Intelligence 58 (2016) 42–53

Contents lists available at ScienceDirect

Intelligence

Linking the Standard and Advanced Raven Progressive Matrices tests to model intelligence covariance in twin families

Bente Otermann *, Stéphanie M. van den Berg

Department of Research Methodology, Measurement and Data Analysis (OMD), Faculty of Behavioral Sciences, University of Twente, The Netherlands

ARTICLE INFO

Article history:
Received 28 November 2015
Received in revised form 22 May 2016
Accepted 23 June 2016
Available online xxxx

Keywords:
Genetic and cultural transmission
Intelligence
Parent-offspring
Bayesian
Item Response Theory
Item data
Test linking
Harmonization

ABSTRACT

An abundance of research shows significant resemblance in standardized IQ scores in children and their biological parents. Twin and family studies based on such standardized scores suggest that a large proportion of the resemblance is due to genetic transmission, rather than cultural transmission. However, most studies used standardized intelligence scores that were based on different tests for different age groups, which makes it hard to say if the exact same construct is measured. Here we re-analyze intelligence data on two different versions of the Raven Progressive Matrices test, collected in Dutch twin children (Standard test version) and their biological parents (Advanced test version). First, the data from parents and their offspring were harmonized using test linking through an item response theory measurement model. This required collecting data from extra participants who were assessed with items from both test versions. Next, the raw item data were analyzed to study transmission of intelligence, correcting for the differences in difficulty of the items in the parental and child test versions and differences in measurement reliability. Results showed a significant difference in the phenotypic variance in intelligence in the two generations. Model fitting showed that the surplus variance in the parental generation is likely due to surplus environmental variance that is not transmitted to the offspring. This could reflect that there was extra measurement error under the parental testing conditions. Genetic modelling showed that intelligence covariance in parents and their children is most likely based on genetic transmission without cultural transmission.

1. Introduction

Individual variation in intelligence tends to cluster within families (Bartels, Rietveld, Van Baal, & Boomsma, 2002; Posthuma, De Geus, Bleichrodt, & Boomsma, 2000). The similarity between parents and their children can be the product of either genetic or cultural (non-genetic) transmission from parent to child, or perhaps both. Twin and adoption studies investigate how much of the variation in intelligence is explained by genetic and non-genetic sources. With cultural transmission we mean the similarity in phenotype across generations that is not due to the transmission of genetic material; it is the residual predictive power of the parents' phenotypes for the child's phenotype over and above the resemblance in genotype. Using adoption designs, cultural transmission can be distinguished from genetic transmission by the fact that there is no genetic transmission from the adoption parents. Previous adoption studies suggest that there is no significant cultural transmission for specific cognitive abilities (Fulker & DeFries, 1983; Plomin, Fulker, Corley, & DeFries, 1997). However, other adoption studies conclude that there is cultural transmission of intelligence. Scarr and Weinberg (1978, 1983) found that the intelligence of adopted children correlates highly with the intelligence of their adoption parents during their childhood, but becomes more correlated with the intelligence of their biological parents as they grow older. Previous twin research showed that 20–50% of the variability of intelligence can be ascribed to genetic effects and the remaining variance to environmental effects (Fulker & DeFries, 1983; Tucker-Drob & Briley, 2014). These studies used designs including twins and their parents, twins and their children and/or twins and their spouses (Eaves et al., 1999; Giubilei et al., 2008; Reynolds, Baker, & Pedersen, 2000; Rijsdijk, Vernon, & Boomsma, 1998). Such designs including family members of twins are vital to check certain important assumptions regarding for instance assortative mating, gene-environmental correlations, and dominance genetic effects.

Most adoption and twin studies are based on standardized test scores: raw test scores are standardized, for instance to have a mean of 100 and standard deviation 15 within a particular age group (i.e. IQ scores). By analyzing correlations of such IQ scores in families, the implicit assumption is that the same phenotype is

* Corresponding author at: Department of Research Methodology, Measurement and Data Analysis (OMD), Faculty of Behavioral Sciences, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.
E-mail address: benteotermann@gmail.com (B. Otermann).

http://dx.doi.org/10.1016/j.intell.2016.06.006
0160-2896/

measured across age. But since there are such huge age effects on test scores, there are different tests or test versions for particular age groups, such as the the standard and advanced versions of the Raven Progressive Matrices test (Raven, 2000). Apart from the assumption that the same phenotype is assessed in children and adults, the standardization leads to the same variance in scores across age. This standardization only allows to model correlations between family members, and information about any differences in variance is lost. This is important since certain phenomena (e.g., spouse similarity, cultural transmission) can lead to differences in variance across generations that have genetic implications and can therefore lead to biased or wrong conclusions. Studying covariation of intelligence in families therefore requires the use of phenotypes that are harmonized (Van den Berg et al., 2014), that is, phenotypes of different family members should be on the same scale. The study of Wicherts and Johnson (2009) states similar critiques of the use of raw scores of the Raven's in behaviour genetic studies.

Here we propose the use of item-response theory (IRT) based test linking in order to map the observed item data from children and parents to a common latent scale. This allows assessing not only mean and variance differences, but the whole covariance structure within twin families. Moreover, we propose to model the covariance structure not of equated test scores, but rather to model the structure at the latent level, using an IRT-based measurement model. This measurement model links the latent model for the covariance structure to the observed raw item data. In that way, we not only correct for the different sets of test items across test version, but also for the different reliabilities of test scores across test versions and individuals (Van den Berg, Glas, & Boomsma, 2007). Van den Berg et al. (2014) published a similar study with the harmonization of phenotypes using IRT with personality data.

Van Leeuwen, Van Den Berg, and Boomsma (2008) published a study on the genetics of intelligence using data on twins and their parents. Parents were assessed using the 36-item Advanced Raven test, while the 9-year-old twins were tested using the 60-item Standard Raven test. The authors dealt with the different test version problem by analyzing raw item scores through an IRT measurement model, an approach that dealt with differences in measurement reliability within and across scales. However, they assumed that phenotypic variance was constant across generations. Another, more implicit, assumption was that the Advanced and the Standard versions of the Raven measured the exact same phenotype. Here we report the results of a test linking study that assessed the possibility of harmonizing the parental Advanced data and the child Standard data to one common scale. This required the collection of Raven data in a new group of individuals that were assessed with both Advanced and Standard test items and IRT-based model fitting. Next, these results were used to re-analyze the Van Leeuwen et al. (2008) data in order to study the covariance structure at the latent common scale and to answer the question how intelligence in the parents is conferred to the children.

2. Materials and methods

2.1. Materials

In this study the Raven Progressive Matrices test (RPM) is used to measure intelligence. The RPM is a widely used nonverbal test of eductive ability and consists of visual problems (Raven, 2000). The items in this test are multiple choice and ranked with regard to difficulty. Here we used two versions of the RPM: the Standard Progressive Matrices (SPM) and the Advanced Progressive Matrices (APM). The SPM consists of five sets (A-E) of 12 items each, resulting in 60 items (Raven, Raven, & Court, 1998a), and the APM consists of 36 other items (Raven et al., 1998a). The test-retest reliability of the SPM is 0.88 in children (Raven et al., 1998a) and for the APM is 0.91 in adults (Raven, Raven, & Court, 1998b).

2.2. Participants

The Van Leeuwen et al. (2008) data consist of item data from 9-year-old twins sampled from the Dutch population of twins registered at the Netherlands Twin Register (NTR) who completed the SPM, and item data from the twins' parents who completed the APM (paper-and-pencil versions). This total data set consists of 112 families (224 children and 189 parents). Mean age of the twins at time of assessment was 9.1 years, ranging from 8.9 to 9.5 years ($N = 327$), of the fathers 43.7 years ($N = 94$, $SD = 3.7$ years) and of the mothers 41.9 years ($N = 95$, $SD = 3.4$ years). Zygosity status of the twin pairs (identical or fraternal) was determined by questionnaire items and DNA polymorphisms. The sample is representative of the Dutch population, albeit that the average IQ in this particular sample was slightly above 100. For more details, see Van Leeuwen (2008).

Additional data of additional participants was collected in 49 Dutch adults at the University of Utrecht in the autumn of 2013, using a snowballing sampling technique. These were given paper-and-pencil tests consisting of a number of SPM items and a number of APM items. In order to optimize the information gained from the above-average intelligent adult participants (working or studying at a university), 16 APM items were selected on the basis of the proportion of correct answers (p-values) in the parental Van Leeuwen et al. (2008) data: between 0.40 and 0.70. A subset of rather difficult SPM test items was selected: the 10 most difficult items from the B, C, D and E sets. Half of the participants (randomly selected) got the APM items first and then the SPM items, while the other half started with the SPM items. For the complete set of items, see Table 1, where the items selected for the test linking data collection are printed in bold. All started with four very easy items for practice (the first two items of the SPM followed by the first two items of the APM). These items were not used in the data-analysis. Data were collected in 18 males and 30 females (plus one participant that did not disclose information on sex), aged between 19 and 63 years. Thirty-three participants were students (at higher professional, academic bachelor or master level), 15 had a job (medium professional level and upwards), and one participant was unemployed. The sample size was determined on the basis of a power study using data simulation; details can be obtained from the first author.

2.3. Test linking

The advantage of using IRT models is the possibility to separate the influences of item difficulty and ability level on responses (Baker & Kim, 2004). Differences between persons can be assessed independent of what specific items are in the test, so response data from individuals that were tested with different test versions can be analyzed in one analysis (Van den Berg et al., 2014). In order to do that, one needs to first estimate the differences in difficulty for all items in the test versions. This is called test linking.

There have been previous attempts to link the Advanced and Standard forms using raw score test equating methods (Jensen, Saccuzzo, & Larson, 1988; Styles & Andrich, 1993), but there the fit of one Rasch model to all items was not explicitly tested. In this paper we use the Rasch model, which is a well-known IRT model for dichotomous data (Rasch, 1960). The Rasch model assumes local independence, which implies unidimensionality of ability. Local independence means that correlations among items are absent, once controlled for the latent variable. Previous studies show mixed results concerning the dimensionality of the Raven Progressive Matrices. Whereas studies have shown that the RPM is largely unidimensional (Rost & Gebert, 1980), other studies indicates that the RPM might be multidimensional (Lynn, Allik, & Irwing, 2004; Van der Ven & Ellis, 2000; Vigneau & Bors, 2005). However, multidimensionality of intelligence tests has to be assessed with some care, since when items vary widely in difficulty, linear factor models will

Table 1
Item fit measures and parameter values. Items that were used in the extra data collection for test linking are printed in bold.

Item	Outfit MSQ	Infit MSQ	β parameter
A5	4.52	0.95	−6.54
A6	0.49	0.94	−6.54
A7	0.70	0.90	−3.16
A8	1.44	1.19	−2.14
A9	2.23	1.00	−4.51
A10	1.16	1.06	−3.27
A11	1.12	1.08	−0.84
A12	1.22	1.10	0.11
B1	4.52	0.95	−6.54
B2	0.71	0.98	−4.11
B3	4.60	0.86	−5.41
B4	1.01	0.94	−3.91
B5	1.66	1.04	−3.13
B6	1.10	1.07	−1.89
B7	1.17	1.11	−1.35
B8	1.10	0.96	−1.05
B9	1.14	0.84	−1.56
B10	0.59	0.82	−1.87
B11	0.84	0.84	−1.31
B12	1.38	0.98	0.52
C1	2.12	0.95	−6.54
C2	1.80	1.15	−3.00
C3	1.01	1.06	−1.97
C4	1.24	1.14	−1.47
C5	1.37	0.97	−2.14
C6	2.52	1.13	−0.93
C7	0.69	0.79	−1.11
C8	0.99	1.00	−0.29
C9	1.15	1.03	−0.97
C10	1.05	0.95	0.91
C11	1.09	1.04	1.45
C12	2.36	0.81	2.58
D1	1.07	1.01	−5.11
D2	0.72	0.77	−1.97
D3	0.65	0.79	−1.79
D4	0.79	0.89	−1.41
D5	0.41	0.74	−2.59
D6	1.04	0.90	−1.53
D7	0.96	1.01	−0.62
D8	0.98	0.93	−0.84
D9	0.85	0.91	−0.37
D10	0.82	0.86	−0.31
D11	1.24	1.12	1.68
D12	0.95	0.87	3.06
E1	0.97	1.00	−0.02
E2	0.88	0.93	−0.12
E3	0.93	0.97	0.11
E4	0.86	0.84	0.68
E5	0.75	0.80	0.78
E6	1.07	0.91	1.09
E7	1.10	0.97	1.35
E8	1.00	0.89	1.66
E9	1.49	0.88	2.72
E10	2.11	0.91	2.91
E11	2.97	1.10	3.21
E12	2.30	1.06	3.20
1	0.17	0.67	−0.90
2	1.79	1.12	−0.69
3	0.96	0.88	0.03
4	0.81	0.93	0.66
5	0.65	0.85	0.59
6	2.79	0.90	−0.51
7	0.72	1.00	0.24
8	1.98	0.89	0.14
9	0.59	0.96	0.42
10	0.49	0.87	0.51
11	0.60	0.88	−0.21
12	0.63	0.93	0.73
13	1.67	1.00	1.93
14	0.77	0.88	0.66
15	0.73	0.97	0.59
16	0.79	0.94	1.65
17	1.08	1.04	1.77
18	1.10	1.04	2.60
19	1.29	1.08	1.81
20	1.13	1.08	1.93
21	0.76	0.90	2.32
22	0.94	0.99	3.10
23	0.77	0.89	2.77
24	0.90	0.94	2.82
25	1.28	1.16	3.37
26	1.05	1.07	2.65
27	0.97	1.04	3.03
28	1.08	1.06	3.19
29	0.93	0.97	3.85
30	1.26	1.05	2.53
31	0.80	0.91	2.48
32	1.27	1.08	4.05
33	1.16	1.06	3.43
34	0.75	0.85	3.41
35	0.97	0.91	3.96
36	1.18	0.81	5.27

generally show several factors, one for each difficulty level (Gibson, 1959). A study that reported multidimensionality also reported that the dimensions in the RPM are highly correlated, around 0.90 (Lynn et al., 2004), which supports unidimensionality. Therefore in this study the Rasch model will be assumed as an appropriate model for this data. For further investigation about multidimensionality in the data see section 3.1 Test linking.

2.3.1. Linking design

The above data sets, adult data on the APM, children data on the SPM and adult data on a subset of items from the APM and SPM allows for test linking. Using an Item Response Theory model, the differences in difficulty among the 36 items from the APM can be estimated on the basis of the adult data from Van Leeuwen et al. (2008), the differences in difficulty among all 60 SPM items can be estimated based on the twin data, and the data from the 49 extra participants can be used as extra information on differences in difficulty among SPM items and among APM items, but also to estimate difficulty differences between items that come from different test versions. The IRT model used was the one-parameter logistic model, also known as the Rasch model. This model was fitted to the full data set containing 60 (SPM) plus 36 (APM) equals 96 items. APM item data were treated as missing at random for the twins, SPM items were treated missing at random for the parents, and the items not included in the extra data collection were assumed missing at random for the relevant participants. In the Rasch model the probability of answering an item correct (coded as a 1, rather than a 0) is a function of the difficulty parameter (β) for that particular item and the ability level of the participant that is tested (θ) (Holland & Wainer, 2012):

$$P\left(y_{jk}=1;\theta_j,\beta_k\right)=\frac{e^{\theta_j-\beta_k}}{1+e^{\theta_j-\beta_k}}$$

where y_{jk} is the response of participant j on item k (1 = correct, 0 = incorrect), θ_j is the ability level of participant j, and β_k is the difficulty of item k. This model allows not only for estimating differences in ability level among participants, but also estimating differences in difficulty level among items, without a need for assumptions about the population of participants (that is why a snowballing sampling technique does not cause problems here). The model is identified by fixing the mean ability level to an arbitrary value, say 0, or by fixing the mean difficulty level. Estimating this Rasch model on the data sets described above results in a set of item parameters in such a way that they quantify differences in item difficulty. Conditioning on these estimated difficulty levels, the ability levels of twins and parents can be modelled in the next phase.

2.3.2. Assessing quality of test linking

The question then is to assess to what extent the data linking was successful: does a Rasch model indeed fit both the SPM and APM data and can the equated item parameters be used to quantify differences in ability across subgroups (i.e., twins, their parents, and the extra participants)? To answer this question, Andersen's likelihood ratio (LR) test was carried out to test whether the estimates of the difficulty parameters would be different between groups. First, it was tested whether the differences in item difficulty among the 10 SPM items were different in the twin data than among the respective items in the linking data set. Second, it was tested whether the differences in difficulty among the 16 APM items were the same in the twins' parents as in the extra participants.

Next, overall fit of the Rasch model to all three data sets at once was assessed using outfit and infit MSQ measures (Christensen & Kreiner, 2013). MSQ stands for the mean of the standardized squared residuals. Infit refers to inlier-sensitive fit. Infit is sensitive to the pattern of responses to items targeted on the person, that is, items with a β value close to the θ value of the test-taker. Outfit refers to outlier-sensitive fit. Outfit is sensitive to responses to items with difficulty levels far removed from the ability level of a person. For example, outfit reports overfit when responses are imputed, and underfit for lucky guesses and careless mistakes such as generally observed when very intelligent people have to answer very simple questions. Mean-square fit statistics reflect the relative amount of randomness that is either too high or too low for a given item. Statistically, mean-squares are chi square statistics divided by their degrees of freedom and are therefore always positive. Their expected value is 1. Values less than 1 indicate that item responses are too predictable; values greater than 1 indicate too much unpredictability. MSQs larger than 2 are generally regarded as problematic (www.rasch.org). Test linking analyses were carried out using the eRm package (Mair, Hatzinger, & Maier, 2015).

2.4. Modelling transmission of intelligence

The complete model in this study for the familial transmission of intelligence consists of two parts: the measurement model and the biometric model. The link between the two models is formed by parameter θ that represents an intelligence score that explains the variation in performance on the observed item data. The covariance structure of this parameter θ within families is then modelled by the biometric model. Thus, as in structural equation modelling (SEM), a distinction is made between the structural model and the measurement model. In the sections below the measurement model, the biometric (i.e. structural) model and the framework of Bayesian estimation are described successively.

2.4.1. Measurement model

For each *k*th item ($k = 1,\ldots,K$) there is one difficulty parameter β_k influencing the response of the *j*th individual from the *i*th twin family with a latent score θ_{ij} ($j = 1.2$; $i = 1,\ldots,N$). The probability of a correct response, p_{ijk} , is modelled as

$$p_{ijk} = \frac{e^{\theta_{ij}-\beta_k}}{1+e^{\theta_{ij}-\beta_k}}.$$

Response y_{ij} is then Bernoulli distributed, $y_{ij}\ Bern(p_{ijk})$.

Since in this study test linking is used to harmonize difficulty parameters across test versions, we will impute the linked values for the difficulty parameter values (see Table 1) into the measurement model (i.e., assuming they are known), so that the scale for θ is identified. The covariance structure of this θ will then be modelled through the structural model, that is, the biometric model.

2.4.2. Biometric model

In quantitative genetic studies where the variance of an observed phenotype is studied, a distinction is made between variation caused by additive genetic effects (*A*) and variation caused by environmental effects (*E*). See Falconer (1960), for an introduction to quantitative genetics. Note that in contrast to standard analyses, here we have a *latent phenotype*, θ, that is identified through the measurement model described above. If one assumes additive genetic effects and environmental effects to be standard normally distributed, we can write as our basic model for family member *j* of family *i:*

$$\theta_{ij} = h \times A_{ij} + e \times E_{ij},$$

where *h* and *e* are the factor loadings for the *A* and E random effects.

In quantitative genetics, the expected genotypic value in offspring is the average of the parental genotypes, $E(A) = \frac{1}{2}(A_{\text{mother}} + A_{\text{father}})$, to which a random term is added, known as the Mendelian sampling term. If mating is random, that is, if phenotypes are uncorrelated in parents, the variance of this Mendelian sampling term equals half the genetic variance. However, intelligence scores of the parents are known to be correlated. Assuming that this correlation is the result of spouse selection that is at least partly based on similarity in intelligence level (*phenotypic assortment*), this resulting similarity in phenotypic values in parents is accompanied with similarity in genotypic value. This leads to an increase of the genetic variance in the next generation, since the genetic variance is the sum of the variance by direct transmission of genetic material plus twice the covariance in genetic effects in the parents, $Var(A_{\text{offspring}})\frac{1}{2} = (Var(A_{\text{father}}) + Var(A_{\text{mother}})) + 2Cov(A_{\text{father}}, A_{\text{mother}})$.

The size of this genetic covariance in the parents depends on how much phenotypic variance is explained by genetic variance, indicated by parameter *h*, and on the size of the spouse correlation, ρ. Under phenotypic assortment, the genetic correlation in parents, γ, equals $\gamma = \rho(h + se)^2$, where *s* is the correlation between genetic effects and environmental effects. Such a correlation *s* can be induced in situations where parent-child correlations are not only due to genetic transmission, but where there is a residual correlation between parents and their children that cannot be explained by genetic effects, a residual correlation that is usually termed cultural transmission, and is usually modelled as a direct regression of the child's environmental effect onto the parental phenotypes:

$$E_{\text{offspring}} \sim N(z(P_{\text{mother}} + P_{\text{father}}), \sigma^2_{\text{Eres}})$$

where *z* is the parameter for the cultural transmission and σ^2_{Eres} is the residual variance of the environmental random effect *E* after the regression on the parental phenotypes. Since parents not only transmit their genes, but can also affect the child's environment, based on their phenotypic value, the genetic effect and the environmental effects become correlated in the child, $Cov(A_{\text{offspring}}, E_{\text{offspring}}) = s \neq 0$. Thus, the total phenotypic variance is larger in the offspring than in the parents due to correlated parental phenotypes and cultural transmission.

However, it seems unlikely that phenotypic assortment and cultural transmission only play a role in data from parents and their children in this generation. The individuals who are parents now are the children of yesterday's parents. Therefore, we expect that in these prior generations similar processes played a role as in the latest generation. Wright (1968) showed that after a limited number of generations, variance becomes stable. If we assume that cultural transmission and phenotypic assortment have been going on for a number of generations, one can therefore assume that all parameters have reached their equilibrium values (Wright, 1968).

Such a model for equilibrium parameter values has been described by Fulker and DeFries (1983) and has often been applied to twin-parent data and more complex family designs. Here we have an added level to

the modelling: this model for both cultural and genetic transmission needs to be combined with an IRT measurement model. Van den Berg et al. (2007) showed how genetic models are best combined with measurement models using Markov-chain Monte Carlo (MCMC) techniques. These are most pragmatically applied through off-the-shelf software packages such as JAGS (Plummer, 2003). Van den Berg (2009) showed how the biometric model described by Fulker and DeFries (1983) can be estimated in such packages. Therefore, we follow the approach described in Van den Berg (2009), except that the observed phenotypic value is replaced by a latent phenotype θ that is equal to $hA_{ij} + eE_{ij}$ and that is identified by the addition of an IRT measurement model.

2.4.3. Bayesian framework

To estimate the parent-offspring model and the IRT measurement model simultaneously, Bayesian statistical modelling is used, as in Van den Berg et al. (2007) and Van den Berg (2009). In the Bayesian framework, inference is based on the posterior probability distribution of the model parameters (e.g., h and e) or functions thereof (e.g., $\frac{h^2}{h^2+e^2}$). The posterior probability distribution is the probability distribution of a parameter or a set of parameters given the data. Such a probability distribution is defined, using Bayes' theorem, as a function of the likelihood function (the distribution of the data given the model parameters) and prior distributions.

The posterior probability distributions can be easily inspected by drawing randomly from these through the use of Markov-chain Monte Carlo (MCMC) methods (Van den Berg, Beem, & Boomsma, 2006). To use off-the-shelf packages like JAGS one only needs to specify the biometric model and specify prior distributions for those parameters that are not a function of other parameters (i.e. h, e, ρ, z and the means). For the structural parameters h, e, ρ, and z we used uniform prior distributions, that is, non-informative priors. Furthermore, for the means in the model, normal prior distributions are used with expectation 0 and variance 10. The software used in this study is JAGS version 3.4.0 (Plummer, 2003) and R version 3.1.0 (R Core Team, 2014). In R the package rjags is used to run JAGS scripts from R (Plummer, 2014). The JAGS script of one of the models (the first one described below) can be found in the Appendix A.

2.4.4. Model comparison

In this paper, various models are compared. The first model includes phenotypic assortment and genetic transmission, but no cultural transmission (Model 1). This was the model that came out as the preferred model in Van Leeuwen et al. (2008). In this model, cultural transmission parameter z is fixed to zero, which leads to s (genotype-environment correlation) also becoming zero. The fit of this model was assessed by posterior predictive checks, where the posterior density of model characteristics is compared to the expected density of those characteristics under the model. The idea is that a posterior predictive distribution is sampled, which is the distribution of future observations that could arise from the fitted model (Lynch, 2007). If a model fits the current data well, future data simulated from the model should show similar features as the current data (Lynch, 2007). Data is simulated from the posterior predictive distribution of θ and compared to the posterior distribution of θ. Different posterior predictive checks were performed, focusing on the sufficient statistics for biometric models: variances and covariances of θ. First, for each iteration the correlation between posterior samples of θ in monozygotic twin pairs, dizygotic twin pairs, mother-father pairings and parent-child pairings were computed. These posterior distributions of correlations were plotted, and compared with the respective correlations based on newly simulated θ values (i.e., the posterior predictive distributions of these correlations). Similarly, the posterior variance of θ in twins and parents were computed and compared with the respective posterior predictive distributions. If the modes of posterior predictive distributions are very close to their respective modes of the posterior distributions, one can infer that the fitted model makes predictions that are actually observed in the data.

Based on these posterior predictive checks, two alternative model improvements were made: one improvement that allowed for different values of h in parents and children (allowing for more genetic variance in one of the generations, Model 2), and one improvement that allowed for different values for e) in parents and children (allowing for more environmental variance in one of the generations, Model 3). In the Results section these improvements are discussed in more detail.

3. Results

3.1. Test linking

The Likelihood Ratio test results are displayed in Figs. 1 and 2. Fig. 1 shows that the 10 SPM items that were assessed both in twins and in the extra participants had similar values for item difficulty (the line represents equality, the red ellipses represent Bonferroni-corrected confidence intervals, i.e. 1–5%/10). All of the confidence ellipses overlapped with the equality line, meaning that none of the differences in item difficulty was different across the two groups. The same was observed for the 16 APM items that were assessed both in the parents and in the extra participants (ellipses represent Bonferroni-corrected confidence intervals, i.e. 1–5%/16). In sum, the Likelihood Ratio test results showed that item parameters were not different across groups: the differences in difficulty level among 10 SPM items were the same in the children as in the adults from the extra data collection, and the differences in difficulty level among the 16 APM items were the same in the parents as in the extra adults. In other words, there was no differential item functioning, at least not for the items that were in the overlapping data set. We therefore believe it is reasonable to assume that the SPM is measurement invariant across age groups, and that the APM and SPM assess the same intelligence dimension. A side-note of the results is the

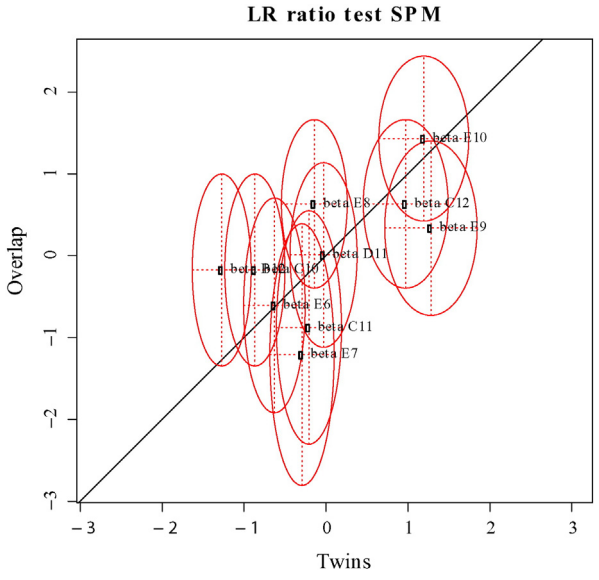

Fig. 1. Confidence intervals for difficulty levels of the SPM 10 items that were administered both to the twin children and the extra 49 participants. The line represents equality, the red ellipses represent Bonferroni-corrected confidence intervals, i.e. 1–5%/10. The different beta values represent the difficulty levels of the corresponding items. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

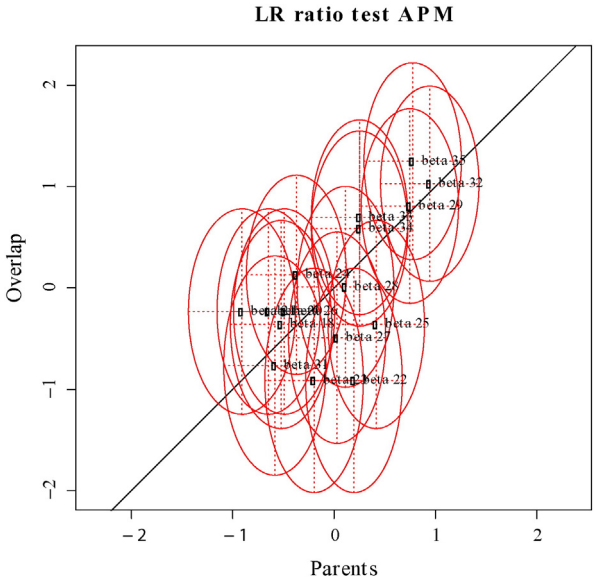

Fig. 2. Confidence intervals for difficulty levels of the 16 APM items that were administered both to the twins' parents and the extra 49 participants. The line represents equality, the red ellipses represent Bonferroni-corrected confidence intervals, i.e. 1–5%/10. The different beta values represent the difficulty levels of the corresponding items. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

possibility that the likelihood ratio tests could be underpowered because of the limited sample size.

Next, the Rasch model was applied to the full data set (three groups combined) for all 96 SPM and APM items. Table 1 shows the estimated item parameters and the infit and outfit MSQ measures. Note that items A1 through A4 are missing, as there were no incorrect responses to these items. Among the STM items, there were 10 with an outfit MSQ larger than 2. This means that answers to these items showed twice as much randomness than expected under the Rasch model. All of these items were either very easy items or very difficult items (see the β values in the third column), and are therefore sensitive to outfit problems that can arise when very intelligent people have to answer very easy questions, and

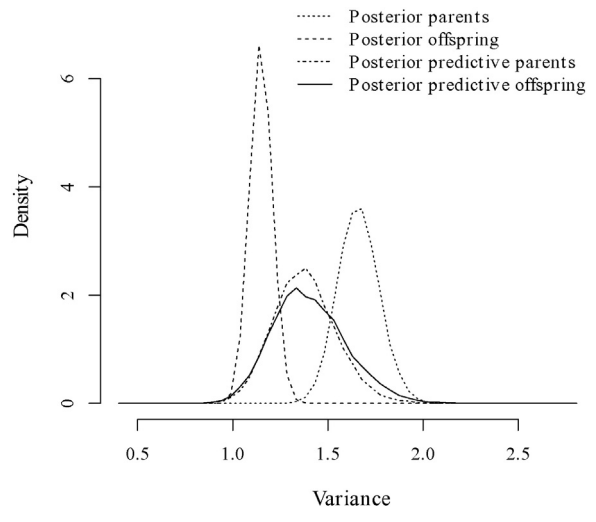

Fig. 3. Posterior and posterior predictive distributions of the variance of intelligence θ of parents and their offspring, based on Model 1.

less intelligent people have to answer very hard questions. Among the APM, only item 6 had an outfit MSQ larger than 2, a relatively easy APM item: most persons give the correct answer anyway so it will not lead to substantial bias. Moreover, outfit problems are less of a threat to measurement than Infit ones (www.rasch.org). There were no items with infit MSQ measures larger than 1.19.

3.2. Transmission of intelligence

Fig. 3 shows the posterior and posterior predictive distributions of the variance of θ of parents and their offspring, based on Model 1.

The posterior predictive distribution of the variance of θ overlap for parents and their offspring, that is, the model predicts similar variances for θ in children and parents. However, the posterior distributions of these variances are not equal, their modes are clearly different. The posterior variance of θ is much larger in parents than in the offspring, and both variances are not predicted well by the model.

Fig. 4a and b show the posterior and posterior predictive distributions of the correlation between θ of monozygotic and dizygotic twins, based on Model 1. In both the posterior distribution and the posterior predictive distribution the correlation between θ in monozygotic twins is higher than the correlation between θ in dizygotic twins (looking at the modes). For monozygotic twins the model predicts a correlation that is a bit lower than the posterior correlation. For dizygotic twins the model predicts a correlation between θ of the offspring similar to its posterior distribution. Fig. 4c shows the posterior and posterior predictive densities of the correlation between θ of father and mother revealing that the model predicts a correlation between θ of the parents and offspring that is a bit lower than the correlation found in the posterior distribution. Finally, Fig. 4d shows that Model 1 predicts a correlation between θ of the parents and θ of the offspring that is a bit higher than the correlation found in the posterior distribution.

In sum, the results of the posterior predictive checks show that Model 1 predicts the four types of familial correlations reasonably well. There is no indication that dominance would lead to better fit (genetic dominance would increase the MZ twin correlation relative to both the DZ twin correlation and the parent-child correlation), nor would cultural transmission (cultural transmission would lead to more similar correlations in MZ and DZ twin pairs and parent-child pairings). However, for the variances of θ there clearly is model misfit.

There are two possible explanations for the difference in variance of θ between parents and offspring. The first one is that the genetic influences of intelligence are different for parents and offspring (different h). Several studies show that the genetic variance component of intelligence is relatively higher for adults than for children (Bouchard & McGue, 2003; Briley & Tucker-Drob, 2013; Patrick, 2000; Plomin & Spinath, 2004; Reynolds, Finkel, & Zavala, 2014). It might be the case that the genetic variance component increases with age also in an absolute sense, thereby increasing total phenotypic variance. The alternative explanation is that the absolute size of the environmental variance component increases with age, so that the value of e is larger in parents than in their children. Two new models were fitted, each including a different improvement of Model 1 based on these two explanations. Model 2 includes different h for parents and offspring and Model 3 includes different e for parents and offspring. These two models were compared to see which model showed best model fit. DIC values were computed for the different models. Model 1 had a DIC of 15,240, Model 2 had a DIC of 15,161 and Model 3 had a DIC of 15,155, so Model 3, with different e for parents and children, showed the lowest DIC value. Posterior predictive checks were next performed to compare model fit more closely.

Figs. 5 and 6 show the posterior and posterior predictive distributions of the variance of θ of parents and their offspring for different h

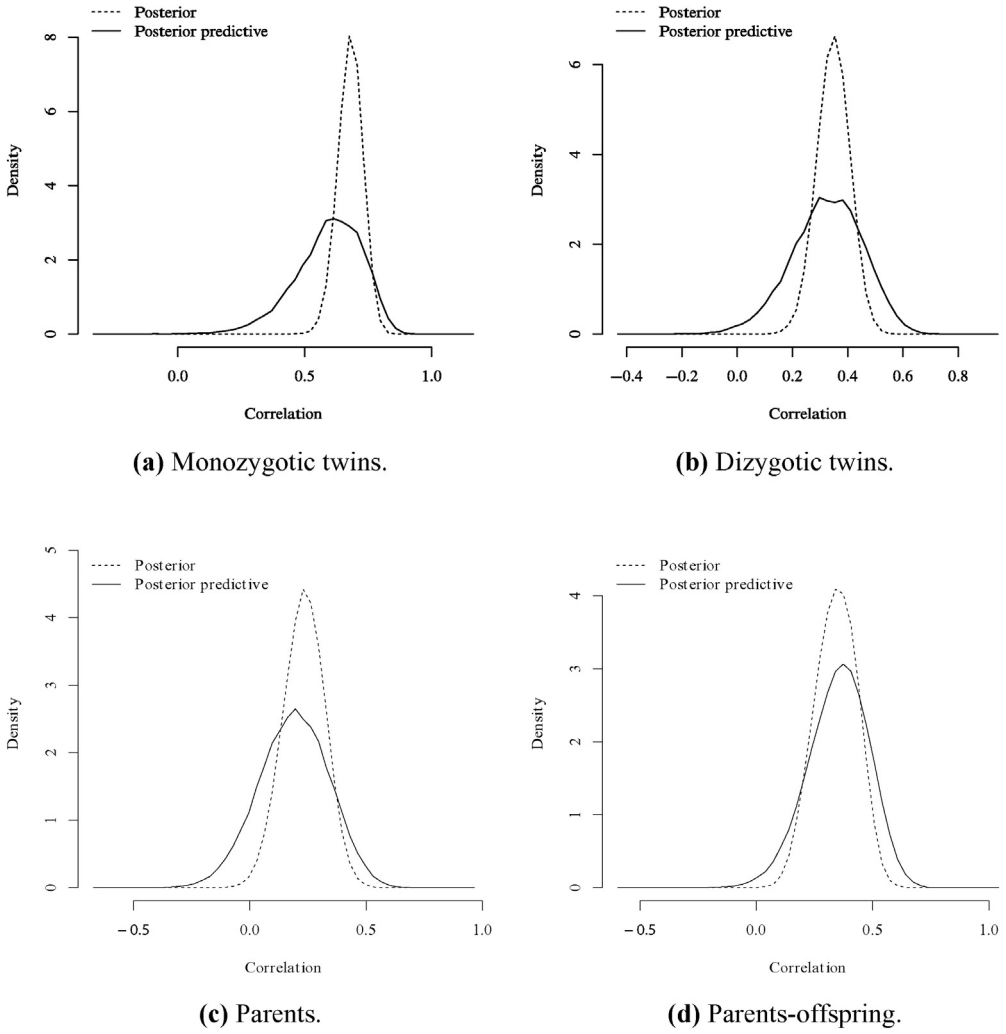

Fig. 4. Posterior and posterior predictive distributions of the correlation between intelligence θ of monozygotic and dizygotic twins, parents and between parents and offspring, based on Model 1.

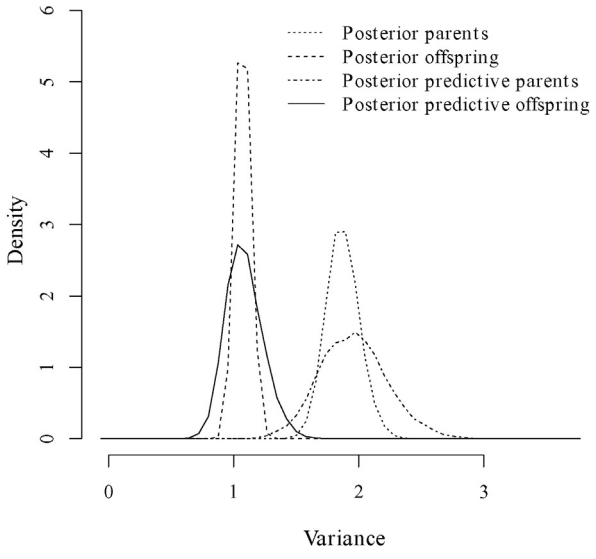

Fig. 5. Posterior and posterior predictive distribution of the variance of intelligence θ of parents and their offspring, based on Model 2.

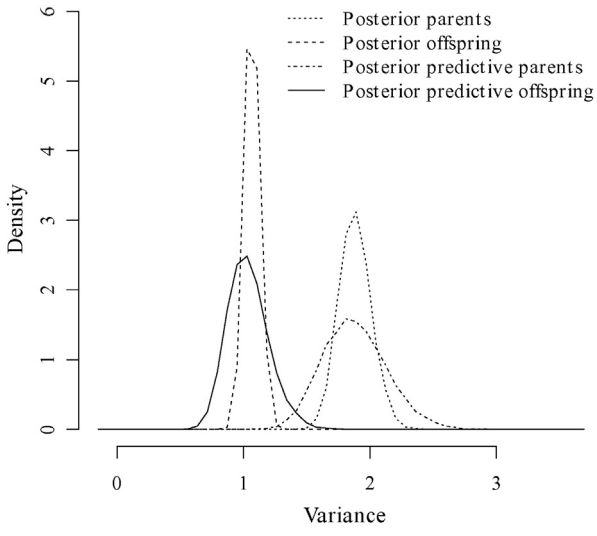

Fig. 6. Posterior and posterior predictive distribution of the variance of intelligence θ of parents and their offspring, based on Model 3.

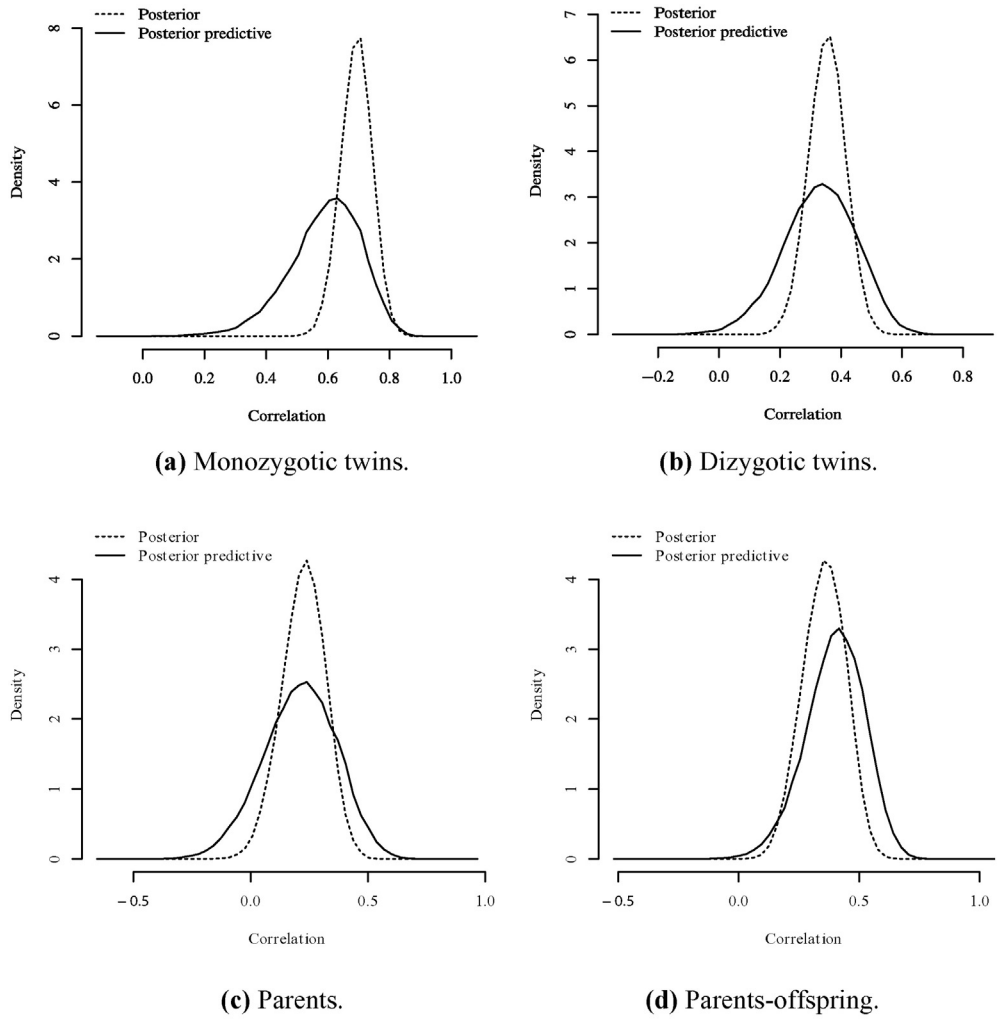

Fig. 7. Posterior and posterior predictive distributions of the correlation between intelligence θ of monozygotic and dizygotic twins, parents and between parents and offspring, based on Model 2.

and e. Fig. 5 shows that for Model 2, the modes of the posterior predictive distributions of the variance of θ are not equal for parents and their offspring. The same pattern is shown in Fig. 6 for Model 3. As these posterior predictive modes almost coincide with the respective posterior modes, both these two models fit the data much better than Model 1.

Fig. 7a, b, c and d show that a model with different parameter h for parents and children (Model 2) makes good predictions for the correlation in parents and in dizygotic twins, but less so for the correlation in monozygotic twins and parent-offspring pairings.

Fig. 8a, b, c and d show that Model 3, with different parameter e for parents and children makes much better predictions regarding all familial correlations.

In sum, both Models 2 and 3 fit much better than Model 1 with regard to the phenotypic variances of parents and twins. Both Models 2 and 3 predict the correlations in parents and dizygotic twins well, but the model with different e for parents and offspring (Model 3) fits the data better compared to the model with different h (Model 2), in terms of reproducing the parent-offspring correlation and monozygotic twin correlations.

For Model 3, the posterior mean of $e^2_{\text{offspring}}$ (i.e., the environmental variance) is 0.31 (SD = 0.056) with 95% credibility interval [0.305, 0.307], the posterior mean of e^2_{parents} is 1.21 (SD = 0.158) with 95% credibility interval [1.203, 1.209] and the posterior mean of h^2 (i.e., genetic variance) is 0.74 (SD = 0.094) with 95% credibility interval [0.734, 0.738].

4. Discussion

Previous studies show mixed results concerning the transmission of intelligence from parents to children: some showed that resemblance is mainly due to the transmission of genes, some showed it is mainly through cultural transmission, and many studies showed that there is a mix of these two processes. However, earlier research is hard to interpret, since in most studies, children and their biological or adoptive parents were tested with different tests. Thus, all results relied on the important assumption that the same phenotype was assessed in both generations. In order to draw definitive conclusions, one prerequisite is that either parents and children are assessed with the same measurement instrument (obviously with comparable reliabilities for both groups), or that one can show evidence that indeed the same phenotype is assessed with the two tests (and control for any differences in test reliability).

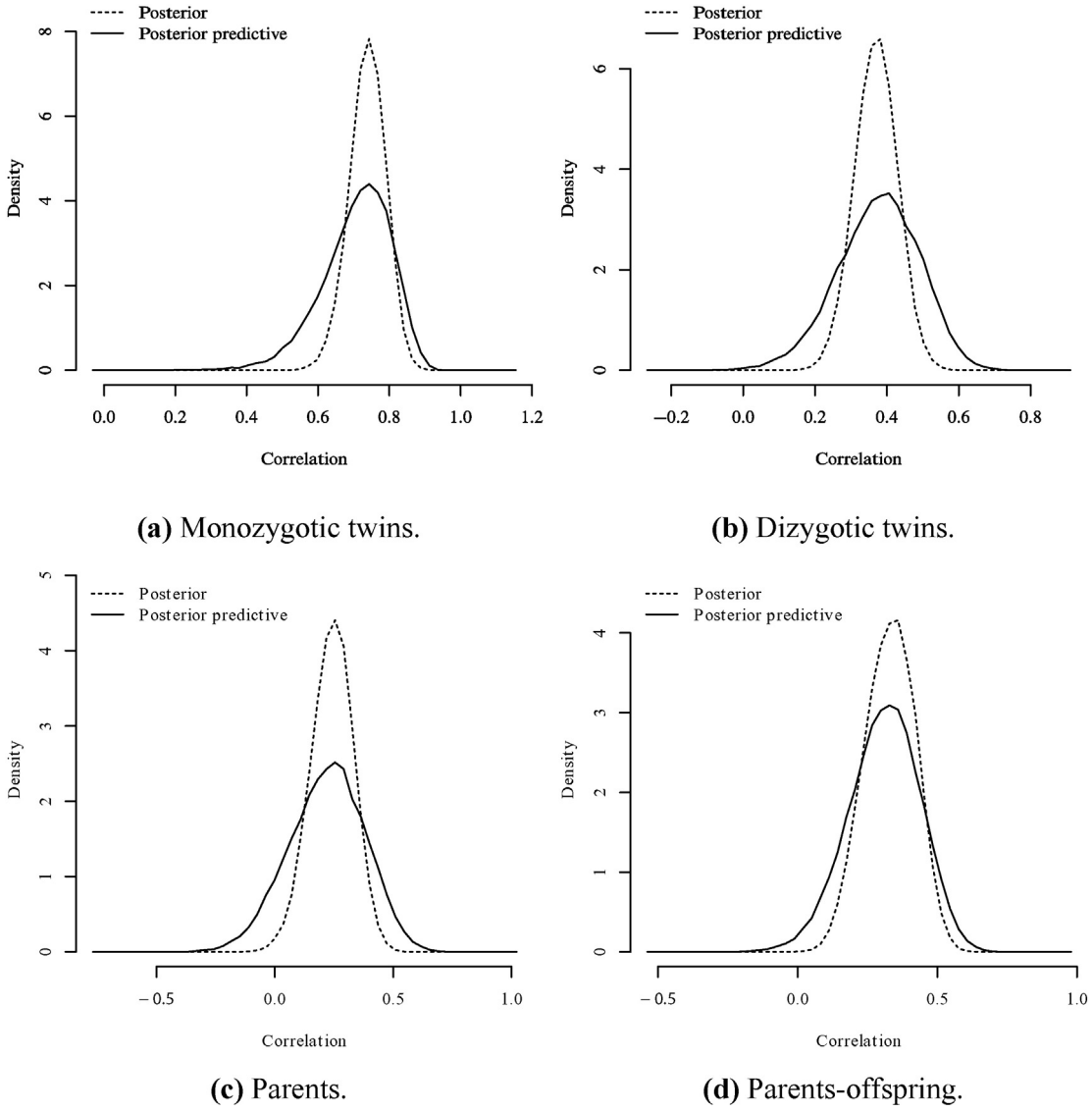

Fig. 8. Posterior and posterior predictive distributions of the correlation between intelligence θ of monozygotic and dizygotic twins, parents and between parents and offspring, based on Model 3.

Here, we re-analyzed data from a study by Van Leeuwen et al. (2008) on Dutch twins and their biological parents, where children plying a test linking analysis, that indeed these two test versions measure the same phenotype. Secondly, by incorporating an IRT-based measurement model, we were able to correct for differences in test reliabilities and we were able to study the covariance structure in twin families. Using this approach we observed a difference in phenotypic variance across the two generations. Note that such variance differences go unnoticed when working with standardized IQ scores. Studying variance differences is important since phenomena such as assortative mating and cultural transmission generally predict an *increase* in variance from one generation to the next. Interestingly, here we found a *decrease* in phenotypic variance. Further model fitting and model checking showed that this larger phenotypic variance in the parents was most likely due to a larger environmental variance component: environmental factors explain more variance in parental intelligence than in children. Such a very simple model of only additive genetic effects, without dominance genetic effects and without cultural transmission, fitted the data nicely, as indicated by predictive posterior checks for the sufficient statistics. Thus, under this model, all similarity in intelligence among family members can be explained by additive genetic effects, whereas environmental effects only contribute to differences among family members.

We found that the model with differently sized environmental variance components for parents and children fitted the data better than the model with different sized genetic variance components for parents and children. There are two possible interpretations of this finding: either this means that in general, intelligence in the adults is more influenced by environmental effects than in children, or that in this particular data set, the environmental component was larger due to the data collection set-up. Regarding the first explanation, previous research shows that the *relative* size of the genetic variance component in intelligence increases with age (Bouchard & McGue, 2003; Plomin & Spinath, 2004) and thus that the *relative* environmental variance component decreases with age. Even though

these studies cannot say anything about the absolute sizes of genetic and environmental variance components, it nevertheless seems unlikely that we found a general increase of environmental variance in parents and not of genetic variance. We feel a more likely option is that the way intelligence was measured in the parents and the children in the Van Leeuwen et al. (2008) study contributed to the increased environmental variance in the parents. In the Van Leeuwen et al. (2008) study, the Raven was assessed in the children under supervision of a research assistant, whereas the Raven was assessed unsupervised in the parents. In the data collection, children were tested during a couple of hours, continuously coached and motivated to perform to the best of their ability, while parents were simply asked to do the Raven test while they were waiting for their children to finish, it being clear that they were not the primary target for data collection. We therefore believe, without any further evidence, that the best explanation of the difference in phenotypic variance in intelligence that we found here is that there was more measurement noise in the parents' data than in the children's data. Therefore, for future research on covariance among family members, it is vital to make sure that not only the same phenotype is studied in all generations (exactly the same, or at least phenotypes that can be linked), but also that test circumstances are exactly the same for all participating family members.

The current analysis was possible because we were able to link two Raven test versions to one common scale. There have been previous attempts to link the Advanced and Standard forms using raw score test equating methods (Jensen et al., 1988; Styles & Andrich, 1993), but there the fit of one Rasch model to all items and was not explicitly tested. The current IRT-based test linking was possible after the collecting of data on a subset of SPM items and a subset of APM items in an extra group of adults. Statistical tests showed there was no significant differential item functioning of these items across groups. Most importantly, the differences in difficulty level of the SPM item in this subset was the same in the children as in the adults. This was a prerequisite to link the APM and the SPM items to one and the same common scale. Further model fitting on the full set of 96 items showed that there were quite a few SPM items that showed large outfit measures. However this is generally to be expected with intelligence tests, that aim to measure differences across a large range. For such tests where intelligent people have to answer easy questions and dull people have to answer hard questions it is to be expected that intelligent people make casual mistakes and that dull people accidentally give the correct answer. Outfit problems are less important than infit problems (www.rasch.org). Fortunately there were no items with problematic infit measures. Thus, apart from problems that are inevitable with a wide range in difficulty level (and ability level), these result support the idea that SPM and APM items measure the same underlying construct. However it should be pointed out that this conclusion is based on an overlapping data set of 49 adults and only 26 items of the 96 items. Future research could look at whether the same conclusions hold when a more varied set of individuals is given a subset of APM and SPM items (i.e., children, adolescents, and adults).

Concluding, there is simple transmission of intelligence from parent to child, consisting of only additive genetic transmission and no cultural transmission. Furthermore, the variance of intelligence of parents and children is not the same. This result is explained by different test circumstances for parents and children.

Acknowledgments

The authors are grateful to Prof. D.I. Boomsma for kindly making the data available and would like to thank Prof. Boomsma and Dr. I. Schwabe for commenting on an earlier draft of the manuscript. The data collection was supported by grants NWO 051-02-060, 668-772, NWO/SPI 56-464-14192, NWO 575-25-006, and NWO 480-04-004.

Appendix A. JAGS script for model 1

Model for monozygotic and dizygotic twins and their parents. Data columns are father items, mother items, twin1 items, twin2 items. Model includes phenotypic assortment, no dominance, and no cultural transmission.

```
model {
# MZ
for (i in 1:Nmz)
{ MZ.covparents[i] ~ dnorm(meanparents, tau.cov)

theta.mz[i,1] ~ dnorm(MZ.covparents[i], tau.parentsres)
theta.mz[i,2] ~ dnorm(MZ.covparents[i], tau.parentsres)

A.MZ.father[i] ~ dnorm(h_inv * theta.mz[i,1], tau.Aparentsres)
A.MZ.mother[i] ~ dnorm(h_inv * theta.mz[i,2], tau.Aparentsres)

A.MZ.offspring[i] ~ dnorm(0.5*(A.MZ.mother[i] + A.MZ.father[i])+ mu.indv, tau.Ares)

E.MZ.offspring1[i] ~ dnorm(0,1)
E.MZ.offspring2[i] ~ dnorm(0,1)
# E.MZ.offspring1[i] ~ dnorm(0,tau.Etwin) #Use only for different e
# E.MZ.offspring2[i] ~ dnorm(0,tau.Etwin) #Use only for different e

theta.mz[i,3] <- e*E.MZ.offspring1[i] + h*A.MZ.offspring[i]
theta.mz[i,4] <- e*E.MZ.offspring2[i] + h*A.MZ.offspring[i]
# theta.mz[i,3] <- e*E.MZ.offspring1[i] + h_twin*A.MZ.offspring[i] #Use only for different h
# theta.mz[i,4] <- e*E.MZ.offspring2[i] + h_twin*A.MZ.offspring[i] #Use only for different h

for (item in 1:NItemsadv) #father items
{       logit(pmz[i,item]) <- theta.mz[i,1]-beta[item]
        MZ[i,item] ~ dbern(pmz[i,item])}

for (item in (NItemsadv+1):(2*NItemsadv)) #mother items
{       logit(pmz[i,item]) <- theta.mz[i,2]-beta[item-NItemsadv]
        MZ[i,item] ~ dbern(pmz[i,item])}

for (item in (2*NItemsadv+1):(2*NItemsadv+NItemsstan)) #twin1 items
{       logit(pmz[i,item]) <- theta.mz[i,3]-beta[item-NItemsadv]
        MZ[i,item] ~ dbern(pmz[i,item])}

for (item in (2*NItemsadv+NItemsstan+1):(2*NItemsadv+2*NItemsstan)) #twin2 items
{       logit(pmz[i,item]) <- theta.mz[i,4]-beta[item-NItemsadv-NItemsstan]
        MZ[i,item] ~ dbern(pmz[i,item])}}
```

```
#DZ
for (i in 1:Ndz)
{ DZ.covparents[i] ~ dnorm(meanparents, tau.cov)

theta.dz[i,1] ~ dnorm(DZ.covparents[i], tau.parentsres)
theta.dz[i,2] ~ dnorm(DZ.covparents[i], tau.parentsres)

A.DZ.father[i] ~ dnorm(h_inv * theta.dz[i,1], tau.Aparentsres)
A.DZ.mother[i] ~ dnorm(h_inv * theta.dz[i,2], tau.Aparentsres)

A.DZ.offspring1[i] ~ dnorm(0.5*(A.DZ.mother[i] + A.DZ.father[i])+mu.indv, tau.Ares)
A.DZ.offspring2[i] ~ dnorm(0.5*(A.DZ.mother[i] + A.DZ.father[i])+mu.indv, tau.Ares)

E.DZ.offspring1[i] ~ dnorm(0,1)
E.DZ.offspring2[i] ~ dnorm(0,1)
# E.DZ.offspring1[i] ~ dnorm(0,tau.Etwin) #Use only for different e
# E.DZ.offspring2[i] ~ dnorm(0,tau.Etwin) #Use only for different e

theta.dz[i,3] <- e*E.DZ.offspring1[i] + h*A.DZ.offspring1[i]
theta.dz[i,4] <- e*E.DZ.offspring2[i] + h*A.DZ.offspring1[i]
# theta.dz[i,3] <- e*E.DZ.offspring1[i] + h_twin*A.DZ.offspring[i] #Use only for different h
# theta.dz[i,4] <- e*E.DZ.offspring2[i] + h_twin*A.DZ.offspring[i] #Use only for different h

for (item in 1:NItemsadv) #father items

{       logit(pdz[i,item]) <- theta.dz[i,1]-beta[item]
        DZ[i,item] ~ dbern(pdz[i,item])}

for (item in (NItemsadv+1):(2*NItemsadv)) #mother items
{       logit(pdz[i,item]) <- theta.dz[i,2]-beta[item-NItemsadv]
        DZ[i,item] ~ dbern(pdz[i,item])}

for (item in (2*NItemsadv+1):(2*NItemsadv+NItemsstan)) #twin1 items
{       logit(pdz[i,item]) <- theta.dz[i,3]-beta[item-NItemsadv]
        DZ[i,item] ~ dbern(pdz[i,item])}

for (item in (2*NItemsadv+NItemsstan+1):(2*NItemsadv+2*NItemsstan)) #twin2 items
{       logit(pdz[i,item]) <- theta.dz[i,4]-beta[item-NItemsadv-NItemsstan]
        DZ[i,item] ~ dbern(pdz[i,item])}}


                #Priors
                meanparents ~ dnorm(0,.1)
                mu.indv ~ dnorm(0,.1)
                e ~ dunif(0,4)
                e2 <- e*e
                h ~ dunif(0,4)
                h2 <- h*h

                # tau.Etwin ~ dgamma(1,.1) #Use only for different e
                totvar <- h2+e2
                h_inv <- h/totvar
                # h_twin ~ dunif(0,4) #Use only for different h
                mu ~ dunif(0,1)
                gamma <- mu*h_inv*h_inv
                tau.cov <- 1/(mu * totvar)
                tau.parentsres <- 1/((1-mu)*totvar)
                tau.Aparentsres <- 1/(1-(h2/totvar))
                tau.Ares <- 1/(0.5-0.5*gamma)}
```

Appendix B. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.intell.2016.06.006.

References

Baker, F., & Kim, S. -H. (2004). *Item response theory: Parameter estimation techniques.* New York: Marcel Dekker.

Bartels, M., Rietveld, M., Van Baal, G., & Boomsma, D. (2002). Genetic and environmental influences on the development of intelligence. *Behavior Genetics*, *32*(4), 237–249. http://dx.doi.org/10.1023/A:1019772628912.

Bouchard, T. J., & McGue, M. (2003). Genetic and environmental influences on human psychological differences. *Journal of Neurobiology*, *54*(1), 4–45. http://dx.doi.org/10.1002/neu.10160.

Briley, D. A., & Tucker-Drob, E. M. (2013). Explaining the increasing heritability of cognitive ability across development: A meta-analysis of longitudinal twin and adoption studies. *Psychological Science*, *24*, 1704–1713. http://dx.doi.org/10.1177/0956797613478618.

Christensen, K. B., & Kreiner, S. (2013). *Item fit statistics, in Rasch models in health.* Hoboken: John Wiley & Sons, Inc.

Core Team, R. (2014). *R: A language and environment for statistical computing [computer software manual].* Vienna: Austria (Retrieved from) http://www.R-project.org/

Eaves, L. J., Heath, A., Martin, N., Maes, H., Neale, M., Kendler, K., ... Corey, L. (1999). Comparing the biological and cultural inheritance of personality and social attitudes in the Virginia 30,000 study of twins and their relatives. *Twin Research*, *2*(2), 62–80. http://dx.doi.org/10.1375/twin.2.2.62.

Falconer, D. S. (1960). *Introduction to quantitative genetics.* (DS Falconer).

Fulker, D. W., & DeFries, J. (1983). Genetic and environmental transmission in the Colorado adoption project: Path analysis. *British Journal of Mathematical and Statistical Psychology*, *36*(2), 175–188. http://dx.doi.org/10.1111/j.2044-8317.1983.tb01123.x.

Gibson, W. A. (1959). Three multivariate models: Factor analysis, latent structure analysis, and latent profile analysis. *Psychometrika*, *24*(3), 229–252. http://dx.doi.org/10.1007/BF02289845.

Giubilei, F., Medda, E., Fagnani, C., Bianchi, V., De Carolis, A., Salvetti, M., ... Stazi, M. A. (2008). Heritability of neurocognitive functioning in the elderly: Evidence from an Italian twin study. *Age and Ageing*, *37*(6), 640–646. http://dx.doi.org/10.1093/ageing/afn132.

Holland, P. W., & Wainer, H. (2012). *Differential item functioning.* Hoboken: Taylor and Francis.

Jensen, A. R., Saccuzzo, D. P., & Larson, G. E. (1988). Equating the standard and advanced forms of the raven progressive matrices. *Educational and Psychological Measurement*, *48*(4), 1091–1095. http://dx.doi.org/10.1177/0013164488484026.

Lynch, S. M. (2007). *Introduction to applied Bayesian statistics and estimation for social scientists.* New York: Springer.

Lynn, R., Allik, J., & Irwing, P. (2004). Sex differences on three factors identified in Raven's standard progressive matrices. *Intelligence*, *32*(4), 411–424. http://dx.doi.org/10.1016/j.intell.2004.06.007.

Mair, P., Hatzinger, R., & Maier, M. J. (2015). eRm: Extended Rasch modeling. 0.15–5 [computer software manual]. (Retrieved from) http://erm.r-forge.r-project.org/

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*(1), 79–108. http://dx.doi.org/10.1016/S0022-4405(99)00038-2.

Plomin, R., & Spinath, F. M. (2004). Intelligence: Genetics, genes, and genomics. *Journal of Personality and Social Psychology*, *86*(1), 112–129. http://dx.doi.org/10.1037/0022-3514.86.1.112.

Plomin, R., Fulker, D. W., Corley, R., & DeFries, J. C. (1997). Nature, nurture and cognitive development from 1 to 16 years: A parent-offspring adoption study. *Psychological Science*, *8*(6), 442–447. http://dx.doi.org/10.1111/j.1467-9280.1997.tb00458.x.

Plummer, M. (2003). *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.*

Plummer, M. (2014). *rjags: Bayesian graphical models using MCMC [Computer software manual].* Retrieved from http://CRAN.R-project.org/package=rjags (R package version 3–13).

Posthuma, D., De Geus, E. J., Bleichrodt, N., & Boomsma, D. I. (2000). Twin-singleton differences in intelligence? *Twin Research*, *3*(2), 83–87. http://dx.doi.org/10.1375/twin.3.2.83.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Danish Institute for Educational Research.

Raven, J. (2000). The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology*, *41*(1), 1–48. http://dx.doi.org/10.1006/cogp.1999.0735.

Raven, J., Raven, J. C., & Court, J. H. (1998a). *Manual for Raven's progressive matrices and vocabulary scales.* Oxford: Oxford Psychologists Press.

Raven, J., Raven, J. C., & Court, J. H. (1998b). *Raven manual section 4: Advanced progressive matrices.* Oxford: Oxford Psychologists Press.

Reynolds, C. A., Baker, L. A., & Pedersen, N. L. (2000). Multivariate models of mixed assortment: Phenotypic assortment and social homogamy for education and fluid ability. *Behavior Genetics*, *30*(6), 455–476. http://dx.doi.org/10.1023/A:1010250818089.

Reynolds, C. A., Finkel, D., & Zavala, C. (2014). Gene by environment interplay in cognitive aging. In D. Finkel, & C. A. Reynolds (Eds.), *Behavior genetics of cognition across the lifespan. Vol. 1.* (pp. 169–199). New York: Springer.

Rijsdijk, F., Vernon, P., & Boomsma, D. (1998). The genetic basis of the relation between speed-of-information-processing and IQ. *Behavioural Brain Research*, *95*(1), 77–84. http://dx.doi.org/10.1016/S0166-4328(97)00212-X.

Rost, D., & Gebert, A. (1980). Zum problem der Faktoreninterpretation bei Raven's coloured progressive matrices [on the problem of factor interpretation in Raven's coloured progressive matrices]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *1*, 255–273.

Scarr, S., & Weinberg, R. A. (1978). The influence of "family background" on intellectual attainment. *American Sociological Review*, 674–692.

Scarr, S., & Weinberg, R. A. (1983). The Minnesota adoption studies: Genetic differences and malleability. *Child Development*, 260–267. http://dx.doi.org/10.2307/1129689.

Styles, I., & Andrich, D. (1993). Linking the standard and advanced forms of the Raven's progressive matrices in both the pencil-and-paper and computer-adaptive-testing formats. *Educational and Psychological Measurement*, *53*(4), 905–925. http://dx.doi.org/10.1177/0013164493053004004.

Tucker-Drob, E. M., & Briley, D. A. (2014). Continuity of genetic and environmental influences on cognition across the life span: A meta-analysis of longitudinal twin and

adoption studies. *Psychological Bulletin*, *140*(4), 949. http://dx.doi.org/10.1037/a0035893.

Van den Berg, S. M. (2009). Imposing nonlinear constraints when estimating genetic and cultural transmission under assortative mating: A simulation study using Mx and BUGS. *Behavior Genetics*, *39*(1), 123–131. http://dx.doi.org/10.1007/s10519-008-9239-7.

Van den Berg, S. M., Beem, L., & Boomsma, D. I. (2006). Fitting genetic models using WinBUGS. *Twin Research and Human Genetics*, *9*, 334–342. http://dx.doi.org/10.1375/twin.9.3.334.

Van den Berg, S. M., Glas, C. A., & Boomsma, D. I. (2007). Variance decomposition using an IRT measurement model. *Behavior Genetics*, *37*(4), 604–616. http://dx.doi.org/10.1007/s10519-007-9156-1.

Van den Berg, S. M., de Moor, M. H., McGue, M., Pettersson, E., Terracciano, A., Verweij, K. J., et al. (2014). Harmonization of neuroticism and extraversion phenotypes across inventories and cohorts in the genetics of personality consortium: An application of item response theory. *Behavior Genetics*, *44*(4), 295–313. http://dx.doi.org/10.1007/s10519-014-9654-x.

Van der Ven, A. H. G. S., & Ellis, J. L. (2000). A Rasch analysis of Raven's standard progressive matrices. *Personality and Individual Differences*, *29*(1), 45–64. http://dx.doi.org/10.1016/S0191-8869(99)00177-4.

Van Leeuwen, M. (2008). *A study of cognition in pre-adolescent twins.* (Doctoral dissertation) Amsterdam, The Netherlands: VU University Amsterdam.

Van Leeuwen, M., Van Den Berg, S. M., & Boomsma, D. I. (2008). A twin-family study of general IQ. *Learning and Individual Differences*, *18*(1), 76–88. http://dx.doi.org/10.1016/j.lindif.2007.04.006.

Vigneau, F., & Bors, D. A. (2005). Items in context: Assessing the dimensionality of ravens advanced progressive matrices. *Educational and Psychological Measurement*, *65*(1), 109–123. http://dx.doi.org/10.1177/0013164404267286.

Wicherts, J. M., & Johnson, W. (2009). Group differences in the heritability of items and test scores. *Proceedings of the Royal Society of London B: Biological Sciences*, *276*(1667), 2675–2683. http://dx.doi.org/10.1098/rspb.2009.0238.

Wright, S. (1968). *Evolution and the genetics of populations.* Chicago: The University of Chicago Press.