

# 1 Single construct models

## 1.1 Models and coefficients

Two parallel test halves were build, corresponding to sum scores of half of the trials of the same time point per task. Trials were alternately assigned to the first and the second test half. For tasks repeated 12 times per measurement time point this procedure resulted in two test halves assuming 7 possible values (0 to 6 correctly solved trials), for tasks repeated 8 times per measurement time point test halves could maximally assume 5 possible values (0 to 4 correctly solved trials). Not all categories were observed at all time points and categories had to be collapsed in some cases (see descriptions below). The two test halves served as indicators for a common latent construct per measurement time point, assuming parallel test halves (i.e., factor loadings set to 1 and assuming equal reliabilities). Due to only few observed categories, indicators were modeled as ordered categorical, using a probit link function. The models thereby correspond to Graded response models. For model parsimony, to improve estimation accuracy (see simulation studies) and in order to test for latent mean differences across time, thresholds for all indicators were set equal across time (resulting in the assumption of strict measurement invariance) as well as across test halves.

### 1.1.1 Latent State models

#### Measurement Equation

Measurement equation for parcel  $i$  at time point  $t$ :

$$Y_{it} = S_t + \varepsilon_{it} \quad (1)$$

At each time point  $t$ , a latent state variable  $S_t$ , underlying the two observed indicators  $Y_{1t}$  and  $Y_{it}$  is estimated. Latent state variables are allowed to freely correlate across time, with latent (measurement-error free) correlations serving as indirect indicators of stability across time. The model is depicted for 6 measurement time points in Figure 1.

### 1.1.2 Latent State Trait (LST) models

#### Measurement Equation

Measurement equation for parcel  $i$  at time point  $t$ :

$$Y_{it} = T + S_t + \varepsilon_{it} \quad (2)$$

where  $T$  is a stable latent trait variable,  $S_t$  captures time-specific deviations of the respective true score from the stable trait at time  $t$ , and  $\varepsilon_{it}$  is a measurement error variable, with  $Var(\varepsilon_{it}) = 1 \quad \forall i, t$  (probit parameterization; Graded response model). The model is depicted for 6 measurement time points in Figure 2.

Note that by assuming strong measurement invariance (i.e., loading parameters are set to 1 at all time points, residual variances are equal to 1 by definition of the Graded response model with a probit link, threshold parameters are set invariant across time points, variances of latent state residual factors are set invariant across time points), the specified LST model (without autoregressive effects) corresponds to a multilevel model with a latent trait factor

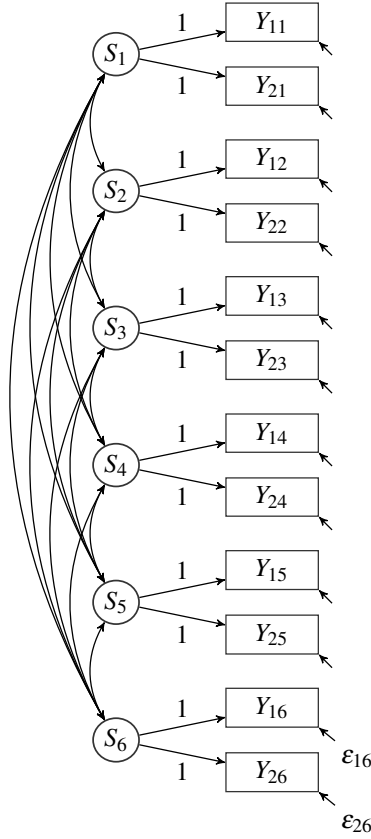


Abbildung 1: Latent State model for two indicators and six measurement time points.

at the between-level (person-level) and a latent state residual factor at the within-level (time-specific) level.

In order to test for possible mean changes across time, latent state models are estimated in a first step. LST models as single-level models are estimated to test whether measurement invariance assumptions across time can be reasonably assumed. Once measurement invariance can be established, the models can alternatively be estimated as multilevel SEMs.

The following variance components can be computed for the presented LST model (without autoregressive effects).

### Consistency

Proportion of true variance (i.e., measurement-error free variance) that is due to true inter-individual stable trait differences.

$$Con(Y_{it}) = \frac{Var(T)}{Var(T) + Var(S_t)} \quad (3)$$

### Occasion specificity

Proportion of true variance (i.e., measurement-error free variance) that is due to true inter-

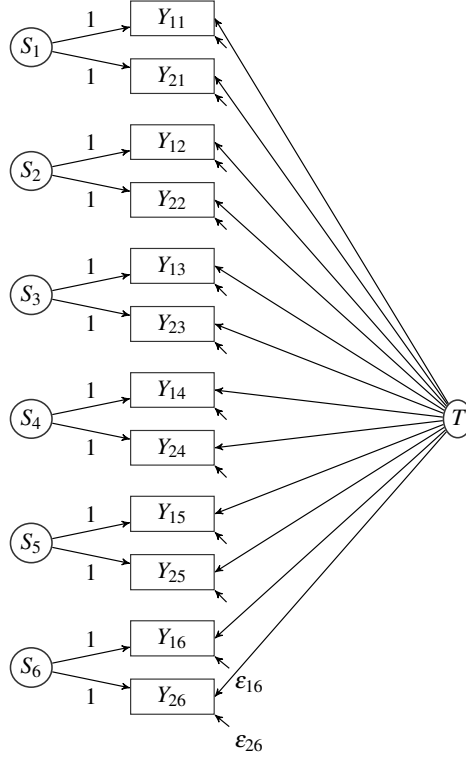


Abbildung 2: Latent State Trait model for two indicators and six measurement time points.

individual differences in the state residual variables (i.e. occasion-specific variation).

$$OS(Y_{it}) = 1 - Con(Y_{it}) = \frac{Var(S_t)}{Var(T) + Var(S_t)} \quad (4)$$

As strong measurement invariance is assumed and  $Var(S_t)$  is set equal across time,  $OS(Y_{it})$  is constant across time as well as across item parcels  $i$ .

### 1.1.3 Latent State Trait (LST) models with autoregressive effects

Model corresponds to model described in Eid, Holtmann, Santangelo, & Ebner-Priemer (2017). The model is depicted for 6 measurement time points in Figure 3.

#### Measurement Equation

Measurement equation for parcel  $i$  at time point  $t$ :

$$Y_{it} = T + O_t + \varepsilon_{it} \quad (5)$$

where  $T$  is a stable latent trait variable,  $O_t$  captures time-specific deviations of the respective true score from the stable trait at time  $t$ , and  $\varepsilon_{it}$  is a measurement error variable, with  $Var(\varepsilon_{it}) = 1 \quad \forall i, t$  (probit parameterization; Graded response model).  $O_t$  is assumed to follow an autoregressive process of order 1 across time (within subjects), that is:

$$O_t = S_t \quad t = 1$$

$$O_t = \beta O_{(t-1)} + S_t \quad t > 1$$

where the latent state residual variables  $S_t$  capture true occasion-specific inter-individual differences that can not be explained by states at previous measurement time points.

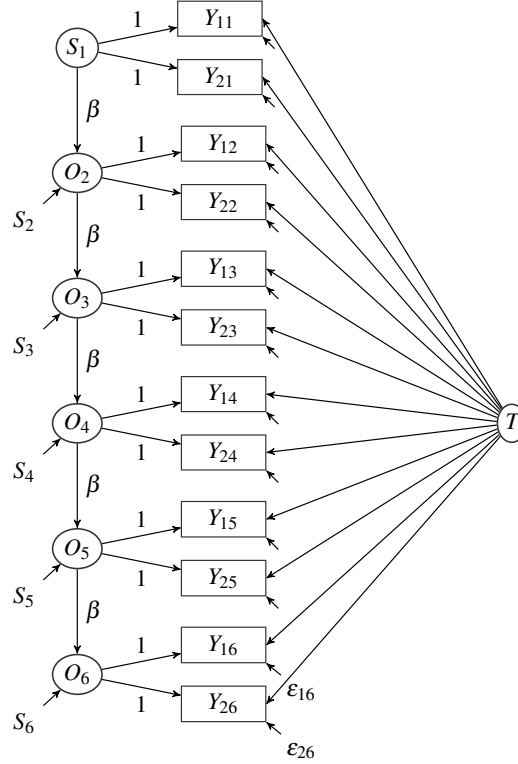


Abbildung 3: Latent State Trait model with autoregressive effects for two indicators and six measurement time points.

The following variance coefficients can be computed.

### Consistency

Proportion of true variance (i.e., measurement-error free variance) that is due to true inter-individual stable trait differences.

$$Con(Y_{it}) = \frac{Var(T)}{Var(T) + \beta^2 Var(O_{(t-1)}) + Var(S_t)} \quad (6)$$

### Occasion specificity

Proportion of true variance (i.e., measurement-error free variance) that is due to true inter-individual differences in the state residual variables, that is occasion-specific variation that is not explained by the autoregressive process.

$$OS(Y_{it}) = \frac{Var(S_t)}{Var(T) + \beta^2 Var(O_{(t-1)}) + Var(S_t)} \quad (7)$$

As the proportion of variance explained by the autoregressive process stabilizes over time, all coefficients have converged to a relatively stable value at  $t = 14$ , indicating the long-term

proportions of variance that are to be expected.

### **Predictability**

Proportion of true variance that is explained by carry-over effects from previous measurement time points.

$$Pred(Y_{it}) = \frac{\beta^2 Var(O_{(t-1)})}{Var(T) + \beta^2 Var(O_{(t-1)}) + Var(S_t)} \quad (8)$$

## **1.2 Estimation**

Models were estimated with MPlus version 8.4, using Bayesian Markov-Chain Monte-Carlo sampling, with the Mplus default priors (also see simulation studies). Using inverse gamma priors [IG(0.001, 0.001); see simulation study] for LST models did not substantially change the parameter estimates. Therefore, only the results with respect to the MPlus default priors are reported. We used two chains with a minimum of 10,000 iterations per chain, with a thinning of 10 (corresponds to a minimum of 100,000 drawn samples per chain of which every 10th is used for the construction of the posterior distribution). The first half of each chain is discarded as burn-in. Convergence was assumed and estimation stopped when the Potential Scale Reduction (PSR) factor well below a threshold of 1.01 for the first time after the minimum number of iterations was reached.

## **1.3 Results**

### **1.3.1 Causality**

The first two categories (0 and 1 correctly solved trials) were collapsed into one category due to sparsity. Thresholds could not be set equal for indicator 2 at time point 3 and 11 as well as indicator 1 at time points 4 and 12 due to a different number of observed categories for the respective indicator and time-point combination. Latent means can still be compared across time for the respective state factors based on the respective other indicator. At time point 7 thresholds of both indicators could not be set measurement invariant across time (due to a divergent number of observed categories) and latent mean differences with respect to the latent state variable at time point 7 should be interpreted with caution.

#### **Latent State model:**

The model fit the data well, with a Posterior Predictive P-value (PPP) of 0.242, and a 95% confidence interval for the difference between the observed and replicated chi-square values of CI = [-74.402; 161.090].

Latent State means varied between -0.204 and 0.569 (with the exception of the 7th time point; latent state mean = -0.922). None of the latent state means differed significantly from zero. That is, no systematic mean change across time is observed. Estimated thresholds corresponded to -2.706, -1.717, -1.080, -0.078, and 0.915.

**Tabelle 1:** Latent variance covariance matrix of latent state variables, LS model for causality

Time point	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1	[.734; .967]												
2	.905	1												
3	.894	.920	1											
4	.902	.939	.908	1										
5	.856	.900	.839	.926	1									
6	.871	.902	.858	.925	.908	1								
7	.857	.884	.820	.907	.931	.918	1							
8	.894	.921	.873	.929	.916	.917	.923	1						
9	.902	.897	.841	.896	.908	.892	.933	.921	1					
10	.869	.869	.823	.881	.907	.890	.931	.906	.918	1				
11	.893	.886	.854	.869	.859	.865	.892	.901	.924	.889	1			
12	.891	.874	.827	.871	.893	.860	.909	.901	.936	.910	.918	1		
13	.865	.878	.835	.875	.884	.859	.897	.897	.917	.884	.915	.918	1	
14	.912	.892	.863	.880	.879	.867	.893	.901	.930	.890	.925	.928	.919	1

Note.

### Latent State Trait model with autoregressive effect (AR(1)):

The model fit the data well, with a PPP of 0.262, and a 95% confidence interval for the difference between observed and replicated chi-square values of CI = [-80.040; 156.563]. Around 82.3% of true interindividual differences at time point 14 go back to stable trait differences, around 10.6% of the interindividual differences can be explained by carry-over effects from previous time points (i.e. inertia in the within-person process) and only 6.1% of the variance is due to time-specific variance between persons, that is, variance in the time-specific true scores from the stable trait level that could not be predicted by the autoregressive process.

**Tabelle 2:** Latent variance covariance matrix of latent state variables, LS model for causality

Parameter	Estimate	Credibility Interval
<b>time point 2</b>		
Consistency	.774	[.248; .920]
Predictability	.162	[.017; .712]
Occasion specificity	.057	[.022; .122]
Reliability	.755	[.642; .841]
<b>time point 14</b>		
Consistency	.823	[.290; .931]
Predictability	.106	[.026; .654]
Occasion specificity	.061	[.026; .123]
Reliability	.731	[.629; .820]
<b>Primary parameters</b>		
Trait variance	2.145	[.704; 3.912]
State residual variance t = 1	0.821	[0.084; 2.896]
State residual variance t > 1	0.166	[0.072; 0.332]
AR	.794	[.527; .965]

Note.

### Latent State Trait model without AR(1):

The model fit the data well, with a PPP of 0.224, and a 95% confidence interval for the difference between observed and replicated chi-square values of CI = [-72.051; 172.452]. The consistency (i.e., stability) coefficient is estimated to be around .903, that is around 90%

of true inter-individual differences are attributable to stable differences between persons,, while approximately 10% are due to variance in time-point specific deviations from the stable trait. Reliability is high with an estimate of .725 (LST model without AR effects) / .733 (LST-AR model).

**Tabelle 3:** Latent variance covariance matrix of latent state variables, LS model for causality

Parameter	Estimate	Credibility Interval
Consistency	.903	[.820; .961]
Occasion specificity	.097	[.039; .179]
Reliability	.725	[.622; .818]
<b>Primary parameters</b>		
Trait variance	2.367	[1.442; 4.165]
State residual variance	0.261	[0.098; 0.472]

*Note.*

### 1.3.2 Inference

Thresholds could not be set equal for indicator 2 at time point 6 as well as indicator 1 at time points 7 and 14, due to a different number of observed categories for the respective indicator and time-point combination. Latent means can still be compared across time for the respective state factors based on the respective other indicator.

#### **Latent State model:**

The model fit the data well, with a PPP of 0.336, and a 95% confidence interval for the difference between observed and replicated chi-square values of CI = [-88.002; 137.279].

Latent means varied between -0.213 and 0.529 (i.e., all falling in the same threshold boundaries as 0), none differing significantly from zero. Ordered thresholds: -2.795, -1.599, -0.715, 0.628, 1.444 and 2.672

**Tabelle 4:** Latent variance covariance matrix of latent state variables, LS model for inference

Time point	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1														
2														
3		.												
4														
5														
6														
7														
8														
9														
10														
11														
12														
13														
14	.880	.722	.584	.800	.782	.809	.788	.871	.795	.845	.901	.899	.925	1

*Note.*

#### **Latent State Trait model with autoregressive effect (AR(1)):**

The model fit the data well, with a PPP of 0.197, and a 95% confidence interval for the

difference between observed and replicated chi-square values of  $CI = [-65.369; 165.291]$ . Around 79.4% of true interindividual differences at time point 14 go back to stable trait differences, around 8.3% of the interindividual differences can be explained by carry-over effects from previous time points (i.e. inertia in the within-person process) and only 11.3% of the variance is due to time-specific variance between persons, that is, variance in the time-specific true scores from the stable trait level that could not be predicted by the autoregressive process.

**Tabelle 5:** Latent variance covariance matrix of latent state variables, LS model for inference

Parameter	Estimate	Credibility Interval
<b>time point 2</b>		
Consistency	.834	[.686; .910]
Predictability	.037	[.003; .193]
Occasion specificity	.120	[.063; .202]
Reliability	.807	[.723; .873]
<b>time point 14</b>		
Consistency	.794	[.567; .888]
Predictability	.083	[.020; .330]
Occasion specificity	.113	[.059; .191]
Reliability	.815	[.746; .876]
<b>Primary parameters</b>		
Trait variance	3.435	[1.940; 6.038]
State residual variance $t = 1$	0.419	[0.029; 1.607]
State residual variance $t > 1$	0.508	[0.255; 0.811]
AR	.648	[.372; .902]

*Note.*

### Latent State Trait model without AR(1):

The model fit the data well, with a PPP of 0.145, and a 95% confidence interval for the difference between observed and replicated chi-square values of  $CI = [-48.420; 183.311]$ .

The consistency (i.e., stability) coefficient is estimated to be around .859, that is around 86% of true inter-individual differences are attributable to stable differences between persons,, while approximately 14% are due to variance in time-point specific deviations from the stable trait. Reliability is high with an estimate of .815.

**Tabelle 6:** Latent variance covariance matrix of latent state variables, LS model for inference

Parameter	Estimate	Credibility Interval
Consistency	.859	[.781; .917]
Occasion specificity	.141	[.083; .219]
Reliability	.815	[.745; .877]
<b>Primary parameters</b>		
Trait variance	3.762	[2.352; 6.432]
State residual variance	0.622	[0.398; 0.905]

*Note.*

### 1.3.3 Gaze following

Only 8 observed trials per measurement occasion. The highest two categories (3 and 4 correctly solved trials) were collapsed into one category due to sparsity. Thresholds could not



be set equal for indicator 2 at time point 9 as well as indicator 1 and indicator 2 at time point 8, due to a different number of observed categories for the respective indicator and time-point combination. Latent means can still be compared across time with the exception of time point 8.

### Latent State model:

The model fit the data well, with a PPP of .535, and a 95% confidence interval for the difference between observed and replicated chi-square values of  $CI = [-124.040; 111.499]$ .

Latent state means varied between -0.990 and -2.153 (for  $t = 8$  the latent state mean -2.77, but thresholds of this time point are not measurement invariant). All of the latent state means are significantly different from zero! That is, we observe a decrease in average ability after the second time point.

Thresholds correspond to -1.204, 0.057, and 1.163.

*Table 7: Latent variance covariance matrix of latent state variables, LS model for gaze.*

Time point	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1														
2														
3		.												
4														
5														
6														
7														
8														
9														
10														
11														
12														
13														
14	.783	.815	.810	.823	.935	.942	.881	.908	.931	.906	.901	.854	.904	1

*Note.*

### Latent State Trait model with autoregressive effect (AR(1)):

The model fit the data well, with a PPP of .485, and a 95% confidence interval for the difference between observed and replicated chi-square values of  $CI = [-114.891; .126.697]$ .

Around 85.5% of true interindividual differences at time point 14 go back to stable trait differences, around 3.3% of the interindividual differences can be explained by carry-over effects from previous time points (i.e. inertia in the within-person process) and only 10.9% of the variance is due to time-specific variance between persons, that is, variance in the time-specific true scores from the stable trait level that could not be predicted by the autoregressive process.

The state residual variance at  $t = 1$  is estimated with great uncertainty (large credible interval). Maybe better use LST model without autoregressive effects (predictability is very low in this model).

**Tabelle 8:** Latent variance covariance matrix of latent state variables, LS model for gaze following.

Parameter	Estimate	Credibility Interval
<b>time point 2</b>		
Consistency	.512	[.286; .728]
Predictability	.419	[.189; .665]
Occasion specificity	.064	[.018; .145]
Reliability	.843	[.763; .901]
<b>time point 14</b>		
Consistency	.855	[.716; .955]
Predictability	.033	[.010; .088]
Occasion specificity	.109	[.033; .217]
Reliability	.759	[.640; .850]
<b>Primary parameters</b>		
Trait variance	2.668	[1.455; 5.029]
State residual variance $t = 1$	9.177	[3.986; 22.435]
State residual variance $t > 1$	0.349	[0.087; 0.761]
AR	.490	[.334; .636]

*Note.*

### Latent State Trait model without AR(1):

The model fit the data well, with a PPP of .360, and a 95% confidence interval for the difference between observed and replicated chi-square values of  $CI = [-99.091; 136.892]$ .

The consistency (i.e., stability) coefficient is estimated to be around .758, that is around 76% of true inter-individual differences are attributable to stable differences between persons,, while approximately 24% are due to variance in time-point specific deviations from the stable trait. Reliability is high with an estimate of .792.

**Tabelle 9:** Latent variance covariance matrix of latent state variables, LS model for gaze following

Parameter	Estimate	Credibility Interval
Consistency	.758	[.622; .863]
Occasion specificity	.242	[.137; .378]
Reliability	.792	[.696; .868]
<b>Primary parameters</b>		
Trait variance	2.855	[1.576; 5.424]
State residual variance	0.924	[0.505; 1.507]

*Note.*

### 1.3.4 Quantity

The lowest three categories (0, 1 and 2 correctly solved trials) were collapsed into one category due to sparsity. Thresholds could not be set equal for indicator 1 at time point 5, due to a different number of observed categories for the respective indicator and time-point combination. Latent means can still be compared across time.

### Latent State model:

The model fit the data well, with a PPP of .485, and a 95% confidence interval for the difference between observed and replicated chi-square values of  $CI = [-103.635; 119.705]$ .

Latent state means varied between -0.058 and 0.369. None of the state means differs significantly from zero.

Thresholds correspond to -1.364, -0.752, 0.356, and 1.411

**Tabelle 10:** Latent variance covariance matrix of latent state variables, LS model for quantity

Time point	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1														
2														
3														
4														
5														
6														
7														
8														
9														
10														
11														
12														
13														
14	.643	.687	.780	.747	.786	.816	.802	.796	.750	.780	.811	.767	.752	1

*Note.*

### Latent State Trait model with autoregressive effect (AR(1)):

The model fit the data well, with a PPP of .520, and a 95% confidence interval for the difference between observed and replicated chi-square values of CI = [-116.228; .108.538].

Around 95.8% of true interindividual differences at time point 14 go back to stable trait differences, around 0.7% of the interindividual differences can be explained by carry-over effects from previous time points (i.e. inertia in the within-person process) and only 2.9% of the variance is due to time-specific variance between persons, that is, variance in the time-specific true scores from the stable trait level that could not be predicted by the autoregressive process.

**Tabelle 11:** Latent variance covariance matrix of latent state variables, LS model for quantity.

Parameter	Estimate	Credibility Interval
<b>time point 2</b>		
Consistency	.858	[.504; .985]
Predictability	.106	[.000; .472]
Occasion specificity	.026	[.002; .109]
Reliability	.486	[.399; .716]
<b>time point 14</b>		
Consistency	.958	[.603; .997]
Predictability	.007	[.000; .355]
Occasion specificity	.029	[.002; .117]
Reliability	.456	[.342; .588]
<b>Primary parameters</b>		
Trait variance	0.781	[0.419; 1.366]
State residual variance t = 1	0.485	[0.055; 1.560]
State residual variance t > 1	0.024	[0.002; 0.099]
AR	.471	[-.361; .957]

*Note.*

### Latent State Trait model without AR(1):

The model fit the data well, with a PPP of .508, and a 95% confidence interval for the difference between observed and replicated chi-square values of CI = [-116.335; 117.757].

The consistency (i.e., stability) coefficient is estimated to be around .964, that is around 96% of true inter-individual differences are attributable to stable differences between persons,, while approximately 3.6% are due to variance in time-point specific deviations from the stable trait. Reliability is rather low with an estimate of .446.

*Tabelle 12:* Latent variance covariance matrix of latent state variables, LS model for quantity.

Parameter	Estimate	Credibility Interval
Consistency	.964	[.860; .998]
Occasion specificity	.036	[.002; .140]
Reliability	.446	[.334; .576]
<b>Primary parameters</b>		
Trait variance	0.768	[0.474; 1.313]
State residual variance	0.029	[0.002; 0.118]

*Note.*

## 2 Multi-construct models

For the combined analysis of several tasks (constructs), we estimated six separate Latent-State-Trait models for the combination of two constructs, respectively. In the combined models, the submodels for each task / construct are equivalent to the LST models described above. For ease of model specification the respective LST models were estimated as multilevel models. As LST models for single tasks did show good fit under the assumption of strict measurement invariance (MI), the single-level LST model with strict MI and the respective multilevel LST model are equivalent.

As detailed information on the parameter estimates obtained in LST models for each separate task is provided above (e.g., estimated stabilities, reliabilities,...), the results reported in the following focus on the latent correlations only. Additional parameters that are of interest in the models combining two tasks are latent correlations between a) the traits belonging to the different tasks, indicating associations between stable latent cognitive ability as estimated by the different tasks, and b) correlations between occasion-specific state residual variables belonging to the same measurement time point, as an indicator of time-specific associations between latent abilities across the two tasks, above and beyond stable inter-individual differences.

As simulation studies suggested that LST models in which latent correlations between time-specific state residual variables are estimated to be time-point specific (i.e., covariances and variances of state residual variables can freely vary across time) do not show good estimation performance under the given conditions (sample size, ordinal indicator variables, etc.), we chose a model with fixed correlations between state residual variables across time. That is, a model in which it is assumed that associations between latent time-specific cognitive abilities across two different tasks within each time point is equal at all time points. This assumption is considered reasonable in the present application. The respective model showed

good estimation performance under the given conditions in the simulation study. For details on MCMC estimation see section 1.1 of the single-task models. Due to increased model complexity, in the estimation of multi-construct models the minimum number of iterations per Markov chain was increased to 20,000 (with a thinning of 10, that is, 200,000 iterations per chain).

## 2.1 Causality and Inference

PPP= .273, 95% confidence interval for the difference between the observed and replicated chi-square values of CI = [-14.831; 28.641].

Trait correlation estimate  $r = .153$  with CI=[-.177; .448] (n.s.).

State correlation estimate  $r = -.551$  with CI=[-.749; -.299] (sign.).

## 2.2 Causality and Gaze following

PPP= .371, 95% confidence interval for the difference between the observed and replicated chi-square values of CI = [-17.732; 24.373].

Trait correlation estimate  $r = .067$  with CI=[-.266; .388] (n.s.).

State correlation estimate  $r = -.276$  with CI=[-.590; .079] (n.s.).

## 2.3 Causality and Quantity

Due to a low PPP value, the model was modified such that indicator-specific trait factors were estimated on the between-level for both causality and quantity, including indicator-specific trait correlations between the two tasks.

PPP= .419, 95% confidence interval for the difference between the observed and replicated chi-square values of CI = [-19.264; 23.539].

Trait correlation estimates between causality and quantity:  $r = .096$  with CI=[-.235; .404] (n.s.),  $r = .313$  with CI=[-.027; .589] (n.s.),  $r = .258$  with CI=[-.084; .546] (n.s.),  $r = .466$  with CI=[.162; .688] (sign.).

State correlation estimate  $r = -.205$  with CI=[-.533; .159] (n.s.).

## 2.4 Inference and Gaze following

PPP= .419, 95% confidence interval for the difference between the observed and replicated chi-square values of CI = [-18.550; 24.279].

Trait correlation estimate  $r = .018$  with CI=[-.298; .338] (n.s.).

State correlation estimate  $r = .116$  with CI=[-.146; .379] (n.s.).

## 2.5 Inference and Quantity

PPP= .341, 95% confidence interval for the difference between the observed and replicated chi-square values of CI = [-16.100; 25.863].

Trait correlation estimate  $r = .436$  with CI=[.135; .665] (sign.).

State correlation estimate  $r = .294$  with CI=[-.067; .580] (n.s.).

## 2.6 Quantity and Gaze following

PPP= .402, 95% confidence interval for the difference between the observed and replicated chi-square values of CI = [-18.957; 23.435].

Trait correlation estimate  $r = .098$  with CI=[-.361; .598] (n.s.).

State correlation estimate  $r = -.026$  with CI=[-.142; .081] (n.s.).

# 3 Simulation Studies

## 3.1 Simulation setup

Data were generated and estimated using MPlus 8.4. Data-generating values are based on the real-data application of the models to the available subset of the data at the time of conducting the simulation study. That is, data were simulated for 40 individuals (N) observed across 9 or 12 measurement occasions, with 5 or 7 observed categories per indicator. 1000 replications were simulated. Data estimation took place using the MPlus default priors. In case of LST models for one construct, default priors were compared with IG(0.001, 0.001) priors set on all variance parameters (model did not include latent covariances). Two chains with a minimum of 5,000 iterations per chain and a thinning factor of 10 was applied (i.e. at last 50,000 iterations of which only every 10th was used for constructing the posterior distribution). Convergence was assumed and estimation stopped when the PSR fell below 1.05 for the first time after the minimum number of iterations was reached.

## 3.2 Summary of the results

In the following, the 95% coverage rate, the relative parameter estimation bias (deviation between average estimate and population parameter divided by the population parameter), the Mean Squared Error, absolute bias, as well as relative standard error bias are displayed for every simulated model.

Latent state models for one construct work well, with small biases (below a cutoff of 10% bias) and good coverage rates, irrespective of simulating 7 or 5 observed ordered categories for the observed indicators.

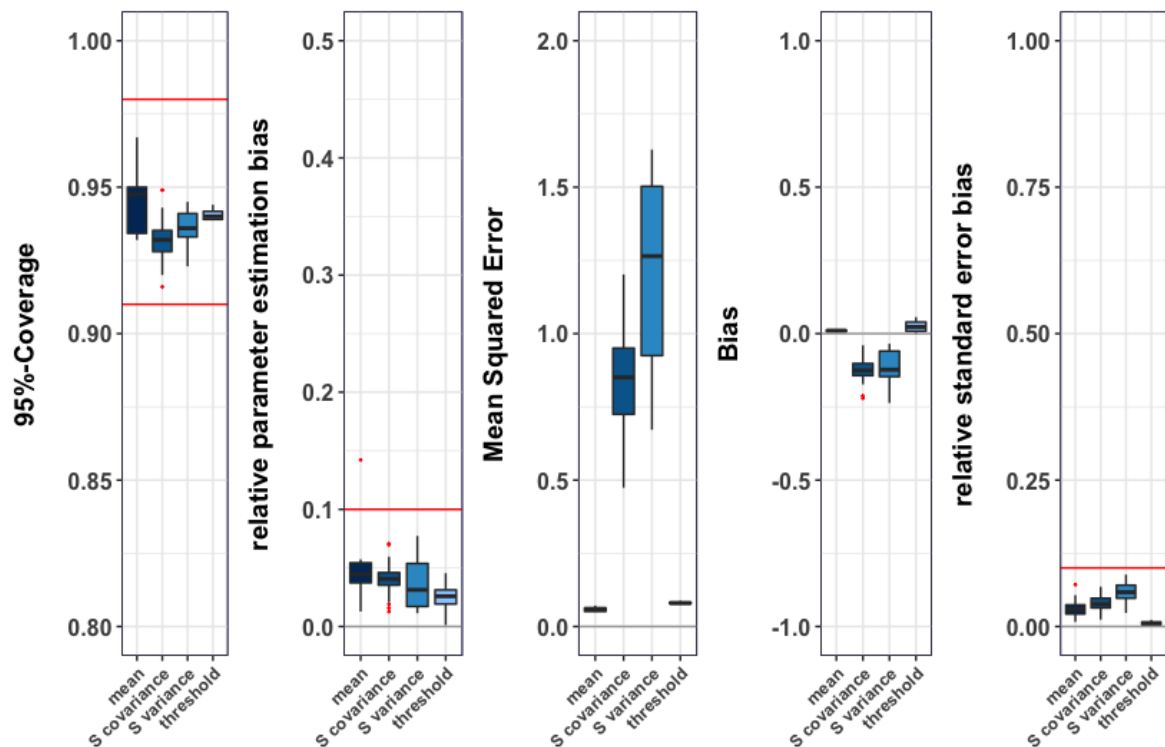
Latent State Trait models for one construct with latent state residual variances fixed across time show good estimation performance, with both default or adapted inverse gamma priors. When freely estimating latent state residual variances across time points (i.e., no restrictions on variances), model parameters are not estimated accurately, irrespective of the prior choice.

Latent State Trait models for a combination of two constructs with latent variances and covariances of the state residual variances restricted to equality across time points work well. Models with freely estimated variances and covariances do not show good estimation performance. The same holds for the Latent State model with two constructs and freely estimated variances and covariances.

In conclusion, Latent State models for one construct (freely estimated variances) as well as Latent State Trait models for one or two constructs with state residual variances restricted across time exhibit good estimation performance (low biases, high coverage) and application under simulated samples sizes should be feasible in practice.

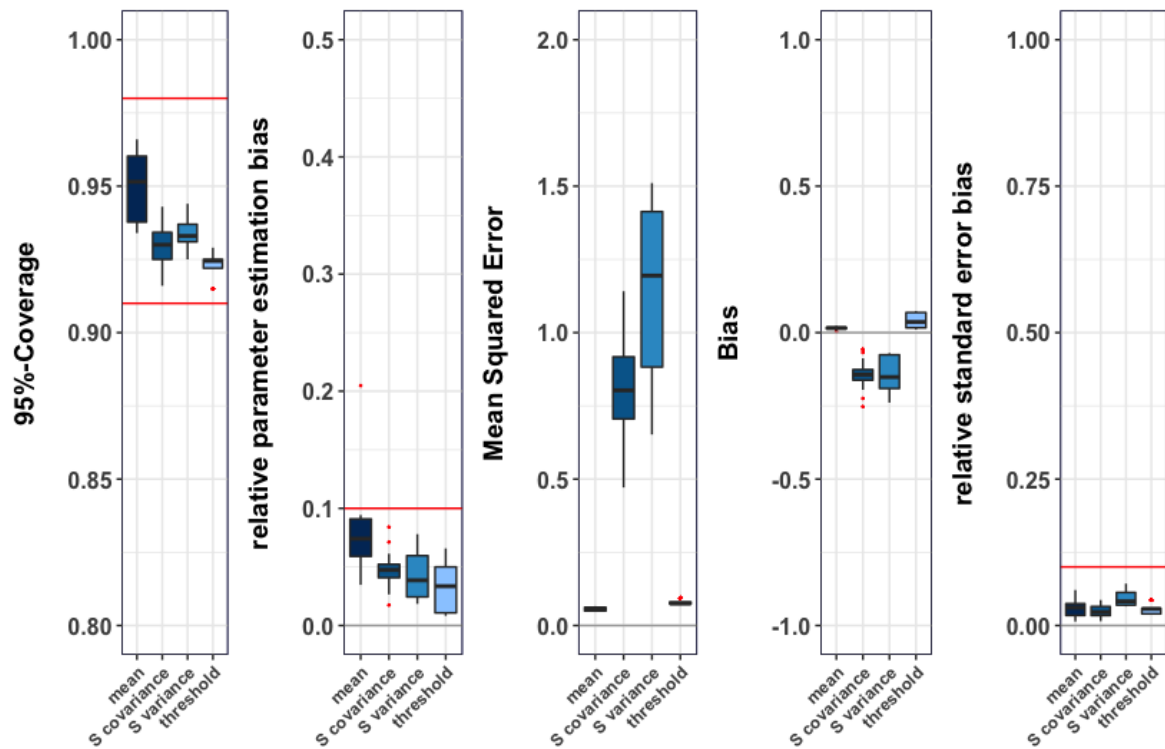
### 3.3 Latent State models: One construct

#### 3.3.1 Freely varying state variances and covariances across time points. 5 ordered categories.



**Abbildung 4:** Results of the simulation study for the Latent State (LS) model including one construct with freely estimated latent State variances and covariances, spanning 9 measurement time points. Ordinal indicators were simulated with 5 ordered categories. Boxplots display the distribution of the respective statistic across different parameters of the same parameter type.

### 3.3.2 Freely varying state variances and covariances across time points. 7 ordered categories.

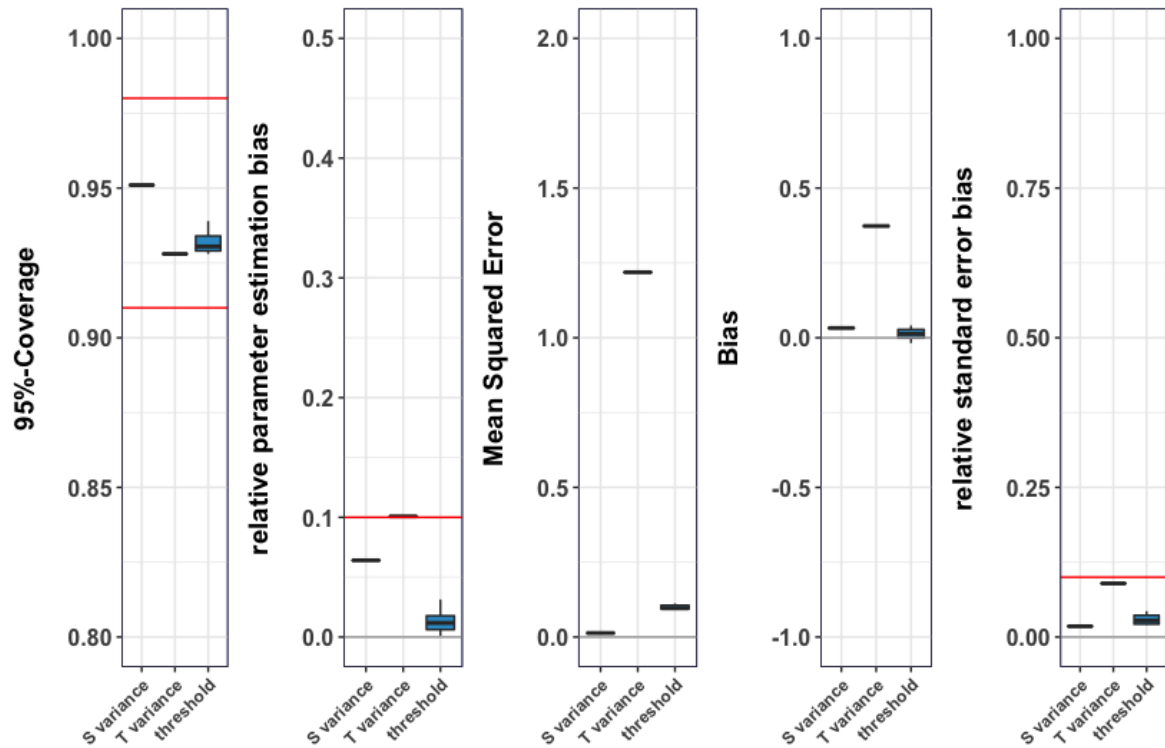


**Abbildung 5:** Results of the simulation study for the Latent State (LS) model including one construct with freely estimated latent State variances and covariances, spanning 9 measurement time points. Ordinal indicators were simulated with 7 ordered categories. Boxplots display the distribution of the respective statistic across different parameters of the same parameter type.



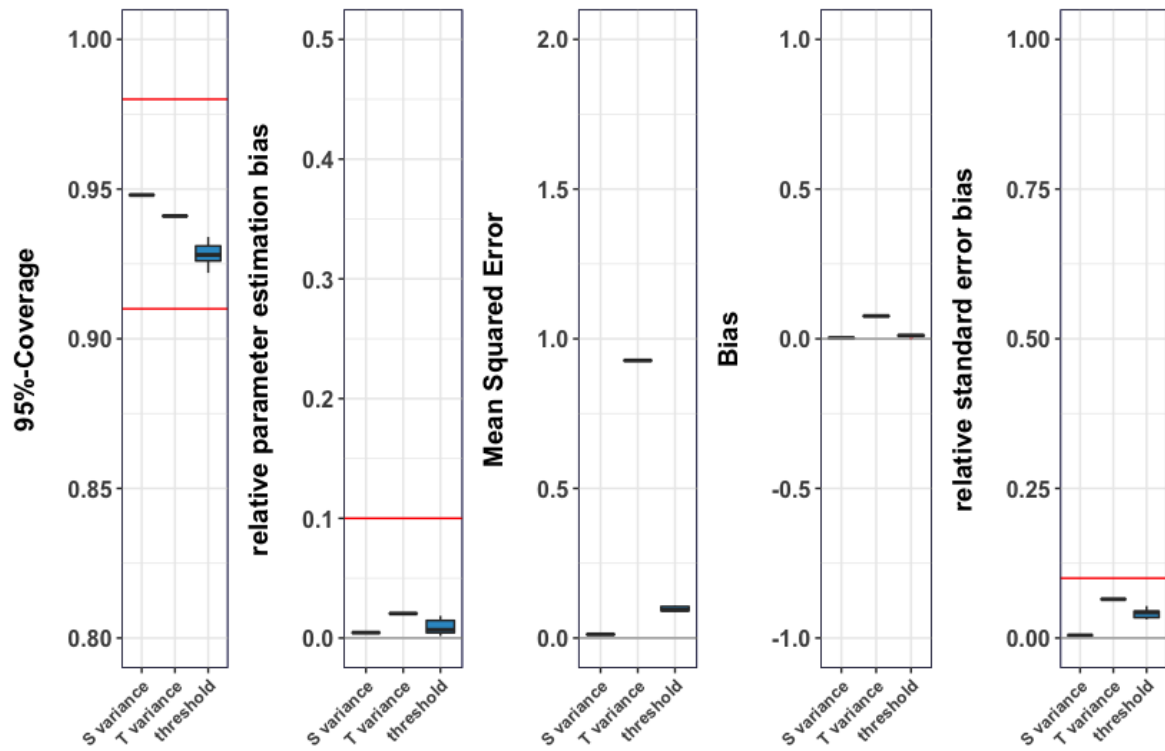
### 3.4 Latent State Trait models: One construct

#### 3.4.1 Fixed state variances across time points. Default priors.



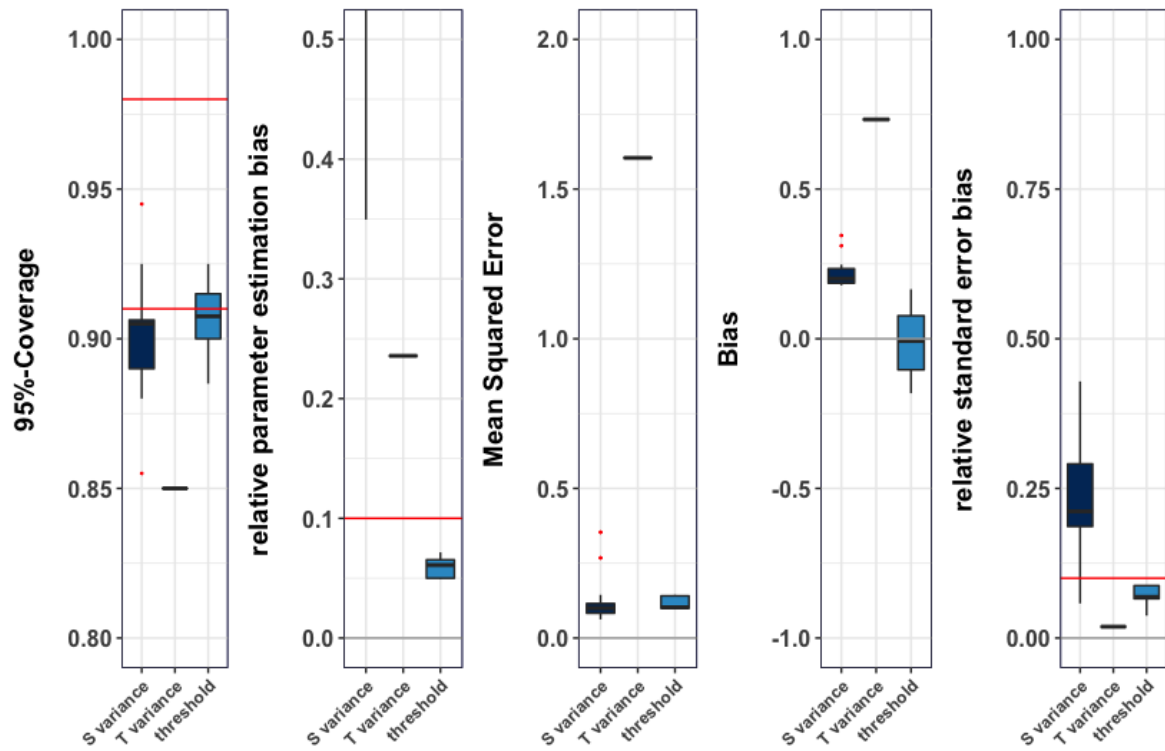
**Abbildung 6:** Results of the simulation study for the Latent State Trait (LST) model including one construct with latent state residual variances fixed to be equal across time, spanning 9 measurement time points. MPlus default priors. Ordinal indicators were simulated with 7 ordered categories. Boxplots display the distribution of the respective statistic across different parameters of the same parameter type.

### 3.4.2 Fixed state variances across time points. Inverse gamma priors.



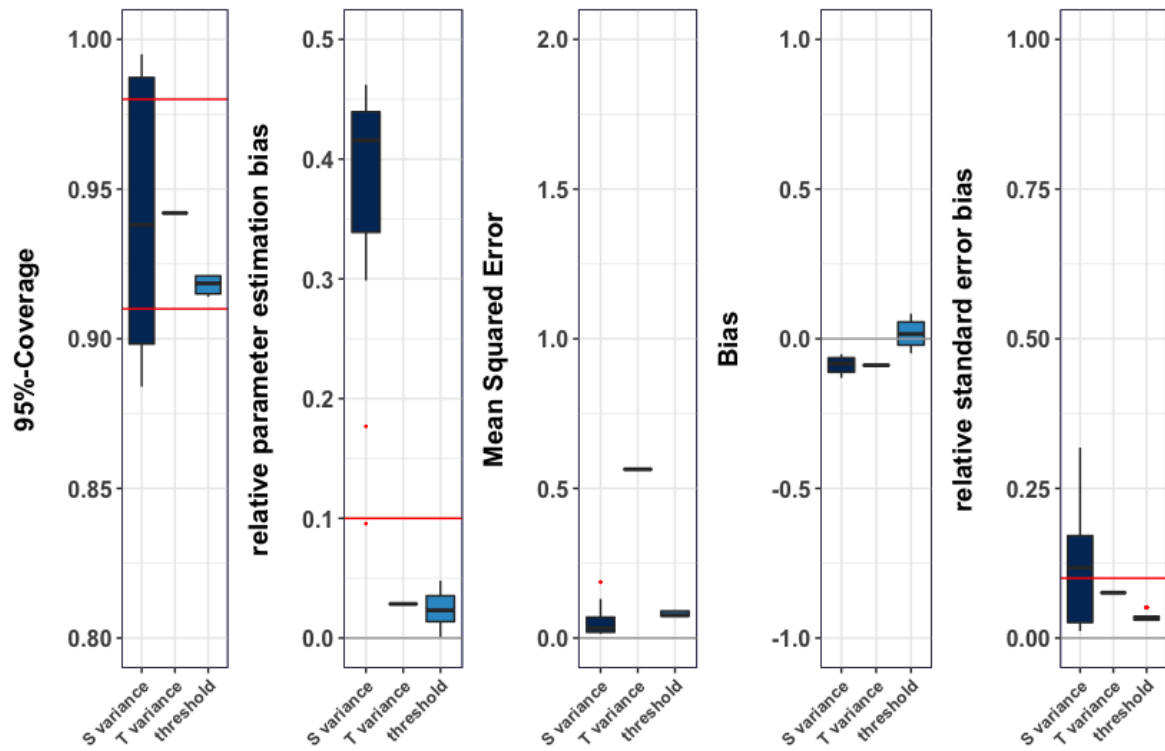
**Abbildung 7:** Results of the simulation study for the Latent State Trait (LST) model including one construct with latent state residual variances fixed to be equal across time, spanning 9 measurement time points. Inverse gamma priors  $IG(0.001, 0.001)$  for all variances. Ordinal indicators were simulated with 7 ordered categories. Boxplots display the distribution of the respective statistic across different parameters of the same parameter type.

### 3.4.3 Free state variances across time points. Default priors.



**Abbildung 8:** Results of the simulation study for the Latent State Trait (LST) model including one construct with latent state residual variances freely estimates across time, spanning 9 measurement time points. Ordinal indicators were simulated with 7 ordered categories. Boxplots display the distribution of the respective statistic across different parameters of the same parameter type.

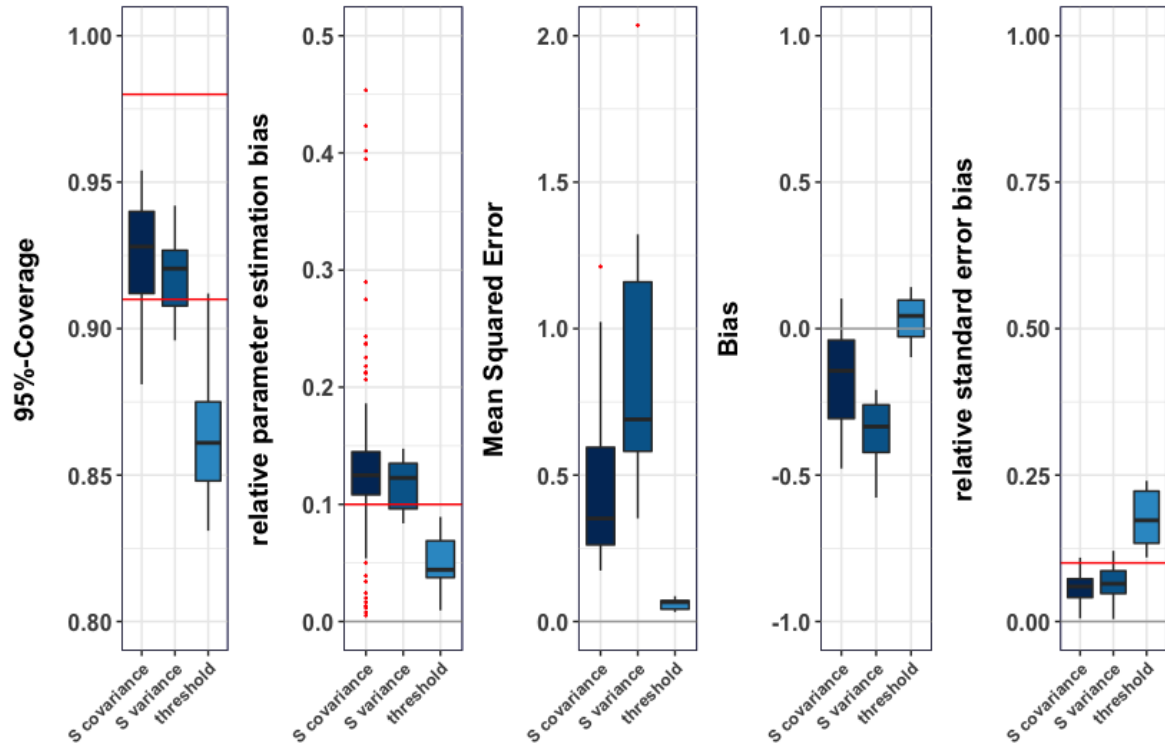
### 3.4.4 Free state variances across time points. Inverse gamma priors.



**Abbildung 9:** Results of the simulation study for the Latent State Trait (LST) model including one construct with latent state residual variances freely estimates across time, spanning 9 measurement time points. Inverse gamma priors  $IG(0.001, 0.001)$  for all variances. Ordinal indicators were simulated with 7 ordered categories. Boxplots display the distribution of the respective statistic across different parameters of the same parameter type.

### 3.5 Latent State models: Two constructs

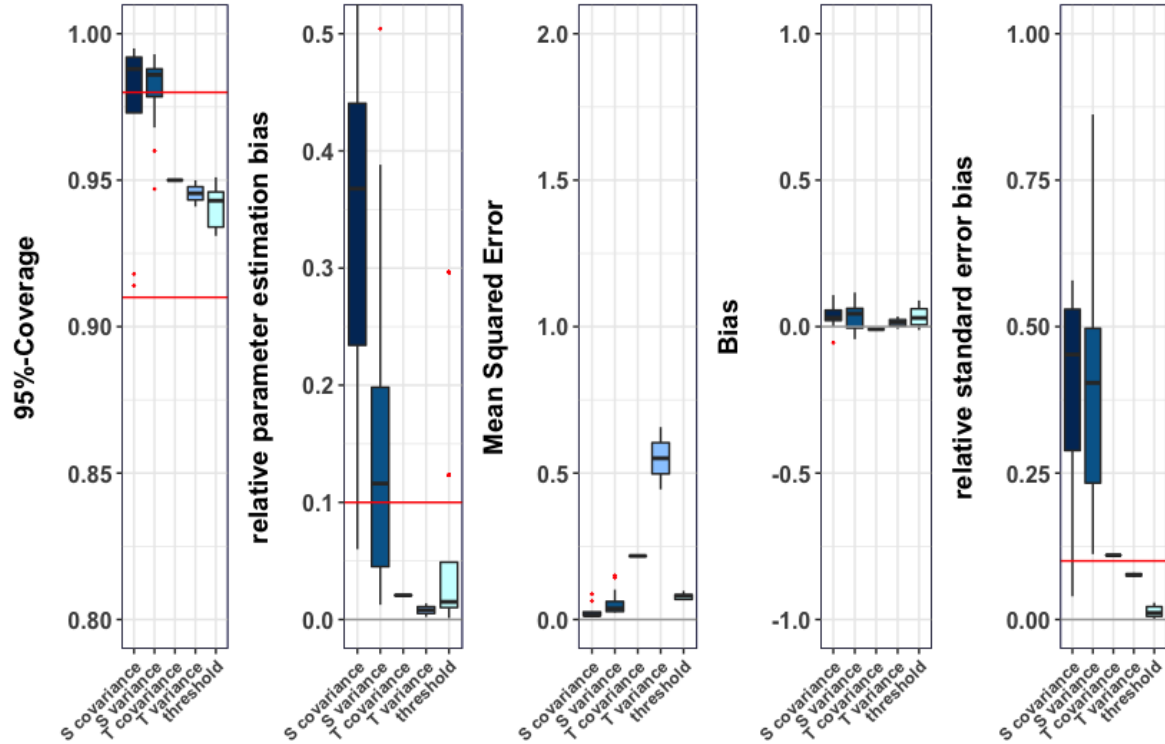
#### 3.5.1 Free state variances and covariances across time points. Default priors.



**Abbildung 10:** Results of the simulation study for the Latent State model including two constructs with latent state variances freely estimates across time, spanning 9 measurement time points. Ordinal indicators were simulated with 7 ordered categories. Boxplots display the distribution of the respective statistic across different parameters of the same parameter type.

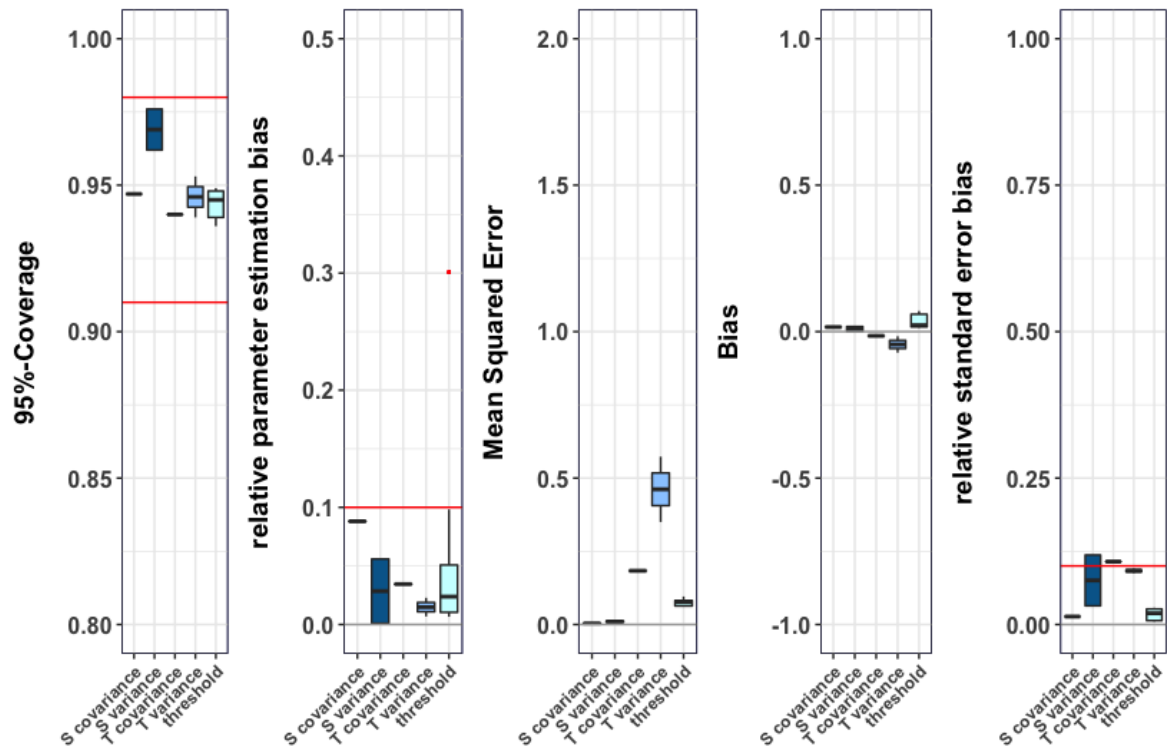
### 3.6 Latent State Trait models: Two constructs

#### 3.6.1 Free state variances and covariances across time points.



**Abbildung 11:** Results of the simulation study for the Latent State Trait (LST) model including two constructs with free latent state residual variances across time, spanning 9 measurement time points. Ordinal indicators were simulated with 7 ordered categories. Boxplots display the distribution of the respective statistic across different parameters of the same parameter type.

### 3.6.2 Fixed state variances and covariances across time points.



**Abbildung 12:** Results of the simulation study for the Latent State Trait (LST) model including two constructs with fixed latent state residual variances across time, spanning 9 measurement time points. Ordinal indicators were simulated with 7 ordered categories. Boxplots display the distribution of the respective statistic across different parameters of the same parameter type.