

Structural equation modeling in the genetically informative study of the covariation of intelligence, working memory and planning

I. Voronin^{1,a}, V. Ismatullina¹, I. Zakharov¹, G. Vasin¹ and S. Malykh¹

¹Psychological Institute of Russian Academy of Education, Laboratory of developmental behaviour genetics, 125009 Moscow, Russia

Abstract. Structural equation modelling (SEM) has become an important tool in behaviour genetic research. The application of SEM for multivariate twin analysis allows revealing the structure of genetic and environmental factors underlying individual differences in human traits. We outline the framework of twin method and SEM, describe SEM implementation of a multivariate twin model and provide an example of a multivariate twin study. The study included 901 adolescent twin pairs from Russia. We measured general cognitive ability and characteristics of working memory and planning. The individual differences in working memory and planning were explained mostly by person-specific environment. The variability of intelligence is related to genes, family environment, and person specific environment. Moderate and weak associations between intelligence, working memory, and planning were entirely explained by shared environmental effects.

1 Introduction

Behavioural genetic research aims to disentangle the genetic and environmental influences on the individual differences in behavioural traits. The information about exact genetic code of a person is still hardly accessible and the exact mechanisms of gene-environment interplay are mostly unknown. However the methods of behaviour genetics provide indirect estimates of the variance accounted for genetic and environmental factors. Such methods compare family members with known genetic relatedness. Twin method is most used among them.

The combination of twin method and structural equation modelling allows addressing complex research questions: the role of genes and environment in the relationship between several traits, the sources of stability and change of certain trait in time etc. This paper provides an example of the application of structural equation modelling for multivariate twin research.

2 Twin method

Twin method compares two groups of twins to disentangle genetic and environmental effects on the measured trait (phenotype). **Monozygotic** (MZ) twins develop from the same egg and have similar genetic code. **Dizygotic** (DZ) twins develop from two different eggs fertilized at the same time, so they share 50% of genetic variation [1]. Twin method consider following sources of individual differences:

- Additive genetic effects (A) represent the additive effects of many genetic variants which influence the trait.

- Non-additive genetic effects (D) represent the interaction between alleles (dominance and epistasis).
- Environmental effects shared by the members of the same family (C): socioeconomic status, parenting style etc.
- Individual experience of family members (person-specific environmental effects, E), and also measurement error.

The similarity of twins within pair sources from shared genes and shared family environment. MZ twins share genetic effects and family environment to the full extent. DZ twins share 50% of additive genetic effects, 25% of non-additive genetic effects, and 100% of family environment. MZ twins differ because of person-specific environment, DZ twins – because of unique environment and genes. This set of assumptions defines the set of equations:

$$\begin{aligned} r_{MZ} &= A + D + C \\ r_{DZ} &= 0.5A + 0.25D + C \\ 1 &= A + D + C + E \end{aligned} \quad (1)$$

The set of three equations (1) allows computing three components at the same time: A, C, and E, or A, D, and E:

$$\begin{aligned} E &= 1 - r_{MZ} \\ A &= 2(r_{MZ} - r_{DZ}) \\ C &= 2r_{DZ} - r_{MZ} \end{aligned} \quad (2)$$

$$\begin{aligned} E &= 1 - r_{MZ} \\ A &= 4r_{DZ} - r_{MZ} \\ D &= 2(r_{MZ} - 2r_{DZ}) \end{aligned} \quad (3)$$

^a Corresponding author: ivan.voronin@pirao.ru

Twin method provides reliable evidence about genetic and environmental effects under following assumptions:

- The twins do not differ from general population in terms of a certain phenotype.
- MZ and DZ twins share their environments to the same extent.
- There are no gene-environment correlations or interactions.
- The mating in population is random.

3 Structural equation modelling

3.1 Basic principles of structural equation modelling

Structural equation modeling (SEM) is a set of methods which allows checking the hypotheses about the structure of relationships between observed and unobserved (latent) variables [2].

The model is usually represented by path diagram which include variables and linear relationships between them. On such a diagram circles are latent variables, squares are observed variables, single-headed arrows are one-way (causal) relationships, and double-headed arrows are two-way relationships (covariances). Figure 1 provides an example of a path diagram.

Path coefficients (the values of causal effects and covariances) constitute the set of model parameters. The estimation is possible only when the number of estimated parameters is less or equal than number of available data (number of observed variances and covariances). The difference between the two is known as number of degrees of freedom of the model (df).

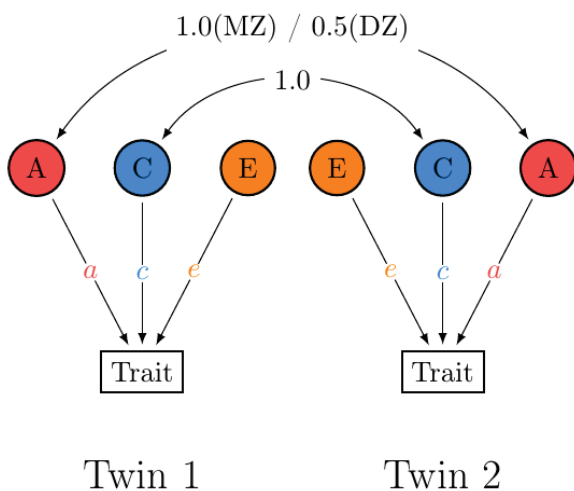


Figure 1: Path diagram for the univariate twin model

The expected covariance between two variables is computed as a sum of the covariances accounted for all possible paths between these variables. Such path must not go through the same variable twice, and must go first through one-way paths backwards, then through the one two-way path, then through one-way paths forward. The value of covariance accounted for the path is the product of path coefficients along this path.

The model can include several sub-models. For example, any twin model comprises separate sub-models for MZ and DZ twins. The models use same set of parameters but fit on different datasets (Figure 1).

The difference between actual and expected covariance matrices indexes general fit of a model: the model with better fit explains the covariance structure more accurately. The programs for structural equation modeling estimate model parameters by minimizing the difference between actual and expected covariance matrices. Least squares and maximum likelihood are the most used methods of estimation. Least squares suggest minimizing sum of the squared differences between actual and expected data. Maximum likelihood estimation maximizes the likelihood of the data given the model with certain parameters). For twin analysis the full information maximum likelihood estimation is most appropriate as it uses each single data point to compute the likelihood. Likelihood-based fit statistics can also be used for the estimation of parameters' confidence intervals.

The main measure of general model fit is the difference between the model likelihood and the likelihood of perfectly fitting (Saturated) model. This difference distributed as χ^2 with df equal the difference in df between two models. χ^2 test shows if the model fits the data well. A statistically significant p-value ($p < 0.05$) means that expected and observed covariance matrices differ significantly, so the model does not fit the data. The opposite means that the model explains data well.

χ^2 test can also be used to compare the nested models – when the set of parameters of one model is the subset of the parameters of another. The model with fewer parameters (more parsimonious) is preferable when two models have similar goodness-of-fit. The comparison of nested models is often used to exclude non-significant parameters.

3.2 Application of structural equation modelling in twin research

Although SEM is a useful tool in a univariate twin study, it provides even more opportunities for the study of genetic and environmental effects underlying the variance shared by several phenotypes. This approach is based on the comparison of cross-trait cross-twin correlations (correlations between one trait in one twin and another trait in another twin) in MZ and DZ twins. The discrepancy between the correlations indexes genetic effects which affect both phenotypes.

Basic multivariate twin model can be defined via Cholesky decomposition (Figure 2) [3]. The advantage of Cholesky decomposition is that each component of the covariance matrix has a compact representation as matrix algebra:

$$\begin{aligned} A &= \begin{pmatrix} a_1 & 0 \\ x & a_2 \end{pmatrix} \times \begin{pmatrix} a_1 & x \\ 0 & a_2 \end{pmatrix} \\ C &= \begin{pmatrix} c_1 & 0 \\ y & c_2 \end{pmatrix} \times \begin{pmatrix} c_1 & y \\ 0 & c_2 \end{pmatrix} \\ E &= \begin{pmatrix} e_1 & 0 \\ z & e_2 \end{pmatrix} \times \begin{pmatrix} e_1 & z \\ 0 & e_2 \end{pmatrix} \end{aligned} \quad (4)$$

This allows computing cross-twin covariance matrices in matrix algebra form as well:

$$\begin{aligned} cov(MZ) &= \begin{pmatrix} A + C + E & A + C \\ A + C & A + C + E \end{pmatrix} \\ cov(DZ) &= \begin{pmatrix} A + C + E & 0.5A + C \\ 0.5A + C & A + C + E \end{pmatrix} \end{aligned} \quad (5)$$

The correlations between genetic and environmental factors affecting phenotypic variables can be computed upon the parameters of the Cholesky model.

4 The study of genetic and environmental factors of the relationships between intelligence, working memory, and planning

4.1 Introduction

Intelligence (general cognitive ability) is a stable cognitive ability which explains individual differences in a broad range of intellectual activities. Working memory is a cognitive system aimed to concurrently maintain access to information relevant to a task. The strong relationship between intelligence and working memory has been established in many studies [4, 5] and raised the discussion about the nature of their covariance. At the same time intelligence is assessed by means of complex multi-step tasks which require the elaboration of plan and strategy. Indeed, the study of Gilhooly et al. [6] demonstrates the relationships between individual differences in intelligence, working memory, and planning ability.

The current study investigates genetic and environmental sources of phenotypic relationships between planning intelligence, working memory, and planning ability in Russian adolescents.

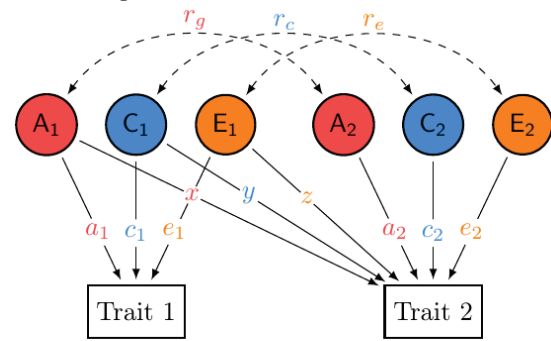
4.2 Sample and methods

The sample included 377 MZ and 524 DZ pairs of Russian twins (mean age 12.9 years, SD=3.0 years, 45% of twins were female).

General cognitive ability was assessed by means of Raven's Standard Progressive Matrices. The test comprises 5 sub-scales (12 items each) of increasing difficulty. Each task requires to choose an element which completes the pattern (matrix). The sum of correct responses was used as an index of intelligence.

Working memory and planning ability were measured by means of Spatial Working Memory (SWM) and Stocking of Cambridge (SOC) tests from the automated neuropsychological test battery CANTAB. SWM test includes the series of trials where a participant has to collect the tokens searching for them in the boxes on the screen. Within a trial each box contains the token once, so efficient strategy of the search assumes opening the boxes which have not been opened within this search. For twin analysis we used the index of working memory

productivity which takes into account both number of errors and response time.



Twin 1

Figure 2: Path diagram for the bivariate twin model (one twin, dashed paths are computed upon other parameters)

SOC test include the tasks which require repeating the pattern shown in the bottom part of the panel by moving balls from one 'sock' in minimum number of moves. The participant is encouraged to plan the moves beforehand. For twin analysis we used the index of planning efficiency which takes into account number of problems solved in minimum moves and the time of initial planning.

As a first step we computed phenotypic Pearson correlations between intelligence, working memory, and planning ability measures. The correlations were computed on the sample of one twin selected randomly from each twin pair. We fit three univariate twin models to estimate genetic and environmental effects separately for each measure. Then we fit a trivariate twin model with correlated factors (Figure 3) to find out whether genetic or environmental factors explain phenotypic correlations between the measures. We constrained the correlations between the genetic and environmental components to be positive as previous research gave no evidence that they could be negative.

Before performing twin analyses we adjusted phenotypic variables for the effects of sex and age to avoid possible confounds [7] and standardized them.

4.3 Results

The measure of intelligence showed moderate correlation with SWM score (0.231, p = 0.006) and small correlation with SOC score (0.127, p = 0.105). Working memory and planning ability measures correlated weakly (0.105, p = 0.166).

The univariate twin analysis showed that the variability of Raven's score was associated with genes, family environment, and person-specific environment (A = 30%, C = 46%, E = 24%). The individual differences in SOC score was explained by environmental factors (A = 7%, C = 30%, E = 63%). The variance of SWM score was explained by genes and person-specific environment (A = 34%, C = 8%, E = 58%). The results of the univariate twin analysis

Table 1. The estimates of genetic and environmental correlations across intelligence, working memory, and planning ability measures and their impact to phenotypic correlations (with 95% confidence intervals).

	A	C	E	Phenotypic*	A (%)	C (%)	E (%)
Raven ↔ SOC	0.00 [0.00, 1.00]	0.77 [0.43, 1.00]	0.00 [0.00, 0.08]	0.29 [0.17, 0.38]	0% [0%, 32%]	100% [69%, 100%]	0% [0%, 11%]
Raven ↔ SWM	0.06 [0.00, 1.00]	0.77 [0.09, 1.00]	0.00 [0.00, 0.20]	0.24 [0.12, 0.36]	0% [0%, 93%]	94% [7%, 100%]	0% [0%, 31%]
SOC ↔ SWM	0.02 [0.00, 1.00]	0.17 [0.00, 1.00]	0.14 [0.00, 0.29]	0.13 [0.03, 0.24]	2% [0%, 100%]	33% [0%, 100%]	65% [0%, 100%]

* Estimated phenotypic correlation

suggests that the phenotypic correlation between intelligence and working memory can be explained by common genetic effects, as genes has effect on both phenotypes. Another possible source is person-specific environment; however, the factors of person-specific environment of intelligence usually represent measurement error which is independent for different assessment tools.

The results of the trivariate twin modelling are shown on Figure 3 and Table 1. The estimates for the individual impact of genetic and environmental factors to the variation of Raven's, SOC, and SWM scores differ from ones obtained from the univariate analysis. The reason for this is that in multivariate twin analysis the estimation of the effects of genes and environment on individual phenotypic variables takes into account cross-twin cross-trait correlations as well as cross-twin within-trait correlations. It should also be noted that the estimates of the phenotypic correlations obtained from the trivariate model are higher and all statistically significant according to the 95% confidence intervals

(Table 1). The reason for this is that these estimates are computed upon the model's parameters and take into account larger amount of data (both twins).

The correlations across genetic and person-specific environmental components of intelligence, planning ability, and working memory measures are small and non-significant. There are statistically significant relationships of size 0.77 between family environmental components of Raven's and SOC, and Raven's and SWM. These relationships explain the corresponding phenotypic correlations entirely.

4.4 Discussion

The multivariate twin study shows that intelligence, working memory, and planning ability in Russian adolescent twins show different structure of phenotypic variability. Individual differences in planning ability and working memory are explained mostly by unique environment, but also to some extent by family environment (planning ability) and genes (working

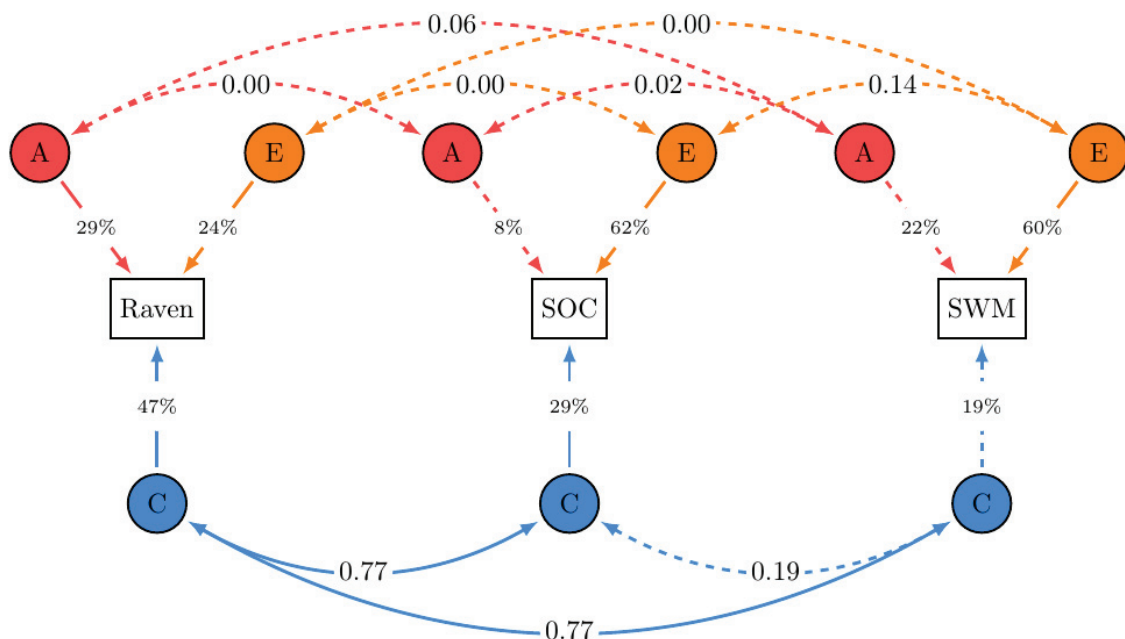


Figure 3: Trivariate twin model with correlated factors (one twin). Non-significant relationships are dashed.

memory). The variability of intelligence is accounted for all three sources of individual differences. The relationships between the phenotypes are explained by shared environmental correlations.

Previous research generally shows higher genetic effects on general and specific cognitive abilities and functions. The large twin study of intelligence performed by [8] gives the estimate of heritability of intelligence 55%. In the study of Ando, Ono, and Wright [9] genes explained 43-49% of individual differences in working memory characteristics.

In our study we observe high impact of person-specific environmental factors to individual differences in cognitive measures (working memory and planning ability). Person-specific environmental factors include measurement error, so low reliability of composite indices can explain unexpectedly high impact of E and low impact of A.

Another possible reason of the discrepancy is age of participants and cross-cultural differences across studies. The increase of heritability of intelligence through the lifespan is a well-established fact [10]. The possible explanation for this is gene-environment correlation which amplifies genetic effects in favourable environments [11]. This effect manifests both in the increase of heritability through the lifespan and in the variability of estimates of heritability in twin studies.

The study of genetic covariance among measures of general cognitive ability and working memory shows that the covariance between two phenotypes can be explained by shared genetic effects [12]. However, the authors used AE model which included genetic and person-specific effects only. In our study genetic effects are moderate or low, so they cannot explain substantial amount of the relationships between intelligence, working memory, and planning ability.

Further research is required to make firm conclusion about aetiology of relationships between intelligence, working memory, and planning ability because the study of moderate or small phenotypic relationships requires large sample size. Another possible option is to use other measures of the phenotypes to reduce measurement error.

5 General discussion

We aimed to outline the application of structural equation modelling in twin study and to provide an example of multivariate twin study.

Multivariate twin study allows to reveal the structure of generic and environmental factors which underlie the variability of several phenotypes. One of the prominent results of multivariate twin study is the hypothesis of generalist genes for cognitive abilities [13, 14]. Cholesky decomposition was used to show that various cognitive processes are associated with same genetic factors.

The flexibility of structural equation modelling tools allows to verify the wide range of hypotheses in twin study. However, the methods are demanding in terms of sample size and the reliability of measures, especially when phenotypic relationships are moderate or weak.

Acknowledgement

The research was supported by the grant from Russian Foundation for Basic Research No.15-36-20902.

References

1. F.V. Rijdsdijk, P.C. Sham, *Brief. Bioinform.* **3**, 119 (2002)
2. J.C. Loehlin, *Latent variable models: An introduction to factor, path, and structural equation analysis* (Psychology Press, 2004).
3. J.C. Loehlin, *Behav. Genet.* **26**, 65 (1996)
4. M.J. Kane, D.Z. Hambrick and A.R. Conway, *Psychol. Bull.* **131**, 66 (2005)
5. K. Oberauer, R. Schulze, O. Wilhelm and H.M. Süß, *Psychol. Bull.*, **131**, 61 (2005)
6. K.J. Gilhooly, V. Wynn, L.H. Phillips, R.H. Logie and S.D. Sala, *Think. Rreasoning*, **8**, 165 (2002)
7. M. McGue, T.J. Bouchard and Jr, *Behav. Genet.*, **14**, 325 (1984)
8. C.M.A. Haworth, M.J. Wright, M. Luciano, N.G. Martin, E.J.C. De Geus, C.E.M. Van Beijsterveldt, ... Y. Kovas, *Mol. Psychiatr.* **15**, 1112 (2010)
9. J. Ando, Y. Ono, M.J. Wright, *Behav. genet.*, **31**, 615 (2001)
10. T.J. Bouchard, *Twin Res. Hum. Genet.*, **16**, 923 (2013)
11. E.M. Tucker-Drob, D.A. Briley and K.P. Harden, *Curr. Dir. Psychol. Sci.*, **22**, 349 (2013)
12. M. Luciano, M.J. Wright, G.A. Smith, G.M. Geffen, L.B. Geffen and N.G. Martin, *Behav. Genet.* **31**, 581 (2001)
13. Y. Kovas and R. Plomin, *Trends Cogn. Sci.* **10**, 198 (2006)
14. Y. Kovas, C.M. Haworth, P.S. Dale, R. Plomin, R.A. Weinberg, J.M. Thomson and K.W. Fischer, *Monogr. Soc. Res. Child.*, **72**, (2007)