



**Universität  
Zürich<sup>UZH</sup>**

Psychologisches Institut

---

# Studie 1

**Forschungsdesign**

Alain Stocker

2021

# INHALT

---

1	Studie 1.....	3
1.1	Hypothesen .....	3
1.2	Methoden.....	6
1.2.1	Datengenerierung.....	6
1.2.2	Versuchsordnung.....	9
1.2.3	Verwendete Modelle.....	10
2	Literaturverzeichnis .....	12

# 1 STUDIE 1

---

Die Forschungsfrage soll mit Hilfe von Monte-Carlo-Simulationen beantwortet werden. In diesem Kapitel wird die erste von insgesamt zwei Simulationsstudien behandelt. Ziel des Kapitels soll es sein, nicht nur einen kurzen Bericht über Studie 1 abzuliefern, sondern gemäss den Empfehlungen von Ripley (1987, S. 4) bezüglich Simulationsstudien auch eine allfällige Replikation und Überprüfung der Ergebnisse basierend auf den hier berichteten Beschreibungen zu ermöglichen.

## 1.1 HYPOTHESEN

Die Fragestellung soll mittels Hypothesen untersucht werden, was sich auf die von Boomsma (2013) empfohlene Vorgehensweise für Monte-Carlo-Studien mit SEMs stützt. Für Studie 1 wurden vier Hypothesen formuliert. Dabei stützen sich die ersten beiden Hypothesen auf Autoren wie Cochran (1968), Fuller (1987, S. 3), Ree und Carretta (2006), Liu und Salvendy (2009), Hutcheon et al. (2010) oder Carroll (2014), gemäss denen ein einfaches lineares Regressionsmodell (engl. *linear model*, LM) bei einer abnehmender Reliabilität der unabhängigen Variable dazu tendiert, den Regressionskoeffizienten zunehmend zu unterschätzen.<sup>1</sup> Es kann angenommen werden, dass unter Idealbedingungen – also bei perfekten Reliabilitäten der unabhängigen und abhängigen Variablen sowie einem perfekten Zusammenhang zwischen den gemessenen Variablen und den latenten Konstrukten – beide Modelle eine exakte Schätzung des Regressionskoeffizienten ermöglichen. Die in Formel (1) dargestellte Abweichung zwischen dem geschätzten ( $\hat{\beta}_1$ ) und dem wahren Regressionskoeffizienten ( $\beta_1$ ) sollte somit für beide Modelle im Mittel null betragen.

---

<sup>1</sup> Während im Theorieteil dieser Arbeit aus didaktischen Gründen die in ihrem jeweiligen Kontext korrekte Nomenklatur zur Bezeichnung der Variablen verwendet wird – sprich Prädiktor und Kriterium bei Regressionsmodellen sowie exogene und endogene Variablen bei SEMs – sei darauf hingewiesen, dass in den übrigen Kapiteln einfachheitshalber lediglich von der unabhängigen respektive abhängigen Variable gesprochen wird.

*Hypothese 1:* Wenn die Reliabilitäten der abhängigen sowie unabhängigen Variablen 1.0 betragen und deren Faktorladungen auf 1 fixiert werden, dann beträgt das Bias von  $\hat{\beta}_1$  beim LM sowie beim SEM im Mittel null.

$$\text{Bias } \hat{\beta}_1 = \hat{\beta}_1 - \beta_1 \quad (1)$$

Da das SEM die Messfehler, welche die geringe Reliabilität der unabhängigen Variable verursachen, als Teil des Modelles schätzt und von den anderen zu schätzenden Parametern zu trennen weiss (Nachtigall et al., 2003), kann darüber hinaus davon ausgegangen werden, dass diese Unterschätzung des Regressionskoeffizienten beim SEM nicht vorkommt. Bei einer geringen Reliabilität der unabhängigen Variable sollte das SEM also kein Bias von  $\hat{\beta}_1$  aufweisen, während das Bias beim LM aufgrund Unterschätzung des Regressionskoeffizienten negativ ausfallen sollte.

*Hypothese 2:* Mit abnehmender Reliabilität der unabhängigen Variablen wird das Bias von  $\hat{\beta}_1$  beim LM im Mittel in zunehmendem Ausmass negativ ausfallen, während das SEM im Mittel ein Bias von null aufweist.

Neben der Schätzung des Regressionskoeffizienten sowie deren Abweichung vom wahren Wert sollen auch der Standardfehler des Regressionskoeffizienten sowie dessen Abweichung von der empirisch ermittelten Streuung untersucht werden.

Da bei einer abnehmenden Reliabilität der abhängigen Variable die vertikale Streuung der gemessenen Daten zunimmt, wird die Schätzung des Regressionskoeffizienten ungenauer (Hutcheon et al., 2010). In solchen Fällen sollte der angegebene Standardfehler für  $\hat{\beta}_1$  beim LM zunehmen. Das SEM hingegen sollte sich auch bei geringer Reliabilität der abhängigen Variable nicht in seiner Schätzung des Regressionskoeffizienten beeinflussen lassen, da auch hier die verursachenden Messfehler bereits Teil des Modells sind.

*Hypothese 3a:* Mit abnehmender Reliabilität der abhängigen Variablen ist der Standardfehler von  $\hat{\beta}_1$  beim LM grösser als beim SEM.

Wie im Theorieteil erörtert, gibt der Standardfehler des Regressionskoeffizienten die Genauigkeit an, mit der der Regressionskoeffizient geschätzt wurde. Die Genauigkeit des Standardfehlers wiederum kann über das relative Bias des Standardfehlers von  $\hat{\beta}_1$  quantifiziert werden. Dieses errechnet sich, wie in Formel (2) gezeigt, aus der Differenz der Standardabweichung von  $\hat{\beta}_1$  vom Erwartungswert des Standardfehlers von  $\hat{\beta}_1$  dividiert durch die Standardabweichung von  $\hat{\beta}_1$  (Hoogland & Boomsma, 1998; Beaujean, 2014; McNeish, 2014). Das relative Bias des Standardfehlers von  $\hat{\beta}_1$  gibt an, wie stark der Mittelwert der angegebenen Standardfehler von der empirisch ermittelten Standardabweichung über multiple Stichproben abweicht. Kurzum: Der Wert zeigt auf, um wie viel sich die vom LM oder SEM angegebenen Standardfehler von der tatsächlichen Standardabweichung unterscheiden. Hoogland und Boomsma (1998) geben an, dass eine Abweichung von betragsmässig 10% oder weniger akzeptabel sei.

*Hypothese 3b:* Mit abnehmender Reliabilität der abhängigen Variablen ist das relative Bias des Standardfehlers von  $\hat{\beta}_1$  (Hoogland & Boomsma, 1998; Beaujean, 2014; McNeish, 2014) beim linearen Regressionsmodell stärker als beim Strukturgleichungsmodell.

$$\text{Relatives Bias von } SE(\hat{\beta}_1) = \frac{E(SE[\hat{\beta}_1]) - SD(\hat{\beta}_1)}{SD(\hat{\beta}_1)} \quad (2)$$

Da bei einer geringeren Reliabilität der unabhängigen Variable das LM den Regressionskoeffizienten unterschätzt (Cochran, 1968), werden tatsächlich vorhandene Effekte weniger häufig entdeckt. Die Power sinkt mit abnehmender Reliabilität der unabhängigen Variable für das LM, während sie für das SEM im Vergleich höher ausfallen sollte. Die Power ergibt sich gemäss Beaujean (2014) in Monte Carlo Studien aus dem Verhältnis zwischen der Anzahl signifikanter Schätzungen und der Anzahl Replikationen (siehe Formel (3)).

*Hypothese 4:* Die Power (Beaujean, 2014) wird mit geringerer Reliabilität der unabhängigen Variablen beim LM niedriger ausfallen als beim SEM.

$$\text{Power} = \text{Trefferrate} = \frac{\text{Anzahl signifikanter Schätzungen}}{\text{Anzahl Replikationen}} \quad (3)$$

Zusätzlich zu den vier genannten Hypothesen sollen in explorativen Analysen weitere Einflussfaktoren untersucht werden, auf die im Folgenden kurz eingegangen werden soll.

Da mit zunehmender Stichprobengrösse die Standardfehler immer kleiner ausfallen und die statistische Power immer grösser wird (Sedlmeier & Renkewitz, 2013, S. 370–371), könnten für das LM die Effekte einer geringen Reliabilität der abhängigen Variable durch eine grosse Stichprobe ausgeglichen werden. Für SEMs existieren in der Literatur keine eindeutigen Angaben zur Mindestgrösse der Stichprobe. Allerdings wird darauf hingewiesen, dass SEMs grosse Stichproben benötigen und bei kleineren Stichprobengrössen mit Ungenauigkeiten sowie technischen Problemen zu kämpfen haben (Bentler & Chou, 1987; Kline, 2016, S. 14). Aus diesen Gründen soll auch die Auswirkung der Stichprobengrösse auf die Schätzung des Regressionskoeffizienten sowie dessen Standardfehler und die Power zumindest explorativ für das LM sowie das SEM untersucht werden.

Autoren wie Bentler und Chou (1987), Kline (2016, S. 201) oder Bagozzi und Yi (2012) warnen bei einer zu geringen Anzahl Indikatoren pro latenter Variable – insbesondere bei einer kleinen Stichprobengrösse – vor technischen Problemen bei der Analyse sowie Unteridentifikation des Modells und empfehlen bei Modellen mit mehr als einer latenten Variable eine Mindestanzahl von 2, besser aber 3 bis 5 Indikatoren pro latenter Variable. Aus diesem Grund sollen auch der Einfluss der Anzahl Indikatoren auf die geschätzten Regressionskoeffizienten, dessen Standardfehler sowie die Power in explorativer Art untersucht werden.

## 1.2 METHODEN

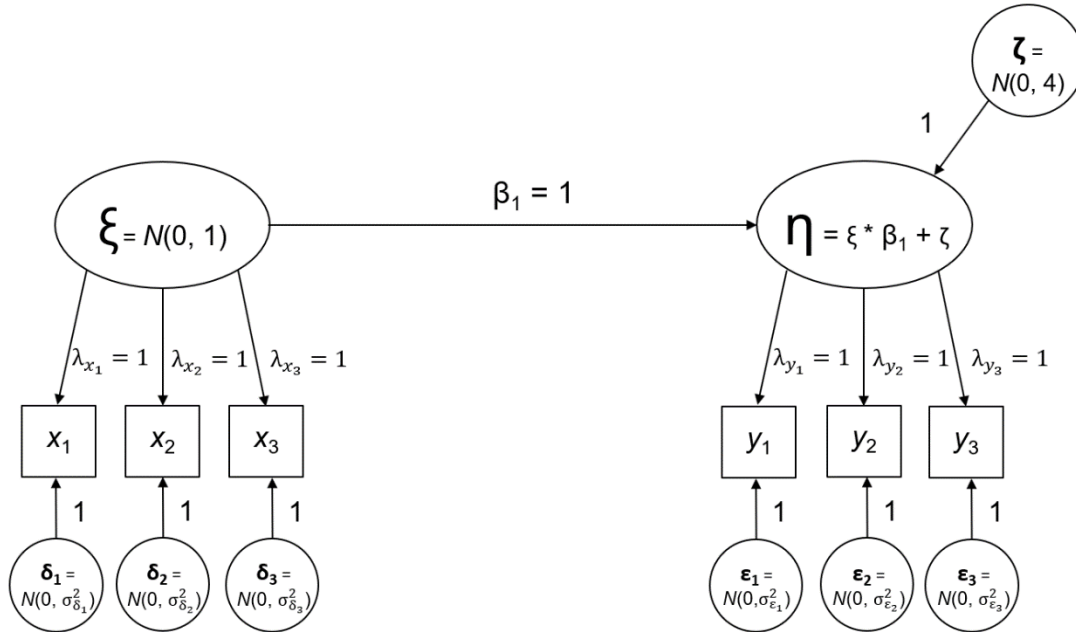
### 1.2.1 Datengenerierung

Die Datengenerierung orientierte sich am Schema in Abbildung 1. Es wurden zwei latente Variablen generiert, wobei die exogene Variable  $\xi$  aus einer Standardnormalverteilung gezogen wurde, während die endogene Variable  $\eta$  mit einem Koeffizienten von  $\beta_1 = 1$  auf  $\xi$  regressierte. Die endogene, latente Variable  $\eta$  setzte sich für Per-

son  $n$  wie in Formel (4) dargestellt zusammen, wobei  $\zeta_n$  die mit  $\zeta \sim N(0, 4)$  normalverteilte Störgrösse darstellte. Für  $\zeta$  wurde eine Varianz von  $\sigma_\zeta^2 = 4$  verwendet, um den prozentualen Anteil der aufgeklärten Varianz von  $\eta$  auf 20% zu fixieren.

### Abbildung 1

*Schematische Darstellung der Datengenerierung*



*Anmerkung.* Schema der Datengenerierung für die Bedingungen mit jeweils 3 Indikatoren pro latenter Variable.

Es wurden für  $\xi$  die manifesten Variablen  $x_m$  und für  $\eta$  die manifesten Variablen  $y_m$  als Indikatoren generiert, wobei  $m = \{1, \dots, M\}$  entsprach und je nach Versuchsbedingung  $M = 3$  oder  $M = 5$  galt. Somit betrug die Anzahl Indikatoren je nach Versuchsbedingung entweder 3 oder 5 pro latenter Variable. Die Faktorladungen der Indikatoren  $\lambda_{x_m}$  und  $\lambda_{y_m}$  wurden auf 1 fixiert, wodurch ein sogenanntes paralleles Messmodell aus der Klassischen Testtheorie (Graham, 2006; Bühner, 2011, S. 149–150) hergestellt wird. Gemäss diesem Modell messen die Indikatoren  $x_m$  die latente Eigenschaft  $\xi$  alle genau gleich gut und auf derselben Skala (Lord & Novick, 2008, S. 48). Dasselbe gilt auch für die Indikatoren  $y_m$  und die latente Variable  $\eta$ .

Damit ergaben sich folgende Modellgleichungen für die Datengenerierung endogener Variablen, wobei  $m$  den Index der manifesten Variable und  $n$  die  $n$ -te Person aus der Stichprobe repräsentieren:

Latent

$$\eta_n = \xi_n * \beta_1 + \zeta_n$$

Manifest

$$\begin{aligned} x_{mn} &= \xi_n * \lambda_{x_m} + \delta_{mn} \\ y_{mn} &= \eta_n * \lambda_{y_m} + \varepsilon_{mn} \end{aligned} \tag{4}$$

Die Reliabilität der manifesten Variablen wurde über deren Fehlervarianz operationalisiert. Für die  $m$ -te manifeste Variable einer jeden Person  $n$  wurde jeweils aus einer Normalverteilung ein Messfehler  $\delta_{mn} \sim N(0, \sigma_{\delta_m}^2)$  für  $x_{mn}$  respektive  $\varepsilon_{mn} \sim N(0, \sigma_{\varepsilon_m}^2)$  für  $y_{mn}$  generiert. Die Fehlervarianzen aller Indikatoren  $x_m$  wurden innerhalb einer Versuchsbedingung mit  $\sigma_{\delta_m}^2 = \sigma_{\delta}^2$  gleichgesetzt, dasselbe galt auch für die Fehlervarianzen von  $y_m$  mit  $\sigma_{\varepsilon_m}^2 = \sigma_{\varepsilon}^2$ . Dabei wurden zur Manipulation der Reliabilität die Fehlervarianzen  $\sigma_{\delta}^2$  sowie  $\sigma_{\varepsilon}^2$  so gewählt, dass McDonalds Omega ( $\omega$ ) nach McDonald (1999) als Reliabilitätswert der aufsummierten Indikatoren  $x_1$  bis  $x_M$  beziehungsweise  $y_1$  bis  $y_M$  mit  $M = 3$  oder  $M = 5$  unter Annahme der Eindimensionalität die Werte aus Tabelle 1 annahm. Der Reliabilitätswert  $\omega$  errechnet sich dabei gemäss McDonald (1999) wie folgt:

$$\omega = \frac{(\sum_{m=1}^M \lambda_m)^2}{(\sum_{m=1}^M \lambda_m)^2 + \sum_{m=1}^M \sigma_{e_m}^2} \tag{5}$$

Hierbei stellen  $\lambda_m$  die Faktorladung und  $\sigma_{e_m}^2$  die Fehlervarianz des  $m$ -ten Indikators dar. Die Summe wird jeweils von  $m = 1$  bis zur Anzahl Indikatoren  $M$  gebildet. Somit entstehen zwei Reliabilitätswerte, namentlich  $\omega_x$  für die aufsummierten  $x$ -Variablen mit  $\sigma_{e_m}^2 = \sigma_{\delta_m}^2$  und  $\omega_y$  für die Summe der  $y$ -Variablen mit  $\sigma_{e_m}^2 = \sigma_{\varepsilon_m}^2$ .

Da gemäss Reise (2012) der Reliabilitätswert  $\omega$  analog zu Cronbachs Alpha ( $\alpha$ ) nach Cronbach (1951) ist, wurden die Werte von  $\omega_x$  respektive  $\omega_y$  so gewählt, dass in der Literatur empfohlene Mindestwerte von Cronbachs Alpha ( $\alpha$ ) für eine akzeptable interne Konsistenz von Tests mit Summenscores abgebildet werden konnten (Nunnally & Bernstein, 1994, S. 264–265). Zusätzlich wurden die Reliabilitätswerte von 0.5 sowie 1.0 als Ober- beziehungsweise Untergrenze ausgewählt, um eine Untersuchung des Verhaltens der Modelle unter Extrembedingungen zu ermöglichen. Die für  $\omega_x$  respektive  $\omega_y$  gewählten Werte decken sich somit auch mit dem von Reise et al. (2013) genannten Minimum für  $\omega$  von 0.5 sowie dem empfohlenen Mindestwert von 0.75.



Tabelle 1 zeigt die Operationalisierung der Reliabilitäten  $\omega_x$  und  $\omega_y$  durch die Fehlervarianzen  $\sigma_{\delta_m}^2$  und  $\sigma_{\varepsilon_m}^2$  in Abhängigkeit zur Anzahl Indikatoren pro latenter Variable ( $M$ ). Die Fehlervarianzen  $\sigma_{\delta_m}^2$  sowie  $\sigma_{\varepsilon_m}^2$  nahmen je nach Anzahl Indikatoren  $M$  bei gleichbleibender Reliabilität  $\omega_x$  respektive  $\omega_y$  andere Werte an. So entsprach beispielsweise eine Reliabilität von  $\omega_x = 0.8$  einer Fehlervarianz von  $\sigma_{\delta_m}^2 = 0.75$  bei einer Versuchsbedingung mit  $M = 3$  Indikatoren pro latenter Variable, während bei einer Bedingung mit  $M = 5$  die Fehlervarianz  $\sigma_{\delta_m}^2 = 1.25$  betrug.

**Tabelle 1**

*Fehlervarianzen den entsprechenden Werten von McDonalds  $\omega$  zugeordnet*

Fehlervarianz	M	$\omega_x$ bzw. $\omega_y$					
		0.5	0.7	0.8	0.9	0.95	1.0
$\sigma_{\delta_m}^2$ bzw. $\sigma_{\varepsilon_m}^2$	3	3	1.29	0.75	0.33	0.16	0
	5	5	2.14	1.25	0.56	0.26	0

*Anmerkung.* Fehlervarianzen auf zwei Nachkommastellen gerundet.

## 1.2.2 Versuchsanordnung

Es wurden die beiden Ausprägungen für die Anzahl Indikatoren pro Faktor, die Stichprobengrößen in Anlehnung an Brunner und Austin (2009), die Reliabilitäten  $\omega_x$  respektive  $\omega_y$  sowie die beiden Methoden zur Parameterschätzung genutzt und miteinander kombiniert, was in total  $2 \times 2 \times 5 \times 6 \times 6 = 720$  Versuchsbedingungen resultierte. Zusammengefasst wurden also folgende 5 Faktoren mit den unten aufgelisteten Stufen untersucht:

- *Methode:* LM, SEM
- *Anzahl Indikatoren pro latenter Variable:* 3, 5
- *Stichprobengrösse:* 50, 100, 250, 500, 1000
- *Reliabilität  $\omega_x$ :* 0.5, 0.7, 0.8, 0.9, 0.95, 1.0
- *Reliabilität  $\omega_y$ :* 0.5, 0.7, 0.8, 0.9, 0.95, 1.0

In jeder der 720 Versuchsbedingungen wurden 1'000 Datensätze generiert. Für jeden einzelnen Datensatz wurden sowohl ein LM als auch ein SEM gefittet und die abhängigen Variablen der Studie – z.B. das Bias von  $\hat{\beta}_1$  – berechnet respektive die dafür

benötigten Daten – z.B. der Standardfehler von  $\hat{\beta}_1$  zur Berechnung des relativen Bias – extrahiert. Bei der Methode handelt es sich somit um einen Within-Faktor.

### 1.2.3 Verwendete Modelle

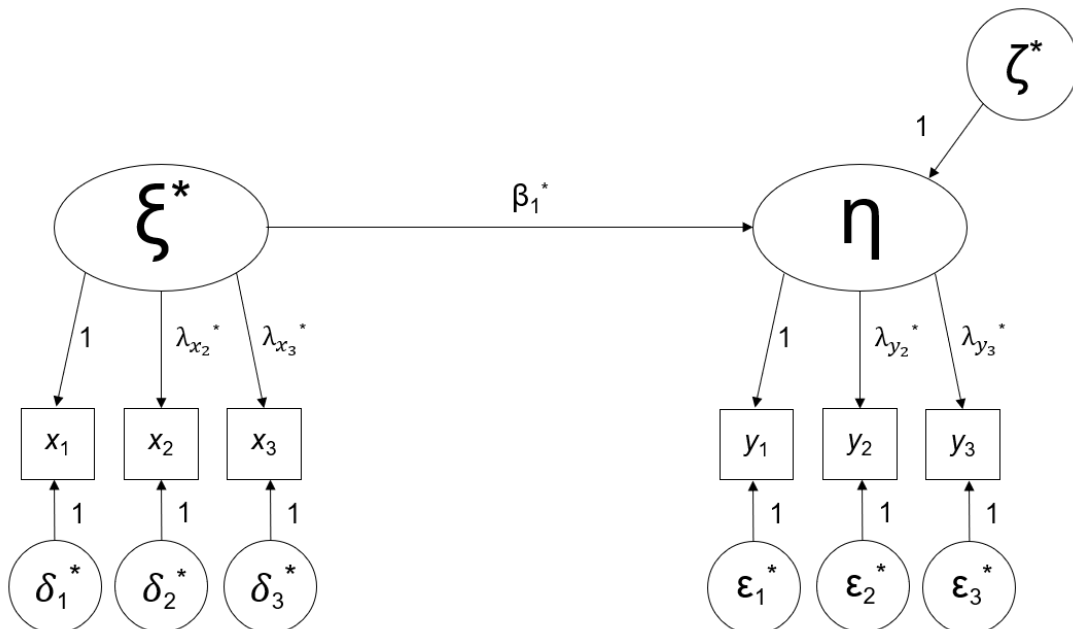
Bei den an die generierten Daten anzupassenden Modelle handelte es sich zum einen um ein lineares Regressionsmodell und zum anderen um ein lineares Strukturgleichungsmodell. Das lineare Regressionsmodell verfügte über folgende Modellgleichung:

$$y_{sum_n} = \beta_0 + \beta_1 * x_{sum_n} + e_n \quad (6)$$

Dabei gilt für Person  $n$   $x_{sum_n} = \sum_{m=1}^M x_{m_n}$  und  $y_{sum_n} = \sum_{m=1}^M y_{m_n}$  mit  $m = \{1, \dots, M\}$  als Index für den  $m$ -ten Indikator. Der Term  $e_n$  steht für das Residuum der  $n$ -ten Person, welches einer Schätzung der Störgrösse  $\zeta$  aus Abbildung 1 entspricht. Der Achsenabschnitt  $\beta_0$  fällt weg, da dieser null betrug.

#### Abbildung 2

*Pfaddiagramm für das SEM*



*Anmerkung:* Verwendetes Pfaddiagramm für Versuchsbedingungen mit je 3 Indikatoren pro latenter Variable. Asteriske kennzeichnen zu schätzende Modellparameter.

Für das SEM wurde das R-Paket *lavaan* genutzt, welches nach Eingabe eines Pfaddiagramms die Berechnung der Parameter ermöglicht. Für die Bedingungen mit 3

Indikatoren pro latenter Variable wurde das Pfaddiagramm in Abbildung 2 verwendet und für die Versuchsbedingungen mit 5 Indikatoren pro Faktor je 2 weitere manifeste Variablen inklusive deren Fehler sowie Faktorladungen hinzugefügt. Zu schätzende Parameter sind jeweils mit einem Asterisk gekennzeichnet. Das Pfaddiagramm für das SEM entspricht dem Modell der Datengenerierung aus Abbildung 1, um einer möglichen Beeinflussung der Schätzungen durch eine schlechte Passung zwischen Modell und Daten vorbeugend entgegenzuwirken. Schätzungen durch das SEM sollten gemäss Schinka et al. (2003, S. 610) nämlich nur dann Ernst genommen werden, wenn die Passung des Modells sehr gut ist.

Die Anzahl zu schätzender Parameter  $q$  betrug im Pfaddiagramm für Versuchsbedingungen mit einer Anzahl Indikatoren  $M$  pro latenter Variable  $q = 13$  bei  $M = 3$  und  $q = 17$  bei  $M = 5$ . Die Anzahl beobachteter Variablen  $p$  betrug  $p = 6$  für  $M = 3$  respektive  $p = 10$  bei  $M = 5$ . Damit liess sich die Anzahl Freiheitsgrade des Modells nach Formel berechnen. Beide Modelle waren mit  $df = 8$  für  $M = 3$  respektive  $df = 13$  für  $M = 5$  überidentifiziert.

## 2 LITERATURVERZEICHNIS

---

- Bagozzi, R. P. & Yi, Y. (2012). Specification, evaluation, and interpretation of structural equation models. *Journal of the Academy of Marketing Science*, 40(1), 8–34.  
<https://doi.org/10.1007/s11747-011-0278-x>
- Beaujean, A. A. (2014). Sample size determination for regression models using Monte Carlo methods in R. *Practical Assessment, Research, and Evaluation*, 19(1).  
<https://doi.org/10.7275/d5pv-8v28>
- Bentler, P. M. & Chou, C.-P. (1987). Practical Issues in Structural Modeling. *Sociological Methods & Research*, 16(1), 78–117.  
<https://doi.org/10.1177/0049124187016001004>
- Boomsma, A. (2013). Reporting Monte Carlo Studies in Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 20(3), 518–540.  
<https://doi.org/10.1080/10705511.2013.797839>
- Brunner, J. & Austin, P. C. (2009). Inflation of Type I error rate in multiple regression when independent variables are measured with error. *Canadian Journal of Statistics*, 37(1), 33–46. <https://doi.org/10.1002/cjs.10004>
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3. Aufl.). PS Psychologie. Pearson Studium.
- Carroll, R. J. (2014). Measurement Error in Epidemiologic Studies. In N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri & J. L. Teugels (Hg.), *Wiley StatsRef: Statistics Reference Online* (Bd. 28, S. 273). John Wiley & Sons.  
<https://doi.org/10.1002/9781118445112.stat05178>
- Cochran, W. G. (1968). Errors of Measurement in Statistics. *Technometrics*, 10(4), 637–666. <https://doi.org/10.1080/00401706.1968.10490621>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555>
- Fuller, W. A. (1987). *Measurement Error Models*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc. <https://doi.org/10.1002/9780470316665>

- Graham, J. M. (2006). Congeneric and (Essentially) Tau-Equivalent Estimates of Score Reliability. *Educational and Psychological Measurement*, 66(6), 930–944.  
<https://doi.org/10.1177/0013164406288165>
- Hoogland, J. J. & Boomsma, A. (1998). Robustness Studies in Covariance Structure Modeling. *Sociological Methods & Research*, 26(3), 329–367.  
<https://doi.org/10.1177/0049124198026003003>
- Hutcheon, J. A., Chiolero, A. & Hanley, J. A. (2010). Random measurement error and regression dilution bias. *BMJ (Clinical research ed.)*, 340, c2289.  
<https://doi.org/10.1136/bmj.c2289>
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4. Aufl.). The Guilford Press.
- Liu, Y. & Salvendy, G. (2009). Effects of measurement errors on psychometric measurements in ergonomics studies: Implications for correlations, ANOVA, linear regression, factor analysis, and linear discriminant analysis. *Ergonomics*, 52(5), 499–511. <https://doi.org/10.1080/00140130802392999>
- Lord, F. M. & Novick, M. R. (2008). *Statistical theories of mental test scores*. Addison-Wesley Series in Behavioral Science. Information Age Publishing.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum Associates.
- McNeish, D. M. (2014). Modeling sparsely clustered data: design-based, model-based, and single-level methods. *Psychological Methods*, 19(4), 552–563.  
<https://doi.org/10.1037/met0000024>
- Nachtigall, C., Kroehne, U., Funke, F. & Steyer, R. (2003). (Why) Should We Use SEM? Pros and cons of structural equation modeling. *Methods Psychological Research Online*, 8(2), 1–22.
- Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory* (3. Aufl.). McGraw-Hill Series in Psychology. McGraw-Hill.
- Ree, M. J. & Carretta, T. R. (2006). The Role of Measurement Error in Familiar Statistics. *Organizational Research Methods*, 9(1), 99–112.  
<https://doi.org/10.1177/1094428105283192>

- Reise, S. P. (2012). The Rediscovery of Bifactor Measurement Models. *Multivariate Behavioral Research*, 47(5), 667–696. <https://doi.org/10.1080/00273171.2012.715555>
- Reise, S. P., Bonifay, W. E. & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, 95(2), 129–140. <https://doi.org/10.1080/00223891.2012.725437>
- Ripley, B. D. (1987). *Stochastic Simulation*. John Wiley & Sons.  
<https://doi.org/10.1002/9780470316726>
- Schinka, J. A., Velicer, W. F. & Weiner, I. B. (2003). *Research methods in psychology. Handbook of Psychology: / Irving B. Weiner, editor-in-chief ; Vol. 2*. John Wiley & Sons.
- Sedlmeier, P. & Renkewitz, F. (2013). *Forschungsmethoden und Statistik für Psychologen und Sozialwissenschaftler* (2. Aufl.). *Always Learning*. Pearson.