Supplementary Material

Sebastian Castro-Alvarez, Jorge N. Tendeiro, Rob R. Meijer, Laura F. Bringmann

University of Groningen

Using Structural Equation Modeling to Study Traits and States in Intensive Longitudinal Data

This document includes the supplementary material of the study named *Using Structural Equation Modeling to Study Traits and States in Intensive Longitudinal Data*. The material is supplementary to the simulation study and the empirical example. In relation to the simulation study, we provide (a) the true population parameters used during the simulation to generate the data, (b) a sensitivity analysis of the Bayesian models, (c) the results of the simulation that was performed to manipulate and explore the effect of the number of indicators, (d) the extended results of the simulation taking into account the information criteria and the particular effect of the trait-state variance ratio in some conditions, (e) the results of the additional simulation that was performed to explore why the multilevel CUTS model failed to fit to TSO data, and (f) a plot of the average bias of the autoregressive effect for the conditions where the TSO model was fitted to TSO data. Regarding the empirical example, we present the results of the analyses when non-stationary time series were excluded from the analyses, which reduced the sample size to 376 individuals. In particular, we added four tables that summarize these results: The first table shows the ppp and the DIC of the models fitted to these data, the second and third tables show the estimates of the multilevel TSO model and their credibility intervals for each set of items, and the fourth table shows the estimates of the variance coefficients.

True Population Parameters

Table S1
MSST True Parameters per Trait-State Variance Ratio

	MSST 1:3	MSST 1:1	MSST 3:1
Within loading λ_{S_1}	1.00	1.00	1.00
Within loading λ_{S_2}	0.50	0.50	0.50
Within loading λ_{S_3}	1.30	1.30	1.30
Within loading λ_{S_4}	0.80	0.80	0.80
State variance $var(\zeta)$	1.80	1.20	0.60
Error variance $var(\varepsilon_1)$	0.60	0.60	0.60
Error variance $var(\varepsilon_2)$	0.25	0.25	0.25
Error variance $var(\varepsilon_3)$	0.70	0.70	0.70
Error variance $var(\varepsilon_4)$	0.50	0.50	0.50
Between loading λ_{T_1}	1.00	1.00	1.00
Between loading λ_{T_2}	0.50	0.50	0.50
Between loading λ_{T_3}	1.30	1.30	1.30
Between loading λ_{T_4}	0.80	0.80	0.80
Intercept α_1	0.00	0.00	0.00
Intercept α_2	0.20	0.20	0.20
Intercept α_3	0.40	0.40	0.40
Intercept α_4	0.60	0.60	0.60
Trait variance $var(\xi)$	0.60	1.20	1.80
Trait mean $\hat{\xi}$	4.00	4.00	4.00

Table S2

CUTS True Parameters per Trait-State Variance Ratio

	CUTS 1:3	CUTS 1:1	CUTS 3:1
Within loading λ_{S_1}	1.00	1.00	1.00
Within loading λ_{S_2}	0.50	0.50	0.50
Within loading λ_{S_3}	1.30	1.30	1.30
Within loading λ_{S_4}	0.80	0.80	0.80
Common State variance $var(\zeta)$	1.80	1.20	0.60
Unique state variance $var(\varepsilon_1)$	0.60	0.60	0.60
Unique state variance $var(\varepsilon_2)$	0.25	0.25	0.25
Unique state variance $var(\varepsilon_3)$	0.80	0.80	0.80
Unique state variance $var(\varepsilon_4)$	0.50	0.50	0.50
Between loading λ_{T_1}	1.00	1.00	1.00
Between loading λ_{T_2}	0.50	0.50	0.50
Between loading λ_{T_3}	1.30	1.30	1.30
Between loading λ_{T_4}	0.80	0.80	0.80
Intercept α_1	2.00	2.00	2.00
Intercept α_2	2.50	2.50	2.50
Intercept α_3	3.00	3.00	3.00
Intercept α_4	3.50	3.50	3.50
Common Trait variance $var(\xi)$	0.40	1.00	1.60
Unique trait variance $var(\vartheta)$	0.20	0.20	0.20
Unique trait variance $var(\vartheta)$	0.10	0.10	0.10
Unique trait variance $var(\vartheta)$	0.25	0.25	0.25
Unique trait variance $var(\vartheta)$	0.15	0.15	0.15

Table S3

TSO True Parameters per Trait-State Variance Ratio

	TSO 1:3	TSO 1:1	TSO 3:1
Within loading λ_{S_1}	1.00	1.00	1.00
Within loading λ_{S_2}	0.50	0.50	0.50
Within loading λ_{S_3}	1.30	1.30	1.30
Within loading λ_{S_4}	0.80	0.80	0.80
Occasion-specific residual variance $var(\zeta)$	1.80	1.20	0.60
Error variance $var(\varepsilon_1)$	0.60	0.60	0.60
Error variance $var(\varepsilon_2)$	0.25	0.25	0.25
Error variance $var(\varepsilon_3)$	0.70	0.70	0.70
Error variance $var(\varepsilon_4)$	0.50	0.50	0.50
Autoregressive effect β	0.50	0.50	0.50
Intercept α_1	2.00	2.00	2.00
Intercept α_2	2.50	2.50	2.50
Intercept α_3	3.00	3.00	3.00
Intercept α_4	3.50	3.50	3.50
Latent trait indicator variance $var(\xi_1)$	0.30	0.80	1.60
Latent trait indicator variance $var(\xi_2)$	0.10	0.20	0.55
Latent trait indicator variance $var(\xi_3)$	0.40	1.30	2.20
Latent trait indicator variance $var(\xi_4)$	0.20	0.50	1.20
$Cov(\xi_1, \xi_2)$	0.14	0.32	0.75
$Cov(\xi_1, \xi_3)$	0.31	0.92	1.69
$Cov(\xi_2, \xi_3)$	0.16	0.41	0.88
$Cov(\xi_1, \xi_4)$	0.22	0.57	1.25
$Cov(\xi_2, \xi_4)$	0.10	0.22	0.57
$Cov(\xi_3, \xi_4)$	0.20	0.56	1.14

Table S4

MSST True Variance Coefficient Components per Trait-State Variance Ratio

	MSST 1:3	MSST 1:1	MSST 3:1
Reliability Y_1	0.80	0.80	0.80
Reliability Y_2	0.71	0.71	0.71
Reliability Y_3	0.85	0.85	0.85
Reliability Y_4	0.75	0.75	0.75
Consistency Y_1	0.20	0.40	0.60
Consistency Y_2	0.18	0.35	0.53
Consistency Y_3	0.21	0.43	0.64
Consistency Y_4	0.19	0.38	0.57
Occasion Specificity Y_1	0.60	0.40	0.20
Occasion Specificity Y_2	0.53	0.35	0.18
Occasion Specificity Y_3	0.64	0.43	0.21
Occasion Specificity Y_4	0.57	0.38	0.19

Table S5

CUTS True Variance Coefficient Components per Trait-State Variance Ratio

	CUTS 1:3	CUTS 1:1	CUTS 3:1
Reliability Y_1	0.80	0.80	0.80
Reliability Y_2	0.72	0.72	0.72
Reliability Y_3	0.83	0.83	0.83
Reliability Y_4	0.76	0.76	0.76
Total Consistency Y_1	0.20	0.40	0.60
Total Consistency Y_2	0.22	0.39	0.56
Total Consistency Y_3	0.19	0.41	0.62
Total Consistency Y_4	0.20	0.38	0.57
Common Consistency Y_1	0.13	0.33	0.53
Common Consistency Y_2	0.11	0.28	0.44
Common Consistency Y_3	0.14	0.35	0.57
Common Consistency Y_4	0.12	0.31	0.50
Unique Consistency Y_1	0.07	0.07	0.07
Unique Consistency Y_2	0.11	0.11	0.11
Unique Consistency Y_3	0.05	0.05	0.05
Unique Consistency Y_4	0.07	0.07	0.07
Occasion Specificity Y_1	0.60	0.40	0.20
Occasion Specificity Y_2	0.50	0.33	0.17
Occasion Specificity Y_3	0.64	0.43	0.21
Occasion Specificity Y_4	0.56	0.37	0.19

Table S6

TSO True Variance Coefficient Components per Trait-State Variance Ratio

	TSO 1:3	TSO 1:1	TSO 3:1
Reliability Y_1	0.82	0.80	0.80
Reliability Y_2	0.74	0.71	0.75
Reliability Y_3	0.86	0.85	0.84
Reliability Y_4	0.78	0.75	0.77
Consistency Y_1	0.27	0.40	0.60
Consistency Y_2	0.26	0.35	0.60
Consistency Y_3	0.27	0.42	0.60
Consistency Y_4	0.26	0.37	0.60
Predictability by Trait Y_1	0.09	0.27	0.53
Predictability by Trait Y_2	0.11	0.24	0.55
Predictability by Trait Y_3	0.08	0.28	0.52
Predictability by Trait Y_4	0.09	0.25	0.54
Unpredictability by Trait Y_1	0.18	0.13	0.07
Unpredictability by Trait Y_2	0.16	0.12	0.05
Unpredictability by Trait Y_3	0.20	0.14	0.08
Unpredictability by Trait Y_4	0.17	0.13	0.06
Occasion Specificity Y_1	0.55	0.40	0.20
Occasion Specificity Y_2	0.47	0.35	0.15
Occasion Specificity Y_3	0.59	0.43	0.24
Occasion Specificity Y_4	0.52	0.38	0.17

Sensitivity Analysis

When doing Bayesian estimation, researchers have to be careful in the selection of their prior distributions because they can influence the estimates of the model. In general, the effect of the prior distributions is diminished the larger the data are. For the simulation study, we used the default prior distributions available in Mplus for our Bayesian analyses. These default priors are uninformative priors, for example, loadings are given normal priors $N(0, 10^{10})$. To verify that a different selection of priors would not affect our results, we conducted a sensitivity analysis on a random sample of the analyses of the simulation. We selected one random replication from 20 random conditions, and fitted the Bayesian models using the default priors and some weak priors selected by us (e.g., $N(0, 5)$). The results from these analyses showed that the estimates are basically the same regardless of the priors. Differences between the estimates at the within-level were not larger than 0.01 and differences between the estimates at the between-level were not larger than 0.15. These larger differences in the between-level were mainly associated to the variances, which are harder to estimate because there was less information at this level due to the sample size. As an example, Table S7 presents the estimates obtained by using both uninformative and weak priors of one of the models in one of the replications.

Table S7

Estimates of the Multilevel TSO Model using Uninformative and Weak Priors with the Base Model as the TSO, 30 Measurement Occasions, 0% Missing Values, and a Trait-State Variance Ratio of 1:1.

	Uninformative Priors	Weak Priors
Within loading λ_{S_1}	**1.00**	**1.00**
Within loading λ_{S_2}	0.479	0.479
Within loading λ_{S_3}	1.258	1.259
Within loading λ_{S_4}	0.780	0.780
Occasion-specific residual variance $var(\zeta)$	1.347	1.348
Error variance $var(\varepsilon_1)$	0.599	0.599
Error variance $var(\varepsilon_2)$	0.263	0.262
Error variance $var(\varepsilon_3)$	0.644	0.642
Error variance $var(\varepsilon_4)$	0.510	0.510
Autoregressive effect β	0.493	0.495
Intercept α_1	1.930	1.935
Intercept α_2	2.477	2.479
Intercept α_3	2.903	2.918
Intercept α_4	3.479	3.481
Latent trait indicator variance $var(\xi_1)$	0.829	0.696
Latent trait indicator variance $var(\xi_2)$	0.209	0.180
Latent trait indicator variance $var(\xi_3)$	1.295	1.110
Latent trait indicator variance $var(\xi_4)$	0.505	0.438
$Cov(\xi_1, \xi_2)$	0.339	0.279
$Cov(\xi_1, \xi_3)$	0.927	0.773
$Cov(\xi_2, \xi_3)$	0.424	0.352
$Cov(\xi_1, \xi_4)$	0.570	0.477
$Cov(\xi_2, \xi_4)$	0.235	0.192
$Cov(\xi_3, \xi_4)$	0.528	0.422

Simulation: Effect of the Number of Indicators

When using the single-level state-trait SEMs, increasing the number of indicators has a similar effect as increasing the number of measurement occasions in relation to the number of observed variables that are included in the models. For example, a study with six indicators and five measurement occasions has the same number of observed variables as a study with three indicators and ten measurement occasions. Because of this, it is reasonable to think that increasing the number of indicators also impacts the performance of the models. However, this factor was kept fixed in the simulation design because intensive longitudinal studies are not likely to include a lot indicators to measure the same construct. Nevertheless, to verify how the number of indicators might affect the performance of the models, we ran a small-scale simulation where we manipulated this factor. In these analyses, the number of indicators was varied between 4, 7, and 10; and the number of measurement occasions was varied between 10, 20, 30, and 60. Moreover, we only analyzed the data with the same model that was used to generate them.

The number of analyses that finished successfully given each condition are shown in Figures S1 to S3. These plots clearly show that the number of indicators does not impact the performance of the models but the number of measurement occasions does. This difference can be explained by how each of these factors directly affects the models. On the one hand, increasing the number of measurement occasions makes the models more complex because it introduces more latent variables that have to be modeled for each additional occasion. For example, if we are using the CUTS model in a design with 3 indicators and increase the number of measurement occasions from 10 to 20, the number of latent variables in the model increases from 14 to 24. On the other hand, while increasing the number of indicators increases the number of loadings and within-residual variances, it barely increases the number of latent variables that have to be modeled. For example, if we are applying the CUTS model and increase the number of indicators from 3 to 6 in a design with 10 measurement occasions, we go from a model with 14 to a model with 17 latent variables. In conclusion, with large datasets, single-level state-trait SEMs are more likely to run into convergence issues the more latent variables there are in the model. Therefore, as increasing the number of indicators barely increases the number of latent variables in the model, including this factor in the simulation study was unnecessary.

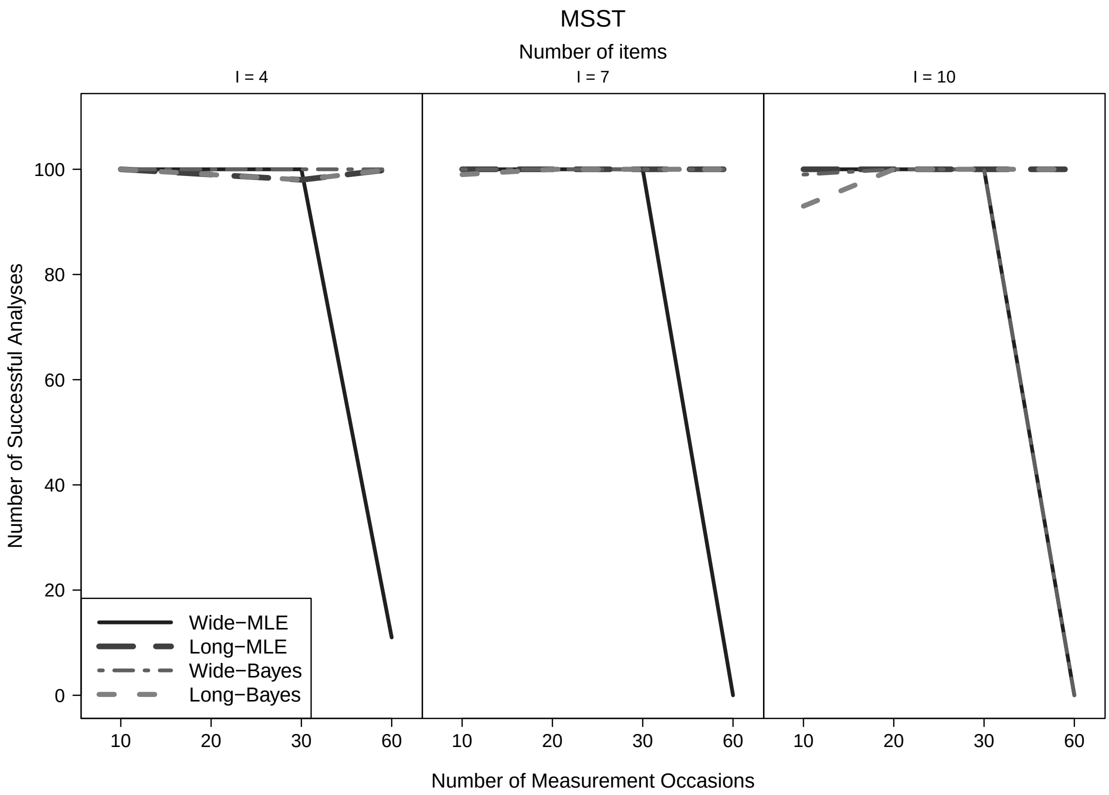

Figure S1. Number of Successful Analyses varying the Number of Indicators with the MSST Model

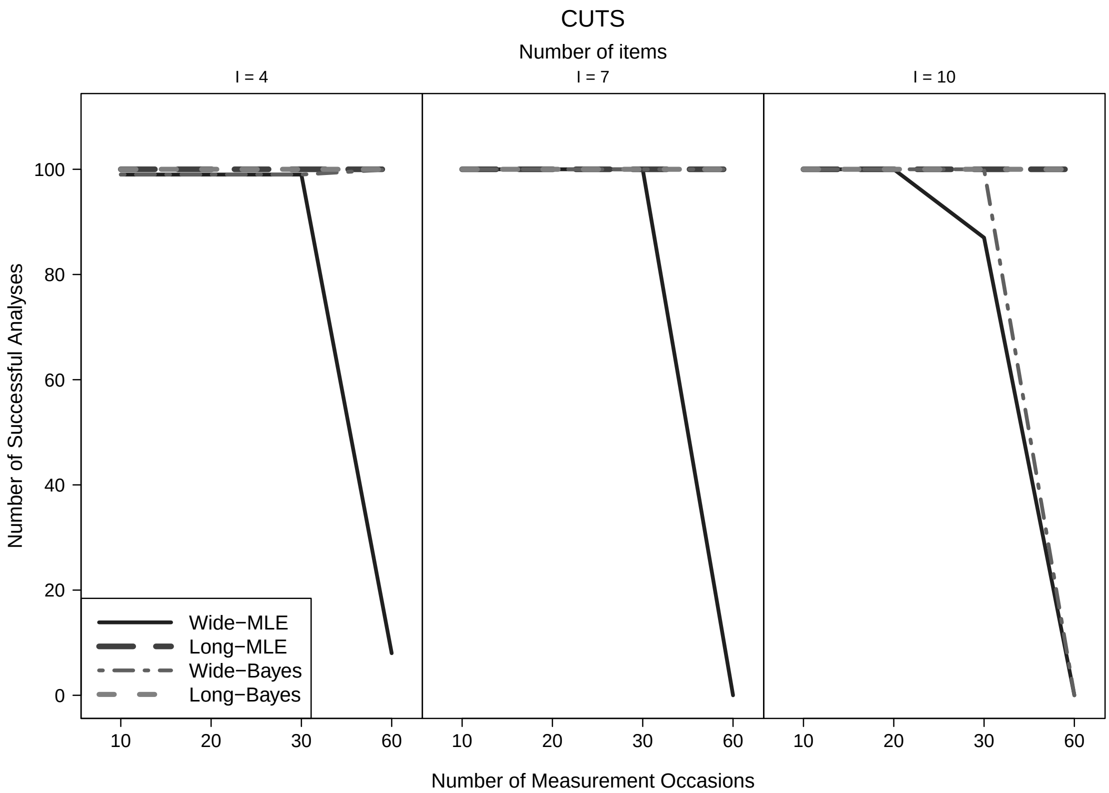

Figure S2. Number of Successful Analyses varying the Number of Indicators with the CUTS Model

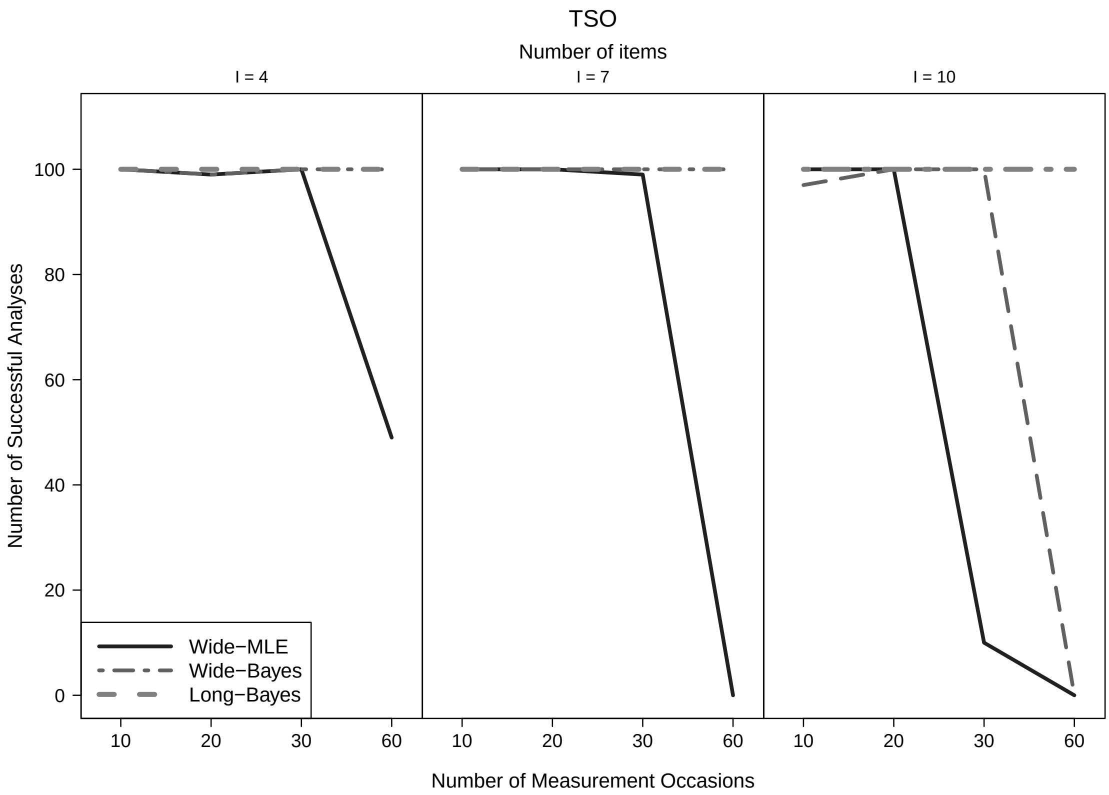

Figure S3. Number of Successful Analyses varying the Number of Indicators with the TSO Model

Simulation: Extended Results

In this section, we provide extended results of the simulation study. Firstly, we give a brief summary of the information criteria indices, which are useful to decide which model fitted the data best. Secondly, we include the results of the number of successful analyses and the quality of the estimates related to the conditions with a trait-state variance ratio different from 1:1.

Information Criteria

To decide which model fitted the data best, we used the information criterion indices available in Mplus. This includes the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the adjusted Bayesian Information Criterion (aBIC) when the estimation method was MLE; and the Deviance Information Criterion (DIC) when the estimation method was MCMC Gibbs sampling.

Concerning the information criteria, we were only able to compare models estimated with the same method. When the estimation method was MLE, we selected among five models (MSST, ML-MSST, CUTS, ML-CUTS, and TSO) by means of the AIC, the BIC, and the aBIC. The percentage of the number of times a model was selected as the best model is shown in Table S8 given the model used to generate the data. In most of the cases, the correct model was selected as the best model regardless of whether it was the single level or the multilevel version. However, when the data were generated based on the TSO model, the model selected as the best model was actually the ML-MSST in about a third of the analyses. This happened because when the number of measurement occasions was large and the base model was the TSO, the only model that converged by means of MLE was the ML-MSST model (see Figures S4 to S9). Hence, it was the only model to pick from as the best model.

Table S8

Percentage of Times a Model was Selected as Best Model According to the Information Criteria Indices Available with Maximum Likelihood Estimation

Base Model	Information Criterion	Fitted Model				
		MSST	ML-MSST	CUTS	ML-CUTS	TSO
MSST	AIC	35.9	63.6	0.1	0.4	0
	BIC	58.4	41.6	0	0	0
	aBIC	58.2	41.6	0.2	0	0
CUTS	AIC	0	0	50.1	42.4	7.4
	BIC	0	0	56.5	38.4	5.1
	aBIC	0	0	42.9	38.4	18.7
TSO	AIC	0.1	42.8	0.7	1.1	55.4
	BIC	6.6	36.2	0.7	1.1	55.4
	aBIC	6.6	36.2	0.7	1.1	55.4

When the estimation method was Bayesian, we compared the models by means of the DIC. The ML-TSO model was almost always selected as the best model across all the conditions independently of the base model. For example, when the data were generated based on the MSST model, the ML-TSO model was selected as the best model 5383 times out of 5400 regardless of the number of measurement occasions, the proportion of missing values, or the trait-state variance ratio. Note that if the data were in wide format and had 30 measurement occasions or more, Mplus was unable to compute the DIC.[1]This means that when the number of measurement occasions was 30 or more, the best model according to the DIC was selected only from the multilevel models.

Effect of the Trait-State Variance Ratio

We included the trait-state variance ratio in the simulation design expecting little to no differences when manipulating this factor. While this was generally true, there are some particular conditions where the trait-state variance ratio interacts with other manipulated factors and shows some effects in the performance of the models and the quality of the estimates. In this section, we further explain these results.

First of all, in Figures S4 to S9, we present the number of analyses that finished successfully in the conditions with a trait-state variance ratio different from 1:1 given each proportion of missingness. In particular, Figures S7 to S9 show how having more state-like variables can affect the performance of the MSST when analyzing CUTS data and the performance of the TSO model when analyzing its own data. Firstly, the single-level and the multilevel MSST models performed very poorly when analyzing CUTS data across all conditions. This means that the MSST model is problematic if the variables of interest are more state-like and if real method effects are present in the data. Secondly, the single-level TSO model with MLE seemed to

[1] When doing the analyses in wide-format by means of the Gibbs sampling algorithm, the following message was always printed in the Mplus output: "Problem occurred in the computation of the posterior predictive p-value. This may be due to a singular sample variance-covariance matrix, such as with zero sample variances." This happened because of the large number of observed variables that were included in some of our analyses (120, 240, 360), which led to huge sample variance-covariance matrices. These huge matrices might make it impossible for Mplus to compute the deviance and the posterior predictive p-value of the model.

improve as the number of measurement occasions increased. One possible explanation is that, because the variables were more state-like, more data was needed to correctly capture the structure of the latent trait indicator variables (the trait component of the variables).

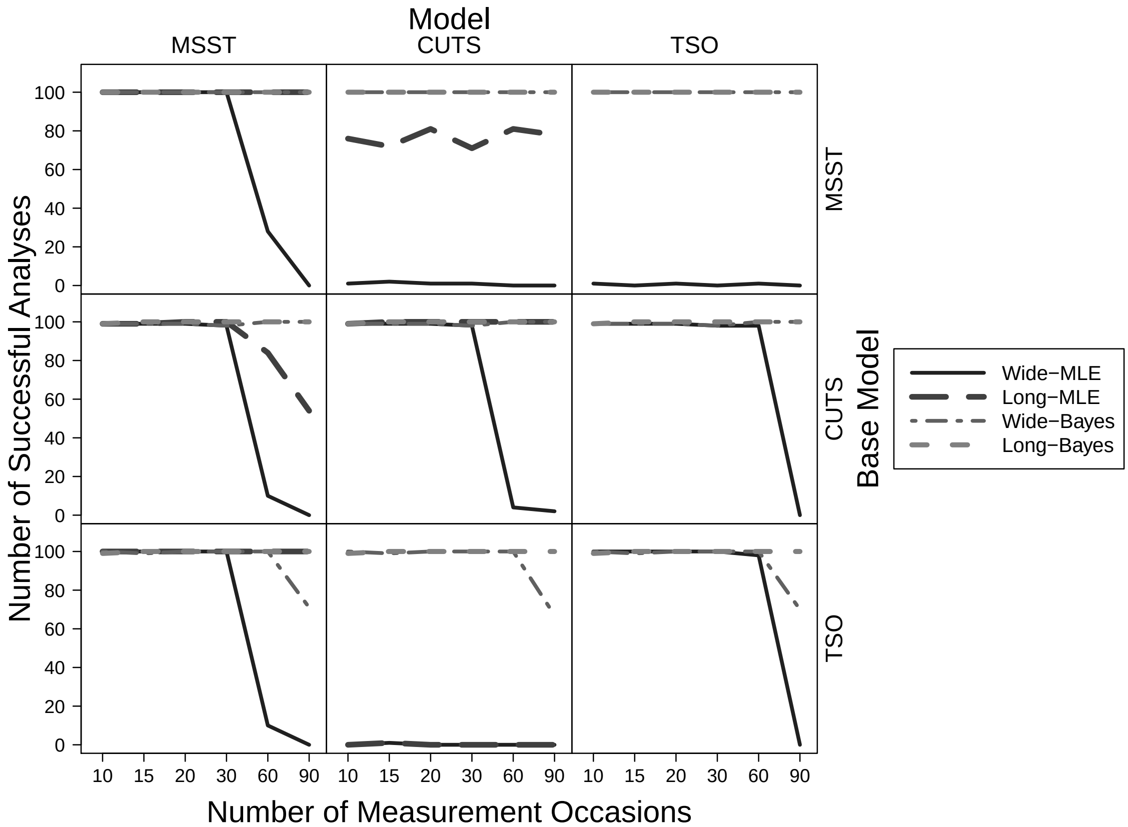

Figure S4. Number of Successful Analyses per Condition with 0% Missingness and 3:1 Trait-State Variance Ratio

Finally, varying the trait-state variance ratio also had some specific effects in the quality of the estimates of some of the parameters. Firstly, while the estimates of the within factor loadings λ_{S_j} were practically unbiased in most conditions, they tended to be overestimated when analyzing TSO data in conditions that had a trait-state variance ratio of 3:1 with the MSST model. Moreover, as mentioned in the main

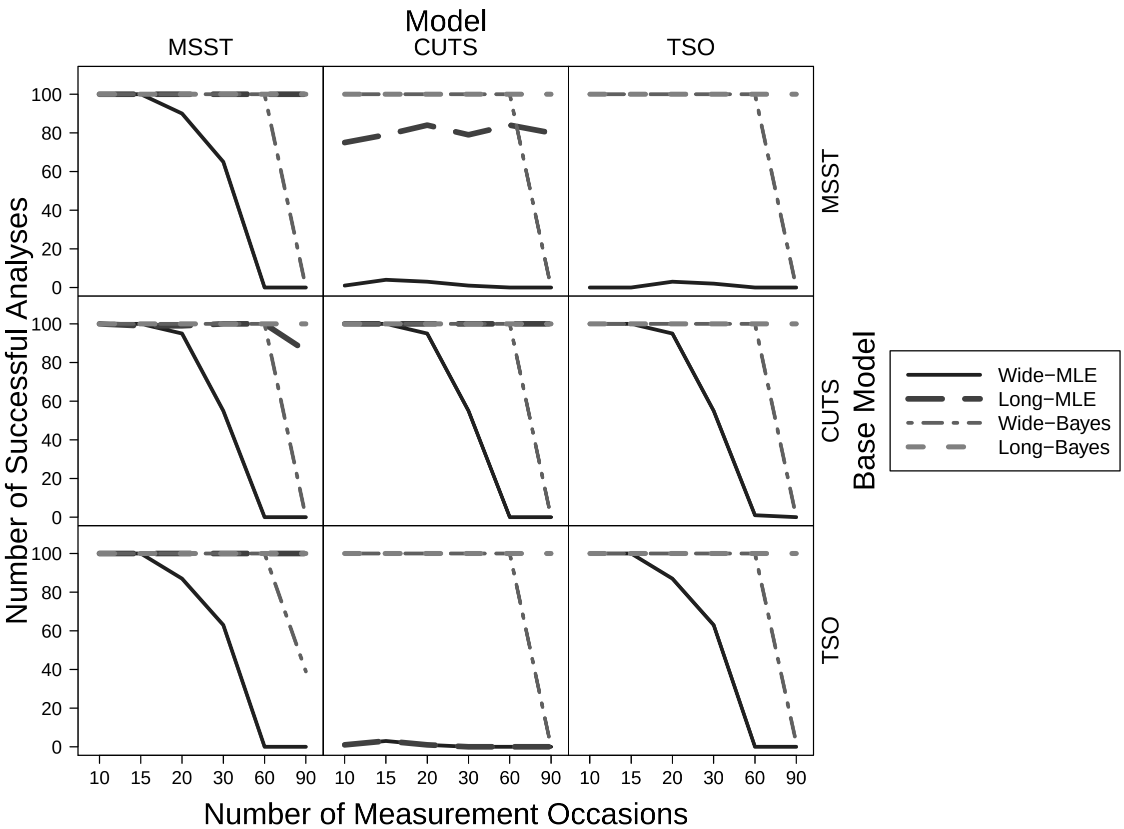

Figure S5. Number of Successful Analyses per Condition with 10% Missingness and 3:1 Trait-State Variance Ratio

text, the estimates of the consistencies tended to show more bias when the variables were more state-like. In contrast, the estimates of the occasion-specificities tended to show more bias when the variables were more trait-like.

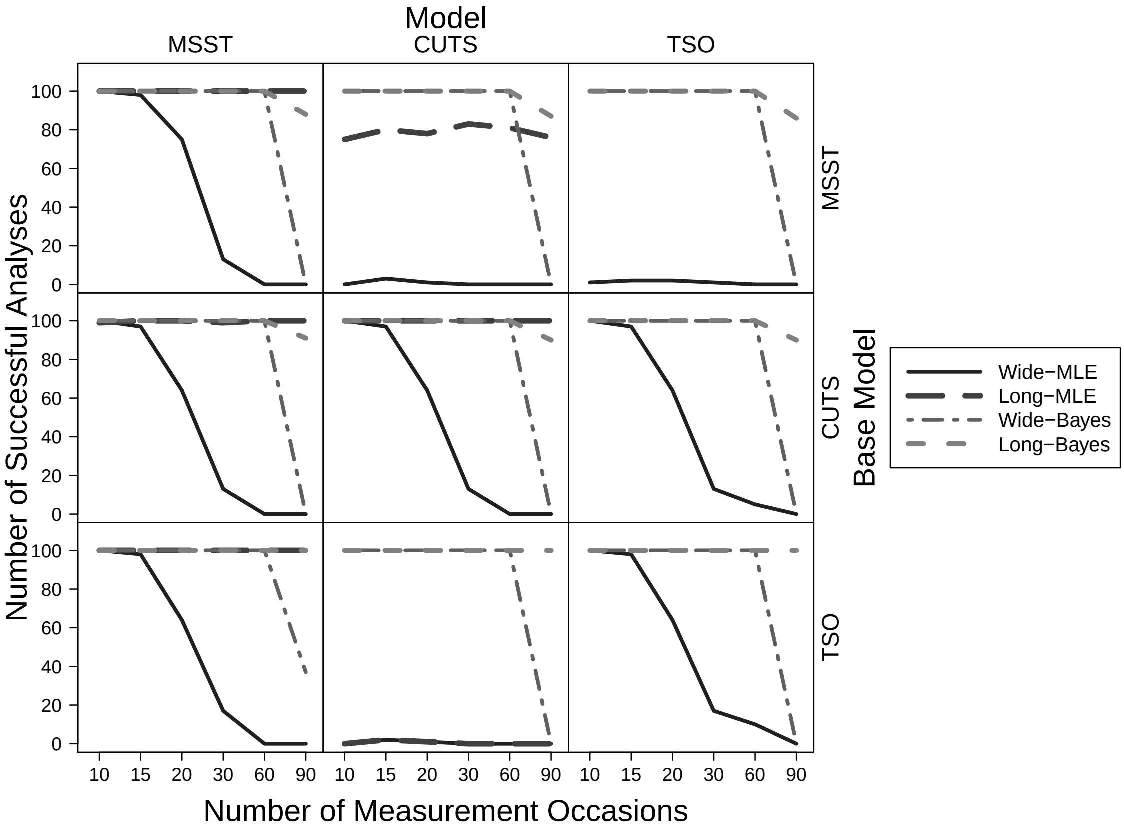

Figure S6. Number of Successful Analyses per Condition with 20% Missingness and 3:1 Trait-State Variance Ratio

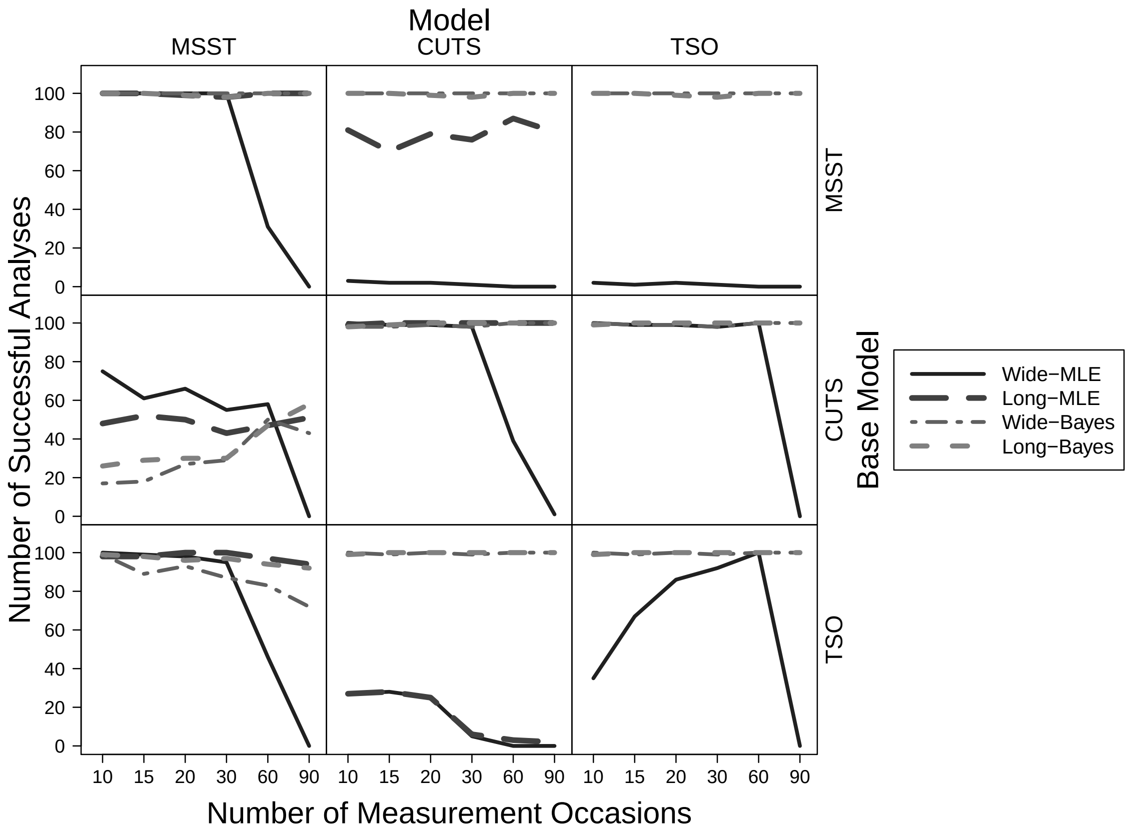

Figure S7. Number of Successful Analyses per Condition with 0% Missingness and 1:3 Trait-State Variance Ratio

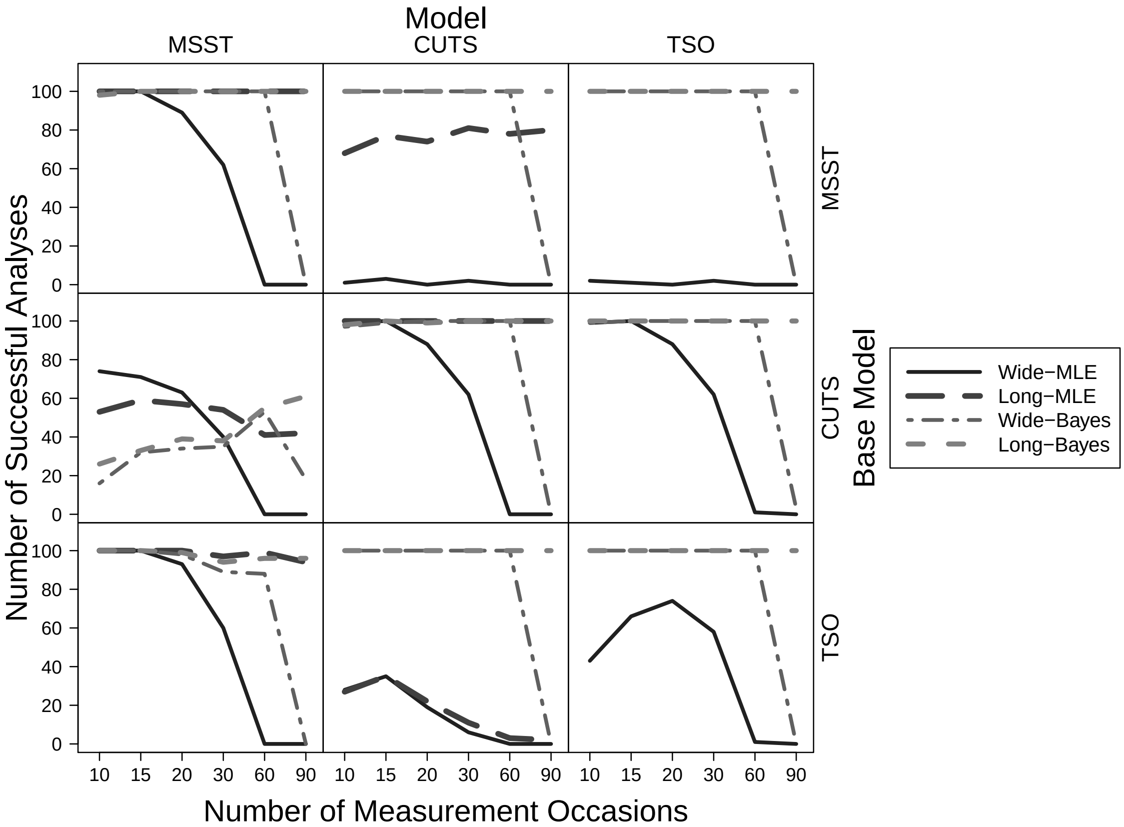

Figure S8. Number of Successful Analyses per Condition with 10% Missingness and 1:3 Trait-State Variance Ratio

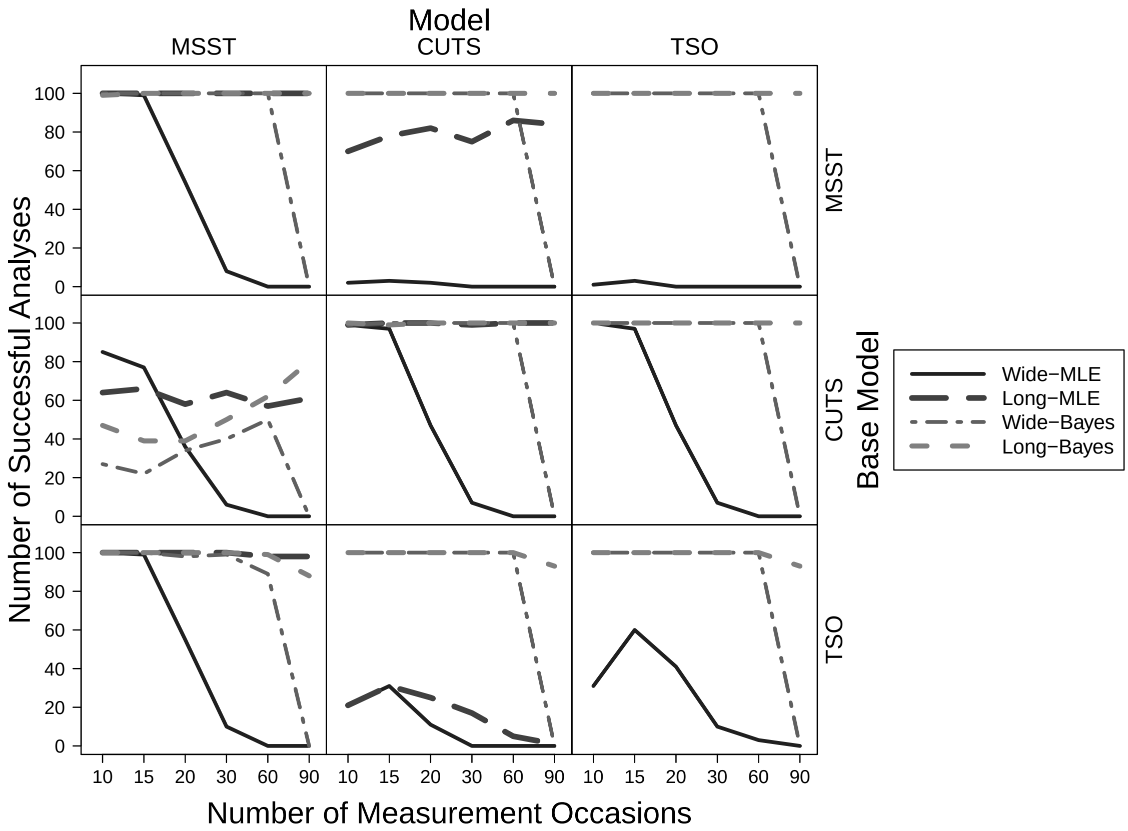

Figure S9. Number of Successful Analyses per Condition with 20% Missingness and 1:3 Trait-State Variance Ratio

Simulation: Fitting the Multilevel CUTS model to TSO data

In the simulation study, it was unexpected that the multilevel version of the CUTS model failed to converge to a reasonable solution in all the replications of certain conditions. Particularly, the CUTS model failed when it was estimated with maximum likelihood estimation (MLE) and the TSO model was used to generate the data. As a consequence, we decided to further investigate these results.

To start with, we repeated some analyses under the problematic conditions. We noted that the CUTS model was estimating at least one negative variance, which resulted in an improper solution. To find what was causing these results, we repeated part of the simulation changing the true parameters of the TSO model. Specifically, we decided to modify the correlation matrix used to simulate the latent trait indicator variables of the TSO. We followed the next process: (a) Generate a new correlation matrix based on a specific seed, (b) simulate 50 datasets with 30 measurement occasions and without missing values, (c) fit the multilevel CUTS model to the data by means of MLE, (d) count how many analyses failed. This procedure was repeated 14 times, using seven different seeds and with high (i.e., 0.7, 0.8, 0.9) and low (i.e., 0.5, 0.6, 0.7) correlations.

The results of the analyses with high correlations are shown in Table S9. This table presents the seed used to generate the correlation matrix, the correlation matrix, its determinant, and the number of analyses that failed out of the 50 replications. The first row is actually the correlation matrix used in the simulation of this study, which failed in 50 out of 50 analyses. Next, by generating the correlation matrix based on a different seed, the results change dramatically. In some cases, the CUTS model did an excellent job (seeds 13002 and 666), in other cases, the CUTS model was mediocre (seeds 13003, 13004, and 2019), and in others, it failed completely (seeds 13001 and 13014).

We also repeated the previous simulations but with low correlations. The results are shown in Table S10. In these cases, the CUTS model did, generally speaking, a good job. There were only two situations were there was at least one analysis that failed, and the number of analyses that failed was rather small.

To sum up, it is a fact that the correlation matrix used to generated the TSO data has an effect on the number of analyses that fail when fitting the CUTS model by means of MLE. Importantly, if the correlations in the correlation matrix are high then it is more likely that analyses will fail. However, it is unclear what exactly in the correlation matrix results in more failed analyses.

Table S9

Results of Simulations Varying the Correlation Matrix of High Correlations used to Generate TSO Data

High Correlations			
Seed	Correlation matrix	Determinant	# of Analyses that Failed
13001	$\begin{bmatrix} 1 & & & \\ 0.8 & 1 & & \\ 0.9 & 0.8 & 1 & \\ 0.9 & 0.7 & 0.7 & 1 \end{bmatrix}$	0.0077	50
13002	$\begin{bmatrix} 1 & & & \\ 0.7 & 1 & & \\ 0.8 & 0.9 & 1 & \\ 0.8 & 0.8 & 0.8 & 1 \end{bmatrix}$	0.0168	1
13003	$\begin{bmatrix} 1 & & & \\ 0.8 & 1 & & \\ 0.8 & 0.7 & 1 & \\ 0.9 & 0.9 & 0.8 & 1 \end{bmatrix}$	0.0117	15
13004	$\begin{bmatrix} 1 & & & \\ 0.9 & 1 & & \\ 0.7 & 0.9 & 1 & \\ 0.9 & 0.8 & 0.7 & 1 \end{bmatrix}$	0.0032	9
13014	$\begin{bmatrix} 1 & & & \\ 0.9 & 1 & & \\ 0.8 & 0.9 & 1 & \\ 0.7 & 0.9 & 0.8 & 1 \end{bmatrix}$	0.0045	49
666	$\begin{bmatrix} 1 & & & \\ 0.8 & 1 & & \\ 0.7 & 0.7 & 1 & \\ 0.7 & 0.7 & 0.7 & 1 \end{bmatrix}$	0.066	0
2019	$\begin{bmatrix} 1 & & & \\ 0.7 & 1 & & \\ 0.7 & 0.8 & 1 & \\ 0.7 & 0.7 & 0.9 & 1 \end{bmatrix}$	0.0288	7

Table S10

Results of Simulations Varying the Correlation Matrix of Low Correlations used to Generate TSO Data

Low Correlations			
Seed	Correlation matrix	Determinant	# of Analyses that Failed
13001	$\begin{bmatrix} 1 & & & \\ 0.6 & 1 & & \\ 0.7 & 0.6 & 1 & \\ 0.7 & 0.5 & 0.5 & 1 \end{bmatrix}$	0.1469	3
13002	$\begin{bmatrix} 1 & & & \\ 0.5 & 1 & & \\ 0.6 & 0.7 & 1 & \\ 0.6 & 0.6 & 0.6 & 1 \end{bmatrix}$	0.1616	0
13003	$\begin{bmatrix} 1 & & & \\ 0.6 & 1 & & \\ 0.6 & 0.5 & 1 & \\ 0.7 & 0.7 & 0.6 & 1 \end{bmatrix}$	0.1421	0
13004	$\begin{bmatrix} 1 & & & \\ 0.7 & 1 & & \\ 0.5 & 0.7 & 1 & \\ 0.7 & 0.6 & 0.5 & 1 \end{bmatrix}$	0.1236	0
13014	$\begin{bmatrix} 1 & & & \\ 0.7 & 1 & & \\ 0.6 & 0.7 & 1 & \\ 0.5 & 0.7 & 0.6 & 1 \end{bmatrix}$	0.1205	1
666	$\begin{bmatrix} 1 & & & \\ 0.6 & 1 & & \\ 0.5 & 0.5 & 1 & \\ 0.5 & 0.5 & 0.5 & 1 \end{bmatrix}$	0.28	0
2019	$\begin{bmatrix} 1 & & & \\ 0.5 & 1 & & \\ 0.5 & 0.6 & 1 & \\ 0.5 & 0.5 & 0.7 & 1 \end{bmatrix}$	0.21	0

Average Bias of the Autoregressive Effect

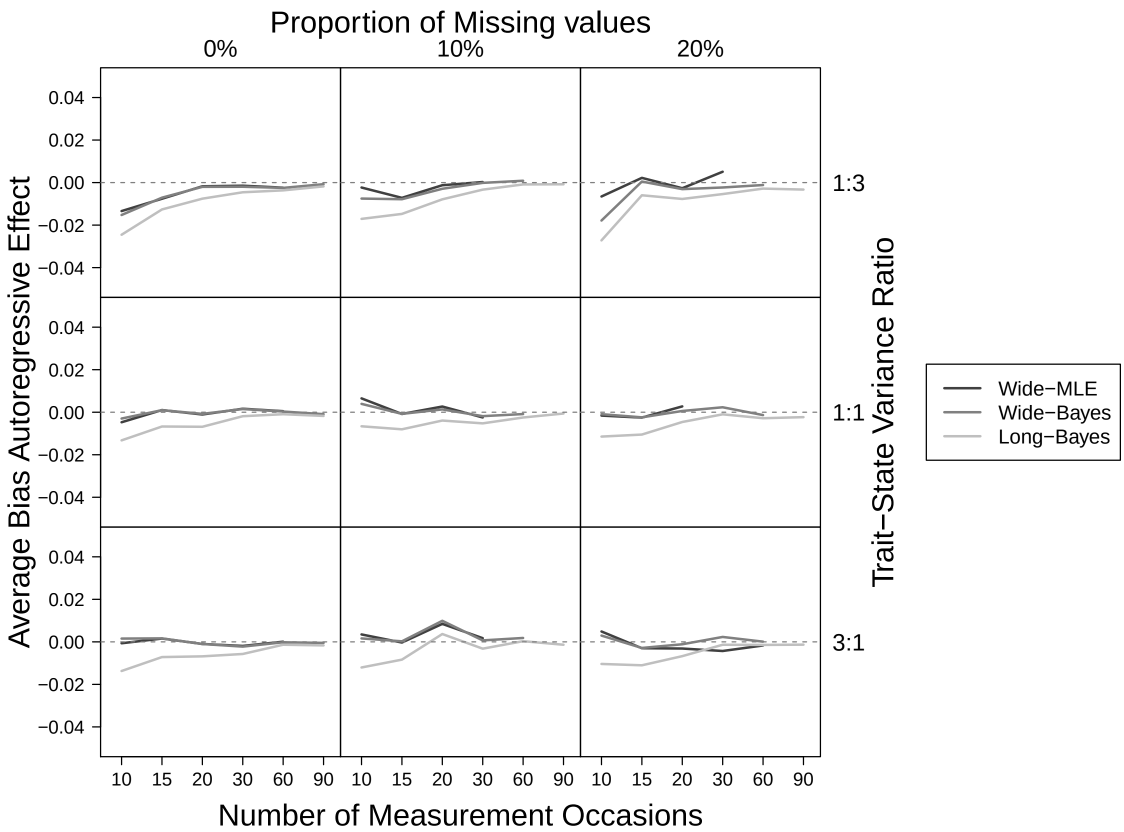

Figure S10. Average Bias of the Autoregressive Effect when fitting the single-level and multilevel TSO models to TSO data.

Empirical Example Excluding Non-stationary Time Series

Stationarity is an important assumption of state-trait SEMs when analyzing intensive longitudinal data. For this reason, the HND data were analyzed a second time but excluding individuals with at least one non-stationary time series. To test for stationarity, we used the Kwiatkowski–Phillips–Schmidt–Shin test (Kwiatkowski, Phillips, Schmidt, & Shin, 1992). This procedure reduced our sample size from 644 to 376 individuals. Next, we present the fit measures of the MSST, the CUTS, and TSO models in Table S11. The DICs indicated that the TSO model fit best the data. Moreover, Tables S12 and S13 present the estimates of the TSO model for the two samples: The sample with 644 individuals and the sample with 376 individuals that excludes non-stationary time series. Finally, Table S14 presents the estimates of the variance coefficients for the TSO model on the sample without non-stationary time series.

Although there were some small differences in the estimates, they were not large enough to change our interpretation of the results. For example, the autoregressive effect of the multilevel TSO model fitted to the items of positive affect deactivation went from 0.37 with the whole sample to 0.32 with the stationary sample (see S12). However, the fact that the differences were not substantial does not mean that stationarity is not a required assumption for state-trait SEMs. Simply, in this particular data the violations of stationarity were not large enough to bias our results. In general, we advice practitioners to always test for stationarity and to fit the models with both the whole sample and the stationary sample. If the results from these analyses lead to different interpretations, the results from the stationary sample should be preferred.

Table S11

ppp and DIC of the Three Models for the Two Sets of Items

		MSST	CUTS	TSO
PAD	ppp	0.614	0.655	-
	DIC	688083.833	680948.219	656303.508
PAA	ppp	0.603	0.650	-
	DIC	687287.995	680842.489	647618.688

Table S12

Estimates of the Multilevel TSO for the Items of Positive Affect Deactivation

	N = 644	N = 376
Parameter	$\hat{\theta}$ (Credibility Interval)	$\hat{\theta}$ (Credibility Interval)
Within Loading R	**1**	**1**
Within Loading Co	0.89 (0.88, 0.9)	0.89 (0.87, 0.91)
Within Loading Ca	0.89 (0.87, 0.9)	0.86 (0.84, 0.88)
Autoregressive Effect	0.37 (0.36, 0.38)	0.32 (0.3, 0.33)
Within Variance R	74.15 (72.01, 76.23)	71.74 (68.87, 74.58)
Within Variance Co	135.45 (133.12, 137.79)	136.67 (133.63, 139.93)
Within Variance Ca	130.34 (128.14, 132.6)	132.21 (129.34, 135.27)
Latent Occasion-Specific Residual Variance: PAD	143.56 (140.3, 146.98)	153.02 (148.56, 157.57)
Latent Trait Indicator Variance: R	124.38 (111.26, 139.93)	104.89 (90.66, 122.35)
Latent Trait Indicator Variance: Co	157.47 (141.13, 176.96)	141.34 (121.83, 164.55)
Latent Trait Indicator Variance: Ca	138.43 (123.97, 155.18)	121.22 (104.68, 141.38)
Covariance R-Co	119.75 (106.39, 135.84)	104.95 (89.11, 123.77)
Covariance R-Ca	119.04 (105.89, 134.48)	100.89 (86.2, 118.49)
Covariance Co-Ca	117.98 (104.05, 134.32)	104.87 (88.3, 124.26)

Table S13

Estimates of the Multilevel TSO for the Items of Positive Affect Activation

	N = 644	N = 376
Parameter	$\hat{\theta}$ (Credibility Interval)	$\hat{\theta}$ (Credibility Interval)
Within Loading Eg	**1**	**1**
Within Loading Et	1.13 (1.11, 1.14)	1.13 (1.11, 1.14)
Within Loading Ch	1.05 (1.04, 1.06)	1.06 (1.05, 1.08)
Autoregressive Effect	0.32 (0.31, 0.33)	0.28 (0.26, 0.29)
Within Variance Eg	140.1 (137.91, 142.36)	140.66 (137.75, 143.57)
Within Variance Et	68.83 (66.92, 70.72)	68.51 (66, 71.04)
Within Variance Ch	92.11 (90.21, 94.06)	92.41 (89.9, 94.99)
Latent Occasion-Specific Residual Variance: PAD	158.55 (155.06, 162.17)	163.11 (158.39, 167.91)
Latent Trait Indicator Variance: Eg	144.93 (129.81, 162.75)	119.77 (103.65, 139.86)
Latent Trait Indicator Variance: Et	148.82 (133.23, 167.41)	128.13 (110.17, 149.22)
Latent Trait Indicator Variance: Ch	152.34 (136.63, 170.84)	125.63 (108.52, 146.89)
Covariance Eg-Et	130.74 (116.22, 147.84)	110.42 (94.28, 129.97)
Covariance Eg-Ch	127.22 (112.71, 144.05)	105.07 (89.49, 124.2)
Covariance Et-Ch	138.49 (123.49, 156.24)	115.51 (98.71, 135.75)

Table S14

Variance Coefficients of the Three Models for the Two Sets of Items

	Variance Coefficient	Items		
		Relaxed	Content	Calm
PAD	Reliability	0.79	0.67	0.65
	Consistency	0.35	0.38	0.35
	Predictability by Trait	0.30	0.34	0.32
	Unpredictability by Trait	0.05	0.03	0.03
	Occasion Specificity	0.44	0.29	0.30
		Energetic	Enthusiastic	Cheerful
PAA	Reliability	0.68	0.84	0.78
	Consistency	0.30	0.34	0.34
	Predictability by Trait	0.27	0.30	0.30
	Unpredictability by Trait	0.03	0.04	0.04
	Occasion Specificity	0.37	0.49	0.44

References

Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. How sure are we that economic time series have a unit root? *Journal of Econometrics*, *54*(1-3), 159–178. doi:10.1016/0304-4076(92)90104-Y