CHAPTER 3

THE DETECTION OF GENOTYPE-ENVIRONMENT INTERACTION IN LONGITUDINAL GENETIC MODELS

Peter C.M. Molenaar, Dorret I. Boomsma and Conor V. Dolan

University of Amsterdam and Free University, Amsterdam, The Netherlands

INTRODUCTION

Recently a number of models have been suggested for the analysis of longitudinal twin data (see Loehlin, 1991 for a brief overview). These models can be used to study the genetic and environmental contributions to the variances and covariances of phenotypic measures at a single measurement occasion and to the stability and change of individual differences over time. McArdle (1986) applied the latent growth curve model to longitudinal twin data, Boomsma and Molenaar (1987) proposed modeling development by means of autoregressive or simplex models and Eaves, Long and Heath (1986) suggested a model combining both autoregressive and confirmatory factor analysis models (see also Boomsma, Martin and Molenaar, 1989; Hewitt, Eaves, Neale and Meyer, 1988; Loehlin, Horn and Willerman, 1989; Molenaar, Boomsma and Dolan, 1991). Although these models incorporate different developmental hypotheses, they share the basic assumption that the phenotypic deviation score at each measurement occasion is related to latent genetic and environmental deviation scores according to a simple linear (or additive) model:

$$P = hG + eE + cC \qquad (1)$$

where P stands for the phenotypic deviation score of an individual (subject subscript is discarded) and G, E and C stand for the genetic, unshared and shared (common to family members) environmental deviation scores. The coefficients h, e and c are standardized regression coefficients (factor loadings). All variables are expressed as deviations from

the mean so that their expected values: E[P] = E[E] = E[G] = E[C] = 0. If the unobserved or latent genetic and environmental factors are standardized to have unit variance then the variance of the phenotype is equal to :

$$V_P = h^2 + e^2 + c^2 \tag{2}$$

In this linear measurement model, possible non-linear effects arising through the interaction among any combination of G, E and C are assumed to be absent. The absence of genotype-environment interaction implies that an environmental effect (or 'treatment') has the same effect regardless of the genotype of the individual upon whom it is imposed (Neale and Cardon, 1992, page 22). Plomin, DeFries and McClearn (1990) define genotype-environment (GxE) interaction as follows: 'Genotype-environment interaction denotes an interaction in the statistical sense of a conditional relationship: The effect of environmental factors depend on the genotype' (page 250).

Statistical analysis of genotype-environment interaction can be conducted by means of various approaches, including analysis of variance and regression analysis, in combination with direct or indirect measures of the environment (Neale and Cardon, 1992, Chapter 11; Eaves, 1984; Freeman, 1973) or of the genotype (Martin, Eaves and Heath, 1987) or both (Plomin, 1986, Chapter 5). If measures of either environment or genotype are not available -i.e. in the majority of the quantitative genetic studies of metric human phenotypes- the detection of genotype-environment interaction is more difficult. One test for genotype by environment suggested by Jinks and Fulker (1970) involves examing the association between the means and standard deviations of MZ twin pairs. For MZ twins reared together, the difference between members of a twin pair reflects the magnitude of environmental differences within families and the sum of their scores reflects genetic (or environmental) differences between families. For MZ twins reared together, this test thus detects interactions between genotype and individual-specific environmental factors. For MZ twins reared apart, interactions with all postnatal environmental effects are included in the test.

Another approach to test for genotype-environment interaction when measures of the environment or the genotype are not available, has recently been suggested and is based on the analysis of the higher-order moments of genetic and environmental factor scores. Molenaar and Boomsma (1987) and Molenaar, Boomsma, Neeleman and Dolan (1990) have shown that the effects of certain types of interaction cannot be detected at the level of second-order moments (i.e. variances and covariances), but do lead to specific values of the third- and fourth-order moments (i.e. skewness and kurtosis) of genetic and environmental factor scores. These methods do not require measurements of the environment or the genotype, but require multiple indicators of the phenotype for the calculation of factor scores (Boomsma, Molenaar and Orlebeke, 1990) and the estimation of the higher-order moments of these factor scores.

The object of this paper is to study the effects of genotype-environment interaction in the context of longitudinal data using the genetic simplex model (Boomsma and Molenaar, 1987). To explore the effects of interaction in the standard genetic model based on second-order statistics we assume that measurements are available at three time points on three congeneric tests (i.e. tests that are, except for errors in measurement,

perfectly correlated). At each occasion a common additive genetic and a common unshared environmental factor account for the covariance between the observations. The variance specific to each variable at each occasion is error variance. The simplex part of the model consists of the covariance between factors across time being attributable to the first-order autoregressions of the common additive genetic and unshared environmental factors. In a first-order autoregressive process latent factors are only influenced by the latent factors directly preceeding them, so that the partial correlation between factors at time points i and k $r_{ik.j} = 0$, whenever i<j<k. We choose this somewhat simple model because our main objective is a theoretical exploration of the consequences of genotype-environment interaction in developmental data. In the illustrative simulations we will look at the possibility that part of the developmental process consists of genotype x unshared environment (G x E) interaction at each time-point. This simple scenario gives rise to several interesting and unexpected results of genotype-environment interaction. In order to arrive at a somewhat self-contained chapter, we first present a summary of results from simulation studies on the estimation of individual factor scores in genetic covariance structure models and on the detection of different types of interaction using these factor scores. Next we introduce the analysis of longitudinal twin data by means of the genetic simplex model and discuss the estimation of factor scores in a longitudinal twin design. In the last part, we consider the detection of genotype-environment interaction in this longitudinal model.

Summary of Results Relating to the Calculation of Individual Genetic and Environmental Factor Scores and Statistical Tests of Genotype-Environment Interaction

We consider the multivariate version of Equation 1 (Martin and Eaves, 1977) in matrix notation (discarding the subject index):

$$P = hG + eE + cC + \varepsilon \tag{3}$$

where $P = (P_1,...,P_p)'$ denotes a random p-dimensional vector of zero means phenotypes and ' denotes transposition. The vectors G, E, and C represent common (i.e. to the components of P) genetic, within family (unshared) and between family (shared) environmental factors with p-dimensional loadings h, e, and c. The unique part in each phenotype, e, is a random p-dimensional vector composed of influences unique to each phenotype P_j, j=1,...,p. The common factors G, E, and C are taken to be mutually uncorrelated normally distributed variables with zero mean and unit variance. On an individual basis G, E and C represent individual factor scores, i.e. an individual's genetic, unshared environmental and shared environmental deviation scores. Let Ψ denote the correlation matrix of the 3 common factors G, E, and C, and let Θ denote the covariance matrix of the unique components ε. The (p x p) covariance matrix, Σ_P, of P is then:

$$\Sigma_P = \Lambda\Psi\Lambda' + \Theta \tag{4}$$

Within individuals, the 3 x 3 matrix Ψ is an identity matrix and the p x 3 matrix Λ contains the factor loadings: Λ = [h, e, c]. Assuming that all parameters are known (i.e. the elements of Λ and Θ), we may calculate an individuals deviation scores on the common factors in a number of distinct ways (see McDonald and Burr, 1967; Lawley and Maxwell, 1971; Saris, De Pijper and Mulder, 1978). We limit the discussion to the regression method for estimating factor scores which is investigated extensively in Boomsma, Molenaar and Orlebeke (1990) for the genetic common factor model and in Boomsma, Molenaar and Dolan (1991) for the genetic simplex model. Factors scores of individual i are calculated according to the regression method by multiplying the observations P_i for individual i by a weight matrix W. W is constructed in such a way that the sum of squares of the difference between estimated and true factor scores is minimized:

$$\eta_i = W P_i \tag{5}$$

$$W = \Psi\Lambda' \Sigma_P^{-1} \tag{6}$$

where $\eta_i' = [G_i, E_i, C_i]'$. With data from genetically related individuals the weight matrix W can be extended to include the observations from family members in the construction of the factor scores. Standard errors of these scores can also be calculated so that confidence intervals can be constructed around the individual genetic and environmental deviation scores (Boomsma, Molenaar and Orlebeke, 1990). As an example of this technique, Table 1 shows correlations between simulated factor scores and factor scores estimated by the regression method for MZ and DZ twins. The simulated data consisted of a 5-variate factor model with low unique variance for all 5 variables. From simulations such as these, it is clear that individual factor scores may be estimated reliably, given that a good-fitting multivariate model has been obtained on the original observations that supplies the parameter estimates of Λ and Θ needed to construct the weight matrix for the computation of factor scores.

Table 1. Correlations of simulated and estimated factor scores for MZ and DZ twins (decimal point omitted) using the regression method. Simulations were based on 100 MZ and 100 DZ twin pairs, using a 5-variate common factor model for G, E, and C. Unique variances were between 5 and 11% for each variable.

	MZ			DZ		
	E(G)	E(E)	E(C)	E(G)	E(E)	E(C)
G	910*	081	211	884*	213	100
E	020	879*	096	295*	773*	369*
C	124	100	936*	045	341*	897*

* $p < .001$

If each of the zero-mean unit-variance common factors G, C or E were replaced by interaction terms, the presence of such interactions would not show up in the second-order moments of the data. However, estimates of the higher-order moments of the

distribution of factor scores would reveal genotype-environment interactions in the data even when both the genotype and the environment are not directly observed. Molenaar and Boomsma (1987) considered the case in which interactions give rise to additional factors in the standard genetic covariance model. For instance, replacement of a second genetic factor G by G*=GxE or a second shared environmental factor C by C*=CxE gives rise to a model with two common unshared environmental factors. Replacement of C by C*=CxG gives rise to a second common genetic factor. Molenaar and Boomsma used a factor rotation method devised by McDonald (1967) to distinguish between a true second environmental (or genetic) common factor and a second common factor attributable to the mentioned forms of interaction. The test is based on a special rotation of the multiple within family environmental (or genetic) factors that maximizes the third-order moments of factor scores in order to determine whether the second factor that behaves as an additional E (or G) factor really is an interaction factor.

Table 2. Characteristics of second and fourth-order moments of latent interaction factors

Model	interaction	characteristics of 2nd-order statistics	characteristics of 4th-order moments
G*,E,C	G*=GxC	var(G*)=1 cor(G*,E)=0 cor(G*,C)=0 cor(G*,G*)mz = 1 cor(G*,G*)dz=0.5 G* behaves like G	$E[G^4]=9$ and $E[G^2C^2]=3$ when G*=GxC, whereas $E[G^4]=3$ and $E[G^2C^2]=1$ when G* is not an interaction factor
G,E*,C	E*=ExG	var(E*)=1 cor(G,E*)=0 cor(C,E*)=0 cor(E*E*)mz = 0 cor(E*E*)dz = 0 E* behaves like E	$E[E^4]=9$ and $E[E^2G^2]=3$ when E*=ExG, whereas $E[E^4]=3$ and $E[G^2E^2]=1$ when E* is not an interaction factor
G,E*,C	E*=ExC	var(E*)=1 cor(G,E*)=0 cor(C,E*)=0 cor(E*E*)mz = 0 cor(E*E*)dz = 0 E* behaves like E	$E[E^4]=9$ and $E[E^2C^2]=3$ when E*=ExC, whereas $E[E^4]=3$ and $E[C^2E^2]=1$ when E* is not an interaction factor

Molenaar, Boomsma, Neeleman and Dolan (1990) presented a more general approach to the test of genotype-environment interactions underlying multivariate observations. This test can be applied to covariance structure models in which only one common genetic, one common within-family and one common between-family environmental factor (or a subset of these factors) is present and in which the factors that make up the interaction term are not present as separate factors in the model. Detection of interaction

in this case requires a test of fourth-order moments of factor scores. Table 2 contains a summary of the characteristics of the second- and fourth-order moments such interaction factors. The interactions described in Table 2 cannot be detected at the level of second-order statistics, but the fourth-order moment expressions can serve as simple tests for the presence of of various forms of interaction. Simulation studies suggest that application of these expectations to estimated factor scores makes it possible to detect genotype-environment interaction even with realistic sample sizes. The test can also be generalized to the case where genes that control sensitivity to the environment are different from genes that control average response over all environments (Mather and Jinks, 1982; Eaves, 1984; Martin, Eaves and Heath, 1987).

SPECIFICATION OF THE GENETIC SIMPLEX MODEL

In this section we introduce the genetic simplex model and discuss the estimation of longitudinal individual genetic and environmental profiles. The stage will then be set for a consideration of interactions of the type shown in Table 2 for developmental data. The basic model that we employ is shown in Figure 1. We assume that MZ and DZ scores are available consisting of three indicators of a phenotype at three measurement occasions. At each time point, individual differences are determined by an additive genetic and an unshared environmental factor. The longitudinal part of the model is represented by autoregressions of the latent variables on earlier latent variables. Let P denote the 6 dimensional vector of phenotypic congeneric deviation scores of twin pair i at occasion t. The phenotypic vector is related to the common genetic and unshared environmental factors through a linear measurement model:

$$P_{ti} = \Lambda_t \eta_{ti} + \varepsilon_{ti} \tag{7}$$

where

$$P'_{ti} = [P_{t11}\ P_{t12}\ P_{t13}\ P_{t21}\ P_{t22}\ P_{t23}]_i$$

$$\eta'_{ti} = [G_{t1}\ E_{ti}\ G_{t2}\ E_{t2}]_i$$

$$\varepsilon'_{ti} = [\varepsilon_{t11}\ \varepsilon_{t12}\ \varepsilon_{t13}\ \varepsilon_{t21}\ \varepsilon_{t22}\ \varepsilon_{t23}]_i$$

where i is the twin pair index (i; i=1,...N). The subscripts of the phenotypic variables (P_{tij}),the common latent factors (G_{ti}, E_{ti}) and specific error terms (ε_{tij}) indicate measurement occasion (t; t=1,2,3) and phenotypic variable (j=1,2,3). The matrix of factor loadings at each occasion equals:

$$\Lambda_t = \begin{bmatrix} \Lambda_{gt} & \Lambda_{et} & 0 & 0 \\ 0 & 0 & \Lambda_{gt} & \Lambda_{et} \end{bmatrix}$$

where Λ_{gt} and Λ_{et} are (3 x 1) matrices of genetic and environmental factor loadings. The longitudinal models for G and E are specified as first-order autoregressions:

On the Way to Individuality: Methodological Issues in Behavioral Genetics

CONTRIBUTORS

Dorret I. Boomsma
University of Amsterdam and Free University
Amsterdam, The Netherlands

Thomas J. Bouchard, Jr.
University of Minnesota
Minnesota, USA

Remi J. Cadoret
University of Iowa College of Medicine
Iowa, USA

Alexei A. Chikanian
Yale University
Connecticut, USA
Lebedev Institute for Physics
Moscow, Russia

Hilary Coon
University of Utah School of Medicine
Utah, USA

Conor Dolan
University of Amsterdam and Free University
Amsterdam, The Netherlands

Lindon J. Eaves
Medical College of Virginia
Virginia Commonwealth University
Virginia, USA

Graham G. Giles
The Anti-Cancer Council of Victoria
Victoria, Australia

Elena L. Grigorenko
Yale University
Connecticut, USA
Moscow State University
Moscow, Russia

John K. Hewitt
University of Colorado, Boulder
Colorado, USA

John L. Hopper
The University of Melbourne
Victoria, Australia

Mark A. Jenkins
The University of Melbourne
Victoria, Australia

Michele C. LaBuda
Johns Hopkins Medical Institutions
Maryland, USA

Gregory T. Macaskill
The University of Melbourne
Victoria, Australia

Hermine H. Maes
Medical College of Virginia
Virginia Commonwealth University
Virginia, USA

Millenium Pharmaceuticals

Peter C. M. Molenaar
University of Amsterdam and Free University
Amsterdam, The Netherlands

Nancy Pedersen
Karolinska Institute
Stockholm, Sweden
Pennsylvania State University
Pennsylvania, USA

Andrew Pickles
Institute of Psychiatry
University of London
London, England

Richard Rende
Department of Psychology
Rutgers University
New Jersey, USA

Michael Rutter
Institute of Psychiatry
University of London
London, England

Judy L. Silberg
Medical College of Virginia
Virginia Commonwealth University
Virginia, USA

Emily Simonoff
Institute of Psychiatry
University of London
London, England

Kristian Tambs
National Institute of Public Health
Department of Epidemiology
Oslo, Norway

Edward P. Troughton
University of Iowa College of Medicine
Iowa, USA

George P. Vogler
Pennsylvania State University
Pennsylvania, USA

William R. Yates
University of Iowa College of Medicine
Iowa, USA

$$G_{t+1} = \beta_{gt+1,t}\, G_t + \zeta_{gt+1} \quad (8)$$

$$E_{t+1} = \beta_{et+1,t}\, E_t + \zeta_{et+1} \quad (9)$$

where β_g and β_e represent the regressions of latent factors G and E on the previous genetic and environmental factors and ζ represents a random input term or innovation.

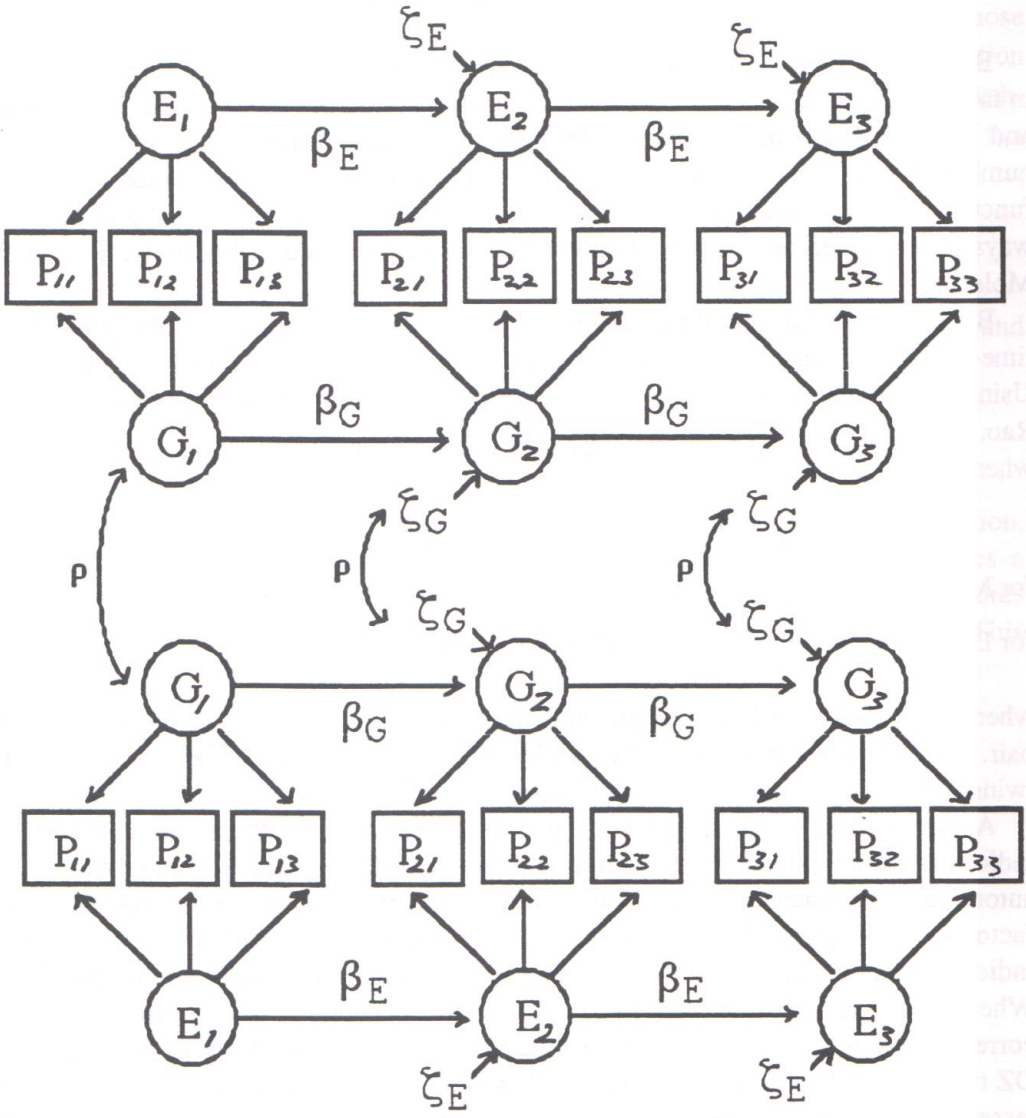

Figure 1. Genetic simplex model: 3 observations (squares) on 3 time-points in MZ and DZ twins pairs. The parameter ρ represents the additive genetic correlation (ρ=1 in MZ and 0.5 in DZ twin pairs). Latent variables G(enotype) and E(nvironment) are represented by circles, the b's represent the influence of a latent variable at an earlier time-point on a latent variable at a later time-point.

The implied covariance structure of the genetic factors is (exactly the same expectations obtain for the environmental factors):

$$\text{var}(G_1) = \text{var}(\zeta_{g1})$$

$$\text{cov}(G_{t+1}, G_t) = \beta_{gt+1,t}\ \text{var}(G_t)$$

$$\text{var}(G_{t+1}) = \beta^2_{gt+1,t}\ \text{var}(G_t) + \text{var}(\zeta_{gt+1})$$

By fitting this model to MZ and DZ covariance matrices, estimates of the parameters in the model can be obtained with standard software packages such as LISREL (Jöreskog and Sörbom, 1988) or Mx (Neale, 1991). To estimate the parameters in the model a number of loss functions can be minimized. Throughout we minimize the likelihood ratio function to obtain maximum likelihood estimates (Neale and Cardon, 1992). Different ways of parameterizing this problem in LISREL are discussed in Boomsma, Martin and Molenaar (1989).

Boomsma, Molenaar and Dolan (1991) have looked at the feasibility of estimating the time-dependent genetic and environmental factor scores in the genetic simplex model. Using a generalization of the regression method described above (Priestley and Subba Rao, 1975; Brown, 1983), factor scores for the i-th twin pair are calculated as: $\eta_i = W\ P_i$, where:

$$W = E[\eta\eta'']\ \Lambda'\ \Sigma^{-1}$$

for MZ $$W = [(I-B)^{-1}\ \Psi_{mz}\ (I-B')^{-1}]\ \Lambda'\ \Sigma^{-1}_{mz} \quad (10)$$

for DZ $$W = [(I-B)^{-1}\ \Psi_{dz}\ (I-B')^{-1}]\ \Lambda'\ \Sigma^{-1}_{dz} \quad (11)$$

where η'_i contains the latent genetic and environmental trajectories of the i-th MZ twin pair. For MZ twins, the genetic trajectories will of course be identical, whereas for DZ twins they will be correlated 0.5 on average at each time-point.

Applying these equations to simulated time-series data with different numbers of indicators for the phenotype at each time point and different genetic and environmental autoregressive parameters, a correlation between the true factor scores and the calculated factor scores of above 0.9 for MZ and DZ twins was obtained when three congeneric indicators for the phenotype were available (Boomsma, Molenaar and Dolan, 1991). When only one indicator was measured for the phenotype at each time point, the correlation between the true and the estimated factor scores was between 0.7 and 0.8 for DZ twins and around 0.8 for MZ twins. However, decomposition of univariate, and to a lesser extent bivariate, time-series yielded estimates of independent G_t and E_t scores that were intercorrelated. These intercorrelations depended somewhat on the difference in size between the genetic and environmental autoregressions, but to obtain independent estimates of individual genetic and non-genetic time-series, at least three measured indicators are needed at each time-point.

INTERACTIONS AND FOURTH-ORDER MOMENTS IN THE GENETIC SIMPLEX MODEL

The model for the latent genetic and unshared environmental trajectories is an autoregression as in equations 8 and 9 ($\eta_t = b_{t,t-1}\,\eta_{t-1} + \zeta_t$). Suppose that both η_{t-1} and ζ_t are the outcome of a multiplicative interaction factor of two standard normal variables where the variables contributing to the innovation η_{t-1} are uncorrelated with those contributing to ζ_t. To test whether η_t is the outcome of a sum of two such interaction terms, we have to derive the fourth-order moment of η_t. We first consider the fourth-order statistics at an arbitrary time point t, where t>1. Expressing the contributions of η_{t-1} and ζ_t to the total variance of η_t as proportions we write:

$$\eta_t = a\eta_{t-1} + b\zeta_t$$

where $a = \sqrt{(\beta^2/(1+\beta^2))}$ and $b = \sqrt{(1-a^2)}$, as $var(\eta_t) = \beta^2_{t,t-1}\ var(\eta_{t-1}) + var(\zeta_t)$ and $var(\eta_{t-1}) = var(\zeta_t) = 1$. The fourth-order moment of η_t then equals:

$$E[\eta_t^4] = 9 - 12a^2 + 12a^4 \quad (12)$$

So we find that $E[\eta_t^4]$ can vary between 9 (a = 1, that is when there is no innovation, or a = 0, that is when b = 0, i.e. no transmission) and 6 (a = 0.7). This implies a dependence of the expected fourth-order moment of the interaction of factor scores $E[\eta_t^4]$ on the value of the innovation ζ_t and the transmission parameter $\beta_{t,t-1}$. This dependency between $E[\eta_t^4]$ and a is pictured in Figure 2.

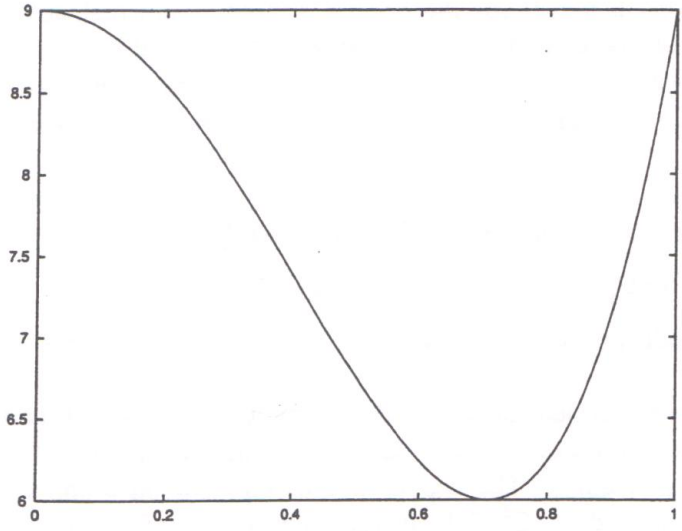

Figure 2. Fourth-order moment of factor scores $E[\eta_t^4]$ as a function of a, where $a = \sqrt{(\beta^2/(1+\beta^2))}$, is the proportion of variance which is transmitted from t-1 to t and where η is an interaction factor.

Even more importantly, we may expect the fourth-order moment to decrease as the number of interaction components contributing to η_t increases. A situation then arises where the Central Limit theorem obtains: the distribution of η_t will tend towards normality. According to the Central Limit theorem the distribution of the sum of N independent random variables converges to the normal distribution if N approaches infinity. In fact there are several variants of the Central Limit theorem that apply under still weaker conditions (e.g. sums of autocorrelated random variables). We will only need the standard version of the Central Limit theorem, however, applying to sums of independently identically distributed (i.i.d.) random variables. The i.i.d. variables concerned are the ζ_t innovations in equations 8 and 9. These equations can be rewritten in such a way that the latent factor η_{t+1} is expressed as an infinite weighted sum of ζ_{t+1}, ζ_t, ... Application of the Central Limit theorem to this so-called moving-average of infinite order shows that the distribution of η always will converge to the normal distribution, irrespective of the distribution of ζ. In the case presently considered the distribution of ζ, where ζ is a pure interaction innovation process, is rather complex. Yet the distribution of the latent genetic and environmental factors still will converge to a normal distribution as time proceeds.

At present there are two interaction components contributing to η_t, viz. η_{t-1} and ζ_t. But if the measurement at t=1 is recorded late in the developmental process, η_{t-1} will consist of an accumulation of interaction terms built up during development prior to t=1 so that the Central Limit theorem applies and η_t tends to normality. Surprisingly, this implies that the presence of fourth-order moments that equal the expected value under normality does not mean that the developmental process is not the outcome of an accumulation of interaction components.

Given the presence of an additive genetic and unshared environmental series, we may consider the fourth-order moment $[G^2_t E^2_t]$. As above the model for the latent variables is given by equations 8 and 9 for the genetic and environmental autoregressions. We assume that E_t is attributable to an interaction between G_{t-1} and E_{t-1} (both standard normal variables) and that ζ_{et} is the outcome of an interaction between two standard normal variables at time t. As usual, G_t and ζ_{gt} are defined as random zero-mean unit-variance variables.

For E_t we may write (as above): $E_t = a\, E_{t-1} + b\, \zeta_{et}$

and similarly for G_t: $G_t = c\, G_{t-1} + d\, \zeta_{gt}$

Their crossproduct $E[G^2_t E^2_t]$ then equals: $E[G^2_t E^2_t] = 3 - 2a^2 - 2c^2 + 4a^2c^2$ (13)

where a and b are as defined above (and thus depend on the value of the innovations and the transmission parameters) and c and d depend on the genetic autoregression parameters and innovations. The dependency of $[G^2_t E^2_t]$ on a and c is illustrated in Figure 3. The crossproduct $[G^2_t E^2_t]$ can vary between 3 (a=0 and c=0 or a=1 and c=1) and 1 (a=0 and c=1 or a=1 and c=0). We now have a fourth-order moment that is dependent on the autoregressive coefficients b_g.and b_e. In this case the Central Limit theorem also applies.

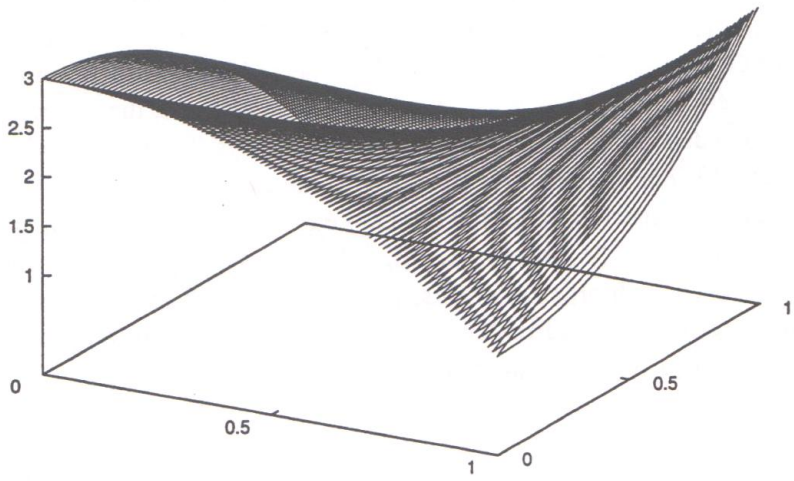

Figure 3. The crossproduct $[G_t^2E_t^2]$ as a function of a and c, where $a = \sqrt{(\beta_e^2/(1+\beta_e^2))}$ and $c = \sqrt{(\beta_g^2/(1+\beta_g^2))}$, are the proportions of respectively G x E variance and genetic variance that are transmitted from t-1 to t.

ILLUSTRATION USING SIMULATED DATA

To illustrate the application of fourth-order moments in the detection of genotype-environment interaction as outlined above, we provide some results obtained by analyzing simulated data. Data sets were simulated with IMSL subroutine FTGEN (IMSL, 1979) for 500 monozygotic and 500 dizygotic twin pairs at three occasions. The latent factors influencing the phenotype at each occasion were a common additive genetic and a common unshared environmental factor. The environmental and additive genetic factors were orthogonal. The longitudinal model for the genetic and environmental factors was a first-order autoregression. The phenotype consisted of three congeneric phenotypic tests (see Figure 1).

The factor loadings of the phenotypic variables on the latent variables were constant over time and equaled 0.9, 0.7, and 0.5 for the additive genetic factor and 0.5, 0.7, and 0.9 for the unshared environmental factor. Unique variance (error variance) was 1 for each phenotypic variable at each measurement occasion. Series of five data sets were simulated according to a G,E model without genotype-environment interaction and according to an interaction model where the E factor was made up of G x E interaction. Each of the five data sets were generated with the following values of the autoregressive coefficients, β_g.and β_e: 1, 1.5, 2, 2.5, and 3. The initial latent variances var(G_1) and var(E_1) and all innovation variances [var(ζ_{g2}), var(ζ_{g3}), var(ζ_{e2}) and var(ζ_{e3})] were all set to equal 1.

The G x E interaction was introduced at the first time point by multiplying the environmental deviation scores with the additive genetic deviation scores ($E^*_1 = E_1 \times G_1$). This interaction factor was transmitted to subsequent time points by the

autoregressive process. At time point t = 2 and t = 3, the innovation terms also consisted of G x E interaction ($\zeta^*_{e2} = \zeta_{e2}$ x ζ_{g2} and $\zeta^*_{e3} = \zeta_{e3}$ x ζ_{g3}). The expected values fourth-order moments of G_1, G_2 and G_3 are 3. At t = 1, the expected fourth-order moments $[E_1^4]$ and $[E_1^2G_1^2]$ are given in Table 2 as 9 and 3, respectively. At t = 2, the expected values for $[E_2^4]$ and $[E_2^2G_2^2]$ depend on the autoregressive coefficients (β_g.and β_e) and are given in equations 12 and 13. For example, for $\beta_g = \beta_e = 3$, the expected fourth-order moment for $[E_2^4]$ is 7.9 and the expected value for $[E_2^2G_2^2]$ equals 2.64. For $\beta_g = \beta_e = 2$, these expected values are 7.1 and 2.36, and for $\beta_g = \beta_e = 1$, they are respectively 6 and 2. The expected values of the fourth-order moments, $[E_3^4]$ and $[E_3^2G_3^2]$ can be derived in the same manner as those for $[E_2^4]$ and $[E_2^2G_2^2]$ (these are not derived in the present paper, but will be considered in a future publication).

Table 3A. Fourth-order moments of genetic and environmental factor scores from simulated longitudinal data on 500 MZ and 500 DZ twin pairs, without G x E interaction. Estimated moments are given separately for twin 1 and twin 2, and are pooled over zygosities. Expected values for G_t^4 and E_t^4 equal 3, and expected values for the crossproduct $[E_t^2G_t^2]$ equal 1. Results are presented from five simulations with different values for the autoregressive coefficients β_g and β_e (3,2.5,2,1.5, and 1).

		Twin 1			Twin 2	
	I $\chi^2(333) = 348.0$ (p=.27)		$\beta_g = 3.014$ (3)	$\beta_e = 3.014$ (3)		
t	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$
1	2.80	3.08	1.08	2.94	2.97	0.94
2	2.79	3.06	1.08	2.87	2.94	0.95
3	2.78	3.05	1.08	2.87	2.93	0.94
	II $\chi^2(333) = 349.8$ (p=.25)		$\beta_g = 2.469$ (2.5)	$\beta_e = 2.506$ (2.5)		
t	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$
1	2.75	2.98	1.06	2.83	2.94	0.93
2	2.76	3.00	1.06	2.88	2.93	0.93
3	2.77	3.00	1.06	2.89	2.91	0.93
	III $\chi^2(333) = 381.9$ (p=.03)		$\beta_g = 2.007$ (2)	$\beta_e = 2.019$ (2)		
t	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$
1	2.99	2.98	1.03	2.89	3.08	1.12
2	3.00	2.97	1.03	2.95	3.04	1.06
3	3.04	2.97	1.03	2.97	3.05	1.07
	IV $\chi^2(333) = 368.8$ (p=.09)		$\beta_g = 1.491$ (1.5)	$\beta_e = 1.492$ (1.5)		
t	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$
1	3.05	2.88	1.08	2.96	2.78	1.02
2	2.95	2.83	1.11	2.97	2.77	1.02
3	2.93	2.81	1.13	2.97	2.78	0.97
	V $\chi^2(333) = 362.0$ (p=.13)		$\beta_g = .997$ (1)	$\beta_e = 0.976$ (1)		
t	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$
1	3.18	3.27	1.27	3.25	2.81	1.30
2	3.16	3.27	1.19	3.23	2.83	1.15
3	2.94	2.98	1.04	3.06	2.86	1.23

Table 3B. Fourth-order moments of genetic and environmental factor scores from simulated longitudinal data on 500 MZ and 500 DZ twin pairs for a simulation model including genotype-environment interaction, where E* = G x E. Estimated moments are given separately for twin 1 and twin 2, and are pooled over zygosities. Results are shown for well-fitting models.

	Twin 1			Twin 2		
	I $\chi^2(333) = 364.9$ (p=.11)		$\beta_g = 2.962$ (3)		$\beta_e = 3.016$ (3)	
t	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$
1	3.18	7.04	2.60	3.17	8.26	2.92
2	3.17	6.95	2.58	3.12	8.05	2.68
3	3.19	6.94	2.57	3.13	8.01	2.66
	II $\chi^2(333) = 336.3$ (p=.44)		$\beta_g = 2.499$ (2.5)		$\beta_e = 2.482$ (2.5)	
t	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$
1	2.69	5.72	2.17	2.87	6.10	2.10
2	2.65	5.65	2.11	2.81	6.06	2.06
3	2.67	5.60	2.10	2.82	6.11	2.08
	III $\chi^2(333) = 349.0$ (p=.26)		$\beta_g = 1.993$ (2)		$\beta_e = 1.997$ (2)	
t	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$
1	3.55	6.83	2.39	3.66	5.72	2.47
2	3.29	6.59	2.16	3.42	5.24	2.18
3	3.38	6.51	2.16	3.49	5.18	2.14
	IV $\chi^2(333) = 353.0$ (p=.21)		$\beta_g = 1.479$ (1.5)		$\beta_e = 1.525$ (1.5)	
t	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$
1	3.35	6.55	2.64	3.42	6.75	2.32
2	2.89	5.69	1.80	2.88	5.38	1.96
3	2.88	5.43	1.73	2.83	5.08	1.92
	V $\chi^2(333) = 348.3$ (p=.27)		$\beta_g = .986$ (1)		$\beta_e = 1.006$ (1)	
t	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$
1	3.79	5.70	3.07	3.98	5.90	2.97
2	3.23	5.29	2.37	3.44	5.40	2.40
3	2.90	4.29	1.80	2.97	4.56	1.87

Table 3A contains the results obtained by analyzing the simulated data without G x E interaction. For each of the five data sets we first give the χ^2 and associated probability and the parameter estimates for β_g.and β_e obtained from fitting the true model. Next, the the fourth-order moments of the estimated factor scores for G_t and E_t are shown for each data set (whose expected values are 3 in data without interaction) and the values obtained for the crossproduct $[E_t^2G_t^2]$ (which equals 1 in data without interaction). It can be seen that for all values of b the fourth-order moments of factor scores are close to their expected values and would lead to the correct conclusion of no G x E interaction.

Tables 3B and 3C contain the fourth-order moments obtained by analyzing simulated data containing a latent factor that behaves in the analysis of second-order moments (the covariance analysis) as an unshared environmental factor E, but is made up of G x E interaction terms according to the simulation model outlined above. If we compare the estimated moments for $[E_t^4]$ and for $[E_t^2G_t^2]$ with their expected values in the absence of

G x E interaction (respectively 3 and 1), it is clear that the values reported in Tables 3B and 3C indicate that E should be regarded as an interaction factor, instead of pure environmental one. Although these values for the fourth-order moments are in some instances not as high as their expectations, it clear that they are substantially higher than the values in Table 3A for the case of no G x E interaction. Comparing Tables 3B and 3C it may be seen that there is 'trade-off' between χ^2 goodness-of-fit and the value of the fourth-order moments: these are closer to their expected values under interaction when χ^2 is higher (Table 3C). This probably is related the amount of kurtosis in the raw data, where a higher kurtosis makes it easier to detect G x E interaction, but also leads to a higher χ^2. In contrast, the observed fourth-order moments $[G_t^4]$ are close to their expected value of 3 in all simulations, indicating that G is a pure genetic factor.

Table 3C. Fourth-order moments of genetic and environmental factor scores from simulated longitudinal data on 500 MZ and 500 DZ twin pairs for a simulation model including genotype-environment interaction, where E* = G x E. Estimated moments are given separately for twin 1 and twin 2, and are pooled over zygosities. Results are shown for the case where the true model does not show a good fit (as indicated by a significant χ^2) to the data.

		Twin 1			Twin 2	
	I $\chi^2(333) = 415.1$ (p=.001)		$\beta_g = 3.024$ (3)		$\beta_e = 2.981$ (3)	
t	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$
1	2.79	9.03	2.63	2.79	8.50	2.61
2	2.75	8.61	2.46	2.72	8.29	2.50
3	2.75	8.52	2.43	2.71	8.26	2.48
	II $\chi^2(333) = 418.9$ (p=.001)		$\beta_g = 2.473$ (2.5)		$\beta_e = 2.515$ (2.5)	
t	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$
1	2.95	7.17	2.61	2.98	7.35	2.62
2	2.93	6.87	2.42	2.95	7.43	2.61
3	2.94	6.80	2.39	2.95	7.41	2.56
	III $\chi^2(333) = 422.8$ (p=.001)		$\beta_g = 1.976$ (2)		$\beta_e = 1.992$ (2)	
t	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$
1	3.00	5.26	1.97	3.34	8.98	3.16
2	3.05	4.84	1.89	3.26	8.17	2.85
3	3.11	4.71	1.79	3.22	7.86	2.69
	IV $\chi^2(333) = 408.8$ (p=.003)		$\beta_g = 1.463$ (1.5)		$\beta_e = 1.498$ (1.5)	
t	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$
1	3.75	5.87	2.74	3.60	5.24	2.27
2	3.46	5.17	2.16	3.16	4.78	1.91
3	3.45	5.06	2.05	3.21	4.59	1.72
	V $\chi^2(333) = 391.5$ (p=.015)		$\beta_g = 1.001$ (1)		$\beta_e = .958$ (1)	
t	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$	$E[G_t^4]$	$E[E_t^4]$	$E[E_t^2G_t^2]$
1	3.86	5.37	3.66	3.98	5.77	2.99
2	3.87	5.45	2.78	3.44	5.51	2.03
3	2.46	4.84	2.07	2.97	4.33	1.65

DISCUSSION

It was shown that the change between two consecutive time points t-1 and t in the fourth-order moments of the factor scores associated with an interaction factor series depends upon the proportion of variance which is transmitted from t-1 to t. In the genetic simplex, which involves first-order autoregressions describing the latent genetic and environmental factor series, the total amount of transmitted variance between two consecutive time points is a simple function of the autoregressive beta-coefficients: $\beta^2 var(\eta_{t-1})$. Hence the proportion of transmitted variance of an interaction factor series at time t depends upon the relative magnitude of the autoregressive β-coefficient in comparison with the standard-deviation of the innovation term at t. For this rather simple scheme, explicit expressions for the change in fourth-order moments were derived. In particular it was found that this change is absent only if the variance of the innovation term is zero. In that special case the fourth-order moments keep their initial values (9 and 3, respectively) at all time points. Notice that if the innovation variance of a latent autoregression in our simplex model becomes zero, then this part of the simplex model reduces to a common factor model as described by Eaves, Long and Heath (1986).

Only the change in the fourth-order moments of an interaction factor between two consecutive time points t-1 and t were derived. The change between t-1 and t+1 (spanning two lags) then follows immediately by a recursive application of our derivation. In fact, the change between an arbitrary number of lags can thus be derived. Moreover, the same principles can be used to determine the changes concerned for interaction factor series obeying more complex time-series models such as higher-order autoregressions and moving-averages.

The simulations of G x E interaction in longitudinal models that were presented in this paper are not exhaustive, but serve to illustrate the possibility of detecting such interactions without measures of the environment or the genotype. We showed that estimates of fourth-order moments of factor scores are close to their expected values of 3 and 1 in data without interaction. In simulated longitudinal data with interaction the value of these fourth-order moments are larger indeed indicate the presence of interaction.

Our approach to the detection of interaction factors hinges upon the estimation of fourth-order moments. Unfortunately, the sampling variability of these estimates is very high and therefore one will need a large sample of phenotypical values in order to secure the reliability of the detection tests. In our simulation studies (where it is certain that the generated phenotypical values constitute a homogeneous sample) it was found that estimates of fourth-order moments strongly depend upon the extreme phenotypical values in a sample and that removal of these extreme observations (interpreted as outliers) could lead to severe bias. In future explorations of the approach we intend to consider alternative L-statistics characterizing the kurtosis of latent factor series (Hosking, 1990). It appears that, compared with the conventional fourth-order moments, L-kurtosis is less subject to bias in estimation, approximates its asymptotic distribution more closely in finite samples, and is more robust to the presence of genuine outliers in the phenotypical data.

Perhaps the most surprising conclusion of the present study, at least to us, is that the interactive nature of a developmental process becomes invisible after a sufficient amount of time, even under the most favorable circumstances. Started as a pure interaction factor (with fourth-order moments of 9 and 3, respectively), the repetitive addition of pure interaction innovations in time combined with non-zero transmission over time leads to a developmental process whose interactive nature becomes impossible to detect (fourth-order moments of 3 and 1, respectively). It could be that such processes account for the absence of a shared environmental component in many twin and family studies since a G x C interaction factor will look like a genetic factor. Also, this might be part of the explanation for the high contributions of unshared environmental factors to many characteristics (Plomin and Daniels, 1987) since all interactions with E look like E. We believe that this phenomenon can be best understood by an appeal to the Central Limit theorem: because of the repeated addition of independently and identically distributed interaction innovations the random process approaches normality as time proceeds. Notice that the applicability of the Central Limit theorem does not depend upon the details of our interaction detection procedure, but pertains to any developmental process involving nonzero transmission in combination with innovations. That is, the presence of interactive causes underlying such processes will become invisible as development proceeds, even if the causal interaction is enduring, stable, and effective during the entire life span. As far as we know this is the first time that this result, which may have far-reaching implications transcending the field of behavior genetics proper, has been noticed in the published literature.

REFERENCES

Boomsma, D.I., Martin, N.G., & Molenaar, P.C. (1989). Factor and simplex models for repeated measures: application to two psychomotor measures of alcohol sensitivity in twins. *Behav Genet, 19*, 79-96.

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

Boomsma, D.I., Molenaar, P.C., & Dolan, C.V. (1991). Estimation of individual genetic and environmental profiles in longitudinal designs. *Behav Genet, 21*, 243-255.

Boomsma, D.I., Molenaar, P.C.M., & Orlebeke, J.F. (1990). Estimation of individual genetic and environmental factor scores. *Genet Epidemiol, 7*, 83-91.

Brown, R.G. (1983). *Introduction to random signal analysis and Kalman filtering*. New York: John Wiley & Sons.

Eaves, L.J. (1984). The resolution of genotype x environment interaction in segregation analysis of nuclear families. *Genet Epidemiol, 1*, 215-228.

Eaves, L.J., Long, J., & Heath, A.C. (1986). A theory of developmental change in quantitative phenotypes applied to cognitive development. *Behav Genet, 16*, 143-162.

Freeman, G.H. (1973). Statistical methods for the analysis of genotype-environment interactions. *Heredity, 31*, 339-354.

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*, 133-151.

Hosking, J.R.M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society B, 52*, 105-124.

IMSL, Inc. (1979). *IMSL library reference manual, edition 7*. Houston, TX: IMSL Inc..

Jinks, J.L. & Fulker, D.W. (1970). Comparison of the biometrical genetical, MAVA, and classical approaches to the analysis of human behavior. *Psychol Bull, 73*, 311-349.

Joreskog, K.G. & Sorbom, D. (1988). *LISREL VII: A guide to the program and applications*. Chicago: Spss Inc..

Lawley, D.N. & Maxwell, A.E. (1971). *Factor analysis as a statistical method*. London: Butterworths.

Loehlin, J. (1991). Behavior genetic studies of change. In L.M. Collins & J.L. Horn (Eds.). *Best methods for the analysis of change. Recent advances, unanswered questions, future directions* Washington, D.C.: American Psychological Association

Loehlin, J.C., Horn, J.M., & Willerman, L. (1989). Modeling IQ change: evidence from the Texas Adoption Project. *Child Dev, 60*, 993-1004.

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

Martin, N.G., Eaves, L.J., & Heath, A.C. (1987). Prospects for detecting genotype X environment interactions in twins with breast cancer. *Acta Genet Med Gemellol (Roma), 36*, 5-20.

Mather, K. & Jinks, J.L. (1982). *Biometrical Genetics: The study of continuous variation*. London: Chapman and Hall.

McArdle, J.J. (1986). Latent variable growth within behavior genetic models. *Behav Genet, 16*, 163-200.

McDonald, R.P. (1967). Factor interaction in nonlinear factor analysis. *Br J Math Stat Psychol, 20*, 205-215.

McDonald, R.P. & Burr, E.J. (1967). A comparison of four methods of constructing factor scores. *Psychometrika, 32*, 381-401.

Molenaar, P.C. & Boomsma, D.I. (1987). Application of nonlinear factor analysis to genotype-environment interaction. *Behav Genet, 17*, 71-80.

Molenaar, P.C., Boomsma, D.I., Neeleman, D., & Dolan, C.V. (1990). Using factor scores to detect G X E interactive origin of "pure" genetic or environmental factors obtained in genetic covariance structure analysis. *Genet Epidemiol, 7*, 93-100.

Molenaar, P.C.M., Boomsma, D.I. & Dolan, C.V. (1991). Genetic and environmental factors in a developmental perspective. In D. Magnusson, L.R. Bergman, G. Rudinger & B. Thorestad (Eds.). *Problems and methods in longitudinal research: Stability and change* (pp. 250-273). Cambridge: Cambridge University Press

Neale, M.C. (1991). *Statistical modeling*. Box 3, MCV, Richmond, VA 23298: Department of Human Genetics.

Neale, M.C. & Cardon, L.R. (1992). *Methodology for genetic studies of twins and families (NATO ASI Series D: Behavioral and Social Sciences-vol 67)*. Dordrecht, The Netherlands: Kluwer Academic Publishers B. V..

Plomin, R. (1986). *Development, genetics and psychology*. Hillsdale: Lawrence Erlbaum Associates.

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

Plomin, R., DeFries, J.C. & McClearn, G.E. (1990). *Behavior Genetics: A primer*. San Francisco: W.H.Freeman and Company.

Priestly, M.B. & Subba Rao, T. (1975). The estimation of factor scores and Kalman filtering for discrete parameter stationary processes. *International Jounal of Control*, *21*, 971-975.

Saris, W.E., dePijper, M., & Mulder, J. (1978). Optimal procedures for estimation of factor scores. *Sociological Methods & Research*, *7*, 85-106.