Journal of Mathematical Psychology 98 (2020) 102400

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

A comparison of correlation and regression approaches for multinomial processing tree models

Lisa J. Jobst [a,*], Daniel W. Heck [b,c,1], Morten Moshagen [a]

[a] Ulm University, Germany
[b] University of Mannheim, Germany
[c] Philipps-University of Marburg, Germany

ARTICLE INFO

Article history:
Received 30 May 2019
Received in revised form 18 May 2020
Accepted 20 May 2020
Available online 1 June 2020

Keywords:
Multinomial modeling
Individual differences
Parameter correlations
Hierarchical Bayesian modeling
Correction for attenuation

ABSTRACT

Multinomial processing tree (MPT) models are a class of stochastic models for categorical data that have recently been extended to account for heterogeneity in individuals by assuming separate parameters per participant. These extensions enable the estimation of correlations among model parameters and correlations between model parameters and external covariates. The present study compares different approaches regarding their ability to estimate both types of correlations. For parameter–parameter correlations, we considered two Bayesian hierarchical MPT models – the beta-MPT approach and the latent-trait approach – and two frequentist approaches that fit the data of each participant separately, either involving a correction for attenuation or not (corrected and uncorrected individual-model approach). Regarding parameter-covariate correlations, we additionally considered the latent-trait regression. Recovery performance was determined via a Monte Carlo simulation varying sample size, number of items, extent of heterogeneity, and magnitude of the true correlation. The results indicate the smallest bias regarding parameter–parameter correlations for the latent-trait approach and the corrected individual-model approach and the smallest bias regarding parameter-covariate correlations for the latent-trait regression and the corrected individual-model approach. However, adequately recovering correlations of MPT parameters generally requires a sufficiently large number of observations and sufficient heterogeneity.

1. Introduction

Multinomial processing tree (MPT) models are stochastic models that aim at explaining observed categorical data in terms of a finite number of underlying latent processes (Batchelder, 1998). Such models are popular in psychology and beyond because their statistical properties are thoroughly understood (Riefer & Batchelder, 1988), they can easily be adapted to various research paradigms (for reviews, see Batchelder & Riefer, 1999; Erdfelder et al., 2009), and can be used as a measurement device to obtain information about latent processes and states that cannot be observed otherwise (Riefer & Batchelder, 1988). William H. Batchelder played a pioneering role by early identifying the potential of MPT models with respect to the assessment of individual differences and made major contributions to the application of MPT models in the field of cognitive psychometrics (e.g., Batchelder, 2010; Batchelder & Riefer, 2007; Riefer, Knapp, Batchelder, Bamber, & Manifold, 2002). As such, his contributions and more recent extensions of basic MPT models allow for the assessment of individual differences in cognitive processes.

In MPT modeling, it is often of interest to estimate correlations between various MPT parameters or to relate MPT parameters to external covariates (e.g., Arnold, Bayen, & Böhm, 2015; Calanchini, Sherman, Klauer, & Lai, 2014; Heck & Moshagen, 2018; Heck, Thielmann, Moshagen, & Hilbig, 2018; Klein, Hilbig, & Heck, 2017; Müller & Moshagen, 2018). Correlations among cognitive processes are more the rule rather than the exception due to the concept of positive manifold according to which different abilities measured in cognitive tests are generally positively correlated (e.g., Barbey, 2018; van der Maas et al., 2006). Moreover, it is well-known that organismic variables like age can be related to cognitive abilities such as memory performance (e.g., Arbuckle, Gold, Andres, Schwartzman, & Chaikelson, 1992; Naveh-Benjamin, 2000). Obviously, a meaningful analysis of such relations requires the unbiased and reliable estimation of correlations.

We herein investigate the performance of different approaches to estimate bivariate correlations or, more generally, bivariate

* Correspondence to: Research Methods, Institute of Psychology and Education, Ulm University, 89069 Ulm, Germany
E-mail address: lisa.jobst@uni-ulm.de (L.J. Jobst).

1 Supported by the research training group Statistical Modeling in Psychology, funded by the German Research Foundation (GRK 2277).

https://doi.org/10.1016/j.jmp.2020.102400
0022-2496/

associations involving MPT parameters. In particular, we considered two individual-model approaches (often referred to as "no pooling" (e.g., Chechile, 2009)), one of them applying a correction for attenuation, two Bayesian hierarchical approaches ("partial pooling"), the beta-MPT approach (Smith & Batchelder, 2010) and the latent-trait approach (Klauer, 2010), as well as one regression approach, the latent-trait regression (Klauer, 2010). Note that regression approaches genuinely estimate regression slopes, which, however, can be transformed into correlations by means of standardization. In the following, we first describe the basic structure of MPT models and the different approaches for obtaining correlations of MPT parameters. We then present the methods and results of the simulation study and conclude with a discussion of our findings and limitations.

1.1. MPT models

MPT models comprise one or more branches connecting a particular manifest input with a response category yielding the typical tree-like structure of MPT models. These branches represent a sequence of latent states, which usually have a direct interpretation as psychological processes (Batchelder & Riefer, 1999). The probability that a branch i leads to category j is given by:

$$p_{ij}\left(\Theta\right) = c_{ij} \prod_{s=1}^{S} \theta_s^{a_{ijs}} (1-\theta_s)^{b_{ijs}}, \tag{1}$$

where the vector Θ collects S functionally independent parameters $\theta_1, \ldots, \theta_S$, each representing a transition probability with potential values in the interval [0, 1]. Within each branch, c_{ij} is the product of all constants in that branch (e.g., a fixed guessing probability of .50) and the variables a_{ijs} and b_{ijs} count the occurrences of the parameter θ_s and $1 - \theta_s$, respectively (Hu & Batchelder, 1994). The probability of a particular category j is then given by the sum of the probabilities of all branches i terminating in this category.

MPT models are very flexible and can be adapted to many research paradigms. Their specific structure depends on the underlying paradigm. Here, we use the two-high threshold (2HT) model of recognition memory shown in Fig. 1 as a running example (for more details, see Snodgrass & Corwin, 1988). In a typical recognition paradigm, participants have to learn a list of words. After a retention interval, participants are provided with another list of words composed of old words from the learned list (targets) and new words (lures) and are asked to judge whether a presented word is old or new. Target words can lead to two possible observable outcomes: they are either correctly judged as old (hit) or incorrectly judged as new (miss). Similarly, lures can lead to a false alarm if participants incorrectly judge the word as old, and to a correct rejection if participants correctly judge the word as new. According to the 2HT model, two thresholds divide the cognitive processes linking the input items with the four output categories. If an item is recognized as old or new with certainty, it crosses the recognition threshold (for targets) or the rejection threshold (for lures), respectively. The area between both thresholds represents the state of uncertainty, in which case the observed outcome is determined by a guessing process (Bayen, Erdfelder, & Murnane, 1996).

In the 2HT model, each latent process is represented by one parameter: the probability of a correct recognition of targets or a correct rejection of lures is denoted by the parameter d. Conditional on being in a state of uncertainty, which is reached with probability $1-d$, a guessing process is initiated with g being the probability of guessing that an old item has been presented.[2] The 2HT model thus is defined by the following model equations:

$$p(hit \mid old\ item) = d + (1-d) \cdot g, \tag{2}$$

$$p(false\ alarm \mid new\ item) = (1-d) \cdot g. \tag{3}$$

Traditionally, data analysis in the context of MPT modeling relies on maximum likelihood estimation applied on aggregated data (i.e., the sum of response frequencies in each category across trials/items and individuals) leading to parameter estimates on the group level (Hu & Batchelder, 1994). The aggregation of data rests on the assumption that there are no systematic differences in individuals or items regarding the latent processes. However, heterogeneity can occur even in highly standardized research paradigms (e.g., Smith & Batchelder, 2008). Ignoring those differences by assuming independent and identically distributed (*i.i.d.)* observations may lead to biased parameter estimates and standard errors as well as inflated Type I error rates (Klauer, 2010). For this reason, different approaches have been developed to deal with parameter heterogeneity (e.g., Klauer, 2006, 2010; Matzke, Dolan, Batchelder, & Wagenmakers, 2015; Rouder, Lu, Morey, Sun, & Speckman, 2008; Smith & Batchelder, 2010). Those approaches enable the estimation of separate parameters θ_{sn} for the nth individual, in turn allowing for the estimation of correlations involving MPT parameters.

1.2. The individual-model approach

Given the availability of well-established methods for parameter estimation in traditional MPT modeling, an obvious approach is to apply these methods on the individual level to obtain separate parameter estimates for each individual (Chechile, 2009). Thereby, bivariate correlations involving MPT parameters can be computed similarly as for observable variables by using the MPT parameter estimates for each individual. However, estimating one model per individual can be disadvantageous because the precision of parameter estimates hinges on the number of observations per category (i.e., the number of trials or items). Depending on the response behavior and the number of available categories, it is possible that some individuals show an insufficient number of responses in one or more categories. This may in turn lead to very large standard errors or to an empirically underidentified model, meaning that the parameters cannot uniquely be estimated (Moshagen, 2010).

As a consequence of estimation uncertainty, individual MPT parameters are associated with unsystematic measurement error, so that correlations involving such variables are attenuated downwards. Moreover, even for the same MPT parameter, the standard error will usually vary across individuals because estimation accuracy may depend not only on the number of items but also on other parameters of the model. For example, in the 2HT model, the standard error of the guessing probability g increases as the memory performance d increases. To obtain unbiased estimates of correlations, it is thus necessary to correct simple bivariate correlations for attenuation. In case of the correlation between two MPT parameters d and g, the correction for attenuation needs to account for both, the reliability of estimation of the two parameters (r_{dd} and r_{gg}) and for the fact that their estimation errors may be correlated (Zimmerman & Williams, 1977):

$$r_{T_d T_g} = \frac{r_{dg} - r_{E_d E_g}\sqrt{1-r_{dd}}\sqrt{1-r_{gg}}}{\sqrt{r_{dd}}\sqrt{r_{gg}}}, \tag{4}$$

[2] It is also possible to define three parameters by distinguishing between the detection probabilities d for targets and lures. However, identifying such a model requires parameter constraints such as $g = .50$ or alterations of the experimental paradigm.

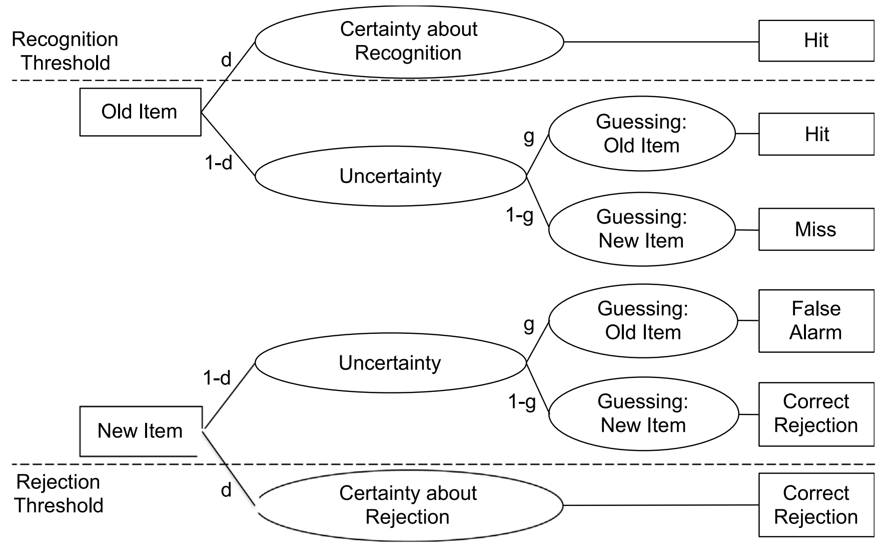

Fig. 1. Visualization of the 2HT model for a classical recognition-test paradigm with the detection parameter d and the guessing parameter g. The latent processes (in ellipses) mediate between the input and the outcome (in boxes).

where r_{dg} is the attenuated, simple correlation of the parameter estimates and r_{EdEg} denotes the correlation between the errors (see below for details on implementation). A similar correction for attenuation can be applied for correlations between an MPT parameter and an external covariate. Given that the reliability of the covariate is often unknown, a conservative approach is to assume perfect reliability for the covariate, so that only uncertainty in the estimated MPT parameter is considered in the correction for attenuation.

1.3. The beta-MPT approach

Unlike the individual-model approach, hierarchical MPT approaches assume that the individual parameters follow a continuous group-level distribution. The beta-MPT approach (Smith & Batchelder, 2010) is based on independent beta distributions as a multivariate hyperdistribution of the individual MPT parameters. The beta distribution is defined on the interval [0, 1] which equals the range of probabilities represented by MPT parameters. Moreover, the two parameters α and β ($0 < \alpha, \beta$) determine the shape of the distribution: the corresponding density function is uniform if α and $\beta = 1$, unimodal if α and $\beta > 1$ and U-shaped if α and $\beta < 1$. The density of the beta distribution for the parameter θ_s is defined by:

$$g\left(\theta_s|\alpha_s, \beta_s\right) = \frac{\Gamma\left(\alpha_s + \beta_s\right)}{\Gamma(\alpha_s) + \Gamma(\beta_s)} \theta_s^{\alpha_s - 1}(1 - \theta_s)^{\beta_s - 1} \tag{5}$$

with $\Gamma(\mathrm{x})$ representing the gamma function. The mean of the beta distribution refers to the average probability of a latent process across individuals (Smith & Batchelder, 2010).

In the Bayesian framework, Markov chain Monte Carlo (MCMC) methods allow to estimate both the individual and group-level parameters simultaneously. Based on the resulting posterior samples, bivariate correlations involving MPT parameters can easily be derived by computing the correlation of the individual MPT parameters within each MCMC iteration. However, given that the beta-MPT approach defines the hyperdistribution by independent beta distributions, the model assumes that MPT parameters are independent across participants. Thus, this approach – by design – is arguably not well-suited to estimate non-zero correlations between different parameters of an MPT model (Smith & Batchelder, 2010). Nonetheless, even if this approach will likely fall short when estimating correlations between MPT parameters, it might still be possible to obtain unbiased correlations to external covariates.

1.4. The latent-trait approach

In contrast to the beta-MPT approach, the latent-trait approach (Klauer, 2010) explicitly incorporates correlated parameter structures. This is achieved by assuming that the probit-transformed individual parameters follow a multivariate normal distribution with mean μ and variance–covariance matrix Σ on the group level. The group-level parameters μ and Σ need to be estimated from the observed data assuming appropriate hyperpriors (for details, see Klauer, 2010; Matzke et al., 2015). The latent-trait approach specifies three different levels: the person-level model linking the parameter vector θ_s of an individual to the MPT category probabilities, the group-level model specifying the distribution of individual MPT parameters, and the hyperpriors enabling the estimation of μ and Σ, so that inter-parameter correlations can be immediately obtained from the covariance matrix Σ. Similarly as in the beta-MPT approach, bivariate correlations of the probit-transformed individual MPT parameters and observable covariates can be computed within each MCMC iteration (e.g., Heck, Arnold, & Arnold, 2018; Ly et al., 2017).

The latent-trait approach can also be used to model a probit regression of the individual MPT parameters on external covariates (Klauer, 2010; Matzke et al., 2015) via

$$\Phi^{-1}\left(\theta_{sn}\right) = \mu_s + \boldsymbol{X}_{sn} \cdot \boldsymbol{\beta}_s + \delta_{sn}, \tag{6}$$

where Φ indicates the cumulative distribution function of the standard normal distribution, θ_{sn} represents the individual MPT parameter s of participant n, $\boldsymbol{X}_{sn}$ denotes a vector of covariate values to predict parameter s of participant n, μ_s is the intercept, $\boldsymbol{\beta}_s$ is a vector containing the regression coefficients, and δ_{sn} represents the random effect of participant n on the respective MPT parameter. The regression slopes express the expected change in the probit value of the sth MPT parameter given a unit-change in the respective covariate, so that – in the case of a single covariate – the bivariate correlation between the covariate and the MPT parameter on probit level can be obtained by standardizing the regression slope.

1.5. The present study

The present simulation study is directed towards a systematic evaluation of the individual-model approach (both with and without correction for attenuation), the beta-MPT approach, and the latent-trait approach with respect to their ability to recover correlations between two MPT parameters as well as between an MPT parameter and an external covariate. Additionally, the latent-trait regression was evaluated regarding its ability to assess parameter-covariate correlations.[3] To this end, we varied the magnitude of the "true", data-generating correlations, the degree of heterogeneity across individuals, the sample size, and the number of items. As a dependent variable, we assessed the recovery performance in terms of absolute bias defined as the difference between the estimated and the true correlation value in the population.

Based on the theoretical foundation of the different approaches, we expected the latent-trait approach to show the smallest recovery bias in conditions with substantial correlations between parameters, because this approach explicitly models and estimates the correlation of the individual parameters by assuming a multivariate normal distribution with the variance–covariance matrix Σ as a free parameter. In contrast, the assumption of independent beta distributions inherent in the beta-MPT approach is at odds with correlated parameters, so a stronger bias is to be expected. We expected the performance of the individual-model approach to depend strongly on the available sample size and number of items due to estimation issues arising from the analysis of sparse data and the corresponding uncertainty in the individual parameter estimates. In any case, the correction for attenuation should improve the performance of the individual-model approach. Regarding parameter-covariate correlations, we expected the latent-trait regression to outperform the remaining approaches because it incorporates the covariate directly into model estimation. Instead of including the covariate into the estimation process, an alternative realized in the beta-MPT approach and the latent-trait approach is to compute parameter-covariate correlations based on posterior estimates (e.g., Heck, Arnold, & Arnold, 2018; Ly et al., 2017), which, however, is likely to lead to biased estimates as the hierarchical structure is not modeled appropriately (Boehm, Marsman, Matzke, & Wagenmakers, 2018). For this reason, we were especially interested to which extent performance differences between the latent-trait regression and the latent-trait approach would occur for sample sizes and number of items commonly used in cognitive research.

2. Methods

A Monte Carlo simulation was conducted to assess the ability of the different approaches to recover the true correlations between two MPT parameters and a parameter and an external covariate, respectively, under varying conditions. As the underlying MPT model, we used the 2HT model as defined in Eqs. (2) and (3).

The population model as well as the experimental conditions were chosen to closely mimic sample sizes and numbers of items (e.g., Erdfelder, Küpper-Tetzel, & Mattern, 2011; Singmann, Kellen, & Klauer, 2013) as well as parameter values (e.g., Dube, Starns, Rotello, & Ratcliff, 2012) commonly occurring in recognition research. In the population, the probit-transformed population values of the two MPT parameters (d and g) and the covariate (Cov) followed a multivariate normal distribution with group-level means $\mu_d = 0.385$, $\mu_g = 0.000$, and $\mu_{Cov} = 0.000$. The true standard deviation of the covariate was $1/\sqrt{2}$, whereas the standard deviation of the probit values of d and g was varied across conditions ($\sigma_d = \sigma_g = 0.10$ versus $\sigma_d = \sigma_g = 0.40$) to analyze the effect of the degree of heterogeneity on recovery performance. The true correlation ρ for the probit-transformed MPT parameters d and g was 0, .30, or .60 and always equaled the correlation between the probit-transformed parameter d and the covariate. The covariate was always uncorrelated to g. To ensure a sufficiently large number of observations in each response category, an equal proportion of targets and lures was used. We manipulated the sample size (N = 50, 100, or 300) and the number of items (M = 50, 100, or 200) to evaluate the different approaches under realistic experimental conditions. Based on the population values on the probit scale, individual probabilities were computed and used to generate data sets comprising the aggregated frequencies in each response category for each individual. These data sets were used as input in the estimation process. The design thus realized 54 experimental conditions (2 levels of heterogeneity × 3 magnitudes of the correlation × 3 sample sizes × 3 numbers of items), which were replicated 250 times each.

To fit the beta-MPT models, we chose a gamma distribution with shape 1.00 and rate 0.10 as prior distribution for the parameters α and β (Heck, Arnold, & Arnold, 2018). Regarding prior distributions for the group-level parameters μ and Σ of the latent-trait approach, we followed the suggestions of Matzke et al. (2015) to rely on weakly informative priors by using standard normal distributions for μ and scaled inverse-Wishart priors for Σ. The scaled inverse-Wishart distribution had $S + 1$ degrees of freedom and an $S \times S$ identity matrix as scale matrix with each scaling parameter ξ_s following a uniform distribution on the interval [0, 10] (Heck, Arnold, & Arnold, 2018). For parameter-covariate correlations estimated by the latent-trait approach and both types of correlations estimated by the beta-MPT approach, a correction was applied to adjust the distribution of MCMC samples for each correlation by accounting for the sampling error depending on sample size (see Ly, Marsman, & Wagenmakers, 2018). This was necessary because the correlation of interest was not directly defined as a model parameter, but rather obtained as a generated quantity based on the MCMC samples of the individual MPT parameters. As a consequence, the distribution of these correlation samples only reflected the posterior uncertainty due to the estimation error of the individual MPT parameters, but not the uncertainty due to the specific number of participants in the sample. As a remedy, the distribution of correlations obtained via MCMC sampling needs to be corrected depending on the sample size N. For this correction, we assumed a scaled uniform distribution on the interval $[-1, 1]$ as a prior distribution for the correlation in both hierarchical approaches.

For all hierarchical approaches, three MCMC chains with different starting values were run in parallel using 10,000 adaptation samples of each chain to increase the sampling efficiency of the MCMC sampler JAGS. We used thinning to reduce autocorrelation and stored only every 10th sample of each chain. All chains ran with at least 50,000 iterations of which the first 10,000 iterations were discarded as burn-in. If the Gelman–Rubin statistic (Gelman & Rubin, 1992) did not indicate convergence (i.e., $\hat{R} \geq 1.05$), the sampling was continued until the convergence criterion was reached.

For the latent-trait approach, the parameter–parameter correlation was estimated based on the posterior samples for the covariance matrix Σ in each iteration. For the beta-MPT approach, we computed the correlation of the probit-transformed posterior samples of the individual parameters in each iteration corrected for sampling error as explained above. In both hierarchical approaches, the posterior median of the posterior samples was then used as an estimate for the parameter–parameter correlation. For the individual-model approach, we relied on the individual maximum likelihood estimates. Because the population was

[3] Another regression approach relating MPT parameters to covariates has been introduced by Coolin, Erdfelder, Bernstein, Thornton, and Thornton (2015). We do not consider this particular approach as it aims to quantify the relative influence of several covariates on MPT parameters assuming no unexplained heterogeneity in the outcome except sampling error, which is beyond the scope of this simulation study.

drawn from parameters on the probit scale, we computed the Pearson correlation coefficient based on the probit-transformed estimated parameters.[4] If extreme probit values (i.e., $\pm$ infinity) occurred in one of the approaches, they were truncated to -5 and 5, respectively. To disattenuate the correlations as by Eq. (4), the reliability r_{ss} of each MPT parameter was estimated as the proportion of true-score variance relative to the total variance of the parameter estimates

$$r_{ss} = \frac{var(\hat{\theta}_s) - \overline{SE}(\hat{\theta}_s)^2}{var(\hat{\theta}_s)}, \tag{7}$$

where *var* denotes the variance of the parameter estimates across individuals and $\overline{SE}(\hat{\theta}_s)^2$ is the mean across individuals of the variance of the individual parameter estimates. The correlation between the error terms $r_{E_d E_g}$ Eq. (4) was obtained as the average over individuals of the off-diagonal element of the inverted and standardized observed Fisher information matrices. We excluded cases in which one or more of the reliability values was below .01 from further analyses to avoid extreme corrections by a factor larger than 10. Furthermore, the disattenuated correlation estimates were truncated to lie in the interval $[-1.00, 1.00]$. A detailed overview of the percentage of excluded and truncated values can be found in Appendix A (Tables A.1 and A.2).

The point estimates for the correlation between d and the covariate were similarly obtained as described above and only differed in some minor aspects from the parameter–parameter correlations: the sample-size correction for the posterior correlation samples (Ly et al., 2018) was used for both the beta-MPT approach and the latent-trait approach because correlations were estimated by correlating the MCMC samples of the individual parameters on the probit scale and the covariate. Regardless of the approach, extreme probit values were again truncated. Regarding the individual-model approach, we assumed perfect reliability of the covariate and independent errors. To estimate the slopes of the latent-trait regression, we z-standardized the covariate and assumed a univariate normal prior with a mean of zero and a variance of v_s for each standardized β_s. For the variance parameter v_s, we chose an inverse gamma prior with shape parameter of 0.5 and scale parameter of 0.25 (Heck, Arnold, & Arnold, 2018). Since the predictor in the regression model was z-standardized, MCMC sampling of β_s provided partially standardized slope estimates. To obtain correlation estimates, we computed completely standardized slope estimates for each MCMC sample by dividing the slope estimate β_s by the square root of the variance of the dependent variable which is given by

$$var\left(\Phi^{-1}(d)\right) = var\left(\beta_s \cdot Cov + \delta\right) = \beta_s^2 \cdot var\left(Cov\right) + var\left(\delta\right), \tag{8}$$

where $var\left(\Phi^{-1}(d)\right)$ denotes the total variance of the individual MPT parameters on the probit scale, *var(Cov)* represents the variance of the covariate, which equals one due to standardization, and $var(\delta)$ is the residual variance estimated by the parameter σ_d^2.

For each approach, we defined the bias as the difference between the correlation estimates and the true, data-generating correlation. As respective point estimate, we considered the median estimate over the 250 replications. All analyses were performed within the R environment (R Core Team, 2018) using the packages TreeBUGS (Heck, Arnold, & Arnold, 2018), MPTinR (Singmann & Kellen, 2013), runjags (Denwood, 2016), and MASS (Venables & Ripley, 2002) as well as the MCMC sampler JAGS (Plummer, 2003).

[4] Note that computing the correlations based on the probability scale yielded virtually identical results.

3. Results

3.1. Mean estimates

We first verified the ability of the different approaches to recover the true group-level means of the parameters d and g. The results for parameter d refer to all analyzed approaches whereas the results for parameter g refer to the respective approaches used to estimate parameter–parameter correlations (i.e., the latent-trait approach, the beta-MPT approach, the uncorrected, as well as the corrected individual-model approach). For the latent-trait approach and the beta-MPT approach, we assessed the posterior median of the respective group-level parameter referring to the mean across individuals, whereas we assessed the group-level average of the individual parameter estimates for d and g in the individual-model approach. For the latent-trait regression, we used the estimate of the intercept μ as a group-level estimate for d on the probit scale. For all approaches, we then defined the median of the respective estimate across all replications as a point estimate for the true mean. All approaches showed very small biases (defined as the difference between the estimated and the true value) ranging from -0.017 to 0.010 for d and from -0.027 to 0.008 for g. The standard deviation of the estimates across the replications of one condition ranged from 0.004 to 0.074 for d and from 0.001 to 0.089 for g.

3.2. Correlation estimates

As shown in Figs. 2 and 3, the ability to recover the true correlation between two MPT parameters strongly depended on the experimental conditions. Generally, virtually no biases were observed if the true correlation was zero regardless of the particular condition and approach used. Furthermore, all approaches indicated a stronger bias in conditions with $\rho > 0$ for both parameter–parameter and parameter-covariate correlations when there was only a minor degree of heterogeneity ($\sigma = 0.10$) compared to a moderate degree of heterogeneity ($\sigma = 0.40$). This can be attributed to the general difficulty of estimating correlations when there is little true-score variance in the first place.

In conditions with non-zero correlations in the population, the latent-trait approach closely recovered the true correlation with biases close to zero if moderate heterogeneity of $\sigma = 0.40$ was present. Also the corrected individual-model approach was able to recover the true correlation and showed almost no biases if a larger number of items was available. In contrast, the beta-MPT approach and the uncorrected individual-model approach revealed substantial biases in all conditions and systematically underestimated the true, positive correlation. For example, with $\sigma = 0.40$, $N = 100$, and $M = 100$, a true correlation of .60 was estimated at .27 by the beta-MPT approach and at .39 by the uncorrected individual-model approach.

Recovery performance generally improved as the number of items increased due to the lower measurement error of the individual MPT parameter estimates. In contrast, increasing sample size rarely had an effect on recovery performance in terms of bias with the exception of the latent-trait approach for which recovery performance improved for larger N. However, larger sample sizes generally resulted in a lower variance of the correlation estimates in all conditions. When comparing the approaches within these conditions, the corrected individual-model approach resulted in the largest variance across replications, followed by the latent-trait approach and the individual-model approach without correction. The beta-MPT approach resulted in the most consistent (though clearly biased) estimates for the correlation. To test whether the applied correction for the sample size N (Ly

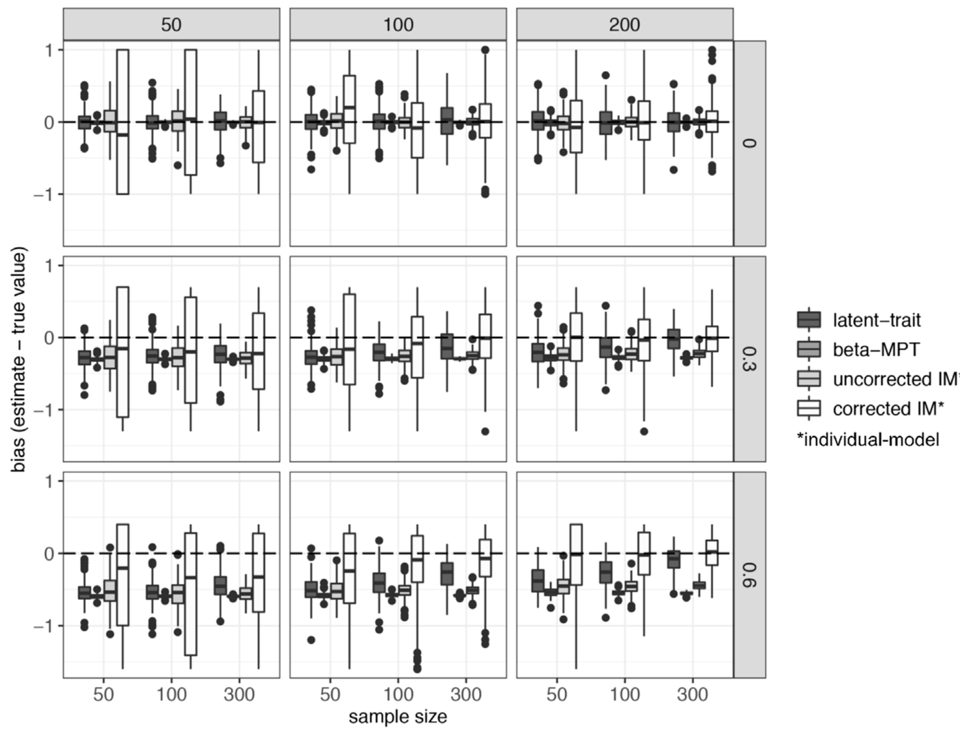

Fig. 2. Difference between the estimated and the true correlation between the two parameters d and g of the 2HT model for a minor level of heterogeneity ($\sigma = 0.10$). The boxplots show the distribution of estimates obtained across 250 replications. The labels on top of the panels indicate the number of items per individual, the labels on the right of the panels indicate the true correlation.

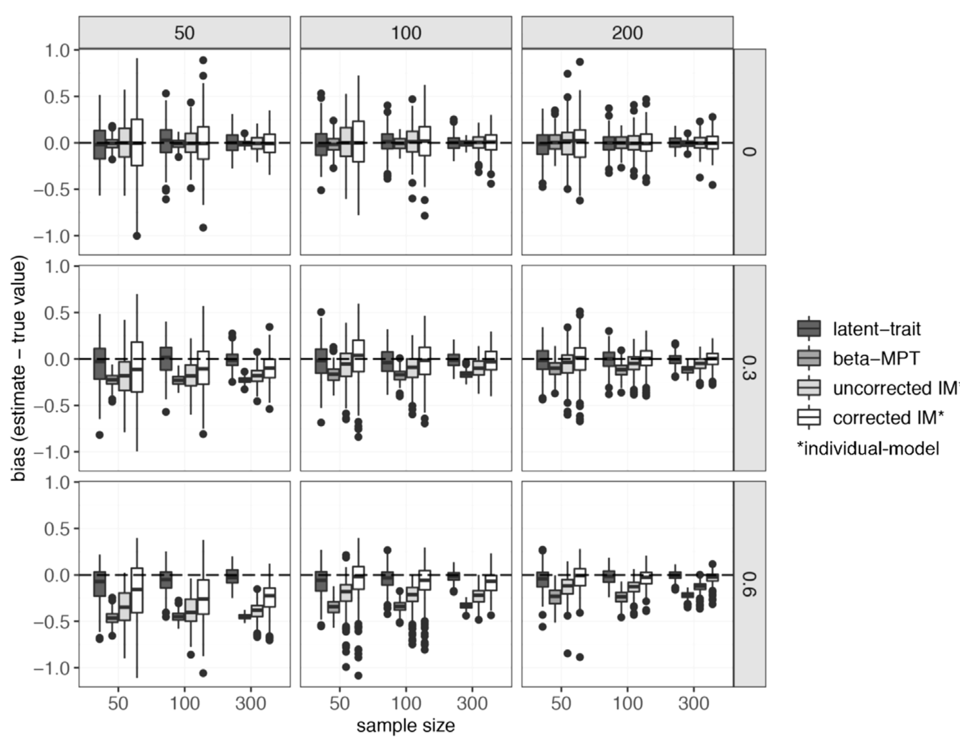

Fig. 3. Difference between the estimated and the true correlation between the two parameters d and g of the 2HT model for a moderate level of heterogeneity ($\sigma = 0.40$). The boxplots show the distribution of estimates obtained across 250 replications. The labels on top of the panels indicate the number of items per individual, the labels on the right of the panels indicate the true correlation.

et al., 2018) led to the small variance of beta-MPT estimates, we also considered the estimates obtained without the correction. However, this led to virtually identical results.

Figs. 4 and 5 show the estimation bias for the correlation of the MPT parameter d and the external covariate. Again, all approaches were virtually unbiased in conditions with $\rho = 0$. However,

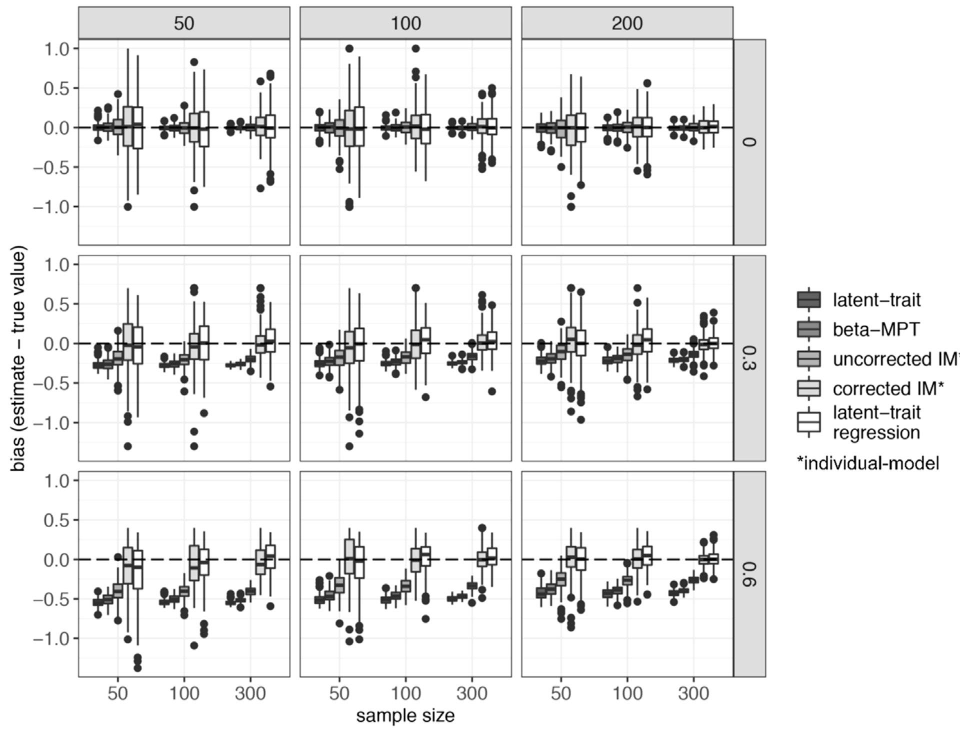

Fig. 4. Difference between the estimated and the true correlation between the parameter *d* of the 2HT model and the covariate *Cov* for a minor level of heterogeneity ($\sigma = 0.10$). The boxplots show the distribution of estimates obtained across 250 replications. The labels on top of the panels indicate the number of items per individual, the labels on the right of the panels indicate the true correlation.

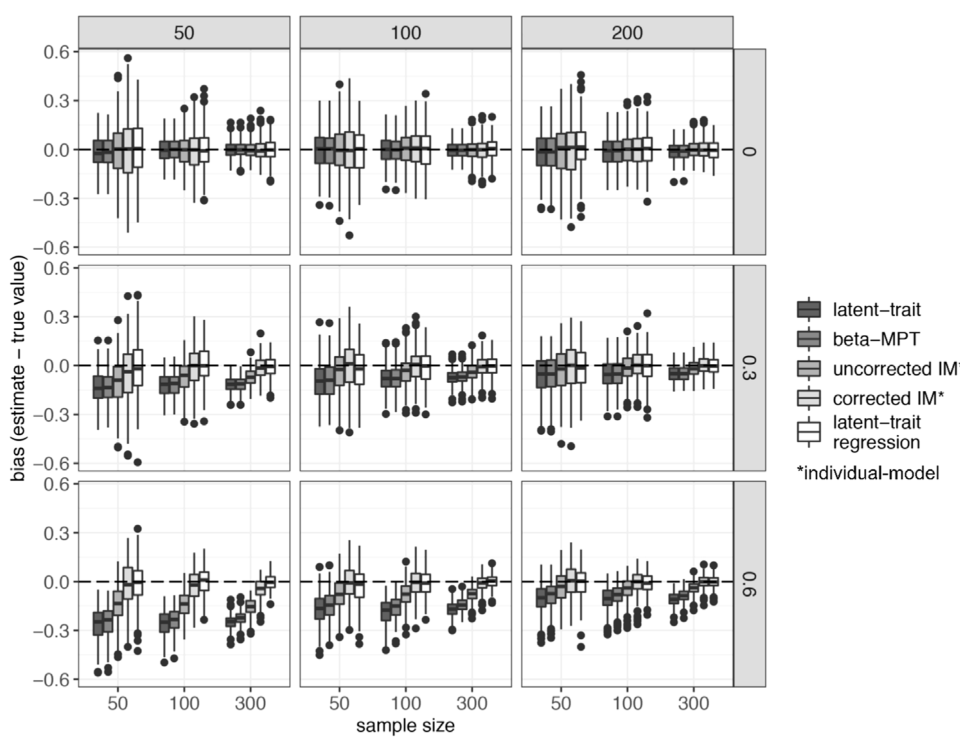

Fig. 5. Difference between the estimated and the true correlation between the parameter *d* of the 2HT model and the covariate *Cov* for a moderate level of heterogeneity ($\sigma = 0.40$). The boxplots show the distribution of estimates obtained across 250 replications. The labels on top of the panels indicate the number of items per individual, the labels on the right of the panels indicate the true correlation.

substantial biases occurred when the MPT parameter and the external covariate were correlated: only the point estimates of the corrected individual-model approach and the latent-trait regression closely recovered the true correlation, whereas all other approaches underestimated the correlation between the MPT parameter *d* and the external covariate to a substantial degree. For instance, with $\sigma = 0.10$, $N = 300$, and $M = 200$, a true correlation of .60 was estimated at .18 by the latent-trait approach, at .21 by the beta-MPT approach, and at .34 by the uncorrected individual-model approach.

As observed for parameter–parameter correlations, a larger number of items improved the accuracy of the approaches. Moreover, larger sample sizes resulted in a reduced variance of the correlation estimates. Within each condition, the smallest bias resulted for the corrected individual-model approach and the latent-trait regression followed by the uncorrected individual-model approach, the beta-MPT approach, and the latent-trait approach. Approaches with lower accuracy were associated with less variance in the estimates: the estimates of the uncorrected individual-model approach, the beta-MPT approach, and the latent-trait approach were less variable compared to those of the corrected individual-model approach and the latent-trait regression. In conditions with a low heterogeneity of $\sigma = 0.10$, reliability of the individual MPT parameter estimates was necessarily low (average r_{dd} = .16, .21, and .32 for $M = 50$, 100, and 200 items), and accordingly, the correction for attenuation substantially increased the variance of correlation estimates compared to the uncorrected individual-model approach by a factor ranging from 3.1 to 14.9 depending on the particular condition. However, this effect almost disappeared in conditions with $\sigma = 0.40$, for which reliability was much higher (average r_{dd} = .62, .78, and .87 for $M = 50$, 100, and 200 items).

To quantify how well the different methods accounted for estimation uncertainty, we assessed the percentage of replications for which the frequentist confidence interval and the Bayesian credible interval, respectively, covered the true correlation value. For parameter–parameter correlations, the latent-trait approach yielded coverage rates close to the expected level of 95% throughout all conditions. The corrected individual-model approach performed similarly well in conditions with a large number of items. For parameter-covariate correlations, the latent-trait regression and the corrected individual-model approach resulted in satisfactory coverage rates across all conditions. The uncorrected individual-model approach yielded comparable coverage rates, but only in conditions with moderate heterogeneity and at least 100 items. The beta-MPT approach and the latent-trait approach performed highly contextual and revealed high coverage rates only for specific conditions. Appendix A provides a detailed overview of the coverage rates for all approaches and conditions (see Tables B.1 to B.3).

Finally, we investigated whether the specific choice of prior distribution for the latent-trait regression contributed to its low bias and the high recovery rates. Hence, we conducted additional simulations based on two different prior widths for the slope parameter of the latent-trait regression. We only considered conditions with $M = 100$ items and defined one wider prior with variance $v_s = 1$ and one narrower prior with variance $v_s = 0.5$ for the standardized regression coefficient (the corresponding implied prior distributions on the correlation can be found in Appendix A, see Fig. A.1). Although we used different prior distributions, virtually identical results were obtained (see Fig. A.2 in Appendix A).

4. Discussion

MPT models provide a powerful method for disentangling latent processes in psychology and may prove especially useful for gaining new insights for measuring individual differences (i.e., cognitive psychometrics; (Riefer et al., 2002). Given that associations between different latent processes or between latent processes and covariates like organismic variables can be expected to be more the rule rather than the exception, it is often of key interest to estimate correlations between different MPT parameters or to relate MPT parameters to external covariates. In the present study, we evaluated the performance of different approaches for recovering stochastic relationships involving MPT parameters, namely the beta-MPT approach (Smith & Batchelder, 2010), the latent-trait approach (Klauer, 2010), the individual-model approach (Chechile, 2009), an attenuation-corrected individual-model approach, as well as the latent-trait regression (Klauer, 2010).

The results indicate that the recovery performance of the approaches under scrutiny depends on several factors. All approaches were unbiased in cases of uncorrelated variables and showed an underestimation tendency for positive correlations, which is consistent with previous simulation studies (e.g., Klauer, 2010; Matzke et al., 2015; Smith & Batchelder, 2010). Unsurprisingly, the performance generally improved with increasing sample size (thus reducing the variance of the correlation estimates) and number of items (thus reducing bias). Also, recovering correlations became increasingly difficult when there was little heterogeneity of the MPT parameters, mirroring the fact that a reliable estimation of correlations requires sufficient variance of the variables in the first place. In rather homogenous samples, a large sample of individuals along with a large number of items is required to yield unbiased and sufficiently reliable results.

Considering the recovery performance regarding correlated parameters, the latent-trait approach – if a moderate degree of heterogeneity was present – and the corrected individual-model approach – if a sufficiently large number of items was available – provided a close recovery of the true correlation. However, the corrected individual-model approach exhibited a large variance of the correlation estimates in conditions with a low level of interindividual heterogeneity. It seems advisable to prefer the latent-trait approach in conditions with substantial interindividual differences as it outperformed the corrected individual-model approach in terms of bias in conditions with few items. Nevertheless, in rather homogeneous samples, the corrected individual-model approach is the best option if a large number of items (to increase estimation accuracy) and large sample sizes (to reduce the variance of the correlation estimates) are available. Concerning parameter-covariate correlations, the latent-trait regression as well as the corrected individual-model approach (for conditions with a sufficiently large number of items) provided a good recovery, albeit both were associated with a substantial variance for small sample sizes.

Regarding the overall performance, the corrected individual-model approach addresses many needs of researchers interested in analyzing correlations in MPT models. Throughout many experimental conditions, this approach was able to adequately recover parameter–parameter as well as parameter-covariate correlations and exhibited satisfactory coverage rates. Compared to the latent-trait approach, the corrected individual-model approach has the advantages that it does not rely on distributional assumptions of the individual parameters at the group level and that it enables faster parameter estimation. However, the correction for attenuation easily breaks down when there is little interindividual variance leading to reliability estimates that are close to zero or even negative (which, in turn, may result in disattenuated correlation estimates larger than one or smaller than minus one), so that a certain extent of individual differences is required. Moreover, as the accuracy and the variance of the estimates depend on the number of items and the sample size, respectively, both variables have to be carefully considered in the design of an experiment. Although the corrected individual-model approach yielded a good recovery performance in most of the analyzed conditions, we therefore recommend caution when the reliability estimate for an MPT parameter is low, which might be due to minor individual differences or to a small number of items. However, if these aspects are addressed, the corrected individual-model approach seems to be a useful tool to analyze correlations in MPT models.

The direct comparison between the latent-trait approach and the latent-trait regression supports the contention that parameter-covariate correlations computed based on MCMC samples of the posterior distribution distinctly underestimate the true value in conditions with non-zero correlations. Even though this underestimation tendency can be reduced with an increasing number of items, the required number of items to obtain unbiased estimates will likely exceed the practical limits in substantive research. The biased estimation of parameter-covariate correlations in the latent-trait approach can be attributed to the fact that the correlation coefficient of interest is not included in the model specification as a free parameter with an appropriate prior distribution (for details, see Boehm et al., 2018). Instead, the correlation is merely computed based on the posterior samples of the individual MPT parameters. We therefore recommend the latent-trait regression be used to estimate correlations between external covariates and MPT parameters as the correlation is explicitly modeled as a free (slope) parameter with a corresponding prior distribution, thus improving estimation accuracy.

An alternative approach we did not consider herein has been suggested by Coolin et al. (2015). This approach uses a probit (or logit) regression of an MPT parameter on one or more external covariates, thus assuming a deterministic relation between the MPT parameter and the covariates. This in turn means that true-score person heterogeneity is completely attributed to variance in the covariates. Hence, unexplained heterogeneity as assumed by the latent-trait model violates the model assumptions, thereby arguably distorting estimation accuracy for the scenario considered in the present simulation. For this reason, this approach is suitable if the most important covariates contributing to almost all of the variance in an MPT parameter are included in the model and if only their relative influence (but not a correlation estimate) is of interest. In our study, we considered only one covariate explaining between 0% and 36% of the variance in the MPT parameter. Given that this approach did obviously not match our research focus, it was not included in the simulation. However, if the model assumptions hold, which can be tested by a goodness-of-fit test, this regression approach may be used to investigate the relative influence of various covariates on the MPT parameters, even though the corresponding estimates cannot be interpreted as or transformed to correlations. Future research should evaluate this approach in comparison with the latent-trait regression in settings where multiple predictors explain most of the variance in an MPT parameter.

Beyond the accuracy of an approach, other application-oriented aspects like convenience and speed of parameter estimation also matter in substantive research. One advantage of the latent-trait approach, the beta-MPT approach, the uncorrected individual-model approach, and the latent-trait regression over the corrected individual-model approach is the ease of application. The open-source software R (R Core Team, 2018) offers different packages – like the package TreeBUGS (Heck, Arnold, & Arnold, 2018) for hierarchical MPT modeling – enabling user-friendly and comprehensive analyses of MPT models. In contrast, the suggested correction for attenuation is not yet implemented in a software package and has therefore to be programmed manually (R scripts are available in the Open Science Framework (OSF) repository at https://osf.io/85duk/). However, implementing the correction for attenuation in Eq. (4) is arguably rather simple, given that standard software for MPT modeling provides the covariance matrix of the individual parameter estimates. Another aspect is the required computing time, which is distinctly faster for the frequentist approaches compared to the Bayesian approaches. In our simulation study, parameter estimation of the frequentist approaches required mere seconds for all experimental conditions, whereas the processing time of the Bayesian approaches could take hours depending on the condition (in particular with large sample sizes).

One limitation of our study is that it is unclear how the choice of the data-generating, multivariate distribution in the population influenced our results. It is possible that the underlying multivariate normal distribution of the population values favored the latent-trait approach because it is based on the same distributional assumptions. However, another explanation for its good performance regarding parameter–parameter correlations can be attributed to the fact that – unlike the other approaches – it explicitly incorporates free parameters for the covariance matrix of the probit-transformed individual MPT parameters. As a consequence, the MCMC sampler directly provides a summary of the posterior distribution of the correlation parameter instead of post-processing the posterior samples of the individual-level parameters (i.e., by computing a correlation for all MCMC iterations). Furthermore, our findings rest on the 2HT model for recognition tests and refer to a population in which the parameter g was uncorrelated with the covariate, and the correlation between d and g always equaled the correlation between d and the covariate. Another limiting aspect concerns the comparability of Bayesian and frequentist estimation methods: whereas the analyzed frequentist approaches use maximum likelihood estimation, the latent-trait regression, the latent-trait approach, and the beta-MPT approach rely on Bayesian statistics. However, for the purpose of assessing parameter recovery (e.g., as in Klauer, 2010), we aimed to minimize the impact of the prior distribution by using weakly informative priors. Nevertheless, it is possible that the specific choice of priors can account for the results, although our corresponding sensitivity simulations did not lead to meaningfully different results depending on the prior.

5. Conclusion

Taken together, different approaches are available for considering individual differences in MPT modeling, each of which enables the estimation of bivariate correlations between model parameters or between a model parameter and an external covariate. The accuracy in terms of bias of these approaches depends on the extent of heterogeneity in particular, on the number of items, and on whether the correlation referred to two MPT parameters or to an MPT parameter and an external covariate. In contrast, the variance of the correlation estimates mainly depends on sample size.

When estimating correlations between MPT parameters, we suggest using the latent-trait approach if there is at least a moderate degree of heterogeneity in participants: it directly models the correlations as free parameters, results in high coverage rates, and is often associated with a comparatively smaller variance of the correlation estimates, in particular with a smaller number of items. However, unless there are only few items, the corrected individual-model approach generally provides a reasonable alternative. In particular, this approach outperformed the latent-trait approach in terms of accuracy when the samples are rather homogenous, but also required large sample sizes to reduce the variance of the correlation estimates. To estimate parameter-covariate correlations, both the latent-trait regression and the corrected individual-model approach showed a high level of accuracy throughout various experimental conditions, but were associated with a large variance in small samples.

Table A.1
Percentage of excluded values per condition in the corrected individual-model approach.

			$N = 50$			$N = 100$			$N = 300$		
Type of correlation	σ	ρ	$M = 50$	$M = 100$	$M = 200$	$M = 50$	$M = 100$	$M = 200$	$M = 50$	$M = 100$	$M = 200$
Parameter–parameter	0.10	0	46.8	34.0	16.4	28.0	22.8	4.4	10.0	4.4	0.0
	0.10	.30	46.8	40.4	17.2	29.2	26.0	6.0	8.0	4.0	0.0
	0.10	.60	44.0	40.0	15.2	29.6	19.2	7.6	9.2	4.8	0.0
Parameter-covariate	0.10	0	28.0	14.8	3.6	19.6	9.6	0.4	8.4	1.2	0.0
	0.10	.30	29.6	20.8	3.2	19.6	7.6	0.8	6.4	0.4	0.0
	0.10	.60	27.2	16.4	4.0	19.6	8.4	0.8	7.6	0.8	0.0

Note. Correlation estimates were excluded if reliability values were below .01 to avoid extreme corrections. In conditions with $\sigma = 0.40$, no estimates were excluded.

Table A.2
Percentage of truncated values per condition in the corrected individual-model approach.

			$N = 50$			$N = 100$			$N = 300$		
Type of correlation	σ	ρ	$M = 50$	$M = 100$	$M = 200$	$M = 50$	$M = 100$	$M = 200$	$M = 50$	$M = 100$	$M = 200$
Parameter–parameter	0.10	0	52.6	22.4	12.9	45.6	16.1	5.9	17.3	3.8	0.4
	0.10	.30	46.6	27.5	16.9	37.3	21.1	8.9	23.9	7.9	0.0
	0.10	.60	48.6	25.3	27.8	42.0	22.3	15.6	25.1	16.0	7.2
	0.40	0	1.6	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0
	0.40	.30	1.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.40	.60	5.6	1.2	0.0	0.4	0.0	0.0	0.0	0.0	0.0
Parameter-covariate	0.10	0	3.9	2.8	1.2	0.5	0.4	0.0	0.0	0.0	0.0
	0.10	.30	8.0	6.1	2.9	3.0	2.6	0.4	1.3	0.0	0.0
	0.10	.60	8.8	16.7	6.7	14.4	6.6	3.2	6.9	1.6	0.0

Note. Disattenuated correlation estimates were truncated to lie in the interval [−1.00, 1.00]. In conditions with $\sigma = 0.40$, no estimates for parameter-covariate correlations had to be truncated.

Table B.1
Coverage rates of parameter–parameter correlations.

			$N = 50$			$N = 100$			$N = 300$		
Approach	σ	ρ	$M = 50$	$M = 100$	$M = 200$	$M = 50$	$M = 100$	$M = 200$	$M = 50$	$M = 100$	$M = 200$
Latent-trait	0.10	0	1.00	.996	1.00	1.00	1.00	.996	1.00	.980	.972
	0.10	.30	1.00	1.00	1.00	1.00	.996	.992	.996	.992	.968
	0.10	.60	.992	.992	.992	.988	.996	.980	.984	.992	.984
	0.40	0	.980	.976	.952	.956	.960	.960	.948	.932	.960
	0.40	.30	.960	.960	.960	.956	.920	.940	.948	.948	.964
	0.40	.60	.960	.964	.940	.968	.940	.964	.940	.948	.952
Beta-MPT	0.10	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.10	.30	.992	.972	.956	.072	.172	.332	.000	.000	.000
	0.10	.60	.000	.000	.000	.000	.000	.000	.000	.000	.000
	0.40	0	1.00	1.00	.988	1.00	1.00	.996	1.00	1.00	1.00
	0.40	.30	.976	.968	.984	.668	.844	.948	.004	.260	.668
	0.40	.60	.012	.220	.576	.000	.012	.196	.000	.000	.000
Uncorrected individual-model	0.10	0	.784	.928	.908	.720	.928	.944	.744	.892	.924
	0.10	.30	.520	.560	.648	.328	.292	.369	.052	.020	.028
	0.10	.60	.116	.060	.112	.012	.000	.004	.000	.000	.000
	0.40	0	.792	.776	.880	.796	.756	.884	.856	.776	.828
	0.40	.30	.644	.820	.908	.528	.740	.908	.224	.520	.788
	0.40	.60	.360	.708	.888	.116	.404	.736	.000	.040	.264
Corrected individual-model	0.10	0	.737	.915	.895	.689	.917	.946	.738	.858	.916
	0.10	.30	.714	.926	.942	.706	.930	.940	.678	.871	.916
	0.10	.60	.807	.893	.929	.716	.906	.948	.678	.929	.940
	0.40	0	.788	.768	.880	.792	.752	.864	.852	.772	.828
	0.40	.30	.748	.816	.920	.748	.800	.912	.780	.732	.888
	0.40	.60	.836	.932	.976	.696	.852	.964	.460	.728	.948

In conclusion, correlations between MPT parameters as well as between an MPT parameter and a covariate can be closely recovered in a Bayesian setting relying on the latent-trait approach and the latent-trait regression, respectively. In the frequentist setting, the corrected individual-model approach offers a reasonable alternative: it is suitable to estimate parameter–parameter correlations (given large numbers of items and individuals) as well as parameter-covariate correlations (given a sufficiently large sample size) with an accuracy similar to the latent-trait approaches, and it allows for much faster parameter estimation compared to the Bayesian approaches relying on MCMC sampling. Generally, however, sufficiently large numbers of participants and items are required in either approach to yield reliable results.

Appendix A

See Tables A.1 and A.2, Tables B.1–B.3 and Figs. A.1 and A.2.

Table B.2
Coverage rates of parameter-covariate correlations for the correlation approaches.

Approach	σ	ρ	$N = 50$			$N = 100$			$N = 300$		
			$M = 50$	$M = 100$	$M = 200$	$M = 50$	$M = 100$	$M = 200$	$M = 50$	$M = 100$	$M = 200$
Latent-trait	0.10	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.10	.30	.992	.988	.988	.368	.572	.752	.000	.000	.036
	0.10	.60	.000	.004	.052	.000	.000	.000	.000	.000	.000
	0.40	0	1.00	.992	.988	1.00	.992	.988	.996	.996	.988
	0.40	.30	.968	.968	.968	.912	.928	.964	.644	.872	.932
	0.40	.60	.524	.792	.912	.124	.444	.800	.000	.052	.340
Beta-MPT	0.10	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.10	.30	.928	.968	.972	.504	.652	.772	.000	.004	.068
	0.10	.60	.000	.016	.124	.000	.000	.000	.000	.000	.000
	0.40	0	1.00	.996	.988	1.00	.996	.988	.992	.996	.984
	0.40	.30	.964	.972	.972	.920	.936	.968	.664	.872	.944
	0.40	.60	.588	.844	.948	.204	.584	.868	.004	.104	.524
Uncorrected individual-model	0.10	0	.948	.932	.960	.976	.960	.964	.948	.936	.944
	0.10	.30	.776	.772	.880	.476	.584	.704	.044	.168	.296
	0.10	.60	.116	.344	.540	.004	.040	.156	.000	.000	.000
	0.40	0	.932	.944	.920	.948	.940	.944	.952	.936	.948
	0.40	.30	.904	.944	.972	.916	.912	.944	.780	.912	.944
	0.40	.60	.856	.960	.984	.680	.912	.960	.124	.704	.920
Corrected individual-model	0.10	0	.950	.934	.963	.975	.969	.964	.948	.935	.944
	0.10	.30	.949	.929	.959	.960	.944	.944	.953	.928	.932
	0.10	.60	.951	.967	.954	.910	.969	.935	.840	.944	956
	0.40	0	.932	.944	.920	.948	.940	.944	.952	.936	.948
	0.40	.30	.932	.944	.968	.928	.948	.972	.972	.956	.952
	0.40	.60	.988	.992	.984	.976	.984	.980	.908	.964	.984

Table B.3
Coverage rates of parameter-covariate correlations for the regression approach.

Approach	σ	ρ	$N = 50$			$N = 100$			$N = 300$		
			$M = 50$	$M = 100$	$M = 200$	$M = 50$	$M = 100$	$M = 200$	$M = 50$	$M = 100$	$M = 200$
Latent-trait regression	0.10	0	.964	.948	.972	.964	.960	.964	.944	.944	.956
	0.10	.30	.976	.944	.936	.996	.980	.936	.956	.968	.952
	0.10	.60	.968	.984	.984	.992	.976	.956	.964	.940	.928
	0.40	0	.968	.992	.944	.944	.940	.972	.940	.936	.940
	0.40	.30	.948	.968	.956	.944	.952	.940	.936	.944	.964
	0.40	.60	.948	.948	.952	.936	.916	.928	.940	.928	.920

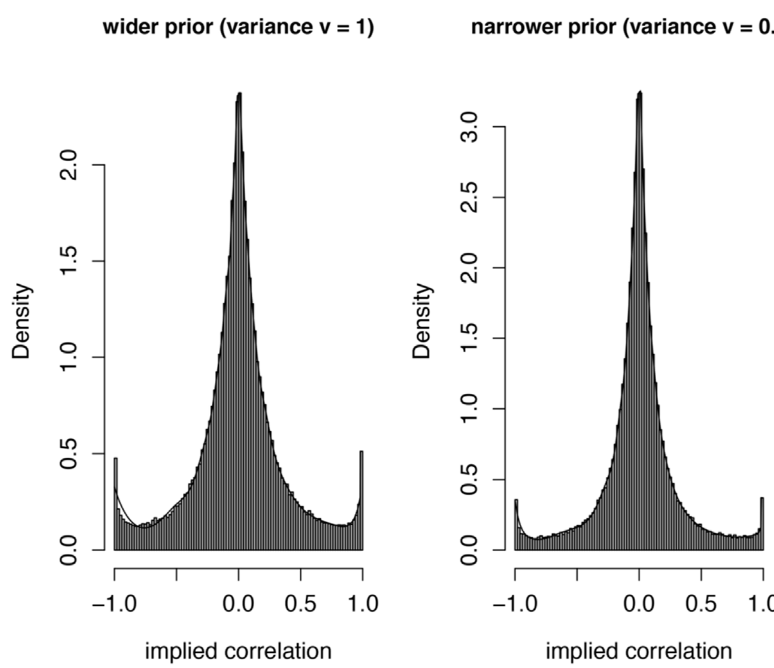

Fig. A.1. Implied prior distributions on the correlation for the latent-trait regression. The distribution was approximated by (1) drawing independent samples from the prior distribution for all model parameters and (2) applying the standardization in Eq. (8) to all samples.

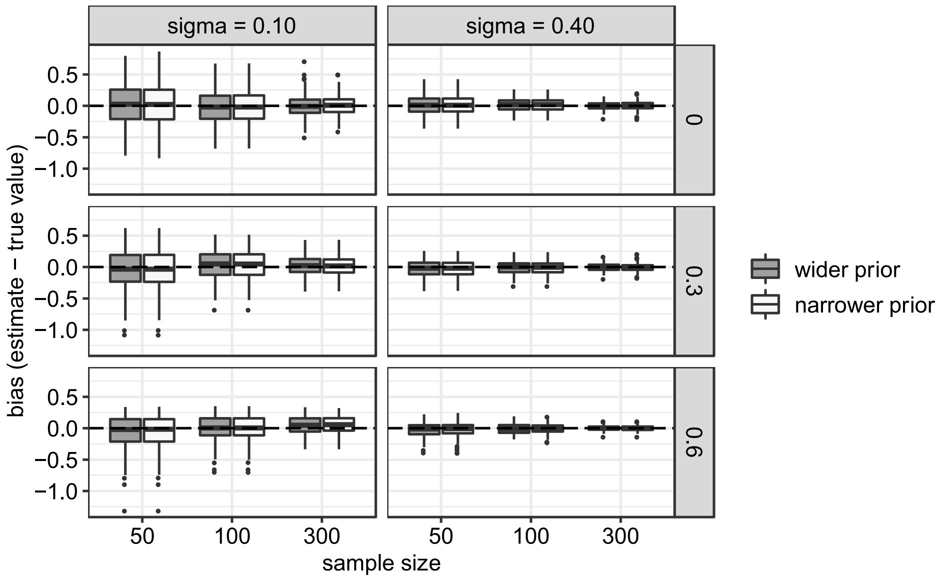

Fig. A.2. Difference between the estimated and the true correlation between the parameter *d* of the 2HT model and the covariate *Cov* obtained by the latent-trait regression based on two different priors (a wider prior with variance $v_s = 1$ and one narrower prior with variance $v_s = 0.5$) in conditions with $M = 100$ items. The boxplots show the distribution of estimates obtained across 250 replications. The labels on top of the panels indicate the extent of heterogeneity, the labels on the right of the panels indicate the true correlation.

References

Arbuckle, T. Y., Gold, D. P., Andres, D., Schwartzman, A., & Chaikelson, J. (1992). The role of psychosocial context, age, and intelligence in memory performance of older men. *Psychology and Aging*, *7*(1), 25–36. http://dx.doi.org/10.1037/0882-7974.7.1.25.

Arnold, N. R., Bayen, U. J., & Böhm, M. F. (2015). Is prospective memory related to depression and anxiety? A hierarchical MPT modelling approach. *Memory*, *23*(8), 1215–1228. http://dx.doi.org/10.1080/09658211.2014.969276.

Barbey, A. K. (2018). Network neuroscience theory of human intelligence. *Trends in Cognitive Sciences*, *22*(1), 8–20. http://dx.doi.org/10.1016/j.tics.2017.10.001.

Batchelder, W. H. (1998). Multinomial processing tree models and psychological assessment. *Psychological Assessment*, *10*(4), 331–344.

Batchelder, W. H. (2010). Cognitive psychometrics: Using multinomial processing tree models as measurement tools. In S. E. Embretson (Ed.), *Measuring psychological constructs: Advances in model-based approaches* (pp. 71–93). American Psychological Association, http://dx.doi.org/10.1037/12074-004.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin and Review*, *6*(1), 57–86.

Batchelder, W. H., & Riefer, D. M. (2007). Using multinomial processing tree models to measure cognitive deficits in clinical populations. In R. W. J. Neufeld (Ed.), *Advances in clinical cognitive science: Formal modeling of processes and symptoms* (pp. 19–50). American Psychological Association, http://dx.doi.org/10.1037/11556-001.

Bayen, U. J., Erdfelder, E., & Murnane, K. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *22*(1), 197–215.

Boehm, U., Marsman, M., Matzke, D., & Wagenmakers, E. -J. (2018). On the importance of avoiding shortcuts in applying cognitive models to hierarchical data. *Behavior Research Methods*, *50*, 1614–1631. http://dx.doi.org/10.3758/s13428-018-1054-3.

Calanchini, J., Sherman, J. W., Klauer, K. C., & Lai, C. K. (2014). Attitudinal and non-attitudinal components of IAT performance. *Personality and Social Psychology Bulletin*, *40*(10), 1285–1296. http://dx.doi.org/10.1177/0146167214540723.

Chechile, R. A. (2009). Pooling data versus averaging model fits for some prototypical multinomial processing tree models. *Journal of Mathematical Psychology*, *53*(6), 562–576. http://dx.doi.org/10.1016/j.jmp.2009.06.005.

Coolin, A., Erdfelder, E., Bernstein, D. M., Thornton, A. E., & Thornton, W. L. (2015). Explaining individual differences in cognitive processes underlying hindsight bias. *Psychonomic Bulletin and Review*, *22*(2), 328–348. http://dx.doi.org/10.3758/s13423-014-0691-5.

Denwood, M. J. (2016). Runjags: An R package providing interface utilities, model templates, parallel computing methods and additional distributions for MCMC models in JAGS. *Journal of Statistical Software*, *71*(9), 1–25. http://dx.doi.org/10.18637/jss.v071.i09.

Dube, C., Starns, J. J., Rotello, C. M., & Ratcliff, R. (2012). Beyond ROC curvature: Strength effects and response time data support continuous-evidence models of recognition memory. *Journal of Memory and Language*, *67*(3), 389–406. http://dx.doi.org/10.1016/j.jml.2012.06.002.

Erdfelder, E., Auer, T. -S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Journal of Psychology*, *217*(3), 108–124. http://dx.doi.org/10.1027/0044-3409.217.3.108.

Erdfelder, E., Küpper-Tetzel, C. E., & Mattern, S. D. (2011). Threshold models of recognition and the recognition heuristic. *Judgment and Decision Making*, *6*(1), 7–22.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. http://dx.doi.org/10.1214/ss/1177011136.

Heck, D. W., Arnold, N. R., & Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods*, *50*(1), 264–284. http://dx.doi.org/10.3758/s13428-017-0869-7.

Heck, D. W., & Moshagen, M. (2018). RRreg: An R package for correlation and regression analyses of randomized response data. *Journal of Statistical Software*, *85*(2), 1–29. http://dx.doi.org/10.18637/jss.v085.i02.

Heck, D. W., Thielmann, I., Moshagen, M., & Hilbig, B. E. (2018). Who lies? A large-scale reanalysis linking basic personality traits to unethical decision making. *Judgment and Decision Making*, *13*(4), 356–371.

Hu, X., & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika*, *59*(1), 21–47. http://dx.doi.org/10.1007/BF02294263.

Klauer, K. C. (2006). Hierarchical multinomial processing tree models: A latent-class approach. *Psychometrika*, *71*(1), 7–31. http://dx.doi.org/10.1007/s11336-004-1188-3.

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: a latent-trait approach. *Psychometrika*, *75*(1), 70–98. http://dx.doi.org/10.1007/s11336-009-9141-0.

Klein, S. A., Hilbig, B. E., & Heck, D. W. (2017). Which is the greater good? A social dilemma paradigm disentangling environmentalism and cooperation. *Journal of Environmental Psychology*, *53*, 40–49. http://dx.doi.org/10.1016/j.jenvp.2017.06.001.

Ly, A., Boehm, U., Heathcote, A., Turner, B. M., Forstmann, B., Marsman, M., et al. (2017). A flexible and efficient hierarchical Bayesian approach to the exploration of individual differences in cognitive-model-based neuroscience. In A. A. Moustafa (Ed.), *Computational models of brain and behavior* (pp. 467–480). Hoboken: John Wiley & Sons, http://dx.doi.org/10.1002/9781119159193.ch34.

Ly, A., Marsman, M., & Wagenmakers, E. -J. (2018). Analytic posteriors for Pearson's correlation coefficient. *Statistica Neerlandica*, *72*(1), 4–13. http://dx.doi.org/10.1111/stan.12111.

van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*(4), 842–861. http://dx.doi.org/10.1037/0033-295X.113.4.842.

Matzke, D., Dolan, C. V., Batchelder, W. H., & Wagenmakers, E. -J. (2015). Bayesian estimation of multinomial processing tree models with heterogeneity in participants and items. *Psychometrika*, *80*(1), 205–235. http://dx.doi.org/10.1007/s11336-013-9374-9.

Moshagen, M. (2010). Multitree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*(1), 42–54. http://dx.doi.org/10.3758/BRM.42.1.42.

Müller, S., & Moshagen, M. (2018). Overclaiming shares processes with the hindsight bias. *Personality and Individual Differences*, *134*, 298–300. http://dx.doi.org/10.1016/j.paid.2018.06.035.

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *26*(5), 1170–1187.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*. Vienna, Austria. Retrieved from https://www.r-project.org/conferences/DSC-2003/Proceedings/Plummer.pdf.

R Core Team (2018). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*(3), 318–339.

Riefer, D. M., Knapp, B. R., Batchelder, W. H., Bamber, D., & Manifold, V. (2002). Cognitive psychometrics: Assessing storage and retrieval deficits in special populations with multinomial processing tree models. *Psychological Assessment*, *14*(2), 184–201. http://dx.doi.org/10.1037/1040-3590.14.2.184.

Rouder, J. N., Lu, J., Morey, R., Sun, D., & Speckman, P. (2008). A hierarchical process-dissociation model. *Journal of Experimental Psychology: General*, *137*, 370–389. http://dx.doi.org/10.1037/0096-3445.137.2.370.

Singmann, H., & Kellen, D. (2013). MPTinR: Analysis of multinomial processing tree models in R. *Behavior Research Methods*, *45*(2), 560–575. http://dx.doi.org/10.3758/s13428-012-0259-0.

Singmann, H., Kellen, D., & Klauer, K. C. (2013). Investigating the other-race effect of Germans towards Turks and Arabs using multinomial processing tree models. In *Proceedings of the annual meeting of the cognitive science society*: *vol. 35*, (35), (pp. 1330–1335).

Smith, J. B., & Batchelder, W. H. (2008). Assessing individual differences in categorical data. *Psychonomic Bulletin and Review*, *15*(4), 713–731. http://dx.doi.org/10.3758/PBR.15.4.713.

Smith, J. B., & Batchelder, W. H. (2010). Beta-MPT: Multinomial processing tree models for addressing individual differences. *Journal of Mathematical Psychology*, *54*, 167–183. http://dx.doi.org/10.1016/j.jmp.2009.06.007.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*(1), 34–50.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York: Springer.

Zimmerman, D. W., & Williams, R. H. (1977). The theory of test validity and correlated errors of measurement. *Journal of Mathematical Psychology*, *16*(2), 135–152.