

## Abstract

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

**Document Version:** Generated using [AdversarialSimulation](#) in state of Git commit SHA e45bcd6

## Table of contents

|   |           |
|---|-----------|
| <b>Introduction</b>                                       | <b>3</b>  |
| <b>Methods</b>  | <b>4</b>  |
| A Framework for Adversarial Collaboration . . . . .       | 4         |
| Methods of Individual Simulation Studies . . . . .        | 4         |
| Studies by Kriegmair . . . . .                            | 4         |
| Studies by Kosanke . . . . .                              | 9         |
| Methods of “Joint” Simulation Study . . . . .             | 12        |
| <b>Results</b>  | <b>14</b> |
| Results of Individual Simulation Studies . . . . .        | 14        |
| Results of Kriegmair’s Simulation Studies . . . . .       | 14        |
| Results of Kosanke’s Simulation Studies . . . . .         | 16        |
| Results of the “Joint” Simulation Study . . . . .         | 16        |
| Results of the Adversarial Collaboration . . . . .        | 16        |
| <b>Discussion</b>   | <b>16</b> |
| Discussing the substantial results . . . . .              | 16        |
| Disucssing the Adversarial Collaboration . . . . .        | 16        |
| <b>References</b>   | <b>16</b> |
| <b>Appendix</b>   | <b>17</b> |
| Appendix A: Simulation Protocol . . . . .                 | 17        |
| Appendix B: Supplementary Figures . . . . .               | 27        |
| Appendix C: Detailed Error and Warning Messages . . . . . | 29        |

## Introduction

Work in progress, just a placeholder for now: Karl Popper once described science as the art of “systematic over-simplification.” This term ironically yet accurately describes the very basic cycle of empirical research, where we lay out general claims about the world as hypotheses, translate them into measurable constructs, and choose how to gather data from certain populations, which finally, in turn, updates our beliefs about general verbal claims about the world. One core challenge in every research endeavor is mapping the general to the specific when designing and conducting a study, or from the specific and empirical to the verbal and general when interpreting the results. This mapping appears to be the crux of most research and is where most, if not all, of a researcher’s degrees of freedom lie. Divergences in these mappings can be the source of ambiguity and verbal disputes. When differences in mappings from the verbal plane are not recognized or lack transparency, persistent and seemingly unresolvable disagreements in the general and verbal domains can occur. This challenge transfers to Monte Carlo simulation studies. These are commonly used tools to test statistical methods in simulated data to evaluate any method against a known ground truth. As it is impossible to simulate and test every possible data and analysis model combination, researchers are confronted with a multitude of degrees of freedom and decisions about what “prototypical” models to test in which “prototypical” data and settings. Especially the comparison of different methods in their “general” applicability and performance for various research settings is prone to conflicting verbal claims based on diverging simulation decisions. Biases for a specific method developed by one researcher might additionally amplify these divergences, not only at the step of interpreting results but importantly also when designing a simulation. To address these challenges of entrenched disagreements, the practice of adversarial collaboration has been proposed to unveil discrepancies in underlying methodological decisions and assumptions. It was famously pioneered by Ralph Hertwig and Daniel Kahneman, who tried to settle a persistent scientific disagreement about frequency representation and consulted Barbara Mellers as a neutral arbiter. Today, it is recognized as a potent tool in the social empirical research community. The basic idea is for two researchers in disagreement to first identify a general verbal dispute and agree on a research question to settle the debate. Based on this, they collaboratively work on operationalizing, testing, and interpreting this verbal claim. This process aims to unveil and concretize underlying disagreements and thus reduce ambiguity and increase generalizability. In this project, we aimed to transfer the concept of adversarial collaboration from the empirical domain to Monte Carlo simulation studies and assess its feasibility and viability in a case study in this context. To conduct such an exemplary adversarial collaboration, we first need a framework that structures the collaborative process tailored to the outline of simulation studies. Traditional SEM methods, like maximum likelihood estimation, optimize all parameters of a model simultaneously under the

assumption of multivariate normality. While powerful and although robust estimation techniques relax the normality assumption, all system wide estimators suffer from several shortcomings, they often face issues such as non-convergence, improper solutions (with parameters out of definitional range), and biases from local measurement misspecifications that affect the entire model. They also typically require large sample sizes for adequate performance, especially in complex models.

## Methods

### A Framework for Adversarial Collaboration

We developed a specific adversarial simulation framework and structured the collaboration into two rounds. In the first round, each collaborator independently conducts a separate simulation study. In the second round, they come together to work on a joint study, building on the findings from the first round. This two-step approach is designed to highlight differences in a systematic way and to establish a virtual foundation for collaboration before engaging in a joint effort in our case study.

### Methods of Individual Simulation Studies

#### *Studies by Kriegmair*

The methodological setup of my individual simulation studies follows the structure we established for our *adversarial simulation* framework to facilitate stepwise collaboration. It is based on a preregistered protocol but includes some deviations from the preregistration ([See Appendix A](#) for the full protocol and all deviations from the preregistration).

In the initial phase of our case study, I independently conducted two separate simulation studies without my collaborator’s involvement with the goal to conceptually replicate the findings regarding SAM compared to standard SEM estimation of Rosseel & Loh (2022) and Dhaene & Rosseel (2023). However, there are several differences in the design and setup of the studies compared to the original studies as outlined below.

**Aims, objectives and research questions** Both studies aimed to evaluate the performance of vanilla SEM (with maximum likelihood) compared to global SAM (gSAM), local SAM with maximum likelihood (lSAM-ML), and local SAM with unweighted least squares (lSAM-ULS) under various conditions. The two research questions we jointly established prior to conducting the studies served as general basis for both studies:

1. How do SAM and traditional SEM methods (including ML and ULS) compare in terms of bias, Mean Squared Error (MSE), and convergence rates in small to moderate samples?
2. What is the impact of model misspecifications, such as residual correlations and cross-loadings, on the performance of SAM compared to traditional SEM methods?

## Population Models and Data Generation Mechanisms

**Study 1** Data were generated based on a 5-factor population structural model with 3 indicators for each factor. Four different models were simulated (see figure 1-4). In line with Rosseel & Loh (2022) this model design was chosen to represent a realistic model with sufficient complexity to pose a challenge for the estimation methods, especially in the presence of misspecifications:

- Model 1.1: Correctly specified model.
- Model 1.2: Misspecified with cross-loadings in the population model that are ignored in the estimation model (model 1.1)
- Model 1.3: Misspecified with correlated residuals and a reversed structural path between the third and fourth latent factors in the population model that are ignored in the estimation model (model 1.1)
- Model 1.4: Misspecified with a bidirectional structural relation between factors 3 and 4 specified as only one directional

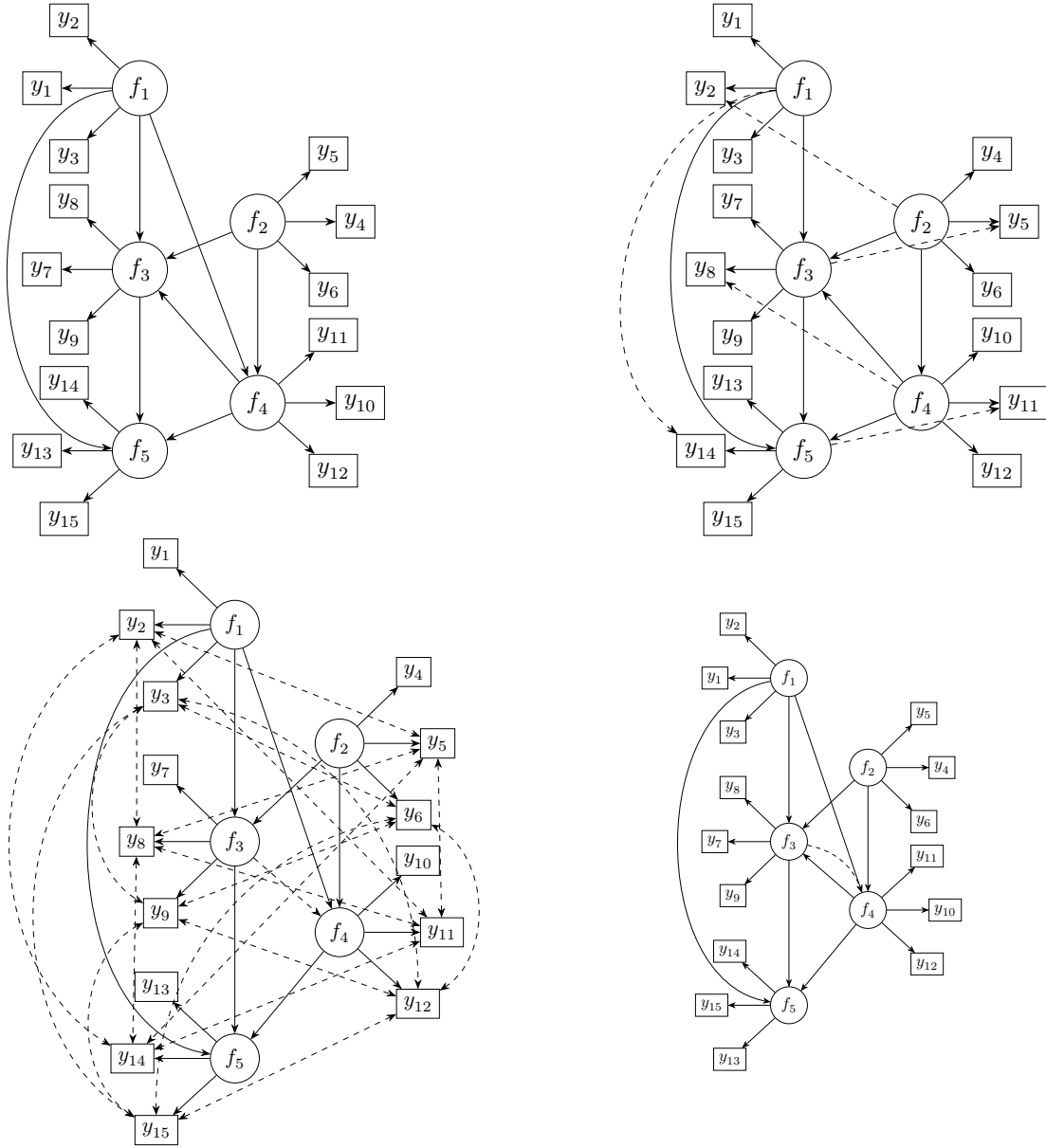
Factor loadings were fixed across all reliability conditions, with the first indicator of each factor serving as the scaling indicator ( $\lambda = 1.0$ ), and the other two indicators having loadings of 0.7. Indicator reliability levels were manipulated by adjusting the measurement error variances in the  $\Theta$  matrix. Specifically, the a reliability value was set at different levels (low = 0.3, moderate = 0.5 or high = 0.7) to compute the respective error variances on the diagonal of  $\Theta$ :  $\Theta^* = \text{Var}(\eta)\Lambda^T \times \frac{1}{r-1}$ .

To investigate additional possible and realistic scenarios beyond the ones studied by Rosseel & Loh (2022) model 1.3 included a combination of measurement and structural misspecifications as opposed to only measurement misspecifications to introduce an even more severely misspecified model under which SAM methods might perform even better than traditional SEM. Further, model 1.4 included a (not estimated) bidirectional structural relation between factors 3 and 4 as opposed to the unidirectional reversed one. For all models, the population-level values of the structural parameters were set to 0.1.

**Study 2** Data were generated based on a 5-factor population structural model with 3 indicators for each factor with loadings set to 1, 0.9 and 0.8 for each factor and reliability modulated like in study 1. Regression weights were set to either 0.183 and 0.224 (low) or 0.365 and 0.447 (medium). This should represent varying variance explained ( $R^2$ ) by the endogenous factors set at low ( $R^2 = 0.1$ ) or medium ( $R^2 = 0.4$ ). Note however that the computation of this was a simplification and does

**Figure 1**

*Population Model Variations of Study 1*



*Note.* Error terms are not explicitly shown in the figure. Dashed lines represent relations omitted in the estimation model present in the population model.”

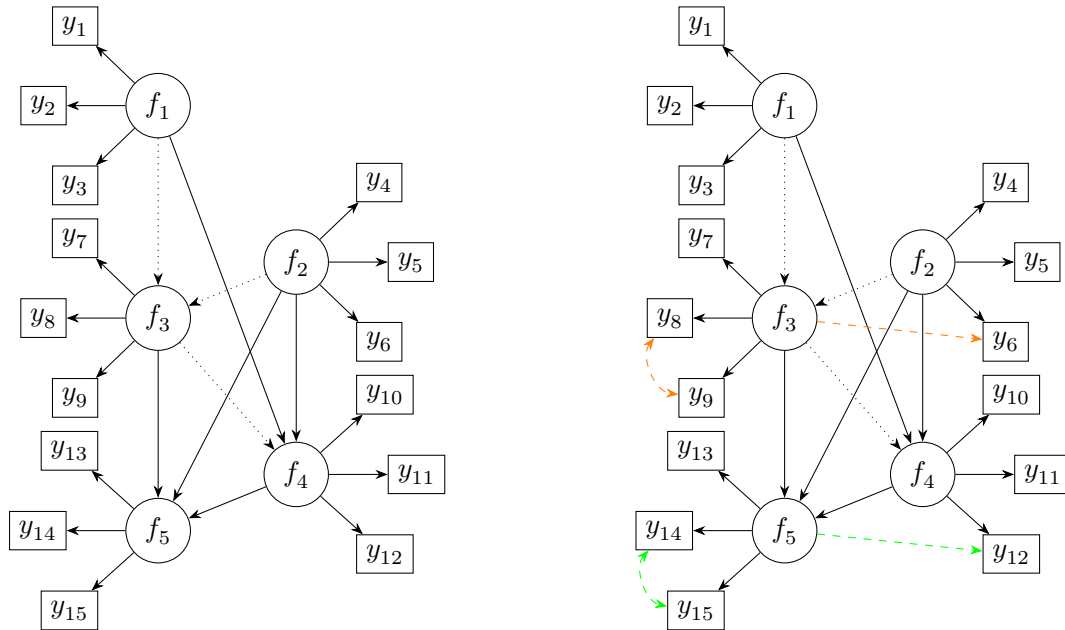
not accurately result in said  $R^2$  values. The aim here was only to generally modulate between lower and higher regression weights. The population models resulted in the following model types with varying misspecification in the estimation model:

- Model 2.1: structural misspecification with falsely specified paths in the estimation model absent in the population model (Figure 5)
- Model 2.2: correlated residuals and a factor cross-loading in either the exogenous (Model 2.2-exo), endogenous (Model 2.2-endo) part of the model or both (Model 2.2-both) with falsely specified paths in the estimation model absent in the population model (see figure 5-6).

To enable the analysis of the impact of falsely specified paths in the estimation model that are not present in the population model and how well the different methods recover these non-existing relations both population models included several such misspecifications in addition to the measurement misspecifications evaluated by Dhaene & Rosseel (2023).

**Figure 2**

*Population Model Variations of Study 2*



*Note.* Error terms are not explicitly shown in the figure. Dotted paths represent relations specified in the estimation model not present in the population model. For Model 2.2, orange lines represent misspecifications in the exogenous part of the model, and green lines represent misspecifications in the endogenous part of the model. These types of misspecifications result in different realizations of model 2.2 when they are modulated as factors in study 2 but are subsumed under one model here.

## Experimental Design of simulation procedures

**Study 1** Study 1 varied three main conditions: (1) sample sizes of small ( $N = 100$ ), moderate ( $N = 400$ ), and large ( $N = 6400$ ); (2) Indicator reliability of low ( $= 0.3$ ), moderate ( $0.5$ ), high ( $= 0.7$ ); (3) Model specifications: correctly specified model and misspecified with not specified cross loadings in the population model (see figure 2), misspecified with not-specified correlated residuals and a reversed structural path between the the third and the fourth latent factor in the population model (see figure 3) and a recursive structural relation between factor 3 and 4 in the population specified as only one directional (see figure 4).

**Study 2** Study 2 varied five conditions: (1) sample sizes: small ( $N = 100$ ), medium ( $N = 400$ ), and large ( $N = 6400$ ). (2) Variance explained by endogenous factors: low ( $R^2 = 0.1$ ) and medium ( $R^2 = 0.4$ ). (3) Indicator reliability: low ( $0.3$ ), moderate ( $0.5$ ), and high ( $0.7$ ). (4) Model misspecifications: varying the population model by omitting a residual covariance and a factor cross-loading in different parts of the model. (5) Number of measurement blocks: separate measurement model per latent variable ( $b = 5$ ) and joint measurement model for all exogenous variables ( $b = 3$ ) for the local SAM condition (ISAM-ML).

**Method Selection** Both studies compared the performance of four estimation methods: Vanilla SEM with maximum likelihood (ML), Global SAM with maximum likelihood (gSAM), Local SAM with maximum likelihood (ISAM-ML), Local SAM with unweighted least squares (ISAM-ULS).

**Performance Measures** For both studies convergence rates were tracked via lavaan’s built-in function that indicates convergence. Further, improper solutions, converged models that showed negative variances (as the only type of improper solution present), were tracked via lavaan warning messages. Next of all converged and proper solutions bias ( $\bar{T} - \theta$ ), and RMSE ( $\sqrt{\frac{1}{K} \sum_{k=1}^K (T_k - \theta)^2}$ ) where  $T_k$  is the estimated parameter,  $\bar{T}$  the mean of the estimated parameters and  $\theta$  the true parameter value, and  $K$  is the number of replications computed. For comparability across varying regression weights for study 2, relative bias ( $\frac{\bar{T} - \theta}{\theta}$ ) and relative RMSE ( $\sqrt{\frac{(\bar{T} - \theta)^2 + S_T^2}{\theta^2}}$ ) were computed. Monte Carlo standard errors (MCSE) were computed for bias and RMSE as well as relative bias and relative RMSE:  $\sqrt{\frac{S_T^2}{K}}$  and  $\sqrt{\frac{S_T^2}{K\theta^2}}$  for bias and relative bias, and  $\sqrt{\frac{K-1}{K} \sum_{j=1}^K (\text{RMSE}_{(j)} - \text{RMSE})^2}$  and  $\sqrt{\frac{K-1}{K} \sum_{j=1}^K (r\text{RMSE}_{(j)} - r\text{RMSE})^2}$  for RMSE and relative RMSE.

**Software** All analyses were conducted in R Core Team (2023). Simulation and estimation was done using Rosseel (2012). To ensure reproducibility and avoid synchronization in parallelized a pre-generated list of seeds was used for all replications. For further details and a complete list of libraries and dependencies, visit <https://github.com/valentinkm/AdversarialSimulation>.

**Analysis and Interpretation plan** Similar to the studies by Rosseel & Loh (2022) and Dhaene & Rosseel (2023) results were interpreted by descriptively comparing the performance measures of



the different estimation methods under varying sample sizes, indicator reliability levels, and model misspecifications without predetermined cut-off values or critical distances. Performance metric values were aggregated across all parameters excluding the misspecified parameters (present in the population but not in the estimation model).

### *Studies by Kosanke*

**Quoted verbatim from Kosanke’s report** (Git commit SHA 4d0e95e):

The structure of this section closely aligns to our agreed upon structure of simulation studies in Table 1.

In a first step, I published a simulation protocol containing all the planned analysis to be replicated from the original paper by Robitzsch (2022). This protocol can be accessed here: [https://github.com/lkosanke/AdversarialSimulation/blob/main/LK/simulation\\_protocol.pdf](https://github.com/lkosanke/AdversarialSimulation/blob/main/LK/simulation_protocol.pdf).

#### **Aims, objectives and research questions**

For my individual study, I replicated parts of Robitzsch (2022) that were relevant to our two substantive research questions. Overall, I conducted 6 simulation studies.

#### **Population Models and Data Generation Mechanisms**

The most important details with regards to the population models and data-generating mechanisms are visible in Table 7. With regards to the population models, all factors in all studies loaded onto 3 indicators each. I chose the population values to align with the original paper by Robitzsch (2022). The multivariate normally distributed data was generated parametrically, based on a specified population model. All simulations were conducted using seeds to allow for the reproducibility of results.

For more details on the exact values of each study, see the simulation scripts in the Github repository.

**Figure 3***Overview of Simulation Studies Conducted by Kosanke*

| Study    | Model        | Correct model included? | Unmodelled RC               | Unmodelled CL               | N Sizes | $\varphi$ / $\beta$   | $\lambda$ |
|----------|--------------|-------------------------|-----------------------------|-----------------------------|---------|-----------------------|-----------|
| Study 1  | 2-factor-CFA | Yes                     | 1 and 2, both pos. and neg. | x                           | 7       | $\varphi = 0.6$       | Fixed     |
| Study 1b | 2-factor-CFA | Yes                     | x                           | x                           | 2       | $\varphi = 0.2 - 0.8$ | Varied    |
| Study 2  | 2-factor-CFA | x                       | x                           | 1 and 2, both pos. and neg. | 7       | $\varphi = 0.6$       | Fixed     |
| Study 3  | 2-factor-CFA | x                       | 1, pos.                     | 1, pos.                     | 7       | $\varphi = 0.6$       | Fixed     |
| Study 4  | 5-factors    | Yes                     | 20, all pos.                | 5, all pos.                 | 7       | $\beta = 0.1$         | Fixed     |
| Study 4a | 5-factors    | x                       | 20, all pos.                | 5, all pos.                 | 7       | $\beta = 0.1 - 0.4$   | Fixed     |

*Note.*  $\Phi$ : factor correlation, N: sample size,  $\lambda$ : factor loading,  $\sigma$ : residual variance,  $\tau$ : factor variance, RC: residual correlations, CL: cross-loadings, CFA: confirmatory factor analysis,  $\beta$ : regression coefficient between factors.

**Experimental Design of simulation procedures**

Overall, 3 different types of factors were varied that can be deduced from Table 7 and are detailed again in the simulation scripts provided.

Firstly, I varied the sample size in all studies, ranging from  $N = 50$  to 100.000. I included a smaller sample size  $N=50$  for all studies, to be able to answer our substantive research questions in more detail. Study 1b explicitly investigated the small sample bias of LSAM estimation in low sample sizes. Thus, only  $N=50$  and  $N=100$  were present in this study.

Additionally, I varied the amount of misspecification in all studies, either via different numbers of unmodelled residual correlations, cross-loadings, or both.

Thirdly, in Studies 1b and 4a, I varied the population values for three model parameters ( $\phi$ ,  $\beta$  and/ or  $\lambda$ ).

Besides studies 1 and 2, I implemented full factorial designs. In Studies 1 and 2 I omitted conditions where both one positive and one negative value would be present. I hypothesize that this was done in Robitzsch (2022) to avoid cancellation of biases, but the authors did not give reasoning for this decision themselves.

In Studies 4 and 4a I investigated the differential performance of the estimators in a model that included a non-saturated structural model (i.e. regressions between some of the factors). These

studies were replications not only of the paper by Robitzsch (2022), but of the first paper on the SAM approach by Rosseel & Loh (2022). In contrast to the other studies, studies 4 and 4a differed in the way the misspecification variation was labelled in Robitzsch (2022). Instead of varying a factor misspecification as in the previous study, they varied 3 different data-generating mechanisms (DGM's) as a whole. Thus the conditions are labelled differently: DGM 1 contained no misspecification. DGM 2 contained 5 cross-loadings in the data-generating model, that were not modelled in the estimated models. DGM 3 contained 20 residual correlations that were not modelled in the models. I extended them to investigate the interaction of beta and N for the 5-factor regression model, as this again was of interest for our substantial research questions. Additionally, I omitted the inclusion of DGM 1 in Study 4a, as it neither contained misspecification (which is central to our research question), nor did it lead to interesting results in the original study.

### **Method Selection**

In terms of estimation methods, I used constrained SEM maximum-likelihood (SEM-ML) and unweighted-least-squares estimation (SEM-ULS), so that loadings and variance parameters were given the constraints that they had to be positive and larger than 0.01. Additionally, I implemented local-SAM (LSAM) and global-SAM (GSAM) estimation, in both maximum-likelihood (LSAM-ML/ GSAM-ML), and unweighted-least-squares estimation (GSAM-ML/ GSAM-ULS) contexts. Exceptions were studies 1b, 4 and 4a, where only LSAM was investigated, as results did not really differ between the two different SAM-methods (Robitzsch, 2022).

### **Performance Measures**

I calculated the bias and RMSE of the estimated factor correlations in all studies, as well as the standard deviation of the one factor correlation present in Studies 1,2 and 3. For the type of bias calculated, I oriented on Robitzsch (2022), besides in Study 1b. Thus, I calculated average relative bias in Studies 1, 2 and 3, and average absolute bias in Studies 1b, 4 and 4a. In Study 1b, I took the absolute value to see if negative and positive biases canceled each other out in the original study for conditions with lower phi values. In addition to what was done in Robitzsch (2022), I calculated confidence intervals for the bias estimates, but omitted them in the results tables for presentation purposes. The exact computation of the performance measures is detailed in the simulation scripts and results.pdf file in my sub-folder of the Github repository.

I did not include a detailed mechanism to capture model convergence as detailed in the first substantive research question. As Robitzsch (2022) argued in their paper, and was shown already in other simulations, using constrained maximum likelihood estimation should resolve convergence issues of classical maximum likelihood estimation in smaller samples (Lüdtke et al., 2021; Ulitzsch et al., 2023). I did include, however, a mechanism to track the total number of warnings for each

estimation and compare it to the total number of estimations as a sanity check.

### **Software**

All analyses were conducted in R (R Core Team, 2023). I used the packages lavaan, purrr, tidyverse, furrr to conduct the simulations, as well as knitr and kableExtra for presenting the results (Rosseel, 2012; Zhu et al., 2024; **purrr**?; **tidyverse**?; **furrr**?; **knitr**?).

### **Analysis and Interpretation plan**

For the interpretation of results, I oriented on cut-offs that were used in the original paper by Robitzsch (2022). For bias, I interpreted differences of 0.05 or higher as substantial. For SD, I explicitly mentioned percentage reductions of more or equal to 5%. For RMSE, the same interpretation was used for differences of 0.03 or higher. The simulation was repeated 1500 times for each Study.

### **Methods of “Joint” Simulation Study**

After collaborating based on the individually conducted studies and the respective results, we did not jointly arrive at the conclusion that conducting a collaborative simulation study as planned was warranted.

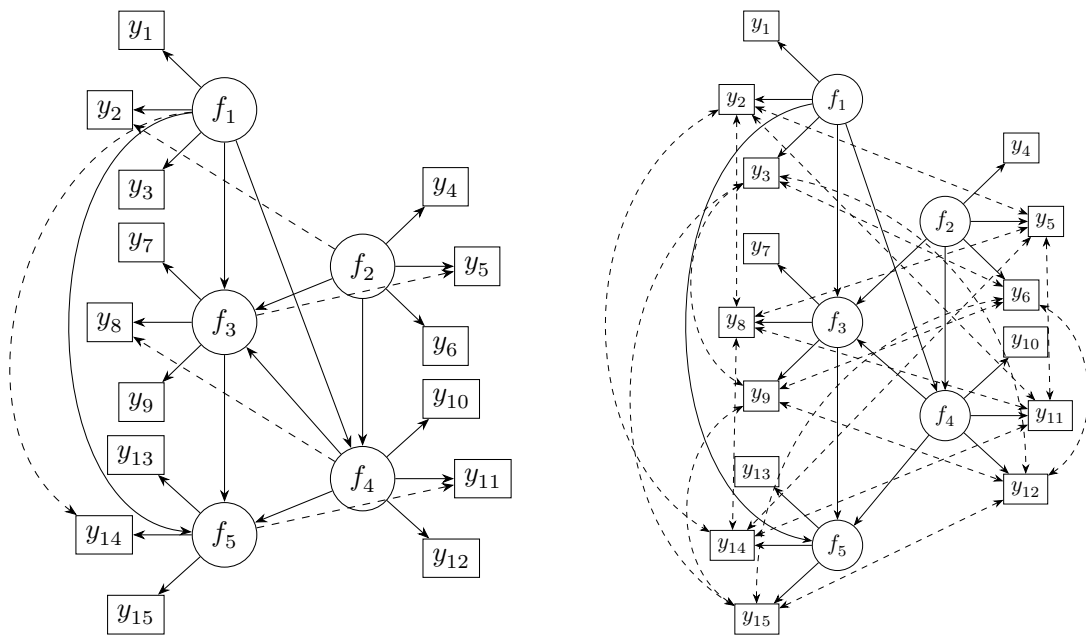
However I identified several reasons for setting up another simulation. Firstly, to test and evaluate the viability and technical feasibility of AC for simulation studies, setting up a study based on the individual studies, their results and with Kosanke can provide valuable...

**Aims, objectives and research questions** Following our framework for collaboration the research questions for the joint study remains the same as specified prior to the individual studies.

**Population Models and Data Generation Mechanisms** As in study 1 and 2 data for study 3 was generated based on a 5-factor population structural model with 3 indicators for each factor. Factor loadings and indicator reliability was computed in the same way as in study 1. Two different population models were simulated that resulted in misspecifications of either omitted crossloadings (model 3.1) or omitted correlated residuals (model 3.2). The population-level values of the structural parameters were set to 0.1. Reliability levels were manipulated as in Study 1. The omitted crossloadings (see figure 7) could either be all positive or negative and were set to be 10% lower in absolute values than the factor loadings. Correlated residuals were also either all positive or all negative and were set to not exceed a factor of 0.6 of the residual variances of the indicators.

**Figure 4**

*Population Model Variations for Study 3*



*Note. Note.* Error terms are not explicitly shown in the figure. Dashed lines represent relations omitted in the estimation model present in the population model. Unspecified crossloadings and correlated residuals could be either positive or negative resulting in 2 modulations of model 3.1 and 3.2 in the study.

**Experimental Design of simulation procedures\*\*** The joint study varied three conditions: (1) sample sizes of very small ( $N = 50$ ), small ( $N = 100$ ) or moderate ( $N = 400$ ). (2) Indicator reliability of low ( $= 0.3$ ), moderate ( $0.5$ ) or high ( $= 0.7$ ); (3) Model misspecifications with not-specified cross loadings in the population model that were positive or negative (see figure ) or not-specified correlated residuals in the population model that were positive or negative (see figure 8).

**Method Selection** To address the question ...

### Performance Measures

**Software** To fully evaluate the effect of bound SEM on convergence convergence rate and rate of proper solutions were tracked condition wise (Kriegmair's individual studies).

**Analysis and Interpretation plan** The analysis was conducted largely in the same way as in the individual studies

## Results

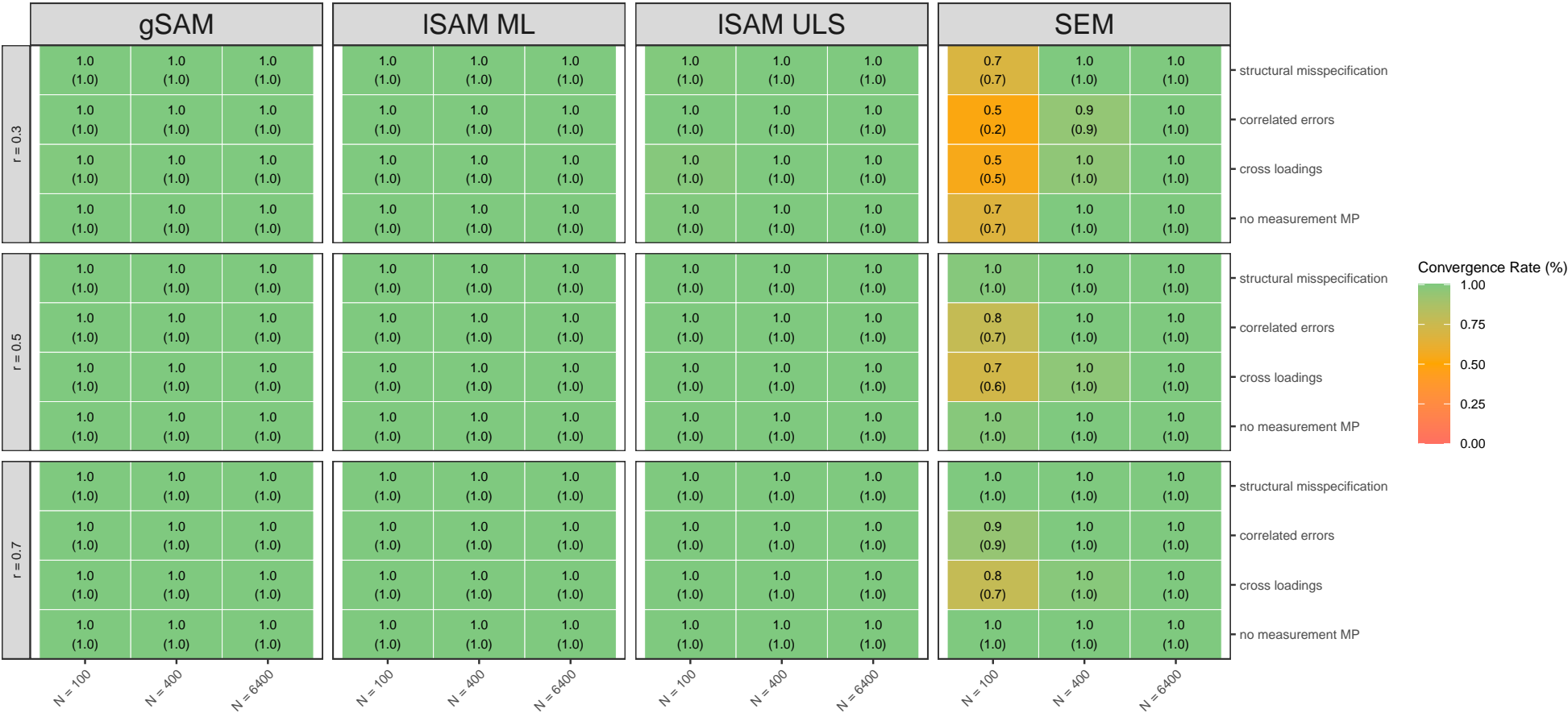
### Results of Individual Simulation Studies

#### *Results of Kriegmair's Simulation Studies*

As shown in Figure 5 there were no convergence issues for all SAM methods (gSAM, lsAM ML and ULS) with a convergence rate of 100% and no improper solutions across all conditions even in small samples with low reliability. Standard SEM however showed severe convergence issues in small samples with low to moderate reliability with a convergence rate of as low as 50% and 50% improper solutions in the cross loading misspecification condition as the most challenging condition.

Figure 5

Convergence Rate and Rate of Proper Solutions in Study 1



Note. Convergence and proper solutions (in parentheses) rates across sample sizes (N), reliability (r), and model misspecifications for global SAM (gSAM), local SAM with Maximum Likelihood (ISAM-ML), Unweighted Least Squares (ISAM-ULS), and SEM.

Convergence rates in study 2 were consistent with this with 100% convergence rates for all SAM methods and as low as 60% for standard SEM with exogenous measurement misspecifications posing more challenges than endogenous misspecifications (see Figure B1 in Appendix B).

## ***Results of Kosanke's Simulation Studies***

### **Results of the “Joint” Simulation Study**

### **Results of the Adversarial Collaboration**

## **Discussion**

### **Discussing the substantial results**

### **Disucssing the Adversarial Collaboration**

Idea: living simulations

## **References**

- A future for r: A comprehensive overview.* (n.d.). Retrieved April 23, 2024, from <https://cran.r-project.org/web/packages/future/vignettes/future-1-overview.html>
- Bartlett, M. S. (1937). The statistical conception of mental factors. *British Journal of Psychology. General Section*, 28(1), 97–104.  
<https://www.proquest.com/docview/1293463650/citation/9FE530638F5041E0PQ/1>
- Bartlett, M. S. (1938). Methods of estimating mental factors. *Nature*.
- Dhaene, S., & Rosseel, Y. (2023). An evaluation of non-iterative estimators in the structural after measurement (SAM) approach to structural equation modeling (SEM). *Structural Equation Modeling: A Multidisciplinary Journal*, 30(6), 926–940.  
<https://doi.org/10.1080/10705511.2023.2220135>
- Heidilohr. (n.d.). *Azure virtual desktop documentation*. Retrieved April 23, 2024, from <https://learn.microsoft.com/en-us/azure/virtual-desktop/>
- Helper functions for simulation studies.* (n.d.). Retrieved April 23, 2024, from <https://meghapsimatrix.github.io/simhelpers/>
- Kriegmair, V. (2024). *Preregistration: Comparing a structural after measurement (SAM) approach to standard structural equation model (SEM) estimation*. Zenodo.  
<https://doi.org/10.5281/zenodo.11459378>



- Lüdtke, O., Ulitzsch, E., & Robitzsch, A. (2021). A comparison of penalized maximum likelihood estimation and markov chain monte carlo techniques for estimating confirmatory factor analysis models with small sample sizes. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.615162>
- R Core Team. (2023). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Robitzsch, A. (2022). Comparing the robustness of the structural after measurement (SAM) approach to structural equation modeling (SEM) against local model misspecifications with alternative estimation approaches. *Stats*, 5(3), 631–672. <https://doi.org/10.3390/stats5030039>
- Rosseel, Y. (2012). Lavaan: An r package for structural equation modeling. *Journal of Statistical Software*, 48, 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Rosseel, Y., & Loh, W. W. (2022). A structural after measurement approach to structural equation modeling. *Psychological Methods*, No Pagination Specified–No Pagination Specified. <https://doi.org/10.1037/met0000503>
- TARDIS core package documentation — tardis*. (n.d.). Retrieved April 23, 2024, from <https://tardis-sn.github.io/tardis/>
- Ulitzsch, E., Lüdtke, O., & Robitzsch, A. (2023). Alleviating estimation problems in small sample structural equation modeling—a comparison of constrained maximum likelihood, bayesian estimation, and fixed reliability approaches. *Psychological Methods*, 28(3), 527–557. <https://doi.org/10.1037/met0000435>
- Zhu, H., Travison, T., Tsai, T., Beasley, W., Xie, Y., Yu, G., Laurent, S., Shepherd, R., Sidi, Y., Salzer, B., Gui, G., Fan, Y., Murdoch, D., Arel-Bundock, V., & Evans, B. (2024). *kableExtra: Construct complex table with 'kable' and pipe syntax* (Version 1.4.0) [Computer software]. <https://cran.r-project.org/web/packages/kableExtra/index.html>

## Appendix

### Appendix A: Simulation Protocol

Here the full simulation protocol of my simulation studies conducted individually prior to collaboration as well as the follow up study I conducted in light of the collaboration with Kosanke after the first round of conducting and evaluating our individual studies is presented. It is based on the preregistration of my individual studies (Kriegmair (2024)) and outlines all deviations from it.

**Preregistration template designed by:** Björn S. Siepe, František Bartoš, Tim P. Morris, Anne-Laure Boulesteix, Daniel W. Heck, and Samuel Pawel

## 1. General Information

### *1.1. What is the title of the project?*

Comparing a Structural After Measurement (SAM) Approach to Standard Structural Equation Model (SEM) Estimation

### *1.2. Who are the current and future project contributors?*

Valentin Kriegmair

### *1.3. Provide a description of the project.*

The studies registered were part of an adversarial collaboration project. The aim was to conceptually (only in part) replicate the results obtained by Dhaene & Rosseel (2023) and Rosseel & Loh (2022). I set out to evaluate the performance of a Structural After Measurement (SAM) approach for estimating structural equation models (SEM) in comparison to standard SEM estimation methods. This served as the basis for the adversarial collaboration with another researcher who evaluated the same research question from the perspective of a conceptual replication of the (in part contradicting) results obtained by Robitzsch (2022). However, the following only describes the first (conceptual) replication.

### *1.4. Did any of the contributors already conduct related simulation studies on this specific question?*

No prior related simulation studies have been conducted by the contributors.

## 2. Aims

Structural After Measurement (SAM) is an estimation method for structural equation models that consists of a stepwise estimation of the measurement and structural parts of a model. The research questions of the current simulation were:

1. How do SAM and traditional SEM methods (including ML and ULS) compare in terms of bias, Mean Squared Error (MSE), and convergence rates in small to moderate samples?
2. What is the impact of model misspecifications, such as residual correlations and cross-loadings, on the performance of SAM compared to traditional SEM methods?

### 3. Data-Generating Mechanism

#### 3.1. How will the parameters for the data-generating mechanism (DGM) be specified?

**3.1.1 Study 1** In study 1 (conceptually replicating Rosseel & Loh (2022)) data was generated parametrically. Four different population structural equation models (SEM) with latent variables and continuous indicators based on the following matrices were simulated:

- $B$  as  $M \times M$  matrix representing latent regression coefficients with all  $b = 0.1$ .

– **Model 1.1 and 1.2:**

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0.1 & 0.1 & 0 & 0.1 & 0 \\ 0.1 & 0.1 & 0 & 0 & 0 \\ 0.1 & 0 & 0.1 & 0.1 & 0 \end{bmatrix}$$

- **Model 1.3** in deviation from the preregistration with a reversed effect between latent factors f3 and f4 to introduce another realistic and more severe misspecification to show the potential of SAM in most challenging conditions:

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0.1 & 0.1 & 0 & 0 & 0 \\ 0.1 & 0.1 & 0.1 & 0 & 0 \\ 0.1 & 0 & 0.1 & 0.1 & 0 \end{bmatrix}$$

- **Model 1.4** in deviation from the preregistration with a bidirectional structural relation between f3 and f4 specified as only one directional instead of just reversing the effect to investigate a different type of misspecification:

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0.1 & 0.1 & 0 & 0.1 & 0 \\ 0.1 & 0.1 & 0.1 & 0 & 0 \\ 0.1 & 0 & 0.1 & 0.1 & 0 \end{bmatrix}$$

- $\Psi$  as  $M \times M$  as diagonal matrix representing the residual variances in deviation from the preregistration not adjusted for the varying structural relations. This was only updated in

the joint study (study 3) to adjust residual variances of all endogenous factors to accurately reflect the number of regressors

– **Model 1.1, 1.2, 1.3, and 1.4:**

$$\Psi = \begin{bmatrix} 1.0 & 0 & 0 & 0 & 0 \\ 0 & 1.0 & 0 & 0 & 0 \\ 0 & 0 & 1.0 & 0 & 0 \\ 0 & 0 & 0 & 1.0 & 0 \\ 0 & 0 & 0 & 0 & 1.0 \end{bmatrix}$$

- $\Lambda$  as  $P \times M$  matrix representing factor loadings.

– **Model 1.1, 1.3 and 1.4:**

$$\Lambda = \begin{bmatrix} 1.0 & 0 & 0 & 0 & 0 \\ 0.7 & 0 & 0 & 0 & 0 \\ 0.7 & 0 & 0 & 0 & 0 \\ 0 & 1.0 & 0 & 0 & 0 \\ 0 & 0.7 & 0 & 0 & 0 \\ 0 & 0.7 & 0 & 0 & 0 \\ 0 & 0 & 1.0 & 0 & 0 \\ 0 & 0 & 0.7 & 0 & 0 \\ 0 & 0 & 0.7 & 0 & 0 \\ 0 & 0 & 0 & 1.0 & 0 \\ 0 & 0 & 0 & 0.7 & 0 \\ 0 & 0 & 0 & 0.7 & 0 \\ 0 & 0 & 0 & 0 & 1.0 \\ 0 & 0 & 0 & 0 & 0.7 \\ 0 & 0 & 0 & 0 & 0.7 \end{bmatrix}$$

– **Model 1.2:** cross loadings will be set to be 10% lower than the factor loadings:

$\Lambda_{ik,jk} = 0.63 = 0.9 \times 0.7$ . They will be generated by the following elements in  $\Lambda$ : (2, 2), (5, 3), (8, 4), (11, 5), (14, 1).

- $\Theta$  as a  $P \times P$  matrix representing the residual variances and covariances of the indicators.

– Model 1.1, 1.2 and 1.4: The diagonal generated as:

$$\Theta^* = \text{Var}(\eta)\Lambda^T \times \frac{1}{r-1}$$

(where  $r$  is the reliability of the indicators) and 0 on all off-diagonal elements  
– Model 1.3:

- \*  $\Theta^*$  on the diagonal.
- \* Correlated residuals generated between specific indicator pairs: for  $i = (2, 5, 8, 11, 14)$  and  $i' = (3, 6, 9, 12, 15)$ , and for each  $k = 1, \dots, 4$  and  $l = k + 1, \dots, 5$ , the entries  $(i_k, i'_l)$  and  $(i'_l, i_k)$  in  $\Theta$  are set to  $0.6 \times \min \Theta^*$ , ensuring correlated errors among selected indicator pairs without exceeding a 0.6 correlation coefficient.

**3.1.2. Study 2:** For study 2, again, different population models were generated parametrically. Further, the different models of study 2 were used for different simulation settings resulting in two sub-studies 2.1 and 2.2 (see simulation settings).

- $B$  as  $M \times M$  matrix representing latent regression coefficients with varying parameter size defined by two conditions of endogenous factor variance explained by the exogenous factors (low:  $R^2 = 0.1$  or medium:  $R^2 = 0.4$  see below under factor):

– Model 2.1 and 2.2:

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \beta_{\eta_4, \eta_1} & \beta_{\eta_4, \eta_2} & 0 & 0 & 0 \\ 0 & \beta_{\eta_5, \eta_2} & \beta_{\eta_5, \eta_3} & \beta_{\eta_5, \eta_4} & 0 \end{bmatrix}$$

- $\Lambda$  as  $P \times M$  matrix representing factor loadings of indicators on the latent factors.

– **Model 2.1:**

$$\Lambda = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.9 & 0 & 0 & 0 & 0 \\ 0.8 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0.9 & 0 & 0 & 0 \\ 0 & 0.8 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0.9 & 0 & 0 \\ 0 & 0 & 0.8 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0.9 & 0 \\ 0 & 0 & 0 & 0.8 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0.9 \\ 0 & 0 & 0 & 0 & 0.8 \end{bmatrix}$$

– **Model 2.2** with cross-loadings either in the exogenous ( $\lambda_{6,3}$ ), endogenous ( $\lambda_{12,5}$ ) or both parts of the model. Which cross loading was present depended on the misspecification simulation factor. The specific magnitude of the endogenous ( $\lambda_{12,5}$ ) loading depended on  $R^2$  (see under 3.2.2):

$$\Lambda = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.9 & 0 & 0 & 0 & 0 \\ 0.8 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0.9 & 0 & 0 & 0 \\ 0 & 0.8 & \lambda_{6,3} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0.9 & 0 & 0 \\ 0 & 0 & 0.8 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0.9 & 0 \\ 0 & 0 & 0 & 0.8 & \lambda_{12,5} \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0.9 \\ 0 & 0 & 0 & 0 & 0.8 \end{bmatrix}$$

- $\Theta$  as a  $P \times P$  matrix representing the residual variances and covariances of the indicators.

This was computed as the portion of the indicator's total variance that is not explained by the latent factors, after accounting for the strength and reliability of its relationship to these factors (factor loadings), as well as the effects of regressions between the latent factors themselves.

- **Model 2.1:** The diagonal of  $\Theta$  generated as:

$$\Theta^* = \text{Var}(\eta)\Lambda^T \times \frac{1}{r-1}$$

(where  $r$  is the reliability of the indicators) and 0 on all off-diagonal elements

- **Model 2.2:**

- \*  $\Theta^*$  on the diagonal.
- \* Correlated residuals generated between specific indicator pairs in either the endogenous, exogenous or both parts of the model.

Thus depending on the simulation setting either:

- \*  $\Theta_{8,9}, \Theta_{9,8}$  (exogenous part)
- \*  $\Theta_{14,15}$  and  $\Theta_{15,14}$  (endogenous part)
- \*  $\Theta_{8,9}, \Theta_{9,8}, \Theta_{14,15}$  and  $\Theta_{15,14}$  (both parts)

were set  $0.6 \times \min \Theta^*$ , ensuring correlated errors among selected indicator pairs without exceeding a 0.6 correlation coefficient:

### 3.1.3 Study 3:

## 3.2. What will be the different factors of the data-generating mechanism?

### 3.2.1. Study 1 The first study modulated the following factors:

- **Different misspecifications** of the population model where the population model varies between the different models (1.1, 1.2, 1.3, 1.4) as described above, while the analysis model remains specified as model 1.1.
- **Sample sizes** of small ( $N = 100$ ), medium ( $N = 400$ ), or large ( $N = 6400$ )
- Indicator reliability of low (.3), moderate (.5), or high (.7)

### 3.2.2. Study 2: The second study modulated the following factors of the data generating process across both studies:

- **Sample sizes** of small ( $N = 100$ ), medium ( $N = 400$ ), or large ( $N = 6400$ )

- **Variance explained** ( $R^2$ ) of the endogenous factor variance explained by the exogenous factors: low ( $R^2 = 0.1$ ) or medium ( $R^2 = 0.4$ )

In two sub-studies 2.1 and 2.2, the following additional factors will be modulated, respectively:

- **Study 2.1:**
  - **Indicator reliability** of three indicators per factor: *all high* (.8), *all low* (.5), *average low* (.5) varying between .7 to .3 with the highest reliability for the scaling indicator.
  - **Sample sizes** of small ( $N = 100$ ), medium ( $N = 400$ ), or large ( $N = 6400$ )
  - Deviating from the preregistration distribution (normal vs. non-normal) was not considered in the simulation settings to limit the scope of the study.
- **Study 2.2**
  - **Structural misspecifications** by varying the population model as described and thus omitting either
    - \* a residual covariance in the exogenous part of the model
    - \* a factor loading in the endogenous part of the model
    - \* a factor loading in the exogenous part of the model
    - \* a structural path in the endogenous part of the model
    - \* a structural path in the analysis model.
  - **Number of measurement blocks** (how many separate measurement models are fitted in the first step of SAM):
    - \* Separate measurement model per latent variable ( $b = k = 5$ )
    - \* Joint measurement model for all exogenous variables ( $b = 3$ )

### **Explanation of the chosen factor values:**

The aim was a conceptual replication of the results obtained by Rosseel & Loh (2022) in study 1 and Dhaene & Rosseel (2023) study 2. The values were chosen accordingly. No detailed explanation was provided by the authors.

*If there is more than one factor: How will the factor levels be combined and how many simulation conditions will this create?*

- Study 1:
  - 4 population models x 2 model specifications x 3 sample sizes x 3 reliabilities = 72 conditions
    - \* Note: For each of the 4 population models (1.1, 1.2, 1.3, 1.4), two scenarios are considered:



- Correctly specified analysis model matching the population model.
  - Incorrectly specified analysis model, corresponding to model 1.1, with specific misspecifications for models 1.2 (omitted cross-loadings), 1.3 (correlated item residuals), and 1.4 (structural misspecification).
- Study 2.1:
    - 2 population models (2.1, 2.2) x 3 sample sizes x 3 reliabilities x 2 distributions = 36 conditions
  - Study 2.2:
    - 2 population models x 3 sample sizes x 5 misspecifications = 30 conditions

#### 4. Estimands and Targets

##### *Estimands*

Structural model parameters (path coefficients)

#### 5. Methods

Both studies will compare four different estimation methods for SEMs:

- **“Vanilla” SEM:** (structural and measurement model estimated simultaneously) (rationale: the current standard approach in SEM estimation serving as a baseline with maximum likelihood (ML)):
- **SAM:** (separating the estimation of the measurement and structural model to alleviate the potential for propagation of bias from (e.g. misspecified) measurement part to the structural part of the model)
  - **Local SAM** (Uses summary statistics from the measurement model to derive the model-implied mean vector and variance-covariance matrix of latent variables. These statistics are then utilized to estimate the structural parameters. A mapping matrix (M) is used to transform the observed data into the latent variable space. It can be estimated using different methods.)
    - \* **With ML mapping matrix** (Akin to a factor score approach (Bartlett (1937), Bartlett (1938)))

- \* **With ULS mapping matrix** (Uses the Moore-Penrose pseudoinverse, suitable for scenarios with complex or underdetermined systems, where the K matrix is rank-deficient but requires adjustments for structural constraints.)
- **Global SAM** (rationale: Fixing the parameters obtained from the measurement model in the first step, and then using them as constants in the full SEM during the second step. Suitable for models where local SAM is impractical due to higher-order latent variables or rank deficiencies in  $\lambda$ .)

“Vanilla” SEM as well as both steps in the SAM approach will be estimated using Maximum Likelihood (ML) using the `lavaan` (Rosseel (2012)) package in R (R Core Team (2023)).

## 6. Performance Measures

Across both studies the following performance measures will be captured:

- Convergence rates: Proportions of observed data sets that successfully converged for each estimation method.
- Empirical relative biases: Average difference between an estimate and its true value, normalized by the true value, assessed across all path coefficients.
- Empirical coverage levels of 95% confidence intervals (CIs): Proportion of observed data sets where the constructed CIs included the true value.
- Root Mean Squared Errors (RMSE): Calculated as the square root of the average squared difference between an estimate and its true value, evaluated under conditions of model misspecification.

*How will Monte Carlo uncertainty of the estimated performance measures be calculated and reported?*

Monte Carlo uncertainty will be calculated using the `simhelpers` (*Helper Functions for Simulation Studies* (n.d.)) package in R as Monte Carlo Standard Errors (MCSEs):

For the bias:

$$\sqrt{S_T^2/K}$$

For the RMSEA:

$$\sqrt{\frac{K-1}{K} \sum_{j=1}^K (RMSE_{(j)} - RMSE)^2}$$

$$S_T^2 = \frac{1}{K-1} \sum_{k=1}^K (T_k - \bar{T})^2$$

***How many simulation repetitions will be used for each condition?***

- Replicating Rosseel & Loh (2022) study 1 will consist of 5000 repetitions per condition.
- Replicating Dhaene & Rosseel (2023) study 2 will consist of 10000 repetitions per condition.

***How will missing values due to non-convergence or other reasons be handled?***

As mentioned above the convergence rates will be captured and evaluated.

***How do you plan on interpreting the performance measures? (optional)***

The results will be interpreted in comparison to the results and interpretations obtained by Dhaene & Rosseel (2023) and Rosseel & Loh (2022).

***Which statistical software/packages do you plan to use?***

The simulation will be set up in R Core Team (2023) using the `lavaan` package for generating data based on the population models as well as for applying SEM estimation methods (Rosseel (2012)) as well as the `furrr` ((davis\_furrr\_2022?)) package for parallel simulation execution.

***Which computational environment do you plan to use?***

The simulation will run on a high performance computing cluster using *TARDIS Core Package Documentation* — *Tardis* (n.d.) with *A Future for r* (n.d.) and Heidilohr (n.d.)

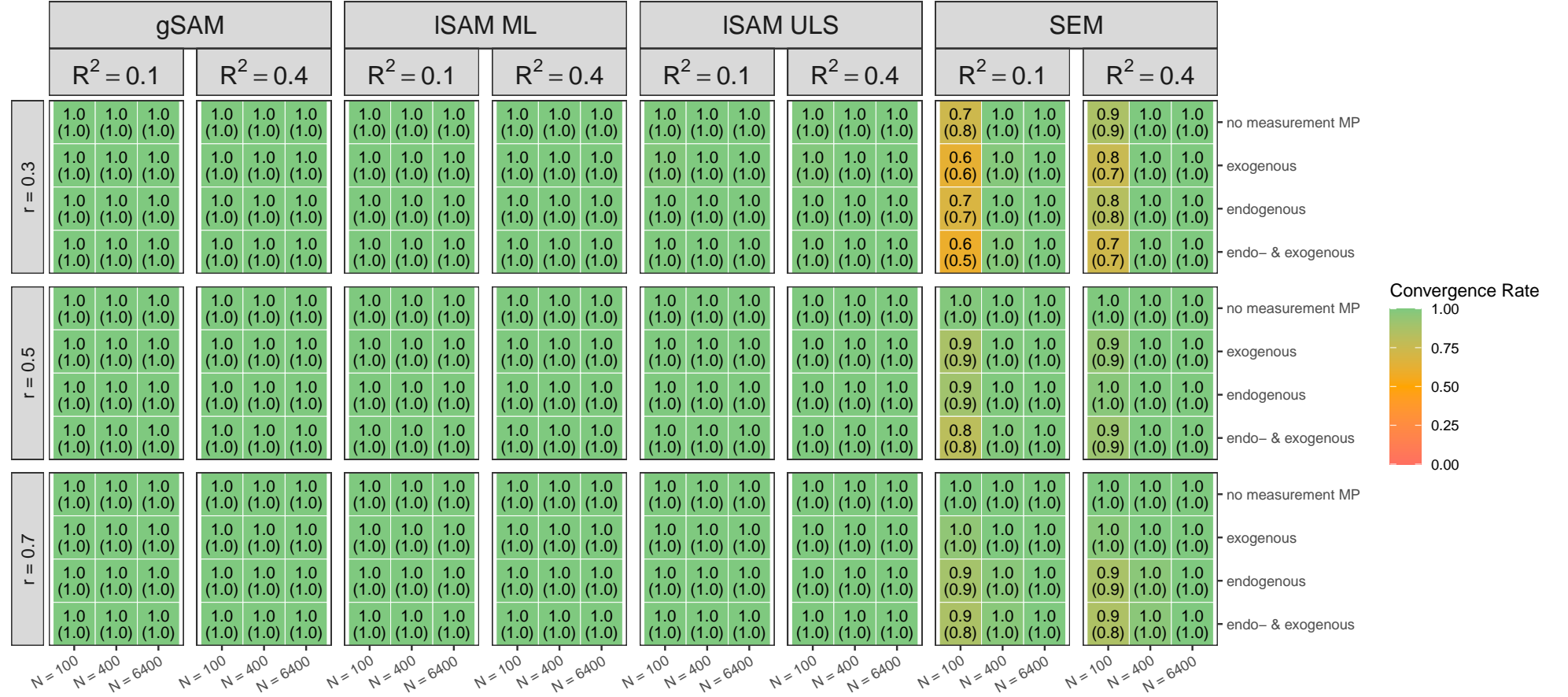
***Which other steps will you undertake to make simulation results reproducible? (optional)***

The code of the simulation will be made available on GitHub (<https://github.com/valentinkm/AdversarialSimulation>)

## **Appendix B: Supplementary Figures**

Figure B1

Convergence Rate and Rate of Proper Solutions in Study 2



Note. Convergence and proper solutions (in parentheses) rates across sample sizes (N), reliability (r), and model misspecification location for global SAM (gSAM), local SAM with Maximum Likelihood (ISAM-ML), Unweighted Least Squares (ISAM-ULS), and SEM.

## Appendix C: Detailed Error and Warning Messages

In the following, all different warning and error messages raised during the studies are listed (see Table C1) and shown how often they occurred under various fitting conditions (see Table C2).

sdlkfjs

**Table C1**

*List of Unique Warnings and Errors*

| ID | Message   |
|----|---|
| 1  | lavaan WARNING: some estimated ov variances are negative  |
| 2  | lavaan WARNING: the optimizer warns that a solution has NOT been found!   |
| 3  | lavaan WARNING: the optimizer (NLMINB) claimed the model converged, but not all elements of the gradient are (near) zero; the optimizer may not have found a local solution use <code>check.gradient = FALSE</code> to skip this check.   |
| 4  | lavaan WARNING: some estimated lv variances are negative  |
| 5  | lavaan WARNING: some estimated ov variances are negative, lavaan WARNING: some estimated lv variances are negative  |
| 6  | number of items to replace is not a multiple of replacement length  |
| 7  | lavaan WARNING: Could not compute standard errors! The information matrix could not be inverted. This may be a symptom that the model is not identified.,<br>lavaan WARNING: some estimated ov variances are negative   |
| 8  | lavaan WARNING: covariance matrix of latent variables is not positive definite; use <code>lavInspect(fit, "cov.lv")</code> to investigate.  |
| 9  | lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be positive definite! The smallest eigenvalue is smaller than or close to zero. This may be a symptom that the model is not identified.,<br>lavaan WARNING: some estimated ov variances are negative |
| 10 | lavaan WARNING: Could not compute standard errors! The information matrix could not be inverted. This may be a symptom that the model is not identified.,<br>lavaan WARNING: some estimated lv variances are negative   |
| 11 | lavaan WARNING: Could not compute standard errors! The information matrix could not be inverted. This may be a symptom that the model is not identified.,<br>lavaan WARNING: some estimated ov variances are negative, lavaan WARNING: some estimated lv variances are negative                           |

**Table C1***List of Unique Warnings and Errors (continued)*

| ID | Message   |
|----|---|
| 12 | lavaan WARNING: some estimated ov variances are negative, lavaan WARNING: covariance matrix of latent variables is not positive definite; use lavInspect(fit, "cov.lv") to investigate.   |
| 13 | lavaan WARNING: Could not compute standard errors! The information matrix could not be inverted. This may be a symptom that the model is not identified.  |
| 14 | lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be positive definite! The smallest eigenvalue is smaller than or close to zero. This may be a symptom that the model is not identified., lavaan WARNING: covariance matrix of latent variables is not positive definite; use lavInspect(fit, "cov.lv") to investigate. |
| 15 | lavaan WARNING: Could not compute standard errors! The information matrix could not be inverted. This may be a symptom that the model is not identified., lavaan WARNING: covariance matrix of latent variables is not positive definite; use lavInspect(fit, "cov.lv") to investigate.   |
| 16 | lavaan WARNING: the covariance matrix of the residuals of the observed variables (theta) is not positive definite; use lavInspect(fit, "theta") to investigate.   |
| 17 | lavaan WARNING: The variance-covariance matrix of the estimated parameters (vcov) does not appear to be positive definite! The smallest eigenvalue is smaller than or close to zero. This may be a symptom that the model is not identified.  |

*Note.* This table lists all unique warnings and errors encountered during the simulation studies.

**Table C2***Summary of Warnings and Errors by Condition with ID for All Studies*

| Study   | Model Type | N   | Reliability | Method | MessageType | Count | Message ID |
|---------|------------|-----|-------------|--------|-------------|-------|------------|
| Study 1 | 1.3        | 100 | 0.3         | SEM    | Warning     | 6860  | 1          |
| Study 1 | 1.2        | 100 | 0.5         | SEM    | Warning     | 3575  | 1          |
| Study 1 | 1.3        | 100 | 0.5         | SEM    | Warning     | 2923  | 1          |
| Study 1 | 1.2        | 100 | 0.7         | SEM    | Warning     | 2903  | 1          |
| Study 1 | 1.2        | 100 | 0.3         | SEM    | Warning     | 2769  | 1          |
| Study 1 | 1.1        | 100 | 0.3         | SEM    | Warning     | 2700  | 1          |
| Study 1 | 1.4        | 100 | 0.3         | SEM    | Warning     | 2577  | 1          |
| Study 1 | 1.2        | 100 | 0.3         | SEM    | Warning     | 2037  | 2          |
| Study 1 | 1.1        | 100 | 0.3         | SEM    | Warning     | 1258  | 2          |

**Table C2***Summary of Warnings and Errors by Condition with ID for All Studies (continued)*

| Study   | Model Type | N   | Reliability | Method       | MessageType | Count | Message ID |
|---------|------------|-----|-------------|--------------|-------------|-------|------------|
| Study 1 | 1.4        | 100 | 0.3         | SEM          | Warning     | 1133  | 2          |
| Study 1 | 1.2        | 100 | 0.3         | SEM          | Warning     | 729   | 3          |
| Study 1 | 1.3        | 100 | 0.3         | SEM          | Warning     | 692   | 3          |
| Study 1 | 1.3        | 100 | 0.7         | SEM          | Warning     | 688   | 1          |
| Study 1 | 1.3        | 400 | 0.3         | SEM          | Warning     | 606   | 1          |
| Study 1 | 1.3        | 100 | 0.3         | SEM          | Warning     | 507   | 2          |
| Study 1 | 1.2        | 400 | 0.3         | SEM          | Warning     | 450   | 1          |
| Study 1 | 1.2        | 400 | 0.5         | SEM          | Warning     | 429   | 1          |
| Study 1 | 1.2        | 100 | 0.3         | SEM          | Warning     | 417   | 4          |
| Study 1 | 1.2        | 400 | 0.7         | SEM          | Warning     | 248   | 1          |
| Study 1 | 1.1        | 100 | 0.3         | SEM          | Warning     | 242   | 3          |
| Study 1 | 1.4        | 100 | 0.3         | SEM          | Warning     | 223   | 3          |
| Study 1 | 1.2        | 100 | 0.5         | SEM          | Warning     | 203   | 2          |
| Study 1 | 1.1        | 100 | 0.5         | SEM          | Warning     | 197   | 1          |
| Study 1 | 1.4        | 100 | 0.5         | SEM          | Warning     | 183   | 1          |
| Study 1 | 1.2        | 100 | 0.3         | ISAM-<br>ULS | Warning     | 150   | 1          |
| Study 1 | 1.2        | 100 | 0.3         | SEM          | Warning     | 146   | 5          |
| Study 1 | 1.4        | 100 | 0.3         | ISAM-<br>ULS | Warning     | 62    | 1          |
| Study 1 | 1.1        | 100 | 0.3         | ISAM-<br>ULS | Warning     | 52    | 1          |
| Study 1 | 1.2        | 100 | 0.3         | gSAM         | Warning     | 50    | 4          |
| Study 1 | 1.4        | 100 | 0.3         | SEM          | Warning     | 50    | 4          |
| Study 1 | 1.2        | 100 | 0.5         | SEM          | Warning     | 42    | 3          |
| Study 1 | 1.2        | 100 | 0.3         | ISAM-<br>ULS | Error       | 38    | 6          |
| Study 1 | 1.1        | 100 | 0.3         | SEM          | Warning     | 29    | 4          |
| Study 1 | 1.1        | 100 | 0.3         | ISAM-<br>ULS | Error       | 25    | 6          |
| Study 1 | 1.4        | 100 | 0.3         | ISAM-<br>ULS | Error       | 24    | 6          |
| Study 1 | 1.2        | 100 | 0.3         | SEM          | Warning     | 23    | 7          |
| Study 1 | 1.2        | 400 | 0.3         | SEM          | Warning     | 15    | 2          |

**Table C2***Summary of Warnings and Errors by Condition with ID for All Studies (continued)*

| Study   | Model Type | N   | Reliability | Method       | MessageType | Count | Message ID |
|---------|------------|-----|-------------|--------------|-------------|-------|------------|
| Study 1 | 1.1        | 100 | 0.3         | SEM          | Warning     | 14    | 7          |
| Study 1 | 1.2        | 100 | 0.3         | gSAM         | Error       | 14    | 6          |
| Study 1 | 1.1        | 100 | 0.3         | SEM          | Warning     | 12    | 5          |
| Study 1 | 1.4        | 100 | 0.3         | SEM          | Warning     | 11    | 5          |
| Study 1 | 1.4        | 100 | 0.3         | SEM          | Warning     | 9     | 7          |
| Study 1 | 1.2        | 100 | 0.7         | SEM          | Warning     | 7     | 2          |
| Study 1 | 1.4        | 100 | 0.7         | SEM          | Warning     | 7     | 1          |
| Study 1 | 1.2        | 100 | 0.3         | SEM          | Warning     | 5     | 8          |
| Study 1 | 1.1        | 100 | 0.5         | SEM          | Warning     | 4     | 2          |
| Study 1 | 1.3        | 100 | 0.5         | SEM          | Warning     | 4     | 3          |
| Study 1 | 1.1        | 100 | 0.3         | SEM          | Warning     | 3     | 9          |
| Study 1 | 1.1        | 100 | 0.7         | SEM          | Warning     | 3     | 1          |
| Study 1 | 1.2        | 100 | 0.3         | SEM          | Warning     | 3     | 10         |
| Study 1 | 1.2        | 100 | 0.3         | SEM          | Warning     | 3     | 11         |
| Study 1 | 1.1        | 100 | 0.3         | gSAM         | Error       | 2     | 6          |
| Study 1 | 1.1        | 100 | 0.3         | gSAM         | Warning     | 2     | 4          |
| Study 1 | 1.1        | 400 | 0.3         | SEM          | Warning     | 2     | 1          |
| Study 1 | 1.2        | 100 | 0.3         | SEM          | Warning     | 2     | 12         |
| Study 1 | 1.2        | 100 | 0.3         | ISAM-<br>ULS | Warning     | 2     | 3          |
| Study 1 | 1.2        | 100 | 0.5         | SEM          | Warning     | 2     | 4          |
| Study 1 | 1.2        | 400 | 0.3         | SEM          | Warning     | 2     | 3          |
| Study 1 | 1.3        | 100 | 0.5         | ISAM-<br>ULS | Error       | 2     | 6          |
| Study 1 | 1.3        | 100 | 0.5         | ISAM-<br>ULS | Warning     | 2     | 1          |
| Study 1 | 1.4        | 100 | 0.3         | SEM          | Warning     | 2     | 9          |
| Study 1 | 1.4        | 100 | 0.3         | gSAM         | Error       | 2     | 6          |
| Study 1 | 1.4        | 100 | 0.3         | gSAM         | Warning     | 2     | 4          |
| Study 1 | 1.4        | 100 | 0.5         | ISAM-<br>ULS | Error       | 2     | 6          |
| Study 1 | 1.1        | 100 | 0.3         | SEM          | Warning     | 1     | 11         |
| Study 1 | 1.1        | 100 | 0.3         | ISAM-<br>ULS | Warning     | 1     | 3          |



**Table C2***Summary of Warnings and Errors by Condition with ID for All Studies (continued)*

| Study   | Model Type | N   | Reliability | Method       | MessageType | Count | Message ID |
|---------|------------|-----|-------------|--------------|-------------|-------|------------|
| Study 1 | 1.1        | 100 | 0.5         | ISAM-<br>ULS | Warning     | 1     | 1          |
| Study 1 | 1.2        | 100 | 0.3         | SEM          | Warning     | 1     | 9          |
| Study 1 | 1.2        | 100 | 0.3         | gSAM         | Warning     | 1     | 8          |
| Study 1 | 1.2        | 100 | 0.5         | ISAM-<br>ULS | Error       | 1     | 6          |
| Study 1 | 1.2        | 100 | 0.5         | ISAM-<br>ULS | Warning     | 1     | 1          |
| Study 1 | 1.3        | 100 | 0.3         | SEM          | Warning     | 1     | 7          |
| Study 1 | 1.3        | 100 | 0.3         | ISAM-<br>ULS | Warning     | 1     | 1          |
| Study 1 | 1.3        | 100 | 0.5         | SEM          | Warning     | 1     | 2          |
| Study 1 | 1.3        | 400 | 0.5         | SEM          | Warning     | 1     | 1          |
| Study 1 | 1.4        | 100 | 0.3         | SEM          | Warning     | 1     | 10         |
| Study 1 | 1.4        | 100 | 0.3         | ISAM-<br>ULS | Warning     | 1     | 3          |
| Study 1 | 1.4        | 100 | 0.3         | ISAM-<br>ULS | Warning     | 1     | 2          |
| Study 1 | 1.4        | 100 | 0.5         | ISAM-<br>ULS | Warning     | 1     | 1          |
| Study 2 | 2.2_both   | 100 | 0.3         | SEM          | Warning     | 5265  | 1          |
| Study 2 | 2.2_exo    | 100 | 0.3         | SEM          | Warning     | 4622  | 1          |
| Study 2 | 2.2_endo   | 100 | 0.3         | SEM          | Warning     | 3615  | 1          |
| Study 2 | 2.2_both   | 100 | 0.7         | SEM          | Warning     | 2904  | 1          |
| Study 2 | 2.2_both   | 100 | 0.5         | SEM          | Warning     | 2743  | 1          |
| Study 2 | 2.1        | 100 | 0.3         | SEM          | Warning     | 2701  | 1          |
| Study 2 | 2.2_endo   | 100 | 0.7         | SEM          | Warning     | 2336  | 1          |
| Study 2 | 2.2_exo    | 100 | 0.5         | SEM          | Warning     | 1814  | 1          |
| Study 2 | 2.2_both   | 100 | 0.3         | SEM          | Warning     | 1625  | 2          |
| Study 2 | 2.2_exo    | 100 | 0.3         | SEM          | Warning     | 1624  | 2          |
| Study 2 | 2.2_endo   | 100 | 0.5         | SEM          | Warning     | 1252  | 1          |
| Study 2 | 2.2_endo   | 100 | 0.3         | SEM          | Warning     | 1211  | 2          |
| Study 2 | 2.1        | 100 | 0.3         | SEM          | Warning     | 1121  | 2          |
| Study 2 | 2.2_exo    | 100 | 0.7         | SEM          | Warning     | 702   | 1          |
| Study 2 | 2.2_both   | 100 | 0.3         | SEM          | Warning     | 675   | 3          |

**Table C2***Summary of Warnings and Errors by Condition with ID for All Studies (continued)*

| Study   | Model Type | N   | Reliability | Method       | MessageType | Count | Message ID |
|---------|------------|-----|-------------|--------------|-------------|-------|------------|
| Study 2 | 2.2_both   | 400 | 0.7         | SEM          | Warning     | 599   | 1          |
| Study 2 | 2.2_endo   | 400 | 0.7         | SEM          | Warning     | 588   | 1          |
| Study 2 | 2.2_exo    | 100 | 0.3         | SEM          | Warning     | 559   | 3          |
| Study 2 | 2.2_endo   | 100 | 0.3         | SEM          | Warning     | 305   | 3          |
| Study 2 | 2.2_exo    | 100 | 0.3         | ISAM-<br>ULS | Warning     | 286   | 1          |
| Study 2 | 2.2_both   | 100 | 0.3         | SEM          | Warning     | 242   | 4          |
| Study 2 | 2.2_both   | 400 | 0.3         | SEM          | Warning     | 239   | 1          |
| Study 2 | 2.1        | 100 | 0.3         | SEM          | Warning     | 220   | 3          |
| Study 2 | 2.2_exo    | 400 | 0.3         | SEM          | Warning     | 195   | 1          |
| Study 2 | 2.2_exo    | 100 | 0.3         | SEM          | Warning     | 175   | 4          |
| Study 2 | 2.1        | 100 | 0.5         | SEM          | Warning     | 165   | 1          |
| Study 2 | 2.2_endo   | 100 | 0.3         | SEM          | Warning     | 157   | 4          |
| Study 2 | 2.1        | 100 | 0.3         | ISAM-<br>ULS | Warning     | 138   | 1          |
| Study 2 | 2.1        | 100 | 0.3         | SEM          | Warning     | 132   | 4          |
| Study 2 | 2.2_both   | 400 | 0.5         | SEM          | Warning     | 130   | 1          |
| Study 2 | 2.2_both   | 100 | 0.3         | ISAM-<br>ULS | Warning     | 128   | 1          |
| Study 2 | 2.2_exo    | 100 | 0.3         | ISAM-<br>ULS | Error       | 121   | 6          |
| Study 2 | 2.2_both   | 100 | 0.5         | SEM          | Warning     | 105   | 2          |
| Study 2 | 2.2_exo    | 100 | 0.5         | SEM          | Warning     | 93    | 2          |
| Study 2 | 2.2_endo   | 100 | 0.3         | ISAM-<br>ULS | Warning     | 78    | 1          |
| Study 2 | 2.2_exo    | 400 | 0.5         | SEM          | Warning     | 77    | 1          |
| Study 2 | 2.2_both   | 100 | 0.3         | SEM          | Warning     | 57    | 5          |
| Study 2 | 2.1        | 100 | 0.3         | ISAM-<br>ULS | Error       | 44    | 6          |
| Study 2 | 2.2_exo    | 100 | 0.3         | SEM          | Warning     | 43    | 5          |
| Study 2 | 2.2_both   | 100 | 0.3         | ISAM-<br>ULS | Error       | 41    | 6          |
| Study 2 | 2.2_endo   | 400 | 0.5         | SEM          | Warning     | 40    | 1          |
| Study 2 | 2.2_both   | 100 | 0.3         | gSAM         | Warning     | 31    | 4          |

**Table C2***Summary of Warnings and Errors by Condition with ID for All Studies (continued)*

| Study   | Model Type | N   | Reliability | Method       | MessageType | Count | Message ID |
|---------|------------|-----|-------------|--------------|-------------|-------|------------|
| Study 2 | 2.2_endo   | 400 | 0.3         | SEM          | Warning     | 29    | 1          |
| Study 2 | 2.2_endo   | 100 | 0.3         | SEM          | Warning     | 26    | 5          |
| Study 2 | 2.2_endo   | 100 | 0.3         | gSAM         | Warning     | 26    | 4          |
| Study 2 | 2.2_exo    | 100 | 0.3         | gSAM         | Warning     | 20    | 4          |
| Study 2 | 2.1        | 100 | 0.3         | gSAM         | Warning     | 18    | 4          |
| Study 2 | 2.2_both   | 100 | 0.5         | SEM          | Warning     | 18    | 3          |
| Study 2 | 2.2_both   | 100 | 0.5         | ISAM-<br>ULS | Warning     | 17    | 1          |
| Study 2 | 2.2_exo    | 100 | 0.5         | ISAM-<br>ULS | Warning     | 16    | 1          |
| Study 2 | 2.1        | 100 | 0.3         | SEM          | Warning     | 15    | 5          |
| Study 2 | 2.1        | 100 | 0.3         | SEM          | Warning     | 14    | 7          |
| Study 2 | 2.2_exo    | 100 | 0.5         | SEM          | Warning     | 14    | 3          |
| Study 2 | 2.2_endo   | 100 | 0.3         | ISAM-<br>ULS | Error       | 14    | 6          |
| Study 2 | 2.2_endo   | 100 | 0.5         | SEM          | Warning     | 13    | 2          |
| Study 2 | 2.2_exo    | 100 | 0.3         | SEM          | Warning     | 10    | 7          |
| Study 2 | 2.2_exo    | 400 | 0.3         | SEM          | Warning     | 10    | 2          |
| Study 2 | 2.2_both   | 100 | 0.3         | SEM          | Warning     | 10    | 7          |
| Study 2 | 2.2_both   | 400 | 0.3         | SEM          | Warning     | 8     | 2          |
| Study 2 | 2.2_exo    | 100 | 0.3         | ISAM-<br>ULS | Warning     | 7     | 3          |
| Study 2 | 2.2_both   | 100 | 0.3         | gSAM         | Error       | 7     | 6          |
| Study 2 | 2.2_exo    | 100 | 0.3         | SEM          | Warning     | 6     | 10         |
| Study 2 | 2.2_exo    | 100 | 0.5         | ISAM-<br>ULS | Error       | 5     | 6          |
| Study 2 | 2.2_both   | 100 | 0.5         | ISAM-<br>ULS | Error       | 5     | 6          |
| Study 2 | 2.2_exo    | 400 | 0.7         | SEM          | Warning     | 4     | 1          |
| Study 2 | 2.2_endo   | 100 | 0.3         | SEM          | Warning     | 4     | 7          |
| Study 2 | 2.1        | 100 | 0.3         | SEM          | Warning     | 3     | 10         |
| Study 2 | 2.1        | 100 | 0.5         | ISAM-<br>ULS | Warning     | 3     | 1          |
| Study 2 | 2.2_exo    | 100 | 0.3         | SEM          | Warning     | 3     | 11         |
| Study 2 | 2.2_both   | 100 | 0.3         | SEM          | Warning     | 3     | 10         |

**Table C2***Summary of Warnings and Errors by Condition with ID for All Studies (continued)*

| Study   | Model Type | N   | Reliability | Method       | MessageType | Count | Message ID |
|---------|------------|-----|-------------|--------------|-------------|-------|------------|
| Study 2 | 2.2_both   | 100 | 0.3         | SEM          | Warning     | 3     | 11         |
| Study 2 | 2.2_both   | 100 | 0.7         | SEM          | Warning     | 3     | 2          |
| Study 2 | 2.1        | 100 | 0.7         | SEM          | Warning     | 2     | 1          |
| Study 2 | 2.2_exo    | 100 | 0.3         | gSAM         | Error       | 2     | 6          |
| Study 2 | 2.2_exo    | 100 | 0.3         | ISAM-<br>ULS | Warning     | 2     | 2          |
| Study 2 | 2.2_exo    | 100 | 0.7         | SEM          | Warning     | 2     | 2          |
| Study 2 | 2.2_exo    | 400 | 0.3         | SEM          | Warning     | 2     | 3          |
| Study 2 | 2.2_endo   | 100 | 0.5         | SEM          | Warning     | 2     | 3          |
| Study 2 | 2.2_both   | 100 | 0.3         | gSAM         | Warning     | 2     | 3          |
| Study 2 | 2.2_both   | 100 | 0.5         | SEM          | Warning     | 2     | 4          |
| Study 2 | 2.1        | 100 | 0.3         | gSAM         | Error       | 1     | 6          |
| Study 2 | 2.1        | 400 | 0.3         | SEM          | Warning     | 1     | 1          |
| Study 2 | 2.2_exo    | 100 | 0.3         | SEM          | Warning     | 1     | 13         |
| Study 2 | 2.2_exo    | 100 | 0.3         | SEM          | Warning     | 1     | 12         |
| Study 2 | 2.2_exo    | 100 | 0.5         | ISAM-<br>ULS | Warning     | 1     | 3          |
| Study 2 | 2.2_exo    | 100 | 0.7         | SEM          | Warning     | 1     | 3          |
| Study 2 | 2.2_exo    | 400 | 0.3         | ISAM-<br>ULS | Error       | 1     | 6          |
| Study 2 | 2.2_endo   | 100 | 0.3         | SEM          | Warning     | 1     | 9          |
| Study 2 | 2.2_endo   | 100 | 0.3         | ISAM-<br>ULS | Warning     | 1     | 3          |
| Study 2 | 2.2_endo   | 100 | 0.5         | ISAM-<br>ULS | Error       | 1     | 6          |
| Study 2 | 2.2_endo   | 100 | 0.5         | ISAM-<br>ULS | Warning     | 1     | 1          |
| Study 2 | 2.2_endo   | 400 | 0.3         | SEM          | Warning     | 1     | 2          |
| Study 2 | 2.2_both   | 100 | 0.3         | ISAM-<br>ULS | Warning     | 1     | 2          |
| Study 2 | 2.2_both   | 100 | 0.3         | ISAM_ML      | Error       | 1     | 6          |
| Study 2 | 2.2_both   | 100 | 0.5         | ISAM-<br>ULS | Warning     | 1     | 2          |
| Study 2 | 2.2_both   | 100 | 0.7         | ISAM-<br>ULS | Warning     | 1     | 1          |

**Table C2***Summary of Warnings and Errors by Condition with ID for All Studies (continued)*

| Study   | Model Type   | N   | Reliability | Method       | MessageType | Count | Message ID |
|---------|--------------|-----|-------------|--------------|-------------|-------|------------|
| Study 2 | 2.2_both     | 400 | 0.3         | SEM          | Warning     | 1     | 3          |
| Study 2 | 2.2_both     | 400 | 0.3         | SEM          | Warning     | 1     | 4          |
| Study 3 | 3.1          | 50  | 0.3         | ISAM-<br>ULS | Warning     | 248   | 1          |
| Study 3 | 3.1          | 50  | 0.3         | gSAM         | Warning     | 209   | 4          |
| Study 3 | 3.1_negative | 50  | 0.3         | ISAM-<br>ULS | Warning     | 208   | 1          |
| Study 3 | 3.1_negative | 50  | 0.3         | gSAM         | Warning     | 164   | 4          |
| Study 3 | 3.2          | 50  | 0.3         | ISAM-<br>ULS | Warning     | 82    | 1          |
| Study 3 | 3.2_negative | 50  | 0.3         | ISAM-<br>ULS | Warning     | 82    | 1          |
| Study 3 | 3.1_negative | 50  | 0.3         | ISAM-<br>ULS | Error       | 79    | 6          |
| Study 3 | 3.1          | 50  | 0.3         | ISAM-<br>ULS | Error       | 72    | 6          |
| Study 3 | 3.2          | 50  | 0.3         | SEM          | Warning     | 71    | 8          |
| Study 3 | 3.2_negative | 50  | 0.3         | SEM          | Warning     | 71    | 8          |
| Study 3 | 3.1          | 100 | 0.3         | ISAM-<br>ULS | Warning     | 62    | 1          |
| Study 3 | 3.1_negative | 50  | 0.3         | SEM          | Warning     | 61    | 8          |
| Study 3 | 3.1          | 50  | 0.3         | SEM          | Warning     | 53    | 8          |
| Study 3 | 3.1_negative | 100 | 0.3         | ISAM-<br>ULS | Warning     | 52    | 1          |
| Study 3 | 3.1          | 50  | 0.5         | ISAM-<br>ULS | Warning     | 51    | 1          |
| Study 3 | 3.1_negative | 50  | 0.3         | gSAM         | Error       | 47    | 6          |
| Study 3 | 3.1_negative | 50  | 0.5         | ISAM-<br>ULS | Warning     | 37    | 1          |
| Study 3 | 3.1          | 50  | 0.3         | gSAM         | Error       | 36    | 6          |
| Study 3 | 3.2          | 50  | 0.3         | SEM          | Warning     | 31    | 14         |
| Study 3 | 3.2          | 100 | 0.3         | ISAM-<br>ULS | Warning     | 31    | 1          |
| Study 3 | 3.2_negative | 50  | 0.3         | SEM          | Warning     | 31    | 14         |

**Table C2***Summary of Warnings and Errors by Condition with ID for All Studies (continued)*

| Study   | Model Type   | N   | Reliability | Method       | MessageType | Count | Message ID |
|---------|--------------|-----|-------------|--------------|-------------|-------|------------|
| Study 3 | 3.2_negative | 100 | 0.3         | ISAM-<br>ULS | Warning     | 31    | 1          |
| Study 3 | 3.1          | 100 | 0.3         | ISAM-<br>ULS | Error       | 28    | 6          |
| Study 3 | 3.2          | 50  | 0.3         | ISAM-<br>ULS | Warning     | 19    | 2          |
| Study 3 | 3.2_negative | 50  | 0.3         | ISAM-<br>ULS | Warning     | 19    | 2          |
| Study 3 | 3.1          | 50  | 0.5         | ISAM-<br>ULS | Error       | 18    | 6          |
| Study 3 | 3.2          | 50  | 0.3         | ISAM-<br>ULS | Warning     | 18    | 3          |
| Study 3 | 3.1_negative | 100 | 0.3         | ISAM-<br>ULS | Error       | 18    | 6          |
| Study 3 | 3.2_negative | 50  | 0.3         | ISAM-<br>ULS | Warning     | 18    | 3          |
| Study 3 | 3.1          | 100 | 0.3         | gSAM         | Warning     | 17    | 4          |
| Study 3 | 3.2          | 50  | 0.3         | ISAM-<br>ULS | Error       | 17    | 6          |
| Study 3 | 3.1_negative | 50  | 0.3         | SEM          | Warning     | 17    | 14         |
| Study 3 | 3.2_negative | 50  | 0.3         | ISAM-<br>ULS | Error       | 17    | 6          |
| Study 3 | 3.1_negative | 50  | 0.5         | ISAM-<br>ULS | Error       | 13    | 6          |
| Study 3 | 3.1_negative | 100 | 0.3         | gSAM         | Warning     | 12    | 4          |
| Study 3 | 3.2          | 100 | 0.3         | ISAM-<br>ULS | Warning     | 11    | 3          |
| Study 3 | 3.2_negative | 100 | 0.3         | ISAM-<br>ULS | Warning     | 11    | 3          |
| Study 3 | 3.1          | 50  | 0.3         | SEM          | Warning     | 10    | 15         |
| Study 3 | 3.1          | 50  | 0.3         | gSAM         | Warning     | 10    | 3          |
| Study 3 | 3.2          | 50  | 0.3         | gSAM         | Warning     | 10    | 4          |
| Study 3 | 3.1_negative | 100 | 0.3         | SEM          | Warning     | 10    | 8          |
| Study 3 | 3.2_negative | 50  | 0.3         | gSAM         | Warning     | 10    | 4          |
| Study 3 | 3.1          | 50  | 0.3         | gSAM         | Warning     | 9     | 8          |

**Table C2***Summary of Warnings and Errors by Condition with ID for All Studies (continued)*

| Study   | Model Type   | N   | Reliability | Method       | MessageType | Count | Message ID |
|---------|--------------|-----|-------------|--------------|-------------|-------|------------|
| Study 3 | 3.2          | 100 | 0.3         | ISAM-<br>ULS | Error       | 9     | 6          |
| Study 3 | 3.2_negative | 100 | 0.3         | ISAM-<br>ULS | Error       | 9     | 6          |
| Study 3 | 3.1          | 50  | 0.5         | gSAM         | Warning     | 8     | 4          |
| Study 3 | 3.1          | 50  | 0.7         | ISAM-<br>ULS | Warning     | 7     | 1          |
| Study 3 | 3.2          | 100 | 0.3         | ISAM-<br>ULS | Warning     | 7     | 2          |
| Study 3 | 3.2_negative | 100 | 0.3         | ISAM-<br>ULS | Warning     | 7     | 2          |
| Study 3 | 3.1          | 50  | 0.3         | SEM          | Warning     | 6     | 14         |
| Study 3 | 3.1          | 100 | 0.3         | SEM          | Warning     | 6     | 8          |
| Study 3 | 3.2          | 50  | 0.3         | SEM          | Warning     | 6     | 15         |
| Study 3 | 3.1_negative | 50  | 0.3         | SEM          | Warning     | 6     | 15         |
| Study 3 | 3.1_negative | 50  | 0.3         | gSAM         | Warning     | 6     | 8          |
| Study 3 | 3.1_negative | 50  | 0.7         | ISAM-<br>ULS | Warning     | 6     | 1          |
| Study 3 | 3.2_negative | 50  | 0.3         | SEM          | Warning     | 6     | 15         |
| Study 3 | 3.1          | 50  | 0.7         | ISAM-<br>ULS | Error       | 5     | 6          |
| Study 3 | 3.2          | 50  | 0.5         | ISAM-<br>ULS | Warning     | 5     | 1          |
| Study 3 | 3.1_negative | 100 | 0.3         | gSAM         | Error       | 5     | 6          |
| Study 3 | 3.2_negative | 50  | 0.5         | ISAM-<br>ULS | Warning     | 5     | 1          |
| Study 3 | 3.1          | 50  | 0.3         | ISAM-<br>ULS | Warning     | 4     | 2          |
| Study 3 | 3.2          | 50  | 0.5         | ISAM-<br>ULS | Error       | 4     | 6          |
| Study 3 | 3.2          | 100 | 0.3         | SEM          | Warning     | 4     | 8          |
| Study 3 | 3.1_negative | 50  | 0.3         | ISAM-<br>ULS | Warning     | 4     | 2          |
| Study 3 | 3.1_negative | 50  | 0.5         | SEM          | Warning     | 4     | 8          |

**Table C2***Summary of Warnings and Errors by Condition with ID for All Studies (continued)*

| Study   | Model Type   | N   | Reliability | Method       | MessageType | Count | Message ID |
|---------|--------------|-----|-------------|--------------|-------------|-------|------------|
| Study 3 | 3.2_negative | 50  | 0.5         | ISAM-<br>ULS | Error       | 4     | 6          |
| Study 3 | 3.2_negative | 100 | 0.3         | SEM          | Warning     | 4     | 8          |
| Study 3 | 3.1          | 50  | 0.3         | ISAM-<br>ULS | Warning     | 3     | 16         |
| Study 3 | 3.1_negative | 50  | 0.3         | gSAM         | Warning     | 3     | 3          |
| Study 3 | 3.1_negative | 100 | 0.5         | ISAM-<br>ULS | Warning     | 3     | 1          |
| Study 3 | 3.1          | 50  | 0.3         | ISAM-<br>ULS | Warning     | 2     | 3          |
| Study 3 | 3.1          | 50  | 0.5         | SEM          | Warning     | 2     | 8          |
| Study 3 | 3.1          | 100 | 0.3         | gSAM         | Error       | 2     | 6          |
| Study 3 | 3.1          | 100 | 0.5         | ISAM-<br>ULS | Error       | 2     | 6          |
| Study 3 | 3.1          | 250 | 0.3         | ISAM-<br>ULS | Warning     | 2     | 1          |
| Study 3 | 3.2          | 100 | 0.3         | SEM          | Warning     | 2     | 14         |
| Study 3 | 3.1_negative | 50  | 0.3         | ISAM-<br>ULS | Warning     | 2     | 3          |
| Study 3 | 3.1_negative | 50  | 0.3         | ISAM-<br>ULS | Warning     | 2     | 16         |
| Study 3 | 3.1_negative | 50  | 0.5         | gSAM         | Warning     | 2     | 4          |
| Study 3 | 3.1_negative | 100 | 0.3         | SEM          | Warning     | 2     | 14         |
| Study 3 | 3.1_negative | 100 | 0.3         | ISAM-<br>ULS | Warning     | 2     | 3          |
| Study 3 | 3.2_negative | 100 | 0.3         | SEM          | Warning     | 2     | 14         |
| Study 3 | 3.1          | 50  | 0.5         | gSAM         | Error       | 1     | 6          |
| Study 3 | 3.1          | 50  | 0.5         | ISAM-<br>ULS | Warning     | 1     | 3          |
| Study 3 | 3.1          | 50  | 0.5         | ISAM-<br>ULS | Warning     | 1     | 2          |
| Study 3 | 3.1          | 100 | 0.3         | SEM          | Warning     | 1     | 14         |
| Study 3 | 3.1          | 100 | 0.3         | gSAM         | Warning     | 1     | 8          |
| Study 3 | 3.1          | 100 | 0.3         | ISAM-<br>ULS | Warning     | 1     | 3          |



**Table C2***Summary of Warnings and Errors by Condition with ID for All Studies (continued)*

| Study   | Model Type   | N   | Reliability | Method       | MessageType | Count | Message ID |
|---------|--------------|-----|-------------|--------------|-------------|-------|------------|
| Study 3 | 3.1          | 100 | 0.5         | ISAM-<br>ULS | Warning     | 1     | 1          |
| Study 3 | 3.1          | 250 | 0.3         | ISAM-<br>ULS | Error       | 1     | 6          |
| Study 3 | 3.2          | 50  | 0.3         | gSAM         | Error       | 1     | 6          |
| Study 3 | 3.2          | 50  | 0.3         | ISAM-<br>ULS | Warning     | 1     | 16         |
| Study 3 | 3.2          | 50  | 0.5         | gSAM         | Warning     | 1     | 4          |
| Study 3 | 3.2          | 50  | 0.5         | ISAM-<br>ULS | Warning     | 1     | 3          |
| Study 3 | 3.2          | 50  | 0.7         | ISAM-<br>ULS | Warning     | 1     | 1          |
| Study 3 | 3.2          | 250 | 0.3         | ISAM-<br>ULS | Warning     | 1     | 3          |
| Study 3 | 3.2          | 250 | 0.3         | ISAM-<br>ULS | Warning     | 1     | 2          |
| Study 3 | 3.2          | 250 | 0.3         | ISAM-<br>ULS | Warning     | 1     | 1          |
| Study 3 | 3.2          | 400 | 0.3         | ISAM-<br>ULS | Warning     | 1     | 1          |
| Study 3 | 3.1_negative | 50  | 0.5         | gSAM         | Error       | 1     | 6          |
| Study 3 | 3.1_negative | 50  | 0.5         | ISAM-<br>ULS | Warning     | 1     | 16         |
| Study 3 | 3.1_negative | 50  | 0.7         | ISAM-<br>ULS | Error       | 1     | 6          |
| Study 3 | 3.1_negative | 50  | 0.7         | ISAM-<br>ULS | Warning     | 1     | 3          |
| Study 3 | 3.1_negative | 100 | 0.3         | SEM          | Warning     | 1     | 13         |
| Study 3 | 3.1_negative | 100 | 0.3         | SEM          | Warning     | 1     | 17         |
| Study 3 | 3.1_negative | 100 | 0.3         | gSAM         | Warning     | 1     | 8          |
| Study 3 | 3.1_negative | 250 | 0.3         | ISAM-<br>ULS | Error       | 1     | 6          |
| Study 3 | 3.1_negative | 250 | 0.3         | ISAM-<br>ULS | Warning     | 1     | 1          |
| Study 3 | 3.2_negative | 50  | 0.3         | gSAM         | Error       | 1     | 6          |

**Table C2***Summary of Warnings and Errors by Condition with ID for All Studies (continued)*

| Study   | Model Type   | N   | Reliability | Method       | MessageType | Count | Message ID |
|---------|--------------|-----|-------------|--------------|-------------|-------|------------|
| Study 3 | 3.2_negative | 50  | 0.3         | lSAM-<br>ULS | Warning     | 1     | 16         |
| Study 3 | 3.2_negative | 50  | 0.5         | gSAM         | Warning     | 1     | 4          |
| Study 3 | 3.2_negative | 50  | 0.5         | lSAM-<br>ULS | Warning     | 1     | 3          |
| Study 3 | 3.2_negative | 50  | 0.7         | lSAM-<br>ULS | Warning     | 1     | 1          |
| Study 3 | 3.2_negative | 250 | 0.3         | lSAM-<br>ULS | Warning     | 1     | 3          |
| Study 3 | 3.2_negative | 250 | 0.3         | lSAM-<br>ULS | Warning     | 1     | 2          |
| Study 3 | 3.2_negative | 250 | 0.3         | lSAM-<br>ULS | Warning     | 1     | 1          |
| Study 3 | 3.2_negative | 400 | 0.3         | lSAM-<br>ULS | Warning     | 1     | 1          |

*Note.* This table summarizes the count of warnings and errors for each condition in all three simulation studies with the respective ID number corresponding to Table 1.