation of inattentive, impulsive, and hyperactive behaviors is known as the Attention Deficit / Hyperactivity Disorder (ADHD). The disorder arises before the age of 7 years, and is persistent across situations like the home environment and school. Furthermore, the presence of symptoms has a significant impact on the child's daily functioning (DSM IV; American Psychiatric Association, 1994). The variables of interest in this thesis, overactive and attention problem behaviors are related to ADHD (e.g., Chen et al., 1994). From a developmental point of view, overactivity is reported to be most prominent

in pre-school children whereas inattentive behavior is more prominent in children who attend primary school (Biederman, 1998; Biederman et al., 2000; Ross & Ross, 1982; Rutter et al., 1998).

In this thesis, Overactivity and Attention Problem behavior were measured with the Child Behavior Checklist (CBCL). The CBCL is a standardized questionnaire for parents to report the frequency and intensity of behavioral and emotional problems exhibited by their child in the past six months. Twins were assessed repeatedly, close to their 3rd, 7th, 10th, and 12th birthday. The questionnaire that is used to assess the 3-year-old twins (Overactivity, CBCL/2-3; Achenbach, 1992) differs in item content from the questionnaire that is used to assess the 7-, 10-, and 12-year-old twins (Attention Problems, CBCL/4-18; Achenbach, 1991). The item-content of the problem scales is reported in Chapter 7. The total Overactivity score and Attention Problem score is obtained by summing the relevant item scores. Reports of problem behaviors were collected in birth cohorts 1986 through 1993. In total, we collected 11938 maternal reports of 3-year-old twins, 10657 maternal reports of 7-year-old twins, 6192 maternal reports of 10-year-old twins, and 3124 maternal reports of 12-year-old twins. Besides the large sample sizes at each age, additional strengths of the study is the equal representation of boys and girls, and the broad age range from 3 to 12 years. There are no other longitudinal behavior genetic studies that cover such a long and important developmental period. This study therefore makes an important contribution to the behavior genetic literature on child psychopathology. These data provide important information concerning Overactive and Attention Problem behavior as it is manifested at these specific ages. Further, the longitudinal approach of the study makes it possible to assess the causes of developmental stability of overactive and inattentive behavior. In all analyses of the age-specific and longitudinal data, sex differences are evaluated.

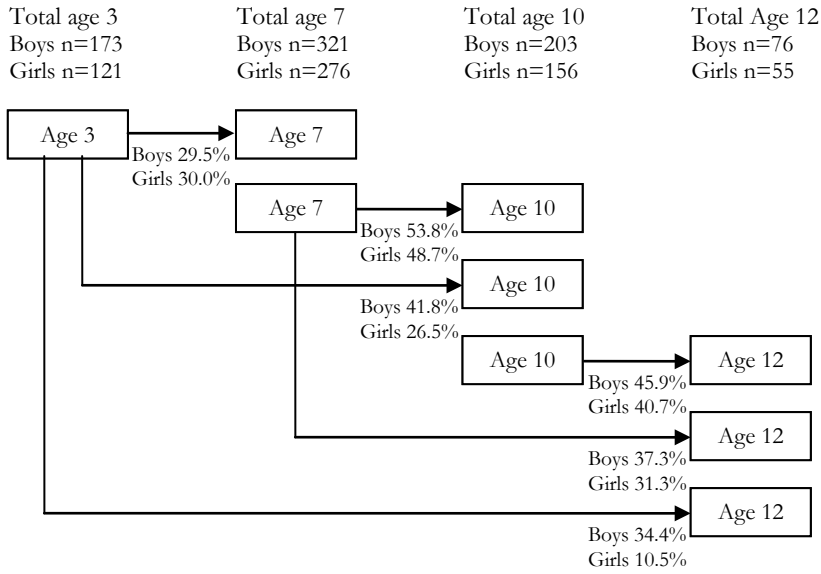
Methodological challenge: missing data in a longitudinal study

Loss of respondents at follow-up measurements is a major concern in longitudinal studies (Kessler et al., 1995). A reduction in sample size reduces the precision of estimates. Furthermore, nonresponse may lead to bias in the data when participants and nonparticipants differ with respect to variables of interest. Such a bias may hinder the generalization of the results obtained in this study.

The present thesis includes two longitudinal studies; (1) Intelligence measured at ages 5, 7, and 10 years, and (2) Overactivity / Attention Problem behavior assessed at ages 3, 7, 10, and 12 years. The non-response rate for the study of Intelligence is small, 8% at age 7 (n = 17 families) and 6% at age 10 (n = 12 families). Of these, only 5 families refused participation at both measurement occasions. No difference in general intelligence at age 5

was found between nonparticipating twins and twins who continued to participate. This finding suggests that the data collected at ages 7 and 10 are representative of the data collected at age 5. In the longitudinal study of problem behavior in twins at ages 3, 7, 10, and 12 refusal rates are larger compared to the longitudinal study of Intelligence. From one age to the next, around 20% of the sampled families did not respond. This percentage includes families, who changed residence without giving notice. As in the Intelligence study, families who dropped out at one occasion were likely to rejoin the study at a later measurement occasion. In contrast to the Intelligence study, families could also enter the study for the first time at age 7, age 10, or age 12. These different patterns of participation complicate the evaluation of missing data. It was established that nonparticipation at follow-up ages (ages 7, 10, and 12) is associated with a higher level of overactivity at the initial measurement occasion (age 3). Further, nonparticipation at ages 10 and 12 is associated with a higher level of attention problems at the previous measurement occasion, ages 7 and 10, respectively. The differences are small, but significant except for the interval between ages 3 and 7. To illustrate the challenges that are encountered with these results, developmental trajectories of children suffering from a relatively large degree of Overactivity / Attention Problems are shown in Figure 1.1.

Figure 1.1: Developmental pathways of children with severe Overactivity / Attention Problems



At each age, the total number of boys and girls with a borderline / clinical score ($T \geq 67$; $\geq 95^{th}$ percentile) on the Overactivity (age 3) and Attention Problem (ages 7, 10, 12) scale was obtained. These numbers are listed at the top of Figure 1.1. The total N refers to an

age-specific sample with at least one maternal report available, whereas the percentages refer to a longitudinal sample with at least two available maternal reports. The percentages indicate the number of children who remained borderline / clinical from one age to the next. From age 7 onward, around half of the boys and girls continue to have high Attention Problem scores. This compares to a smaller rate reported for the intervals including age 3, in both girls and boys. From these percentages it is tempting to conclude that a considerable group of children that suffered from Overactivity at age 3 did not have serious Attention Problems at a later age. It is also tempting to conclude that when children displayed an increased degree of Attention Problems at age 7, half of these children continue to have these problems at a later age. However, the children who drop out from the study display significantly more Overactivity and Attention Problems compared to the children who continue to participate. Thus, because these children leave the study it is not possible to draw a firm conclusion about the degree of stability of increased Overactivity and Attention Problems across childhood age.

The identification of the processes that generate the missing data is crucial for the analysis of the longitudinal data. Little and Rubin (1987) distinguish three processes that generate missing data; missing completely at random (MCAR), missing at random (MAR), and non-ignorable missing data. Under MCAR, the distributions of the missing and observed variable are identical. In other words, the selected sample of twins is a random sample of the target twin sample. This situation applies to the longitudinal study of Intelligence. Because of the significant difference in problem behavior between nonparticipants and participants, MCAR is not applicable to the longitudinal study of Overactivity and Attention Problems. Under the less stringent MAR, the distributions of the missing and measured variable are identical, conditional on a set of predictor variables. Here, we assume that the missing data in the study of Overactivity and Attention Problems are MAR. There are different approaches to analyze incomplete, longitudinal datasets. The method of raw data Maximum likelihood (ML), which is implemented in the Mx structural equation modeling program (Neale et al., 1999), combines information on collected data, together with information on the mean and variance of the missing data (e.g., Attention Problems at ages 10 and 12) given the presence of other data (thus, Overactivity at age 3 and Attention Problems at age 7). Wothke (2000) explored several treatments of missing data and reported that raw data ML produces unbiased estimates under MAR, and is highly efficient under MCAR. To avoid biased results, longitudinal data analyses took place by raw data ML estimation in both the Intelligence and Overactivity / Attention Problem studies.

Methodological challenge: classification of twin zygosity on the basis of questionnaire data

Zygosity determination of twin pairs is of crucial importance in behavior genetic research and genetic epidemiology. A correct estimation of genetic and environmental influences depends on the accurate determination of zygosity of the participating twins. Several procedures are available of which DNA typing using polymerase chain reaction amplification (PCR), blood group polymorphisms, and similarity questionnaires are most common (Smith & Penrose, 1955; Sutton et al., 1955; Jeffreys et al., 1985; Vlietinck, 1986; Goldsmith, 1991). DNA PCR and blood group polymorphisms are the most reliable methods, but both have the disadvantages of being time-consuming and expensive. Alternative methods like examination of photographs (Carter-Saltzman & Scarr, 1977), a phone questionnaire (Peeters et al., 1998), evaluation of anthropometric characteristics (Dahlberg, 1926; Gedda et al., 1976; Fairpo, 1979), the use of dermatoglyphic variables (Reed et al., 1977), examination of placental membranes (Bruner, 1970; Mathews et al, 1981), or various combinations of these means (Dencker et al., 1961; Sammalisto, 1961) have been employed. However, since these methods are not equally reliable or practical for large-scale twin registers, the use of a mailed questionnaire is often preferred given the large numbers of participants. Several studies combining genetic marker information and questionnaire data have shown that the determination of zygosity based on mailed questionnaires is 95% correct. The majority of these studies have been performed with participants varying widely in age. Given the variation in age, the possibility of an age-effect on the correct assignment of zygosity to twins has to be considered. There are no studies that explore how to make optimal use of information provided by multiple caregivers. Chapter 4 addresses zygosity classification by questionnaire by comparing the accuracy of classification by different informants and in twins of different ages.

Methodological challenge: the power to detect a contrast effect

The literature on behavior problems in childhood suggests that for certain traits, siblings interact with one another (Carey, 1986; Eaves, 1976). A negative sibling interaction is often referred to as a contrast effect. In studies of impulsivity, inattentive, and hyperactive behaviors the presence of a contrast effect in the twin data is often detected. A contrast effect may represent a true interaction between siblings. However, the prevailing interpretation of this effect is that it represents a bias introduced by the rater of the child's behavior. The bias is limited to parental ratings (Kuntsi et al., 2000; Sherman et al., 1997b; Simonoff et al., 1998; Thapar et al., 1995). Presumably, parents emphasize the behavioral difference displayed by their twins. For a heritable trait such as ADHD, observed differences in behavior are (already) larger in DZ pairs compared to MZ pairs. Thus, parental emphasis on behavioral differences leads to reduced resemblance in both MZ and DZ twin pairs, and this effect is greater in DZ than in MZ twin pairs. An accompanying statis-

tical consequence is that the variance in DZ twins is greater than in MZ twins. The pattern in twin covariances is also observed in the presence of genetic dominance. However, the discrepancy between MZ and DZ variances is not observed in case of genetic dominance, thereby providing the means to distinguish between genetic dominance and a contrast effect. If a contrast effect is present, but ignored in the analyses of the data, heritability estimates are biased upwards. Detection of a contrast effect and the distinction between a contrast effect and genetic dominance is problematic due to the requirement of large samples composed of twins. The required sample sizes are achieved by only a few twin studies (Eaves et al., 1997; Nadder et al., 1998; Nadder et al., 2001; Price et al., submitted; Thapar et al., 2000; Van Beijsterveldt et al., submitted), including the study of Overactive / Attention Problem behavior. The power problems that are encountered by twin studies of ADHD phenotypes are explored in Chapter 5.

The prevalence of Overactive / Attention Problem behavior

The DSM (diagnostic and statistical manual of mental disorders) system represents a categorical approach to the classification of disorders. In this system, children are characterized by the presence or absence of ADHD. In contrast to a categorized approach, a quantitative approach views ADHD as the extreme of a continuous distribution. The CBCL is an example of the quantitative approach. To enable a comparison between the prevalence rates of ADHD and the prevalence rates of Overactivity / Attention Problem behavior, a cut-off score that corresponds to a T-score of 67 is applied to the distribution of the CBCL problem scales (Chen et al., 1994). In so doing, the percentage of children who are likely to receive an ADHD diagnosis by DSM criteria are obtained. The results of this calculation are reported in Chapters 6 and 7. It is estimated that between 2% to 10% of the population is likely to receive the diagnosis ADHD (Faraone & Doyle, 2001). In community samples, the disruptive behaviors that are related to ADHD are among the most frequent complaints made by parents and teachers (e.g., Rutter & Garmezy, 1983). Although boys display more problem behaviors than girls, the gender ratio varies considerable among non-clinical studies of ADHD symptoms (Centraal Buro voor de Statistiek, 2001; Pineda et al., 1999; Szatmari et al., 1989) and between non-clinical and clinical studies. On average, the ratio girl : boy varies from 1 : 1 to 1 : 10 with the largest differences between boys and girls reported among referred children (Barkley et al., 1990).

Heritability of Overactive / Attention Problem behavior

Behavior genetic studies have adopted both the categorized and quantitative approaches to assess genetic influences on ADHD (Hudziak, 2001; Levy et al., 1997; Sherman et al., 1997a; Sherman et al., 1997b), but the analysis of quantitative differentiation has received

most attention. The heritability of ADHD and related behaviors is now firmly established by these studies (but see Joseph, 2000). Table 1.1 gives a summary of most cited twin studies that include maternal reports of behavioral symptoms that relate closely to ADHD. The variation in behavior problems displayed by children was assessed by analyses of summed item scores, instead of categories. Results that relate to additional measured phenotypes and informants are not reported here. Unless mentioned otherwise, data were collected by questionnaire.

Column-wise inspection of Table 1.1 reveals that the ADHD phenotype has become a much studied phenotype in behavior genetic research during the last decade. The assessment instruments, and the age and number of the participants vary greatly among studies. Given this variety, the consistency in outcomes is noteworthy. On average, 75% of the observed variation in ADHD phenotypes between children and adolescents is accounted for by genetic influences. The remaining 25% of the observed variation is explained by variation in environmental influences. These environmental influences are unique to the individual child or adolescent, that is, they contribute to observable differences in behavior between individuals that live in the same household. The environmental influences that are shared by individuals in the same household are negligible in explaining the variation in ADHD behaviors. With the exception of the study by Rhee et al. (1999), sex differences in the magnitudes of the genetic and environmental effects appear to be absent. The final column in Table 1.1 shows that the presence of a contrast effect is often reported by twin studies.

Table 1.1 Twin studies of ADHD and related phenotypes.

Study [1]	Age [2]	N [3]	Assessment Instrument [4]	Heritability [5]	Gender [6]	Contrast [7]
Eaves et al., 1997	8-16	1412	ADHD (Child and Adolescent Psychiatric Assessment, CAPA). Interview.	70% - 75%	Not tested	Yes
Edelbrock et al., 1995	7-15	181	Attention Problems (Child Behavior CheckList, CBCL).	66%	Not tested	Not tested
Gjone et al., 1996	5-9 12-15	390 526	Attention Problems (CBCL).	73% - 79%	Not tested	Not tested
Hudziak et al., 2000	8-12	492	Attention Problems (CBCL).	68% - 76%	Not tested	Yes
Kuntsi et al., 2000	7-11	125	Impulsive-Hyperactive (Connors' Parent Rating Scales, CPRS).	71%	Not tested	Yes
Martin et al., 2002	5-16	656	ADHD (CPRS). Hyperactivity (Strength & Difficulties Questionnaire, SDQ).	69% - 74%	Not tested	Yes, hyperactivity
Nadder et al., 1998	7-13	900	ADHD (CAPA; DSM-III-R/-IV; International Classification of Diseases, ICD-10). Telephone survey.	51% - 61%	No	Yes
Price et al., subm.	2 3 4	5491 5406 4451	Hyperactivity (Revised Rutter Scale for Preschool Children; SDQ).	75% - 79%	Yes, age 3	Yes
Rhee et al., 1999	3-18	2391	ADHD (DSM-III-R).	85% - 90%	Yes	No (Levy et al., 1997)
Schmitz et al., subm.	7 12	429 325	Attention Problems (CBCL).	65% (age 7) 48% (age 12)	Not tested	No
Sherman et al., 1997b	11-12	576	Inattention, Impulsive-hyperactive (Diagnostic Interview for Children and Adolescents-Revised). Interview.	69% - 91%	Boys only	Not tested
Simonoff et al., 1998	8-16	1044	Hyperactivity (Rutter parents' scales, Rutter A).	69% - 75%	No	Yes
Spinath, et al., 1998	2-14	354	Activity (EAS Temperament Survey).	63%	Not tested	Yes
Thapar et al., 2000	5-17	2082	Overactivity, impulsivity, inattention, ADHD (DSM-III-R/IV, ICD-10, DuPaul). Hyperactivity (Rutter A).	73% - 83%	No	Yes, inatt. & hyperact.
Vd Oord et al., 1996	3	1281	Overactivity (CBCL).	41% - 50%	No	Yes
ZahnWaxler et al., 1996	5	184 ?	Attention Problems (CBCL).	72%	No	Not tested

Note: [1] The most updated study is reported in case of multiple publications with the same instrument and sample. [2] Selection of studies with participants of 3 years and older. [3] N in pairs. Adoptees and sibling pairs are included in the studies by Schmitz et al. (submitted) and Rhee et al. (1999), respectively. [4] Studies have either used all or a selection of items. [5] Broad heritability, that is, summed over all genetic sources of variance. [6] Differences in absolute genetic and environmental variance components. [7] Negative interaction between siblings.

The development of Overactive / Attention Problem behavior

Both clinical and community studies have established that numerous children who display severe hyperactive, impulsive, and inattentive behavior continue to do so throughout adolescence and into adulthood (Biederman, 1998; Klein & Manuzza, 1991; Moffit, 1990). Among the ADHD behaviors, the inattentive subtype has shown to be more stable across development compared to the hyperactive and impulsive subtypes (Biederman et al., 2000).

With the exception of the studies by Price et al., (submitted), Schmitz et al. (submitted), and Van der Valk et al. (1998), behavior genetic studies that incorporate a longitudinal design to assess ADHD behaviors are lacking. The advantage of such longitudinal studies is that they provide insight into the causes of stability of observed inattentive, impulsive, and hyperactive behavior. These causes may be either genetic, environmental, or both. The observed stability of hyperactivity from age 2 to age 3 and 4 varies between .55 and .60 (Price et al., submitted). Observed stability in attention problems during late childhood and adolescence is estimated between .54 and .60 (Schmitz et al., submitted; Van der Valk et al., 1998). The importance of genes in explaining the stability of behavior is reported by all three studies. The adoption study by Van der Valk et al. (1998) reported the contribution of the unique environment to stability in Attention Problem behavior. Whilst the entire school-age period is considered important in the expression and continuation of inattentive, overactive, and impulsive behavior, there is no longitudinal twin study that explores the role of genes and environment during this developmental period. This makes the study of the development of Overactive/ Attention Problem behavior unique.

Genetic analyses of Overactivity / Attention Problems

The same approach that was taken to study the covariance among the RAKIT subtests was taken in the analyses of the longitudinal dataset of Overactivity / Attention Problems. The genetic and environmental covariance structures were modeled by a Cholesky decomposition. The literature on ADHD behaviors and the twin correlations of Overactivity / Attention Problems suggest the contribution of genetic dominance to individual variation. Therefore, genetic dominance was specified to contribute to the phenotypic (co)variance instead of the shared environment. Further, the observed difference in DZ and MZ variances suggested the presence of a contrast effect. In the analyses of the age-specific and longitudinal data, a contrast effect was specified at each age. A contrast effect is hypothesized to operate at the phenotypic level and is therefore specified between the observed Overactivity / Attention Problems scores of the twins. A path diagram of the univariate model including the contrast effect is depicted in Chapters 6 and 7.

The present thesis

Intelligence, research questions

- Is the factor structure of intelligence at age 5 the same as the factor structure at ages 7 and 10?
- Is the observed clustering of abilities explained by genetic and / or environmental effects?
- Do the genetic and environmental contributions exhibit a factor structure identical to the phenotypic factor structure?
- Do genetic and environmental effects display a pattern of differentiation?
- Is stability in intellectual performance explained by genetic and / or environmental effects?
- Do the genetic and environmental results differ by sex?

Overactivity / Attention Problems, research questions

- What is the prevalence rate of severe Overactive / Attention Problem behavior at ages 3, 7, 10, and 12?
- Is Overactive / Attention Problem behavior stable across the age-range studied?
- What are the genetic and environmental contributions to Overactivity / Attention Problems at ages 3, 7, 10, and 12?
- Are these contributions of equal importance at each age, or do they display a developmental pattern?
- To what degree is the cause of stability of Overactivity / Attention Problems due to genetic and environmental sources of influence?
- Are the results at the phenotypic, genetic, and environmental level different in boys and girls?

Outline

Chapters 2 and 3 describe the study of Intelligence. The factor structure, the genetic and environmental influences on intelligence at age 5 are presented first. Next, the stability and change in the factor structure, and the genetic and environmental influences across ages 5, 7, and 10 are described in the subsequent chapter. Prior to the study of Overactivity / Attention Problems, zygosity ascertainment methods and statistical power to detect a contrast effect in ADHD phenotypes are considered in Chapters 4 and 5, respectively. Age-specific analyses of Overactivity at age 3, and Attention Problems at ages 7, 10, and 12 are reported in Chapter 6, followed by longitudinal analyses presented in Chapter 7. Chapter 8 contains a summary, discussion and synthesis of these results.

Behavior Genetics, Vol. 30, No. 1, 2000

Genetic Factor Analyses of Specific Cognitive Abilities in 5-Year-Old Dutch Children

M. J. H. Rietveld,[1,3] G. C. M. van Baal,[1] C. V. Dolan,[2] and D. I. Boomsma[1]

Received 13 Jan. 1999—Final 3 Sept. 1999

The genetic and environmental factor structures of intellectual abilities in 5-year-old Dutch twins were examined. Six subtests of the RAKIT, a Dutch intelligence test, were administered to 209 twin pairs. The subtests were categorized as either verbal or nonverbal. The genetic covariance structure displayed a two-common factor structure including specific factors to account for subtest residual variance. The correlation between the genetic Verbal and genetic Nonverbal factors did not differ significantly from zero. The shared environmental influence displayed a single-common factor structure. Unique environmental influences did not contribute to the covariance between subtests and were specific in origin. Estimates of heritability of the subtests ranged from 15% to 56%. Shared environmental influences were significantly present, but were modest in magnitude. The phenotypic data was best described by an oblique two-factor model. This model was not mirrored in the factor structures found for either the genetic or environmental covariances.

KEY WORDS: Specific cognitive abilities; factor models; heritability; children.

INTRODUCTION

The factor structure of genetic and environmental influences on specific cognitive abilities measured at one point of time in childhood is investigated in the present study. Six subtests of a Dutch intelligence test were administered to 5-year-old twins. The examination of cognitive abilities is particularly interesting during childhood when rapid accumulation of learning and experience takes place. In school, children are exposed to many novel environmental effects which act specifically on the development of their intellectual abilities. The accumulating effects of these environmental influences may result in a different pattern of cognitive abilities in groups of individuals varying in age (Reinert, 1970; Schaie *et al.,* 1989; Werdelin and Stjernberg, 1995). The association between age and differentiation of abilities relates to the differentiation hypothesis (Garret, 1946) which suggests that abilities tend to cohere strongly in infancy and childhood, insofar as it is testable. When maturation proceeds, the factorial pattern of intelligence changes and intellectual abilities become more independent from each other.

A useful approach to study the structure of intelligence is factor analysis. A factor model includes a set of common factors to explain the variance shared by various measurements and test-specific factors to explain any residual variance. In a second-order factor model the intercorrelation among the first-order common factors is explained by positing one or more second-order common factors. The intercorrelation among the variables is then decomposed into a part that is attributable to the first-order common factors and into a part that is attributable to the second-order common factors. The hierarchical factor model is a popular multivariate factor technique to examine the structure of

[1] Department of Biological Psychology, Vrije Universiteit, Amsterdam, The Netherlands.
[2] Department of Psychology, Universiteit van Amsterdam, Amsterdam, The Netherlands.
[3] To whom correspondence should be addressed at Vrije Universiteit, Department of Biological Psychology, De Boelelaan 1111, 1081 HV Amsterdam, Fax: +31.20.444.8832. e-mail: mjh.rietveld@psy.vu.nl.

0001-8244/00/0100-0029$18.00/0

intellectual abilities (Carroll, 1993; Gustafsson, 1984). A hierarchical model presents a general intelligence factor, usually referred to as *g*, as a second-order general factor common to first-order factors. These first-order factors represent various dimensions of intelligence, e.g., verbal and spatial intelligence. As pointed out by Vernon (1965) this model integrates both Spearman's concept of a general cognitive ability (Spearman, 1927) and the Primary Mental Abilities theory of Thurstone (1938).

Studies on the differentiation of cognitive abilities have been summarized by Carroll (Carroll, 1993, p. 677). In his extensive survey of factor-analytic studies, he concluded that the same cognitive ability factors are present at various ages, from the early school years to adulthood and beyond, and that there is no evidence to support the differentiation hypothesis. Two possible explanations that can account for the early differentiation of abilities were proposed. One is that young children already experience a variety of environmental influences which promotes the development of different abilities. Alternatively, the observed distinction between cognitive ability factors at young age may be the result of genetic specificity. Genetic specificity implies that intellectual abilities have no genetic source in common and that genetic effects result from independent sets of genes. The identification of a particular factor model at the observed, phenotypic level is not necessary reflected in the same factor model at the underlying genetic and environmental level. The structure of these genetic and environmental influences can be studied through use of multivariate genetic factor models and requires genetically informative subjects such as twins or adoptees. Application of factor models to such designs can reveal whether an observed association between measures results from a shared genetic source or a shared environmental source, or both. So far, a handful of genetic studies on intellectual abilities and elementary cognitive abilities in children and adults have been carried out using hierarchical or first-order factor models (Cardon *et al.,* 1992; Cardon, 1994; Casto *et al.,* 1995; Finkel *et al.,* 1995; LaBuda *et al.,* 1987; Luo *et al.,* 1994; Pedersen *et al.,* 1994; Petrill *et al.,* 1996; Petrill *et al.,* 1998; Rijsdijk *et al.,* under revision). Results suggest that one general factor, group factors, and test-specific factors are all required to account for the genetic covariance structure. Shared environmental influences display either a single-common factor structure or a multiple factor structure. Unique environmental influences to variation in cognitive abilities are largely subtest-specific effects. The magnitude of the genetic and environmental influences appears to be a function of both the tests used and of the sample investigated.

The field of analyzing individual differences in specific cognitive abilities in young children with genetic factor models is relatively unexplored. To our knowledge, the only recent reports on genetic factor models applied to data collected in childhood are reports from the Colorado Adoption Project (CAP). The CAP is a prospective longitudinal adoption study on behavior development starting in infancy (e.g., Plomin and DeFries, 1985). Cardon (1994) reported results of analyses of longitudinal cognitive data collected in adoptees and nonadopted sibling pairs at various ages. A hierarchical model with one second-order factor and four first-order factors was fitted to the data. At age 4, a substantial genetic effect common to the first-order ability factors was found. Heritable variation on the specific subtests was dominated by the effects of these first-order factors and little test-specific genetic effects were found. The unique environmental influences showed considerable ability-specific effects on the first-order common factors. Also, in contrast to the genetic influences, large unique environmental effects specific to each subtest were found. Notably, shared environmental factors contributed little to the observed phenotypic variation at this young age. At the age of 7, the magnitude of genetic and environmental influences changed slightly. Again, the influence of shared environmental factors was nearly absent. The genetic covariance matrix was dominated by a general second-order common factor and the first-order group factors. Compared to age 4, ability-specific genetic effects were found to be greater. Unique environmental influences were primarily subtest-specific in origin.

In an ongoing longitudinal study of brain function (Van Baal, 1997) and cognition in young Dutch twins data on six intelligence subtests were collected in 5-year-old boys and girls. This paper explores the genetic and environmental structure of specific cognitive abilities by evaluating a number of alternative factor models to the six measures of verbal and nonverbal IQ.

METHODS

Participants

The participants were 209 twin pairs recruited from the Netherlands Twin Register. The register contains around 50% of all Dutch twins born after 1986 (Boomsma *et al.,* 1992). All 209 pairs participated in a

study of the development of brain-activity and cognitive development at ages 5 and 7 (Van Baal *et al.,* 1996; Boomsma & Van Baal, 1998). Families were selected on the basis of the age and zygosity of the twins, and of their city of residence. Mean age of the children in the present study was 5 years and 3 months (80% within range 5 years and 1 month to 5 years and 6 months). School attendance was 100%; all children were in their first year of formal education.

Zygosity of the same-sex twins was determined by analysis of bloodgroup (142 pairs) or DNA polymorphisms (20 pairs), and in a few cases by physical resemblance (8 pairs). Because one twin did not complete all subtests, the pair was excluded from the genetic analyses. The complete sample comprised 47 monozygotic female (MZF), 37 dizygotic female (DZF), 42 monozygotic male (MZM), 44 dizygotic male (DZM), and 39 dizygotic pairs of opposite sex (DOS). Approximately two years prior to the experiment, information on parental occupation and education was obtained from 186 families by questionnaire. Socioeconomic status (SES) was assessed on a 5-step scale relating to current occupation of the fathers (NCBS, 1993a). Of 186 families, information on occupation was obtained in 178 fathers. The majority of these families (48%) was of middle SES, 24% was of lower SES, and 28% was of higher SES. Parental education was rated on a 7-points scale (NCBS, 1993b). Midparent score for education was positively correlated with the average offspring IQ score ($r = .29$). A correlation of .47 was found for level of education between the spouses.

Procedure

The brain-activity study required participants to come to the university. After arrival at the laboratory the protocol was explained to the twins and their parents. While one of the twins participated in the electrophysiological experiment, the other twin received the intelligence test which took approximately one hour to administer. IQ of each child was assessed individually by an experienced test-administrator. The whole session, including a break, lasted between three and four hours. All children received a present afterwards.

Measures

The RAKIT, a Dutch intelligence test, was used in the present study to assess cognitive abilities (Bleichrodt *et al.,* 1984). The RAKIT manual defines intelligence as a multidimensional construct composed of different factors (Bleichrodt *et al.,* 1987). This definition is closely related to Thurstone's theory of intellectual functioning (Thurstone, 1938). The full-scale test comprises 12 subtests. The concurrent validity with the WISC-R is .86 for total IQ. Raw subtest-scores are standardized to facilitate comparison among performances on subtests. In this study, six subtests of the RAKIT were employed to assess cognitive functioning. The Exclusion subtest measures reasoning by assessing the child's ability to induce a relationship between four figures, and the ability to determine that one of the figures is deviant; the Discs subtest measures spatial orientation and speed of spatial visualization; the Hidden Figures subtest relates to transformation of a visual field, convergence/flexibility of closure; the Verbal Meaning subtest is a vocabulary index and a measure of passive verbal learning; the Learning Names subtest measures active learning and remembering meaningful pictures; the Idea Production measures verbal fluency. The combination of these six subtests has been shown to correlate .93 with the full-scale test score within this age group and is generally accepted as a shortened version of the full-scale test (Bleichrodt *et al.,* 1987). Regarding the nature of the subtests and the way in which the child has to respond it was hypothesized that three subtests load on a Nonverbal factor (Exclusion, Discs, Hidden Figures) and three subtests load on a Verbal factor (Learning Names, Verbal Meaning, Idea Production). The response of the child when performing the nonverbal subtests is to point at the right picture or to move blocks to the proper place. The mode of response to the verbal subtests relates more strongly to language skills.

Statistical Analyses

Means, standard deviations, variance-covariance matrices, and correlations among IQ subtests were calculated separately for the first and second born twins (twin a and twin b respectively) using SPSS/Windows 7.5. Twin a refers to the boy in opposite-sex twin pairs. We fitted four factor models to the phenotypic data, namely, an oblique two-factor model, an orthogonal two-factor model, a one-factor model, and a model with specifics only. These analyses were conducted in Mx (Neale, 1997), using maximum likelihood estimation of parameters.

PRELIS 2 (Jöreskog and Sörbom, 1993) was used to compute the variance-covariance matrices of the observations, separately for each sex-by-zygosity group. These covariance matrices as well as the cross-twin cross-trait correlations are to be found at the website

of the Netherlands Twin Register, http://www.psy.vu.nl:80/ntr/.

Genetic Modeling

To obtain an estimate of additive genetic (A), shared environmental (C), and unique environmental (E) contributions to the observed variances and covariances between measures, structural equation modeling was employed (Neale and Cardon, 1992). Let Σp represent the expected phenotypic covariance matrix between traits. The expected covariance matrix can be partitioned into additive genetic covariance (Σa), into common environmental covariance (Σc) and into unique environmental covariance (Σe) as follows:

$$\Sigma p = \Sigma a + \Sigma c + \Sigma e$$

First, these three covariance matrices were estimated by means of Cholesky decompositions. Next, a series of oblique first-order factor models was fitted to the data. The oblique factor model is equivalent to a hierarchical factor model with one second-order factor and two first-order factors, where factor loadings of these first-order factors on the second-order factor are constrained to be equal.

The path diagram in Figure 1 represents the most general phenotypic factor model. In view of the nature of the subtests, we considered three different factor models derived from this general model. We started with fitting correlated genetic two-factor models. The path diagram in Figure 2 shows the decomposition of the covariance between subtests into two group factors. This model also includes test-specific effects.

Next, we viewed the subtests as indicators of general intelligence (Jensen, 1998), that is, we fitted a common single-factor model to the covariance matrices. The importance of one general genetic and one general shared and one general unique environmental factor is represented in the path diagram in Figure 3. In this model it is suggested that the observed covariation between the subtests results from one shared, underlying factor at each of the genetic and shared and unique environmental levels. Factors specific to each subtest were added to test for any residual variance.

Finally, the third model that was considered is one in which there are no sources of variation common to the subtests.

The structure of the shared environment was examined first. The various factor structures were evaluated for C, while genetic and unique environmental influences were modeled by Cholesky decompositions.

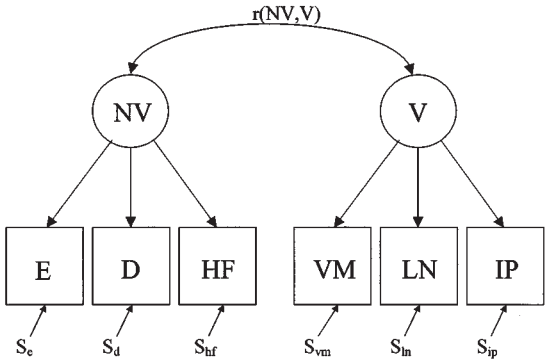


Fig. 1. Phenotypic first-order factor model with a Nonverbal factor (NV), a Verbal factor (V) and Specific factors (S_e to S_{ip}) unique to each subtest. NV and V are oblique factors, represented by r (NV, V). Exclusion, E; Discs, D; Hidden Figures, HF; Verbal Meaning, VM; Learning Names, LN; Idea Production, IP.

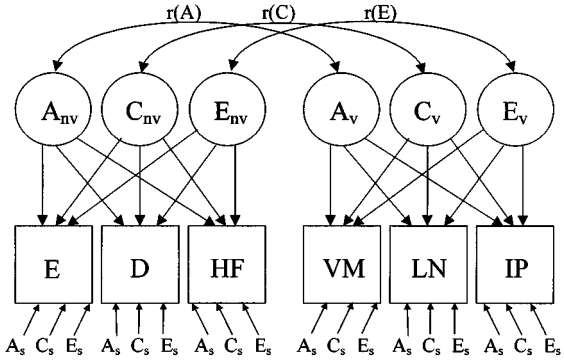


Fig. 2. This two-factor model (subscripts nv and v) with test-specifics (subscripts s) suggests that two factors are needed to explain variance among subtests. Proportion of total variance due to genetic, shared and nonshared environmental influences upon factors are represented by A, C, and E, respectively. The double-headed arrow between the Nonverbal and Verbal factor represents a correlation.

The model that best accounted for the shared environmental component was retained in subsequent analyses of the unique environment. Having established the best fitting model for E, this again was retained in examining the structure of the genetic part of the model.

The genetic analyses were performed with the program Mx (Neale, 1997). Mx provides estimates of the parameters in the model and an overall chi-square (χ^2) goodness-of-fit index. If the χ^2 has a probability smaller than a predetermined value (in this study, for α=.05), then the model is rejected and requires modification. The best fitting model in a sequence of models is determined by means of the hierarchical χ^2 tests. When a more restricted model does not describe the data significantly

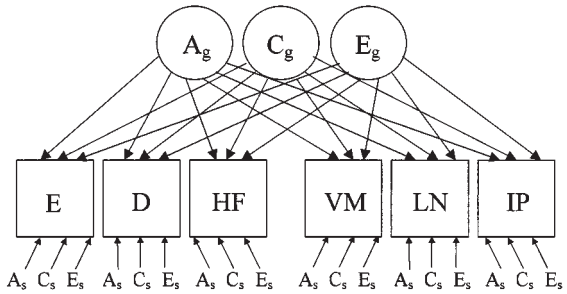


Fig. 3. This one-factor model (subscripts g) with test-specifics (subscripts s) suggests that one general factor explains all covariance among subtests. Proportion of total variance due to genetic, shared and nonshared environmental influences upon factors are represented by A, C, and E, respectively.

worse than the more general model, the most parsimonious model is chosen.

RESULTS

Descriptive Statistics

Inspection of the scores on 5 RAKIT subscales showed that the variables were approximately normally distributed (z-tests for skewness and kurtosis were nonsignificant). The distribution of scores on Hidden Figures differed slightly but significantly from a normal distribution (z-tests for skewness and kurtosis exceeded 1.96 and –1.96). The data were scanned for outliers by visual inspection, but none was found. Mean scores and standard deviations for total IQ and subtests are presented in Table I. No differences in means were found between sexes, between zygosities, or between twin a and twin b. The normative mean score for IQ is 100 (SD = 15) and the normative score on each subtest is 15 (SD = 5). The majority of the observed means were slightly higher than the population means. However, these differences were not significant for either males or females.

Phenotypic Analyses

Phenotypic correlations among the subtests are presented in Table II. As mentioned earlier, based on the nature of the subtests and mode of response we hypothesized that Exclusion, Discs, and Hidden Figures load on a Nonverbal factor and Verbal Meaning, Learning Names, and Idea Production load on a Verbal factor. The intercorrelations among subtests loading on the Nonverbal factor (r ranging from .26 to .40), and among subtests loading on the Verbal factor (r ranging from .16 to .55) were generally higher than the intercorrelations among subtests from different factors (r ranging from .07 to .36). The Nonverbal subtest Exclusion, however, showed a relatively large degree of overlap with Verbal subtests.

Various factor models were specified to investigate the structure of the phenotypic data. These models were fitted to the variance-covariance matrix of twin a and twin b separately. We first applied an oblique two-factor model to the data. The results of this analysis indicated that the fit of the model was acceptable (twin a, $\chi^2 = 8.23$, df = 8, $p = .41$; twin b, $\chi^2 = 16.64$, df = 8, $p = .03$). Inspection of the data of twin b revealed no systematic source for the deviation from the expected model. The mean intercorrelation among Verbal and Nonverbal subtests was .58 for twin a, and .57 for twin b. This two-factor model is represented by the path diagram in Figure 1. Subsequent reduced models, like an orthogonal two-factor model (twin a, $\chi^2 = 42.75$, df = 9, $p = .00$; twin b, $\chi^2 = 38.87$, df = 9, $p = .00$), an one-factor model (twin a, $\chi^2 = 34.04$, df = 9, $p = .00$; twin b, $\chi^2 = 37.51$, df = 9, $p = .00$), and a model with residuals only (twin a, $\chi^2 = 206.21$, df = 15, $p = .00$; twin b, $\chi^2 = 151.97$, df = 15, $p = .00$) did not lead to an improvement in fit.

Genetic Analyses

Correlations between twin a and twin b for each subtest for each zygosity group are shown in Table III. Inspecting the correlations of the twins we found higher values for monozygotic twins for all subtests compared to dizygotic twins. Except for Discs, the correlations of monozygotic twins were less than twice the correlations of dizygotic twins suggesting the presence of shared environmental influences.

Results of the Cholesky decomposition and subsequent multivariate factor models are presented in Table IV. No significant difference was found between a Cholesky model with sex differences and a Cholesky model in which estimates for A, C and E were constrained to be equal across sexes (χ^2 difference = 53.42, df = 63, $p = .80$). The genetic and environmental factor loadings derived from the Cholesky decomposition are reported in Table V. No clear pattern emerged from the genetic factor loadings. Formal testing of various factor models is necessary to resolve which factor model best describes the genetic structure. Clearly, a genetic one-factor structure is unlikely because a number of large factor loadings of the second through sixth

Table I. Means and Standard Deviations for Subtests and Total IQ-Score for Females and Males. Descriptives are Calculated Separately for Twin a (First Row) and Twin b (Second Row)

	Females (N = 207)[a]		Males (N = 211)	
	Mean	SD	Mean	SD
Exclusion	15.86	4.27	15.40	4.36
	15.69	4.15	15.40	4.38
Discs	13.85	4.67	13.92	5.09
	14.62	4.74	13.45	5.44
Hidden figures	17.03	4.50	16.27	4.07
	16.60	4.11	16.54	5.10
Verbal meaning	15.14	4.93	16.08	4.20
	15.83	4.33	16.29	3.98
Learning names	16.38	4.87	16.73	4.51
	16.80	4.66	16.54[b]	4.62[b]
Idea production	15.59	4.38	14.90	4.39
	15.38	4.33	15.02	3.52
Total IQ	102.95	13.98	102.48	13.33
	103.70	12.92	101.55[b]	12.74[b]

[a] N = number of participants
[b] $N - 1$. SD = standard deviation.

Table II. Phenotypic Pearson Correlations for Subtests for Twin a (Below Diagonal) and Twin b (Above Diagonal)[a]

	Exclusion	Discs	Hidden figures	Verbal meaning	Learning names	Idea production
Exclusion	1.00	.32	.33	.33	.21[b]	.07[ns]
Discs	.40	1.00	.26	.09	.07[b,ns]	.12[ns]
Hidden figures	.26	.26	1.00	.29	.19[b]	.08[ns]
Verbal meaning	.36	.14	.23	1.00	.43[b]	.16
Learning names	.32	.14	.24	.55	1.00	.29
Idea production	.21	.10[ns]	.10[ns]	.33	.32	1.00

[a] N = 209.
[b] $N - 1$. [ns] = nonsignificant correlation, $p > .05$.

factor are found. In addition, a model with only genetic specifics is not expected either since the off-diagonal factor loadings are substantial. The interpretation of the pattern of shared environmental factor loadings seems more straightforward. Since large factor loadings are found on the first factor, a one-factor structure is plausible. Regarding the unique environmental effects, specific factors appear important since the largest factor loadings are on the diagonal.

As indicated by the genetic factor loadings derived from the Cholesky decomposition, the genetic correlations varied substantially, ranging from –.56 to .91. The correlations among subtests which load on either the Nonverbal or Verbal factor did not exhibit consistently higher estimates compared to the intercorrelations among subtests from different factors. As expected, the correlations calculated for shared environmental effects showed more coherence (range from .27 to .94). Correlations between subtests for unique environmental effects were very low (range from –.12 to .20).

The Cholesky decomposition without sex differences was taken as a reference for evaluating changes in χ^2 and associated degrees of freedom of more parsimonious factor models. The structure of the common environmental contribution was investigated first while additive genetic and unique environmental influences were modeled by means of the Cholesky decomposition. Since the shared environmental correlations estimated from the full Cholesky decomposition suggested a one-factor structure, first a model including a General factor and Specifics was applied to the data (Model 3). The fit of this model did not lead to a significant increase

Table III. Twin Correlations for IQ Subtests for All Zygosity Groups

	MZF[a] (N = 47)	DZF (N = 37)	MZM (N = 42)	DZM (N = 44)	DOS (N = 39)
Exclusion	.47	.41	.71	.30	.45
Discs	.30	.16	.56	.20	.34
Hidden figures	.61	.56	.63	.23	.49
Verbal meaning	.64	.38	.55	.47	.54
Learning names	.80	.48	.68	.47[b]	.37
Idea production	.69	.38	.59	.44	.35

[a] N = number of twin pairs.
[b] N – 1.

Table IV. Model Fit Indices for Cholesky Decomposition and Nested Sequence of Factor Models[a]

Model	χ^2	df	P	Tested against	$\Delta\chi^2$	Δdf
1. Cholesky decomposition ACE, + sex differences	242.06	264	.83			
2. Cholesky decomposition ACE, no sex differences	295.48	327	.89	1	53.42	63
3. A one-factor structure imposed on C, + Specifics	299.08	336	.92	2	3.60	9
4. Only one factor	300.04	342	.95	3	0.96	6
5. A two-factor structure imposed on E, + Specifics	309.18	350	.94	4	9.14	8
6. A one-factor structure imposed on E, + Specifics	311.46	351	.94	4	11.42	9
7. Only Specifics	318.52	357	.93	6	7.06	6
8. A two-factor structure imposed on A, + Specifics	324.02	365	.94	7	5.5	8
9. Correlation between genetic factors constrained to zero	324.33	366	.94	8	0.31	1
10. A one-factor structure imposed on A, + Specifics	337.16	366	.86	7	18.64*	9
11. Only Specifics	385.29	372	.31	9	61.27*	6

[a] χ^2 = chi-square, $\Delta\chi^2$ = change in chi-square, df = degrees of freedom, Δdf = change in number of degrees of freedom.
* $p < .03$. p-value for all other $\Delta\chi^2$ tests > .25.

in χ^2. The test-specific effects explaining residual variance could be omitted from this model without deterioration in fit. This resulted in Model 4 including only one General factor explaining the common environmental variance and covariance (χ^2 = 300.04, df = 342, p = .95). Next, taking Model 4 as a point of departure, unique environmental effects were examined in a similar way. Model 5 and Model 6 postulated a two-factor structure with Specifics and a one-factor structure with Specifics, respectively. Although both models gave a good description of the data, the most parsimonious model, Model 7, included only Specifics (χ^2 = 318.52, df = 357, p = .93). This indicated that unique environmental effects did not contribute to the observed covariance between subtests, but only to subtest-specific variance. The concluding analyses investigated the structure of A, additive genetic effects. Model 7 was taken as the new reference model. In Model 7 the shared environmental structure was defined by one General factor, unique environmental influences were defined by Specifics only, and genetic effects were still specified as a Cholesky decomposition. Application of Model 8, the reduced model including a Nonverbal genetic factor and a Verbal genetic factor with Specifics resulted in a good fit. The correlation between the genetic Nonverbal and Verbal factor was estimated at .10 with the 95% confidence interval ranging from –.26 to .10. This outcome provided strong evidence for a genetic factor structure with two independent factors. The independency was explicitly tested in Model 9 and this resulted in an adequate description of the data (χ^2 = 324.33, df = 366, p = .94). More parsimonious models did not fit the data. The importance of each source of variance was tested in various sequences of model fitting. The outcome of each order was identical in the sense that A, C, and E were all found to be indispensible.

Thus, specifying a General factor for the shared environmental effects, Specific factors for the unique environmental effects, and a Nonverbal factor, a Ver-

Table V. Genetic (First Panel), Shared Environmental (Second Panel), and Unique Environmental (Third Panel) Factor Loadings Estimated for the Full Cholesky Decomposition[a]

	$A\text{-}Chol_1$	$A\text{-}Chol_2$	$A\text{-}Chol_3$	$A\text{-}Chol_4$	$A\text{-}Chol_5$	$A\text{-}Chol_6$
Exclusion	2.40					
Discs	.16	2.02				
Hidden figures	2.16	.36	1.14			
Verbal meaning	–.57	2.76	.98	.07		
Learning names	.37	1.22	–1.99	–.19	0.02	
Idea production	–.96	–.19	1.27	2.00	–.60	.04
	$C\text{-}Chol_1$	$C\text{-}Chol_2$	$C\text{-}Chol_3$	$C\text{-}Chol_4$	$C\text{-}Chol_5$	$C\text{-}Chol_6$
Exclusion	2.27					
Discs	2.45	.96				
Hidden figures	.81	–.16	1.78			
Verbal meaning	2.27	–.78	–.40	.00		
Learning names	1.82	–.74	1.56	.00	.00	
Idea production	1.80	.39	.57	.00	.00	.00
	$E\text{-}Chol_1$	$E\text{-}Chol_2$	$E\text{-}Chol_3$	$E\text{-}Chol_4$	$E\text{-}Chol_5$	$E\text{-}Chol_6$
Exclusion	2.77					
Discs	.18	2.79				
Hidden figures	.25	–.24	3.83			
Verbal meaning	.49	–.27	–.10	2.43		
Learning names	.22	–.35	.48	–.38	2.74	
Idea production	.28	.19	.13	.33	.47	2.52

[a] $A\text{-}Chol_1$ to $A\text{-}Chol_6$ = genetic Cholesky factors, $C\text{-}Chol_1$ to $C\text{-}Chol_6$ = shared environmental Cholesky factors, $E\text{-}Chol_1$ to $E\text{-}Chol_6$ = unique environmental Cholesky factors.

bal factor and Specific factors for the additive genetic effects best accounted for the variances and the covariances among subtests. The model is illustrated in Figure 4, with parameter estimates.

The genetic correlations between subtests can easily be calculated from the parameter estimates depicted in Figure 4. Correlations among Nonverbal subtests ranged from .22 to .69, and correlations among Verbal subtests ranged from .32 to .66. The variation in strength of the correlation reflected the influence of test-specific genetic effects. These effects only contributed to the observed variance and not to the observed covariance. Since no significant correlation was found between the genetic Verbal and Nonverbal factor, the intercorrelations among subtests loading on these two factors were zero.

Table VI contains the estimates for the influence of the genetic factors and estimates for shared and unique environmental influences based on Model 9. The parameter estimates for the genetic factors showed no clear pattern. Contribution to the total variance of

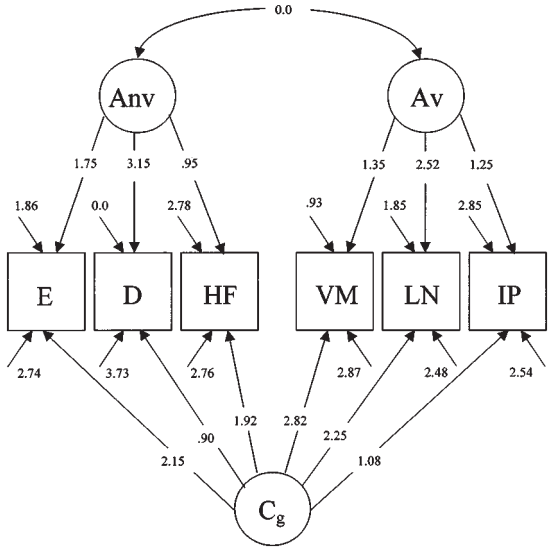


Fig. 4. Parameter estimates for best fitting model: two genetic factors (A_{nv} and A_v), one shared environmental factor (C_g), genetic specifics (single headed arrows placed above squares) and unique environmental specifics (single headed arrows placed below squares).

Table VI. Percentages of Total Variance Explained by Nonverbal Genetic Factor, Verbal Genetic Factor, Specific Genetic Factors, and Environmental Factors[a]

	Variance accounted for by genetic and environmental effects (%)					
	A_{nv}	A_v	A_{sp}	h^2	c^2	e^2
Exclusion	16	-	18	34	25	40
Discs	40	-	0	40	3	57
Hidden figures	5	-	39	44	18	38
Verbal meaning	-	10	5	15	42	44
Learning names	-	30	16	46	24	29
Idea production	-	9	47	56	7	37

[a] A_{nv} = Nonverbal genetic factor, A_v = Verbal genetic factor, A_{sp} = Specific genetic factors. h^2 = proportion of total variance explained by genetic factors, c^2 = proportion of total variance explained by shared environmental factors, e^2 = proportion of total variance explained by unique environmental factors.

the Nonverbal factor ranged from 5% to 40% and the contribution to the total variance of the Verbal factor ranged from 9% to 30%. Hidden Figures shared little variance with other Nonverbal subtests, just as a relatively small loading on the Verbal factor was found for Idea Production. Discs seemed to be a typical Nonverbal subtest; all genetic variance was shared with the other Nonverbal subtests. Looking at the pattern of heritability estimates, no clear distinction could be made between Verbal and Nonverbal tests. On average, the influences of genetic factors and unique environmental factors were of similar magnitude. The influence of shared environmental factors was much more modest. Verbal Meaning was the exception: a large estimate for shared environmental influences was found (42%).

DISCUSSION

We examined the factorial structure of genetic and environmental influences on specific cognitive abilities by fitting a number of oblique first-order factor models to data on specific cognitive abilities in 5-year-old Dutch twins. Bearing in mind that our sample size is not very large, we first summarize the results and subsequently compare these with findings reported in other studies. An oblique two-factor model gave a good description of the phenotypic covariance matrices. This does not necessarily imply, however, that a two-factor model will adequately describe the genetic and non-genetic covariance structure. It was found that the data were best described by a model with different factor structures for A, C and E. The genetic component of covariance displayed a two-factor structure. The Nonverbal tests loaded on one common genetic factor and the Verbal tests loaded on the other genetic factor. These two factors were independent. Genetic subtest-specific effects were needed to explain residual variance. The C covariance matrix was dominated by a single factor, accounting for the observed covariance among all subtests. Unique environmental influences were subtest-specific in origin, contributing to the variance specific to the subtests. The estimates for the genetic influences calculated separately for the Nonverbal, Verbal and Specific factors did not exhibit a clear pattern. The Nonverbal genetic factor explained only 5% of the total variance of the Hidden Figures subtest. In contrast, this factor explained all genetic variation in Discs. The genetic two-factor model included a correlation between the genetic Nonverbal and genetic Verbal factor. This correlation was found not to differ significantly from zero. This indicates that the genes influencing Verbal tests are independent from the genes influencing Nonverbal tests. This independency among the two genetic factors implies that the observed covariance between Nonverbal and Verbal tests is solely due to shared environmental influences. The standardized estimates for genetic influences and unique environmental influences were on average of the same magnitude. Genetic estimates ranged from 15% to 56% and unique environmental estimates ranged from 29% to 57%. A large influence of shared environment was found for Verbal Meaning (42%). For the other five subtests this influence was only modest (ranging from 3% to 25%).

The number of reports of genetic analyses on specific cognitive abilities in early childhood and in early school years is limited (Boomsma, 1993; Cardon *et al.*, 1992; Cardon, 1994). The method used in older studies conducted in childhood involved comparison of MZ

correlations with DZ correlations for subtest scores or factor scores (Foch and Plomin, 1980; Garfinkle, 1982; Segal, 1985; Wilson, 1975). While the results in these studies were indicative of genetic variation for some of the measures, not all separate subtest scores displayed heritability. In addition to the report of analyses of separate subtests, Segal (1985) and Wilson (1975) both reported on greater concordance in the pattern of subtest scores in MZ twins compared to DZ twins. The application of various factor models to examine the genetic and environmental covariance structure of specific cognitive abilities using a model fitting approach has just recently gained more attention. A small number of recent studies examining cognitive abilities involved the application of a hierarchical model to data collected in children and adults (Cardon *et al.*, 1992; Cardon, 1994; Luo *et al.*, 1994; Petrill *et al.*, 1996).

Results obtained from twin studies on intelligence in childhood suggest differential heritability for verbal, spatial, memory and perceptual speed tests (reviewed by Plomin, 1986). We did not observe this distinction in estimates of genetic influences between different subtests. Not all subtests in the administered Dutch intelligence test can be classified as purely verbal or purely nonverbal. Although the mode of response is distinctively verbal or nonverbal, the execution of the tasks may partly involve, for example, memory (Learning Names) and, for example, verbal fluency (Idea Production). Therefore, a difference in heritability estimates may arise compared to studies in which more pure subtests or factors are used.

Our results resemble those obtained by Plomin and Vandenberg (1980) in a reanalysis of Koch's (1966) Primary Mental Abilities data obtained in 5- to 7-year-old twins. Plomin and Vandenberg (1980) reported that verbal and spatial abilities, both showing a large genetic influence, are genetically independent at this stage in development. Cardon (1994) applied a hierarchical model to cognitive ability data collected in CAP participants at various ages. Both at ages 4 and 7 years, a genetic general factor was apparent and genetic subtest-specific effects were small. These results are in clear contrast to our findings. Our analyses revealed the presence of two independent genetic factors and genetic subtest-specific effects contributed significantly to the genetic variance. The hierarchical model (Cardon, 1994) also included genetic effects on the group factors independent of *g* with increasing influence from 4 to 7 years. This last finding is more in line with our observation of two separate genetic factors at age 5. In our study, shared environmental influences were modest but significantly present while Cardon (1994) reported a minimal contribution of C. This is quite remarkable since several reviews on developmental intelligence agree upon a large influence of shared environmental factors in early life (e.g., Boomsma, 1993; McCartney *et al*, 1990; Thompson, 1993).

Most genetic factor analyses have been performed on cognitive ability data collected in participants of older age. Our findings here are to some extent consistent with those reported in studies on specific intellectual abilities in which older samples were examined, even in the elderly (Petrill *et al.*, 1998). Genetic influences are significantly present and environmental influences are largely not shared by members of the same family. Regarding the structure of the latent factors more differences between studies are observed. Some studies reported on substantial ability-specific genetic effects (Casto *et al.*, 1995; LaBuda *et al.*, 1987) while others reported a dominant role of *g* (Pedersen *et al.*, 1994; Petrill, *et al.*, 1998). The significant effects of both a general genetic factor and genetic group factors were reported in studies in which the data were analyzed through application of hierarchical models (Cardon *et al.*, 1992; Luo *et al.*, 1994; Petrill *et al.*, 1996). As in the majority of these studies, we found that a one-factor structure best explained the shared environmental covariance matrix. Among these reports of multivariate analyses most agreement was on unique environmental influences. Those influences are predominantly subtest-specific only.

Boomsma and Van Baal (1998) performed a longitudinal analysis of total IQ data collected in the same sample at ages 5 and 7. Shared environmental factors contributed to half of the observed variance at age 5 while we found a modest contribution of shared environmental factors on specific subtests. This apparent inconsistency arises because the impact of genetic and environmental influences shared by subtests is augmented, and the genetic and environmental influences specific to subtests are decreased when analyzing a composite IQ score (Eaves *et al.*, 1989; p. 201–202). Since our analysis revealed the importance of C in explaining the covariance between verbal and nonverbal subtests, a higher estimate of shared environmental factors was detected when analyzing total IQ.

We also collected information on occupation and education in the parents. Parental education showed a positive but moderate correlation with total IQ in their children. The correlation between education of the mother and education of the father was strong (.46).

Although education level is not equivalent to IQ, a strong positive relationship exists (Ceci, 1991). The relationship between education and IQ in parents may have an impact on the estimate of genetic and environmental influences on intelligence in their children. Since IQ is a heritable phenotype in adults, the estimates of shared environmental contributions may be inflated if the resemblance in IQ of spouses is based on phenotypic assortment.

Results from recent studies which focused on the differentiation hypothesis indicated that the same factorial pattern of cognitive abilities is maintained across time (Schaie *et al.*, 1989; Werdelin and Stjernberg, 1995). Behavior genetic factor models provide the means to examine whether this structural invariance at the phenotypic level also applies to the underlying genetic and environmental level. Our results show that at age 5, the phenotypic factor structure differs from the factor structure found for the genetic, shared environmental, and unique environmental sources of variance.

A related issue to the exploration of the factor structure of intellectual abilities is the relationship between cognitive functioning and behavior disorders. This negative association is found in clinical (e.g., Frick *et al.*, 1991) and in healthy populations (e.g., Dietz *et al.*, 1997). The covariance between intelligence and problem behaviors is indicative of a shared source, either environmental or genetic in origin, or both. Considering the findings in this study, an association of behavior problems with both verbal and nonverbal aspects of intelligence would suggest that environmental factors which are shared by children in the same family play a role in explaining the association. A stronger association of either verbal or nonverbal IQ with behavior problems would suggest a common genetic source.

This study is part of an ongoing longitudinal project in which intelligence and behavior problems in childhood are investigated (Van den Oord *et al.*, 1996; Van der Valk *et al.*, 1998). Therefore, we aim to further explore the stability of the genetic and environmental factor structure of cognitive abilities in a longitudinal design, and to examine the relationship between intelligence and problem behaviors.

ACKNOWLEDGMENTS

This work was financially supported by the USF (grant number 96/22). The authors gratefully acknowledge the assistance of Sophia Kramer with the administration of the intelligence test. The Netherlands Organization for Scientific Research is acknowledged for funding the work of Caroline van Baal (575-65-052). The research of Conor Dolan was made possible by a fellowship of the Royal Netherlands Academy of Arts and Sciences.

REFERENCES

Bleichrodt, N., Drenth, P. J. D., Zaal, J. N., and Resing, W. C. M. (1984). *Revisie Amsterdamse Kinder Intelligentie Test* [Revised Amsterdam Child Intelligence Test], Swets & Zeitlinger B. V., Lisse, The Netherlands.

Bleichrodt, N., Resing, W. C. M., Drenth, P. J. D., and Zaal, J. N. (1987). *Intelligentie-Meting bij Kinderen. Empirische en Methodologische Verantwoording van de Gereviseerde Kinder Intelligentie Test* [Intelligence-Measurement in Children. Empirical and Methodological Justification of the Revised Child Intelligence Test], Swets & Zeitlinger B. V., Lisse, The Netherlands.

Boomsma, D. I. (1993). Current status and future prospects in twin studies of the development of cognitive abilities: Infancy to old age. In Bouchard, T. J., Jr., and Propping, P. (eds.), *Twins as a Tool of Behavioral Genetics*, Wiley, Chichester, UK, pp. 67–82.

Boomsma, D. I., and van Baal, G. C. M. (1998). Genetic influences on childhood IQ in 5- and 7-year-old Dutch twins. *Dev. Neuropsychol.* **14**:115–126.

Boomsma, D. I., Orlebeke, J. F., and van Baal, G. C. M. (1992). The Dutch Twin Register: Growth data on weight and height. *Behav. Genet.* **22**:247–251.

Cardon, L. R. (1994). Specific cognitive abilities. In DeFries, J. C., Ploming R., and Fulker, D. W. (eds.), *Nature and Nurture During Middle Childhood*, Blackwell, Oxford, UK, pp. 57–76.

Cardon, L. R., Fulker, D. W., DeFries, J. C., and Plomin, R. (1992). Multivariate genetic analysis of specific cognitive abilities in the Colorado Adoption Project at age 7. *Intelligence* **16**: 383–400.

Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*, Cambridge University Press, Cambridge.

Casto, S. D., DeFries, J. C., and Fulker, D. W. (1995). Multivariate genetic analysis of Wechsler Intelligence Scale for Children-Revised (WISC-R) factors. *Behav. Genet.* **25**:25–32.

Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Dev. Psychol.* **27**:703–722.

Dietz, K. R., Lavigne, J. V., Arend, R., and Rosenbaum, D. (1997). Relation between intelligence and psychopathology among preschoolers. *J. Clin. Child Psychol.* **26**:99–107.

Eaves, L. J., Eysenck, H. J., and Martin, N. G. (1989). *Genes, Culture and Personality: An Empirical Approach*, Oxford University Press, London.

Finkel, D., Pedersen, N. L., McGue, M., and McClearn, G. E. (1995). Heritability of cognitive abilities in adult twins: Comparison of Minnesota and Swedish data. *Behav. Genet.* **25**:421–431.

Foch, T. T., and Plomin, R. (1980). Specific cognitive abilities in 5- to 12-year-old twins. *Behav. Genet.* **10**:507–520.

Frick, P. J., Kamphaus, R. W., Lahey, B. B., Loeber, R., Christ, M. A. G., Hart, E. L., and Tannenbaum, L. E. (1991). Academic underachievement and the disruptive behavior disorders. *J. Consult. Clin. Psychol.* **59**:289–294.

Garfinkle, A. S. (1982). Genetic and environmental influences on the development of Piagetian locico-mathematical concepts and other specific cognitive abilities: A twin study. *Acta Genetic. Med. Gemellol.* **31**:10–61.

Garret, H. E. (1946). A developmental theory of intelligence. *Am. Psychol.* **1**:372–378.

Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence* **8**:179–203.
Jensen, A. R. (1998). *The g Factor: The Science of Mental Ability,* Preager Publishers, Westport, CT.
Jöreskog, K. D., and Sörbom, D. (1993). New Features in PRELIS 2, Scientific Software International, Inc., Chicago.
Koch, H. L. (1966). *Twins and Twin Relations.* University of Chicago Press, Chicago.
LaBuda, M. C., DeFries, J. C., and Fulker, D. W. (1987). Genetic and environmental covariance structures among WISC-R subtests: A twin study. *Intelligence* **11**:233–244.
Luo, D., Petrill, S. A., and Thompson, L. A. (1994). An exploration of genetic *g:* Hierarchical factor analysis of cognitive data from the Western Reserve Twin Project. *Intelligence* **18**:335–347.
McCartney, K., Harris, M. J., and Bernieri, F. (1990). Growing up and growing apart: A developmental meta-analysis of twin studies. *Psychol. Bull.* **107**:226–237.
Neale, M. C., (1997). *Mx: Statistical Modeling* (4th Ed.), Department of Psychiatry, Box 126 MCV, Richmond, VA 23298.
Neale, M. C., and Cardon, L. R. (1992). *Methodology for Genetic Studies of Twins and Families,* Kluwer Academic, Dordrecht, The Netherlands.
Netherlands Central Bureau of Statistics (1993a). *Standard Beroepenclassificatie 1992* [Standardized Classification of Occupations 1992], Netherlands Central Bureau of Statistics, Voorburg/ Heerlen, The Netherlands.
Netherlands Central Bureau of Statistics (1993b). *Standaard Onderwijsindeling SOI-1978* [Standardized Classification of Education 1978], Netherlands Central Bureau of Statistics, Voorburg/Heerlen, The Netherlands.
Pedersen, N. L., Plomin, R., and McClearn, G. E. (1994). Is there G beyond *g?* (Is there genetic influence on specific cognitive abilities independent of genetic influence on general cognitive ability?). *Intelligence* **18**:133–143.
Petrill, S. A., Luo, D., Thompson, L. A., and Detterman, D. K. (1996). The independent prediction of general inteligence by elementary cognitive tasks: Genetic and environmental influences. *Behav. Genet.* **26**:135–147.
Petrill, S. A., Plomin, R., Berg, S., Johansson, B, Pedersen, N. L., Ahern, F. M., and McClearn, G. E. (1998). The genetic and environmental relationship between general and specific cognitive abilities in twins age 80 and older. *Psychol. Sci.* **9**: 183–189.
Plomin, R. (1986). *Development, Genetics, and Psychology,* Erlbaum Associates, Hillsdale, NJ.
Plomin, R., and DeFries, J. C. (1985). *Origins of Individual Differences in Infancy: The Colorado Adoption Project,* Academic Press, Orlando, FL.
Plomin, R., and Vandenberg, S. G. (1980). An analysis of Koch's (1966) primary mental abilities test data for 5- to 7-year-old twins. *Behav. Genet.* **10**:409–412.
Reinert, G. (1970). Comparative factor analytic studies of intelligence throughout the human life-span. In Goulet L. R., and Baltes, P. B. (eds.), *Life-Span Developmental Psychology,* Academic Press, New York, pp. 467–484.
Rijsdijk, F. V., Vernon, P. A., and Boomsma, D. I. Application of multivariate genetic models to Raven and WAIS subtests: A Dutch twin study. *Behav. Genet.* (under revision).
Schaie, K. W., Willis, S. L., Jay, G., and Chipuer, H. (1989). Structural invariance of cognitive abilities across the adult life span: A cross-sectional study. *Dev. Psychol.* **25**:652–662.
Segal, N. L. (1985). Monozygotic and dizygotic twins: A comparative analysis of mental ability profiles. *Child Dev.* **56**: 1051–1058.
Spearman, C. (1927). *The Abilities of Man,* MacMillan, London.
SPSS Inc. (1997). *SPSS Base 7.5 for Windows User's Guide,* SPSS Inc., Chicago, IL.
Thompson, L. A. (1993). Genetic contributions to intellectual development in infancy and childhood. In Vernon, P. A. (ed.), *Biological Approaches to the Study of Human Intelligence,* Ablex, Norwood, NJ, pp. 95–138.
Thurstone, L. L. (1938). *Primary Mental Abilities*, Psychometric Monographs, No 1.
Van Baal, G. C. M. (1997). A Genetic Perspective on the Developing Brain, Doctoral thesis, Vrije Universiteit, Amsterdam.
Van Baal, G. C. M., de Geus, E. J. C., and Boomsma, D. I. (1996). Genetic architecture of EEG power spectra in early life. *Electroencephalogr. Clin. Neurophysiol.* **98**:502–514.
Van den Oord, E. J. C. G., Verhulst, F. C., and Boomsma, D. I. (1996). A genetic study of maternal and paternal ratings of problem behaviros in 3-year-old twins. *J. Abnorm. Psychol.* **105**: 349–356.
Van der Valk, J. C., Verhulst, F. C., Stroet, T. M., and Boomsma, D. I. (1998). Quantitative genetic analysis of internalising and externalising problems in a large sample of 3-year-old twins. *Twin Res.* **1**:25–33.
Vernon, P. A. (1965). Ability factors and environmental influences. *Am. Psychol.* **20**:723–733.
Werdelin, I., and Stjernberg, G. (1995). Age differences in factorial structure: A study of the "differentiation hypothesis." *Interdisciplinaria* **12**:79–97.
Wilson, R. R. (1975). Twins: Patterns of cognitive development as measured on the Wechsler Preschool and Primary Scale of Intelligence. *Dev. Psychol.* **11**:126–134.

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

A Twin Study of Differentiation of Cognitive Abilities in Childhood

M. J. H. Rietveld,[1,3] C. V. Dolan,[2] G. C. M. van Baal,[1] and D. I. Boomsma[1]

Received 31 Oct. 2002—Final 17 Feb. 2003

The differentiation hypothesis in cognitive development states that cognitive abilities become progressively more independent as children grow older. Studies of phenotypic development in children have generally failed to produce convincing support for this hypothesis. The aim of the present study is to investigate the issue of differentiation at the genetic and environmental level. Six psychometric measures assessing verbal and nonverbal cognitive abilities were administered to 209 Dutch twin pairs at ages 5, 7, and 10 years. Longitudinal results provided little evidence for the differentiation hypothesis. Stability in subtest performance is due mainly to genetic influences. The shared environment contribution to phenotypic stability is small. The unique environment contributes to age-specific variance only.

KEY WORDS: Differentiation; cognitive abilities; heritability; longitudinal; childhood.

INTRODUCTION

The structure of individual differences in cognitive abilities during development has received considerable attention in cognitive developmental theory (Schaie, 1994; Vernon, 1976). One important hypothesis, dating back to Garrett (1946; see also Carroll, 1993; Reinert, 1970; Wohlwill, 1973), states that cognitive abilities become increasingly more differentiated during development.[4] In operational terms, this means that the intercorrelations among psychometric measures of ability decrease during normal cognitive development in children. So far, support for this hypothesis has been poor. In an early review of about 60 factor analytic studies, Reinert (1970) suggested that a trend toward increased differentiation was present. However, this conclusion was based on the selection of studies that actually reported a change. In a more recent review, Carroll (1993) failed to find clear evidence for the differentiation of abilities. More recent cross-sectional studies of the changes in abilities in both children and adults are consistent with Carroll's finding in that age-related differentiation of cognitive abilities was not observed (Bickley *et al.,* 1995; Deary *et al.,* 1996; Juan-Espinosa *et al.,* 2000; Werdelin and Stjernberg, 1995).

Although support for the differentiation is apparently weak, comparative factor analytic studies have two drawbacks. First, studies vary in the definition of differentiation within the factor model and in the criteria used to evaluate such change. A number of approaches have been suggested to assess factorial invariance. A comprehensive summary of theories of factorial invariance is given in Cunningham (1991; see also Horn, 1991). Factorial change and stability across time may be judged by criteria that vary in restrictiveness. A second drawback is the limitation of most studies to analyses at the phenotypic level. It has long been recognized in behavior genetics that the phenotypic covariance structure does not always reflect the latent covariance

[1] Department of Biological Psychology, Vrije Universiteit, Amsterdam, The Netherlands.
[2] Department of Psychology, Universiteit van Amsterdam, Amsterdam, The Netherlands.
[3] To whom correspondence should be addressed at Vrije Universiteit, Dept. of Biological Psychology, Van der Boechorststraat 1, 1081 BT Amsterdam. e-mail: mjh.rietveld@psy.vu.nl
[4] As pointed out by a reviewer, another way to view differentiation is in terms of the divergence during development of phenotypic mean trends of, for instance, fluid and crystallized intelligence. Mean trends cannot be addressed with the present data as the IQ used is tailored to each age group. This gives rise to age-standardized test scores.

0001-8244/03/0700-0367/0 © 2003 Plenum Publishing Corporation

structures due to genetic and(or) environmental factors (Plomin, 1983). The failure to establish phenotypic differentiation does not necessarily imply the absence of differentiation at the genetic and environmental level. Many papers have been devoted to the development of longitudinal behavior genetic models (Boomsma and Molenaar, 1987; DeFries and Fulker, 1986; Meyer *et al.,* 1999; Plomin and DeFries, 1981; McArdle, 1986; McArdle and Goldsmith, 1990). Several behavior genetic studies have addressed the issue of stability and change in cognitive development (Bishop *et al.,* 2003; Bartels *et al.,* 2002; Cardon and Fulker, 1993; Eaves *et al.,* 1986; Fulker *et al.,* 1993; Hay and O'Brien, 1983; Hewitt *et al.,* 1988; Plomin *et al.,* 1994; Reznick *et al.,* 1997; Wilson, 1983). Although these studies vary in the age range of the samples, psychometric instruments used to assess cognitive abilities, and statistical methods, the results show a considerable degree of agreement. Overall, genetic effects acting on specific cognitive abilities display both developmental stability (variance common to different ages) and change (age-specific variance). Genetic effects become increasingly more important in explaining individual differences as children grow older. In contrast to the differentiation hypothesis, the correlation between genetic factors that represent different domains of intelligence appear to increase with age (Casto *et al.,* 1995; Price *et al.,* 2000). The shared environmental effects contribute both to the correlation among psychometric subtests and to the correlation among these subtests over time. However, these effects diminish as children grow older (Patrick, 2000). Unshared environmental effects are important at every age but show little to no stability over time. Thus, with respect to the genetic and environmental contributions, no simple picture emerges with respect to the differentiation hypothesis. However, most research has involved univariate measures of cognitive abilities and has usually focused on the analyses of general intelligence (e.g., Bartels *et al.,* 2002; Bishop *et al.,* 2003). The number of behavior genetic studies incorporating a longitudinal design and multivariate measures is limited.

In the present study, we investigate the differentiation hypothesis using twin data. A multidimensional test of cognitive abilities was administered to 209 twin pairs at ages 5, 7, and 10 years. In part these are the same data as analyzed in Bartels *et al.,* (2002) and Boomsma and Van Baal (1998). However, rather than modeling a single repeated measure of general intelligence, here we model repeated measures of six specific cognitive abilities. We consider the least strict version of factorial invariance, namely configural invariance (Horn and McArdle, 1992; Schaie *et al.,* 1998). We present detailed results relating to the changes over time in genetic and environmental effects and to their contribution to the phenotypic stability. In these longitudinal analyses, we hope to establish the contributions, if any, of genetic and environmental effects to differentiation.

METHODS

Sample

This study is part of a longitudinal study of the development of intelligence and problem behavior. The sample was obtained from the Netherlands Twin Registry, which is maintained at the Vrije Universiteit in Amsterdam. The registry contains around 50% of all twins born in the Netherlands since 1986 (Boomsma, 1998). The recruitment of the initial sample of 209 twin pairs took place on the basis of age, zygosity, and city of residence. Mean ages at the three measurement occasions were 5.3 years (80% ranging from 5 years, 1 month to 5 years, 6 months), 6.8 years (80% ranging from 6 years, 6 months to 7 years, 1 month), and 10 years (80% ranging from 9 years, 11 months to 10 years, 1 month). All children had started elementary school at the beginning of the study. Zygosity of the same-sex twins in the initial sample was established by either blood group polymorphism (137 pairs) or DNA analyses (24 pairs), and in 9 twin pairs by physical resemblance as assessed by the test administrator. The sample comprised 47 monozygotic (MZ) female pairs, 37 dizygotic (DZ) female pairs, 42 MZ male pairs, 44 DZ male pairs, and 39 DZ opposite-sex pairs. At the three measurement occasions the number of participating pairs was 209 (age 5), 192 (age 7), and 197 (age 10). Five families did not participate at both age 7 and age 10. At the age of 5, one child failed one subtest due to difficulties during testing. One twin pair was assigned missing values on the verbal subtests measured at all three ages, because they suffered from hearing difficulties. A sample of 184 pairs provided complete data on all subtests at all three ages.

The demographic characteristics of the employed twin mothers agree with those in the norm population (Central Bureau of Statistics, 2001). However, the occupational status of the employed twin fathers is not quite representative of the Dutch population. The twin sample is somewhat underrepresented in the lower end of the socioeconomic scale.

Procedure

When the twins were 5 and 7 years old, they participated in a combined study of the development of cognitive abilities and brain function (Van Baal *et al.,* 1996). At these occasions, the children and their parents visited the laboratory at the university. While one of the twins participated in the electro-physiological experiment, the co-twin completed the intelligence test. At the third measurement occasion (when the twins were about 10 years old), parents were invited by letter to make an appointment for testing in their own home or at the university. The majority of the families preferred testing at home (70%). No difference in full IQ score was observed between children tested at home and children tested at the university.

Intelligence Test

A shortened version of the Revised Amsterdam Children Intelligence Test, or the RAKIT, to use the Dutch abbreviation (Bleichrodt *et al.,* 1984) was administered at all three occasions. The RAKIT is a Dutch psychometric intelligence test for children, with subtests covering a broad spectrum of intellectual capabilities. The test is designed for children between about 4 and 11 years. The items in the subtests are organized by both age level and difficulty, and are arranged in overlapping sets. Each set is tailored to a specific age group. The overlap in items between the three measurement occasions decreases with the increasing interval between ages. The shortened version of the RAKIT includes six subtests measuring verbal and nonverbal abilities:

Exclusion measures understanding of figural classes or similarities. The child is presented with four figures, of which three figures have some common characteristic. After establishing the relationship between these figures, the child is asked to identify the odd one out. Total score is the number of correct decisions. Since children of older ages are shown more items, the raw total score is not comparable across ages.

Discs is a measure of spatial orientation and speed of spatial visualization. This test consists of a wooden board with pins fixed in a particular shape and of small wooden discs with holes in a particular shape. The child is required to fit the right discs onto the pegs at the right place and in the proper position as quickly as possible. The score is the total time required to place all the blocks. Children at age 5 were given only 12 discs out a set of 18 discs.

Hidden Figures requires visual analysis, pattern recognition, matching, and the ability to ignore irrelevant stimuli. A large drawing consisting of many lines and six smaller simplified drawings are shown to the child. The child has to find out which one of the smaller drawings is present in the more complex drawing. The number of correct judgments is the total score. Besides a difference in the number of items presented, children of older age are allowed less time to reach a decision.

Verbal Meaning assesses knowledge of concepts, and verbal conceptualization. In this vocabulary test, the child is shown four drawings representing familiar objects or acts. The tester reads the name of one object or act aloud and the child is asked to choose which of the drawings best suits the word. The number of correct items is the total score. Older children are presented with a large number of items than the younger children.

Learning Names is a verbal memory task, which is used to assess the ability to learn and recall names with pictures. The tester shows the child a book containing pictures of cats and butterflies and states the proper (fictional) name of each animal. At times, the tester explains why a given animal was given a particular name. The child has to recall the names of the animals within a certain time limit when shown the pictures. Five-year-old children are shown fewer pictures than the older children. The total score is the number of correct responses.

Idea Production is a measure of verbal fluency. Several simple questions are posed, like "what items can you put in your coat pocket?". The child has to respond by producing as many answers as possible within a certain time limit. The questions asked, the maximum time for answering, and the evaluation of the answers is the same for each age group. The number of acceptable answers forms the total score.

Age-corrected norms are based on an interval of four months. Raw subtest total scores are corrected for age and transformed into standardized scores with a mean of 15 and a standard deviation of 5. The reliabilities (internal consistency) of the subtests are stable across age, varying from .69 to .90. The concurrent validity with the WISC-R is .86 for total IQ (see also Bartels *et al.,* 2002).

Modeling of the Data

Standard genetic (co)variance modeling of twin data was used to estimate the size and the structure of the genetic and environmental effects (Neale and Cardon, 1992). Phenotypic variance is assumed to be due to shared environmental (C), unique environmental (E), and additive genetic (A) effects (Plomin *et al.,* 2001). Although meta-analyses suggest that nonadditive effects do contribute to individual differences in IQ test scores (15% of the variance according to Daniels, Devlin, and Roeder, 1997), we do not consider possible nonadditive genetic effects (dominance, epistasis) here. There is in fact little evidence of dominance effects in the present sample (Bartels *et al.,* 2002; Boomsma and Van Baal, 1998), while the correlations do suggest the presence of shared environmental effects (see Table I below). In addition, the present sample sizes do not confer sufficient power to detect relatively small nonadditive genetic effects (Eaves, 1972; Posthuma and Boomsma, 2000; Rietveld *et al.,* in press). In view of these considerations, and the fact that the present twin design does not provide sufficient information to fit a model including C and dominance effects simultaneously, we limit our attention to models including A, C, and E.

On the individual level an observed phenotype (P) can be represented as a function of a subject's additive genetic, common environmental, and unique environmental deviations:

$$P_{ij} = aA_{ij} + cC_{ij} + eE_{ij}$$

where $i = 1,2$ (members within a twin pair) and $j = 1, \ldots N$ (number of twin pairs). The coefficients *a, c,* and *e* are population invariant and can be considered as regression coefficients or factor loadings of P on the latent factors A, C, and E. If the latent factors are constrained to have unit variance (Neale and Cardon, 1992), the decomposition of the phenotypic variance is

$$V_p = a^2 + c^2 + e^2$$

The different degree of genetic relatedness between monozygotic (MZ) and dizygotic (DZ) twin pairs is used to estimate the contributions of the latent factors to the phenotypic variation in cognitive abilities. Similarity (covariances) between MZ twins can be due to additive genetic influences (a^2) and/or environmental influences that are shared by both twins (c^2). DZ covariances equal $1/2a^2$, in addition to the shared environmental influences (c^2). Environmental influences that make MZ and DZ twins different from one another are unique environmental influences (e^2). If multiple measures on the same twins are available, the association between measurements is also analyzed as a function of A, C and E (Boomsma and Molenaar, 1986; Martin and Eaves, 1977), where A, C, and E themselves may be uni- or multidimensional.

Model Fitting

All models were fit using the program Mx (Neale *et al.,* 1999). Parameter estimates were obtained by maximizing the raw data likelihood. Given the presence of missing data, albeit limited in number in the present study (8% at age 7; 6% at age 10), this method is preferable to the analysis of covariance matrices (e.g., Wothke, 2000). To test various hypotheses concerning the genetic and environmental covariance structures, we fitted a number of models to the longitudinal, multivariate data. Some of these models were nested, in the sense that one model was derivable from a second, less restrictive model by fixing selected parameters in the second model to equal certain values or by imposing equality constraints on selected parameters in the second model. Given this nesting, minus twice the difference in log-likelihood between the models may be used to evaluate the imposed restrictions. This statistic is asymptotically chi-square distributed if the restrictions are tenable. The number of degrees of freedom of this test statistic equals the difference in the number of estimated parameters between the two models. If the chi-square test is not significant, we consider the restrictions tenable (e.g., Loehlin, 1992).

Two series of model fitting took place; analyses of the cross-sectional data followed by analyses of the longitudinal data. The aim of the cross-sectional analyses was to provide an initial insight into the multivariate data, to test for sex differences in covariance structure, and to obtain a starting point for the longitudinal analyses. The cross-sectional data were initially modeled as a Cholesky decomposition. Sex differences were evaluated by constraining the covariance structures to be equal for boys and girls. Cross-sectional analyses proceeded with the specification of a more parsimonious and hypothesis-driven factor model, derived from model-fitting results obtained in our previous study of cognitive abilities at age 5 (Rietveld *et al.,* 2000). The genetic covariance structure was specified as a correlated two-factor structure with subtest-specific factors. The shared environmental covariance structure was modeled as one general factor, without subtest-specific factors. The unique environmental covariance structure was modeled

as subtest-specific factors only. This model served as a reference for the final three analyses, in which the importance of the latent factors was evaluated by constraining the relevant factor loadings at zero.

To establish that the same trait is being measured at different ages, we imposed the minimal condition of configural invariance. Configural invariance implies that the factor pattern is invariant over time but that the factor loadings may vary in size. Below we do consider the more restrictive hypothesis that factor loadings are in fact equal over time. Configural invariance does not imply stability, in that similar factor patterns at each age do not imply high correlations over time between the genetic and environmental factors. The issue of stability was addressed by model fitting of the longitudinal dataset. These analyses were initiated with the specification of an autoregressive, or simplex model (Boomsma and Molenaar, 1987; Guttman, 1954). This particular model was chosen because it provides a straightforward account of the stability of individual differences over time. Figure 1 represents a path diagram of the simplex model that was applied to the genetic covariance matrix. The model supposes that genetic variance at any age is partly a function of genetic variance at the previous age, via paths *t*, and partly determined by other genetic factors, via paths *i*. The former paths represent transmission effects and the latter represent innovation effects. The Nonverbal and Verbal innovation factors are correlated within age 7 and 10 (paths $r(I_{nv}, I_v)$). To accommodate shared variance between the genetic Nonverbal and Verbal factor at age 5, we added an interfactor correlation, path $r(A_{nv}, A_v)$.

Figure 2 illustrates the environmental parts of the model. Since the shared environmental structure was originally modeled as a single common factor, only two transmission parameters were added to the model to account for the variance transmitted from age 5 to 7 and from age 7 to 10 (paths *t*). Innovation effects or unshared variance at age 7 and age 10 are represented by paths *i*. The unique environmental contributions were found specific in origin. As with the genetic specifics, each subtest-specific factor was allowed to correlate over time in the longitudinal model.

Having established that this model was acceptable by comparison with a Cholesky decomposition, longitudinal analyses proceeded by testing various hypotheses within this model. First, the correlations that were specified between the genetic specific factors and the unique environmental specific factors were constrained at zero. In so doing we investigated whether the stability in subtest scores is accounted for completely by the common factors. Second, the importance of the genetic and familial environmental group factors was evaluated by constraining the relevant factor loadings at zero. Third, it was explored whether the influences of the genetic group factors and the shared environmental group factors were of equal magnitude at different ages.

RESULTS

Phenotypic Analyses

Evaluation of skewness and kurtosis of each subtest measured at each age indicated no serious departure from normality. Visual inspection revealed no outliers, so we included all data in subsequent analyses. Mean differences due to gender, zygosity, and birth order of the twins were found to be absent. Nonparticipation at ages 7 ($N = 17$) and 10 ($N = 12$) was not related to the twins' full-score IQ at age 5. By confirmatory age-specific factor analyses using structural equation modeling, the covariance matrix of the six subtest scores was best described by an oblique two-factor model with subtest specifics (age 5, $\chi^2 = 8.09$ (8), $p = .43$; age 7, $\chi^2 = 2.74$ (8), $p = .95$; age 10, $\chi^2 = 11.35$ (8), $p = .18$). The subtests Verbal Meaning, Learning Names, and Idea Production loaded on one common factor. Exclusion, Discs, and Hidden Figures loaded on the other common factor. The common factors thus represented verbal and nonverbal (spatial) abilities. The consistency of these results across measurement occasions suggests that configural invariance at the phenotypic level is tenable.

Cross-Sectional Genetic Analyses

A first impression of the relative magnitude of genetic and environmental influences is obtained by the inspection of twin correlations. Because correlations of male, female, and opposite twin pairs did not differ by formal testing, they were pooled. Correlations for each subtest at the three ages are provided in Table I.

Although there is considerable variation in the observed correlations, the MZ twin correlations are consistently larger than the DZ twin correlations. With the possible exception of scores on the Learning Names and Idea Production subtests at age 10, nonadditive genetic effects appear to be absent. The mean correlations show that MZ twins increase and DZ twins decrease in similarity over time. This suggests that genetic effects gain importance in explaining phenotypic individual differences. Shared environmental effects appear to decrease over time. Because the MZ correlations are well

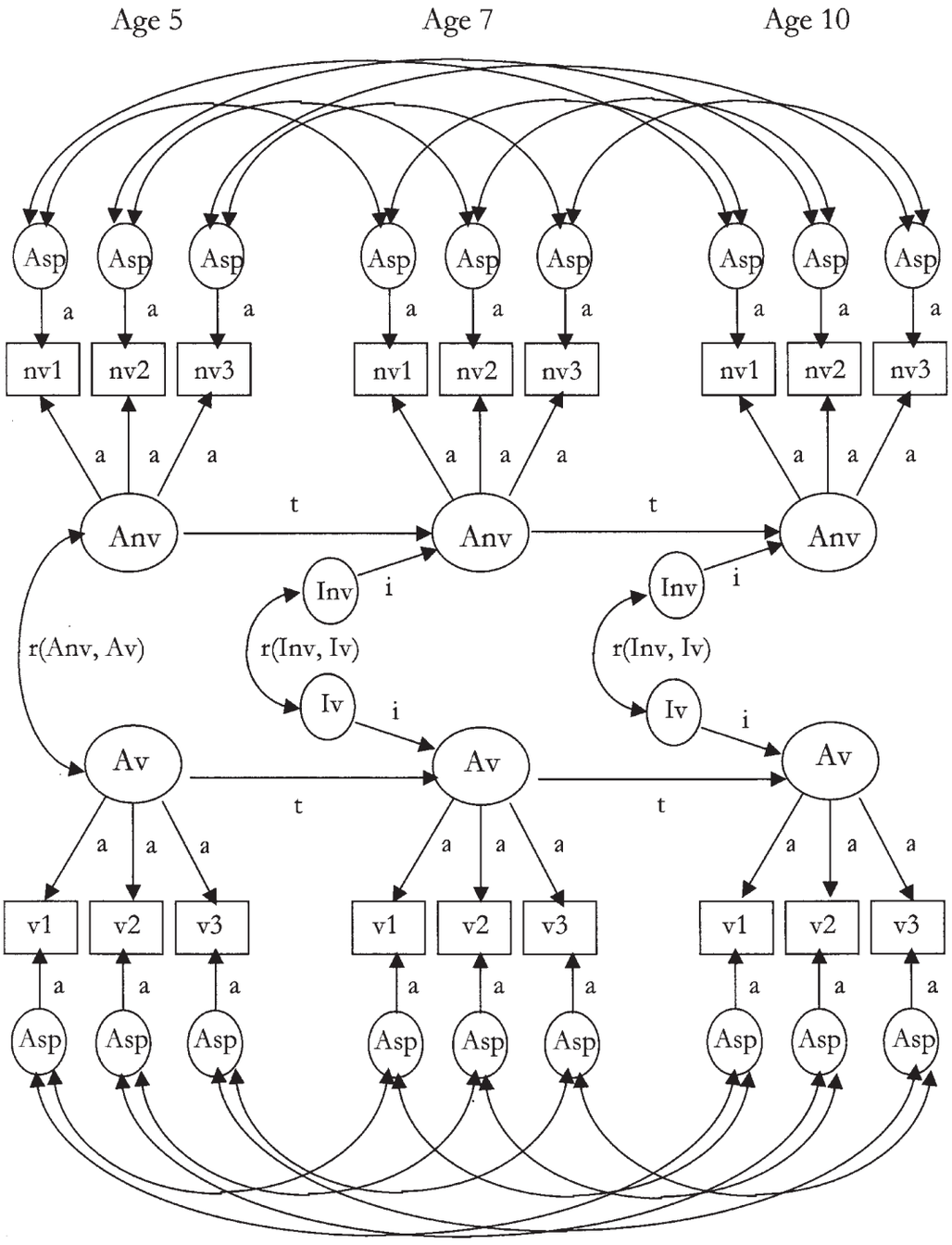


Fig. 1. Path diagram depicting the genetic part of the model. The simplex structure with correlated specifics (A_{sp}) suggests that both genetic common factors and subtest-specific residuals are needed to explain variance in subtest performance. The correlation between the Nonverbal (A_{nv}) and Verbal (A_v) factor at age 5, and the correlation between innovation factors (I) at age 7 and age 10 accommodate the covariation between nonverbal and verbal subtests. The variance that is shared across age is accounted for by autoregressive parameters (*t*), as specified between the common factors, and by the correlations as specified between the subtest-specific residuals. Each individual parameter (*a, t, i,* and *r*) is freely estimated. Exclusion, nv_1; Discs, nv_2; Hidden Figures, nv_3; Verbal Meaning, v_1; Learning Names, v_2; Idea Production, v_3.

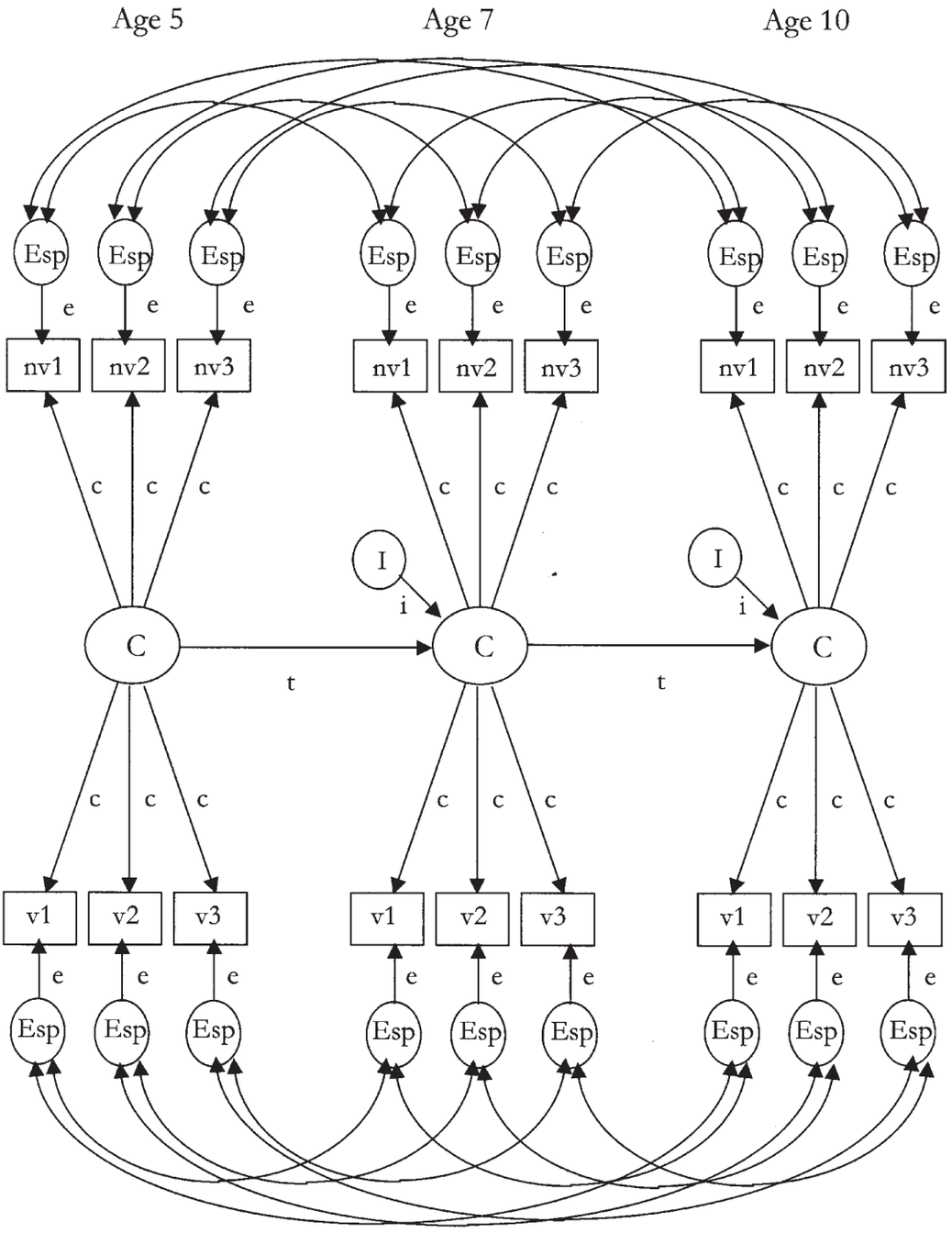


Fig. 2. Path diagram depicting the environmental part of the model. The simplex structure with three shared environmental common factors (C) suggests that the verbal and nonverbal subtests share all their age-specific variance. The variance that is shared among verbal and nonverbal subtests across age is accounted for by the autoregressive parameters (*t*) specified between the general C factors. Innovation factors (I) suggest time-specific variance for C. The subtest-specific unique environmental factors (E_{sp}) explain all variance in subtest performance at a specific age. The correlations as specified between the subtests-specific E factors account for variance shared by the same subtest across age. Each individual parameter (*c, e, t,* and *i*) is freely estimated. Exclusion, nv_1; Discs, nv_2; Hidden Figures, nv_3; Verbal Meaning, v_1; Learning Names, v_2; Idea Production, v_3.

Table I. Pearson Correlations for MZ and DZ for 6 Subtests, Measured at Age 5, 7 and 10

	Age 5		Age 7		Age 10	
	MZ $N = 89$	DZ $N = 120$	MZ $N = 79$	DZ $N = 113$	MZ $N = 82$	DZ $N = 115$
Exclusion	.60	.38	.42	.29	.65	.40
Discs	.44	.23	.61	.30	.39	$.18^{ns}$
Hidden Figures	.63	.43	.53	.32	.62	.31
Verbal Meaning	$.61^{a}$	$.46^{a}$	$.32^{a}$	.22	$.76^{a}$	.48
Learning Names	$.73^{a}$	.44	$.69^{a}$	.50	$.81^{a}$	.30
Idea Production	$.63^{a}$	.38	$.52^{a}$	.44	$.68^{a}$	$.17^{ns}$
Mean	.61	.39	.52	.35	.65	.31

Note: MZ = monozygotic twins; DZ = dizygotic twins; N = number of twin pairs.
[a] Based on $N - 1$ pairs; *ns* = nonsignificant at $p < .05$.

below unity, moderate to large unique environmental effects are expected to be present at each age (Plomin *et al.,* 2001). Systematic differences in results between the verbal and nonverbal subtests are absent. In Table II a summary of the cross-sectional analyses is given.

The nonsignificant difference between Model 1 (Cholesky decomposition with sex differences) and Model 2 (Cholesky decomposition without sex differences) suggests that sex differences are absent at all three ages. This result is consistent with other studies on general IQ in which the same twins were included (Bartels *et al.,* 2002; Boomsma and Van Baal, 1998). Model 3 (different factor structure for A, C, and E) gives an acceptable description of the data at all ages. Indicated by the significant detoriation in goodness of fit, we found that both the genetic common factors (Model 4, drop A correlated factors) and the genetic specific factors (Model 5, drop A correlated specific factors) are important at every age. The one-factor shared environmental structure accounts for a significant part of the total variation at ages 5 and 7, but not at age 10 (Model 6, drop C). On the basis of the agreement in model fits across age, we note that configural invariance appears to be tenable, with the exception of the apparent absence

Table II. Cross-Sectional Analyses

	Age 5		Age 7		Age 10	
Model	-2 LL (df)	χ^2 (df)	-2 LL (df)	χ^2 (df)	-2 LL (df)	χ^2 (df)
1. Cholesky decomposition ACE with sex differences	13785.9 (2315)		13192.1 (2112)		13307.2 (2172)	
2. Cholesky decomposition ACE, no sex differences	13840.5 (2378)	54.5 (63)	13263.1 (2175)	71.0 (63)	13361.0 (2235)	53. (63)
3. A: correlated two-factor structure + specifics C: one-factor structure E: specifics only	13868.9 (2416)	28.4 (38)	13304.1 (2213)	40.9 (38)	13395.1 (2273)	34.1 (38)
4. Drop A correlated factors, keep C and E	13929.0 (2423)	60.0 (7)*	13392.4 (2220)	88.3 (7)*	13523.4 (2280)	128.2 (7)*
5. Drop A specifics, keep C and E	13958.1 (2422)	89.1 (6)*	13325.8 (2219)	21.7 (6)*	13444.3 (2279)	49.1 (6)*
6. Drop C, keep A and E	13901.2 (2422)	32.2 (6)*	13317.2 (2219)	13.1 (6)*	13402.6 (2279)	7.44 (6)

Note: LL = log-likelihood, χ^2 = chi-square, difference in -2LL between nested models. Model 2 is nested with Model 2; Model 3 is nested with Model 2; Model 4, 5, and 6 are nested with Model 3. * = significant at $p < .05$.

of the shared environmental factor at age 10. The genetic structure displays a close resemblance to the phenotypic structure in that the subtests cluster around two correlated genetic factors, which are interpretable as the verbal and nonverbal abilities. We limit the detailed discussion of results in terms of the actual parameters to the longitudinal analyses.

Longitudinal Genetic Analyses

Although the shared environmental factor appeared to be absent at the age of 10, it was decided to retain this factor in the initial stage of the longitudinal analyses. Model fit indices are given in Table III.

The log-likelihood difference between Model 1 (Cholesky decomposition) and Model 2 (simplex structure) suggested that the longitudinal model (see Figs. 1 and 2) provided an acceptable point of departure. Three series of analyses were performed. First, the correlations between the subtest-specific factors were evaluated. The correlations obtained after fitting the factor model to the longitudinal data ranged from −1.00 to 1.00 for the genetic specifics and from −.13 to .23 for the unique environmental specifics. The worsening in goodness of fit of Models 3 (drop A correlations between specifics at ages 5 and 7), Model 4 (drop A correlations between specifics at ages 5 and 10), and Model 5 (drop A correlations between specifics at ages 7 and 10) suggested that the three sets of genetic correlations were significantly different from zero. Tests of the three sets of unique environmental correlations indicated that the correlations between ages 5 and 10 (Model 7) and between ages 7 and 10 (Model 8) were insignificant and that the correlations between ages 5 and 7 were marginally significant (Model 6). In view of this marginal effect, we dropped the correlations between subtest-specific unique environmental residuals. This implies that unique environmental effects do not contribute to the stability in subtest performance between ages (Model 9; E correlations between specifics at ages 5, 7, and 10 are dropped). This model served as the new reference model for subsequent analyses. The second series of analyses evaluated the importance of latent group factors. It was decided not to evaluate the genetic effects because their significant contribution to the observed (co)variance was already established by the age-specific analyses. Evaluation of the shared environmental effects confirmed the earlier obtained age-specific result, that is, the significant contribution at ages 5 and 7 (Model 10, drop C at age 5; Model 11, drop C at age 7). Here, as opposed to the cross-sectional analyses, the shared environment explained a significant part of the (co)variance at age 10 (Model 12, drop C at age 10). Presumably this was due to the fact that the shared environmental common factor served both to explain the covariance between the subtest scores at age 10 and part of the covariance between ages 7 and 10. The third and final analyses were carried out to explore the varying magnitude of the genetic and shared environmental influences at the three occasions. The log-likelihood

Table III. Longitudinal Analyses

Model	−2 LL (*df*)	Comparison	χ^2 (*df*)
1. Cholesky decomposition ACE, no sex differences	39463.0 (6608)		
2. A: simplex structure with 3 verbal and 3 nonverbal factors, 18 correlated specifics C: simplex structure with 3 general factors E: 18 correlated specifics	39801.5 (7025)	Model 1	338.5 (417), *ns*
3. Drop A correlations specifics between ages 5–7	39835.6 (7031)	Model 2	34.1 (6), $p < .05$
4. Drop A correlations specifics between ages 5–10	39818.3 (7031)	Model 2	16.8 (6), $p < .05$
5. Drop A correlations specifics between ages 7–10	39843.6 (7031)	Model 2	42.1 (6), $p < .05$
6. Drop E correlations specifics between ages 5–7	39814.1 (7031)	Model 2	12.7 (6), $p = .05$
7. Drop E correlations specifics between ages 5–10	39807.2 (7031)	Model 2	5.7 (6), *ns*
8. Drop E correlations specifics between ages 7–10	39812.4 (7031)	Model 2	10.9 (6), *ns*
9. Drop E correlations specifics between every age-interval	39828.3 (7043)	Model 2	26.8 (18), *ns*
10. Drop C age 5	39860.5 (7032)	Model 9	32.2 (8), $p < .05$
11. Drop C age 7	39860.6 (7032)	Model 9	32.3 (8), $p < .05$
12. Drop C age 10	39850.7 (7032)	Model 9	22.4 (8), $p < .05$
13. A factor loadings age 5 = age 7	39846.9 (7049)	Model 9	18.7 (6), $p < .05$
14. A factor loadings age 7 = age 10	39847.8 (7049)	Model 9	19.6 (6), $p < .05$
15. C factor loadings age 5 = age 7	39843.9 (7049)	Model 9	15.6 (6), $p < .05$
16. C factor loadings age 7 = age 10	39847.0 (7049)	Model 9	18.7 (6), $p < .05$

Table IV. Percentages of Total Variance Explained by Additive Genetic and Environmental Factors

		% Variance accounted for by genetic and environmental effects with 95% confidence intervals								
Subtest	Age	A_{nv}	A_v	A_{sp}	h^2	95%	c^2	95%	e^2	95%
Exclusion	5	28		10	38	.23–.53	19	.09–.31	43	.32–.55
	7	37		3	40	.28–.53	5	.01–.13	55	.42–.66
	10	55		3	58	.46–.69	7	.01–.17	35	.26–.45
Discs	5	24	—	22	46	.35–.57	3	.00–.09	51	.41–.62
	7	30	—	28	58	.45–.68	5	.01–.16	37	.28–.47
	10	29	—	15	44	.34–.54	4	.00–.11	52	.42–.61
Hidden	5	8	—	40	48	.33–.60	11	.03–.23	41	.31–.54
Figures	7	29	—	8	37	.20–.50	14	.06–.29	49	.38–.62
	10	36	—	12	48	.34–.61	8	.02–.19	44	.32–.56
Verbal	5	—	11	10	21	.08–.37	39	.24–.53	40	.31–.49
Meaning	7	—	9	10	19	.06–.33	17	.06–.29	64	.53–.75
	10	—	25	17	42	.23–.60	32	.16–.48	26	.18–.36
Learning	5	—	49	4	53	.38–.66	17	.06–.31	30	.23–.39
Names	7	—	23	16	39	.23–.60	32	.12–.48	29	.22–.39
	10	—	58	18	74	.65–.83	4	.00–.13	22	.15–.29
Idea	5	—	8	50	58	.45–.68	5	.01–.14	37	.28–.42
Production	7	—	2	56	58	.43–.68	4	.00–.14	38	.29–.51
	10	—	11	49	60	.45–.71	4	.00–.12	36	.26–.50

Note: A_{nv} = nonverbal genetic factor, A_v = verbal genetic factor, A_{sp} = specific genetic factors. h^2 = proportion of total variance explained by genetic factors ($A_{nv} + A_v + A_{sp}$), c^2 = proportion of total variance explained by the shared environmental general factor, e^2 = proportion of total variance explained by unique environmental specific factors. $h^2 + c^2 + e^2 = 100\%$.

difference tests suggested that the equality constraints on the factor loadings were tenable neither in the genetic part of the model (Model 13, A factor loadings age 5 equal to age 7; Model 14, A factor loadings age 7 equal to age 10) nor in the shared environmental part of the model (Model 15, C factor loadings age 5 equal to age 7; Model 16, C factor loadings age 7 equal to age 10). Strictly speaking, changes in factor loading indicate qualitative changes in the traits that are measured by the RAKIT. However, the interpretation of these results is complicated by the fact that the item content of the RAKIT is tailored to each age group. The overlap in items is large but not complete. The observed results may therefore be due to a true developmental effect and/or changes in item content.

Based on the best-fitting longitudinal model (Model 9), we calculated the decomposition of phenotypic variance of each subtest at each occasion. The results are shown in Table IV.

The contribution of the shared environment is either small at all three occasions (Discs, Idea Production), or decreases with time (Exclusion, Hidden Figures, Learning Names). The subtest Verbal Meaning forms an exception in that the contribution is variable, but consistently large. Generally, the subtest variance explained by the common genetic factors increases over time. Judging by the explained variance in the subtest Idea Production, it appears that this subtest is not well represented by the verbal common factor. With respect to the genetic subtest specifics, the results relating to the nonverbal and verbal abilities differ. As opposed to the verbal subtests, the nonverbal subtests display a decrease in the variance of these genetic residuals. Unique environmental influences are large, accounting for around one third to one half of the total variance for each measurement at all ages.

The longitudinal structure of the common genetic and shared environmental factors were modeled using (first-order) autoregressions.

The regressions were used to calculated the correlations among these factors over time. The correlations and confidence intervals are shown in Table V. [5]

[5] In the earlier study (Rietveld *et al.,* 2000) the 95% confidence interval of the nonsignificant genetic correlation was reported incorrectly (−.26 to .10). The correct confidence interval is −.26 to +.43. The confidence intervals between the present and earlier study overlap greatly. As opposed to the age-specific study, the genetic correlation between A-verbal and A-nonverbal at age 5 differs significantly from zero in the present study.

Table V. Correlations for Genetic Common Factors (top) and Shared Environmental Common Factors (bottom). The boundaries of 95% confidence intervals are shown in parentheses.

	Genetic correlations					
	Age 5 Nonverbal	Age 5 Verbal	Age 7 Nonverbal	Age 7 Verbal	Age 10 Nonverbal	Age 10 Verbal
Age 5 Nonverbal	1.00					
Age 5 Verbal	.25 (.07–.41)	1.00				
Age 7 Nonverbal	.92 (.80–1.00)	.23 (.07–.39)	1.00			
Age 7 Verbal	.25 (.07–.41)	.99 (.88–1.00)	.28 (.09–.43)	1.00		
Age 10 Nonverbal	.87 (.73–1.00)	.22 (.06–.37)	.94 (.86–1.00)	.26 (.08–.41)	1.00	
Age 10 Verbal	.24 (.07–.40)	.94 (.79–1.00)	.26 (.08–.42)	.94 (.80–1.00)	.30 (.13–.45)	1.00

	Shared environmental correlations		
	Age 5	Age 7	Age 10
Age 5	1.00		
Age 7	.82 (.67–.92)	1.00	
Age 10	.79 (.60–.92)	.97 (.80–1.00)	1.00

Judging by the correlations between the genetic common factors over occasions, the stability of genetic individual differences on these factors is large (correlations between .87 and .99). The common shared environmental factor likewise contributes greatly to the stability of individual differences over time (correlations equal .82, .97, and .79).

Table VI lists the contributions of genetic and environmental effects to the between-occasion subtest correlations.

Most of the expected phenotypic correlations are estimated between about .40 and .60. The correlations between ages 7 and 10 are larger than those between ages 5 and 7 in four subtests (Exclusion, Discs, Hidden Figures, Idea Production). The decomposition of the correlations reveals that genetic effects are an important source of stability. This is due to both the relative magnitude of the genetic effects at each occasion and to the large genetic correlations over time. On average, genetic effects explain around 74% of the observed stability in subtest performance from age 5 to age 7 and over 80% from age 7 to age 10.

As already suggested by the results in Table IV, specific genetic factors contribute substantially to the expected correlations in the subtests Discs and Idea Production. With the exception of the subtest Verbal Meaning, the contribution of shared environment to the stability is relatively small. Thus, although the correlations between the shared environmental factors are large (Table V), the relatively minor and, over time, diminishing contributions of the shared environment at each occasion (Table IV) render the contribution to the phenotypic stability of subtest performance.

DISCUSSION

The aim of the present study is to investigate the differentiation hypothesis at the level of the genetic and environmental covariance structure. To this end, multivariate data were analyzed of 209 twin pairs, who were tested at ages 5, 7, and 10 years. As a point of departure of comparison of factor structures over time, we considered configural invariance (Horn and McArdle, 1992). Configural invariance implies that the same observed variables load on the same factors across measurement occasions. Configural invariance was established in phenotypic factor analyses of the data observed at each occasion. An oblique two-common-factor model fit the data

Table VI. Within-Trait, Across-Age Expected Correlations, Partitioned into Genetic and Environmental Contributions

	Age interval	E(r_p)	A group	A specifics	C
Exclusion	Age 5 to 7	.41	.30	.03	.08
	Age 7 to 10	.50	.43	.01	.06
Discs	Age 5 to 7	.47	.25	.19	.03
	Age 7 to 10	.53	.28	.20	.05
Hidden Figures	Age 5 to 7	.27	.14	.02	.11
	Age 7 to 10	.47	.31	.05	.11
Verbal Meaning	Age 5 to 7	.41	.10	.10	.21
	Age 7 to 10	.43	.14	.06	.23
Learning Names	Age 5 to 7	.61	.34	.08	.19
	Age 7 to 10	.58	.35	.13	.10
Idea Production	Age 5 to 7	.39	.04	.31	.04
	Age 7 to 10	.47	.04	.39	.04

Note: E(r_p) = expected phenotypic correlation based on final model, being the sum of $a_i * r_a * a_k$ (calculated separately for A group and A specifics), and $c_i * r_c * c_k$ (C). Parameters a and c represent the unstandardized path loadings at a specific age, i and k represent the initial and subsequent test occasion, r represents the correlation between factors. The unique environment does not contribute to stability of subtest performance.

relatively well. The common factors represented verbal (Verbal Meaning, Learning Names, Idea Production) and nonverbal cognitive abilities (Exclusion, Discs, Hidden Figures). Configural invariance was also found to be tenable in the genetic and environmental factor models, established in analyses of the cross-sectional data. The unique environment did not contribute to the covariance between the subtests. The genetic factor structure comprised two correlated, verbal and nonverbal common factors and six subtest-specific factors. A single common factor without residuals accounted for the covariance structure of the shared environmental effects. The comparison of the phenotypic results with the genetic and environmental results indicates that the phenotypic two-factor solution is due mainly to the genetic structure.

The analyses of the longitudinal dataset produced various interesting results. With respect to the structure and importance of the genetic and environmental effects, the results presented here are in agreement with those presented by other twin studies (Cardon and Fulker, 1993; McCartney *et al.,* 1990; McGue *et al.,* 1993; Patrick, 2000). When outcomes from two large infant studies (Price *et al.,* 2000; Reznick *et al.,* 1997) are combined with the outcomes obtained here, it emerges that between the age of 2 and the age of 5, and up to the age of 10, genes become increasingly important in explaining variation in verbal and nonverbal intellectual abilities. Genetic effects not only increase in importance as a source of individual differences at specific ages but also as a source of stability over ages.

Shared environmental effects are highly stable from infancy to middle and late childhood, but diminish in importance during this time period. At age 10, the relative influence of the familial environment is estimated at less than 10% of the total variance of five subtests. In our previous paper (Rietveld *et al.,* 2000) we reported a correlation of .46 between parents for highest attained education. Given the association with IQ, marital assortment for educational level may have contributed to the detected shared environmental effects. However, we do not believe that assortment can account for the presence of shared environmental effects in these data. First, on the assumption that assortment is phenotypic, the induced genetic correlation will be lower than .46. Second, the phenotypic correlation between educational attainment and general intelligence is .5 (median value; e.g., Kline, 1991). Taken together, these findings suggest that the genetic correlation for IQ tests induced by assortment for educational attainment will be low. In addition, we find that shared environmental effects decrease over time. This decrease is known to continue beyond 10 years. This is inconsistent with the possible consequences of phenotypic assortment at the genetic and environmental level.

The unique environmental influences are specific to each age and to each subtest. Although the unique environmental effects lack stability, they do remain the most important environmental influence in explaining individual differences in cognitive abilities. Specific factors account for one third to half of the total variance for each subtest at each age. Numerous studies of developmental intelligence have pointed at the importance of the environment that make children in the same family different from one another (for discussion see Plomin and Daniels, 1987; Plomin *et al.,* 1996; Turkheimer and Waldron, 2000). The large estimates for unique environmental effects at each age cannot be explained by the degree of unreliability of the administered RAKIT subtests (Bleichrodt *et al.,* 1984). From the reported internal consistencies that vary between about .70 and .90, it is suggested that a substantial part of the variance unique to the individual must result from influences different from test unreliability. The unique environmental effect may include an effect due to the interaction between genetic and environmental deviations (Boomsma and Martin, 2002; Plomin *et al.,* 1977). This interaction implies that variations in the environment affect individuals differently depending on their genotypes. If such an interaction is between genetic effects and unique environmental effects, the unique environmental effects are overestimated.

The essence of the differentiation hypothesis, as discussed in the Introduction, concerns the correlations among the subtest scores. The phenotypic correlations are expected to decrease during normal cognitive development. In investigating this hypothesis, we must consider genetic and environmental sources of individual differences to determine whether differentiation has taken place. First, as noted above, we find that the correlations between the common genetic factors at each occasion display little support for this hypothesis at the genetic level (.25, .28 and .30; see Table V). Second, we find that the shared environmental effects are either constant or decrease over time. The shared environmental common factor at each occasion is certainly a source of correlation among the subtest scores. Any decline in these effects over time may thus be viewed as a contribution to differentiation by decreasing the phenotypic correlations among subtests. Clear decrease in factor loadings is limited to the subtests Exclusion, Hidden Figures, and Learning Names. However, the shared environment in general explains relatively little of the variance, and thus the differentiation due to the decrease in shared environmental effects is thus weak at best. The results concerning the unique environmental effects are straightforward. These effects are specific to the subtests and they do not contribute to the stability of individual differences over time. An increase in the unique environmental effect over time will result in differentiation at the phenotypic level, by lowering the phenotypic intersubtest correlations. The results show that there is certainly no increase in the relative contributions to the variance. We emphasize that these results are limited to the age range from 5 to 10 years; we cannot extrapolate beyond the age of 10. However, within this age range, we find little support for the differentiation hypothesis at the genetic or the environmental level.

ACKNOWLEDGMENTS

This work was financially supported by the USF (grant number 96/22). The research of Conor Dolan was made possible by a fellowship of the Royal Netherlands Academy of Arts and Sciences. The Netherlands Organization for Scientific Research is acknowledged for funding the work of Caroline van Baal (575-65-052).

REFERENCES

Bartels, M., Rietveld, M. J. H., Van Baal, G. C. M., and Boomsma, D. I. (2002). Genetic and environmental influences on the development of intelligence. *Behavior Genetics* **32**:237–249.

Bickley, P. G., Keith, T. Z., and Wolfle, L. M. (1995). The three-stratum theory of cognitive abilities: test of the structure of intelligence across the life span. *Intelligence* **20**:309–328.

Bishop, E. G., Cherny, S. S., Corley, R., Plomin, R., DeFries, J. C., and Hewitt, J. K. (2003). Development genetic analysis of general cognitive ability from 1 to 12 years in a sample of adoptees, biological siblings, and twins. *Intelligence* **31**:31–49.

Bleichrodt, N., Drenth, P. J. D., Zaal, J. N., and Resing, W. C. M. (1984). *Revisie Amsterdamse Kinder Intelligentie Test* [Revised Amsterdam Child Intelligence Test]. Lisse, The Netherlands: Swets & Zeitlinger B. V.

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

Boomsma, D. I., and Molenaar, P. C. M. (1986). Using LISREL to analyze genetic and environmental covariance structure. *Behavior Genetics* **16**:237–250.

Boomsma, D. I., and Molenaar, P. C. M. (1987). The genetic analysis of repeated measures. I. Simplex models. *Behavior Genetics* **17**:111–123.

Boomsma, D. I., and Van Baal, G. C. M. (1998). Genetic influences on childhood IQ in 5- and 7-year-old Dutch twins. *Developmental Neuropsychology* **14**:115–126.

Cardon, L. R., and Fulker, D. W. (1993). Genetics of specific cognitive abilities. In: Plomin, R., and McClearn, G. E. (eds.), *Nature, Nurture & Psychology,* American Psychological Association, Washington, D.C., pp. 99–120.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge: Cambridge University Press.

Casto, S. D., DeFries, J. C., and Fulker, D. W. (1995). Multivariate genetic analysis of Wechsler Intelligence Scale for Children—Revised (WISC-R) factors. *Behavior Genetics* **25**:25–32.

Central Bureau of Statistics (2001), *Enquête Beroepsbevolking 1999* [Official Inquiry of Working Population 1999], Central Bureau of Statistics, Voorburg/Heerlen, The Netherlands (information obtained at http://statline.cbs.nl/statweb/index_NL.stm).

Cunningham, W. R. (1991). Issues in factorial invariance. In L. M. Collings and J. L. Horn (eds.), *Best methods for the analysis of change* (pp. 106–113). Washington, D.C.: American Psychological Association.

Deary, I. J., Egan, V., Gibson, G. J., Austin, E. J., Brand, C. R., and Kellaghan, T. (1996). Intelligence and the differentiation hypothesis. *Intelligence* **23**:105–132.

DeFries, J. C., and Fulker, D. W. (1986). Multivariate behavioral genetics and development. *Behavior Genetics* **16**:1–10.

Daniels, M., Devlin, B., and Roeder, K. (1997). Of genes and IQ. In: *Intelligence, genes & success. Scientists respond to the bell curve*. Devlin, B., Fienberg, S. E. Resnick, D. P., and Roeder, K., (eds.). New York: Springer Verlag.

Eaves, L. J. (1972). Computer simulation of sample size and experimental design in human psychogenetics. *Psychological Bulletin* **77**:144–152.

Eaves, L. J., Long, J., and Heath, A. C. (1986). A theory of developmental change in quantitative phenotypes applied to cognitive development. *Behavior Genetics* **16**:143–162.

Fulker, D. W., Cherny, S. S., and Cardon, L. R. (1993). Continuity and change in cognitive development. In Plomin, R., and McClearn, G. E., (eds.), *Nature, nurture & psychology*. Washington, D.C.: American Psychological Association.

Garrett, H. E. (1946). A developmental theory of intelligence. *American Psychologist* **1**:372–378.

Guttman, L. (1954). A new approach to factor analysis: The radex. In P. F. Lazarsfeld (ed.), *Mathematical thinking in the social sciences* (pp. 258–348). Glencoe, IL: The Free Press.

Hay, D. A., and O'Brien, P. J. (1983). The La Trobe twin study: A genetic approach to the structure and development of cognition in twin children. *Child Development* **54**:317–330.

Hewitt, J. K., Eaves, L. J., Neale, M. C., and Meyer, J. M. (1988). Resolving causes of developmental continuity or "tracking." I. Longitudinal twin studies during growth. *Behavior Genetics* **18**:133–152.

Horn, J. L. (1991). Comments on issues in factorial invariance. In L. M. Collings and J. L. Horn (eds.), *Best methods for the analysis of change* (pp. 114–125). Washington, D.C.: American Psychological Association.

Horn, J. L., and McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research* **18**:117–144.

Juan-Espinosa, M., García, L. F., Colom, R., and Abad, F. J. (2000). Testing the age-related differentiation hypothesis through the Wechsler's scales. *Personality and Individual Differences* **29**:1069–1075.

Kline, P. (1991). *Intelligence: The psychometric view*. London: Routledge.

Loehlin, J. C. (1992). *Latent Variable Models: An Introduction to Factor, Path, and Structural Analysis* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

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

McArdle, J. J. (1986). Latent variable growth within behavior genetic models. *Behavior Genetics* **16**:163–200.

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

McCartney, K., Harris, M. J., and Bernieri, F. (1990). Growing up and growing apart: A developmental meta-analysis of twin studies. *Psychological Bulletin* **107**:226–237.

McGue, M., Bouchard, T. J., Jr., Iacono, W. G., and Lykken, D. T. (1993). Behavioral genetics of cognitive ability: A life-span perspective. In: Plomin, R., and McClearn, G. E. (eds.), *Nature, nurture, & psychology* (pp. 59–76). Washington, D.C.: American Psychological Association.

Meyer, J. M., Silberg, J. L., Eaves, L. J., Maes, H. H., Simonoff, E., Pickles, A., Rutter, M. L., and Hewitt, J. K. (1999). Variable age of gene expression: Implications for developmental genetic models. In LaBuda, M. C., and Grigorenko, E. L. (eds.), *On the way to individuality: Current methodological issues in behavioral genetics* (pp. 23–52). Commack: Nova Science Publishers, Inc.

Neale, M. C., Boker, S. M., Xie, G., and Maes, H. H. (1999). *Mx: Statistical Modeling* (5th Ed.), Medical College of Virginia Department of Psychiatry, Box 126 MCV, Richmond, VA 23298.

Neale, M. C., and Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Boston, MA: Kluwer.

Patrick, C. L. (2000). Genetic and environmental influences on the development of cognitive abilities: Evidence from the field of developmental behavior genetics. *Journal of School Psychology* **38**:79–108.

Plomin, R. (1983). Developmental Behavior Genetics. *Child Development* **54**:253–259.

Plomin, R., and Daniels, D. (1987). Why are children in the same family so different from one another? *Behavioral and Brain Sciences* **10**:1–60.

Plomin, R., and DeFries, J. C. (1981). Multivariate behavioral genetics and development: Twin studies. In Gedda, L., Parisi, P., and Nance, W. E., (eds.). *Twin research 3: Part B* (pp. 25–33). New York: Alan R. Liss.

Plomin, R., DeFries, J. C., and Loehlin, J. C. (1977). Genotype-environment interaction and correlation in the analysis of human behavior. *Psychological Bulletin* **84**:309–322.

Plomin, R., DeFries, J. C., McClearn, G. E., and McGuffin, P. (2001). *Behavioral genetics* (4th ed.). New York: W. H. Freeman.

Plomin, R., Pedersen, N. L., Lichtenstein, P., and McClearn, G. E. (1994). Variability and stability in cognitive abilities are largely genetic later in life. *Behavior Genetics* **24**:207–215.

Plomin, R., Petrill, S. A., and Cutting, A. L. (1996). What genetic research on intelligence tells us about the environment. *Journal of Biosocial Science* **28**:587–606.

Posthuma, D., and Boomsma, D. I. (2000). A note on the statistical power in extended twin designs. *Behavior Genetics* **30**:147–158.

Price, T. S., Eley, T. C., Dale, P. S., Stevenson, J., Saudino, K., and Plomin, R. (2000). Genetic and environmental covariation between verbal and nonverbal cognitive development in infancy. *Child Development* **71**:948–959.

Reinert, G. (1970). Comparative factor analytic studies of intelligence throughout the human life-span. In Goulet, L. R., and Baltes, P. B., (eds.). *Life-span developmental psychology* (pp. 467–484). New York: Academic Press.

Reznick, J. S., Corley, R., and Robinson, J. A. (1997). A longitudinal twin study of intelligence in the second year. *Monographs of the Society for Research in Child Development,* serial no. 249, **62**(1).

Rietveld, M. J. H., Posthuma, D., Dolan, C. V., and Boomsma, D. I. (in press). ADHD: Sibling interaction or dominance, an evaluation of statistical power. *Behavior Genetics*.

Rietveld, M. J. H., Van Baal, G. C. M., Dolan, C. V., and Boomsma, D. I. (2000). Genetic factor analyses of specific cognitive abilities in 5-year-old Dutch children. *Behavior Genetics* **30**:29–40.

Schaie, K. W. (1994). The course of adult intellectual development. *American Psychologist* **49**:304–313.

Schaie, K. W., Maitland, S. B., Willis, S. L., and Intrieri, R. C. (1998). Longitudinal invariance of adult psychometric ability factor structures across 7 years. *Psychology and Aging* **13**:8–20.

Turkheimer, E., and Waldron, M. C. (2000). Nonshared environment: A theoretical, methodological, and quantitative review. *Psychological Bulletin* **126**:78–108.

Van Baal, G. C. M., de Geus, E. J. C., and Boomsma, D. I. (1996). Genetic architecture of EEG power spectra in early

life. *Electroencephalography and Clinical Neurophysiology* **98**:502–514.
Vernon, P. E. (1976). Development of intelligence. In Hamilton, V., and Vernon, M. D., (eds.), *The development of cognitive processes* (pp. 507–541). London: Academic Press Inc.
Werdelin, I., and Stjernberg, G. (1995). Age differences in factorial structure: A study of the "differentiation hypothesis." *Interdisciplinaria* **12**:79–97.
Wilson, R. S. (1983). The Louisville twin study: Developmental synchronies in behavior. *Child Development* **54**:298–316.
Wohlwill, J. F. (1973). *The Study of Behavior Development*. New York: Academic Press.
Wothke, W. (2000). Longitudinal and multigroup modeling with missing data. In Little, T. D., Schnabel, K. U., and Baumert, J., (eds.), *Modeling longitudinal and multilevel data; Practical issues, applied approaches and specific examples*. Mahwah, N.J.: Lawrence Erlbaum Associates, Publishers.

Edited by Stacey Cherny

Twin Research (2000) 3, 134–141

www.nature.com/tr

Zygosity diagnosis in young twins by parental report

MJH Rietveld[1], JC van der Valk[1,2], IL Bongers[1], TM Stroet[1], PE Slagboom[3] and DI Boomsma[1]

[1]*Vrije Universiteit, Department of Biological Psychology, Amsterdam*
[2]*Erasmus University, Department of Child and Adolescent Psychiatry, Rotterdam*
[3]*TNO-PG, Gaubius Laboratory, Department of Vascular and Connective Tissue Research, Leiden, The Netherlands*

This study reports on zygosity determination in twins of childhood age. Parents responded to questionnaire items dealing with twin similarity in physical characteristics and frequency of mistaking one twin for another by parents, relatives and strangers. The accuracy of zygosity diagnosis was evaluated across twins aged 6, 8, and 10 and across parents. In addition, it was examined whether the use of multiple raters and the use of longitudinal data lead to an improvement of zygosity assignment. Complete data on zygosity questions and on genetic markers or blood profiles were available for 618 twin pairs at the age of 6 years. The method used was predictive discriminant analyses. Agreement between zygosity assigned by the replies to the questions and zygosity determined by DNA markers/blood typing was around 93%. The accuracy of assignment remained constant across age and parents. Analyses of data provided by both parents and collected over multiple ages did not result in better prediction of zygosity. Details on the discriminant function are provided. *Twin Research* (2000) **3**, 134–141.

Keywords: twin zygosity, childhood, questionnaire, review, discriminant analysis

Introduction

In 1927, Siemens[1] suggested that the diagnosis of zygosity in twins can take place by evaluating the degree of resemblance on genetically determined traits. Development of this method resulted in the frequent use of questionnaires, often including those criteria originally proposed by Siemens, for example.[2] Several studies have shown that the establishment of zygosity based on mailed questionnaires is of considerable accuracy, with around 95% correctly classified compared with blood or DNA typing. Studies on the diagnosis of zygosity by mailed questionnaires are summarised in Appendix 1.[3–23]

The purpose of this paper is twofold. First, the validity of zygosity classification across childhood is examined in a large sample. One might expect the physical dissimilarity between dizygotic twins to become more obvious as they grow up. If so, the accuracy of classification is likely to improve with increasing age of the participants. A few studies have reported on this issue by evaluating the precision of zygosity diagnosis between samples varying in age,[8,19,23] and by test–retest estimation.[9] With the exception of the study of Cohen *et al*,[9] the findings are suggestive of an increased precision in zygosity prediction for older participants. However, findings may have suffered from a lack of statistical power due to a relatively small number of co-operating twins and parents.

To our knowledge there are no studies investigating this issue in a longitudinal sample. Since the availability of longitudinal data of various birth cohorts is increasing in several twin registers,[24] the establishment of zygosity incorporating longitudinal data deserves our attention. The Netherlands Twin Register collects questionnaire data on zygosity items at multiple ages in the same children by parental report. By making use of this longitudinal dataset it is possible to examine whether analysing all available data collected at different ages increases the precision of classification or whether it is sufficient or possibly advisable to rely on information obtained at a specific age only. We are especially interested to determine if reliable classification of zygosity can take place as early as age 6.

The second objective is to investigate how to make optimal use of information provided by multiple carers. The majority of participating families registered with the Netherlands Twin Register returns two completed questionnaires, usually filled in by the mother and father of the twin pair. In other twin studies of young children, typically the mother is used as primary informant.[17] It is of interest to find out whether the precision of the establishment of zygosity can further improve if information provided by a second informant is included in the analyses.

The Netherlands Twin Register has access to complete data on bloodgroup typing or DNA polymorphism and zygosity questionnaires collected in a sample of 618 twin pairs at age 6. This large number

Correspondence: MJH Rietveld, Vrije Universiteit, Department of Biological Psychology, De Boelelaan 1111, 1081 HV Amsterdam, The Netherlands. Fax: +31 20 444 8832; Tel: +31 20 444 8812; E-mail: mjh.rietveld@psy.vu.nl
Received 28 February 2000; accepted 24 March 2000

of participants provides sufficient statistical power to investigate the above issues.

Materials and methods

Subjects

The Netherlands Twin Register (NTR) is a population-based register, which contains 40%–50% of all multiple births after 1986.[25] As part of a current longitudinal study on the development of behaviour problems, two questionnaires are sent to the registered parents or primary carers at multiple points in time with an average interval of 2 years. The present study used information by parental report on twin similarity and twin confusion at three ages in childhood, for cohorts born between 1986 and 1991. On the first occasion of data collection, around the sixth birthday of the twins (mean = 6.36 years, SD = 0.95), information on zygosity by report of the father was not requested. At the second and third assessment, age 8 (mean = 7.90 years, SD = 0.50) and age 10 (mean = 10.27 years, SD = 0.40) respectively, both parents provided information on zygosity items. For this study, only pairs of same sex with DNA/blood zygosity data were included in the analyses (n = 691 pairs). Twin pairs with missing items on the parental zygosity questions were excluded. Table 1 reports on the numbers of same sex twin pairs with complete data on the zygosity items and DNA/blood typing at each age.

Complete longitudinal data were available from 253 mothers (age 6, 8, and 10), and from 224 fathers (age 8 and 10). Data from both raters were collected in 316 twin pairs at age 8, and in 257 twin pairs at age 10. The sample participating in this study was predominantly of Caucasian origin, with around 2% classified into other ethnic groups.

Zygosity questionnaire

The questionnaire used in the present study asked for information regarding similarity of the children and experiences of mistaking one for another (Appendix 2). When the twins were aged 6, parents provided information on eight items. In addition, a question concerning knowledge of zygosity classification based on DNA/blood testing was included. This item was used to identify those families with knowledge of zygosity prior to completing the questions. Two more items were added to the zygosity questionnaire at the second and third measurement occasion.

Table 1 Number of twin pairs participating in the present study

	Age 6 Mother	*Age 8 Mother*	*Age 8 Father*	*Age 10 Mother*	*Age 10 Father*
Questionnaire and DNA/blood data	618	394	335	324	279
MZ	388	243	210	200	163
DZ	230	151	125	124	116

Genotyping and blood polymorphism

A total of 691 same sex twin pairs participated in DNA/blood testing; 62% donated blood samples for analyses of blood grouping profiles and 38% provided a mouth swab sample for DNA isolation. Zygosity determination was performed using eight highly polymorphic di-, tri- and tetranucleotide genetic markers. The zygosity testing included a multiplex PCR of markers D2S125, D8S1130, D1S1609, D5S816 and a second multiplex reaction of markers 15 ActC, D21S1437, D7S2846, and D10S1423. These two multiplex PCR reactions were performed essentially by the protocol provided in the website of the Marshfield Institute (http://www.marshmed.org/genetics/). For the purpose of zygosity determination based on blood grouping profiles, red cells were typed with test sera for the following red cell blood group antigens: AB, CcDEe, MNSs, P_1, Kk, Kp^aKp^b, Fy^aFy^b, Jk^aJk^b, Lu^aLu^b. More details on the collection and treatment of these blood samples are given by Van Dijk *et al*.[26]

Statistical procedures

All parents of twins with DNA/blood data were informed about the zygosity results. Since the employment of DNA/blood testing varied across age, two groups of families could be distinguished. One group of parents with knowledge of the DNA/blood test results before completion of the questionnaire, and one group of parents whose twin pair had not yet participated in the DNA/blood testing. Since prior knowledge of the test results may affect responses to the zygosity questions, it was established first whether the two groups of parents differed in their item response pattern. If so, generalisation of the application of the statistical function to samples for which no information on biological indices is available is seriously hampered. The tests were performed on each item separately by employment of χ^2 tests.

Predictive discriminant analysis was used for classifying subjects into MZ and DZ groups.[27,28] In the present study, the discriminant analysis generated a linear function of the weighted sum of the questionnaire items with the weightings chosen, such that the distinction between MZ and DZ twins was optimal. The estimated success of classification or hit rate is the proportion of correctly classified

observations in the sample. It is sometimes argued that this hit rate is optimistically biased since the classification rule is derived from and applied to the same sample. This bias can be avoided in two ways, either through use of large samples or through application of an external classification analysis. In this study, both routes are taken. As a criterion for sample size, it is proposed that the minimum of observations in the smallest group should be at least five times the number of questionnaire items. As can be seen in Table 1, this requirement was easily met by each individual dataset. The leave-one-out procedure was chosen as the preferred external analysis. This method omits an observation, recalculates the classification rule from the remaining observations, classifies the deleted observation, and repeats these steps for each observation in the sample. The number of deleted observations correctly classified are counted and reported as cross-validated hit rates. Considering the proportion of same-sex MZ and DZ twins in the population, equal prior probabilities of group membership were used. To define the underlying construct that the discriminant function represents, inspection of the correlations between the discriminant function and each of the questionnaire variables was performed. The discriminant function and descriptive statistics were calculated using Statistical Package for Social Sciences/ Windows 9.0.

Results

At age 6, out of 618 pairs with DNA/blood data, 411 mothers knew the result of zygosity testing and 199 mothers had not yet received a request for DNA/ blood testing for their twins. Eight mothers had not answered the question. The ratio MZ : DZ was equal in both groups and data were pooled across zygosities to examine mothers' responses between groups. A difference in response pattern was observed for 1 item only, 'do strangers have difficulty telling them apart?' (χ^2 = 5.17 (1), P = 0.02). A positive answer was given by 65% of those mothers who were ignorant of zygosity, compared with 75% among mothers with knowledge of the DNA/blood test result. Overall, the two groups did not seem to differ allowing the discriminant function to be applied to both groups simultaneously.

A summary of the results of the first series of discriminant analyses is given in Table 2. Each analysis indicated a very accurate hit rate. Between 91.6% and 94.2% of all twin pairs were assigned the correct zygosity by the discriminant function. The precision of classification was not equally distributed across zygosities. Irrespective of age, correct classification for MZ twins was estimated around 97%, whereas around 88% of DZ twins were identified correctly.

Table 2 Classification results by use of discriminant function analyses

		Age 6 Mother	*Age 8 Mother*	*Age 8 Father*	*Age 10 Mother*	*Age 10 Father*
Correctly classified	MZ	96.6%	95.1%	97.1%	97.5%	96.9%
	DZ	90.0%	86.8%	85.6%	88.7%	89.7%
Cross-validated	Total	94.2%	91.6%	91.9%	92.6%	93.9%

Next, twin pairs with longitudinal questionnaire data were considered. The analysis of data collected at age 6, 8, and 10 by report from the mother resulted in a hit rate of 93.7%. Analysis of fathers' reports collected at the twins' age of 8 and 10 yielded a correct classification of 94.2%. Finally, data from mother and father were analysed jointly. At age 8, 93.4% of all twin pairs were classified correctly. A hit rate of 93.8% was obtained at age 10.

The above cross-validated hit rates indicated a minimal difference in the precision of assignment across the use of various datasets. The use of multiple raters and longitudinal data did not lead to an increased precision of zygosity prediction. Because the majority of twin studies are performed within cross-sectional designs, we believe it is of much practical use to report upon the discriminant function coefficients resulting from the first series of analyses. These parameter values together with the associated classification scores are given in Appendix 3. For interpreting the discriminant function, we have listed the correlations between each function and each questionnaire item in Table 3.

Across age and parent, the majority of the correlations ranged from 0.50 to 0.80. Identification of those questionnaire items that show the largest overlap with the function helps to determine the underlying construct that the discriminant function represents. The zygosity questionnaire was developed along two dimensions, similarity of physical characteristics and confusion of identity. At either age and for either parent, the most informative correlations were not clustered in a sense that the function could easily be defined along one of these dimensions. Closer inspection revealed a few interesting details. With the exception of item 1 (facial appearance) and item 2 (hair colour), a relatively large degree of overlap was observed between mothers and fathers within age 8 and age 10 of the twins. Looking at the ranking of the items, parents evaluated the questions in the same general manner. When the percentage of correctly classified twins was taken into consideration, this indicated that parents are interchangeable in assessing identity and fraternity in their children. Another interesting finding was the very small correlation found for item 5 ('peas in a pod'). In contrast to numerous other studies, for example,

Table 3 Correlations between discriminant function and individual questionnaire items

Item	*Age 6 Mother*	*Ranking*	*Age 8 Mother*	*Ranking*	*Age 8 Father*	*Ranking*	*Age 10 Mother*	*Ranking*	*Age 10 Father*	*Ranking*
Facial appearance	0.72	1	0.67	3	0.72	2	0.62	6	0.66	3
Hair colour	0.67	3	0.70	2	0.67	4	0.71	2	0.58	6
Face colour	0.66	4	0.63	6	0.65	6	0.68	5	0.63	5
Eye colour	0.52	6	0.53	7	0.50	7	0.51	7	0.52	7
'Two peas'	0.47	7	0.46	8	0.43	8	0.39	8	0.40	8
Mother/father	0.32	8	0.27	9	0.28	9	0.24	9	0.28	9
Family members	0.68	2	0.64	4	0.66	5	0.70	3	0.63	4
Strangers	0.62	5	0.64	5	0.71	3	0.75	1	0.82	1
Photograph			0.15	10	0.15	10	0.12	10	0.23	10
Hair structure			0.76	1	0.75	1	0.70	4	0.68	2

Magnus *et al*,[16] this item was of minor importance in defining the discriminant function. Even smaller correlations were observed for item 6 (confusion by mother or father) and item 9 (tell twins apart in photograph). The association among these three items seems obvious given that these questions rely on parental impression of global similarity and parental confusion of twins' identities. Apparently, parents themselves did not have difficulties in telling who is who.

Discussion

The primary focus of this study was to evaluate the accuracy of zygosity determination in young children. As young as age 6, the precision in zygosity prediction was high, with 94% agreement between zygosity assigned by the parental replies to the questionnaire items and zygosity determined by blood typing or analyses of genetic markers. It was found that the accuracy of classification remained stable across childhood. The suggestion that determination improves with increasing age due to more obvious dissimilarities in dizygotic twin pairs was not confirmed. It was also found that mothers and fathers were equally effective in diagnosing their children.

Although the questionnaire items allow an accurate determination of zygosity, the accuracy resulting from the discriminant analyses was not equally distributed in monozygotic and dizygotic pairs. At each age and for both parents, a bias towards classification as monozygotic twins took place. This may have resulted either from a tendency by parents to overestimate similarities in their twin children or from a lack of sensitivity of these questions to detect fraternity. The former case seems less plausible, considering assessment of parental replies to a question that deals with their personal opinion of the twins' zygosity. This item is included in a questionnaire sent to parents shortly after registration with the NTR (before the twins' first birthday). Correct in 80% of the cases, parents misclassified true MZ twins more than four times as often as true DZ. This result may reflect either the fact that parents are misinformed by physicians or the parents' wish for fraternity, or a combination of both. A preference towards labelling a twin as dizygotic is commonly found both by use of parental report, as in Cohen *et al*[9] and self report.[29]

The sample used in the analyses was mainly Caucasian. This may imply that the use of the zygosity questionnaire and the application of the discriminant functions do not generalise to groups of non-Caucasian ethnic origin.

Concluding, the use of the zygosity questions and the employment of discriminant analysis as multivariate tool for classification seem of value in determining zygosity in young twins.

Acknowledgements

This work was financially supported by USF grant 96/22. The work of Jolande van der Valk was made possible by grant M165 from the Sophia Foundation for Medical Research.

References

1 Siemens HW. The diagnosis of identity in twins. *J Hered* 1927; **18**: 201–209.
2 Goldsmith HH. A zygosity questionnaire for young twins: A research note. *Behav Genet* 1991; **21**: 257–269.
3 Cederlöf R, Friberg L, Jonsson E, Kaij L. Studies on similarity diagnosis in twins with the aid of mailed questionnaires. *Acta Genet Stat Med* 1961; **11**: 338–362.
4 Nichols RC, Bilbro WC. The diagnosis of twin zygosity. *Acta Genet Stat Med* 1966; **16**: 265–275.
5 Jablon S, Neel JV, Gershowitz H, Atkinson GF. The NAS-NRC twin panel: Methods of construction of the panel, zygosity diagnosis, and proposed use. *Am J Hum Genet* 1967; **19**: 133–161.
6 Hauge M, Harvald B, Fischer M, Gotlieb-Jensen K, Juel-Nielsen N, Raebild I, Shapiro R, Videbech T. The Danish Twin Register. *Acta Genet Med Gemellol (Roma)* 1968; **17**: 315–332.
7 Schoenfeldt LF. A comparison of two analytic procedures for estimating twin zygosity. *Hum Hered* 1969; **19**: 343–353.

8 Cohen DJ, Dibble E, Grawe JM, Pollin W. Separating identical from fraternal twins. *Arch Gen Psychiatry* 1973; **29**: 465–469.
9 Cohen DJ, Dibble E, Grawe JM, Pollin W. Reliably separating identical from fraternal twins. *Arch Gen Psychiatry* 1975; **32**: 1371–1375.
10 Martin NG, Martin PG. The inheritance of scholastic abilities in a sample of twins. I. Ascertainment of the sample and diagnosis of zygosity. *Ann Hum Genet* 1975; **39**: 213–218.
11 Kasriel J, Eaves L. The zygosity of twins: Further evidence on the agreement between diagnosis by blood groups and written questionnaires. *J Biosoc Sci* 1976; **8**: 263–266.
12 Sarna S, Kaprio J, Sistonen P, Koskenvuo M. Diagnosis of twin zygosity by mailed questionnaire. *Hum Hered* 1978; **28**: 241–254.
13 Torgersen S. The determination of twin zygosity by means of a mailed questionnaire. *Acta Genet Med Gemellol (Roma)* 1979; **28**: 225–236.
14 King M-C, Friedman GD, Lattanzio D, Rodgers G, Lewis AM, Dupuy ME, Williams H. Diagnosis of twin zygosity by self-assessment and by genetic analysis. *Acta Genet Med Gemellol (Roma)* 1980; **29**: 121–126.
15 Sarna S, Kaprio J. Use of multiple logistic analysis in twin zygosity diagnosis. *Hum Hered* 1980; **30**: 71–80.
16 Magnus P, Berg K, Nance WE. Predicting zygosity in Norwegian twin pairs born 1915-1960. *Clin Genet* 1983; **24**: 103–112.
17 Bønnelykke B, Hauge M, Holm N, Kristoffersen K, Gurtler H. Evaluation of zygosity diagnosis in twin pairs below age seven by means of a mailed questionnaire. *Acta Genet Med Gemellol (Roma)* 1989; **38**: 305–313.
18 Eisen S, Neuman R, Goldberg J, Rice J, True W. Determining zygosity in the Vietnam Era Twin Registry: an approach using questionnaires. *Clin Genet* 1989; **35**: 423–432.
19 Ooki S, Yamada K, Asaka A, Hayakawa K. Zygosity diagnosis of twins by questionnaire. *Acta Genet Med Gemellol (Roma)* 1990; **39**: 109–115.
20 Ooki S, Yamada K, Asaka A. Zygosity diagnosis of twins by questionnaire for twins' mothers. *Acta Genet Med Gemellol (Roma)* 1993; **42**: 17–22.
21 Spitz E, Moutier R, Reed T, Busnel MC, Marchaland C, Roubertoux PL, Carlier M. Comparative diagnosis of twin zygosity by SSLP variant analysis, questionnaire, and dermatoglyphic analysis. *Behav Genet* 1996; **26**: 55–63.
22 Charlemaine C, Duyme M, Aubin J-T, Guis F, Marquiset N, de Pirieux I, Strub N, Brossard Y, Jarry G, Le Groupe Romulus, Frydman R, Pons JC. Twin zygosity diagnosis by mailed questionnaire below age twelve months. *Acta Genet Med Gemellol (Roma)* 1997; **46**: 147–156.
23 Chen WJ, Chang H-W, Wu M-Z, Lin CCH, Chang C, Chiu Y-N, Soong W-T. Diagnosis of zygosity by questionnaire and polymerase chain reaction in young twins. *Behav Genet* 1999; **29**: 115–124.
24 Boomsma DI. Twin registers in Europe: an overview. *Twin Research* 1998; **1**: 34–51.
25 Boomsma DI, Orlebeke JF, Van Baal GCM. The Dutch twin register: growth data on weight and height. *Behav Genet* 1992; **22**: 247–251.
26 Van Dijk BA, Boomsma DI, De Man AJM. Blood group chimerism in human multiple births is not rare. *Am J of Med Genet* 1996; **61**: 264–268.
27 Huberty CJ. *Applied Discriminant Analysis*. John Wiley & Sons: New York, 1994.
28 Panel on Discriminant Analysis, Classification, and Clustering. Discriminant analysis and clustering. *Stat Sci* 1989; **4**: 34–69.
29 Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ. A test of the equal-environment assumption in twin studies of psychiatric illness. *Behav Genet* 1993; **23**: 21–27.

Appendix 1 Summary of studies on zygosity determination by written questionnaire

Study	*Subjects*	*Mailed questionnaire*	*Method of classification*[1]	*Results*
Cederlöf, Friberg, Jonsson, Kaij, 1961[3]	200 pairs, age 35–75	1 similarity item,[2] 1 multivariate,[3] confusion item; completed by both twins	decision rules	98% of MZ correct; 92% of DZ correct; 10% of total sample left unclassified
Nichols, Bilbro, 1966[4]	123 pairs, high school juniors	5 similarity items and 1 multivariate confusion item; completed by both twins	decision rules; intuitive decision was made in case the previous method left cases unclassified (7%)	93% of total sample correct
Jablon, Neel, Gershowitz, Atkinson, 1967[5]	232 pairs, age 30–45	A short description of 'identical' and 'non-identical' was given by the investigators, followed by one single item that dealt with twins' own opinion; completed by both twins (complete agreement within pair) or individual twins	Evaluation of zygosity diagnosis was performed on one item only: the joint opinion of a pair, and the opinion of the individual twin	No difference in accuracy, between individual twins and pairs. 89% of MZ correct, 97% of DZ correct
Hauge, Harvald, Fischer, Gotlieb-Jensen, Juel-Nielsen, Raebild, Shapiro, Videbech, 1968[6]	335 pairs, adults	Not clearly specified: multiple similarity items as well as 1 multivariate confusion item; completed by one twin or both twins, or by relatives	decision rules	97% of total sample correct
Schoenfeldt 1969[7]	124 pairs, sample is identical to Nichols, Bilbro, 1966	Identical to Nichols and Bilbro, (1966)	decision rules based on one single score calculated from scores of both twins; discriminant analyses on same single score	decision rules: 92% of total sample correct (cross-validated 79%); discriminant: 88% of total sample correct (cross-validated 88%)
Cohen, Dibble, Grawe, Pollin, 1973[8]	Two samples: (1) 120 pairs, mean age 9.4 (2) 35 pairs mean age 4.2	7 similarity items and 1 multivariate confusion item, completed by the mother. Samples differed in age and in knowledge of zygosity by the mother	discriminant analyses; cutting point on summed raw scores	No difference in response pattern between groups varying in age and informed mothers. Groups were pooled; discriminant: 98% of total sample correct; cutting point: 93% of MZ correct and 73% of DZ correct, with the remaining left unclassified
Cohen, Dibble, Grawe, Pollin, 1975[9]	275 pairs, age 1–6	Identical to Cohen *et al*, (1973), completed by the mother	discriminant analyses; cutting point on summed raw score; principal component factor analysis	hit rate is estimated at 90%
Martin, Martin, 1975[10]	47 pairs, age 15	A description of 'identical' and 'non-identical' was given by the investigator, followed by one single item that dealt with the twins' own opinion; their joint answer had to be confirmed by the parents	Since parents and twins all had to agree on the zygosity of the pair, evaluation of zygosity diagnosis was performed on one item only	100% of total sample correct
Kasriel, Eaves, 1976[11]	178 pairs, adults	1 similarity item and 1 univariate[3] confusion item: completed by both twins	decision rules	96% of total sample correct
Sarna, Kaprio, Sistonen, Koskenvuo, 1978[12]	104 pairs, age 20–69	1 similarity item and 1 univariate confusion item; completed by both twins	deterministic decision tree	93% of total sample correct with 7% unclassified
Torgersen, 1979[13]	215 pairs, age 18–67	1 similarity item and 1 multivariate confusion item; completed by both twins	cutting point on single summed raw score composed of scores of both twins; discriminant analyses on same summed raw score; decision tree	cutting point: 95% of total sample correct; discriminant: 94% of MZ correct, 96% of DZ correct; decision tree: 96% of total sample correct

Continued on next page

Appendix 1 *continued from previous page*

Study	*Subjects*	*Mailed questionnaire*	*Method of classification*[1]	*Results*
King, Friedman, Lattanzio, Rodgers, Lewis, Dupuy, Williams, 1980[14]	173 pairs, adults	1 similarity item that dealt with twins' own opinion; completed by both twins	Evaluation of zygosity diagnosis was performed on one item only	83% of MZ correct, 97% of DZ correct
Sarna, Kaprio, 1980.[15] This study is a follow-up of Sarna *et al*, 1978.[12]	Two samples: (1) 52 pairs previously left unclassified (2) 104 pairs	Identical to Sarna *et al*, (1978), completed by both twins	logistic regression, with (1) 0.50 and (2) 0.70 limit for *a posteriori* probability discriminant analyses	logistic regression: (1) all cases classified with 75% correct of total sample, cross-validated, (2) 100% correct of total sample with 53% left unclassified, cross-validated; discriminant: identical results
Magnus, Berg, Nance, 1983[16]	207 pairs, age 33–61	Originally[4] composed of 13 similarity items, 1 multivariate confusion item, and 1 item reflecting twins' own opinion: completed by one twin or both twins	discriminant analyses applied to 2 groups: (1) data from one twin only, (2) data from both twins. Intrapair means of scores was used in case both twins responded	(1) 96% of total sample correct, cross-validated, (2) 98% of total sample correct, cross-validated
Bønnelykke, Hauge, Holm, Kristoffersen, Gurtler, 1989[17]	125 pairs, age 0.5–6.5	4 similarity items and 1 univariate confusion item; completed by the mother	decision rules	91% of total sample correct, 4% misclassified, and 5% left unclassified
Eisen, Neuman, Goldberg, Rice, True, 1989[18]	4774 male pairs with insufficient blood typing data, adults	Identical to Magnus, (1983), completed by both twins	discriminant analyses as employed by Magnus (1983); 3 types of logistic regression including race-specific analysis	By combining the various methods, 9% of MZ twins were classified incorrectly. Variation in discriminating questions was observed for race
Ooki, Yamada, Asaka, Hayakawa, 1990[19]	Two samples: (1) 189 pairs age 12–16; (2) 93 pairs age 52–77	Identical to Torgersen, (1979), completed by both twins	cutting point on single summed raw score composed of scores of both twins; discriminant analyses on same summed raw score	cutting point: (1) 92% of MZ correct, 88% of DZ correct, (2) 100% of MZ correct, 77% of DZ correct; discriminant: (1) 92% of total sample correct, cross-validated in older sample resulted in 95% correct, (2) 94% of total sample correct, cross-validated in younger sample resulted in 67% correct
Ooki, Yamada, Asaka, 1993[20]	74 pairs, high-school age	Identical to Torgersen, (1979), completed by both twins and by the mother	cutting point on single summed raw score composed of (1) scores of both twins, and of (2) scores by mother	(1) 98% of MZ correct, 77% of DZ correct, (2) 93% of MZ correct, 92% of DZ correct
Spitz, Moutier, Reed, Busnel, Marchaland, Roubertoux, Carlier, 1996[21]	79 pairs, age 8–12.5	Adapted from Goldsmith, (1991), originally composed of 18 items, completed by one parent	cutting point on mean score obtained by summing raw scores and dividing by number of items answered; logistic regression	cutting point: 97% of total sample correct; logistic regression: 92% of total sample correct

Continued on next page

Appendix 1 *continued from previous page*

Study	*Subjects*	*Mailed questionnaire*	*Method of classification*[1]	*Results*
Charlemaine, Duyme, Aubin, Guis, Marquiset, De Pirieux, Strub, Brossard, Jarry, Le Group Romulus, Frydman, Pons, 1997[22]	76 pairs, age < 1	Adapted from Bønnelyke *et al*, (1989), originally composed of 26 items; completed by one parent or both parents together	decision rules, various approaches; cutting point on summed raw score	decision rules: ranging from 87% to 99% of total sample correct; cutting point: 96% of total sample correct
Chen, Chang, Wu, Lin, Chang, Chiu, Soong, 1999[23]	Two samples: (1) 105 pairs age 12–16, (2) 47 pairs age 2–12	Adapted from Cohen *et al*, (1975), Goldsmith, (1991), and culture-specific items. Originally composed of 20 (parental report) and 27 (self report) items; completed by (1) both parents and both twins, (2) one parent	logistic regression; cutting point on 3-item profiles for (1) only	logistic regression: (1) 97% of total sample correct by parental report, 96% of total sample correct by self report; (2) 93% of total sample correct; cutting point: (1) identical to logistic regression

[1]Each study compares the assignment of zygosity based on questionnaire to the classification obtained through blood polymorphism or DNA markers, or a combination of both. [2]The question 'are twins alike as two peas in a pod?' is considered a similarity item. [3]Univariate versus multivariate: this reflects the number of sub-questions that deal with confusion of twin identity. Univariate: the occurrence of twin confusion is limited to one type of person, for instance 'strangers'. Multivariate: the occurrence of twin confusion by multiple types of persons, like 'parents', 'family members', 'teachers', etc. [4]'Originally' implies that the final analyses were performed on a reduced number of items.

Appendix 2 Translation of zygosity questionnaire, sent to parents when twins reach the age of 6

How much are the twins alike with respect to:				
1	Facial appearance	not	somewhat	exactly
2	Hair colour	not	somewhat	exactly
3	Face colour	not	somewhat	exactly
4	Eye colour	not	somewhat	exactly
5	Are they as alike as two peas in a pod?	no	yes	
6	Does the mother or father mistake one for the other?	no	yes	
7	Do other family members mistake one for the other?	no	yes	
8	Do strangers have difficulty telling them apart?	no	yes	
At age 8 and 10 of the twins, two more questions are added				
9	Do you have difficulty in correctly identifying each twin on new photographs?	no	yes	
10	Do the twins have the same hair structure?	not	somewhat	exactly

Appendix 3 Unstandardised canonical discriminant function coefficients, constants and classification score to construct the classification rule

Item	*Age 6 Mother*	*Age 8 Mother*	*Age 8 Father*	*Age 10 Mother*	*Age 10 Father*
Facial appearance	0.618128	0.424786	0.546325	0.166356	0.522894
Hair colour	0.431205	0.562038	0.385539	0.465518	0.176443
Face colour	0.521933	0.059957	0.156256	0.170350	0.218696
Eye colour	0.252118	0.242795	0.271036	0.192224	0.119514
Two peas	0.349174	0.329923	0.190973	0.086300	0.165164
Mother/father	0.025022	0.086795	–0.10002	0.061590	–0.00264
Family members	1.098133	0.343303	0.638154	0.825344	0.452154
Strangers	0.358312	0.432926	0.568857	1.054857	1.688902
Photograph		–0.10844	–0.03261	–0.07711	–0.26824
Hair structure		0.778413	0.601257	0.611719	0.459194
Constant	–7.30262	–6.58742	–6.76956	–6.92407	–6.68708

Items are rated 0, 1, or 2 on a three-point scale. Dichotomous items are rated 0 or 1. By multiplying each coefficient with the item score and summing these products with the constant, a zygosity score is obtained for each individual pair. This zygosity score is compared with the classification score that is generated by the discriminant function analysis. In this study, the classification score is 0.4 for each individual dataset. Pairs whose zygosity score is greater than 0.4 are assigned the label monozygotic, pairs with scores below this classification score are considered dizygotic.

Behavior Genetics, Vol. 33, No. 3, May 2003 (© 2003)

ADHD: Sibling Interaction or Dominance: An Evaluation of Statistical Power

M. J. H. Rietveld,[1,3] D. Posthuma,[1] C. V. Dolan,[2] and D. I. Boomsma[1]

Received 25 Apr. 2002—Final 11 Nov. 2002

Sibling interaction effects are suggested by a difference in phenotypic variance between monozygotic (MZ) twins and dizygotic (DZ) twins, and a pattern of twin correlations that is inconsistent with additive genetic influences. Notably, *negative* sibling interaction will result in MZ correlations which are more than twice as high as DZ correlations, a pattern also seen in the presence of genetic dominance. Negative sibling interaction effects have been reported in most genetic studies on Attention Deficit Hyperactivity Disorder (ADHD) and related phenotypes, while the presence of genetic dominance is not always considered in these studies. In the present paper the statistical power to detect both negative sibling interaction effects and genetic dominance is explored. Power calculations are presented for univariate models including sources of variation due to additive genetic influences, unique environmental influences, dominant genetic influences and a negative sibling interaction (i.e., contrast effect) between phenotypes of twins. Parameter values for heritability and contrast effects are chosen in accordance with published behavior genetic studies on ADHD and associated phenotypes. Results show that when both genetic dominance and contrast effects are truly present and using a classical twin design, genetic dominance is more likely to go undetected than the contrast effect. Failure to detect the presence of genetic dominance consequently gives rise to slightly biased estimates of additive genetic effects, unique environmental effects, and the contrast effect. Contrast effects are more easily detected in the absence of genetic dominance. If the significance of the contrast effect is evaluated while also including genetic dominance, small contrast effects are likely to go undetected, resulting in a relatively large bias in estimates of the other parameters. Alternative genetic designs, such as adding pairs of unrelated siblings reared together to a classical twin design, or adding non-twin siblings to twin pairs, greatly enhances the statistical power to detect contrast effects as well as the power to distinguish between genetic dominance and contrast effects.

KEY WORDS: Statistical power; sibling interaction; genetic dominance; ADHD; power calculation; twin study; heritability.

INTRODUCTION

The effects of phenotypic interaction among twins and siblings to individual differences in behavior was first introduced by Eaves (1976) and later discussed by Carey (1986). This interaction can either be cooperative or competitive. In the former case, behavior in one twin leads to similar behavior in his or her co-twin. In the latter case, behavior in one twin leads to opposite behavior in his or her co-twin. For common childhood psychopathology, cooperation and competition effects have both been reported (for a review, see Garcia *et al.,* 2000). In data obtained from parental ratings the effects of cooperation and competition may be mimicked (Eaves *et al.,* 2000; Neale and Stevenson, 1989; Simonoff *et al.,* 1998). When parents are asked to evaluate and report upon their children's phenotype,

[1] Department of Biological Psychology, Vrije Universiteit, Amsterdam, the Netherlands.
[2] Department of Psychology, Universiteit van Amsterdam, Amsterdam, the Netherlands.
[3] To whom correspondence should be addressed at Vrije Universiteit, Department of Biological Psychology, Van der Boechorststraat 1, 1081 BT Amsterdam, the Netherlands. Tel: +31.20.444.8812. Fax: +31.20.444.8832. E-mail: mjh.rietveld@psy.vu.nl

0001-8244/03/0500-0247/0

they may compare the twins' behavior. In this way, the behavior of one twin becomes the standard against which the behavior of the co-twin is rated. Parents may either stress the similarities or differences between the children, resulting in an apparent cooperation or competition effect. The presence of an interaction effect, either true sibling interaction or rater bias, is indicated by differences in MZ and DZ variances. If the interaction effect is cooperative (either true cooperation or due to rater bias), the variances of MZ and DZ twins are both inflated, and this effect is greatest on the MZ variance. The opposite is observed if the effect is competitive; MZ and DZ variances are both deflated and again this effect is greatest on the MZ variance. In the present paper our main interest is in a *competition* effect (either true competition or due to rater bias), also referred to as a contrast effect. In addition to heterogeneity in MZ and DZ variances, the presence of a contrast effect affects MZ and DZ covariances, resulting in a characteristic pattern of MZ and DZ correlations in which MZ correlations are much larger than DZ correlations. This pattern of correlations is not only consistent with contrast effects, but also with genetic nonadditivity such as dominance effects. In order to distinguish between genetic dominance and contrast effects it is therefore crucial to consider MZ and DZ variance-covariance structures in addition to MZ and DZ correlations.

Until recently, the analysis of twin and sibling data by use of contrast models has received limited attention in the behavior genetic literature. However, this has changed with recent publications on common childhood psychopathology measured in large twin samples of around 1000 pairs and more (Eaves *et al.,* 1997; Nadder *et al.,* 1998, 2001; Price *et al.,* submitted; Rhee *et al.,* 1999; Rietveld *et al.,* 2003; Thapar *et al.,* 2000, Van Beijsterveldt *et al.,* submitted). These studies reported that for behavioral problems like inattention, hyperactivity, and impulsivity the DZ correlation is small and sometimes even negative. Overall, the MZ correlation is large and more than twice the DZ correlation, suggesting either the presence of genetic dominance or the presence of contrast effects (or both). In addition, however, the DZ variance is often reported larger than the MZ variance, favouring the presence of a contrast effect over the presence of genetic dominance. The consistency in finding evidence for contrast effects among studies on ADHD is noteworthy. With the exception of Rhee *et al.* (1999) and Levy *et al.* (1997) the largest twin studies report the presence of contrast for at least one of their measures (Van Beijsterveldt *et al.,* submitted; Eaves *et al.,* 1997; Nadder *et al.,* 1998, 2001; Price *et al.,* submitted; Rietveld *et al.,* in press; Thapar *et al.,* 2000). The estimate for the contrast effect ranges from −.05 to −.25. Consistent with studies of smaller samples (Gjone *et al.,* 1996; Hudziak *et al.,* 2000; Kuntsi *et al.,* 2000; Martin *et al.,* 2002; Schmitz *et al.,* 1995; Spinath and Angleitner, 1998; Thapar *et al.,* 1995, Saudino *et al.,* 2000; Stevenson, 1992), broad heritability for traits related to ADHD is estimated between 60%–85% with the residual variance explained by environmental experiences unique to the individual.

Both genetic dominance and contrast effects may act on a given trait. In such a model, four sources of variance are specified; additive genetic (A), genetic dominance (D), unique environmental (E) and the interaction effect (*b*). Twin models including these effects have sometimes been accused of overparametrization or underidentification (Martin *et al.,* 2002). An empirical examination of the identification of the ADE-b model, however, has previously shown and concluded that the model is indeed identified (Nadder *et al.,* 1998). In the present study the identification of the ADE-b model is explored both formally and empirically.

The exploration of an ADE-b model is hindered by a lack of power. With MZ and DZ twins reared together, parameter estimates are highly correlated. The detection of dominance requires large samples, preferably including pairs of varying genetic relatedness (Eaves, 1972; Posthuma and Boomsma, 2000). Eaves (1976) noted that the inclusion of pairs of unrelated individuals reared together facilitates the detection of contrast effects. In the present study we discuss the statistical power to detect genetic dominance and contrast effects using a classic twin design as well as alternative designs.

METHODS

Model

The algebraic derivation of the expectation for the variances and covariances in the presence of contrast effects is found in Eaves (1976) and Neale and Cardon (1992).[4] The expected twin variance and twin covariance are modeled as follows, see Table I.

The present power study is limited to the analysis of the covariance structure. The sibling interaction as

[4] The expected additive genetic variance and genetic dominance variance are listed incorrectly in Table 10.3 (Neale and Cardon, 1992, p. 208). The α is missing for additive genetic variance (i.e., 1 for MZ twins and 0.5 for DZ twins), and δ is missing for genetic dominance variance (i.e., 1 for MZ twins and 0.25 for DZ twins).

Table I. Algebraic Representation of Expected Variances, Covariances, and Correlation under ADE-b and ADE Models

	ADE-b		ADE	
	Expected variance	Expected covariance	Expected variance	Expected covariance
Additive genetic	$\frac{a^2 (1 + 2b\alpha + b^2)}{(1 - b^2)^2}$	$\frac{a^2 (\alpha + 2b + \alpha b^2)}{(1 - b^2)^2}$	a^2	$\alpha\, a^2$
Dominant genetic	$\frac{d^2 (1 + 2b\delta + b^2)}{(1 - b^2)^2}$	$\frac{d^2 (\delta + 2b + \delta b^2)}{(1 - b^2)^2}$	d^2	$\delta\, d^2$
Unique environment	$\frac{e^2 (1 + b^2)}{(1 - b^2)^2}$	$\frac{e^2(2b)}{(1 - b^2)^2}$	e^2	0
Total	$\frac{(1 + b^2)(a^2 + d^2 + e^2) + 2b\,(\alpha\, a^2 + \delta d^2)}{(1 - b^2)^2}$	$\frac{(1 + b^2)(\alpha\, a^2 + \delta d^2) + 2b\,(a^2 + d^2 + e^2)}{(1 - b^2)^2}$	$a^2 + d^2 + e^2$	$\alpha\, a^2 + \delta\, d^2$
Expected correlation	$\frac{(1 + b^2)(\alpha\, a^2 + \delta d^2) + 2b\,(a^2 + d^2 + e^2)}{(1 + b^2)(a^2 + d^2 + e^2) + 2b\,(\alpha\, a^2 + \delta d^2)}$		$\frac{\alpha\, a^2 + \delta d^2}{a^2 + d^2 + e^2}$	

Note: a^2 denotes the additive genetic variance, d^2 denotes the dominant genetic variance, e^2 denotes the unique environmental variance, b denotes the contrast effect, α denotes the additive genetic relation between two individuals of a pair (i.e., 1 for MZ twins and 0.5 for DZ twins), and δ denotes the dominant genetic relation between two individuals of a pair (i.e., 1 for MZ twins and 0.25 for DZ twins).
Adapted from Table 10.3, Neale and Cardon (1992).

considered in this paper does not result in phenotypic mean differences between pairs of relatives varying in genetic relatedness (Carey, 1986, p. 324).

Calculations

Variance-covariance matrices were calculated for MZ and DZ twins for ADE-b models. Because the majority of twin studies on ADHD report no differences in heritability between males and females, evaluation of sex differences were not included in the power calculations. Estimates were 0.50 (a^2), 0.25 (d^2), and 0.25 (e^2) for additive genetic, genetic dominance, and unique environmental sources of variance, respectively. These estimates are usually reported for ADHD phenotypes, after correction of the total variance for the increase in variance due to the contrast effect. In other words, the sum of these three sources of variance equaled 1.00 in the absence of a contrast effect ($b = .00$). The contrast effect b was fixed at varying values, .00, −.05, −.10, −.15 and −.20. We considered univariate phenotypes only. All analyses were carried out using the statistical software package Mx (Neale *et al.,* 1999).

Identification of the ADE-b Model

Identification of parameters in the ADE-b model can be established formally by the method of Bekker *et al.* (1993). This method involves calculating the nullspace of the Jabobian of the covariance structure model. The Jacobian is the matrix of the derivatives of each element in the expected covariance matrices with respect to the parameters in the model. Specifically, this matrix has as many rows as there are elements in the expected covariance matrices, and as many column as there are (to be estimated) parameters. Each entry in the matrix is the derivative of the expected (co-) variance with respect to the parameters. The model is identified if and only if the null space of the Jacobian is zero, that is, if there are no linear dependencies among the columns of the Jacobian. In other words, the Jacobian should have full column rank for the model to be identified. Forming the Jacobian of the ADE-b model and calculating the null space is carried out using the program Maple (Heck, 1997). The formal approach to establish identification does not exclude the possibility that empirical underidentification may be encountered with this model (Kenny, 1979). As empirical underidentification may be apparent in computational difficulties, we also adopted an empirical approach to investigate whether such difficulties were encountered in fitting the ADE-b model (Neale *et al.,* 1999, p. 92). Again, MZ and DZ variance-covariance matrices were calculated with a set of fixed values a^2, d^2, e^2 and varying values for b. In an attempt to retrieve parameter values equal to those which were used as input in the calculation, varying starting values for a^2, d^2, e^2 and b were used. If the true parameter values are recovered regardless of starting values and the chi-square is zero, this provides an indication that empirical underidentification is not a problem.

Power in a Classic Twin Design

Power calculations were carried out by fitting the known model to the exact (population) covariance matrices as described in Neale and Cardon (1992). Constraining a certain set of parameters to zero and refitting the model provides the non-centrality parameter related to that particular constraint. From this non-centrality parameter the sample size required to reject the constrained (i.e. false) model with a probability (i.e., a power) of 0.80 and a significance level α of .05 can be calculated (Martin *et al.,* 1978; Hewitt and Heath, 1988) and is conveniently supplied by Mx. For the true model ADE-b, three series of power analyses were conducted. i) to establish the statistical power to detect contrast effects for given sample sizes, ADE models were fitted to variance-covariance matrices from the true ADE-b model. ii) to establish the statistical power to detect genetic dominance for given sample sizes false AE-b models were fitted. iii) to establish the power to detect a contrast effect after dominance had already been dropped from the model, the fit of an AE model was compared to the fit of an AE-b model while the true model was ADE-b.

DZ twins usually outnumber MZ twins due to the inclusion of opposite-sex twins. We therefore maintained a 1:2 ratio of MZ to DZ. Sample sizes are 300, 1500, 3000 and 6000 twin pairs, corresponding to empirical sample sizes for ADHD. The largest three sample sizes correspond to publications by the Virginia Twin Study of Adolescent Behavioral Development (Eaves *et al.,* 1997; Nadder *et al.,* 2001) and the Greater Manchester Twin Register (Thapar *et al.,* 2000); the Netherlands Twin Registry (Rietveld *et al.,* 2003); and the Twins Early Development Study (Price *et al.,* submitted), respectively. A sample of 300 twin pairs falls within the range of sample sizes reported upon by several smaller studies (Gjone *et al.,* 1996; Hudziak *et al.,* 2000; Kuntsi *et al.,* 2000; Martin *et al.,* 2002; Saudino *et al.,* 2000; Schmitz *et al.,* 1995; Sherman *et al.,* 1997).

Power in Alternative Twin Designs

Following the suggestion of Eaves (1976), the effects of additionally including pairs of genetically unrelated siblings (UR) reared together on statistical power were investigated. These unrelated siblings were considered as an extra group, not being part of the twin families. Given that shared environment is absent for ADHD, it was assumed that any phenotypic relation between these siblings is due to the contrast effect. These additional power calculations were conducted for the most unfavorable conditions, i.e., for the smaller sized twin studies, using the same fixed parameter values. We also considered the increase in power when non-twin siblings were added to the twin pairs. That is, here we considered the additional siblings being part of the twin families. The ratio of twin pairs with a sibling to twin pairs only was fixed at 2:1 while the ratio MZ to DZ twins was maintained at 1:2.

Positive Interaction

Whereas a negative *b* is confounded with D, a positive *b* is confounded with C. A model including shared environmental variance (C) and a positive *b* (i.e., ACE+b) is expected to encounter more power problems compared to an ADE model with a negative *b*. The expected shared environmental variance derived from the ACE+b model is of equal magnitude for MZ and DZ twins. As a result, the total variances of MZ and DZ twins do not differ as much from one another in the presence of a positive *b* compared to a negative *b*. To illustrate the magnitude of the power problem to detect a positive interaction effect, variance-covariance matrices were calculated with variance estimates of .50 for a^2, .25 for c^2 and .25 for e^2. The interaction effect *b* was fixed at +.15. An ACE model was fit to these variance-covariance matrices.

RESULTS

Expected Descriptives

The consequences of the presence a contrast effect on the total expected variances, covariances and correlations can be deducted from the equations given previously. In Table II the expected values of the MZ, DZ, and UR variances, covariances and correlations are given for each of the five true ADE-b models that are used in the power analyses.

This exercise clearly shows the consequences for variances, covariances and correlations in the presence of a contrast effect. Whereas the DZ variance is only minimally affected by a contrast effect, the DZ covariance reduces rapidly. As opposed to the DZ variance, the MZ variance is greatly affected by the contrast effect. The statistics for the UR pairs illustrate how the contrast effect corresponds to a negative correlation.

Identification of the ADE-b Model

Formal identification of the ADE-b model was established by calculation of the nullspace of the Jacobian

Table II. Expected Variances (Var), Covariances (Cov), and Correlations (Cor) for MZ Twins, DZ Twins, and UR Sibling Pairs in the Presence of a Contrast Effect (*b*)

	True model				MZ			DZ			UR		
	a^2	d^2	e^2	b	Var	Cov	Cor	Var	Cov	Cor	Var	Cov	Cor
(1)	.50	.25	.25	.00	1.00	.75	.75	1.00	.31	.31	1.00	.00	.00
(2)				−.05	.93	.66	.70	.98	.21	.22	1.01	−.10	−.10
(3)				−.10	.88	.57	.65	.97	.12	.12	1.03	−.20	−.20
(4)				−.15	.83	.49	.59	.97	.02	.02	1.07	−.31	−.29
(5)				−.20	.80	.41	.51	.99	−.08	−.08	1.13	−.43	−.38

Note: Expected variances and covariances were calculated in Mx. Due to rounding errors the listed descriptives may vary slightly from those calculated by hand.

of the covariance structure (the Maple input is available upon request). Data were generated assuming broad heritability of .75 with the values of the contrast effect *b* varying between −.05 and −.20. The calculated variance – covariance matrices were used as input data in a series of modeling in which starting values were varied for *b* and A, D, and E. Occasionally, we found that Mx converged to the incorrect maximum. This is perhaps due to the fact that Mx optimizes the log-likelihood using only function values. The specification of appropriate bounds ($-1 < b < 1$) avoided these problems and identical parameter values with zero chi-square were retrieved. We therefore conclude that the ADE-b model is empirically identified.

Power in a Classic Twin Design

For our first series of analyses, the effect of interest is the contrast effect *b*. As presented in Table III, a small contrast effect of −.05 remains undetected, irrespective of the number of twin pairs. Even with 6000 pairs, the power is only .30 to detect this effect. A large contrast effect (−.20) is detected reliably with a sample size of only 300 pairs. Comparison of the difference between estimates (see footnote, Table III) and the true values for A, D, and E indicates that when a contrast effect is ignored, estimates for the remaining sources of variance deviate from the true values, most notably when the true value of the contrast effect is larger than −.05.

In the second series of analyses, the power to detect genetic dominance was investigated by using the non-centrality parameter from the AE-b model, fitted to MZ and DZ covariance matrices generated from the ADE-b model. The requirement of very large sample sizes to detect D was confirmed (Table IV). The power to detect D is independent of the size of the contrast effect, clearly illustrated by identical power estimates within each twin design with varying values for *b*.

The third series of power calculations were based on a presumed 'realistic' scenario (Table V). The true parameters, again, are a^2 (.50), d^2 (.25), e^2 (.25) and *b* (.00, −.05, −.10, −.15, and −.20).

Following the usual procedure of comparing nested submodels to a full model (ADE-b), one may first decide that the variance due to D may be omitted from the model. Inability to detect D happens even with the largest sample size of 6000 twin pairs (see Table IV). The estimates that result from the fit of the AE-b model are slightly biased (these are identical to those listed in the footnote of Table IV). Dropping D from the ADE-b

Table III. Power to Detect a Contrast Effect *b* (*df* = 1)

	True model				100 MZ 200 DZ	500 MZ 1000 DZ	1000 MZ 2000 DZ	2000 MZ 4000 DZ
	a^2	d^2	e^2	b				
(1)	.50	.25	.25	−.05	.06	.11	.17	.30
(2)				−.10	.16	.59	.87	.99
(3)				−.15	.61	1.00	1.00	1.00
(4)				−.20	.94	1.00	1.00	1.00

Note: Estimates are (1) $a^2 = .16$, $d^2 = .55$, and $e^2 = .29$; (2) $a^2 = .00$, $d^2 = .66$, and $e^2 = .34$; (3) $a^2 = .00$, $d^2 = .60$, and $e^2 = .40$; (4) $a^2 = .00$, $d^2 = .51$, and $e^2 = .49$.

Table IV. Power to Detect Dominant Genetic Variance (*df* = 1)

	True model							
	a^2	d^2	e^2	b	100 MZ 200 DZ	500 MZ 1000 DZ	1000 MZ 2000 DZ	2000 MZ 4000 DZ
(1)	.50	.25	.25	.00	.06	.08	.12	.18
(2)				−.05	.06	.08	.12	.18
(3)				−.10	.06	.08	.12	.18
(4)				−.15	.06	.08	.12	.18
(5)				−.20	.06	.08	.12	.18

Note: Estimates are $a^2 = .78$, and $e^2 = .22$ for each model, (1) $b = -.04$; (2) $b = -.09$; (3) $b = -.14$; (4) $b = -.19$; (5) $b = -.24$.

model increases power to detect the contrast effect, which is slightly biased upwards. Interestingly, *b* may be detected when *b* is truly absent. Even with data from 3000 twin pairs available, there is a chance of .89 to estimate *b* at .04 when the true value of *b* is .00.

Power in Alternative Twin Designs

Based on Table III, it was decided to perform power analyses for the most unfavorable conditions. Variance-covariance matrices were calculated for 50, 100, and 200 pairs of UR which were combined with the twin data sets including 300 and 1500 pairs. A small contrast of −.05 remained undetected, even if data from 6000 twin pairs and 200 UR pairs are collected (power estimated at .56). We have therefore limited our analyses for a contrast effect of −.10 and −.15. Outcomes are listed in Table VI.

Comparing the results in Table III to the results in Table VI, it is apparent that the power to detect a contrast effect is greatly enhanced by the inclusion of genetically unrelated pairs. To detect a contrast effect of −.10, a small twin study of 300 pairs benefits most from additional information measured in more than 100 UR pairs. Due to the inclusion of pairs of UR siblings reared together, power increases from .16 to .64 (100 UR pairs) and .88 (200 UR pairs). Quite notable, a design including 300 twin pairs and 200 UR pairs is equally powerful as a design including 3000 twin pairs without additional UR siblings (Table III; third twin design). Also, adding 50 UR pairs to a twin study of 1500 pairs rapidly increases the statistical power to detect a contrast effect of −.10 from .59 to .76. Further, when a small twin study is extended with UR pairs, the difficulty to detect a contrast effect of −.15 is no longer encountered.

The increase in power due to the inclusion of non-twin siblings was explored for a contrast effect of −.10 and −.15. To enable a comparison with results shown in Table VI, analyses were performed for sample sizes, identical in number of participating individuals. Irrespective of sample size, power to detect a contrast effect of −.10 is insufficient when only twins and non-siblings are participating. A study of twins and non-twin siblings consisting of 700 individuals from 300 families has sufficient power to detect a contrast effect of −.10. This compares favourable to the power (.61) available in a study of equal family size, consisting of twins only (Table III, third row, first column).

Table V. Power to Detect a Contrast Effect *b* after Dropping D from the ADE-b Model (*df* = 1)

	True model							
	a^2	d^2	e^2	b	100 MZ 200 DZ	500 MZ 1000 DZ	1000 MZ 2000 DZ	2000 MZ 4000 DZ
(1)	.50	.25	.25	.00	.17	.61	.89	.99
(2)				−.05	.56	1.00	1.00	1.00
(3)				−.10	.89	1.00	1.00	1.00
(4)				−.15	.98	1.00	1.00	1.00
(5)				−.20	1.00	1.00	1.00	1.00

Note: Estimates are (1) $a^2 = .74$, $e^2 = .26$; (2) $a^2 = .68$, $e^2 = .32$; (3) $a^2 = .60$, $e^2 = .40$; (4) $a^2 = .47$, $e^2 = .53$.; (5) $a^2 = .31$, $e^2 = .69$.

Table VI. Power to Detect a Contrast Effect ($df = 1$) when Genetically Unrelated Sib Pairs (UR) Are Included

	True model							
	a^2	d^2	e^2	b	100 MZ 200 DZ +50 UR	100 MZ 200 DZ +100 UR	100 MZ 200 DZ +200 UR	500 MZ 1000 DZ +50 UR
(1)	.50	.25	.25	−.10	.43	.64	.88	.76
(2)				−.15	.89	.97	1.00	1.00

Power to Detect Positive Interaction

Expected variances calculated from the ACE +b are 1.31 and 1.23 for MZ and DZ twins, respectively. Expected covariances are 1.12 and .85 for MZ and DZ twins, respectively. The power is .43 to detect an interaction effect of $b = +.15$ with a twin sample of 2000 MZ and 4000 DZ pairs. This suggests that the detection of a positive interaction effect is not feasible given any sample size.

DISCUSSION

The use of models that incorporate interactions between phenotypes has become a popular method to analyze twin data. For overactivity, hyperactivity, impulsivity, inattention, and other phenotypes related to ADHD negative interactions or contrast effects have been reported. For some traits, positive interactions have been observed (Carey, 1992; Patterson, 1984; Rowe *et al.,* 1992). These reports most often concern antisocial tendencies. A single power calculation was performed to illustrate the mission impossible to detect a positive interaction effect. As opposed, the detection of a negative interaction is a feasible operation given a certain sample size. Here, we present power calculations for the detection of genetic dominance and contrast effects in the context of ADHD phenotypes. Sample sizes were fixed at 300, 1500, 3000, and 6000 twin pairs and a MZ to DZ ratio of 1 to 2, based on sample sizes from published reports on ADHD.

Prior to the power calculations, it was formally and empirically established that the ADE-b model is identified when using data from MZ and DZ twins. A model including both these effects as well as additive genetic effects and unique environmental effects was subsequently used as a true model for the power calculations. First, power was calculated for four values of the contrast effect (−.05, −.10, −.15, and −.20). Even with a sample size of 6000 twin pairs it was difficult to detect a relatively small contrast effect of −.05. Larger contrast effects could easily be detected using 3000 twin pairs ($b = -.10$), 1500 twin pairs ($b = -.15$) or even 300 twin pairs ($b = -.20$). The statistical power to detect genetic dominance accounting for 25% of the variance, remained as low as 0.18 even for sample size of 6000 twin pairs, confirming results of previous power analyses (Eaves, 1972; Posthuma and Boomsma, 2000). The difficulty to detect dominance is independent of the presence of either small or large contrast effects. Genetic dominance is estimated by differences in MZ and DZ covariances only and not by differences in MZ and DZ variances.

Usually, in an attempt to explain the data by the most plausible and parsimonious model, each individual source of variance is evaluated for its contribution to the observed total variance. Although reduced models are indeed more parsimonious, these models may give a biased account of the data. If a contrast effect is present but ignored, estimates for genetic dominant sources of variance are inflated whereas estimates for additive genetic sources of variance are deflated. The discrepancies between true values and resulting estimates are large, with unrealistic values for the additive genetic source of variance. Whereas large twin studies have the potential to detect a contrast effect, genetic dominance is more likely to go undetected given any sample size. Under these circumstances, the power to detect the contrast effect is much higher compared to the situation where genetic dominance is not omitted from the model. A small study of 300 pairs has sufficient power (.89) to detect a contrast effect of −.10 when genetic dominance is rejected from the model. This compares to a power estimate of .16 to detect the same effect of *b* when the genetic dominance is still included in the model. An accompanying result of ignoring the presence of genetic dominance is that A, E, and *b* deviate from the actual values. However, this bias is relatively small. When broad heritability is considered, the true values of A and D add up to 75% com-

pared to 78% obtained after fit of the reduced AE-b model. This bias is much smaller compared to the bias that results from rejecting a contrast effect prior to evaluating the presence of genetic dominance. From this we argue that the evaluation of genetic dominance should precede the evaluation of a contrast effect in a sequence of model fitting. If it is decided to exclude D from the model, estimates are close to true values and the likelihood to detect *b* is largely increased. However, with any reduced model, the researcher should be cautioned that the newly obtained estimates are biased.

We showed the advantage of extending the twin design with data from genetically unrelated siblings reared together. The power calculations indicated that a contrast effect is considerably easier to detect when data from such an additional group of informative pairs is available. Here we have assumed that the contrast effect for UR pairs is identical to the effect for twins. This assumption may not always be tenable. For instance, the magnitude of a contrast effect may vary as a function of the age difference between siblings. Since twins are of the same age, and UR most likely not, the magnitude of the contrast effect may differ between twin pairs and sibling pairs. Not only may the magnitude of the effect differ, the interaction may be limited to a one-way effect in UR pairs (e.g., Abramovitch *et al.,* 1979). If the contrast effect in UR pairs may not be constrained to the effect in twin pairs, there is obviously no advantage of including such an extra group of sibling pairs. So, ideally, the UR siblings are close in age. Segal (2000) has reported on such siblings as virtual twins who were either two adoptees, or one biological and one adopted child. Of course, unrelated siblings reared together may also apply to two children with different biological parents who live in the same household because their parents re-married. Although the search for these virtual twins may be troublesome, the increase in power is worth the effort, as are the financial benefits of not having to collect data from a large number of twins. We demonstrated that, in order to detect a moderate contrast effect, a study with 300 twin pairs and 200 sibling pairs has just as much power as a twin study of six times the sample size. The advantage of including UR pairs in the twin study also became evident when we explored the effects on statistical power of adding one non-twin sibling in the twin study (see also Posthuma and Boomsma, 2000).

In conclusion, the detection of a contrast effect is achievable given a certain sample size and composition. Researchers of ADHD and related traits are encouraged to search for unrelated sibling pairs to enhance power to detect a contrast effect. If present twin studies are small and suffer from minimal power to detect a contrast effect, we propose that genetic dominance is evaluated prior to the contrast effect.

ACKNOWLEDGMENTS

This work was financially supported by the Universitair Stimulerings Fonds (grant number 96/22) and by the Human Frontier Science Program (grant number rg0154/1998-B). The work of C. V. Dolan was financially supported by a fellowship of the Royal Netherlands Academy of Arts and Sciences.

REFERENCES

Abramovitch, R., Carter, C., and Lando, B. (1979). Sibling interaction in the home. *Child Dev.* **50**:997–1003.

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

Bekker, P. A., Merckens, A., and Wansbeek, T. J. (1993). *Identification, equivalent models, and computer algebra*. Boston, MA: Academic Press.

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

Carey, G. (1986). Sibling imitation and contrast effects. *Behav. Genet.* **16**:319–341.

Carey, G. (1992). Twin imitation for antisocial behavior: Implications for genetic and family environment research. *J. Abnorm. Psychol.* **101**:18–25.

Eaves, L. J. (1972). Computer simulation of sample size and experimental design in human psychogenetics. *Psychol. Bull.* **77**:144–152.

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

Eaves, L. J., Rutter, M., Silberg, J. L., Shillady, L., Maes, H., and Pickles, A. (2000). Genetic and environmental causes of covariation in interview assessments of disruptive behavior in child and adolescent twins. *Behav. Genet.* **30**:321–334.

Eaves, L. J., Silberg, J. L., Maes, H. H., Simonoff, E., Pickles, A., Rutter, M., Neale, M. C., Reynolds, C. A., Erikson, M. T., Heath, A. C., Loeber, R., Truett, K. R., and Hewitt, J. K. (1997). Genetics and developmental psychopathology: 2. The main effects of genes and environment on behavioral problems in the Virginia Twin Study of Adolescent Behavioral Development. *J. Child Psychol. Psychiatry* **38**:965–980.

Garcia, M. M., Shaw, D. S., Winslow, E. B., and Yaggi, K. E. (2000). Destructive sibling conflict and the development of conduct problems in young boys. *Dev. Psychol.* **36**:44–53.

Gjone, H., Stevenson, J., and Sundet, J. M. (1996). Genetic influence on parent-reported attention-related problems in a Norwegian general population twin sample. *J. Am. Acad. Child Adolesc. Psychiatry* **35**:588–596.

Heck, A. (1997). *Introduction to MAPLE* (2nd ed.). New York: Springer-Verlag.

Hewitt, J. K., and Heath, A. C. (1988). A note on computing the chi-square noncentrality parameter for power analyses. *Behav. Genet.* **18**:105–108.

Hudziak, J. J., Rudiger, L. P., Neale, M. C., Heath, A. C., and Todd, R. D. (2000). A twin study of inattentive, aggressive and anxious/depressed behaviors. *J. Am. Acad. Child Adolesc. Psychiatry* **39**:469–476.

Kenny, D. A. (1979). *Correlation and causality*. New York: Wiley.
Kuntsi, J., Gayán, J., and Stevenson, J. (2000). Parents' and teachers' ratings of problem behaviours in children: genetic and contrast effects. *Twin Res.* **3**:251–258.
Levy, F., Hay, D. A., McStephen, M., Wood, C., and Waldman, I. (1997). Attention-deficit hyperactivity disorder. A category or a continuum? Genetic analysis of a large-scale twin study. *J. Am. Acad. Child Adolesc. Psychiatry* **36**:737–744.
Martin, N. G., Eaves L. J., Kearsey, M. J., and Davies, P. (1978). The power of the classical twin study. *Heredity* **40**:97–116.
Martin, N., Scourfield, J., and McGuffin, P. (2002). Observer effects and heritability of childhood attention-deficit hyperactivity disorder symptoms. *Br. J. Psychiatry* **180**:260–265.
Nadder, T. S., Silberg, J. L., Eaves, L. J., Maes, H. H., and Meyer, J. M. (1998). Genetic effects on ADHD symptomatology in 7- to 13-year-old twins: Results from a telephone survey. *Behav. Genet.* **28**:83–100.
Nadder, T. S., Silberg, J. L., Rutter, M., Maes, H. H., and Eaves, L. J. (2001). Comparison of multiple measures of ADHD symptomatology: A multivariate genetic analysis. *J. Child Psychol. Psychiatry* **42**:475–486.
Neale, M. C., Boker, S. M., Xie, G., and Maes, H. H. (1999). *Mx: Statistical Modeling*. Box 126 MCV, Richmond, VA 23298: Department of Psychiatry. 5th ed.
Neale, M. C., and Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Dordrecht Boston: Kluwer Academic Publishers.
Neale, M. C., and Stevenson, J. (1989). Rater bias in the EASI Temperament Scales: A twin study. *J. Pers. Soc. Psychol.* **56**:446–455.
Patterson, G. R. (1984). Siblings: Fellow travelers in coercive family processes. In: Blanchard R. J. (ed.). *Advances in the study of aggression*, pp. 174–213. New York: Academic Press.
Posthuma, D., and Boomsma, D. I. (2000). A note on the statistical power in extended twin designs. *Behav. Genet.* **30**:147–158.
Price, T. S., Simonoff, E., Waldman, I., Asherson, P., Curran S., and Plomin R. What is stable about hyperactive behaviors in preschool children is genetic: Implications for molecular genetic studies. Submitted.
Rhee, S. H., Waldman, I. D., Hay D. A., and Levy, F. (1999). Sex differences in genetic and environmental influences on DSM-III-R Attention-Deficit/Hyperactivity Disorder. *J. Abnorm. Psychol.* **108**:24–41.
Rietveld, M. J. H., Hudziak, J. J., Bartels, M., van Beijsterveldt, G. C. M., and Boomsma, D. I. (2003) Heritability of attention problems in children: I. Cross-sectional results from age 3 to 12. *Neuropsychiat. Genet.* **1176**:102–113.
Rowe, D. C., Rodgers, J. L., and Meseck-Bushey, S. (1992). Sibling delinquency and the family environment: Shared and unshared influences. *Child Dev.* **63**:59–67.
Saudino, K. J., Cherny, S. S., and Plomin, R. (2000). Parent ratings of temperament in twins: explaining the 'too low' DZ correlations. *Twin Res.* **3**:224–233.
Schmitz, S., Cherny, S. S., and Fulker, D. W. (1998). Increase in power through multivariate analyses. *Behav. Genet.* **28**:357–364.
Schmitz, S., Fulker, D. W., and Mrazek, D. A. (1995). Problem behavior in early and middle childhood: An initial behavior genetic analysis. *J. Child Psychol. Psychiatry* **36**:1443–1458.
Segal, N. (2000). Virtual twins: new findings on within-family environmental influences on intelligence. *J. Edu. Psychol.* **92**: 442–448.
Sherman, D. K., Iacono, W. G., and McGue, M. K. (1997). Attention-deficit hyperactivity disorder dimensions. A twin study of inattention and impulsivity-hyperactivity. *J. Am. Acad. Child Adolesc. Psychiatry* **36**:745–753.
Simonoff, E., Pickles, A., Hervas, A., Silberg, J. L., Rutter, M., and Eaves, L. (1998). Genetic influences on childhood hyperactivity: Contrast effects imply parental rating bias, not sibling interaction. *Psychol. Med.* **28**:825–837.
Spinath, F. M., and Angleitner, A. (1998). Contrast effects in Buss and Plomin's EAS questionnaire: a behavioral-genetic study on early developing personality traits assessed through parental ratings. *Pers. Individ. Diff.* **25**:947–963.
Stevenson, J. (1992). Evidence for a genetic etiology in hyperactivity in children. *Behav. Genet.* **22**:337–344.
Thapar, A., Harrington, R., Ross, K., and McGuffin, P. (2000). Does the definition of ADHD affect heritability? *J. Am. Acad. Child Adolesc. Psychiatry* **39**:1528–1536.
Thapar, A., Hervas A., and McGuffin P. (1995). Childhood hyperactivity scores are highly heritable and show sibling competition effects: Twin study evidence. *Behav. Genet.* **25**:537–544.
Van Beijsterveldt, C. E. M., Verhulst, F. C., Molenaar, P. C. M., and Boomsma, D. I. (submitted). The genetic basis of problem behavior in 5-year-old Dutch twin pairs.

Edited by Pak Sham

American Journal of Medical Genetics Part B (Neuropsychiatric Genetics) 117B:102–113 (2003)

Heritability of Attention Problems in Children: I. Cross-Sectional Results From a Study of Twins, Age 3–12 Years

M.J.H. Rietveld,[1]* J.J. Hudziak,[2] M. Bartels,[1] C.E.M. van Beijsterveldt,[1] and D.I. Boomsma[1]

[1]*Department of Biological Psychology, Vrije Universiteit, Amsterdam, The Netherlands*
[2]*Department of Psychiatry and Medicine (Division of Human Genetics), Center for Children, Youth and Families, University of Vermont, College of Medicine, Burlington*

Multiple twin studies of attention problems (AP) from the Child Behavior Checklist or ADHD from the DSM criteria have reported on the genetic and environmental influences on these behaviors. The majority of these have studied AP and ADHD symptoms in twin samples combined across wide age spans, combined rater information and both genders. Thus, it is possible that the results are complicated by developmental, informant, and gender differences. The purpose of this study was to assess for the genetic and environmental contributions to overactive behavior (a syndrome highly related to AP in 7-, 10-, and 12-years olds) in 3-years olds (3,671 twin pairs), and attention problems in 7- (3,373 twin pairs), 10- (2,485 twin pairs), and 12-years olds (1,305 twin pairs) while controlling for developmental, gender and rater contrast contributions. Using a cross-sectional twin design, contributions from genetic additive, genetic dominance, unique environmental and rater contrast effects were estimated for CBCL maternal reports. We found that genetic influences on overactive behavior and attention problems are high across an age span that covers preschool and elementary school age. Although girls display less problem behavior compared to boys, heritability estimates were found equal for both genders at each age. Environmental experiences that are unique to the individual accounted for the remaining influence. At the age of 3 years, a rater contrast effect was detected. We hypothesize that the contrast effect represents a maternal rater bias effect that is dependent on the age of the twins. The implications of these findings are discussed with reference to the clinical setting and in the context of future research.

KEY WORDS: ADHD; genetics; rater bias; pre-school; schoolage

INTRODUCTION

In order to identify genes that contribute to the etiology of a common disorder such as Attention Deficit Hyperactivity Disorder (ADHD), molecular geneticists must study the relations between the genes of risk in individuals who suffer from the disorder versus individuals who do not. Thus the success of the gene finding expedition depends on the ability to identify those who do and do not have the condition of interest. Although molecular genetic techniques have advanced to the point that identifying genes of risk for child psychiatric disorders is a fairly simple laboratory exercise, our ability to provide accurate diagnoses remains a challenge [Hudziak, 2002]. With recent conflicting reports of molecular genetic contributions to ADHD [Barr et al., 2001] and the almost certain fact that common child psychiatric conditions like ADHD are due to multiple genes and environmental stimuli [Faraone and Doyle, 2001], the need for a clear phenotype (diagnosis) is evident. Obstacles to identifying genetic contributions to child psychopathology include developmental confounds, gender confounds, and rater bias confounds. Each is addressed in some detail.

Developmental confounds exist if the taxonomy does not take into account that behavior changes across development. Clearly, we must understand normal development to design developmentally sensitive phenotypic measures of child psychopathology. For example,

Grant sponsor: NIMH; Grant number: MH52813; Grant sponsor: Universitair Stimulerings Fonds; Grant number: 96/22.

*Correspondence to: M.J.H. Rietveld, Vrije Universiteit, Department of Biological Psychology, Van der Boechhorststraat 1, 1081 BT Amsterdam. E-mail: mjh.rietveld@psy.vu.nl

Received 1 May 2002; Accepted 13 August 2002

DOI 10.1002/ajmg.b.10024

it may not be appropriate to apply some of the criteria for ADHD to 3-years olds (e.g., "often fails to give close attention to details" or "makes careless mistakes in schoolwork"). Similarly, the hyperactivity items of ADHD are widely acknowledged to be less prevalent in older subjects [Biederman, 1998]. Unless we have a strategy that can establish levels of behavior in the normal population and then relate the behavior of 3-years olds to that of 7-, 10-, and 12-years olds, true cases may be misidentified. Failure to control for these differences could lead to incorrect measurements of changing phenotypes. Such measurement errors may produce excessive false positives or false negatives depending on a subject's developmental level. In this study, we analyze data on same aged twins studied together at successive developmental periods. For each age, the Child Behavior CheckList [CBCL; Achenbach, 1991] was used as assessment instrument.

Gender confounds exist when the study fails to consider that the condition may manifest differently in females than in males. Genetic studies of childhood psychopathology should allow for potential gender differences in the manifestation of genotypes. According to Hartung and Widiger [1998], of the 21 disorders usually first diagnosed in infancy, childhood, or adolescence for which sex ratios are provided, 17 have higher prevalence in boys than girls. They enumerated several sources of error which could generate or exaggerate gender differences in rates of psychopathology, most notably sampling biases and biases within the diagnostic criteria, concluding that "there may not be a mental disorder for which there are not important gender differences in the manner in which the disorder is expressed." This seems especially true for the study of ADHD, which is three to six times more prevalent in boys than girls [Offord et al., 1987]. Gaub and Carlson [1997] affirmed that research on gender differences in ADHD is badly needed, with emphasis on the potential confounding effects of referral bias, comorbidity, development, diagnostic procedures, and data source. In this study, we test for gender–genetic and gender–environment interactions in order to determine if there are gender differences in influences on overactivity (3-years olds) and attention problems (7-, 10-, and 12-years olds).

A third confound is rater bias: Does the taxonomic approach bias the selection of true cases? Prior twin studies on ADHD have reported that the best fitting biometric models are ones that include additive genetic (A), unique environmental (E), and rater contrast effects (b). Eaves et al. [1997], and Thapar et al. [2000] have reported large rater contrast effects on ADHD and related symptoms. The presence and magnitude of rater contrast effects can play a major role in the selection of true cases of ADHD [Hudziak, 2001]. Large rater contrast effects result in the underidentification of true cases, and undermine gene-finding efforts [Eaves et al., 2000]. In this study, we test for rater contrast effects by gender and age in large samples of Dutch twins who were recruited to participate in developmental genetic studies shortly after birth.

CURRENT RESEARCH

The purpose of this study is to extend our prior work on genetic influences on behavior problems from the CBCL by analyzing data from an epidemiological twin study, using ratings from mothers obtained when the twins were 3, 7, 10, and 12 years of age. Because of the large sample size, we can test for age, gender, and contrast effects on the genetic and environmental influences on overactivity and attention problems. With this work, we ask the following questions: What are the genetic and environmental contributions to overactivity and attention problems at ages 3, 7, 10, and 12? Are the genetic contributions similar in magnitude? Are there differences in the genetic and environmental influences by gender? Do contrast effects vary across this 9 years developmental period, that is, do mothers have different response habits when rating their 3-year-old versus their 12-year-old children?

MATERIALS AND METHODS

Subjects and Procedure

The study is part of an ongoing twin-family study of the development of behavioral and emotional problems in the Netherlands. The subjects are all part of The Netherlands Twin Registry [NTR; Boomsma, 1998]. For this study, we have assessed a sample of Dutch twin pairs whose parents (or primary caregiver) reported on their behavior when the twins were 3, 7, 10, and 12 years old. Birth cohorts 1986–1991 participated in the collection of 3- and 7-year questionnaires. For the 10-years olds, questionnaire collection was completed for birth cohorts all through 1990 and for 12-year-old twins, questionnaires were available from cohorts 1986 to the first months of 1989. The assessment procedures for the ages 7, 10, and 12 were identical to those at the age of 3 years. Parents were sent a questionnaire and were asked to return it to the NTR by mail. Parents who did not return the forms within 2 months received a reminder and persistent non-responders were contacted by phone 4 months after the initial mailing. This procedure resulted in 80% continued participation from age 3–7, 7–10, and 10–12 years. A small number of families participated at one age only (age 3, 14%; age 7, 12%; age 10, 4%; and age 12, 3%). Families that skipped at least one assessment wave make up the remaining percentages. Around 2% of the total sample of 9,160 individuals suffered from a disease or handicap that interfered severely with daily functioning. These twins and their co-twins were excluded from the genetic analyses. This left a sample of 3,853, 3-year-old pairs; the sample size for 7-year questionnaire data is 3,427 pairs, for the 10-year-old data is 2,504 pairs, and for 12-year-old data is 1,307 pairs. Zygosity was determined by DNA analyses or blood group polymorphisms for 713 same-sex twin pairs. For the remaining same-sex twin pairs, zygosity was determined by discriminant analysis of questionnaire items [Rietveld et al., 2000]. Missing values were assigned to those twin pairs with incomplete information. A small number of pairs were left unclassified because their zygosity status at one age did

not agree with their zygosity status at another age. The total number of pairs, by gender, zygosity, and age are presented in Table I. Twins with unknown zygosity were left out of the genetic analyses.

Measure

The CBCL is a standardized questionnaire for parents to report the frequency and intensity of behavioral and emotional problems exhibited by their child in the past 6 months. The questionnaire that is used for the 3-year-old twins consists of 100 problem items (CBCL/2-3) and the questionnaires that are used for the 7-, 10-, and 12-year-old twins consist of 120 problem items for parents to score (CBCL/4-18). Parents rate each behavior on a 3-point scale: 0 indicates responses of "not true," 1 "somewhat or sometimes true," and 2 "very true or often true." Dutch syndrome scales for the CBCL/2-3 were derived by both exploratory and confirmatory factor analyses using three groups of children; a clinical sample, a community sample and a sample of twins that participate in the present study [Koot et al., 1997]. This series of analyses resulted in the formation of a Dutch overactive scale, for which there is no American equivalent. The overactive scale is composed of 5 items. These items are "can't concentrate," "can't sit still," "constantly seeks help," "quickly shifts activity," and "refuses active games." Twins with more than one item missing were assigned missing value to their overall overactive scale. Around 4% of the mothers provided incomplete information on the overactive items. For the CBCL/4-18, the attention problem scale (AP) was composed according to the 1991 profile [Achenbach, 1991]. Though the majority of the 11 items of the AP scale relate to inattention, some also assess impulsivity and hyperactive behavior. In the present study, subjects with more than three missing items were not included in the analyses. This occurred in less than 2.5% of the returned questionnaires filled out by the mother when the twins were aged 7, 10, and 12 years. To facilitate reading overactive behavior at the age of 3 years is referred to with the abbreviation OA and attention problems at the older ages is referred to with AP. An association between a high score on the AP scale (a T-score of 67 is often applied as the borderline cut-off) and ADHD as assessed by the DSM is reported by numerous studies [Bird et al., 1988; Edelbrock and Costello, 1988; Steingard et al., 1992; Biederman et al., 1993; Chen et al., 1994; Hudziak et al., 2002, in review].

TABLE I. Number of Twin Pairs, by Age, Gender, and Zygosity

	Age (years)			
Zygosity	3	7	10	12
MZM	621	590	452	246
DZM	583	530	392	201
MZF	708	676	526	287
DZF	536	528	380	200
DOS	1,223	1,049	735	371
Unknown	182	54	19	2
Total	3,853	3,427	2,504	1,307

Statistical Analyses

The OA score and AP score was obtained by summing the item scores. Means, standard deviations, and twin correlations among scores were calculated using SPSS/Windows 10.0. Differences in mean scores between genders, between zygosities and across age were assessed by likelihood-ratio chi-square (χ^2) tests using the statistical software program Mx [Neale, 1997]. These tests are performed by taking into account the dependency that exists between scores from twins. PRELIS 2 [Jöreskog and Sörbom, 1993] was employed to compute the variance-covariance matrices of the observations, separately for each sex-by-zygosity group. These matrices were used as input for genetic modeling. Using the program Mx, structural equation modeling was employed to obtain an estimate of the genetic and environmental contributions to the observed variances and covariances between measures. Parameter estimates are produced such that the likelihood of the covariance structure under a given structural model is maximized. A normal distribution of the observed variables is assumed using this method. However, the distribution of OA and AP showed large deviation from normality. To approximate normality, the data were transformed by a square-root transformation. In this way, the method of maximum likelihood was eligible. Technical details of genetic model-fitting analyses are reviewed elsewhere [Neale and Cardon, 1992]. Analyses of the data took place by a cross-sectional design, that is, we analyzed each age-specific dataset separately.

Model Fitting

Observed variation for a particular measured characteristic or behavior can be decomposed into its latent, unobserved genetic and environmental components. The decomposition of variance takes place by comparing the degree of similarities between pairs of individuals who differ in their degree of genetic relatedness. The availability of twin data enables us to obtain estimates of the relative contribution of genes and environment to the observed variation of OA and AP measured at multiple ages. Figure 1 summarizes the fundamental univariate genetic model that underlies these analyses. This model was used to estimate the additive genetic (A, additive effects of genes at multiple loci), dominance genetic (D, interaction of genetic effects at the same loci), and non-shared environment (E, unique to the individual) effects. It was empirically tested that the environment that make members of the same household more alike was absent for each scale at each age. Therefore, we do not refer to this environmental source of variance in this report. The circles represent the latent, unmeasured factors. Correlations of 1.0 for MZ versus 0.5 and 0.25 for DZ genetic influences reflect the zygosity of the pair. Monozygotic twins share all their genetic material, and dizygotic twins share half of their additive genetic values and a quarter of their dominant genetic values. The unique or non-shared environment is by definition, uncorrelated between two members of a pair, either monozygotic or dizygotic. Estimates of the unique

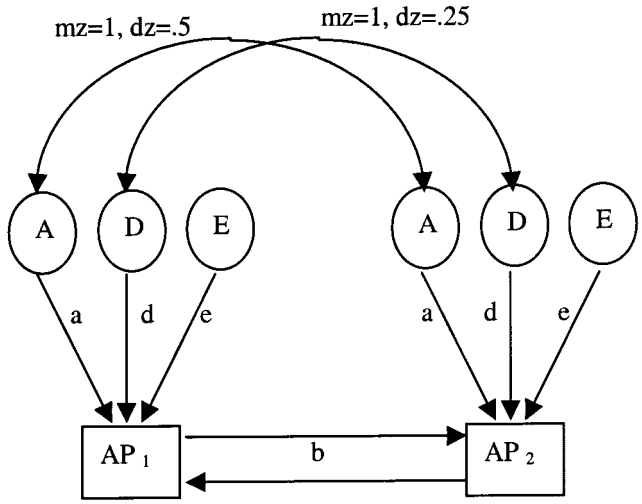


Fig. 1. Univariate path model. Note: A, additive genetic; D, dominance genetic; E, unique environment; AP_1, attention problem score of twin 1; AP_2, attention problem score of twin 2; b, contrast path.

environmental effects also include measurement error [Plomin et al., 2001].

The influence of A, D, and E on the observed variation for AP (or OA, represented in a square) is given by parameters a, d, and e. These can be considered as regression coefficients or factor loadings of AP on the latent factors A, D, and E. Squaring the factor loadings results in the absolute variance explained by each component. The sum of these squared factor loadings makes up the total variance for the observed phenotype as long as there is no evidence of interaction between genes and environment. The absolute variance explained is usually reported in a standardized form, by dividing the absolute variance by the total phenotypic variance. We then obtain a relative estimate which provide a basis for comparison with previous studies [Edelbrock et al., 1995; Schmitz et al., 1995; Rhee et al., 1999; Kuntsi et al., 2000; Thapar et al., 2000; Nadder et al., 2001].

In this univariate model, we have added a path b between the AP scores of the twins. This path implies an interaction between phenotypes. This interaction may be interpreted in two ways [Simonoff et al., 1998]. First, it may be considered a social interaction between siblings [Eaves, 1976; Carey, 1986]. That is, the behavior of one twin has a certain effect on the behavior of his or her co-twin. This effect can be either cooperative or competitive. In the former case, attention problem behavior in one twin leads to likewise behavior in his or her co-twin. In the latter case, increased attention problem behaviors in one twin lead to decreased behaviors in his or her co-twin. Second, the path may also be considered an effect introduced by the rater [Neale and Stevenson, 1989]. When parents are asked to evaluate and report upon the problem behavior of their children, they may very well compare the twins' behaviors against one another, despite instructions on the questionnaire forms. In this way, one twin becomes some kind of standard by which the behavior of the co-twin is rated. Parents may either stress the similarities or differences between the children. In like manner as the hypothesized sibling interaction, the resulting effect is either positive or negative. Based on previous research we expect this path to be negative [Thapar et al., 1995; Nadder et al., 1998; Eaves et al., 2000; Hudziak et al., 2000; Kuntsi and Stevenson, 2000; Martin et al., 2002; Price et al., submitted]. Throughout the remainder of this article, we refer to this parameter as a 'contrast effect,' to capture both mechanisms with an accurate label. Very low DZ correlations compared to MZ correlations give a first indication that a contrast effect is present. Such a pattern in correlations may also indicate the presence of non-additive genetic influences. We are able to distinguish between a contrast effect (AE-b model) and genetic dominance (ADE model) by inspection of the observed variances for MZ and DZ twins. A contrast effect leads to smaller variances for MZ compared to DZ. In case of dominance, MZ and DZ variances are expected to be of equal magnitude.

Model Fitting Procedures

Figure 1 shows the basic model for the analyses. Analyses were initiated with the fit of an ADE-b model. Next, ADE and AE-b models were applied to the data. Knowing that attention problems are not equally distributed in the population, we hypothesized upon a gender difference for the b-path. We asked ourselves if a contrast effect found for boys is equal to such an effect for girls and to opposite-sex twins. The contrast effect of the DOS twins was evaluated first. The contrast parameter was allowed to differ from those estimated for male twin pairs and female twin pairs. We next constrained the DOS contrast parameter to the mean of the contrast effect in male pairs and female pairs. For each age, we found the difference to be non-significant. Because of this uniform result, we do not report the fit of this model in our tables but have adopted this model as a baseline model, like the ADE and ADE-b model. At this stage of model fitting, all parameters were allowed to differ between genders. So, with the ADE, and AE-b models six parameters were estimated, three for each gender. The more general model, ADE-b has eight parameters specified. The ADE and AE-b are both nested within the ADE-b model. The alternative models are evaluated on the basis of their plausibility and goodness-of-fit. The goodness-of-fit index is χ^2, which indicates the statistical significance concerning the fit between expectations (model) and observations (data).

We next proceeded with the fit of reduced models, in which parameters were constrained across gender. That is, we hypothesized that the genetic and environmental effects were of equal magnitude in boys and girls. Likewise, we constrained the contrast effect in boys to equal the effect in girls. These reduced models were all nested within the models in which estimates were allowed to differ by gender. The model that describes the data best is found by comparing the goodness-of-fit of the alternative models. The degrees of freedom for this test equal the difference in the number of parameters between two nested models. If the χ^2 test is not

significant, we considered the restricted model tenable [Loehlin, 1992].

We report here the analyses based on the data from questionnaires that are completed by the mother of the twins because this is the largest dataset, and thus has greater statistical power. Due to funding limitations during some years of the study, CBCL forms were only sent to the mother of the twins. However, the paternal CBCL data were analyzed, and results that are very similar to those based on maternal reports were obtained.

RESULTS

Descriptive Statistics

Means and standard deviations for OA and AP are presented in Table II.

At all ages, mothers report more behavior problems in boys compared to girls. The difference in means was found to be significant at all ages (age 3: $\chi^2_1 = 58.15$, $P = 0.00$; age 7: $\chi^2_1 = 151.29$, $P = 0.00$; age 10: $\chi^2_1 = 119.63$, $P = .00$; age 12: $\chi^2_1 = 60.62$, $P = 0.00$). Next, we were interested whether the observed changes in mean scores across age were significant. To this end, we constrained the mean score at one age to the mean score at the following age. These three tests (age 3–7, 7–10, 10–12 years) were performed separately for each gender. It should be noted that the 3-year-old CBCL OA scale is different than the AP scale for 7, 10, and 12 year olds. For boys, we noted a significant increase in behavior problems from age 3 to 7 years ($\chi^2_1 = 60.81$, $P = 0.00$). As suggested by the very small increase in mean scores of girls, this increase did not reach significance. Likewise, from age 7 to 10 years, both boys and girls displayed stability in their degree of AP. From age 10 and beyond, we observed a decrease in reported attention problems for both genders. This change was found significant (boys: $\chi^2_1 = 25.56$, $P = 0.00$; girls: $\chi^2_1 = 29.79$, $P = .00$)

To estimate the number of children likely to meet an ADHD diagnosis, we first calculated the OA and AP distributions. Next, the raw score that corresponds to a T-score of 67 in a Dutch Community sample [Koot et al., 1997] was imposed on the distribution of the twins. The application of the community cut-off and the resulting percentages allow a comparison with the prevalence of borderline and clinical cases reported for the Dutch and other populations. Since boys differ significantly from girls on the prevalence of problems, the distributions were calculated separately by gender. In the case of the 3-year-old boys and girls, roughly 3% of the boys and 2% of the girls exceed the borderline cut-off of a T-score of greater than 67. While it may be difficult to distinguish true problematic overactive behavior of 3-years olds from age-related manifestations of normal development, the evidence suggests that between 2–3% of the children are rated as deviant on the Dutch OA scale. In comparison with other reports on pre-school rates of overactive behavior and related behavior problems, these percentages appear as quite modest [Campbell, 1995; Koot et al., 1997]. The data for the 7-, 10-, and 12-year-old age groups are more instructive. Inattention, hyperactivity, and impulsivity are more commonly diagnosed in children of these ages [Ross and Ross, 1982]. By imposing the standards of Chen et al. [1994] and using the more rigorous T-score of 67 as cut-off, around 6.5% of the boys would meet DSM criteria for ADHD at age 7 and 10 years. For girls, these numbers are 3.1% and 4.9%, respectively. At the age of 12 years, the number of children for whom DSM criteria would apply has decreased with around 1.3%.

The distributions of OA and AP display a large deviation from normal skewness and kurtosis. After square-root transformation, skewness ranged from −0.39 to 0.17 and kurtosis ranged from −0.82 to −0.59. The means based on raw scores and the variances obtained after transformation of the raw scores are summarized in Table III.

Visual inspection of Table III reveals a somewhat larger variance for dizygotic twins compared to monozygotic twins at the age of 3 years. This is observed for both boys and girls. This pattern is in line with a hypothesis of the presence of a contrast effect. The difference in variances is not observed at subsequent ages. We also note that the variance displayed by boys increases with the increasing age. For girls, the degree of variation appears to stabilize from age 7 and beyond. Means were tested in the following way; means of male DOS twins were compared to the means of DZM twins and next, dizygotic males were compared to the

TABLE II. Means and Standard Deviation (SD) Calculated Separately for Each Gender

	Boys			Girls		
	N	Mean	SD	N	Mean	SD
Age (year)						
3	3,709	2.89	2.24	3,758	2.46*	2.16
7	3,385	3.27	2.93	3,522	2.48*	2.54
10	2,456	3.23	3.09	2,593	2.38*	2.56
12	1,277	2.81	2.92	1,362	1.99*	2.29

N, number of individuals twins.
*Mean score of girls differs significantly from mean score of boys at $P < 0.05$. Reported means are not transformed.

TABLE III. Means and Variances by Zygosity and Gender

					DOS	
Age	MZM	DZM	MZF	DZF	Boy	Girl
3 year						
Mean	3.07	2.73	2.55	2.58	2.83	2.23
Variance	0.64	0.71	0.68	0.74	0.71	0.76
7 year						
Mean	3.25	3.14	2.51	2.65	3.36	2.16
Variance	0.82	0.80	0.81	0.82	0.86	0.85
10 year						
Mean	3.25	3.16	2.24	2.51	3.27	2.10
Variance	0.93	0.89	0.81	0.88	0.97	0.84
12 year						
Mean	2.84	2.72	2.11	1.97	2.86	1.17
Variance	1.02	0.90	0.74	0.78	0.93	0.82

Variances are calculated on square-root transformed means. Reported means in this table are not transformed.

monozygotic males. The same sequence of testing was applied to mean scores of female twins. The dizygotic male scores were found significantly smaller compared to monozygotic male scores at the age of 3 years ($\chi^2_1 = 8.76, P = 0.003$). More interesting is the significant difference between female DOS twins and DZF twins (age 3: $\chi^2_1 = 13.05$, $P = 0.00$; age 7: $\chi^2_1 = 20.43$, $P = 0.00$; age 10: $\chi^2_1 = 8.69$, $P = .00$). With the exception of the age of 12 years, girls with a male co-twin suffer from less OA and AP compared to girls with a female co-twin. The absence of a significant effect at the age of 12 years may have resulted from a reduced lack of power due to a smaller number of participating twins. The evaluation of means was followed by inspection of the gender-by-zygosity correlations. Twin correlations provide the initial insight into the genetic and environmental contributions to variance in OA and AP. Calculated separately for each zygosity group, correlations are reported in Table IV.

The monozygotic twins, both males and females, show large correlations of around 0.70. Quite notable is the stability observed for the monozygotic correlations. Over time monozygotic twins remain rather constant in their resemblance with respect to OA and AP. An estimation of the influence of the environment unique to the individual is obtained after subtracting the monozygotic correlation from unity. This calculation suggests that around one third of the total variance for OA and AP is explained by the unique environment. The dizygotic correlations are small, especially at the age of 3 years. Since the dizygotic correlations are less than half the monozygotic correlations, genetic dominance appears to be important. We note a steady increase in the dizygotic correlations across age. This suggests that the relative dominance genetic effect may decrease in importance when children become older. The somewhat larger dizygotic correlations for girls suggest that genetic dominance may play a larger role in explaining variance in boys as compared to girls. The correlations for the opposite sex twins are in line with the DZM and DZF correlations.

Genetic Modeling

We next turned to formal testing to evaluate the role of additive and dominance genetic effects, gender differences, and the presence of a contrast effect. The specified models and their fit indices are reported upon in Table V.

We fitted two series of models, with and without differences across gender. Because of the small DZ correlations, we argued that the resemblance between twins was more likely due to genetic additive and dominance effects instead of shared environmental effects at each age. This hypothesis was found true for each individual dataset. The model fitting results of models including a shared environmental factor are not reported here. For each individual dataset, we report upon four models that allowed parameter estimates to differ between boys and girls. These models were ADE-b, ADE and AE-b and AE. The good fit of the models that incorporated a contrast effect at the age of 3 years was in line with the earlier noted difference in variance between MZ and DZ twins at this age (see Table III). The contrast effect was found to account for a significant proportion of the total variance, evaluated by the AE model. Given our large sample size and the significant contribution of the contrast effect, there is little doubt upon the presence of the contrast effect at age 3 [Rietveld et al., in press]. However, we lack sufficient power to establish whether D, in addition to the contrast effect, is a necessary source of variance. A model with dominance alone was found insufficient to describe the data.

TABLE IV. Correlations by Age, Gender, and Zygosity

	Age (year)			
	3	7	10	12
MZM	0.63*	0.68*	0.70*	0.75*
DZM	0.08	0.15*	0.20*	0.25*
MZF	0.63*	0.70*	0.70*	0.70*
DZF	0.07	0.23*	0.30*	0.31*
DOS	0.11*	0.26*	0.28*	0.25*

*Correlation is significant at the 0.01 level.

At ages 7, 10, and 12 years, the AE-b model described the data as well as the ADE and ADE-b models. Based on the χ^2, degrees of freedom, and P-value, no distinction between these models could be made. The importance of the contrast parameter and the genetic dominance parameter was formally tested by the application of the AE model. The significant increase in χ^2 indicated the necessary inclusion of the contrast effect or genetic dominance effect or both the effects at the age of 7 and 10 years. This outcome was opposed to the outcome found for the oldest age group; the AE model did not lead to a significant deterioration in fit. However, because the age 12 years comprises the smallest sample size, we may have insufficient power to detect the presence of D. Therefore, we maintain both the ADE and AE model as best fitting models. Parameter b was estimated at 0.01 in the ADE-b model and we consider this effect as non-existent.

Parameters were constrained across gender in the next series of analyses. These reduced models were all nested with the full models, so the difference in χ^2 and degrees of freedom were used to identify the best fitting model. The examination of gender differences resulted in outcomes that varied with the age. Concluding, for the youngest age groups, no gender differences were found. The best fitting model at the age of 3 years was one that included the contrast parameter, AE-b and ADE-b. At the age of 7 years, we could not distinguish between an ADE-b, AE-b and an ADE model. For the oldest two age groups, the magnitude of the genetic and environmental effects was found to depend on gender. At the age of 10 years, in line with the age 7 years, ADE-b, AE-b, and ADE models were considered best fitting. The ADE-b model and the AE-b model was further reduced by specifying the contrast effect in boys to equal the effect in girls ($\Delta\chi^2 = 0.40$, df $= 1, P > 0.05$; $\Delta\chi^2 = 1.02$, df $= 1$, $P > 0.05$, respectively). Since the models differ much in their interpretation and because we lack sufficient power to distinguish between the models we report and discuss all three models. At the age of 12 years,

TABLE V. Model Fitting Results

		χ^2	df	P	Comparison	Change in χ^2 (df)
Age 3 year	With sex. diff.[a]					
1.	ADE-b	1.75	7	0.97		
2.	ADE	12.84	9	0.17	1.	11.09 (2), $P < 0.05$
3.	AE-b	1.93	9	0.99	1.	0.18 (2), ns
4.	AE	101.24	11	0.00	2.	88.4 (2), $P < 0.05$
					3.	99.31 (2), $P < 0.05$
	No sex. diff.[b]					
4.	**ADE-b**	**4.98**	**11**	**0.93**	**1.**	**3.23 (4), ns**
5.	ADE	16.29	12	0.00	5.	11.31 (1), $P < 0.05$
6.	**AE-b**	**5.14**	**12**	**0.95**	**5.**	**0.16 (1), ns**
Age 7 year	With sex. diff.[a]					
1.	ADE-b	7.54	7	0.38		
2.	ADE	7.96	9	0.54	1.	0.42 (2), ns
3.	AE-b	8.71	9	0.44	1.	1.17 (2), $P < 0.05$
4.	AE	32.14	11	0.00	2.	24.18 (2), $P < 0.05$
					3.	23.43 (2), $P < 0.05$
	No sex. diff.[b]					
5.	**ADE-b**	**10.14**	**11**	**0.52**	**1.**	**2.60 (4), ns**
6.	**ADE**	**10.30**	**12**	**0.59**	**5.**	**0.16 (1), ns**
7.	**AE-b**	**10.93**	**12**	**0.54**	**5.**	**0.79 (1), ns**
Age 10 year	With sex. diff.[a]					
1.	**ADE-b**	**6.13**	**7**	**0.53**		
2.	**ADE**	**6.70**	**9**	**0.67**	**1.**	**0.57 (2), ns**
3.	**AE-b**	**6.44**	**9**	**0.70**	**1.**	**0.31 (2), ns**
4.	AE	18.86	11	0.06	2.	12.16 (2), $P < 0.05$
					3.	12.42 (2), $P < 0.05$
	No sex. diff.[b]					
5.	ADE-b	17.19	11	0.10	1.	11.06 (4), $P < 0.05$
Age 12 year	With sex. diff.[a]					
1.	ADE-b	5.28	7	0.63		
2.	**ADE**	**5.57**	**9**	**0.78**	**1.**	**0.29 (2), ns**
3.	AE-b	7.57	9	0.58	1.	2.29 (2), ns
4.	**AE**	**10.17**	**11**	**0.52**	**2.**	**4.60 (2), ns**
					3.	**2.60 (2), ns**
	No sex. diff.					
5.	ADE	25.49	12	0.01	2.	19.92 (3), $P < 0.05$
6.	AE	29.16	13	0.01	4.	18.99 (2), $P < 0.05$

Best fitting models are in bold.
[a]Contrast parameter DOS is constrained to equal mean of male twin pairs and female twin pairs.
[b]Contrast parameter males is equal to females.

gender differences were detected for the ADE and AE model.

In addition to the genetic modeling of transformed data, we also adopted an alternative approach. Given the nature of the questionnaire items, one may argue that the CBCL measures only half of the distribution of OA and AP. That is, the questionnaire items relate to problem behavior only and do not measure the presence of "good concentration skills" or "excellent school work." From this point of view, we considered OA and AP to have an underlying continuity with a normal distribution. This underlying variable has been called the liability [Falconer and Mackay, 1996]. We created multiple categories with thresholds separating them. If a twin's score exceeds a certain threshold, he or she manifests the signs and symptoms that fit in with the next OA or AP category. It is assumed that many causes of variation, both genetic and environmental in origin combine to give the liability. Given that the manifestation of OA and AP is due to multiple causal factors [Faraone and Doyle, 2001], the liability model appears an adequate approach to analyze the data. This alternative approach, the application of the liability model, resulted in identical relative estimates for the genetic and environmental contributions to variance.

The contrast effect, and absolute estimates of the additive genetic, dominance genetic, and unique environmental contributions to the total variance are summarized in Table VI.

Varying with age and specified model, the contrast effect is estimated between -0.14 and -0.02. We find the largest effect at the age of 3 years, which decreases rapidly to insignificant at the age of 12 years. An accompanying interesting outcome is that the contrast effect was found to be of equal importance in boys and girls. This implies that, whether the contrast effect is considered a true interaction between siblings or a bias introduced by the parents, the twins' gender is of no importance. We report upon the absolute estimates of genetic and environmental variance for two reasons. First, these estimates clearly illustrate the significant gender differences that were detected at the age of 10 and 12 years. The total variance, calculated by summing the additive, dominance genetic and unique environmental effects, is much larger for boys as compared to girls. The gender difference appears strongest

TABLE VI. Estimates of Absolute Variance Components and Contrast Parameter

		Additive genetic	Unique environment	Dominance genetic	Total variance	Contrast
Age 3 year						
Boys/girls	ADE-b	0.464	0.200	0.097	0.761	−0.12
	AE-b	0.582	0.190	—	0.772	−0.14
Age 7 year						
Boys/girls	ADE-b	0.373	0.231	0.233	0.837	−0.02
	ADE	0.263	0.242	0.320	0.825	
	AE-b	0.658	0.212	—	0.870	−0.06
Age 10 year						
Boys	ADE-b	0.421	0.263	0.255	0.939	−0.02
Girls	ADE-b	0.526	0.224	0.093	0.843	
Boys	ADE	0.296	0.274	0.356	0.926	
Girls	ADE	0.424	0.234	0.172	0.830	
Boys	AE-b	0.721	0.274	—	0.968	−0.05
Girls	AE-b	0.661	0.208	—	0.869	
Age 12 year						
Boys	ADE	0.344	0.296	0.316	0.956	
Girls	ADE	0.528	0.217	0.035	0.780	
Boys	AE	0.653	0.313	—	0.966	
Girls	AE	0.564	0.221	—	0.785	

A, additive genetic; D, dominance genetic; E, unique environment; −b, contrast parameter.

for the genetic dominance effect. Girls have much smaller estimates as compared to boys. Based on the twin correlations, we had speculated on such a reduced importance of the dominance effect with increasing age for the girls. For the AE and AE-b models, the ratio additive genetic effects to unique environmental effects appears equal across gender. Such an equal ratio will result in similar relative estimates for the boys and girls. A second reason for the report of absolute estimates of variance has to do with the contrast effect. The absolute estimates presented here are applicable to twins, and given that these twins are representative of the population in general, to singletons. The absolute estimates are used for the calculation of the relative influences of genes and environment. With relative we mean the effect of a certain source of variance expressed as a proportion of the total variance. In our "Method" section, we mentioned that the contrast effect causes the total variance of a phenotype to depend on the genetic relationship of the individuals that are being rated. Because the effect is negative, monozygotic twins have smaller variances compared to dizygotic twins. The variance induced by the contrast effect is controlled for when we report upon the genetic and environmental influences for singletons. So, for models including a contrast effect, the calculation of relative estimates results in estimates that differ between MZ twins, DZ twins, and singletons. These relative estimates are shown in Table VII, separately for three groups of individuals who vary in genetic relatedness with one another.

Relative estimates are reported for MZ twins, DZ twins, and singletons for any model including the contrast effect. Because total variances are equal across pairs of varying genetic relatedness in the absence of a contrast effect, estimates derived from the fit of ADE and AE models are reported for singletons only. When the contrast effect is small (age 7 and 10 years), we barely note the differences in relative estimates across twins and singletons. When the contrast effect is large (age 3 years), the proportions of the total variance that is explained by genetic and environmental effects differ between MZ twins, DZ twins and singletons.

Consistent across age, gender, and model, genetic effects are most important in explaining observed variance in OA and AP. Broad heritability, the sum of additive and dominance genetic effects, varies between 68% and 76% of the total variance. From a power study, we have learned that when D is excluded from the model, the heritability and contrast effect estimated by the AE-b model may have been biased upwards [Rietveld et al., in press]. Likewise, ignoring the presence of a contrast effect by accepting an ADE model may have lead to inflated estimates of heritability [Eaves et al., 1997]. The residual variance is explained by the unique environment, whereas the importance of the additive and dominance effect varies with the age, the unique environmental effect continues to account for around one third of the variance across the age. The stability of the magnitude of the unique environmental effects was already suggested by the stable difference between unity and the MZ correlations.

DISCUSSION

In this study, the genetic and environmental contributions to attention problem behavior are estimated by taking advantage of large sample sizes of twin pairs for separate age cohorts. We asked and answered the following questions: What are the genetic and environmental contributions to overactivity at the age of 3 years and attention problems at ages 7, 10, and 12 years? Are these contributions of equal magnitude at each age, or do they vary as a function of age? Are there differences in the genetic and environmental influences by gender? Finally, models were fitted that allowed to test for contrast effects, in order to determine if maternal rating habits or interaction among siblings takes place. Each issue is presented in some detail. The report is completed

TABLE VII. Estimates of Relative Variance Components

		Additive genetic			Unique environment			Dominance genetic		
		MZ (%)	DZ (%)	Singletons (%)[a]	MZ (%)	DZ (%)	Singletons (%)[a]	MZ (%)	DZ (%)	Singletons (%)[a]
Age 3 year	ADE-b	57	59	61	31	28	26	12	13	13
Boys/girls	AE-b	69	73	75	31	27	25	—	—	—
Age 7 year										
Boys/girls	ADE-b	44	44	44	29	28	28	27	28	28
	ADE	—	—	32	—	—	29	—	—	39
	AE-b	73	74	76	27	26	24	—	—	—
Age 10 year										
Boys	ADE-b	44	44	45	28	28	28	27	27	27
Girls	ADE-b	62	62	62	27	27	27	11	11	11
Boys	ADE	—	—	32	—	—	30	—	—	38
Girls	ADE	—	—	51	—	—	28	—	—	21
Boys	AE-b	72	73	74	28	27	26	—	—	
Girls	AE-b	74	75	76	26	25	24	—	—	
Age 12 year										
Boys	ADE	—	—	36	—	—	31	—	—	33
Girls	ADE	—	—	68	—	—	28	—	—	5
Boys	AE	—	—	68	—	—	32	—	—	—
Girls	AE	—	—	72	—	—	28	—	—	—

A, additive genetic; D, dominance genetic; E, unique environment; −b, contrast parameter.
[a]Corrected for any variance induced by the contrast effect; these estimates are usually reported by other behavior genetic studies of ADHD phenotypes for which a contrast effect is detected.

by discussion of clinical implications and limitations of this study.

Phenotypic Findings

With respect to the differences between genders, boys are reported as displaying more overactivity and attention problems as compared to girls. For the subsample that is likely to meet a DSM diagnosis of ADHD, we observed a boy-to-girl ratio of 1.5 to 1 at ages 3, 10, and 12 years. A gender ratio of 2 to 1 was observed at the age of 7 years. In line with our outcome, the meta-analysis by Gaub and Carlson [1997] indicated that non-referred samples of girls with ADHD relative to the boys with ADHD show indeed lower levels of attention problems. Although not all studies agree on the presence of varying degree of behavioral problems between pre-school boys and girls [Campbell, 1995], we report upon a significant gender effect for overactive behavior in 3-year-old children. An increase in symptom counts from preschool age to school age was observed for boys. Entering puberty, both boys and girls displayed a decrease in problem behavior. The age-dependent decline of inattention symptoms confirms the outcomes reported by Biederman et al. [2000]. Unfortunately, even though the inattention symptoms decline, it is known from multiple studies that many of these children continue to have emotional and behavioral problems, sometimes severe [Rutter et al., 1998].

Genetic and Environmental Estimates Across the age

For the overactivity syndrome at the age 3 years and the attention problems syndrome at ages 7, 10, and 12 years, we analyzed several models, and in each case several models fitted the data. Each acceptable model included genetic, either additive or additive and dominance effects, and unique environmental effects. The broad heritability estimates, when dominance and additive relative estimates are summed, were high, ranging from 68% to 76% across age. In the previous research on attention problems and associated phenotypes, additive genetic influences ranging from 60% to 80% were found, depending on the model and the gender of the subjects [Nadder et al., 1998; Simonoff et al., 1998; Rhee et al., 1999; Eaves et al., 2000; Hudziak et al., 2000; Thapar et al., 2000; Martin et al., 2002]. Environmental influences that are unique to the individual contribute to one third of the total variance observed for overactivity and attention problem behavior. Shared environmental factors, those that influence siblings more similar are negligible. The consistency in results across these large twin studies is notable, considering the variety in assessment instruments, method of data collection, methods of analyses, and age of the twins.

As opposed to the large number of behavior genetic studies on problem behavior in school aged children, only few studies have reported upon activity in preschool children [Price et al., submitted; Saudino et al., 2000]. Given the large samples, there is now convincing evidence to conclude that hyperactive/overactive behavior in pre-school children is largely genetic with estimated heritability of around 70%. In addition to the additive genetic effects, contrast effects explain a significant proportion of the variance in hyperactive/overactive behavior at this young age. With respect to the genetic dominance, this source of variance in preschool children can not be excluded. Numerous other studies have reported upon dominance effects for prepubertal children or young adolescents. However, they have done so by means of analyses of even smaller datasets that are often pooled across age [see for review

Thapar et al., 1995; Kuntsi and Stevenson, 2000; Kuntsi et al., 2000]. If data of different ages are merged, a possible diminishing impact of genetic dominance and/or contrast effect at older ages may be masked by the prominent presence of this source of variance at younger ages. Thus it remains unclear from prior studies if dominance and/or contrast effects are prominent when children enter puberty. Despite this shortcoming, given the large agreement across studies, there is little doubt on the importance of (broad) heritability in explaining variance for attention problems and related symptoms, from infancy to young adulthood.

Gender Differences

Another important finding is the presence of gender differences in the variance estimates at ages 10 and 12 years. Here, the evidence supports larger variances for boys at the ages 10 and 12 years. A possible explanation for this difference in variance is that the genetic effect, which is shared by both genders, has been amplified in boys as compared to girls. When variance components are expressed as a proportion of the total variance, we obtained equal estimates for broad heritability and the unique environment. This implies that the pattern of additive genetic, dominance genetic, and unique environmental effects is similar in boys and girls. With the exception of Rhee et al. [1999], this outcome agrees with reports by other large twin studies [Nadder et al., 1998; Thapar et al., 2000].

Contrast Effect

An interesting observation appeared when means across zygosities were evaluated. Girls with a male co-twin were rated to display less overactivity and attention problems compared to girls with a female co-twin. This effect was not reversed; boys with a female co-twin were rated as having an equal degree of problem behavior as other twin boys. This finding suggests that the male twin serve as some kind of protective factor for his female sibling. The mechanism underlying this protective factor is speculative. Given the less likely situation that the opposite-sex female twins truly display less overactivity and attention problems as compared to any other group of twins, the mechanism appears to be related to the person who rates the behavior. Since overactivity and attention problems are more recognized and often occur with other externalizing behaviors in boys [Gaub and Carlson, 1997], mothers may tend to evaluate the behavior of their daughter in comparison to their son as 'not as much.' This speculation implies a bias in the mean values introduced by the mother. In our genetic modeling, evaluation of bias took place by the analysis of variance. We considered three types of contrast parameters, one for male same-sex twins, one for female same-sex twins, and one for the opposite-sex twins. In our preliminary analyses, we established that the magnitude of the contrast effect in opposite-sex twins did not differ from the mean value of the contrast effect in male pairs and female pairs. This finding does not exclude the possibility that, only in opposite-sex pairs, gender does play a role in explaining a part of the variance. So, instead of the specification of one contrast effect, it may be interesting to disentangle the contrast further into an effect from-boy-to-girl and an effect from-girl-to-boy [Eaves et al., 2000]. Post hoc, we performed such an exploration of the contrast effect in 3-year-old opposite-sex twins. In line with our expectation, the effect from-boy-to-girl was found to be much larger compared to the effect from-girl-to-boy (-0.17 vs. -0.05).

The largest contrast effect was detected at the age of 3 years. With respect to our sample size and the magnitude of the effect ($-0.12/-0.14$), we have sufficient power to confirm the presence of the contrast effect. Several studies on ADHD and related phenotypes have reported upon the absence of a contrast effect in teacher- or observer-data [Sherman et al., 1997; Simonoff et al., 1998; Kuntsi et al., 2000; Saudino et al., 2000; Nadder et al., 2001; Martin et al., 2002]. These outcomes confirm our conclusion based on the pattern in means and point to a rater bias effect. The implication of a rater bias in diagnosis and research is clear. Bias can lead to misdiagnoses in the clinical setting and the inclusion of false positives or exclusion of true cases from gene searching efforts, both of which are undesirable. If a gene finding study is designed to select discordant twin pairs and concordant twin pairs, the former group would be over-represented and the latter group would be under-represented due to maternal rater bias.

The Dutch overactivity scale and the attention problem scale of the CBCL appear a suitable instrument for case identification, as part of the sample selection for gene searching research. We do not suggest that these scales should be used as the sole measuring stick for molecular genetic studies of ADHD, but rather as a phenotypic marker to improve our ability to minimize false positives and negatives.

Clinical Implications, Limitations, and Future Research

These data may help us understand similarities and differences between studies of DSM-IV ADHD, that identify four times more boys than girls as having ADHD, versus studies of attention problems, for which the gender ratio is much less marked. The DSM lacks normative data by gender whereas the CBCL compares deviance in boys and girls only against data on other boys and girls, respectively, of the same age. To date there have been no compelling arguments to explain the gender differences in the prevalence of ADHD. Perhaps part of the explanation has to do with the taxonomy and gender-genetic factors. Because the CBCL is normed by age, gender, and is less likely to be affected by rater contrast, the data support the consideration of using the attention problems scale in the clinical setting. The application of the attention problems scale may improve the ability to screen for ADHD in siblings of ADHD children when maternal endorsement of DSM items do not support the diagnoses of ADHD in a 'less affected sibling.' Similarly, because the CBCL is normed by gender, use of this scale may allow the clinician to identify girls who are at risk for ADHD, who might not

appear as deviant as their brothers based on the fact that all girls have fewer symptoms than boys. Finally, consideration of these data in the clinical setting may increase the awareness of the quantitative nature of overactivity and attention problems in children with emotional and behavioral problems. Instead of requiring a child to meet an absolute number of symptoms, the clinician can use this quantitative scale to assess severity, social, personal, and emotional impairment, and treatment response.

With respect to limitations of the study we note the following. Our genetic modeling results indicate that genetic influences are important and fairly stable in magnitude over the course of a 9-year developmental period. The development of attention problems has been well documented [Barkley et al., 1990; Weiss and Hechtman, 1993]. Here we do not comment on longitudinal aspects of the data but focus on heritability at each age. Currently, data from additional birth-cohorts are added to the dataset. The enlarged sample of the same twins assessed at multiple occasions in time makes longitudinal analyses possible. This work is now underway and results will appear in a separate report.

A second limitation from a clinical point of view is that we use the CBCL to assess for AP and not DSM-IV criteria. Thus, although AP is highly predictive of ADHD, it is not ADHD. Although data have been presented on the relations between AP and ADHD, conclusions about AP may not generalize to genetic studies of ADHD. The oucomes obtained by a large twin study confirm the variety in outcomes when different assessment instruments are used [Thapar et al., 2000].

We mentioned that we obtained likewise results by analyses of paternal data, compared to the maternal data which we have reported here. Although it was established that a rater bias effect is of minimal influence beyond age the of 3 years, this does not exclude the possibility that mothers and fathers underscore the presence of inattention in their children [Klein and Mannuzza, 1991]. Recently, the study is extended with the collection of teacher report data, youth self report data, and DSM-IV interview data. This wealth of data enables us to investigate the genetic and environmental influences on the development of attention problems from multiple points of view.

ACKNOWLEDGMENTS

We thank Professor Hans Koot who kindly provided his CBCL dataset on 3-year-old singletons.

REFERENCES

Achenbach TM. 1991. Manual for the child behavior checklist/4-18. Burlington, VT: University of Vermont Department of Psychiatry.

Barkley RA, Fisher M, Edelbrock CS, Smallish L. 1990. The adolescent outcome of hyperactive children diagnosed by research criteria: I. An 8-year prospective follow-up study. J Am Acad Child Adolesc Psychiatr 29:546–557.

Barr C, Swanson J, Kennedy J. 2001. Molecular genetics of ADHD. In: Levy F, Hay DA, editors. Attention, genes and ADHD. East Sussex: Brunner-Routledge. p 173–195.

Biederman J. 1998. Attention-deficit/hyperactivity disorder: A life-span perspective. J Clin Psychiatry 59:4–16.

Biederman J, Faraone SV, Doyle A, Lehman BK, Kraus I, Perrin J, Tsuang MT. 1993. Convergence of the child behavior checklist with structured interview-based psychiatric diagnoses of ADHD children with and without comorbidity. J Child Psychol Psychiatr 34:1241–1251.

Biederman J, Mick E, Faraone SV. 2000. Age-dependent decline of symptoms of attention deficit hyperactivity disorder: Impact of remission definition and symptom type. Am J Psychiatry 157:816–818.

Bird HR, Canino G, Rubio-Stipec M, Gould MS, Ribera J, Sesman M, Woodbury M, Huertas-Goldman S, Pagan A, Sanchez-Lacay A, Moscoso M. 1988. Estimates of the prevalence of childhood maladjustment in a community survey in Puerto Rico. The use of combined measures [published erratum appears in *Arch Gen Psychiatr 1994, 51:429*]. Arch Gen Psychiatr 45:1120–1126.

Boomsma DI. 1998. Twin registers in Europe: An overview. Twin Res 1:34–51.

Campbell SB. 1995. Behavior problems in preschool children: A review of recent research. J Child Psychol Psychiatr 36:113–149.

Carey G. 1986. Sibling imitation and contrast effects. Behav Genet 16:319–341.

Chen WJ, Faraone SV, Biederman J, Tsuang MT. 1994. Diagnostic accuracy of the child behavior checklist scales for attention-deficit hyperactivity disorder: A receiver-operating characteristic analysis. J Consult Clin Psychol 62:1017–1025.

Eaves LJ. 1976. A model for sibling effects in man. Heredity 36:205–214.

Eaves LJ, Silberg JL, Maes HH, Simonoff E, Pickles A, Rutter M, Neale MC, Reynolds CA, Erikson MT, Heath AC, Loeber R, Truett KR, Hewitt JK. 1997. Genetics and developmental psychopathology: 2. The main effects of genes and environment on behavioral problems in the Virginia twin study of adolescent behavioral development. J Child Psychol Psychiatr 38:965–980.

Eaves LJ, Rutter M, Silberg JL, Shillady L, Maes HH, Pickles A. 2000. Genetic and environmental causes of covariation in interview assessments of disruptive behavior in child and adolescent twins. Behav Genet 30:321–334.

Edelbrock C, Costello AJ. 1988. Convergence between statistically derived behavior problem syndromes and child psychiatric diagnoses. J Abnorm Child Psychol 16:219–231.

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

Falconer DS, Mackay TFC. 1996. Introduction to quantitative genetics (4th ed.). Essex: Addison Wesley Longman Limited. 299 p.

Faraone SV, Doyle AE. 2001. The nature and heritability of attention-deficit/hyperactivity disorder. Child Adolesc Psychiatr Clin N Am 10:299–316.

Gaub M, Carlson CL. 1997. Gender differences in ADHD: A meta-analysis and critical review. J Am Acad Child Adolesc Psychiatr 36:1036–1045.

Hartung CM, Widiger TA. 1998. Gender differences in the diagnosis of mental disorders: Conclusions and controversies of the DSM-IV. Psychol Bull 123:260–278.

Hudziak JJ. 2001. The role of phenotypes (diagnoses) in genetic studies of attention-deficit/hyperactivity disorder and related child psychopathology. Child Adolesc Psychiatr Clin N Am 10:279–297.

Hudziak JJ. 2002. The importance of phenotype definition in genetic studies of child psychopathology. In: Helzer JE, Hudziak JJ, editors. Defining psychopathology in the 21st century: DSM-V and beyond. American Psychiatric Publishing Group, Inc. 211–230.

Hudziak JJ, Rudiger LP, Neale MC, Heath AC, Todd RD. 2000. A twin study of inattentive, aggressive and anxious/depressed behaviors. J Am Acad Child Adolesc Psychiatr 39:469–476.

Hudziak JJ, Heath AC, Madden PAF, Reich W, Bucholz KK, Slutske WS, Bierut LJ, Neuman RJ, Todd RD. 2002. In review. The genetic analysis of parental reports of DSM-IV. J Am Acad Child Adolesc Psychiatr.

Jöreskog KD, Sörbom D. 1993. New features in PRELIS 2. Chicago: Scientific Software International, Inc.

Klein RG, Mannuzza S. 1991. Long-term outcome of hyperactive children: A review. J Am Acad Child Adolesc Psychiatr 30:383–387.

Koot HM, van den Oord EJCG, Verhulst FC, Boomsma DI. 1997. Behavioural and emotional problems in young pre-schoolers: Cross-cultural testing of the validity of the child behavior checklist 2/3. J Abnorm Child Psychol 25:183–196.

Kuntsi J, Stevenson J. 2000. Hyperactivity in children: A focus on genetic research and psychological theories. Clin Child Fam Psychol Rev 3: 1–23.

Kuntsi J, Gayán J, Stevenson J. 2000. Parents' and teachers' ratings of problem behaviours in children: Genetic and contrast effects. Twin Res 3:251–258.

Loehlin JC. 1992. Latent variable models: An introduction to factor, path, and structural analysis. Hillsdale, New Yersey: Lawrence Erlbaum Associates.

Martin N, Scourfield J, McGuffin P. 2002. Observer effects and heritability of childhood attention-deficit hyperactivity disorder symptoms. Br J Psychiatr 180:260–265.

Nadder TS, Silberg JL, Eaves LJ, Maes HH, Meyer JM. 1998. Genetic effects on ADHD symptomatology in 7- to 13-year-old twins: Results from a telephone survey. Behav Genet 28:83–100.

Nadder TS, Silberg JL, Rutter M, Maes HH, Eaves LJ. 2001. Comparison of multiple measures of ADHD symptomatology: A multivariate genetic analysis. J Child Psychol Psychiatr 42:475–486.

Neale MC. 1997. Mx: Statistical modeling. Richmond, VA: Department of Psychiatry, Medical College of Virginia.

Neale MC, Cardon LR, North Atlantic Treaty Organization. Scientific Affairs Division. 1992. Methodology for genetic studies of twins and families. Dordrecht: Kluwer Academic Publishers.

Neale MC, Stevenson J. 1989. Rater bias in the EASI temperament scales: A twin study. J Pers Soc Psychol 56:446–455.

Offord DR, Boyle MH, Szatmari P, Rae-Grant NI, Links PS, Cadman DT, Byles JA, Crawford JW, Blum HM, Byrne C, et al. 1987. Ontario child health study. II. Six-month prevalence of disorder and rates of service utilization. Arch Gen Psychiatr 44:832–836.

Plomin R, DeFries JC, McClearn GE, McGuffin P. 2001. Behavior genetics (4th ed.). New York: W.H. Freeman.

Rhee SH, Waldman ID, Hay DA, Levy F. 1999. Sex differences in genetic and environmental influences on DSM-III-R attention-deficit/hyperactivity disorder. J Abn Psychol 108:24–41.

Rietveld MJH, Posthuma D, Dolan CV, Boomsma DI. ADHD: sibling interaction or dominance. An evaluation of statistical power. Beh Genet (in press).

Rietveld MJH, van der Valk JC, Bongers IL, Stroet TM, Slagboom PE, Boomsma DI. 2000. Zygosity diagnosis in young twins by parental report. Twin Res 3:134–141.

Ross DM, Ross SA. 1982. Hyperactivity. Current issues, research, and theory. New York: John Wiley & Sons, Inc.

Rutter M, Giller H, Hagell A. 1998. Antisocial behavior by young people. New York/Cambridge: Cambridge University Press.

Saudino KJ, Cherny SS, Plomin R. 2000. Parent ratings of temperament in twins: Explaining the 'too low' DZ correlations. Twin Res 3:224–233.

Schmitz S, Fulker DW, Mrazek DA. 1995. Problem behavior in early and middle childhood: An initial behavior genetic analysis. J Child Psychol Psychiatr 36:1443–1458.

Sherman DK, Iacono WG, McGue MK. 1997. Attention-deficit hyperactivity disorder dimensions: A twin study of inattention and impulsivity-hyperactivity. J Am Acad Child Adolesc Psychiatr 36:745–753.

Simonoff E, Pickles A, Hervas A, Silberg JL, Rutter M, Eaves L. 1998. Genetic influences on childhood hyperactivity: contrast effects imply parental rating bias, not sibling interaction. Psychol Med 28:825–837.

Steingard R, Biederman J, Doyle A, Sprich-Buckminster S. 1992. Psychiatric comorbidity in attention deficit disorder: Impact on the interpretation of child behavior checklist results. J Am Acad Child Adolesc Psychiatr 31:449–454.

Thapar A, Hervas A, McGuffin P. 1995. Childhood hyperactivity scores are highly heritable and show sibling competition effects: Twin study evidence. Behav Genet 25:537–544.

Thapar A, Harrington R, Ross K, McGuffin P. 2000. Does the defninition of ADHD affect heritability? J Am Acad Child Adolesc Psychiatr 39:1528–1536.

Weiss G, Hechtman L. 1993. Hyperactive children grown up. ADHD in children, adolescents and adutls. (2nd ed.). New York: Guilford Press.

Journal of Child Psychology and Psychiatry 45:3 (2004), pp 577–588

Heritability of attention problems in children: longitudinal results from a study of twins, age 3 to 12

M.J.H. Rietveld,[1] J.J. Hudziak,[2] M. Bartels,[1] C.E.M. van Beijsterveldt,[1] and D.I. Boomsma[1]

[1]Department of Biological Psychology, Vrije Universiteit, Amsterdam, The Netherlands; [2]Department of Psychiatry and Medicine (Division of Human Genetics), Center for Children, Youth and Families, University of Vermont, USA

Background: Twin studies of childhood behavior problems support the conclusion that individual differences in impulsivity, hyperactivity, and inattention are largely due to genetic influences. Non-genetic variation is due to environmental influences that are unique to the individual, and possibly to rater contrast effects. In the present longitudinal twin study, we report on the size of genetic and environmental effects on individual differences in attention problems at ages 3, 7, 10 and 12 years. **Methods:** Mothers were asked to complete the CBCL for their twin offspring when the children were 3 (n = 11,938), 7 (n = 10,657), 10 (n = 6,192), and 12 years old (n = 3,124). We focus on the Overactivity (OA) scale in the Child Behavior Checklist (CBCL/2–3), and on the Attention Problem (AP) scale of the CBCL/4–18. The data were analyzed using longitudinal structural equation modeling. **Results:** Broad heritability of OA and AP is estimated at nearly 75%, at each age. A contrast effect was observed at age 3 only. The results revealed less stability of OA at age 3 to AP at age 7 (r = .40), compared to the stability from AP at age 7 and beyond (r = .70). Genetic effects explained between 76% and 92% of the covariance between OA and AP. **Conclusions:** OA and AP are highly heritable at all ages in both genders. The same set of genes appears to be expressed in boys and girls. The size of genetic and environmental contributions remains the same across the ages studied. Stability in OA and AP is accounted for by genetic influences. Children who do not display OA or AP at a given age are unlikely to develop these problems at a subsequent age. **Keywords:** Overactivity, attention problems, heritability, twin study, repeated measures. **Abbreviations:** OA = overactivity; AP = attention problems; CBCL = Child Behavior Checklist.

Attention Problems (AP) and Overactive behavior (OA) represent syndromes that can be measured by parental ratings on the Child Behavior Checklist (CBCL). OA was identified as a syndrome in factor analyses of responses of parents of Dutch pre-school children (Achenbach, 1992; Koot, Van den Oord, Verhulst, & Boomsma, 1997; Van den Oord, 1993). OA is predictive of, and developmentally related to, AP, a syndrome that emerged in factor analyses of responses of parents rating their children aged 6 to 18 years (Achenbach, 1991). In earlier work (Rietveld, Hudziak, Bartels, Van Beijsterveldt, & Boomsma, 2003a) we demonstrated that genetic influences on maternally rated OA in 3-year-olds explain 75% of the phenotypic variance. Genetic influences remained the most important source of variance on maternally rated AP at ages 7, 10 and 12 years. In this previous work cross-sectional analyses were carried out. Here, we present results of longitudinal analyses of these and additional data.

To date, several behavior genetic studies have incorporated the CBCL. Edelbrock, Rende, Plomin, and Thompson (1995) reported a heritability estimate of 66% for AP measured in twins aged 7 to 15 years. Similar estimates were reported by Schmitz, Saudino, Plomin, Fulker, and DeFries (1996) for 7-year-old children, by Zahn-Waxler, Schmitz, Fulker, Robinson, and Emde (1996) for 5-year-old children, and by Hudziak, Rudiger, Neale, Heath, and Todd (2000) for children in an age range from 8 to 12 years. The study by Gjone, Stevenson, and Sundet (1996) focused on two age ranges, 5 to 9 years and 12 to 15 years, and reported equal results over this period. Van den Oord, Verhulst, and Boomsma (1996) studied OA in 3-year-old Dutch twins, and found that more than half of the observed variation was explained by genetic influences. Their sample is included in the present study. Behavior genetic studies have also addressed the genetic and environmental influences on phenotypes other than OA and AP, but associated with ADHD (Kuntsi, Gayán, & Stevenson, 2000; Martin, Scourfield, & McGuffin, 2002; Thapar, Hervas, & McGuffin, 1995; Saudino, Cherny, & Plomin, 2000; Schmitz, Fulker, & Mrazek, 1995; Sherman, Iacono, & McGue, 1997). These studies have produced varying results, which may be due to the use of different assessment instruments (Nadder, Silberg, Rutter, Maes, & Eaves, 2001; Thapar, Harrington, Ross, & McGuffin, 2000). However, there is considerable consistency in the results. The larger studies including 1,000 or more twin pairs (Eaves et al., 1997; Nadder, Silberg, Eaves, Maes, & Meyer, 1998; Price et al., submitted; Rhee, Waldman, Hay, & Levy, 1999) reported heritability estimates between 60% and 85%. The environmental effects, which account for most of the remaining variance, are unique or unshared. The unique environment encompasses experiences that are specific to an individual and

Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA

make members of the same family different from one another. Environmental experiences, which are shared among family members, are negligible in explaining variation in overactivity, inattention, and impulsivity. Another consistent finding is the presence of a contrast effect in parental data on ADHD phenotypes. This effect shows up as a negative interaction between the twins' phenotypes (Carey, 1986; Eaves, 1976). The interaction appears to be due to bias, which originates in the rating of the parents, rather than a true social interaction between twins (Simonoff et al., 1998). When the contrast effect is present but not included in the analyses, heritability estimates are overestimated. In our previous report, contrast effects were detected at age 3, but the results were inconclusive at older ages. However, contrast effects are very difficult to distinguish from genetic dominance effects in studies of MZ and DZ twins only (Rietveld, Posthuma, Dolan, & Boomsma, 2003b).

In twin studies of AP and ADHD phenotypes, gender differences in heritability have received little attention, and the available results are largely inconclusive. One large study found gender effects in the magnitude of genetic and environmental influences on ADHD (Rhee et al., 1999), whereas two large twin studies reported the absence of gender effects for ADHD and related symptoms (Nadder et al., 1998; Thapar et al., 2000). In our cross-sectional study we found similar heritability estimates in boys and girls at ages 3, 7, 10, and 12 years.

From the literature on developmental child psychology (Biederman, 1998; Biederman, Mick, & Faraone, 2000; Ross & Ross, 1982; Rutter, Giller, & Hagell, 1998) it becomes clear that ADHD phenotypes follow a distinct pattern from infancy to early adolescence. Hyperactivity may be prominent during pre-school age and early school age, but it is less prevalent in older subjects. The inability to sustain attention is most profound during the school age years, but tends to decline in early adolescence, when antisocial behavior becomes more pronounced.

In order to study the pathways into and out of AP, Koot (1993) explored the relations of the Dutch OA syndrome in a community sample of 2- and 3-year-old children, and AP measured at follow-up two years later. Achenbach and Rescorla (2000) performed identical analyses for an adjusted AP scale in pre-school children and AP in school-aged children. The inter-occasion correlations of .40 and .47 reported in these studies suggest that stability is moderate. These correlations are lower than the .60 obtained by Verhulst and Van der Ende (1995) for AP itself over a period that covered 2 and 4 years.

With the exception of the studies by Schmitz et al. (submitted) and Van der Valk, Verhulst, Neale, and Boomsma (1998), longitudinal twin studies of AP and ADHD are lacking. Schmitz et al. (submitted) explored the stability and covariance of temperament and attention problems as measured with the CBCL at ages 7 and 12 years. The phenotypic correlation of maternally reported AP between these ages was .54. This stability in AP was explained mainly by genetic influences. These results are comparable to those reported in a study of adolescent adoptees, where phenotypic stability in AP during adolescence was estimated around .60 (Van der Valk et al., 1998). It was found that genetic effects and environmental effects contributed equally to the stability of attention problems during adolescence. As reported in the study of Schmitz et al. (submitted), environmental effects were unique to the individual.

The contributions of genetic and environmental effects to the (in)stability of OA and AP during childhood remain to be investigated. The aim of the present twin study is to evaluate the genetic and environmental effects on overactivity and attention problem behavior during childhood, in boys and girls. OA was measured in 3-year-old twins, and AP was measured in the same twins at ages 7, 10, and 12 years. This twin study is the first to cover a time period of this length. By using structural equation modeling of the longitudinal twin data, we address the following issues. Is phenotypic stability attributable to genetic influences, environmental influences, or a combination of both? Are the contributions of genetic and environmental factors stable across the entire age range (age 3 to age 12), or do they change during development? Are there gender differences in the genetic and environmental contributions to the inter-occasion correlations?

Materials and methods

Subjects and procedure

The Netherlands Twin Registry (NTR; Boomsma, 1998) was established in 1986 and is maintained by the Department of Biological Psychology at the Vrije Universiteit in Amsterdam. For the present study, we assessed a sample of Dutch twin pairs, whose parents reported on their behavior at ages 3, 7, 10, and 12 years. Data from 3- and 7-year-old twins were collected from birth cohorts 1986 through 1993. The dataset of the 10-year-olds comprised data from birth cohorts 1986 through mid-1991. Data of 12-year-old twins were available from birth cohorts 1986 through mid-1989/1990. We collected 11,938 maternal reports for the 3-year-olds, 10,657 maternal reports for the 7-year-olds, 6,192 maternal reports for the 10-year-olds, and 3,124 maternal reports of the 12-year-olds.

Discriminant analysis of questionnaire items was used to classify the same-sex pairs into MZ and DZ twins (Rietveld et al., 2000). The number of participating twins by maternal report and their assigned zygosity are listed in Table 1.

The number of individuals with complete maternal data at all ages is 2,192. At each age, around 2% of the twins suffered from a disability or disease that

Table 1 Number of individual twins separately by zygosity status

	Age 3	Age 7	Age 10	Age 12
Unknown	282	11	1	2
MZM	1870	1771	1054	579
DZM	1854	1692	954	472
MZF	2146	2012	1249	638
DZF	1743	1639	933	466
DOS	3784	3289	1835	876
Total	11679	10414	6026	3033

Note: The number of individuals with unknown zygosity at age 3 is due to the nonparticipation of these families at later ages. At age 3, these families provided incomplete information on the zygosity questionnaire.

interfered severely with their daily functioning. These twins displayed more than twice as much problem behavior across the entire age range compared to healthy twins. In addition, the co-twins, although unaffected, displayed increased total problem scores. These twin pairs were dropped from the sample.

Parental occupation was assessed using a 5-point scale, ranging from manual labor to academic employment. The level of occupation of the parents of the twins is somewhat higher than the level in the general population (Centraal Bureau voor de Statistiek, 2002a). This may be due to the minimal educational level that is required to read and complete questionnaires, or it may reflect a difference between adults with children and adults in the population as a whole.

Measures

The Child Behavior Checklist (CBCL) is a standardized questionnaire for parents to rate the frequency and intensity of behavioral and emotional problems exhibited by their children in the past six months. Parents rate each item on a 3-point scale. The total problem score is obtained by summing the item scores. The questionnaire that is used for the 3-year-olds (CBCL/2-3; Achenbach, 1992) differs in item content from the questionnaire that is used for the 7-, 10-, and 12-year-olds (CBCL/4-18; Achenbach, 1991). Koot et al. (1997) report the Dutch syndrome scales of the CBCL/2-3, and comparability of these with the American syndrome scales. An overactive scale was identified in a Dutch dataset, which has no American equivalent. Syndrome scales of the CBCL/4-18 were composed according to the profile developed by Achenbach (1991; De Groot, Koot, & Verhulst, 1994). The content of the Dutch overactive scale and attention problem scale is given in Appendix I.

Statistical analyses

The statistical modeling program Mx (Neale, 1997) was used to calculate descriptive statistics of OA and AP, and to test for mean differences between genders and across age. In these tests, likelihood-ratio chi-square (χ^2) tests were used. These tests control for the dependency that exists between the scores from twins. Selective nonparticipation at ages 7, 10, and 12 was evaluated by application of the general linear model (GLM) with repeated measures in SPSS. In these analyses, the OA score of a twin was considered a repeated measure of his or her co-twin's OA score. Participation at follow-up ages (yes/no) was specified as the between-group factor. Selective nonparticipation may change the composition of the sample at ages 7, 10, and 12 compared to the original sample at age 3. A non-significant effect implies that the age-specific samples may be considered to be 'selected completely at random' in the sense of Little and Rubin (1987). If nonparticipation is related to the OA score in the beginning of the study, one may still assume that the sample is 'selected at random'. This assumption is tenable when the willingness of mothers to return the questionnaire depended on variables other than OA but in some way related to OA. If the sample is 'selected completely at random' or 'selected at random', correct maximum likelihood (ML) estimates are obtained if raw data likelihood estimation is used and all available data are included (Wothke, 2000). Here, raw data ML estimation was carried out using Mx. This method requires that the data are normally distributed. The data were square root transformed to render their distribution approximately normal.

Genetic modeling

Figure 1 depicts the baseline model at one time point used for the analyses. The circles represent the latent, unmeasured variables and the squares represent the observed and measured phenotypes.

Phenotypic variance is modeled as the sum of genetic and environmental variances. We distinguish two categories of genetic sources of variance: additive effects of alleles at a large number of loci (A), and non-additive genetic effects reflecting interaction effects between alleles of the same gene locus (dominance, D). The environmental source of variance of OA and AP is con-

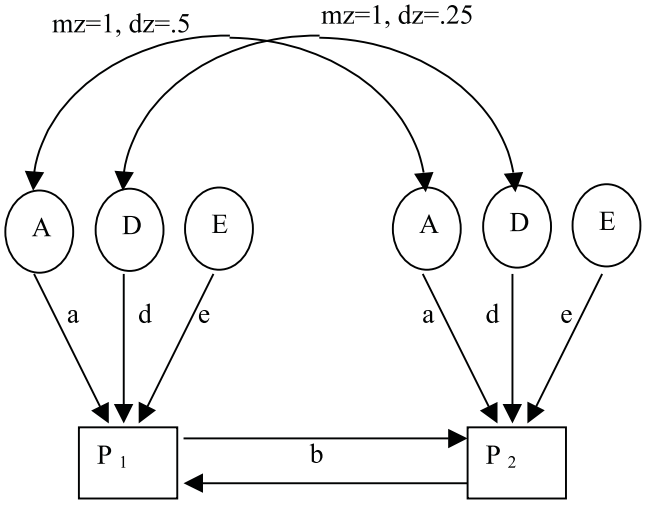


Figure 1 Path diagram of ADE model with contrast effect. P_1 and P_2 represent the observed phenotypes of twin 1 and twin 2, respectively. A, D, and E are additive genetic, genetic dominance, and unique environmental latent factors, respectively; a, d, and e are path-coefficients; b represents the negative contrast effect. Genetic correlations between twins are represented by double-headed arrows between the A and D factors.

sidered to be unique to the individual. These environmental influences (denoted E in Figure 1) contribute to differences between individuals within the same family. Estimates of these unique environmental contributions to the total variance include measurement error. The coefficients a, d, and e are factor loadings of P on the latent factors A, D and E. Given that the latent factors are constrained to have unit variance (Neale et al., 1992), the decomposition of the phenotypic variance (Vp) is:

$$Vp = a^2 + d^2 + e^2$$

The different degree of genetic relatedness of monozygotic (MZ) and dizygotic (DZ) twins is used to identify the contributions of the latent factors to the phenotypic variance of OA and AP (Plomin, DeFries, McClearn, & McGuffin, 2001). Similarity (covariances) of MZ twins may be due to additive genetic influences and genetic dominance, that is, $a^2 + d^2$. DZ twins share on average 50% of their segregating genes. The expectation for the DZ correlations between the genetic additive and dominance deviations are 1/2 and 1/4, respectively (Falconer & Mackay, 1996). Thus, the total DZ covariance equals $1/2\ a^2 + 1/4\ d^2$. Non-shared environmental influences (e^2) do not contribute to the covariance between individuals.

In this univariate model (Figure 1) path *b* is specified between the phenotypic scores of the twins. This path represents a reciprocal interaction between twins (Carey, 1986; Eaves, 1976; Neale & Stevenson, 1989; Simonoff et al., 1998). The literature on ADHD and associated phenotypes indicates that this interaction is negative (i.e., a contrast effect), and that it most likely represents a rater bias (Eaves et al., 2000; Hudziak, 2001; Kuntsi et al., 2000; Martin et al., 2002; Nadder et al., 1998; Price et al., submitted; Rietveld et al., 2003a; Thapar et al., 1995). As the parameter *b* is found to be negative, this suggests that parents stress the differences in behavior between their twins when rating their behavior.

If multiple assessments of the same twins are available, the covariance between measurements can also be modeled as a function of A, D and E (Boomsma & Molenaar, 1986; Martin & Eaves, 1977). When the same twins are measured repeatedly across time, one can evaluate how genes and environment contribute to the observed phenotypic stability during development. Stability is expressed as the correlation or covariance between the repeated measures. To accommodate the observed covariance between measures over time, the age-specific ADE univariate model was extended. Using a Cholesky or triangular decomposition model, the phenotypic covariance of OA and AP can be decomposed into additive genetic, genetic dominance, and unique environmental variance that is common across measurement occasions, and into additive genetic, genetic dominance, and unique environmental variance that is unique to each measurement occasion. The common variance is indicative of the extent to which additive genetic, genetic dominance, and unique environmental variance are shared across measurement occasions. The unique variance is indicative of the extent to which variance is age-specific, that is, not shared across ages. A contrast effect was specified at each age.

Model fitting procedures

Cross-sectional analyses of this and earlier datasets (Rietveld et al., 2003a) produced largely the same results. Longitudinal data analysis of measures collected at all four ages were done by a Cholesky decomposition. This model served as a reference model to judge the significance of genetic dominance and contrast effects.

Alternative models were evaluated on the basis of their plausibility and goodness-of-fit. Restrictions that are imposed in an alternative model are evaluated by minus twice the difference in log-likelihood between the restricted and the more general model. This difference is asymptotically chi-square distributed. If the χ^2 test is not significant, and the alternative model provides a plausible account of the data, the alternative model is considered tenable (Loehlin, 1992).

At all stages of model fitting the parameters were free to differ over gender. To evaluate whether boys and girls share a common set of genes, we tested whether opposite-sex twin correlations were of identical magnitude to same-sex dizygotic correlations. A discrepancy between these sets of correlations points at separate genetic sources of (co)variance for boys and girls, that is, genetic heterogeneity.

Results

Participation rates

Overall, the NTR has enjoyed excellent participation rates. The response rate for age 3 was 75.5% of all families (registered with the NTR) that were contacted when the twins turned 3 years old. This number excludes both 'true' non-responders and families who changed domicile without giving notice. We calculated the continued participation rate by considering complete overlapping birth cohorts for adjoining ages. For example, data from the 1986, 1987, and 1988 birth cohorts were used to calculate the number of maternal reports collected when the twins were age 10 and age 12. Continued participation rate was about 80%, for all age-intervals and in each zygosity group. Despite the large longitudinal response rate, we cannot exclude the possibility that the loss of participants at follow-up assessments was selective. Testing revealed that non-participation at age 7, age 10, and age 12 was positively related to the twins' overactive problem behavior at age 3 by maternal report. At age 7 and age 12, the effect was small but significant (age 7, $F = 11.79$, $p < .01$; age 12, $F = 5.83$, $p = .02$). At age 7, the mean OA scores at age 3 of participants and non-participants were 2.67 and 2.86, respectively. At age 12, the mean OA scores at age 3 of participants and non-participants were 2.59 and 2.81, respectively. This result suggests that the missing data are not missing completely at random. However, the hypothesis that the data are missing at random remains tenable. About half of the mothers, who did not return the questionnaire when the twins were 7, 10, or 12 years of age (i.e., the sample used for the calculation of the

selection bias), participated at least once at another measurement occasion. If a high OA score at age 3 had a direct influence on the mothers' decision to refuse participation at a follow-up age, one would expect the mother to leave the study entirely beyond the twins' age of 3 years. Given that this was not the case in 50% of the mothers suggests that the OA score at age 3 plays a modest role in explaining why mothers dropped out of the study. Although not conclusive, these results support the hypothesis that the data are missing at random.

Ratings of behavior and phenotypic stability

Table 2 contains the OA and AP scores by gender and age.

Boys had higher mean OA and AP scores than girls ($\chi^2_4 = 427.59$, $p < .0001$) at all ages. From a developmental perspective, the changes over age are most interesting. Since OA at age 3 and AP at older ages differ with respect to item content, we did not consider the change in mean scores in this interval. The different pattern from age 3 to age 7 between boys and girls is striking, however. The age interval from 7 to 10 years appears to be a period of minimal change in mean AP, indicated by a non-significant effect for girls and a small significant effect for boys ($\chi^2_1 = 4.61$, $p = .03$). In both girls and boys, the decrease in mean AP from age 10 to age 12 was significant (boys, $\chi^2_1 = 36.43$, $p < .001$; girls, $\chi^2_1 = 37.34$, $p < .001$).

Table 3 contains the inter-occasion correlations.

Two interesting results emerge from Table 3. First, both boys and girls display a moderate degree of stability from ages 3 to 7, and a much larger degree of stability from ages 7 to 10, and from ages 10 to 12. Second, the pattern in correlations is consistent across gender. The correlations between boys and girls do not differ significantly ($\chi^2_6 = 6.66$, $p > .05$). This implies that although boys have higher OA and AP scores, the observed stability across age is equal for boys and girls.

From a clinical perspective, it is of interest to focus on children with a high score on the AP scale. Several studies reported close convergence between high scores on the AP scale and ADHD by DSM criteria (Biederman et al., 1993; Chen, Faraone, Biederman, & Tsuang, 1994; Hudziak et al., in review; Kasius, Ferdinand, Van den Berg, & Verhulst, 1997; Steingard, Biederman, Doyle, & Sprich-Buckminster, 1992). To compare between CBCL AP and DSM ADHD, prevalence rates were calculated by imposing a cut-off score on the distribution of scores of the twins. The cut-off scores were derived from a Dutch community sample (age 3; Koot, 1993) and a norm sample (ages 7, 10, 12; Verhulst, Van der Ende, & Koot, 1996). This cut-off score corresponds to a *T*-score of 67. Children with a larger *T*-score are considered to be borderline or clinical cases (Achenbach, 1991). In taking the Dutch community and norm samples as a reference, we assume that the twin distribution resembles the distribution of scores of the singletons in those samples. The resulting prevalence rates are 3.1%, 6.3%, 7.0%, 5.2% for boys and 2.1%, 5.2%, 5.0%, 3.6% for girls, at ages 3, 7, 10, and 12 years, respectively. Prevalence rates for deviant AP are comparable to the norm rates (Verhulst et al., 1996), but prevalence rates for deviant OA reported here are modest compared to the community rates (Koot, 1993). Slightly more than 400 children participated in the community sample. This relatively small sample, and the inclusion of 10 clinical cases in the community sample, possibly explains part of the discrepancy in prevalence rates between the twin and community sample. In a follow-up study of the community sample (Mesman & Koot, 2001) it was found that the prevalence of DSM-IV diagnoses at ages 10 and 11 was somewhat higher compared to other community reports. Maternal age may further explain the lower number of borderline and clinical cases for OA in the twin sample. Koot (1993) found that age of the mother was a protective factor in the degree of problem behavior displayed by children (see also, Orlebeke, Knol, Boomsma, & Verhulst, 1998). Given that DZ twinning rate increases with age of the mother, the entire twin sample may have been rated by mothers who are on average older than the mothers who participated in the community study. In the present sample DZ twins have older mothers than MZ twins ($t = 7.59$, $df = 5379$, $p < .0001$), and male DZ twins have fewer OA problems than male MZ twins ($\chi^2_1 = 7.46$, $p < .01$). Linear regression revealed a small but significant contribution of maternal age to OA ($R^2 = .025$, $p < .05$). These results suggest that with increasing age, mothers report less problem behavior in their offspring. However, it remains uncertain

Table 2 Means and standard deviation (sd) for OA at age 3 and AP at ages 7, 10 and 12

	Boys			Girls		
	N	Mean	Sd	*N*	Mean	Sd
Age 3	5346	2.92	2.22	5520	2.48 *	2.15
Age 7	5000	3.37	2.95	5211	2.40 *	2.59
Age 10	2866	3.47 #	3.18	3037	2.38 *	2.58
Age 12	1433	3.09 #	2.94	1486	2.06 * #	2.41

Note: Significance was established at $p < .05$. N = number of individual twins. * = mean score of girls differs significantly from mean score of boys. # = mean score differs significantly from mean score at previous age within gender.

Table 3 Maximum likelihood within-person phenotypic correlations by gender. Correlations for boys and girls are reported below and above diagonal, respectively

Boys\Girls	Age 3	Age 7	Age 10	Age 12
Age 3	1.00	.41	.37	.38
Age 7	.41	1.00	.68	.65
Age 10	.38	.69	1.00	.72
Age 12	.35	.67	.75	1.00

whether maternal age explains the difference in prevalence rates *between* studies.

To establish the stability of OA and AP scores within the normal range, we calculated the percentage of boys and girls for whom a diagnosis of DSM ADHD most likely never applies (*T*-score < 67). Around 95% of the children who fall within the normal range of OA at age 3 do not experience AP at age 7. For the following age intervals, from 7 to 10 years and from 10 to 12 years, the percentage of children whose AP score continued to fall below the borderline cut-off score remained at 95%. No effect of gender was observed.

Heritability of OA and AP

Behavior genetic analyses of the data started with the calculation of twin correlations. With longitudinal data, correlations are calculated both within each age and across two ages (cross-correlations). The first give insight in the role of genes and environment at a specific age. The latter are informative with respect to the importance of genes and environment in explaining stability over time. Twin correlations and cross-correlations did not differ between boys and girls and between same-sex dizygotic pairs and opposite-sex dizygotic pairs. Therefore, the within and across age correlations were calculated for all MZ and all DZ twins.

Inspection of the difference between MZ and DZ twin within age correlations (on the diagonal in Table 4) suggests that variance in OA and AP is explained by additive genetic, genetic dominance, and unique environmental effects. MZ cross-correlations are also larger than DZ cross-correlations (see off-diagonals in Table 4). Inspection of the variances reveals that at age 3, DZ variances exceed MZ variances both in girls and boys. The discrepancy in variances is consistent with the presence of a contrast effect.

Table 5 contains an overview of the results of the longitudinal analyses. Model 1 serves as the baseline model. The importance of genetic dominance (Model 2) and the contrast effects (Model 3) at ages 7, 10 and 12 were evaluated against this model. The chi-square test based on difference in log-likelihood and associated degrees of freedom between Model 1 and Model 2 suggested that genetic dominance might be omitted from the model. However, we know from power studies (Eaves, 1972; Posthuma & Boomsma, 2000; Rietveld et al., 2003b) that the current twin sample does not provide much power to detect genetic dominance and it was therefore decided to retain genetic dominance in the model at this stage of model fitting. Estimates provided by Model 1 indicated a larger contrast effect at age 3 (–.08) compared to the contrast effect at older ages (ranging from .02 to .05). The non-significant difference between Model 1 and Model 3 suggested that the three contrast parameters as specified at ages 7, 10, and 12 are negligible. The chi-square based on the difference in log-likelihood of Models 3 and 4 suggests that the contrast parameter at age 3 is significant. We next fitted Model 5 to assess the importance of genetic dominance in explaining observed variation for AP at ages 7, 10, and 12. As indicated by the significant chi-square test, genetic dominance appeared a significant source of variance at these ages. We therefore accepted Model 3 as the best fitting model.

The decomposition of phenotypic variance based on Model 3 is given in the top part of Table 6. The diagonals contain the percentages for boys and girls, respectively. The off-diagonals are the relative contributions of A, D, and E to the observed covariance over time. Estimates for boys and girls are listed in the lower and upper triangular, respectively. These percentages indicate the extent to which the observed covariance of OA with AP from age to age is due to additive genetic, genetic dominance, and unique environmental influences. These results are corrected for the contrast effects. Generally, these results are similar to the results of other twin studies. Genetic and environmental correlations are listed in the lower part of Table 6.

As listed on the diagonals, around one-third to half of the observed variation in OA and AP at each age is explained by additive genetic influences. In

Table 4 Within age twin correlations (diagonal: MZ and DZ correlations) and across age cross-twin correlations (e.g., OA-3-years in twin 1 with AP-7-years in twin 2). MZ correlations below and DZ correlations above diagonal

MZ\DZ	Age 3	Age 7	Age 10	Age 12
Age 3	.66\.13	.13	.13	.12
Age 7	.34	.71\.28	.19	.18
Age 10	.33	.54	.72\.28	.19
Age 12	.32	.52	.55	.72\.26

Table 5 Longitudinal model fitting results by Cholesky decomposition

	Model	Age	–2 LL	*df*	Comparison	χ^2	*df*	*p*-value
1.	ADE + b	all ages	68064.85	29791				
2.	AE + b	all ages	68092.00	29811	1.	27.15	20	.13
3.	ADE + b	age 3	68069.45	29797	1.	4.60	6	.40
	ADE	age 7, 10, 12						
4.	ADE	all ages	68082.38	29799	3.	12.93	2	.00
5.	ADE + b	age 3	68107.31	29815	4.	37.86	18	.00
	AE	age 7, 10, 12						

Table 6 *Top part* includes percentages of total variances (diagonal) and covariances (off-diagonal) explained by additive genetic, genetic dominance, and unique environmental components based on best fitting models. Percentages for boys and girls are reported below and above diagonal, respectively. *Lower part* includes correlations calculated for additive genetic, genetic dominance, and unique environmental sources of variance between different ages. Correlations for boys and girls are reported below and above diagonal, respectively

Relative proportions of variance and covariance

	A %				D %				E %			
Boys\Girls	OA 3	AP 7	AP 10	AP 12	OA 3	AP 7	AP 10	AP 12	OA 3	AP 7	AP 10	AP 12
OA 3	50\41	73	79	75	22\33	17	13	14	28\26	10	8	11
AP 7	59	33\57	50	53	31	39\16	31	28	10	28\27	19	19
AP 10	86	31	41\48	47	6	51	31\25	32	8	18	28\27	21
AP 12	71	24	31	40\54	16	55	45	30\18	13	21	24	30\28

Correlations between different ages

	A				D				E			
Boys\Girls	OA 3	AP 7	AP 10	AP 12	OA 3	AP 7	AP 10	AP 12	OA 3	AP 7	AP 10	AP 12
OA 3	1.00	.60	.66	.57	1.00	.30	.16	.20	1.00	.15	.12	.14
AP 7	.57	1.00	.62	.57	.41	1.00	.99	1.00	.15	1.00	.46	.41
AP 10	.68	.56	1.00	.61	.08	.94	1.00	1.00	.11	.42	1.00	.50
AP 12	.49	.42	.53	1.00	.20	.98	.99	1.00	.14	.45	.58	1.00

girls, additive genetic influences are more important in explaining variation in AP than in boys. With the exception of AP measured in boys at age 7 (A 33%, D 39%), genetic dominance is found to contribute less to the observed variation at each age. Averaged over age and gender, broad heritability (the percentage of variance due to additive and dominance effects) is between 70% and 74%. The residual variance is explained by the unique (unshared) environment.

In both boys and girls, the covariance between OA and AP is explained by additive genetic influences to a much larger extent compared to the covariance among the three measurements of AP. As in the age-specific results, additive genetic effects are important in girls whereas non-additive effects are important in boys. Environmental influences that are unshared by family members contribute little to stability over time (ranging from 8% to 24%).

The correlations in the lower part of Table 6 indicate the degree of overlap between influences at one age and influences at subsequent ages. The additive genetic correlations are estimated between .42 and .68, and thus suggest only partial overlap in additive genetic effects. Correlations genetic dominance effects range from .94 to 1.00 over ages 7, 10, and 12. These correlations suggest a high degree of stability of dominance effects. However, the genetic dominance correlations between OA at age 3 and AP at later ages are much smaller (.08 to .41). A similar pattern of correlations is observed in the case of the unique environmental effects. The environmental correlations between OA and AP are estimated around .13. This is much lower than the environmental correlations (about .46) between AP measured at ages 7, 10, and 12 years.

Discussion

We analyzed maternal ratings of Overactivity (OA) and Attention Problems (AP) in twins at ages 3, 7, 10, and 12 years. The longitudinal allowed the study of the development of problem behavior over this period. We addressed the following questions. To what degree do genetic and environmental influences contribute to variation in OA and AP? Do results vary with the age or gender of the child? How do genetic and environmental influences contribute to the observed stability in OA and AP? Do results depend on the age-interval being studied? Is observed stability in boys and girls explained by an identical pattern of genetic and environmental influences? To what degree do boys and girls share these influences? Age-specific and longitudinal results are discussed at the phenotypic level first, followed by a discussion of genetic and environmental results.

Gender differences in mean scores and prevalences for OA and AP

Studies on OA and AP and other ADHD-related behaviors in girls are lacking (Gaub & Carlson, 1997; Klein & Mannuzza, 1991). In this paper we present data of large samples of twin families, with an equal number of girls and boys. The well-known gender difference with boys displaying more OA and AP was observed at each age. Even at the age of 3, boys display more OA problems than girls. Clinical studies have indicated that severe problem behavior can be identified in very young children (see for review, Campbell, 1995; Keenan & Wakschlag, 2000; Shaw, Owens, Giovannelli, & Winslow, 2001) and that the onset of ADHD is during the pre-school period (Barkley, Fisher, Edelbrock, & Smallish, 1990;

Kadesjö, Kadesjö, Hägglöf, & Gillberg, 2001; McGee, Williams, & Feehan, 1992). Children who suffer from attention deficits, impulsivity, or hyperactivity often display more problematic behavior when they enter the school system. This increase may reflect a true rise in problem behavior, a change in rating behavior of the mother, or a combination of both. Classroom discipline requires children to sit still and pay attention to the teacher. The inability of children suffering from OA and AP to do so may elicit negative reactions from teachers and peers, which in turn may affect the degree of problem behavior. The interaction between the school-age child and others may also lead to an increased awareness in the mother of the degree of AP that her child displays. In our study we observe high mean AP scores in boys at the age of 7 and 10 years. A decline of AP in boys is observed at age 12. Girls display a small but steady decline in maternally reported problem behavior from an earlier age. The difference between boys and girls is also seen at the higher end of the distribution. At every age, the ratio approaches 1.5 to 1, which implies that only a small proportion of boys outnumber girls on DSM criteria in the present study. This gender ratio is not exceptional in non-clinical samples (Biederman, 1998; Pineda et al., 1999; Szatmari, Offord, & Boyle, 1989). In a recent population study of hyperactivity in 1200 Dutch children aged 2 to 11 years (Centraal Bureau voor de Statistiek, 2002b), a similar gender ratio was observed.

Developmental changes from OA to AP

The observed age-specific gender differences in means notwithstanding, the degree of stability in boys and girls is about equal. We observed moderate stability of OA at age 3 to AP at age 7 (correlation .41), and greater stability of AP from age 7 to age 10, and from age 10 to age 12 (correlations about .70). Possible explanations for the moderate degree of stability include the amount of time between assessments, developmental changes, and the change in phenotype (OA at age 3, AP thereafter). The interval between ages 3 and 7 is not smaller than the interval between ages 7 and 12. Given that the observed correlation between ages 7 and 12 is comparable to those between ages 7 and 10, and ages 10 and 12, time interval does not seem to be important in explaining the difference in stability. Developmental change may be a factor, as findings from neurological and neuropsychological studies suggest that the age period from the preschool and early school years is marked by the acquisition of attention skills (Bennett Murphy, Murphy, & Rose, 2001; Woody-Ramsey & Miller, 1988). Thus, 3-year-old children may suffer from OA problems relative to their developmental maturity, but by age 7 years, the majority of children should have progressed through this developmental period and thus have relatively stable attention skills. Any deficits may be more likely to persist. Another reason for the moderate stability may be the differences in assessment instrument. At age 3, the OA scale of the CBCL/2-3 is composed of 5 items. At older ages, the AP scale of the CBCL/4-18 is composed of 11 items. As can be seen in Appendix I, OA and AP share two items (item 5 (OA) with item 8 (AP); item 6 (OA) with item 10 (AP)). If the moderate degree of stability from OA to AP is due to developmental change, the correlation of the overlapping items between ages 3 and 7 should resemble those for the total scores on OA to AP. Specifically, because age is hypothesized to be very relevant and item content is identical, item stability from ages 3 to 7 should approach the observed moderate OA–AP stability from age 3 and beyond (around .40) and should not approach the observed large AP–AP stability from age 7 and beyond (around .70). The observed correlations ranged from .27 to .44, i.e., they resemble the correlation between OA and AP. We therefore suggest that the modest association between OA at age 3 and AP at age 7 represents developmental change. Hart, Lahey, Loeber, Applegate, and Frick (1995) also reported that the decline in hyperactive behavior in a clinical sample was related to the increasing age of the subjects. Artifacts of repeated assessment by the same informant and passage of time between assessments were found to be less plausible in explaining the decrease in hyperactivity.

Our results are comparable with those reported in other studies that used the CBCL. Achenbach and Rescorla (2000) reported a correlation of .40 between AP at age 3 (CBCL/1.5-5; revision of the CBCL/2-3) and AP at age 7. Three out of five items of the Dutch OA scale are included in the AP scale derived in analyses of the CBCL/1.5-5. Koot (1993) reported a correlation of .47 between OA at ages 2–3 years and AP at ages 4–5 years. The high degree of AP stability found here is in line with those reported for repeatedly assessed children varying in age from 7 to 10 years (Verhulst et al., 1995). In a sample of adolescent adoptees, Van der Valk et al. (1998) observed a correlation of around .65 between two assessments that covered a three-year interval.

Age-specific heritability estimates

Individual differences in childhood OA and AP are mainly due to individual differences in genetic factors. The contribution of unique (or unshared) environmental effects to individual differences in OA and AP is considerably smaller. The genetic component also includes any effects of an interaction between shared family environment and genotype (Boomsma & Martin, 2002; Plomin, DeFries, & Loehlin, 1977). Imagine two children with a different score on the AP scale. The children may behave in a similar way when the environment contains only minimal sources of distraction. When the children enter a situation with

many stimuli, they may differ in their response to this new situation. In the classical twin design, it is not possible to distinguish between variance due to the interaction of shared environment and genotype and variance due to genetic effects only. An interaction between genetic effects and unique environmental effects contributes to the unique environmental variance estimate. Because the unique environment was estimated at around a modest 25%, this type of interaction, if present, is unlikely to be large.

Overall, our results are in line with the results of other twin studies of AP (Edelbrock et al., 1995; Gjone et al., 1996; Hudziak et al., 2000; Schmitz et al., 1996; Zahn-Waxler et al., 1996) and phenotypes related to ADHD (Eaves et al., 1997; Kuntsi et al., 2000; Martin et al., 2002; Nadder et al., 1998; Price et al., submitted; Rhee et al., 1999; Sherman et al., 1997; Thapar et al., 1995; Thapar et al., 2000). In many of these studies gender and age effects were not addressed. Here, we found that heritability estimates are equal for boys and girls, although boys display significantly more variation in their behavior at ages 10 and 12. Further, genetic heterogeneity was found to be absent, which suggests that the same set of genes contribute to individual differences in boys and girls.

At age 3, a significant contrast effect was detected. Given the results of other twin studies that included multiple raters (e.g., Simonoff et al., 1998), it is likely that the contrast effect is due to a bias in the ratings of the mother. This bias is due to the mothers comparing the twins with one another and overestimating any perceived differences. As children get older and mothers are exposed to the behavior of children other than their own twins, they rely less on the comparison of the twins in judging the children's behavior. In other words, during children's development, the mothers' frame of reference may change from children inside the home to children outside the home. This would account for the absence of a contrast effect at ages 7, 10, and 12.

Genetic and environmental stability of OA and AP

Additive genetic effects were found to make the greatest contribution to the stability between OA age 3 and AP beyond age 3. Genetic dominance and unique environmental effects contributed relatively little. The differences in these contributions to stability continued over the intervals including ages 7, 10, and 12. After age 3, additive genetic effects were found to be more important in girls and genetic dominance more important in boys. The genetic correlations between the ages suggest that additive genetic effects are far from perfectly stable. Only a subset of genes that operates at one age does so at a later age. This result is consistent over the age-intervals and over gender. In contrast, genetic dominance and the unique environment effects varied with age-interval. Very low correlations are observed for the intervals that include age 3. This indicates that the observed correlation between OA (age 3) and AP (beyond age 3) is nearly entirely due to additive genes. A more complex picture arises for the intervals including ages 7, 10, and 12. When all genetic and environmental correlations are taken into account, it emerges that the observed stability of AP is stability of additive genetic, dominance, and environmental effects common to these 3 ages. These results agree with results reported by Schmitz et al. (submitted).

Limitations of the study

We found that the unique environmental effects explained a quarter of the observed variance in OA and AP at each age. As we did not include an actual measure of the environment, we cannot identify any specific aspect of unique environment that contributes to individual differences in OA and AP, except to say that this 25% also includes measurement error. The nature of these effects is thus partly unknown.

The interval between ages 3 and 7 is large, and possible developmental changes may not have been captured by this study. This seems unlikely for boys, given the sharp increase in problem behavior during this age period. However, girls may first have increased in problem behavior shortly after age 3 and then decreased by the time they were 7 years old.

Neither OA nor AP is a measure of ADHD. Here we used AP as a marker for attention problem behavior, which is an important aspect of ADHD. AP has been shown to be predictive of DSM-III-R ADHD (Chen et al., 1994) and DSM-IV ADHD (Hudziak et al., in review). At present, we are collecting DSM-IV data. These data will enable us to explore the exact relation between OA, AP, and ADHD in the present sample.

Conclusions

Our study supports the findings reported earlier in smaller, cross-sectional studies, that genetic influences are important in the development of AP. The longitudinal data revealed that genetic influences are also important for the stability of these problems. Furthermore, we found that children who do not have AP at one stage of development are unlikely to develop AP at a later stage.

Acknowledgements

The work of the first author is supported by grant 96/22 from the Universitair Stimulerings Fonds to D.I. Boomsma. The work by the second author is supported by grant MH52813 from NIMH to J.J. Hudziak. We thank Professor Hans Koot for kindly providing his CBCL dataset on 3-year-old singletons (referred to in text as the community sample).

Correspondence to

M.J.H. Rietveld, Vrije Universiteit, Department of Biological Psychology, Van der Boechhorststraat 1, 1081 BT Amsterdam, The Netherlands; Email: mjh.rietveld@psy.vu.nl

References

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

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

Achenbach, T.M., & Rescorla, L.A. (2000). *Manual for the ASEBA Preschool Forms and Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth and Families.

Barkley, R.A., Fisher, M., Edelbrock, C.S., & Smallish, L. (1990). The adolescent outcome of hyperactive children diagnosed by research criteria: I. An 8-year prospective follow-up study. *Journal of the American Academy of Child and Adolescence Psychiatry*, *29*, 546–557.

Bennet Murphy, L., Murphy, C.E., & Rose, C.L. (2001). Sustained attention and unintentional injury among preschool-aged children. *Child Neuropsychology*, *7*, 72–83.

Biederman, J. (1998). Attention-Deficit/Hyperactivity Disorder: A life-span perspective. *Journal of Clinical Psychiatry*, *59*, 4–16.

Biederman, J., Faraone, S.V., Doyle, A., Lehman, B.K., Kraus, I., Perrin, J., & Tsuang, M.T. (1993). Convergence of the Child Behavior Checklist with structured interview-based psychiatric diagnoses of ADHD children with and without comorbidity. *Journal of Child Psychology and Psychiatry*, *34*, 1241–1251.

Biederman, J., Mick, E., & Faraone, S.V. (2000). Age-dependent decline of symptoms of Attention Deficit Hyperactivity Disorder: Impact of remission definition and symptom type. *American Journal of Psychiatry*, *157*, 816–818.

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

Boomsma, D.I., & Martin, N.G. (2002). Gene–environment interactions. In H. D'Haenen, J. A. den Boer, & P. Willner (Eds.), *Biological psychiatry* (pp. 181–187). Chichester: John Wiley & Sons, Ltd.

Boomsma, D.I., & Molenaar, P.C.M. (1986). Using LISREL to analyze genetic and environmental covariance structure. *Behavior Genetics*, *16*, 237–250.

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

Carey, G. (1986). Sibling imitation and contrast effects. *Behavior Genetics*, *16*, 319–341.

Centraal Bureau voor de Statistiek. (2002a). *Enquete Beroepsbevolking 1999. Standaard Beroepsclassificatie 1992*. Heerlen: Centraal Bureau voor de Statistiek. [information obtained from the internet address http://www.cbs.nl/].

Centraal Bureau voor de Statistiek. (2002b). *Permanent Onderzoek Leefsituatie, module Gezondheid en Arbeid, specifieke gezondheidsmetingen bij kinderen 2001*. Heerlen: Centraal Bureau voor de Statistiek. [information obtained from the internet address http://www.cbs.nl/].

Chen, W.J., Faraone, S.V., Biederman, J., & Tsuang, M.T. (1994). Diagnostic accuracy of the Child Behavior Checklist scales for attention-deficit hyperactivity disorder: A receiver-operating characteristic analysis. *Journal of Consulting and Clinical Psychology*, *62*, 1017–1025.

De Groot, A., Koot, H.M., & Verhulst, F.C. (1994). Cross-cultural generalizability of the Child Behavior Checklist cross-informant syndromes. *Psychological Assessesment*, *6*, 225–230.

Eaves, L.J. (1972). Computer simulation of sample size and experimental design in human psychogenetics. *Psychological Bulletin*, *77*, 144–152.

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

Eaves, L.J., Rutter, M., Silberg, J.L., Shillady, L., Maes, H.H., & Pickles, A. (2000). Genetic and environmental causes of covariation in interview assessments of disruptive behavior in child and adolescent twins. *Behavior Genetics*, *30*, 321–334.

Eaves, L.J., Silberg, J.L., Maes, H.H., Simonoff, E., Pickles, A., Rutter, M., Neale, M.C., Reynolds, C.A., Erikson, M.T., Heath, A.C., Loeber, R., Truett, K.R., & Hewitt, J.K. (1997). Genetics and developmental psychopathology: 2. The main effects of genes and environment on behavioral problems in the Virginia Twin Study of Adolescent Behavioral Development. *Journal of Child Psychology and Psychiatry*, *38*, 965–980.

Edelbrock, C., Rende, R., Plomin, R., & Thompson, L.A. (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., & Mackay, T.F.C. (1996). *Introduction to quantitative genetics* (4th edn.). Harlow: Longman Group Ltd.

Gaub, M., & Carlson, C.L. (1997). Gender differences in ADHD:A meta-analysis and critical review. *Journal of American Academy of Child and Adolescent Psychiatry*, *36*, 1036–1045.

Gjone, H., Stevenson, J., & Sundet, J.M. (1996). Genetic influence on parent-reported attention-related problems in a Norwegian general population twin sample. *Journal of American Academy of Child and Adolescent Psychiatry*, *35*, 588–596.

Hart, E.L., Lahey, B.B., Loeber, R., Applegate, B., & Frick, P.J. (1995). Developmental change in Attention-Deficit Hyperactivity Disorder in boys: A four-year longitudinal study. *Journal of Abnormal Child Psychology*, *23*, 729–747.

Hudziak, J.J. (2001). The role of phenotypes (diagnoses) in genetic studies of attention-deficit/hyperactivity disorder and related child psychopathology. *Child and Adolescent Psychiatric Clinics of North America*, *10*, 279–297.

Hudziak, J.J., Heath, A.C., Madden, P.A.F., Reich, W., Bucholz, K.K., Slutske, W.S., Bierut, L.J., Neuman, R.J., & Todd, R.D. (in review). The genetic analysis of parental reports of DSM-IV. *Journal of American Academy of Child and Adolescent Psychiatry*.

Hudziak, J.J., Rudiger, L.P., Neale, M.C., Heath, A.C., & Todd, R.D. (2000). A twin study of inattentive, aggressive and anxious/depressed behaviors. *Journal of American Academy of Child and Adolescent Psychiatry*, *39*, 469–476.

Kadesjö, C., Kadesjö, B., Hägglöf, B., & Gillberg, C. (2001). ADHD in Swedish 3- to 7-year-old children. *Journal of American Academy of Child and Adolescent Psychiatry*, *40*, 1021–1028.

Kasius, M.C., Ferdinand, R.F., Van den Berg, H., & Verhulst, F.C. (1997). Associations between different diagnostic approaches for child and adolescent psychopathology. *Journal of Child Psychology and Psychiatry*, *38*, 625–632.

Keenan, K., & Wakschlag, L.S. (2000). More than the terrible twos: The nature and severity of behavior problems in clinic-referred preschool children. *Journal of Abnormal Child Psychology*, *28*, 33–46.

Klein, R.G., & Mannuzza, S. (1991). Long-term outcome of hyperactive children: A review. *Journal of American Academy of Child and Adolescent Psychiatry*, *30*, 383–387.

Koot, H.M. (1993). *Problem Behavior in Dutch Preschoolers*. 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.

Kuntsi, J., Gayán, J., & Stevenson, J. (2000). Parents' and teachers' ratings of problem behaviours in children: Genetic and contrast effects. *Twin Research*, *3*, 251–258.

Little, R.J.A., & Rubin, D.B. (1987). *Statistical analysis with missing data*. New York: Wiley.

Loehlin, J.C. (1992). *Latent variable models: An introduction to factor, path, and structural analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Martin, N., Scourfield, J., & McGuffin, P. (2002). Observer effects and heritability of childhood attention-deficit hyperactivity disorder symptoms. *British Journal of Psychiatry*, *180*, 260–265.

McGee, R., Williams, S., & Feehan, M. (1992). Attention Deficit Disorder and age of onset of problem behaviors. *Journal of Abnormal Child Psychology*, *20*, 487–502.

Mesman, J., & Koot, H.M. (2001). Early preschool predictors of preadolescent internalizing and externalizing DSM-IV diagnoses. *Journal of American Academy of Child and Adolescent Psychiatry*, *40*, 1029–1036.

Nadder, T.S., Silberg, J.L., Eaves, L.J., Maes, H.H., & Meyer, J.M. (1998). Genetic effects on ADHD symptomatology in 7- to 13-year-old twins: Results from a telephone survey. *Behavior Genetics*, *28*, 83–100.

Nadder, T.S., Silberg, J.L., Rutter, M., Maes, H.H., & Eaves, L.J. (2001). Comparison of multiple measures of ADHD symptomatology: A multivariate genetic analysis. *Journal Child Psychology and Psychiatry*, *42*, 475–486.

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

Neale, M.C., Cardon, L.R., & The North Atlantic Treaty Organization. Scientific Affairs Division. (1992). *Methodology for genetic studies of twins and families*. Dordrecht: Kluwer Academic Publishers.

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.

Orlebeke, J.F., Knol, D.L., Boomsma, D.I., & Verhulst, F.C. (1998). Frequency of parental report of problem behavior in children decreases with increasing maternal age of delivery. *Psychological Reports*, *82*, 395–404.

Pineda, D., Ardila, A., Rosselli, M., Arias, B.E., Henao, G.C., Gomez, L.F., Mejia, S.E., & Miranda, M.L. (1999). Prevalence of Attention-Deficit/Hyperactivity Disorder symptoms in 4- to 17-year-old children in the general population. *Journal of Abnormal Child Psychology*, *27*, 455–462.

Plomin, R., DeFries, J.C., & Loehlin, J.C. (1977). Genotype–environment interaction and correlation in the analysis of human behavior. *Psychological Bulletin*, *84*, 309–322.

Plomin, R., DeFries, J.C., McClearn, G.E., & McGuffin, P. (2001). *Behavior genetics* (4th edn). New York: W.H. Freeman.

Posthuma, D., & Boomsma, D.I. (2000). A note on the statistical power in extended twin designs. *Behavior Genetics*, *30*, 147–158.

Price, T.S., Simonoff, E., Waldman, I., Asherson, P., Curran, S., & Plomin, R. (submitted). What is stable about hyperactive behaviors in pre-school children is genetic: Implications for molecular genetic studies.

Rhee, S.H., Waldman, I.D., Hay, D.A., & Levy, F. (1999). Sex differences in genetic and environmental influences on DSM-III-R Attention-Deficit/Hyperactivity Disorder. *Journal of Abnormal Psychology*, *108*, 24–41.

Rietveld, M.J.H., Hudziak, J.J., Bartels, M., Van Beijsterveldt, C.E.M., & Boomsma, D.I. (2003a). Heritability of attention problems in children: I. Cross-sectional results from a study of twins, age 3 to age 12. *American Journal of Medical Genetics B*, *117b*, 102–113.

Rietveld, M.J.H., Posthuma, D., Dolan, C.V., & Boomsma, D.I. (2003b). ADHD: Sibling interaction or dominance: An evaluation of statistical power. *Behavior Genetics*, *33*, 247–255.

Rietveld, M.J.H., Van der Valk, J.C., Bongers, I.L., Stroet, T.M., Slagboom, P.E., & Boomsma, D.I. (2000). Zygosity diagnosis in young twins by parental report. *Twin Research*, *3*, 134–141.

Ross, D.M., & Ross, S.A. (1982). *Hyperactivity. Current issues, research, and theory*. New York: John Wiley & Sons, Inc.

Rutter, M., Giller, H., & Hagell, A. (1998). *Antisocial behavior by young people*. New York/Cambridge: Cambridge University Press.

Saudino, K.J., Cherny, S.S., & Plomin, R. (2000). Parent ratings of temperament in twins: Explaining the 'too low' DZ correlations. *Twin Research*, *3*, 224–233.

Schmitz, S., Corley, R.P., Hewitt, J.K., Zahn-Waxler, C., Emde, R.N., & DeFries, J.C. (submitted). Attention problems and emotionality at ages 7 and 12.

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.

Schmitz, S., Saudino, K.J., Plomin, R., Fulker, D.W., & DeFries, J.C. (1996). Genetic and environmental influences on temperament in middle childhood: Analyses of teacher and tester ratings. *Child Development*, *67*, 409–422.

Shaw, D.S., Owens, E.B., Giovannelli, J., & Winslow, W.B. (2001). Infant and toddler pathways leading to early externalizing disorders. *Journal of American Academy of Child and Adolescent Psychiatry*, *40*, 36–43.

Sherman, D.K., Iacono, W.G., & McGue, M.K. (1997). Attention-deficit hyperactivity disorder dimensions: A twin study of inattention and impulsivity-hyperactivity. *Journal of American Academy of Child and Adolescent Psychiatry*, *36*, 745–753.

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.

Steingard, R., Biederman, J., Doyle, A., & Sprich-Buckminster, S. (1992). Psychiatric comorbidity in attention deficit disorder: Impact on the interpretation of Child Behavior Checklist results. *Journal of American Academy of Child and Adolescent Psychiatry*, *31*, 449–454.

Szatmari, P., Offor, D.R., & Boyle, M.H. (1989). Ontario Child Health Study: Prevalence of Attention Deficit Disorder with Hyperactivity. *Journal of Child Psychology and Psychiatry*, *30*, 219–230.

Thapar, A., Harrington, R., Ross, K., & McGuffin, P. (2000). Does the definition of ADHD affect heritability? *Journal of American Academy of Child and Adolescent Psychiatry*, *39*, 1528–1536.

Thapar, A., Hervas, A., & McGuffin, P. (1995). Childhood hyperactivity scores are highly heritable and show sibling competition effects: Twin study evidence. *Behavior Genetics*, *25*, 537–544.

Van den Oord, E.J.C.G. (1993). *A genetic study of problem behaviors in children*. Doctoral dissertation, Erasmus University, Rotterdam.

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

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

Verhulst, F.C., & Van der Ende, J. (1995). The eight-year stability of problem behavior in an epidemiologic sample. *Pediatric Research*, *38*, 612–617.

Woody-Ramsey, J., & Miller, P.H. (1988). The facilitation of selective attention in preschoolers. *Child Development*, *59*, 1497–1503.

Wothke, W. (2000). Longitudinal and multigroup modeling with missing data. In T.D. Little, K.U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data; Practical issues, applied approaches and specific examples*. Mahwah, NJ: Lawrence Erlbaum Associates.

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 internalizing of control. *Developmental Psychopathology*, *8*, 103–122.

Manuscript accepted 30 May 2003

Appendix I

Description of translated items for problem scales 'overactive behavior' and 'attention problems'

CBCL/2–3 for 3-year-olds Overactive behavior [OA]		CBCL/4–18 for 7-, 10-, and 12-year-olds Attention problems [AP]	
Item no.	Item description	Item no.	Item description
5.	Can't concentrate, can't pay attention for long	1.	Acts too young for his/her age
6.	Can't sit still, restless	8.	Can't concentrate, can't pay attention for long
11.	Constantly seeks help	10.	Can't sit still, restless, or overactive
59.	Quickly shifts from one activity to another	13.	Confused or seems to be in a fog
62.	Refuses to play active games	17.	Daydreams or gets lost in his/her thoughts
		41.	Impulsive or acts without thinking
		45.	Nervous or tense
		46	Nervous movements or twitching
		61.	Poor school work
		62.	Clumsy or poorly coordinated
		80.	Stares blankly

Chapter 8

Summary, discussion, and synthesis

Summary: The structure and stability of Intelligence

In this study, six measures of cognitive abilities assessing two major domains of intelligence were administered to 209 twin pairs at the ages of 5, 7, and 10 years (RAKIT; Bleichrodt et al., 1984). The factor structure of cognitive abilities was studied both by exploratory and confirmatory factor analyses. The longitudinal design of the study permitted the investigation of configural invariance over time, and stability of individual differences in phenotypic scores on the six subtests. The availability of twin data allowed for the exploration of the genetic and environmental contributions to the phenotypic structure of cognitive abilities.

A common factor model with two correlated factors was found to fit the phenotypic covariance matrix of the RAKIT. This factor model was found to fit well and to be amenable a sensible interpretation at ages 5, 7, and 10. The correlation between the verbal and nonverbal group factors was estimated between .40 and .49 at these ages. Age-specific genetic analyses revealed that the genetic covariance matrix, like the phenotypic covariance matrix, was consistent with a two common factor model. A single common factor, without residuals, fitted the shared environmental covariance matrix. The unique environment operated in a subtest-specific manner, that is, these effects contributed only to the residual variances of the test scores. The genetic and environmental factor structures were found to be identical across age.

The factor structure as described was retained in the analyses of the longitudinal data. The genetic nonverbal and verbal factors correlated moderately at each age (.25 at age 5, .28 at age 7, .30 at age 10). The genetic and environmental contributions to the total variance of each subtest at each age are reported in Table 8.1. The genetic contributions to the total variance of each subtest either did not change (Discs, Hidden Figures, Idea Production) or increased (Exclusion, Verbal Meaning, Learning Names) in importance from age 5 to age 10. With the exception of Verbal Meaning, the shared environment contributed little to the phenotypic subtest variance at each age. In contrast, the unique environment explained a relatively large proportion of the total subtest variance. Estimates varied between 22% to 64% depending on subtest and age of measurement. These components of variance include measurement error.

The phenotypic stability of performance on the intelligence subtests was estimated between .35 and .55 from ages 5 to 7, and 7 to 10. These correlations did not vary in any perceivable consistent manner between verbal and nonverbal subtests, or between age intervals. The longitudinal analyses revealed that the stability of subtest performance was found to be mainly due to stable genetic influences. Shared environmental influences con-

tributed to this phenotypic stability, but the contribution was relatively small, and waned as children grew older. Support for the differentiation hypothesis, which states that the inter-correlation among dimensions of cognitive abilities decrease with increasing age, was absent at the genetic and environmental level.

Table 8.1: Percentages of total variance explained by additive genetic (a^2), shared environmental (c^2), and unique environmental factors (e^2).

Subtest	age	a^2	c^2	e^2
	5	38	19	43
Exclusion	7	40	5	55
	10	58	7	35
	5	46	3	51
Discs	7	58	5	37
	10	44	4	52
	5	48	11	41
Hidden Figures	7	37	14	49
	10	48	8	44
	5	21	39	40
Verbal Meaning	7	19	17	64
	10	42	32	26
	5	53	17	30
Learning Names	7	39	32	29
	10	74	4	22
	5	58	5	37
Idea Production	7	58	4	38
	10	60	4	36

Summary: The development of Overactivity and Attention Problems

When twins were aged 3, 7, 10, and 12 years, their parents were asked to complete the CBCL on both twins. At age 3, twins were rated on the Overactivity scale, a syndrome scale of the CBCL/2-3 (Achenbach, 1992) identified by Koot and colleagues (1997). At ages 7, 10, and 12, twins were rated on the Attention Problem scale of the CBCL/4-18 (Achenbach, 1991). Maternal data from complete birth cohorts 1986 through 1993 of 3- and 7-year-old twins were analyzed. Of these cohorts, some twins had yet to reach the age of 10 and 12 years when the present study of Overactivity and Attention Problems began. Resulting data included 11938 reports of 3-year-old twin, 10657 reports of 7-year-old twins, 6192 reports of 10-year-old twins, and 3124 reports of 12-year-old twins.

Boys displayed much more Overactive problem behavior at age 3 and Attention Problem behavior at ages 7, 10, and 12 than girls. Between 2% and 7% of the children displayed extreme Overactive and Attention Problem behavior with a higher prevalence rate in boys at each age.

Genetic analyses revealed a large estimate of heritability of Overactive and Attention Problem behavior. The relative genetic and unique environmental contributions are listed in Table 8.2. A very consistent pattern emerged from these results. Broad heritability (i.e., the sum of additive genetic and genetic dominance effects) is estimated between 70% and 74%. The residual variance in Overactivity and Attention Problems is explained by unique environmental influences, i.e., environmental influences not shared by individuals of the same household.

Table 8.2: Percentages of total variance explained by additive genetic (a^2), genetic dominance (d^2), and unique environmental effects (e^2). In each cell two estimates are given, one for boys and one for girls.

Problem scale	age	a^2	d^2	e^2
Overactivity *	3	50% - 41%	22% - 33%	28% - 26%
Attention Problems	7	33% - 57%	39% - 16%	28% - 27%
Attention Problems	10	41% - 48%	31% - 25%	28% - 27%
Attention Problems	12	40% - 54%	30% - 18%	30% - 28%

Note: * the presence of a contrast effect was suggested at age 3. The prevailing interpretation of the contrast effect is that parents stress the difference in overactive behavior between their twins. In genetic analyses, the contrast effect is difficult to discern from the genetic dominance effect.

The transition from Overactive behavior at age 3 into Attention Problems at age 7 was characterized by a moderate degree of stability (around .40). The stability of Attention Problems from age 7 to age 12 was found to be much larger (around .70). The degree of stability of these behavior problems was found equal in both sexes. Across the entire age range from age 3 to age 12, genetic effects accounted for the observed stability in Overactivity / Attention Problems (range from 76% to 92%). Genetic and environmental contributions to the phenotypic covariance did not differ over the age-intervals. No consistent difference in these contributions emerged between boys and girls. Further, it appeared that boys and girls share the same set of genes and environmental effects.

Discussion: The structure and stability of Intelligence

The following questions concerning the genetic and environmental factors were addressed. Is the organization of the genetic and environmental contributions invariant over development? Do the genetic and environmental contributions exhibit a factor structure identical to that of the phenotypic data? By considering the complete longitudinal data, what can be said concerning the stability of genetic and environmental influences? To what extent is the phenotypic stability due to genetic and environmental factors? These questions are answered in the light of the present results. In addition, the hypothesis of differentiation of cognitive abilities is considered given the present results.

Phenotypic results for psychometric intelligence

Participation rates remained high (>92%), and those who dropped out did not differ in general IQ from participants at the beginning of the study. At each age, subtest scores in the twin sample were comparable to those observed in the norm sample. The distribution of scores was approximately normal. At each age, boys and girls did not differ in mean subtest scores. This result is in line with the literature on childhood intelligence (Helgeson, 2002; Sadker et al., 1991). Stability in subtests performance across intervals was estimated, on average, at r = .42, from ages 5 to 7; at r = .40, from ages 5 to 10; and at r = .50, from ages 7 to 10. The stability estimates for the intervals that include age 5 indicate that intelligence performance at this age is equally predictive of subtest performance at age 7 and age 10 years.

Configural invariance

Configural invariance, as defined in the common factor model, implies that the same observed variables load on the same factors across measurement occasions. To establish that the same theoretical trait is being measured at different ages, configural invariance is a minimal condition. At age 5, the observed covariation among the six subtests was consistent with a correlated two-factor structure, in which a set of three subtests loaded on each factor. In view of the factor loadings, the factors were interpreted as verbal and nonverbal cognitive abilities. This factor structure was found to be stable across age, thus supporting the hypothesis of configural invariance. Next, configural invariance at the genetic and environmental level was studied. The best fitting model at age 5 was derived by exploratory factor analyses. In this model, the genetic and environmental covariance structures differed. The analysis of the genetic covariance matrix suggested the presence of two correlated genetic group factors with subtest-specific factors. The familial environmental covariance structure was consistent with a single common factor model without subtest-

specifics. The unique environmental covariance matrix was described adequately by a model including only subtest-specific factors. In subsequent analyses, a more formal hypothesis testing approach was adopted by imposing the model obtained at age 5 on data collected at ages 7 and 10. Support for the invariance of factor structure over time was provided by the adequate fit of the hypothesized genetic and environmental structures at these ages. The comparison of the phenotypic results with the genetic and environmental results indicated that the phenotypic two-factor solution was due mainly to the genetic structure.

The shared environment

The shared environmental influences correlated strongly over the three ages. The contributions of the shared environment to the (co)variances of the six subtests were small or decreased with increasing age. At age 10, the relative contribution of the shared environment to the phenotypic variances was estimated at less than 10% in five of the six subtests. Thus, in view of the large correlations among the three general factors, shared environmental effects remained largely identical but had a diminishing impact in explaining individual differences in childhood cognitive abilities.

This result is counterintuitive in that parents, upbringing, socioeconomic status, teachers and other environmental aspects shared by children from the same home are expected to contribute to similarity in intellectual development in an accumulative manner. Empirical evidence suggests the opposite. Although the shared environment may be important early in life, the effect on individual differences decreases when children pass the school age period and enter adolescence and adulthood. What can account for the modest importance of the home environment during this age-period? A possibility is that children become more focused on people outside the home as they grow older. During the early school years, children increasingly engage in activities that involve other children. School-aged children come under the influence of the behavior of classmates and friends and instructions from teachers and in activities of their own choosing. Peers and teachers thus may progressively replace parents as influencing agents. To assess the relevance of classmates and friends, information on the children's social relationships is needed. A study by Kindermann (1993) addressed this issue. He examined longitudinal changes in the association between children's school motivation and peer groups. School motivation was conceptualized as engagement in learning activities. It was observed that when children changed from one peer group to another, their school motivation changed as well. Children's academic performance was not measured but one may expect an effect on school performance when the engagement in learning activities changes. Although it remains unclear whether a change in school performance is a cause or effect of a change in social

relationships, Kindermann's study provides some evidence of a link between the social peer context and a child's intellectual performance.

Heritability

The longitudinal genetic covariance structure was modeled using a first order autoregressive, or simplex model. It was assumed that genetic variance at any age was partly a function of genetic variance at the previous age and was partly determined by other genetic factors. The former represents transmission effects and the latter represent innovation effects. Transmission effects refer to variance that is shared by adjacent ages and contributes to stability. In contrast, innovation effects give rise to new variance at a given point in time, and are associated with change. To accommodate shared variance across the verbal and nonverbal domains, correlations were specified between the nonverbal and verbal factor at age 5 and between the innovation factors within age 7 and age 10. In addition to the simplex model, subtest-specific factors were included to account for variance that is not common to the subtests. Correlations over time between these specifics were introduced to account for genetic covariance between the same subtest measured at different ages.

The results indicated that genetic effects reflect both stability and transitory. The verbal and nonverbal genetic factors accounted for most of the stability of subtest performance. The observed covariance between nonverbal and verbal subtests was only partly accounted for by these genetic group factors, as indicated by the correlations between the Nonverbal and Verbal factor (.25 at age 5; .28 at age 7; .30 at age 10). Across-age correlations among the subtest-specific genetic factors varied greatly, making it difficult to discern a general pattern. For some subtests, all genetic effects were shared over time whereas other subtests demonstrated substantial age-specific effects. The complex structure of the genetic contribution was also observed in the study of Reznick and colleagues (1997). Another published twin study reported on the covariation between verbal and nonverbal cognitive development in a sample of two-year-olds (Price et al., 2000). The phenotypic correlation between these two measurements was estimated at .42 in a sample of almost 2000 same-sex twin pairs. Further, it was found that genetic factors contributed less than 25% to the total variance observed in verbal and nonverbal measurements. The genetic effects contributed very little to the phenotypic correlation between the verbal and nonverbal dimensions. Adding our findings to these results it follows that between the age of 2 and the age of 5, and up to the age of 10, genes become increasingly important in explaining individual differences in verbal and nonverbal dimensions of intelligence. An accompanying result is the change in the cause of phenotypic similarity over time. Both studies with infants identified the shared environment as the source of stability in infancy

whereas the present study showed that stability in subtest performance from age 5 onward was largely due to genetic effects. This result is in agreement with those reported by Cardon and Fulker (1993). In their study, an alternative longitudinal multivariate model was applied to specific abilities data obtained in children measured at ages 3, 4, 7, and 9 years. Their findings suggested that domain-specific genetic factors remained identical across age, and that their importance increased over time. In the present study, the high stability of the genetic verbal and nonverbal factors was expressed by the large across-age correlations (range from .87 to .99). Following the initial measurement occasion in the 5-year-olds, these group factors explained an equal or increased proportion of the total variance.

Possibly, a correlation between environment and genotype of the active kind may play a role here (Plomin et al., 1977; Scarr & McCartney, 1983). Such a correlation arises when a child seeks out and creates environmental conditions that match his or her genetic propensities. In other words, the child selects those experiences that are compatible with his or her cognitive abilities. Because the selection of environmental settings depends on the genotype of the individual, MZ twins are expected to display greater similarity in their choices of these environmental conditions than DZ twins. As a result, similarity in MZ twins will increase. Because the difference in resemblance between monozygotic and dizygotic twins gives an indication of the importance of genetic influence, the effect of a correlation between environment and genes emerges as genetic variation in the present design (Plomin, 1994). However, since the increase in heritability is not consistent across all subtests, the effect of such a hypothesized correlation is expected to be moderate. This effect may become appreciable later in adolescence.

The unique environment

The longitudinal findings relating to the unique environment differed greatly from those relating to the shared environment. A lack of stability from one age to the next was indicated by near-zero correlations among subtest specific factors. Although the unique environmental effects lacked stability, they did remain the most important environmental source of individual differences in cognitive abilities. Specific factors accounted for one third to half of the total variance for each subtest at each age. The unique environmental effect accounted for less variance in the verbal subtests than in the nonverbal subtests. The estimate of the unique environment includes measurement error. Studies of general IQ in the same sample of 200 twin pairs find smaller estimates of the unique environmental effect (Bartels et al., 2002; Boomsma & Van Baal, 1998). When a composite IQ score is analyzed, the unique environmental effects generally account for less variance, because the error variance of a composite measure is generally reduced (Eaves et al., 1989; p. 201-202).

Numerous developmental studies of intelligence have indicated the important role of the environment in making children in the same family different from one another (for discussion see Plomin & Daniels, 1987). So far, the exact nature of these environmental factors remains unknown (Plomin et al., 1996; Turkheimer & Waldron, 2000). Considering the occasion-specific nature of the unique environmental effect and the young age of the twins, it may be tempting to ascribe these effects to unreliability of the specific measurement occasion. However, the fact that the covariances between these effects do not differ from zero should not be interpreted to mean that these effects are due to pure measurement error. The large estimates of unique environmental effects at each age are not consistent with the known reliability of the administered RAKIT subtests (Bleichrodt et al., 1984). From the reported internal consistencies it follows that a substantial part of the variance unique to the individual must result from influences other than measurement error. The observed difference in subtest performance between MZ twins is thus most likely a result of both genuine, but transient, environmental effects and measurement error effects (Neale & Cardon, 1992, p.14).

Differentiation of cognitive abilities

Since the early factorial studies of cognition (e.g., Cohen, 1957), it has been acknowledged that the degree of covariation among cognitive abilities may vary as a function of age. It is hypothesized (Garrett, 1946) that during development, the construct of intelligence develops as an increasingly loosely structured group of abilities, that is, intelligence becomes differentiated. A decrease in correlations between different abilities across time is indicative of cognitive differentiation. Phenotypic evidence for such a development is weak at best (Bickley et al., 1995; Carroll, 1993; Deary et al., 1996; Juan-Espinosa et al., 2000; Reinert, 1970; Werdelin & Stjernberg, 1995). In this thesis, the issue of differentiation is studied at the latent genetic and environmental level.

The genetic common two factors represent the verbal and nonverbal domain of intelligence. Any decrease in the magnitude of the correlation between these factors indicates an increase in differentiation. Such a decrease was not observed, that is, the correlations remained relatively constant across age. Although the shared environmental effects were modeled as a single common factor at each occasion, differentiation may still occur if the factor loading decrease in magnitude over the occasions. However, the shared environment explained relatively little stability in subtest performance, and can therefore not be considered a clear source of differentiation of abilities. The presence of specific cognitive abilities is suggested by the subtest-specific structure of the unique environmental effects. Since this structure was established at age 5, and given that the unique environment did not display a distinct developmental pattern, differentiation of intellectual performance

due to the unique environment is not indicated. From the present study it is concluded that the differentiation hypothesis is not supported in the interval between 5 and 10 years.

Discussion: The development of Overactivity and Attention Problems

Prior to the analyses of Overactivity and Attention Problems, two methodological issues needed to be resolved: zygosity determination and the consequences for heritability estimates of the presence of a contrast effect. Given large sample sizes, the determination of zygosity is only feasible by use of questionnaires. In this thesis, assignment of zygosity is obtained by discriminant analysis. The consequences of a contrast effect for heritability estimates, and the feasibility of detecting such an effect were explored. Power calculations were performed with sample sizes and genetic and environmental contributions that are usually reported in the behavior genetic literature on ADHD phenotypes.

The following questions were addressed in the age-specific and longitudinal analyses of Overactivity and Attention Problem behavior. What are the age-specific prevalence rates of children eligible to receive the diagnosis ADHD by DSM criteria? Are sex differences present in these prevalence rates? To what degree is Overactivity at age 3 a precursor of Attention Problems at later ages? Is stability of these traits dependent on sex and age of the twins? How heritable are these traits? In what way contribute genetic and environmental effects to the observed stability in Overactivity and Attention Problems?

Classification of twin zygosity

The influence of age and the effect of informant on zygosity classification were evaluated by the number of correctly classified twins obtained by a series of discriminant analyses. The accuracy of zygosity classification based on questionnaire data was estimated at about 93%. Accuracy did not increase with increasing age of the twins. In addition, mothers and fathers were equally effective in classifying their children. Further, when longitudinal data were pooled over age and when rater-specific data were pooled over fathers and mothers, the accuracy of zygosity classification was not increased. Mothers and fathers answered the zygosity items in a similar way. The use of a single parent and replacing this one parent by the other in case of incomplete data is therefore feasible. Mothers with knowledge of the actual zygosity status did not differ in response pattern from mothers who were lacking this knowledge (see also Jackson et al., 2001). In evaluating the accuracy of parental opinion, it was observed that true MZ twins were more often misclassified as DZ twins than vice versa. This result appeared to reflect the parental preference of labeling a twin pair as dizygotic, and thus the parental tendency to emphasize any difference between twins (see

also Price et al., 2000). The application of discriminant analysis to multiple questionnaire items is therefore a much more reliable method to determine zygosity.

The detection of a contrast effect

Twin studies of ADHD phenotypes have revealed the presence of a parental contrast effect. Most of these studies have adopted model fitting procedures, in which the simultaneous specification of genetic dominance and the contrast effect is avoided. In this thesis, it was established that a model that includes sources of variance due to additive genes, genetic dominance, the unique environment, and the contrast effect (ADE-b) is identified. If neglected, the contrast effect may result in an overestimation of broad-sense heritability. The detection of a large contrast effect (> -.15) is feasible given a reasonable sample size ($N \geq 300$ pairs). Smaller contrast effects can only be detected in large studies including at least 3000 twin pairs. The power to distinguish between a contrast effect and genetic dominance is greatly improved by the inclusion of genetically unrelated sibling pairs.

The summary of twin studies on ADHD phenotypes in Table 1.1 (Chapter 1) shows the diversity among assessment instruments employed in twin studies. Each instrument requires a specific method of assessment, which in turn may also affect the degree to which parental bias takes place (Thapar, 2003). A personal interview may encourage parents to focus on the actual behavior of the individual child, and thus may reduce the bias in their report (Nadder et al., 1998). Another variable relevant to the rater contrast is age of the children. A moderate to large contrast effect for hyperactive and overactive behavior in the preschool child is reported in the present study of Overactivity and the study of Price and colleagues (submitted). In assessing the behavior of a 3-year-old child, a parent may compare the child to his/her co-twin, in the absence of same-aged children who could provide a reference. When children grow up, and especially after they start going to school, parents' exposure to same-aged children is increased, which may lead to a reduction of the parental contrast effect. This developmental aspect has not been given much attention in the behavior genetic literature on ADHD symptoms. Given the longitudinal design, the young age of the twins at the beginning of the study and the focus on specific ages, the most promising studies are those performed at the Netherlands Twin Register.

Sex differences in Overactivity & Attention Problems

The often reported difference in behavior problems between boys and girls (Eme, 1992; Helgeson, 2002) was also observed in the present study. At ages 3, 7, 10, and 12, boys, on average, displayed significantly more Overactive and Attention Problem behavior than girls. The sex effect was also observed in the percentages children with a T-score ≥ 67. The boy-to-girl ratio is about 1.5 to 1. At age 3, between 2% and 3% were considered

borderline / clinical cases. At ages 7, 10, and 12, between 3.1% to 6.5% displayed severe Attention Problem behavior with the smallest estimates at age 12 in both boys and girls.

It is well established that clinical and community studies differ in the report of sex effects in ADHD prevalence rates (Gaub & Carlson, 1997). Such a difference between studies is also observed for other phenotypes, for instance, reading disability (DeFries et al., 1990). One explanation of the discrepancy between studies is that gender prevalence rates may differ across informants. Some studies found that mothers and teachers provided opposite rates of hyperactivity across gender (Costello et al., 1991). By maternal report, girls displayed substantially more hyperactive behavior compared to teacher report (see also, McGee & Feehan, 1991; Verhulst et al., 1994). To obtain a diagnosis by DSM criteria the behavior has to be pervasive across situations (criterion C; APA 1994), that is, both at home and at school. So, parents and teachers have to provide corresponding information to establish ADHD. If this requirement is met more often in boys than in girls, it results in a larger sex effect in clinical studies. A community study by Pineda et al. (1999) provided some evidence in support of this argument. By parental report, the DSM-IV symptoms list (criteria A; APA, 1994) was used to establish the presence of ADHD in a sample of 540 boys and girls. Interestingly, the boy-to-girl ratio was 1.5 to 1. Apparently, excluding the criterion that a child must display problematic behavior across settings led to a prevalence of ADHD that differed only marginally between sexes. The Netherlands Twin Registry is currently collecting teacher report forms (Achenbach, 1991) on 7-, 10-, and 12-year-old twins. On the basis of the data collected so far, it was found that teachers, compared to mothers, report more than twice as much Attention Problem behavior in both boys and girls. The calculation of the number of borderline / clinical cases based on these teacher reports resulted in a much larger sex ratio at ages 7 and 10 (boys : girls, 3 : 1) compared to the ratio based on maternal reports (boys : girls, 1.5 : 1). The difference in degree of Attention Problems and the difference in sex ratio between teacher and maternal reports suggests an interaction between sex of the child and situation, in which the child's behavior is evaluated. Apparently, children display more Attention Problems in the school setting with boys exhibiting more severe Attention Problems than girls. As a consequence, more boys enter a mental health clinic than girls.

The stability of Overactive / Attention Problem behavior

Retrospective clinical studies consistently report that the development of ADHD originates in childhood hyperactivity during the preschool period (Barkley et al., 1990; Biederman, 1998; McGee et al., 1992). The strength of the present work is the use of age-appropriate syndrome scales, Overactivity at preschool age and Attention Problems at school age. The correlation from Overactivity at age 3 to Attention Problems at ages 7, 10,

and 12 was estimated at .41. Stability in Attention Problems was estimated at about .70. Both correlations are in line with those reported in other Dutch and American studies that use the CBCL (Achenbach et al., 2000; Koot, 1993; Van der Valk et al., 1998; Verhulst et al., 1995). In this thesis, it was hypothesized that the moderate degree of stability from age 3 to older ages was due to developmental change. This was assessed by comparing the correlations between identical Overactivity and Attention Problem items to the correlations between the Attention Problem scales at ages 7 and 10, and at ages 10 and 12. Both sets of correlations (i.e., between items and between scales) are calculated between constructs that are identical in content. Thus, if the age-interval is irrelevant, the correlations should resemble one another. This was not the case. The item correlations suggested a very moderate degree of stability. From this, it was concluded that the age-interval being studied captures a period of developmental change.

Around 95% of the children that fall within the normal range of Overactive behavior do not experience severe Attention Problem behavior at age 7. Comparable rates apply to the intervals from ages 7 to 10, and from ages 10 to 12. The developmental pathways of children that were likely to meet DSM criteria at any age are depicted in Figure 1.1 in Chapter 1. Because children with severe symptoms were somewhat more likely to drop out of the study, these pathways should be interpreted with caution. Overall, one out of three children with a high score on the Overactivity scale also had a high score on the Attention Problem scale at any later age. Girls appeared to display more developmental variety than boys. These developmental trajectories are comparable to those reported by a community study known as the Dunedin survey (reviewed by Costello & Angold, 1995; McGee et al., 1991).

Genetic and environmental influences on Overactivity / Attention Problems

From the behavior genetic literature it has become clear that ADHD phenotypes are highly heritable (Table 1.1, Chapter 1). Within the range of childhood psychopathology, the genetic and environmental effects on ADHD phenotypes display a rather typical pattern. In comparison to other syndromes assessed with the CBCL, genetic dominance accounts for a significant proportion of the total observed variation in Overactivity / Attention Problems. In contrast, shared environmental effects are found to be important in explaining variation in other, phenotypically related syndromes like aggressive and oppositional behaviors (Derks et al., submitted; Van der Valk et al., 1998). The result of the present study agrees well with the results of other large twin studies. Genetic effects explained 75% and unique environmental effects explained 25% of the total variation in Overactivity and Attention Problems at each age. Boys displayed more variation in Attention Problems than girls at ages 10 and 12. The genetic and environmental effects are expressed as a

proportion of the total variance, that is, being relative estimates. At age 10 and age 12, genes and environment explained an equal proportion of the total variation in Attention Problems. Further, it appeared that boys and girls share the same set of genetic and environmental effects.

The stability in Overactivity / Attention Problems was attributable to genetic influences (range 76% - 92%; broad heritability). Here, a sex difference emerged. In boys, genetic dominance effects explained most of the covariance in Attention Problems, whereas in girls additive genetic effects were more important. The unique environment accounted for a small proportion of the covariance between Overactivity / Attention Problems (range 8% - 24%). Indicated by the imperfect correlations between the different ages, genetic effects were not completely stable over development. With increasing age, the overlap increased in genetic dominance effects, but remained of equal size in additive genetic effects. The overlap in unique environmental effects across age displayed a similar pattern as the overlap in genetic dominance effects; correlations for intervals including age 3 were substantially smaller compared to the intervals beyond age 3.

Numerous studies have described the environmental influences that are relevant to the development of ADHD phenotypes (Faraone & Doyle, 2001). These influences appear to act in a non-specific way and predispose the child to a range of psychopathology. Risk factors are parenting style (Shaw et al., 20001), negative life events (Achenbach et al., 1995), exposure to parental psychopathology (Barkley et al., 1990; Biederman et al., 2002), and socio-economic status (Pineda et al., 1999; Szatmari et al., 1989). Young age of the mother, maternal education (Mathiesen & Sanson, 2000), smoking during pregnancy, and other pre- and perinatal complications (Milberger et al., 1996; 1997; 1998) are also reported to be associated with an increase in problem behaviors. At first glance, these risk factors appear to be shared by children in the same home, and should thus be regarded as shared environmental influences. However, the results presented in this thesis and by other twin studies have reported the absence of shared environmental effects. The apparent lack of consistency among twin and singleton studies is explained by the use of the term 'shared environmental influence' in behavior genetic studies. An environmental influence is considered shared when children of the same family are equally affected by this influence (Kendler, 2001). It is not unlikely, however, that children respond differently to the same environmental factor to which they are exposed. Such a scenario represents an interaction between genetic and environmental effects (Plomin et al., 1977). Throughout the analyses of Overactivity / Attention Problems it was assumed that such an interaction was absent. Violation of this assumption implies that variation in the environment affects individuals differently depending on their genotypes (Boomsma & Martin, 2002; Caspi et al., 2002). If the interaction takes place between genetic effects and unique environmental

effects, the unique environmental estimate is inflated. The contribution of the unique environment was estimated at 25%. In addition to the actual effects of the unique environment, this estimate includes measurement error. This suggests that an additional component due to the interaction between unique environmental and genetic effects is small at the most and may be considered not important in explaining individual variation in Overactivity / Attention Problems. If the interaction takes place with the shared environment, genetic estimates are increased. Given the heritability estimate of 75%, it is possible that this estimate includes a large component due to interaction between shared environmental risk factors and genetic effects.

Synthesis: Intelligence and Overactivity / Attention Problems

Is Overactive / Attention Problem behavior related to intellectual performance? If so, is the association only observed in children eligible to receive an ADHD diagnosis or is the association with cognitive abilities observed across the entire range of Overactivity / Attention Problem behavior? Is the association influenced by genetic and (or) environmental factors? Do the findings differ by sex? In an attempt to answer these questions, data collected in the study of Intelligence were combined with data collected in the study of Overactive / Attention Problem behavior. Around 350 children provided information on verbal and nonverbal cognitive abilities at ages 5, 7, and 10 *and* Overactive / Attention Problems at ages 3, 7, 10, and 12. There is no difference in degree of problem behavior between children who participated and those who did not participated in this combined study. Based on the literature, it was decided to include additional measures of school performance. First, an introduction to the literature is presented. Second, the questions posed above are answered by analyses of the combined dataset. Finally, results are discussed.

The literature on Intelligence & Overactivity / Attention Problems

The relationship between psychometric intelligence and childhood problem behaviors has yet to be established firmly. Among the studies that report an effect, measures of intelligence usually correlate negatively with externalizing problem behaviors like hyperactivity, impulsivity, and aggressiveness (Cole et al., 1993; Cook et al., 1994; Goodman, 1995). However, among studies based on either a clinical or community based sample, there is little agreement that children displaying hyperactive, impulsive and inattentive behavior are characterized by lower psychometric intelligence (Bonafina et al., 2000; Cohen et al., 2000; Faraone et al., 1993). In the Dunedin community study, measures of IQ and hyperactivity by teacher report were collected in a large sample of 7-year-old children (McGee et al., 1985). The correlations between the behavioral and cognitive scores were significant, estimated between -.17 and -.21 with no difference between boys and girls. This result is in agreement with the results obtained in a community study with 13-year-old children (Goodman et al., 1995): full scale IQ correlated -.15 with hyperactivity as assessed by parental report. The number of community studies, in which psychometric intelligence was assessed is rather limited compared to the number of clinical studies. Despite the severity of behavior problems of referred children, these children do not always display a low IQ. For instance, no difference in full scale IQ emerged between groups differing in degree of persistent ADHD, and controls in a follow-up study of referred boys (Hart et al., 1995). In

contrast, in a clinical study with clinical controls, hyperactive children showed a marked decrease in IQ, with the largest effect observed in girls (James & Taylor, 1990).

In contrast to psychometric intelligence, the association between scholastic achievement and childhood problem behaviors is well established (Frick et al., 1991; Hinshaw, 1992; Kowaleski-Jones & Duncan, 1999; Velez et al., 1989; Wilens et al., 2002). Children who display externalizing, disruptive behaviors are more likely to underachieve in school compared to children who display internalizing problem behaviors. The degree of comorbidity depends on the assessment instrument that is used to quantify underachievement. Measures of reading ability, arithmetic and mathematic abilities, school failure as well as teacher reports have been used as an index of scholastic achievement. The use of these different concepts is confusing. To further complicate the interpretation of scholastic achievement, some studies use the term learning disability. Learning disability is operationalized as a discrepancy between scholastic achievement and intelligence (Fletcher et al., 1999). Scholastic achievement and intelligence are related but different concepts (Ceci, 1991). To illustrate, whereas no difference in IQ was observed between children with and without persistent childhood disorders in the Isle of Wight studies (Graham & Rutter, 1973), the latter group of children displayed significant reading impairment. The comorbidity of intellectual performance, reading ability and learning problems in children with ADHD is discussed by Fletcher et al. (1999). Estimates of the prevalence of comorbidity vary greatly among studies due to differences in sampling strategies and concept operationalization.

In conclusion, the association between psychometric intelligence and hyperactive, impulsive and inattentive behavior seems present but weak. The association between ADHD symptoms and measurements that relate specifically to an educational setting is more firmly established.

Is Overactive / Attention Problem behavior associated with cognitive abilities?

A consistent pattern of negative correlations was observed between Overactivity / Attention Problems and RAKIT intelligence subtests scores. Within and across age intervals, Overactivity / Attention Problems correlated significantly with at least one out of three nonverbal subtests. The subtest Exclusion was found most often associated with Overactivity / Attention Problems. Correlations ranged between -.18 to -.11 with no difference between small age intervals (e.g., Attention Problems and nonverbal subtests both assessed at age 7) and large age intervals (e.g., Overactivity assessed at age 3 and nonverbal subtests assessed at age 10). Among the three verbal subtests, Learning Names correlated with Overactivity / Attention Problems for some but not all age intervals. In an attempt to limit the large number of results, subsequent analyses were based on composite scores

of nonverbal (NV-IQ: Exclusion, Discs, and Hidden Figures)[1] and verbal IQ (V-IQ: Verbal Meaning, Learning Names, and Idea Production)[1]. Since only one marginally significant effect was found for V-IQ, further analyses were limited to NV-IQ. The within-age and across-age correlations between NV-IQ and Overactivity / Attention Problems did not differ by sex. Correlations ranged from -.13 to -.26 with the majority of the correlations estimated around -.15. No difference was observed between within-age and across-age correlations.

From scatterplots and gradual changes in mean NV-IQ, it appears that the negative association is not limited to children at the tail of the Overactivity / Attention Problems distribution.

Cross-twin correlations between NV-IQ and Overactivity / Attention Problems

Cross-twin correlations were calculated to obtain an initial insight into the influences that contribute to the observed association between NV-IQ and Overactivity / Attention Problems. These correlations were calculated for a large number of age-intervals; between the behavioral scores at ages 3, 7, 10, 12, and the intelligence scores at ages 5, 7, 10. MZ correlations were estimated between -.20 and -.32; DZ correlations were estimated between .06 and -.12. The results did not differ by age-interval. Given the small degree of resemblance between DZ twins, the shared environment hardly (if at all) contributes to the observed covariance between NV-IQ and Overactivity / Attention Problems. The large difference between MZ and DZ cross-correlations suggests that the observed covariance is due to genetic effects. However, it should be noted that the correlations are estimated on samples of, on average, 150 pairs. This is a relatively small sample size to establish a firm conclusion about the contribution of the genetic and environmental effects to the covariance of NV-IQ and Overactivity / Attention Problems.

Is School Performance associated with Attention Problems & NV-IQ?

An additional questionnaire, attached to the CBCL/4-18, was used to assess various aspects the twins' functioning in school. On a five-point scale, mothers provided information upon the twins' School Performance with respect to math and language achievement. Whether a child had *ever* experienced learning problems at school was assessed by one item with two response categories. At age 7, around 16% (n = 51) suffered from learning problems or had done so at an earlier age. At age 10, this number increased to 22%. Results are reported for NV-IQ measured at ages 7 and 10. The availability of

[1] NV-IQ and V-IQ were calculated by 15/11 (X-45)+110, in which X is the addition of the three matching subtests.

data of Attention Problems, NV-IQ, math and language achievement, and learning problems permitted the exploration of the five associations that involved School Performance variables. The results are as follows:

(1) Both math (-.28 at age 7; -.32 at age 10) and language (-.37 at age 7; -.34 at age 10) achievement correlated significantly with Attention Problems. (2) The positive relationship between math and language achievement and NV-IQ was estimated at .30 and .28 at age 7, respectively, and at .41 and .35 at age 10, respectively. (3) Children who suffered or had suffered from learning problems displayed significantly more Attention Problems than children who had never experienced learning problems at ages 7 and 10. (4) A comparable difference between categories of learning problems was observed in mean NV-IQ and, (5) in math and language achievement.

Discussion of Overactivity / Attention Problems & NV-IQ & School Performance

In line with the literature, a very modest negative association between Overactive / Attention Problem behavior and psychometric intelligence was observed. Among the RAKIT subtests, the nonverbal subtests displayed a stronger relation with Overactive / Attention Problem behavior than the verbal subtests. This relation appeared to be due to genetic effects. A recent study also suggests the role of genes in explaining the covariance between child psychopathology and cognitive ability (Jacobs et al., 2003). In the literature, there are few studies that focus on specific intellectual abilities. Since ADHD is often reported to coincide with reading problems (Beitchman, 1985; Beitchman et al., 1982; Werry et al., 1987), children's verbal and language abilities are often assessed, whereas measures of other abilities are not included in the study (e.g., Barkley et al., 1990). An exception is the non-clinical study by Dietz et al. (1997) with children of pre-school age. It was established that both lower verbal and performal IQ was associated with increased externalizing problem behavior (see also DuPaul et al., 2001). The study of Dietz et al. (1997) also is one of the few studies that examined the relation between psychometric intelligence and childhood problem behavior in a quantitative dataset, assuming that behavior problems are at the extreme of a latent continuum. In the present thesis it was also found that the negative association between Attention Problems and psychometric intelligence was present over the full range, and not solely at the extreme of the distribution of Attention Problems.

Whereas a lower nonverbal IQ in children with higher Overactivity / Attention Problem scores is established in this thesis, the association between lowered verbal IQ and Overactivity / Attention Problems appears to be absent. This is a puzzling outcome in view of the literature on childhood psychopathology, and in view of the negative association between language achievement and Attention Problems. Post hoc analyses revealed a significant correlation between V-IQ and language achievement of .24 at age 7, and .33 at

age 10 (see also Ceci, 1991). Apparently, the variance that language achievement and Attention Problems have in common is not shared with V-IQ. It was found that each School Performance variable correlated significantly with Attention Problems. An increase in Attention Problem behavior was related to a decrease in math and language achievement and a higher chance of having once experienced learning problems at school. These findings confirm the results obtained in a study of ADHD subtypes in adolescents (Todd et al., 2002). Interestingly, the correlation between Attention Problems and the measures of School Performance appeared much larger than the correlation between Attention Problems and NV-IQ. In view of these results, it may be hypothesized that children with Attention Problems experience more difficulties in the classroom than would be predicted from NV-IQ alone. However, a limitation regarding the measurement of School Performance should be noted. School Performance was assessed by maternal report. Since School Performance takes place *in* the school, maternal reports should certainly be viewed as an indirect measure. A more direct measure may be obtained from teachers. The teacher may also provide more detailed information on the kind of learning problems that a child has ever experienced. The categorization, based on indirect maternal report, of children with and without learning problems is rather crude. In view of this limitation, a firm conclusion regarding the association between School Performance and Attention Problems has yet to be established.

Future directions

The longitudinal study of Intelligence has already been extended with measurements taken at age 12 (Bartels et al., 2002). This year, the twins will be assessed when they reach the age of 17. The continued participation of around 400 children and their parents in the ongoing study of Intelligence is unique in the Netherlands and within the field of behavior genetics. The repeated measurements of psychometric intelligence provides the possibility to understand development of intelligence in a healthy, population-based sample of children. The availability of twin data makes it possible to study the genetic and environmental contributions to stability of intelligence. In the present thesis, sex differences in mean performance and heritability were absent. Entering adolescence, the study of sex differences is highly relevant. Sex differences in cognitive abilities appear to emerge during adolescence and continue to exist into adulthood (Sadker et al., 1991). Anticipating the presence of such an effect, the collection of sex hormones will take place when twins are assessed at the age of 17.

The longitudinal study of Overactivity / Attention Problems is part of the longitudinal study of childhood behavior problems (Van der Valk, 2001). The ongoing collection of CBCL data has resulted in large sample sizes. These sample sizes make it possible to estimate genetic and environmental contributions with good reliability and precision, and to test for the presence of genetic dominance or the contrast effect in Overactivity / Attention Problems. The longitudinal study of childhood behavior problems has been extended with the collection of teacher reports (Teacher Report Form; Achenbach, 1991) and will be extended with self reports of the adolescents (Youth Self Report; Achenbach, 1991). The combination of parental, teacher, and self reports provide a unique perspective on the development of children's behavior. Further, DSM-IV data are being collected which allows to further explore the overlap between measurement instruments.

The additional collection of data in both the study of Intelligence and the study of Overactivity / Attention Problems provides large amount of data when these studies are combined. The combined data may be used to whether Overactivity / Attention Problems proceeds a lowered IQ, or vice versa. The Teacher Report Form provides additional information on how children behave in the school setting. Further, an additional measure of scholastic achievement is being collected in 12-year-old twins (Bartels et al., 2003). This information may be used as a direct estimate of School Performance, and it can be used to explore the association with Overactivity / Attention Problem behavior.

References

Abramovitch, R., Carter, C., & Lando, B. (1979). Sibling interaction in the home. *Child Dev* 50: 997-1003.

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

Achenbach, T.M. (1991b). *Manual for the Teacher's Report Form.* Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T.M. (1991c). *Manual for the Youth Self-Report.* Burlington, VT: University of Vermont, Department of Psychiatry.

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

Achenbach, T.M., Howell, C.T., McConaughy, S.H., & Stanger, C. (1995). Six-year predictors of problems in a national sample of children and youth. *J Am Acad Child Adolesc Psychiatry* 34: 488-498.

Achenbach, T.M., & Rescorla, L.A. (2000). *Manual for the ASEBA Preschool Forms & Profiles.* Burlington, VT: University of Vermont, Research Center for Children, Youth & Families.

Alarcón, M., Plomin, R., Fulker, D.W., Corley, R., & DeFries, J.C. (1998). Multivariate path analysis of specific cognitive abilities data at 12 years of age in the Colorado Adoption Project. *Behav Genet* 28(4): 255-264.

Alarcón, M., Plomin, R., Fulker, D.W., Corley, R., & DeFries, J.C. (1999). Molarity not modularity: Multivariate genetic analysis of specific cognitive abilities in parents and their 16-year-old children in the Colorado Adoption Project. *Cognitive Dev* 14: 175-193.

American Psychiatric Association (1994). Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; 4th edition). Washington DC: American Psychiatric Association.

Baker, L.A., Ho, H.Z., & Reynolds, C. (1994). Sex differences in genetic and environmental influences for cognitive abilities. In: Defries, J.C., Plomin, R., Fulker, D.W. (Eds.). *Nature and Nurture during Middle Childhood.* Oxford: Blackwell.

Barkley, R.A., Fisher, M., Edelbrock, C.S., & Smallish, L. (1990). The adolescent outcome of hyperactive children diagnosed by research criteria: I. An 8-year prospective follow-up study. *J Am Acad Child Adolesc Psychiatry* 29(4): 546-557.

Barr, C., Swanson, J., & Kennedy, J. (2001). Molecular genetics of ADHD. In: Levy, F., Hay, D.A. (Eds.) *Attention, Genes and ADHD* (pp. 173-195). East Sussex: Brunner-Routledge.

Bartels, M., Rietveld, M.J.H., Van Baal, G.C.M., & Boomsma, D.I. (2002). Genetic and environmental influences on the development of intelligence. *Behav Genet* 32(4): 237-249.

Bartels, M., Rietveld, M.J.H., Van Baal, G.C.M., & Boomsma, D.I. (2003). Heritability of educational achievement in 12-year-olds and the overlap with cognitive ability. *Twin res* 5(6): 544-553.

Beitchman, J.H., Patterson, P., Gelfand, B., & Minty, G. (1982). IQ and child psychiatric disorder. *Can J Psychiatry* 27(1) :23-28.

Beitchman, J.H. (1985). Speech and language impairment and psychiatric risk. Toward a model of neurodevelopmental immaturity. *Psychiatr Clin North Am* 8(4): 721-735.

Bekker, P.A., Merckens, A., & Wansbeek, T.J. (1993). *Identification, equivalent models, and computer algebra.* Boston, Mass. : Academic Press.

Bennet Murphy, L., Murphy, C.E., & Rose, C.L. (2001). Sustained attention and unintentional injury among preschool-aged children. *Dev Neuropsychol* 7: 72-83.

Bickley, P.G., Keith, T.Z., & Wolfle, L.M. (1995). The three-stratum theory of cognitive abilities: test of the structure of intelligence across the life span. *Intelligence* 20: 309-328.

Biederman, J., Faraone, S.V., Doyle, A., Lehman, B.K., Kraus, I., Perrin, J., Tsuang, M.T. (1993). Convergence of the Child Behavior Checklist with structured interview-based psychiatric diagnoses of ADHD children with and without comorbidity. *J Child Psychol Psychiatry* 34(7): 1241-1251.

Biederman, J. (1998). Attention-deficit/hyperactivity disorder: a life-span perspective. *J Clin Psychiatry* 59 Suppl 7: 4-16.

Biederman, J., Mick, E., & Faraone, S.V. (2000). Age-dependent decline of symptoms of attention deficit hyperactivity disorder: impact of remission definition and symptom type. *Am J Psychiatry* 157(5): 816-818.

Biederman, J., Faraone, S.V., & Monuteaux, M.C. (2002). Differential effect of environmental adversity by gender: Rutter's index of adversity in a group of boys and girls with and without ADHD. *Am J Psychiatry* 159(9): 1556-1562.

Bird, H.R., Canino, G., Rubio-Stipec, M., Gould, M.S., Ribera, J., Sesman, M., Woodbury, M., Huertas-Goldman, S., Pagan, A., Sanchez-Lacay, A., Moscoso, M. (1988). Estimates of the prevalence of childhood maladjustment in a community survey in Puerto Rico. The use of combined measures. *Arch.Gen. Psychiatry* 45(12): 1120-1126. [published erratum in *Arch Gen Psychiatry* (1994) 51:429]

Bishop, E.G., Cherny, S.S., Corley, R., Plomin, R., DeFries, J.C., & Hewitt, J.K. (2003). Development genetic analysis of general cognitive ability from 1 to 12 years in a sample of adoptees, biological siblings, and twins. *Intelligence* 31: 31-49.

Bleichrodt, N., Resing, W.C.M., Drenth, P.J.D., & Zaal, J.N. (1984). *Revisie Amsterdamse Kinder Intelligentie Test [revised Amsterdam Child Intelligence Test]*. Lisse, the Netherlands: Swets & Zeitlinger B.V.

Bleichrodt, N., Resing, W.C.M., Drenth, P.J.D., & Zaal, J.N (1987). *Intelligentie-meting bij kinderen. Empirische en methodologische verantwoording van de gereviseerde kinder intelligentie test [Intelligence-measurement in children. Empirical and methodological justification of the revised child intelligence test]*. Lisse, The Netherlands: Swets & Zeitlinger B.V.

Bonafina, M.A., Newcorn, J.H., McKay, K.E., Koda, V.H., & Halperin, J.M. (2000). ADHD and reading disabilities: a cluster analytic approach for distinguishing subgroups. *J Learn Disabil* 33: 297-307.

Bønnelykke B., Hauge, M., Holm, N., Kristoffersen, K., & Gurtler, H. (1989). Evaluation of zygosity diagnosis in twin pairs below age seven by means of a mailed questionnaire. *Acta Genet Med Gemellol (Roma.)* 38(3-4): 305-313.

Boomsma, D.I. (1993). Current status and future prospects in twin studies of the development of cognitive abilities: infancy to old age. In: Bouchard, T.J. JR. & Propping, P. (Eds.), *Twins as a Tool of Behavorial Genetics* (pp. 67-82). Chichester, UK: Wiley.

Boomsma, D.I. (1998). Twin registers in Europe: an overview. *Twin Res* 1(1): 34-51.

Boomsma, D.I. & Martin, N.G. (2002). Gene-environment interactions. In: D'Haenen, H., Den Boer, J.A., Willner, P. (Eds.), *Biological Psychiatry* (pp. 181-187). New York: John Wiley & Sons, Ltd.

Boomsma, D.I., & Molenaar, P.C. (1986). Using LISREL to analyze genetic and environmental covariance structure. *Behav Genet* 16(2): 237-250.

Boomsma, D.I., & Molenaar, P.C. (1987). The genetic analysis of repeated measures. I. Simplex models. *Behav Genet* 17(2): 111-123.

Boomsma, D.I., Orlebeke, J.F., & Van Baal, G.C.M. (1992). The Dutch Twin Register: growth data on weight and height. *Behav Genet* 22(2): 247-251.

Boomsma, D.I., & Van Baal, G.C.M. (1998). Genetic influences on childhood IQ in 5- and 7-year-old Dutch twins. *Dev Neuropsychol* 14(1): 115-126.

Bruner, M.G. (1970). *The Biology of Twinning in Man*. Oxford: Clarendon Press.

Campbell, S.B. (1995). Behavior problems in preschool children: a review of recent research. *J Child Psychol Psychiatry* 36(1): 113-149.

Cardon, L.R. (1994). Specific cognitive abilities. In: DeFries, J.C., Plomin, R. & Fulker, D.W. (Eds.), *Nature and Nurture during Middle Childhood* (pp. 57-76). Oxford, UK: Blackwell.

Cardon, L.R. & Fulker, D.W. (1993). Genetics of specific cognitive abilities. In: Plomin, R. & McClearn G.E. (Eds.), *Nature, Nurture & Psychology* (pp. 99-120). Washington: American Psychological Association.

Cardon, L.R., Fulker, D.W., DeFries, J.C., & Plomin, R. (1992). Multivariate genetic analysis of specific cognitive abilities in the Colorado Adoption Project at age 7. *Intelligence* 16: 383-400.

Carey, G. (1986). Sibling imitation and contrast effects. *Behav Genet* 16(3): 319-341.

Carey, G. (1992). Twin imitation for antisocial behavior: implications for genetic and family environment research. *J Abnorm Psychol* 101(1): 18-25.

Carroll, J.B. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. Cambridge: Cambridge University Press.

Carter-Saltzman, L. & Scarr, S. (1977). MZ or DZ? Only your blood grouping laboratory knows for sure. *Behav Genet* 7(4): 273-280.

Caspi, A., McClay, J., Moffitt, T.E., Mill, J., Martin, J., Craig, I.W., Taylor, A., & Poulton,R. (2002). Role of genotype in the cycle of violence in maltreated children. *Science* 297(5582): 851-854.

Casto, S.D., DeFries, J.C., & Fulker D.W. (1995). Multivariate genetic analysis of Wechsler Intelligence Scale for Children--Revised (WISC-R) factors. *Behav Genet* 25(1): 25-32.

Ceci, S.J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Dev Psychol* 27: 703-722.

Cederlöf, R., Friberg, L., Jonsson, E., & Kaij, L. (1961). Studies on similarity diagnosis in twins with the aid of mailed questionnaires. *Acta Genet Stat Med* 11: 338-362.

Centraal Bureau voor de Statistiek (1993a). *Standaard Beroepenclassificatie 1992*. Heerlen: Centraal Bureau voor de Statistiek.

Centraal Bureau voor de Statistiek (1993b). *Standaard Onderwijsindeling SOI-1978*. Heerlen: Centraal Bureau voor de Statisitiek.

Centraal Bureau voor de Statistiek (2001). *Enquête Beroepsbevolking 1999*. Heerlen: Centraal Bureau voor de Statistiek.

Centraal Bureau voor de Statistiek (2002a). *Enquête Beroepsbevolking 1999. Standaard Beroepsclassificatie 1992*. Heerlen: Centraal Bureau voor de Statistiek.

Centraal Bureau voor de Statistiek (2002b). *Permanent Onderzoek Leefsituatie, module Gezondheid en Arbeid, specifieke gezondheidsmetingen bij kinderen 2001*. Heerlen: Centraal Bureau voor de Statistiek.

Charlemaine C., Duyme, M., Aubin, J.-T., Guis, F., Marquiset, N., de Pirieux, I., Strub, N., Brossard, Y., Jarry, G., Le Groupe Romulus, Frydman, R., & Pons, J.C. (1997).

Twin zygosity diagnosis by mailed questionnaire below age twelve months. *Acta Genet Med Gemellol (Roma)* 46(3): 147-156.

Chen, W.J., Faraone, S.V., Biederman, J., & Tsuang, M.T. (1994). Diagnostic accuracy of the Child Behavior Checklist scales for attention-deficit hyperactivity disorder: a receiver-operating characteristic analysis. *J Consult Clin Psychol* 62(5): 1017-1025.

Chen, W.J., Chang, H.-W., Wu, M.-Z., Lin, C.C.H., Chang, C., Chiu, Y.-N., & Soong, W.-T. (1999). Diagnosis of zygosity by questionnaire and polymarker polymerase chain reaction in young twins. *Behav Genet* 29(2): 115-123.

Cohen, J.A. (1957). A factor analytically based rationale for the Wechsler Adult Intelligence Scale. *J Consult Psychol* 21:451-457.

Cohen, D.J., Dibble, E., Grawe, J.M., & Pollin, W. (1973). Separating identical from fraternal twins. *Arch Gen Psychiatry* 29(4): 465-469.

Cohen, D.J., Dibble, E., Grawe, J.M., & Pollin, W. (1975). Reliably separating identical from fraternal twins. *Arch Gen Psychiatry* (32): 1371-1375.

Cohen, N.J., Vallance, D.D., & Barwich, M., Im, N., Menna, R., Horodezky, N.B., & Isaacson, L. (2000). The interface between ADHD and language impairment: an examination of language, achievement, and cognitive processing. *J Child Psychol Psychiatry* 41(3): 353-362.

Cole, P.M., Usher, B.A., & Cargo, A.P. (1993). Cognitive risk and its association with risk for disruptive behavior disorder in preschoolers. *J Clin Child Psychol* 22(2): 154-164.

Cook, E.T., Greenberg, M.T., & Kusche, C.A. (1994). The relations between emotional understanding, intellectual functioning, and disruptive behavior problems in elementary-school-aged children. *J Abnorm Child Psychol* 22(2): 205-219.

Costello, E.J. & Angold, A. (1995). Developmental epidemiology. In: Cicchetti, D., & Cohen, D.J. (Eds.), *Developmental Psychopathology, volume 1: Theory and Methods* (pp. 23-56). New York: John Wiley & Sons, Inc.

Costello, E.J., Loeber, R., & Stouthamer-Loeber, M. (1991). Pervasive and situational hyperactivity--confounding effect of informant: a research note. *J Child Psychol Psychiatry* 32(2): 367-376.

Cunningham, W.R. (1991). Issues in factorial invariance. In: Collings, L.M. & Horn, J.L. (Eds.). *Best Methods for the Analysis of Change* (pp. 106-113). Washington: American Psychological Association.

Dahlberg, G. (1926).*Twin Births and Twins from a Hereditary Point of View* [PhD thesis]. Uppsala, Sweden: University of Uppsala.

Daniels, M., Devlin, B., & Roeder, K. (1997). Of genes and IQ. In: Devlin, B., Fienberg, S.E., Resnick, D.P., & Roeder, K. (Eds.). *Intelligence, Genes, & Success. Scientist respond tot the Bell Curve.* New York: Springer Verlag.

De Groot, A., Koot, H.M., & Verhulst, F.C. (1994). Cross-cultural generalizability of the Child Behavior Checklist cross-informant syndromes. *Psychol Assess* 6: 225-230.

Deary, I.J., Egan, V., Gibson, G.J., Austin E.J., Brand, C.R., & Kellaghan, T. (1996). Intelligence and the differentiation hypothesis. *Intelligence* 23: 105-132.

DeFries, J.C., & Fulker, D.W. (1986). Multivariate behavioral genetics and development: an overview. *Behav Genet* 16(1): 1-10.

DeFries, J.C., Wadsworth, S.J., & Gillis, J.J. (1990). Gender differences in cognitive abilities of reading-disabled twins. *Annals of Dyslexia*, 40: 216-228.

Dencker, S.J., Hauge, M., Kaij, L., & Nielsen, A. (1961). The use of anthropological traits and blood groups in the determination of the zygosity of twins. *Acta Genet* 11: 265-285.

Derks, E.M., Hudziak, J.J., Van Beijsterveldt, C.E.M., & Boomsma, D.I. A study of genetic and environmental influences on CBCBL syndrome scores in a large sample of 3-year-old twins (submitted).

Dietz, K.R., Lavigne, J.V., Arend, R., & Rosenbaum, D. (1997). Relation between intelligence and psychopathology among preschoolers. *J Clin Child Psychol* 26(1): 99-107.

DuPaul, G.J., McGoey, K.E., Eckert, T.L., & VanBrakle, J. (2001). Preschool children with attention-deficit/hyperactivity disorder: impairments in behavioral, social, and school functioning. *J Am Acad Child Adolesc Psychiatry* 40(5): 508-515.

Eaves, L.J. (1972). Computer simulation of sample size and experimental design in human psychogenetics. *Psychol Bull* 77(2): 144-152.

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

Eaves, L.J., Eysenck, H.J., & Martin, N.G. (1989). *Genes, Culture and Personality: An Empirical Approach.* London: Oxford University Press.

Eaves, L.J., Long J., & Heath A.C. (1986). A theory of developmental change in quantitative phenotypes applied to cognitive development. *Behav Genet* 16(1): 143-162.

Eaves, L.J., Rutter, M., Silberg, J.L., Shillady, L., Maes, H.H., & Pickles, A.. (2000). Genetic and environmental causes of covariation in interview assessments of disruptive behavior in child and adolescent twins. *Behav Genet* 30(4): 321-334.

Eaves, L.J., Silberg, J.L., Maes, H.H., Simonoff, E., Pickles, A., Rutter, M., Neale, M.C., Reynolds, C.A., Erikson, M.T., Heath, A.C., Loeber, R., Truett, K.R., & Hewitt, J.K. (1997). Genetics and developmental psychopathology: 2. The main effects of

genes and environment on behavioral problems in the Virginia Twin Study of Adolescent Behavioral Development. *J Child Psychol Psychiatry* 38(8): 965-980.

Edelbrock, C., & Costello A.J. (1988). Convergence between statistically derived behavior problem syndromes and child psychiatric diagnoses. *J Abnorm Child Psychol* 16(2): 219-231.

Edelbrock, C., Rende, R., Plomin, R., & Thompson, L.A. (1995). A twin study of competence and problem behavior in childhood and early adolescence. *J Child Psychol Psychiatry* 36(5): 775-785.

Eisen, S., Neuman, R., Goldberg, J., Rice, J., & True, W. (1989). Determining zygosity in the Vietnam Era Twin Registry: an approach using questionnaires. *Clin Genet* 35(6): 423-432.

Eme, R.F. (1992). Selective female affliction in development of disorders of childhood: a literature review. *J Clin Child Psychol* 21: 354-364.

Fairpo, C.G. (1979). The problem of determining twin zygosity for epidemiological studies. *Acta Genet Med Gemellol (Roma)* 28(1): 21-33.

Falconer, D.S. & Mackay, T.F.C. (1996). *Introduction to Quantitative Genetics.* (4th ed.). Essex: Addison Wesley Longman Limited.

Faraone, S.V., Biederman, J., Lehman, B.K., Keenan, K., Norman, D., Seidman, L.J., Kolodny, R., Kraus, I., Perrin, J., & Chen, W.J. (1993). Intellectual performance and school failure in children with attention deficit hyperactivity disorder and in their siblings. *J Abnorm Psychol* 102(4): 616-623.

Faraone, S.V., & Doyle, A.E. (2001). The nature and heritability of attention-deficit/hyperactivity disorder. *Child Adolesc Psychiatr Clin N Am* 10: 299-316.

Finkel, D., Pedersen, N.L., McGue, M., & McClearn, G.E. (1995). Heritability of cognitive abilities in adult twins: comparison of Minnesota and Swedish data. *Behav Genet* 25(5): 421-431.

Fletcher, J.M., Shaywitz, S.E., & Shaywitz, B.A. (1999). Comorbidity of learning and attention disorders. Separate but equal. *Pediatr Clin North Am* 46(5): 885-97, vi.

Foch, T.T., & Plomin, R. (1980). Specific cognitive abilities in 5- to 12-year-old twins. *Behav Genet* 10(6): 507-520.

Frick, P.J., Kamphaus, R.W., Lahey, B.B., Loeber, R., Christ, M.A.G., Hart, E.L., & Tannenbaum, L.E. (1991). Academic underachievement and the disruptive behavior disorders. *J Consult Clin Psychol* 59: 289-294.

Fulker, D.W., Cherny, S.S., & Cardon, L.R. (1993). Continuity and change in cognitive development. In: Plomin, R., & McClearn, G.E. (Eds.). *Nature, Nurture & Psychology.* Washington: American Psychological Association.

Garcia M.M., Shaw, D.S., Winslow, E.B., & Yaggi, K.E. (2000). Destructive sibling conflict and the development of conduct problems in young boys. *Dev Psychol* 36(1): 44-53.

Garfinkle, A.S. (1982). Genetic and environmental influences on the development of piagetian logico-mathematical concepts and other specific cognitive abilities: a twin study. *Acta Genet Med Gemellol(Roma)* 31(1-2): 10-61.

Garret, H.E. (1946). A developmental theory of intelligence. *Am Psychol* 1: 372-378.

Gaub, M., & Carlson, C.L. (1997). Gender differences in ADHD: a meta-analysis and critical review. *J Am Acad Child Adolesc Psychiatry* 36(8): 1036-1045.

Gedda, L., Milani-Comparetti, M., & D'Alessandro, E. (1976). "Equivocalness" and other empirical methods in zygosity assessment. *Acta Genet Med Gemellol (Roma)* 25:117-120.

Gjone, H., Stevenson, J., & Sundet, J.M. (1996). Genetic influence on parent-reported attention-related problems in a Norwegian general population twin sample. *J Am Acad Child Adolesc Psychiatry* 35(5): 588-596.

Goldsmith, H.H. (1991). A zygosity questionnaire for young twins: a research note. *Behav Genet* 21(3): 257-269.

Goodman, R. (1995). The relationship between normal variation in IQ and common childhood psychopathology: a clinical study. *Eur Child Adolesc Psychiatry* 4(3): 187-196.

Goodman, R., Simonoff, E., & Stevenson, J. (1995). The impact of child IQ, parent IQ and sibling IQ on child behavioural deviance scores. *J Child Psychol Psychiatry* 36(3): 409-425.

Graham, P. & Rutter, M. (1973). Psychiatric disorder in the young adolescent: a follow-up study. *Proc R Soc Med* 66(12): 1226-1229.

Gustafsson, J.E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence* 8: 179-203.

Guttman, L. (1954). A new approach to factor analysis: The radex. In: Lazarsfeld, P.F. (Ed.). *Mathematical Thinking in the Social Sciences* (pp. 258-348). Glencoe, IL: The Free Press.

Hart, E.L., Lahey, B.B., Loeber, R., Applegate, B., & Frick, P.J. (1995). Developmental change in attention-deficit hyperactivity disorder in boys: a four-year longitudinal study. *J Abnorm Child Psychol* 23(6): 729-749.

Hartman, C.A., Hox, J., Auerbach, J., Erol, N., Fonseca, A.C., Mellenbergh, G.J., Novik, T.S., Oosterlaan, J., Roussos, A.C., Shalev, R.S., Zilber, N., & Sergeant, J.A. (1999). Syndrome dimensions of the child behavior checklist and the teacher

report form: a critical empirical evaluation. *J Child Psychol Psychiatry* 40(7): 1095-1116.

Hartung, C.M., & Widiger, T.A. (1998). Gender differences in the diagnosis of mental disorders: conclusions and controversies of the DSM-IV. *Psychol Bull* 123(3): 260-278.

Hauge, M., Harvald, B., Fischer, M., Gotlieb-Jensen, K., Juel-Nielsen, N., Raebild, I., Shapiro, R., & Videbech, T. (1968). The Danish twin register. *Acta Genet Med Gemellol (Roma)* 17(2): 315-332.

Hay, D.A. & O'Brien, P.J. (1983). The La Trobe Twin Study: a genetic approach to the structure and development of cognition in twin children. *Child Dev* 54(2): 317-330.

Heck, A. (1997). *Introduction to MAPLE* (2nd ed.). New York: Springer-Verlag.

Helgeson, V.S. (2002). *The Psychology of Gender.* New Jersey, Upper Saddle River: Pearson Education, Inc.

Hewitt, J.K. & Heath, A.C. (1988). A note on computing the chi-square noncentrality parameter for power analyses. *Behav Genet* 18(1): 105-108.

Hewitt, J.K., Eaves, L.J., Neale, M.C., & Meyer, J.M. (1988). Resolving causes of developmental continuity or "tracking." I. Longitudinal twin studies during growth. *Behav Genet* 18(2): 133-151.

Hinshaw, S.P. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: causal relationships and underlying mechanisms. *Psychol Bull* 111(1): 127-155.

Horn, J.L. (1991). Comments on issues in factorial invariance. In: Collings, L.M., & Horn, J.L. (Eds.). *Best Methods for the Analysis of Change* (pp. 114-125). Washington: American Psychological Association.

Horn, J.L., & McArdle, J.J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Exp Aging Res* 18: 117-144.

Huberty, C.J. (1994). *Applied Discriminant Analysis.* New York: Wiley.

Hudziak, J.J., Rudiger, L.P., Neale, M.C., Heath, A.C., & Todd, R.D. (2000). A twin study of inattentive, aggressive, and anxious/depressed behaviors. *J Am Acad Child Adolesc Psychiatry* 39(4): 469-476.

Hudziak, J.J. (2001). The role of phenotypes (diagnoses) in genetic studies of attention-deficit/hyperactivity disorder and related child psychopathology. *Child Adolesc Psychiatr Clin N Am* 10(2): 279-97, viii.

Hudziak, J.J. (2002). The importance of phenotype definition in genetic studies of child psychopathology. In: Helzer, J.E., Hudziak, J.J. (Eds.). *Defining psychopathology in the 21st century: DSM-V and beyond.* American Psychiatric Publishing Group, Inc.

Hudziak, J.J., Heath, A.C., Madden, P.A.F., Reich, W., Bucholz, K.K., Slutske, W.S., Bierut, L.J., Neuman, R.J., & Todd, R.D. The genetic analysis of parental reports of DSM-IV. *J Am Acad Child Adolesc Psychiatry* (in review).

Jablon, S., Neel, J.V., Gershowitz, H., & Atkinson, G.F. (1967). The NAS-NRC twin panel: methods of construction of the panel, zygosity diagnosis, and proposed use. *Am J Hum Genet* 19(2): 133-161.

Jackson, R.W., Snieder, H., Davis, H., & Treiber, F.A. (2001). Determination of twin zygosity: a comparison of DNA with various questionnaire indices. *Twin Res* 4(1): 12-18.

Jacobs, N., Rijsdijk, F., Derom, C., Danckaerts, M., Thiery, E., Derom, R., Vlietinck, R., & Van Os, J. (2003). Child psychopathology and lower cognitive ability: a general population twin study of the causes of association. *Mol Psychiatry* 7(4): 368-374.

James, A., & Taylor, E. (1990). Sex differences in the hyperkinetic syndrome of childhood. *J Child Psychol Psychiatry* 31(3): 437-446.

Jeffreys, A.J., Wilson, V., & Thein, S.L. (1985). Individual-specific 'fingerprints' of human DNA. *Nature* 316(6023): 76-79.

Jensen, A. R. (1998). *The g Factor: The Science of Mental Ability.* Westport, Connecticut: Preager Publishers.

Jöreskog, K.D. & Sörbom, D. (1993). *New features in PRELIS 2.* Chicago: Scientific Software International, Inc.

Joseph, J. (2000). Not in their genes: A critical view of the genetics of attention-deficit hyperactivity disorder. *Dev Rev* 20(4): 539-567.

Juan-Espinosa, M., Garcia, L.F., Colom, R., & Abad, F.J. (2000).Testing the age related differentiation hypothesis through the Wechsler's scales. *Pers Individ Dif* 29: 1069-1075.

Kadesjö, C., Kadesjö, B., Hägglöf, B., & Gillberg, C. ((2001). ADHD in Swedish 3- to 7-year-old children. *J Am Acad Child Adolesc Psychiatry* 40(9): 1021-1028.

Kasius, M.C., Ferdinand, R.F., Van den Berg, H., & Verhulst, F.C. ((1997). Associations between different diagnostic approaches for child and adolescent psychopathology. *J Child Psychol Psychiatry* 38(6): 625-632.

Kasriel, J. & Eaves, L. (1976). The zygosity of twins: further evidence on the agreement between diagnosis by blood groups and written questionnaires. *J Biosoc Sci* 8(3): 263-266.

Keenan, K. & Wakschlag, L.S. (2000). More than the terrible twos: The nature and severity of behavior problems in clinic-refered preschool children. *J Abnorm Child Psychol* 28: 33-46.

Kenneth, S. & Kendler, K.S. (2001). Twin Studies of Psychiatric Illness. *Arch Gen Psychiatry* 58: 1005-1014.

Kendler, K.S. (1993). Twin studies of psychiatric illness. Current status and future directions. *Arch Gen Psychiatry* 50(11): 905-915.

Kenny, D.A. (1979). *Correlation and Causality*. New York: Wiley.

Kessler, R.C., Little, R.J., & Groves, R.M. (1995). Advances in strategies for minimizing and adjusting for survey nonresponse. *Epidemiol Rev* 17(1): 192-204.

Kindermann, T.A. (1993). Natural peer groups as contexts for individual development-the case of children motivation in school. *Dev Psychol* 29(6): 970-977.

King, M.-C., Friedman, G.D., Lattanzio, D., Rodgers, G., Lewis, A.M., Dupuy, M.E., & Williams, H. (1980). Diagnosis of twin zygosity by self-assessment and by genetic analysis. *Acta Genet Med Gemellol (Roma)* 29(2): 121-126.

Klein, R.G., & Mannuzza, S. (1991). Long-term outcome of hyperactive children: a review. *J Am Acad Child Adolesc Psychiatry* 30(3): 383-387.

Kline, P. (1991). *Intelligence: The Psychometric View*. London: Routledge.

Koch, H.L. (1966). *Twins and Twin Relations*. Chicago: University of Chicago Press.

Koot, H.M. (1993). *Problem Behavior in Dutch Pre-Schoolers* [PhD thesis]. Rotterdam, The Netherlands: Erasmus University.

Koot, H.M., Van den Oord, E.J.C.G., Verhulst, F.C., & Boomsma, D.I. (1997). Behavioral and emotional problems in young preschoolers: cross-cultural testing of the validity of the Child Behavior Checklist/2-3. *J Abnorm Child Psychol* 25(3):183-196.

Kowaleski-Jones, L. & Duncan, G.J. (1999). The structure of achievement and behavior across middle childhood. *Child Dev* 70(4): 930-943.

Kuntsi, J., Gayan, J., & Stevenson, J. (2000). Parents' and teachers' ratings of problem behaviours in children: genetic and contrast effects. *Twin Res* 3(4): 251-258.

Kuntsi, J. & Stevenson, J. (2000). Hyperactivity in children: a focus on genetic research and psychological theories. *Clin Child Fam Psychol Rev* 3(1): 1-23.

LaBuda, M.C., DeFries, J.C., & Fulker, D.W. (1987). Genetic and environmental covariance structures among WISC-R subtests: A twin study. *Intelligence* 11: 233-244.

Levy, F., Hay D.A., McStephen, M., Wood, C., & Waldman, I. (1997). Attention-deficit hyperactivity disorder: a category or a continuum? Genetic analysis of a large-scale twin study. *J Am Acad Child Adolesc Psychiatry* 36(6): 737-744.

Little, R.J.A. & Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. New York: Wiley.

Loehlin, J.C. (1992). *Latent Variable Models: An Introduction to Factor, Path, and Structural Analysis*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Luo, D., Petrill, S.A., & Thompson, L.A. (1994). An exploration of genetic g: hierarchical factor analysis of cognitive data from the Western Reserve Twin Project. *Intelligence* 18: 335-347.

Magnus, P., Berg, K., & Nance, W.E. (1983). Predicting zygosity in Norwegian twin pairs born 1915-1960. *Clin Genet* 24(2): 103-112.

Martin, N.G. & Eaves L.J. (1977). The genetical analysis of covariance structure. *Heredity* 38(1): 79-95.

Martin, N.G., Eaves L.J., Kearsey, M.J., & Davies, P. (1978). The power of the classical twin study. *Heredity* 40(1): 97-116.

Martin, N.G. & Martin, P.G. (1975). The inheritance of scholastic abilities in a sample of twins. I. Ascertainment of the sample and diagnosis of zygosity. *Ann Hum Genet* 39: 213-218.

Martin, N., Scourfield, J., & McGuffin, P. (2002). Observer effects and heritability of childhood attention-deficit hyperactivity disorder symptoms. *Br J Psychiatry* 180: 260-265.

Mathews, J.D., Martin, N.G., & Gibson, J.B. (1981). Zygosity and placentation in twins. *Med J Aust* 2(10): 515.

Mathiesen, K.S. & Sanson, A. (2000). Dimensions of early childhood behavior problems: stability and predictors of change from 18 to 30 months. *J Abnorm Child Psychol* 28(1): 15-31.

McArdle, J.J. (1986). Latent variable growth within behavior genetic models. *Behav Genet* 16(1): 163-200.

McArdle, J.J. & Goldsmith, H.H. (1990). Alternative common factor models for multivariate biometric analyses. *Behav Genet* 20(5): 569-608.

McCall, R.B., Appelbaum, M.I., & Hogarty, P.S. (1973). Developmental changes in mental performance. *Monogr Soc Res Child Dev* 38(3): 1-84.

McCartney, K., Harris, M.J., & Bernieri, F. (1990). Growing up and growing apart: A developmental meta-analysis of twin studies. *Psychol Bull* 107: 226-237.

McGee, R., Williams, S., Bradshaw, J., Chapel, J.L., Robins, A., & Silva, P.A. (1985). The Rutter scale for completion by teachers: factor structure and relationships with cognitive abilities and family adversity for a sample of New Zealand children. *J Child Psychol Psychiatry* 26(5): 727-739.

McGee, R. & Feehan, M. (1991). Are girls with problems of attention under-recognized? *J Psychopathol Behav Assess* 13: 187-198.

McGee, R., Partridge, F., Williams, S., & Silva, P.A. (1991). A twelve year follow-up of preschool hyperactive children. *J Am Acad Child Adolesc Psychiatry* 30: 224-232.

McGee, R., Williams, S., & Feehan, M. (1992). Attention deficit disorder and age of onset of problem behaviors. *J Abnorm Child Psychol* 20(5): 487-502.

McGue, M., Bouchard, T.J. Jr., Iacono, W.G., & Lykken, D.T. (1993). Behavioral genetics of cognitive ability: A life-span perspective. In: Plomin, R., & McClearn, G.E. (Eds.). *Nature, Nurture, & Psychology* (pp. 59-76). Washington: American Psychological Association.

Mesman, J., & Koot, H.M. (2001). Early preschool predictors of preadolescent internalizing and externalizing DSM-IV diagnoses. *J Am Acad Child Adolesc Psychiatry* 40(9): 1029-1036.

Meyer, J.M., Silberg, J.L., Eaves, L.J., Maes, H.H., Simonoff, E., Pickles, A., Rutter, M.L., & Hewitt, J.K. (1999). Variable age of gene expression: implications for developmental genetic models. In: LaBuda, M.C., & Grigorenko, E.L. (Eds.), *On the Way to Individuality: Current Methodological Issues in Behavorial Genetics* (pp.23-52). Commack: Nova Science Publishers, Inc.

Milberger, S., Biederman, J., Faraone, S.V., Chen, L., & Jones, J. (1996). Is maternal smoking during pregnancy a risk factor for attention deficit hyperactivity disorder in children? *Am J Psychiatry* 153(9): 1138-1142.

Milberger, S., Biederman, J., Faraone, S.V., Guite, J., & Tsuang, M.T. (1997). Pregnancy, delivery and infancy complications and attention deficit hyperactivity disorder: issues of gene-environment interaction. *Biol Psychiatry* 41(1): 65-75.

Milberger, S., Biederman, J., Faraone, S.V., & Jones, J. (1998). Further evidence of an association between maternal smoking during pregnancy and attention deficit hyperactivity disorder: findings from a high-risk sample of siblings. *J Clin Child Psychol* 27(3): 352-358.

Moffitt, T.E. (1990). Juvenile delinquency and attention deficit disorder: boys' developmental trajectories from age 3 to age 15. *Child Dev* 61(3): 893-910.

Nadder, T.S., Silberg, J.L., Eaves, L.J., Maes, H.H., & Meyer J.M. (1998). Genetic effects on ADHD symptomatology in 7- to 13-year-old twins: results from a telephone survey. *Behav Genet* 28(2): 83-99.

Nadder, T.S., Silberg, J.L., Rutter, M., Maes, H.H., & Eaves, L.J. (2001). Comparison of multiple measures of ADHD symptomatology: a multivariate genetic analysis. *J Child Psychol Psychiatry* 42(4): 475-486.

Neale, M.C., & Stevenson, J. (1989). Rater bias in the EASI temperament scales: a twin study. *J Pers Soc Psychol* 56(3): 446-455.

Neale, M.C. & Cardon, L.R. (1992). *Methodology for Genetic Studies of Twins and Families.* Dordrecht, The Netherlands: Kluwer Academic.

Neale, M.C. (1997). *Mx: Statistical Modelling* (4th ed.). Richmond, VA: Dept. of Psychiatry, Virginia Commonwealth University.

Neale, M.C., Boker, S.M., Xie, G., & Maes, H.H. (1999). *Mx: Statistical Modeling.* (5th ed.). Richmond, VA: Dept. of Psychiatry, Virginia Commonwealth University.

Nichols, R.C. & Bilbro, W.C. Jr. (1966). The diagnosis of twin zygosity. *Acta Genet Stat Med* 16(3): 265-275.

Offord, D. R., Boyle, M. H., Szatmari, P., Rae-Grant, N. I., Links, P. S., Cadman, D. T., Byles, J. A., Crawford, J. W., Blum, H. M., Byrne, C., Thomas, H., & Woodward, C. (1987). Ontario Child Health Study: Six-month prevalence of disorder and rates of service utilization. *Arch Gen Psychiatry* 44: 832-836.

Ooki, S., Yamada, K., Asaka, A., & Hayakawa, K. (1990). Zygosity diagnosis of twins by questionnaire. *Acta Genet Med Gemellol (Roma)* 39(1): 109-115.

Ooki, S., Yamada, K., & Asaka, A. (1993). Zygosity diagnosis of twins by questionnaire for twins' mothers. *Acta Genet Med Gemellol (Roma)* 42(1): 17-22.

Orlebeke, J.F., Knol, D.L., Boomsma, D.I., & Verhulst, F.C. (1998). Frequency of parental report of problem behavior in children decreases with increasing maternal age at delivery. *Psychol Rep* 82(2): 395-404.

Panel on Discriminant Analysis, Classification and Clustering (1989). Discriminant analysis and clustering. *Stat Sci* 4: 34-69.

Patrick, C.L. (2000). Genetic and environmental influences on the development of cognitive abilities: Evidence from the field of developmental behavior genetics. *J Sch Psychol* 38: 79-108.

Patterson, G.R. (1984). Siblings: Fellow travelers in coercive family processess. In: Blanchard, R.J. (Ed.). *Advances in the Study of Aggression* (pp. 174-213). New York: Academic Press.

Pedersen, N.L., Plomin, R., & McClearn, G.E. (1994). Is there G beyond g? (Is there genetic influence on specific cognitive abilities independent of genetic influence on general cognitive ability). *Intelligence* 18: 133-143.

Peeters, H., Van Gestel, S., Vlietinck, R., Derom, C., & Derom, R (1998). Validation of a telephone zygosity questionnaire in twins of known zygosity. *Behav Genet* 28(3): 159-163.

Petrill, S.A., Luo, D.S., Thompson, L.A., & Detterman D.K. (1996). The independent prediction of general intelligence by elementary cognitive tasks: genetic and environmental influences. *Behav Genet* 26(2): 135-147.

Petrill, S.A., Plomin, R., Berg, S., Johansson, B., Pedersen, N.L., Ahern, F.M., & McClearn, G.E. (1998). The genetic and environmental relationship between general and specific cognitive abilities in twins age 80 and older. *Psychol Sci* 9: 183-189.

Petronis, A. (2001). Human morbid genetics revisited: relevance of epigenetics. *Trends Genet* 17(3): 142-146.

Pineda, D., Ardila, A., Rosselli, M., Arias, B.E., Henao, G.C., Gomez, L.F., Mejia, S.E., & Miranda, M.L. (1999). Prevalence of attention-deficit/hyperactivity disorder symptoms in 4- to 17-year-old children in the general population. *J Abnorm Child Psychol* 27(6): 455-462.

Plomin, R. (1983). Developmental behavioral genetics. *Child Dev* 54(2): 253-259.

Plomin, R. (1986). *Development, Genetics, and Psychology*. Hillsdale, New Jersey: Erlbaum Associates.

Plomin, R. (1994). *Genetics and Experience: The Interplay between Nature and Nurture*. [Sage series on individual differences and development 6]. Thousand Oaks, California: SAGE Publications.

Plomin, R. & Daniels, D. (1987). Why are children in the same family so different from one another? *Behav Brain Sci* 10: 1-60.

Plomin, R., & Defries, J.C. (1981). Multivariate behavioral genetics and development: Twin studies. In: Gedda, L., Parisi, P. & Nance, W.E. (Eds.), *Twin Research 3: Part B* (pp. 25-33). New York: Alan R. Liss.

Plomin, R. & DeFries, J.C. (1985). *Origins of Individual Differences in Infancy: The Colorado Adoption Project*. Orlando, Florida: Academic Press.

Plomin, R., DeFries, J.C., & Loehlin, J.C. (1977). Genotype-environment interaction and correlation in the analysis of human behavior. *Psychol Bull* 84(2): 309-322.

Plomin, R., Defries, J.C., McClearn, G.E., & McGuffin, P. (2000). *Behavioral Genetics*. (4th ed.). New York: Worth Publishers.

Plomin, R., Pedersen, N.L., Lichtenstein, P., & McClearn, G.E. (1994). Variability and stability in cognitive abilities are largely genetic later in life. *Behav Genet* 24(3): 207-215.

Plomin, R., Petrill, S.A., & Cutting, A.L. (1996). What genetic research on intelligence tells us about the environment. *J Biosoc Sci* 28(4): 587-606.

Plomin, R., & Vandenberg, S.G. (1980). An analysis of Koch's (1966) Primary Mental Abilities Test data for 5- to 7-year-old twins. *Behav Genet* 10(4): 409-412.

Posthuma, D. & Boomsma, D.I. (2000). A note on the statistical power in extended twin designs. *Behav Genet* 30(2): 147-158.

Posthuma, D., De Geus, E.J.C., & Boomsma, D.I. (2002). Genetic contributions to anatomical, behavioral, and neurophysiological indices of cognition. In: Plomin, R., Defries, J.C., Craig, I.W., & McGuffin, P. (Eds.), *Behavioral Genetics in the Postgenomic Era*. Washington DC: American Psychological Associaton.

Price, T.S., Eley, T.C., Dale, P.S., Stevenson, J., Saudino, K., & Plomin, R. (2000). Genetic and environmental covariation between verbal and nonverbal cognitive development in infancy. *Child Dev* 71(4): 948-959.

Price, T.S., Petrill, S.A., Dale, P.S., Eley, T.C., & Plomin, R. Modularity and g: A longitudinal multivariate genetic analysis of early cognitive development. (submitted)

Price, T.S., Simonoff, E., Waldman, I., Asherson, P., Curran, S., & Plomin, R. What is stable about hyperactive behaviors in pre-school children is genetic: Implications for molecular genetic studies. (submitted)

Reed, T., Norton, J.A. Jr., & Christian, J.C. (1977). Sources of information for discriminating MZ and DZ twins by dermatoglyphic patterns. *Acta Genet Med Gemellol (Roma)* 26(1): 83-86.

Reinert, G. (1970). Comparative factor analytic studies of intelligence throughout the human life-span. In: Goulet, L.R., & Baltes, P.B. (Eds.). *Life-span Developmental Psychology* (p.p. 467-484). New York: Academic Press.

Reznick, J.S., Corley R., & Robinson, J. (1997). A longitudinal twin study of intelligence in the second year. *Monogr Soc Res Child Dev* 62(1): i-154.

Rhee, S.H., Waldman, I.D., Hay, D.A., & Levy, F. (1999). Sex differences in genetic and environmental influences on DSM-III-R attention-deficit/hyperactivity disorder. *J.Abnorm Psychol* 108(1): 24-41.

Rice, T., Carey, G., Fulker, D.W., & Defries, J.C. (1988). Multivariate path analysis of specific cognitive abilities in the Colorado Adoption Project: Conditional path model of assortative mating. *Behav Genet* 19(2): 195-207.

Rietveld, M.J.H., Hudziak, J.J., Bartels, M., Van Beijsterveldt, C.E.M., & Boomsma, D.I. (2003a). Heritability of attention problems in children: I. cross-sectional results from a study of twins, age 3-12 years. *Am J Med Genet* 117B(1): 102-113.

Rietveld, M.J.H., Posthuma, D., Dolan, C.V., & Boomsma, D.I. (2003b). ADHD: sibling interaction or dominance, an evaluation of statistical power. *Behav Genet* 33: 247-255.

Rietveld, M.J.H, Van Baal, G.C.M., Dolan, C.V., & Boomsma, D.I. (2000a). Genetic factor analyses of specific cognitive abilities in 5-year-old Dutch children. *Behav Genet* 30(1): 29-40.

Rietveld, M.J.H., Van der Valk, J.C., Bongers, I.L., Stroet, T.M., Slagboom, P.E., & Boomsma, D.I. (2000b). Zygosity diagnosis in young twins by parental report. *Twin Res* 3(3): 134-141.

Rijsdijk, F.V., Vernon, P.A., & Boomsma, D.I. (2002). Application of hierarchical genetic models to Raven and WAIS subtests: a Dutch twin study. *Behav Genet* 32(3): 199-210.

Ross, D.M., & Ross, S.A. (1982). *Hyperactivity. Current Issues, Research, and Theory*. New York/Cambridge: Cambridge University Press.

Rowe, D.C., Rodgers, J.L., & Meseck-Bushey, S. (1992). Sibling delinquency and the family environment: shared and unshared influences. *Child Dev* 63(1): 59-67.

Rutter, M., & Garmezy, N. (1983). Developmental psychopathology. In: Mussen, P.H. (Ed.), *Handbook of Child Psychology* (4th ed.). New York: John Wiley & Sons.

Rutter, M., Giller, H., & Hagell, A. (1998). *Antisocial Behavior by Young People*. New York/Cambridge: Cambridge University Press.

Sadker, M., Sadker, D., & Klein, S. (1991). The issue of gender in elementary and secondary education. In: Grant, G. (ed.). *Review of Research in Education*. Washington DC: American Educational Research Association.

Sammalisto, L. (1961).The determination of zygosity in a study of Finnish twins. *Acta Genet* 11: 251-264.

Sarna, S., Kaprio, J., Sistonen, P., & Koskenvuo, M. (1978). Diagnosis of twin zygosity by mailed questionnaire. *Hum Hered* 28(4): 241-254.

Sarna, S., & Kaprio, J. (1980). Use of multiple logistic analysis in twin zygosity diagnosis. *Hum Hered* 30(2): 71-80.

Saudino, K.J., Cherny, S.S., & Plomin, R. (2000). Parent ratings of temperament in twins: explaining the 'too low' DZ correlations. *Twin Res* 3(4): 224-233.

Scarr, S. & McCartney, K. (1983). How people make their own environments: a theory of genotype greater than environment effects. *Child Dev* 54(2): 424-435.

Schaie, K.W., Willis, S.L., Jay, G., & Chipuer, H. (1989). Structural invariance of cognitive abilities across the adult life span: A cross-sectional study. *Dev Psychol* 25: 652-552.

Schaie, K.W. (1994). The course of adult intellectual development. *Am Psychol* 49(4): 304-313.

Schaie, K.W., Maitland, S.B., Willis, S.L., & Intrieri, R.C. (1998). Longitudinal invariance of adult psychometric ability factor structures across 7 years. *Psychol Aging* 13(1): 8-20.

Schmitz, S., Fulker, D.W., & Mrazek, D.A. (1995). Problem behavior in early and middle childhood: an initial behavior genetic analysis. *J Child Psychol Psychiatry* 36(8): 1443-1458.

Schmitz, S., Corley, R.P., Hewitt, J.K., Zahn-Waxler, C., Emde, R.N., & Defries, J.C. Attention problems and emotionality at ages 7 and 12. (submitted)

Schoenfeldt, L.F. (1969). A comparison of two analytic procedures for estimating twin zygosity. *Hum Hered* 19(4): 343-353.

Segal, N.L. (1985). Monozygotic and dizygotic twins: A comparative analysis of mental ability profiles. *Child Dev* 56: 1051-1058.

Segal, N.L. (2000). Virtual twins: new findings on within-family environmental influences on intelligence. *J Edu Psychol* 92: 442-448.

Shaw, D.S., Owens, E.B., Giovannelli, J., & Winslow, E.B. (2001). Infant and toddler pathways leading to early externalizing disorders. *J Am Acad Child Adolesc Psychiatry* 40(1): 36-43.

Sherman, D.K., McGue, M.K., & Iacono, W.G. (1997a). Twin concordance for attention deficit hyperactivity disorder: a comparison of teachers' and mothers' reports. *Am J Psychiatry* 154(4): 532-535.

Sherman, D.K., Iacono, W.G., & McGue, M.K. (1997b). Attention-deficit hyperactivity disorder dimensions: a twin study of inattention and impulsivity-hyperactivity. *J Am Acad Child Adolesc Psychiatry* 36(6): 745-753.

Siemens, H.W. (1927).The diagnosis of identity in twins. *J Hered* 18: 201-209.

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. *Psychol Med* 28(4): 825-837.

Smith, S., & Penrose, L.S. (1955). Monozygotic and dizygotic twin diagnosis. *Ann Hum Genet* 19: 273-289.

Spearman, C.E. (1904). 'General intelligence' objectively determined and measured. *Am J Psychol* 15: 201-293.

Spearman, C. (1927).*The Abilities of Man*. London: MacMillan.

Spinath, F.M., & Angleitner, A. (1998). Contrast effects in Buss and Plomin's EAS questionnaire. *Pers Individ Dif* 25: 947-963.

Spitz, E., Moutier, R., Reed, T., Busnel, M.C., Marchaland, C., Roubertoux, P.L., & Carlier, M. (1996). Comparative diagnoses of twin zygosity by SSLP variant analysis, questionnaire, and dermatoglyphic analysis. *Behav Genet* 26(1): 55-63.

SPSS Inc. (1997). *SPSS Base 7.5 for Windows User's Guide*. Chicago, Illinois: SPSS Inc.

Steingard, R., Biederman, J., Doyle, A., & Sprich-Buckminster, S. (1992). Psychiatric comorbidity in attention deficit disorder: impact on the interpretation of Child Behavior Checklist results. *J Am Acad Child Adolesc Psychiatry* 31(3): 449-454.

Sternberg, R.J. & Grigorenko, E.L. (2002). *The General Factor of Intelligence, how general is it?* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Stevenson, J. (1992). Evidence for a genetic etiology in hyperactivity in children. *Behav Genet* 22: 227-344.

Sutton, H.E., Clark, P.J., & Schull, W.J. (1955). The use of multi-allele genetic characters in the diagnosis of twin zygosity. *Am J Hum Genet* 7: 180-188.

Szatmari, P., Offord, D.R., & Boyle, M.H. (1989). Ontario Child Health Study: prevalence of attention deficit disorder with hyperactivity. *J Child Psychol Psychiatry* 30(2): 219-230.

Tambs, K., Sundet, J.M., & Magnus, P. (1986). Genetic and environmental contributions to the covariation between the Wechsler Adult Intelligence Scale (WAIS) subtests: a study of twins. *Behav Genet* 16(4): 475-491.

Thapar, A. (2002). Attention-Deficit Hyperactivity Disorder: New genetic findings, new directions. In: Plomin, R., Defries, J.C., Craig, I.W., & McGuffin, P. (Eds.), *Behavioral Genetics in the Postgenomic Era.* Washington DC: American Psychological Association.

Thapar, A., Harrington, R., Ross, K., & McGuffin, P. (2000). Does the definition of ADHD affect heritability? *J Am Acad Child Adolesc Psychiatry* 39(12): 1528-1536.

Thapar, A., Hervas, A., & McGuffin, P. (1995). Childhood hyperactivity scores are highly heritable and show sibling competition effects: twin study evidence. *Behav Genet* 25(6) : 537-544.

Thompson, L.A. (1993). Genetic contributions to intellectual development in infancy and childhood. In: Vernon, P.A. (Ed.), *Biological Approaches to the Study of Human Intelligence* (pp. 95-138). Norwood, New Jersey: Ablex.

Thurstone, L.L. (1938). Primary Mental Abilities. *Psychometric Monographs,* No. 1.

Thurstone, L.L. & Thurstone, T.G. (1941). Factorial studies of intelligence. *Psychometric Monographs,* No 2.

Todd, R.D., Sitdhiraksa, N., Reich, W., Ji, T.H.C., Joyner, C.A., Heath, A.C., & Neuman, R.J. (2002). Discrimination of DSM-IV and latent class attention-deficit / hyperactivity disorder subtypes by educational and cognitive performance in a population-based sample of child and adolescent twins. *J Am Acad Child Adolesc Psychiatry* 41(7): 820-828.

Torgersen, S. (1979). The determination of twin zygosity by means of a mailed questionnaire. *Acta Genet Med Gemellol* 28: 225-236.

Turkheimer, E., & Waldron, M. (2000). Nonshared environment: a theoretical, methodological, and quantitative review. *Psychol Bull* 126(1): 78-108.

Van Baal, G.C.M., De Geus, E.J., & Boomsma, D.I. (1996). Genetic architecture of EEG power spectra in early life. *Electroencephalogr Clin Neurophysiol* 98(6): 502-514.

Van Baal, G.C.M. (1997). *A genetic perspective on the developing brain* [PhD thesis]. Amsterdam, The Netherlands: Vrije Universiteit.

Van Beijsterveldt, C.E.M., Verhulst, F.C., Molenaar, P.C.M., & Boomsma, D.I. The genetic basis of problem behavior in 5-year-old Dutch twin pairs. (submitted).

Van den Oord, E.J.C.G. (1993). *Genetic Study of Problem Behaviors in Children* [PhD thesis]. Rotterdam, The Netherlands: Erasmus University.

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. *J Abnorm Psychol* 105(3): 349-357.

Van der Valk, J.C. (2001). *The Genetic and Environmental Contributions to Children's Problem Behaviors: A Developmental Approach* [PhD thesis]. Rotterdam, The Netherlands: Erasmus University.

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

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

Van Dijk, B.A., Boomsma, D.I., & De Man, A.J. (1996). Blood group chimerism in human multiple births is not rare. *Am J Med Genet* 61(3): 264-268.

Velez, C.N., Johnson, J., & Cohen, P. (1989). A longitudinal analysis of selected risk factors for childhood psychopathology. *J Am Acad Child Adolesc Psychiatry* 28(6): 861-864.

Verhulst, F.C. & Van der Ende, J. (1995). The eight-year stability of problem behavior in an epidemiologic sample. *Pediatr Res* 38(4): 612-617.

Verhulst, F.C., Koot, H.M., & Van der Ende, J. (1994). Differential predictive value of parents and teachers reports of childrens problem behaviors – a longitudinal study. *J Abnorm Child Psychol* 22: 531-546.

Verhulst, F.C., Van der Ende, J., & Koot, H.M. (1996). *Handleiding voor de CBCL/4-18.* Rotterdam,Nederland: Afd. Kinder- en Jeugdpsychiatrie/Erasmus Universiteit.

Vernon, P.A. (1965). Ability factors and environmental influences. *Am Psychol* 20: 723-733.

Vernon, P.E. (1976). Development of intelligence. In: Hamilton, V., & Vernon, M.D. (Eds.). *The Development of Cognitive Processes* (pp. 507-541). London: Academic Press Inc.

Vlietinck, R. (1986). *Determination of the Zygosity of Twins* [PhD thesis]. Leuven, Belgium: Katholieke Universiteit Leuven.

Weiss, G. & Hechtman, L. (1993). *Hyperactive Children grown up. ADHD in Children, Adolescents and Adults.* (2nd ed.). New York: Guilford Press.

Werdelin, I., & Stjernberg, G. (1995). Age differences in factorial structure: A study of the "differentiation hypothesis". *Interdisciplinaria* 12: 79-97.

Werry, J.S., Elkind, G.S., & Reeves, J.C. (1987). Attention deficit, conduct, oppositional, and anxiety disorders in children: III. Laboratory differences. *J Abnorm Child Psychol* 15(3): 409-428.

Wilens, T.E., Biederman, J., Brown, S., Tanguay, S., Monuteaux, M.C., Blake, C., & Spencer, T.J. (2002). Psychiatric comorbidity and functioning in clinically referred preschool children and school-age youths with ADHD. *J Am Acad Child Adolesc Psychiatry* 41(3): 262-268.

Wilson, R.S. (1975). Twins: Patterns of cognitive development as measured on the Wechsler Preschool and Primary Scale of Intelligence. *Dev Psychol* 11: 126-134.

Wilson, R.S. (1983). The Louisville Twin Study: developmental synchronies in behavior. *Child Dev* 54(2): 298-316.

Wohlwill, J.F. (1973). *The Study of Behavior Development.* New York: Academic Press.

Woody-Ramsey, J. & Miller, P.H. (1988). The facilitation of selective attention in preschoolers. *Child Dev* 59(6): 1497-1503.

Wothke, W. (2000). Longitudinal and multigroup modeling with missing data. In: Little, T.D., Schnabel, K.U., & Baumert, J. (Eds.), *Modeling Longitudinal and Multilevel Data: Practical Issues, applied Approaches and Specific Examples.* Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

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

Samenvatting

De structuur en stabiliteit van Intelligentie

Ruim 200 tweelingparen hebben meegewerkt aan een onderzoek naar Intelligentie op 5- , 7-, en 10- jarige leeftijd. Bij deze groep zijn cognitieve vaardigheden gemeten met de verkorte versie van de Revisie Amsterdamse Kinder Intelligentie Test (RAKIT; Bleichrodt et al., 1984). Deze versie bestaat uit zes deeltesten die een beroep doen op zowel verbale als non-verbale cognitieve vaardigheden. De samenhang tussen de afzonderlijke deeltesten, d.w.z. de covariantie structuur van Intelligentie, is onderzocht met exploratieve en confirmatieve factor analyses. De samenhang tussen de deeltesten over de tijd heen, met andere woorden de stabiliteit van Intelligentie, is onderzocht middels longitudinale analyses. Door middel van de simultane analyse van de gegevens van één- en twee-eiige tweelingen, is de bijdrage van genen en omgeving aan de fenotypische structuur en stabiliteit van Intelligentie in kaart gebracht. De resultaten van de (longitudinale) factor analyses worden hier kort besproken, zowel op het fenotypische niveau als op het omgevings- en genetische niveau.

Leeftijdsspecifieke factor analyses

Een twee factor model met gecorreleerde groepsfactoren geeft een adequate beschrijving van de fenotypische covariantie structuur van de RAKIT. De deeltesten Woordbetekenis, Namen Leren, en Ideeënproductie laden op een verbale factor en de deeltesten Exclusie, Schijven, en Verborgen Figuren laden op een non-verbale factor. Dit factor model blijkt een goede en coherente beschrijving van de data te geven op 5-, 7-, en 10- jarige leeftijd. De correlatie tussen de groepsfactoren ligt tussen .40 en .49, afhankelijk van de leeftijd. Genetische factor analyses resulteren in verschillende factor modellen voor de genetische en de omgevingsinvloeden. Net als bij de fenotypische covariantie matrix, geeft een twee factor model een goede beschrijving van de additief genetische covariantie matrix. Een één factor model blijkt een goede beschrijving te geven van de gedeelde omgevingscovariantie matrix. Tenslotte blijkt de unieke omgevingscovariantie matrix diagonaal te zijn. Een model met zes deeltest-specifieke (ongecorreleerde) factoren geeft een adequate beschrijving van de unieke omgevingscovariantie matrix. De structuur van deze genetische en omgevings- factor modellen is identiek op de drie verschillende leeftijden.

Longitudinale factor analyses

De genetische factorstructuur en de unieke en gedeelde omgevingsfactorstructuren zijn gehandhaafd bij de analyse van de longitudinale data. De correlatie tussen de genetische groepsfactoren is .25 op 5- jarige leeftijd, .28 op 7- jarige leeftijd, en .30 op 10- jarige leeftijd. De bijdragen van omgevings- en genetische factoren aan de fenotypische deeltest-variantie op elke leeftijd staan vermeld in Tabel I. Gedurende de onderzochte periode

blijven de genetische bijdragen aan de deeltest-variantie constant (Schijven, Verborgen Figuren, Ideeënproduktie) of de genetische bijdragen nemen toe (Exclusie, Woordbetekenis, Namen Leren). Met uitzondering van Woordbetekenis verklaren de gedeelde omgevingsfactoren een klein deel van fenotypische deeltest variantie. Afhankelijk van deeltest en leeftijd verklaart de unieke omgevingsfactoren tussen 22% en 64% van de totale deeltest variantie.

Tabel I: Percentages verklaarde variantie door additief genetische (a^2), gedeelde omgevings- (c^2), en unieke omgevingsfactoren (e^2).

Deeltest	leeftijd	a^2	c^2	e^2
Exclusie	5	38	19	43
	7	40	5	55
	10	58	7	35
Schijven	5	46	3	51
	7	58	5	37
	10	44	4	52
Verborgen Figuren	5	48	11	41
	7	37	14	49
	10	48	8	44
Woordbetekenis	5	21	39	40
	7	19	17	64
	10	42	32	26
Namen Leren	5	53	17	30
	7	39	32	29
	10	74	4	22
Ideeënproduktie	5	58	5	37
	7	58	4	38
	10	60	4	36

De prestaties op de deeltesten vertonen een redelijke mate van stabiliteit van 5 naar 7 jaar, en van 7 naar 10 jaar. Uitgedrukt in een correlatie varieert deze stabiliteit tussen .35 en .55. Er bestaat geen consistent verschil in de mate van stabiliteit van de verbale en non-verbale deeltesten, of tussen de twee leeftijdsintervallen. De genetische analyses van de longitudinale dataset suggereren dat fenotypische stabiliteit voornamelijk is toe te schrijven aan genetische stabiliteit. De gedeelde omgevingsinvloeden spelen een bescheiden rol in het verklaren van stabiliteit. Bovendien neemt deze rol af met het ouder worden van de kinderen. De gemiddelde bijdrage tot de totale deeltest-covariantie van 5 naar 7 jaar is 26%, van 7 naar 10 jaar is de gemiddelde bijdrage 20%.

De longitudinale analyses zijn ook aangewend om de differentiatie hypothese te onderzoeken. De differentiatie hypothese veronderstelt een daling van intercorrelaties tussen cognitieve vaardigheden naar mate kinderen ouder worden. De resultaten van de fenotypische en genetische analyses ondersteunen deze hypothese niet.

Sekse verschillen

Gedurende de leeftijdsspecifieke en longitudinale analyses is onderzocht of er sprake is van sekse verschillen in de covariantie structuur van de testscores op fenotypisch of latent niveau. Wat betreft de structuur van Intelligentie, stabiliteit van Intelligentie, en de bijdrage van genen en omgeving zijn er geen verschillen tussen jongens en meisjes gevonden.

De ontwikkeling van Overactief gedrag en Aandachtsproblemen

De ouders van tweelingen hebben de gehele Child Behavior CheckList (CBCL) ingevuld rond de 3de, 7de, 10de, en 12de verjaardag van hun kinderen. De schaal Overactief gedrag is gebruikt bij de beoordeling van de 3- jarigen (CBCL/2-3; Achenbach, 1992). Deze schaal heeft betrekking op vijf gedragingen, zoals 'kan zich niet concentreren, kan niet lang de aandacht bij iets houden', 'kan niet stil zitten, onrustig, of overactief', en 'verandert snel van activiteit'. Op 7-, 10-, en 12- jarige leeftijd zijn de tweelingen beoordeeld aan de hand van de schaal Aandachtsproblemen (CBCL/4-18; Achenbach, 1991a). Deze schaal omvat elf gedragingen, zoals 'kan zich niet concentreren, kan niet lang de aandacht bij iets houden', 'gedraagt zich te jong voor haar/zijn leeftijd', en 'slechte schoolresultaten'. De 3- en 7- jarigen zijn afkomstig uit de geboortecohorten 1986 tot en met 1993. Gezien het doorlopende karakter van het onderzoek zijn voor de dataverzameling onder 10- en 12- jarigen minder cohorten beschikbaar. Deze kinderen komen uit de cohorten 1986 tot halverwege 1991.

In dit onderzoek is gebruik gemaakt van de vragenlijsten die door de moeder zijn ingevuld. In totaal zijn 11938 vragenlijsten verzameld van 3- jarigen, 10657 van 7- jarigen, 6192 van 10- jarigen, en 3124 van 12- jarigen. Er is gekeken naar gemiddelde verschillen tussen jongens en meisjes, de prevalentie van ernstige vormen van Overactief gedrag en Aandachtsproblemen, en de stabiliteit van deze gedragingen. Daarnaast is gekeken naar de genetische en omgevingsbijdrage tot individuele verschillen in Overactiviteit en Aandachtsproblemen zoals deze voorkomen op elke leeftijd, en naar de genetische en omgevingsbijdragen tot de stabiliteit van gedrag. Voordat deze aspecten konden worden onderzocht, moesten eerst enkele methodologische kwesties behandeld worden. Deze betreffen de bepaling van de zygositeit van tweelingen, de gevolgen voor schattingen van

erfelijkheid in de aanwezigheid van een 'contrast effect' in de data, en het statistische vermogen (*power*) om een dergelijk effect te onderscheiden. De resultaten met betrekking tot deze methodologische kwesties worden eerst gepresenteerd, gevolgd door een samenvatting van de resultaten voor Overactiviteit en Aandachtsproblemen.

Zygositeit

Het bepalen van de zygositeit van een tweelingpaar in het onderzoek naar Overactiviteit en Aandachtsproblemen gebeurt op basis van een korte vragenlijst. De vragen worden beantwoord door beide ouders en hebben betrekking op de uiterlijke overeenkomsten van de tweelingen, en frequentie waarmee familieleden de tweelingen verwarren. Een paar honderd tweelingparen heeft deelgenomen aan bloed- en DNA onderzoek om de zygositeit te bepalen. De nauwkeurigheid waarmee de zygositeit is bepaald op grond van de vragenlijst wordt bepaald door de vragenlijst-, en bloed- en DNA gegevens te gebruiken in een discriminant analyse. In het onderzoek naar de bepaling van de zygositeit stonden twee vragen centraal: 1. Neemt de nauwkeurigheid van het bepalen van de zygositeit toe naar mate de tweelingen ouder worden? 2. Verschilt de nauwkeurigheid van de bepaling tussen moeders en vaders? Beide vragen kunnen ontkennend beantwoord worden. Met een precisie van bijna 95% worden zowel jongere tweelingen als oudere tweelingen correct geclassificeerd als één- of twee-eiig. Moeders en vaders beantwoorden de vragen over gelijkheid en verwisseling van de tweelingen op gelijke wijze. Dit gelijke antwoordpatroon leidt tot een gelijke mate van nauwkeurigheid.

Contrast effect

In tweelingstudies naar ADHD (attention-deficit hyperactivity disorder; DSM-IV) en aanverwante fenotypes, wordt veelal een zogenaamd contrast effect gevonden. Het contrast effect suggereert een negatieve interactie tussen het gedrag van de tweelingen. Met andere woorden, terwijl het ene kind veel impulsief en overactief gedrag vertoont, vertoont het andere kind weinig impulsief en overactief gedrag. De huidige interpretatie van het contrast effect is dat het niet refereert aan het werkelijke gedrag van de tweelingen, maar aan de beoordeling van dit gedrag door de ouders. Ouders benadrukken de fenotypische verschillen tussen de kinderen. Het statistische gevolg van een dergelijke beoordeling is dat de correlaties van de tweelingen een patroon vertonen dat de aanwezigheid suggereert van genetische dominantie. Zowel een negatieve contrast effect als genetische dominantie invloeden leiden tot DZ correlaties die relatief veel lager liggen dan MZ correlaties. Dit maakt het bijzonder moeilijk om in de genetische analyses onderscheid te maken tussen de aanwezigheid van genetische dominantie of een contrast effect. Een andere statistische consequentie van een negatief contrast effect is een daling in de variantie, die vooral op-

treedt bij eeneiige tweelingen. Een dergelijke daling treedt echter niet op bij genetische dominantie invloeden. Deze consequentie biedt dus de mogelijkheid om onderscheid te maken tussen een contrast effect en genetische dominantie. Er is vervolgens onderzocht hoe groot het statistische onderscheidingsvermogen is om een negatief contrast effect te detecteren, in aanwezigheid van additieve genetische, genetische dominantie, en unieke omgevings- variantie. Hierbij is ook gekeken of de bijdragen van genen en omgeving nog wel op de juiste grootte worden geschat. Tenslotte is een oplossing aangedragen om een mogelijk gebrek aan power te compenseren.

Het aantonen van een klein contrast effect (< -.15) in kleine tot middelgrote tweelingonderzoeken (< 1500 paar) is niet mogelijk in aanwezigheid van genetische dominantie. Het onderscheidingsvermogen neemt toe naar mate het effect en de omvang van de steekproef toenemen. Indien een contrast effect niet gespecificeerd wordt terwijl deze wél aanwezig is in de data, leidt dit tot een grote afwijking in de schattingen van de genetische en omgevingsbijdragen. Het eerst weglaten van de genetische dominantie uit het model en het vervolgens toetsen van de aanwezigheid van een contrast effect komt zowel het onderscheidingsvermogen als de nauwkeurigheid van de parameter schattingen ten goede. Het toevoegen van ongerelateerde broers en zussen aan het tweelingdesign geeft een sterke verbetering van het vermogen om een contrast effect (≥ -.10) te onderscheiden. Omdat hiervoor relatief weinig broers en zussen nodig zijn (> 100 paren) is deze optie met name interessant in de kleinere tweelingstudies.

Overactief gedrag en Aandachtsproblemen, sekse verschillen en prevalentie

Vergeleken met meisjes scoren jongens hoger op de schaal Overactief gedrag op 3- jarige leeftijd en op de schaal Aandachtsproblemen op 7-, 10-, en 12- jarige leeftijd. De gemiddelde scores van jongens en meisjes zijn vergelijkbaar met de normgegevens. Tussen de 2% en 7% van de totale groep tweelingen vertoont in ernstige mate Overactief gedrag en Aandachtsproblemen. Deze kinderen komen mogelijk in aanmerking voor de klinische diagnose ADHD. Vergelijkbaar met de gemiddelde scores, zijn het met name de jongens die zeer ernstige Aandachtsproblemen vertonen.

Overactief gedrag en Aandachtsproblemen, genetische analyses

De genetische bijdragen tot individuele verschillen in Overactief gedrag en Aandachtsproblemen zijn groot. In Tabel II staat een overzicht uitgesplitst naar leeftijd en sekse. De totale genetische bijdragen liggen tussen de 70% en 74%. Het resterende deel van de variantie in Overactief gedrag en Aandachtsproblemen kan worden toegeschreven aan unieke omgevingsinvloeden. Gedeelde omgevingsinvloeden zijn afwezig. Er bestaat geen consistent patroon in de verschillen tussen leeftijd en tussen sekse.

Tabel II: Percentages verklaarde variantie door additief genetische (a^2), dominant genetische (d^2), en uniek omgevings- effecten (e^2). In elke cel staan twee schattingen, de eerste is voor jongens en de tweede is voor meisjes.

Probleemschaal	leeftijd	a^2	d^2	e^2
Overactief gedrag *	3	50% - 41%	22% - 33%	28% - 26%
Aandachtsproblemen	7	33% - 57%	39% - 16%	28% - 27%
Aandachtsproblemen	10	41% - 48%	31% - 25%	28% - 27%
Aandachtsproblemen	12	40% - 54%	30% - 18%	30% - 28%

Note: * op 3- jarige leeftijd lijkt een 'contrast effect' aanwezig.

Overactief gedrag en Aandachtsproblemen, stabiliteit

De overgang van Overactief gedrag op 3- jarige leeftijd naar Aandachtsproblemen op latere leeftijd wordt gekenmerkt door een bescheiden mate van stabiliteit. De correlatie tussen Overactief gedrag en Aandachtsproblemen op 7- jarige leeftijd is .40. De correlatie tussen Aandachtsproblemen op 7-, 10-, en 12- jarige leeftijd is ongeveer .70. Er is geen verschil tussen jongens en meisjes wat betreft de stabiliteit van de individuele verschillen in dit gedrag. Het zijn vooral de genetische invloeden die bijdragen tot deze stabiliteit. Deze invloeden verklaren tussen 76% en 92% van de fenotypische correlaties tussen leeftijden. Consistente verschillen in stabiliteit tussen leeftijdsintervallen en tussen sekse zijn afwezig. Bij jongens en meisjes zijn het dezelfde genetische factoren die bijdragen tot individuele verschillen in Overactief gedrag en Aandachtsproblemen.

Intelligentie én Overactiviteit / Aandachtsproblemen

Van ongeveer 150 tweelingparen zijn de complete gegevens van beide onderzoeken, Intelligentie en Overactiviteit / Aandachtsproblemen, gecombineerd. Op basis van de bestaande literatuur wordt een negatief verband tussen (verbale) Intelligentie en probleemgedrag verwacht. Dit negatieve verband is ook hier gevonden. Echter de gevonden correlaties, variërend rond -.15, zijn beperkt tot de nonverbale Intelligentie deeltesten. Leeftijd heeft geen effect op de sterkte van dit verband. Bijvoorbeeld, Overactiviteit op het 3de jaar correleert even sterk met nonverbale Intelligentie op het 5de jaar als op het 10de jaar. Er is geen verschil in deze resultaten tussen jongens en meisjes. Het verband tussen nonverbale Intelligentie en Overactiviteit / Aandachtsproblemen is lineair van aard. Met ander woorden, de daling in nonverbale IQ score is niet beperkt tot de kinderen die ernstige vormen

van Overactief gedrag en Aandachtsproblemen vertonen maar is waarneembaar langs de gehele verdeling van Overactief gedrag en Aandachtsproblemen.

Tenslotte is gekeken naar het verband tussen Aandachtsproblemen en prestatie op taal- en rekenkundige schooltaken op 7- en 10- jarige leeftijd. Ook hier wordt een negatief verband gevonden (variërend rond -.30), en wel veel sterker in vergelijking tot het verband nonverbale Intelligentie en Aandachtsproblemen (variërend rond -.15).

Conclusies

Op het fenotypische niveau bestaat de structuur van Intelligentie uit een clustering van enerzijds verbale en anderzijds nonverbale deeltesten. Deze tweedeling in deeltesten is terug te zien in de genetische factor structuur. De fenotypische correlatie tussen verbale en nonverbale deeltest scores wordt deels verklaard door genetische invloeden en deels door gedeelde omgevingsinvloeden. De unieke omgeving draagt bij aan de deeltest-specifieke variantie, variantie die niet wordt gedeeld met andere deeltesten.

De ontwikkeling van Intelligentie kenmerkt zich door een redelijke mate van stabiliteit op deeltest niveau. Deze stabiliteit is voornamelijk toe te schrijven aan het gegeven dat dezelfde genetische factoren gedurende de ontwikkeling van 5 naar 10 jaar 'actief' zijn. Deze genetische inbreng lijkt toe te nemen met het ouder worden. Van het 5de naar het 7de jaar verklaart de gedeelde omgeving een klein deel van de stabiliteit maar de invloed van de gedeelde omgeving is dan al klein en neemt nog verder af naar het 10de jaar. Een mogelijke verklaring voor het patroon van genetische en omgevingsbijdragen tot de stabiliteit is dat kinderen steeds meer een omgeving opzoeken die aansluit bij hun genetische aanleg. Een dergelijke correlatie tussen genotype en omgeving leidt tot een verhoogde schatting van de genetische bijdrage in tweelingonderzoek.

De differentiatie hypothese veronderstelt dat de samenhang tussen cognitieve vaardigheden afneemt met het ouder worden. Gezien de vele studies is deze hypothese in theoretische zin blijkbaar erg aantrekkelijk. Echter duidelijke empirische ondersteuning ontbreekt. Op fenotypisch, genetisch, en omgevingsniveau is geen ondersteuning voor deze hypothese gevonden. Het is mogelijk dat differentiatie al voor het 5de jaar heeft plaatsgevonden, of pas na het 10de jaar plaatsvindt.

Individuele verschillen in Overactief gedrag en Aandachtsproblemen worden in sterke mate bepaald door verschillen in genetische invloeden. Op 3-, 7-, 10-, en 12- jarige leeftijd wordt de erfelijkheid van deze fenotypen geschat op bijna 75%. De gedeelde omgeving speelt geen rol in het verklaren van individuele verschillen. Het kan niet uitgesloten worden dat er een interactie plaatsvindt tussen gedeelde omgevingsfactoren en genetische

factoren. Kinderen uit hetzelfde gezin kunnen onderling sterk verschillen in hun reactie op bepaalde gezinsinvloeden. Deze vorm van interactie valt onder de genetische invloeden, en verklaart een deel van de erfelijkheid van het gedrag. De overgang van Overactief gedrag op 3- jarige leeftijd naar Aandachtsproblemen op latere leeftijd wordt gekenmerkt door een relatief lage stabiliteit. Het verschil in meetinstrument op 3- en 7- jarige leeftijd (Overactiviteit versus Aandachtsproblemen) lijkt geen verklaring voor de relatief lage stabiliteit. Dit blijkt uit het feit dat de stabiliteit ook laag is voor de items die hetzelfde blijven tussen leeftijd 3 en 7 jaar. De stabiliteit van Aandachtsproblemen van het 7de jaar naar het 10de en 12de jaar is veel groter, en wordt bijna volledig verklaard door genetische invloeden.

Jongens en meisjes verschillen in de mate waarin ze Overactief gedrag en Aandachtsproblemen vertonen. Gemiddelde scores van jongens zijn op elke leeftijd hoger dan de scores van meisjes. Dit geldt ook voor zeer ernstige Overactiviteit / Aandachtsproblemen.

Kinderen met Aandachtsproblemen presteren veel slechter in de schoolsituatie (taal- en rekenkundige taken) dan op basis van hun nonverbale Intelligentie score is te verwachten. Dit suggereert dat kinderen met Aandachtsproblemen niet goed kunnen functioneren op school en dat een specifieke, educatieve aanpak gewenst is.

Dankwoord

Enkele maanden geleden bevond ik mij op een feestje met hoofdzakelijk onbekende, oudere dames en heren. Eén van deze heren vroeg wat mijn promotie onderzoek inhield. Na mijn gebruikelijke lekenpraatje over genen, omgeving en tweelingonderzoek vroeg hij of ik in mijn modelfitting analyses wel rekening had gehouden met een mogelijke interactie tussen genen en omgeving. Bijgekomen van mijn verbazing kreeg ik verdere uitleg van deze heer; in de jaren '70 was hij gepromoveerd op een dierstudie naar voeding bij varkens. Na enige discussie ontstond mijn vraag wat volgens hem, naast de ontwikkeling van statistische mogelijkheden, de grootste vooruitgang is geweest op (gedrags-)genetisch gebied in de tijdsperiode tussen zijn en mijn proefschrift. "Welnu, de varkens zijn mensen geworden", aldus de vriendelijke man.

Bij deze mijn dank aan de mensen: de tweelingen en hun ouders. Het (herhaaldelijk) invullen van de vragenlijsten door zoveel moeders en vaders hebben ervoor gezorgd dat het onderzoek naar overactief gedrag en aandachtsproblemen gebaseerd is op een indrukwekkend grote steekproef. De beste herinneringen tijdens mijn aio-schap bewaar ik aan de dataverzameling in het onderzoek naar intelligentie. Drie van de vier tweelingparen zijn getest in hun eigen huis. De ontvangst met koffie en taart, en het afscheid met tekeningen en zoenen maakten deze logistiek zware periode ook de leukste periode. Ilja Bongers en Saskia vd Berg waren niet alleen enthousiaste maar ook hele goede stagiaires.

De begeleiding van prof. Boomsma en dr. Dolan hebben geleid tot het proefschrift zoals het voor u ligt. Dorret, ondanks jouw drukte is de drempel van je deur altijd laag gebleven voor hulpvragen van aio's/oio's. Ook waardeer ik de prioriteit die je aan een 'bijna-klaar-manuscript' hebt gegeven (en dat zijn er nogal wat geweest). Waar ik je zeker voor wil bedanken is dat je de mogelijkheid hebt geboden om mij verder te bekwamen in het onderwijs. Conor, jouw vaardigheid om moeilijke thema's op een nivo van appels en peren uit te leggen is fantastisch. Tijdens je bliksembezoeken waren niet alleen alle statistische hindernissen genomen, ook raakte ik weer overtuigd van de 'leukheid' van mijn eigen onderzoek. Bedankt hiervoor.

Dr. van Beijsterveldt, Prof. Bleichrodt, dr. Hudziak, prof. Sergeant, dr. Swaab, en prof. Verhulst, de leden van de leescommissie wil ik hartelijk bedanken voor de tijd en moeite die ze in mijn proefschrift hebben gestoken. Jim, your clinical point of view has made an improvement to the papers on attention problems. It was a joy to work with you.

En dan de administratieve kant van mijn werk: Thérèse en haar dames (Monique, Gerda, Ingrid, en Yvon) wil ik bedanken voor de nauwkeurigheid en snelheid waarmee zij het vragenlijst gedeelte van mijn onderzoek hebben ondersteund. Toos, wat was ik blij met jouw aanstelling als datamanager van het jonge tweelingencohort! Het samenstellen van mijn longitudinale dataset werd zowaar een plezierige bezigheid. Natascha en Christina, de dames van het secretariaat bedank ik voor hun algemene hulp en altijd informatieve kletspraat. Natascha, jouw hulp bij het opzetten van mijn referentielijst kan ik alle aio's/oio's aanraden!

En dan de dames-collega's. Gedurende mijn zeven aio-jaren (ik schrok er zelf ook van!) is een aantal vertrokken, gebleven en bijgekomen. Ik durf geen lijst te geven met namen uit angst toch iemand te vergeten. Jullie gaven mij de motivatie om het werk af te maken. Zonder koffie, discussie, wetenschappelijke tips, nationale, internationale en lokale roddels kwam ik de werkdag niet door. Bedankt voor jullie gezelligheid!

In het bijzonder wil ik de paranimfen Meike en Tinca noemen. Meike, toen je mijn kamergenoot werd was ik diep onder de indruk van de doelmatigheid van jouw aanpak. Ik had stille hoop dat ik daar wat van kon leren. Helaas, mijn obsessie om elk mogelijk detail te onderzoeken en te beschrijven is steeds gebleven. Jouw obsessie om never nooit niet van de grote lijn af te wijken bleef ook. Dat maakt ons een wetenschappelijke 'perfect match'. Maar zoals je in je eigen dankwoord schreef, onze weddenschappen vormen toch de kern van de vriendschap. Zolang deze een bourgondisch karakter hebben en ik blijf winnen zullen we elkaar nog vaak zien. Tinca, ook een zeeuwse boerentriene! Jie ben een biezonder aangeneame meid en 'is toch eel erg da me noe nog ma een stuitje samen werke. Ma ons ebbe zo vee tegeare in wat voorbie is da me mekoare nog vee tegen 't lief zulle lope. Da's dan ma gelukkig ee?

Mijn familie wil ik bedanken voor hun aansporende opmerkingen "wanneer ik nou eindelijk ga werken voor mijn geld", en "of mijn scriptie nog niet klaar is". Beste familie, het moment is eindelijk aangebroken. Ik ben net zo blij en trots als jullie.

En toen nog de vrienden. Babette, Josita, Celina, Charlene, en Tonnie, jullie bedank ik voor de niet aflatende emotionele steun en belangstelling. Vier van jullie werk(t)en als therapeut binnen de zorgsector, ik heb het blijkbaar nodig.... Dearest Charlene, you are a great friend. Wonen in Almere had kunnen voldoen aan alle vooroordelen als we het niet zo getroffen hadden met de vele 'buren'. Ik bedank jullie allemaal voor de vele kopjes koffie (of was het toch sherry?), uitstapjes, feestjes, en gedeelde zorg voor de hele horde kinderen. "It takes a village to raise a child" had ons motto kunnen zijn.

Eduard, Guus, en Louis. Jullie bijdrage aan mijn relativeringsvermogen was bijzonder groot de afgelopen jaren, bijna té groot. Toch kan ik het de collega's aanraden: een gezinsleven zorgt ervoor dat je de stress van een proefschrift minder kent. Er zijn wel belangrijker zaken in het leven. Lieve mannen, bedankt.

Marjolein, april 2003