
Die Identifizierung von Missspezifikationen im Strukturmodell von Strukturgleichungsmodellen mit klassischen und neueren Modellfitmaßen

Nadine Juliana Glushko



München, 2018

Die Identifizierung von Misspezifikationen im Strukturmodell von Strukturgleichungsmodellen mit klassischen und neueren Modellfitmaßen

Inauguraldissertation

zur Erlangung des Doktorgrades der Philosophie
an der Ludwig-Maximilians-Universität München

vorgelegt von

Nadine Juliana Glushko

aus Rottenmann

2018

Erstgutachter: Prof. Dr. Moritz Heene
Zweitgutachter: Prof. Dr. Markus Bühner
Datum der mündlichen Prüfung: 05.02.2018

Inhaltsverzeichnis

Zusammenfassung	vi
Studie 1	
1 Einleitung zu Strukturgleichungsmodellen	1
1.1 Allgemeiner Aufbau eines Strukturgleichungsmodells und dessen Verwendung	1
1.2 Möglichkeiten der Modellüberprüfung	4
1.2.1 Der exakte Modellfit – der Chi-Quadrat-Test	6
1.2.2 Der approximative Modellfit – die Fit-Indizes	8
1.2.2.1 Die composite Fit-Indizes	8
1.2.2.2 Die structural Fit-Indizes	11
1.2.3 Cutoff-Werte bei der Verwendung von Fit-Indizes	16
1.3 Probleme bzw. Einschränkungen bei der Verwendung von SEM speziell unter Berücksichtigung der Fit-Indizes	17
1.4 Fragestellung	20
2 Methode des simulierten Strukturgleichungsmodells	21
2.1 Das Populationsmodell	22
2.1.1 Das Strukturmodell	22
2.1.2 Das Populationsmessmodell	23
2.2 Die Missspezifikationen im Struktur- und Messmodell.	25
3 Ergebnisse des exakten und approximativen Modellfits	30
3.1 Ergebnisse des Modellfits zum Modell ohne Fehlerkorrelationen im Messmodell	30
3.1.1 Die Ergebnisse des Chi-Quadrat-Tests – der exakte Modellfit	30

3.1.2 Die Ergebnisse der Fit-Indizes – der approximative Modellfit	34
3.1.2.1 Die composite Fit-Indizes	34
3.1.2.2 Die structural Fit-Indizes	38
3.2 Ergebnisse des Modellfits zum Modell mit Missspezifikationen sowohl im Mess- als auch Strukturmodell	45
4 Diskussion des exakten und approximativen Modellfits	59
4.1 Der exakte Modellfit – der Chi-Quadrat-Test	59
4.2 Der approximative Modellfit – die composite Fit-Indizes	60
4.3 Der approximative Modellfit – die structural Fit-Indizes	61
Studie 2	
5 Einleitung zu den differenzbasierten Fitmaßen	65
5.1 Der two-step approach von Anderson und Gerbing (1988)	65
5.2 Der Δ CFI approach von Widaman (1985)	66
6 Methode der differenzbasierten Fitmaße	67
6.1 Die Methode des two-step approachs von Anderson und Gerbing (1988)	67
6.2 Die Methode des Δ CFI approach von Widaman (1985)	68
7 Ergebnisse der differenzbasierten Fitmaße	69
7.1 Die Ergebnisse des two-step approachs von Anderson und Gerbing (1988)	69
7.2 Die Ergebnisse des Δ CFI approachs von Widaman (1985)	71
8 Diskussion der differenzbasierten Fitmaße	73
8.1 Der two-step approach von Anderson und Gerbing (1988)	73
8.2 Der Δ CFI approach von Widaman (1985)	74

9 Allgemeine Diskussion der durchgeführten Simulationsstudien	75
9.1 Allgemeine Diskussion	75
9.2 Einschränkungen der durchgeführten Simulationsstudien	77
9.3 Implikation für die Praxis	78
Literaturverzeichnis	82
Danksagung	87

Tabellenverzeichnis

Tabelle 1	Vorgeschlagene Cutoff-Werte der composite Fit-Indizes laut Hu und Bentler (1999) sowie Beauducel und Wittmann (2005) und der structural Fit-Indizes laut Hancock und Mueller (2011), McDonald und Ho (2002) und Williams und O'Boyle (2011)	16
Tabelle 2	Übersicht über die Bedingungen	29
Tabelle 3	Alpha-Fehlerraten und Powerraten des Chi-Quadrat-Tests	32
Tabelle 4	Correct rejection rates und wrong rejection rates des RMSEA	34
Tabelle 5	Correct rejection rates und wrong rejection rates des SRMR	36
Tabelle 6	Correct rejection rates und wrong rejection rates des CFI	37
Tabelle 7	Correct rejection rates und wrong rejection rates des RMSEA-P	39
Tabelle 8	Correct rejection rates und wrong rejection rates des RMSEA _s	41
Tabelle 9	Correct rejection rates und wrong rejection rates des SRMR _s	42
Tabelle 10	Correct rejection rates und wrong rejection rates des CFI _s	43
Tabelle 11	Correct rejection rates und wrong rejection rates des RMSEA-P mit Missspezifikationen der Fehlerkorrelationen	47
Tabelle 12	Correct rejection rates und wrong rejection rates des RMSEA _s mit Missspezifikationen der Fehlerkorrelationen	50
Tabelle 13	Correct rejection rates und wrong rejection rates des SRMR _s mit Missspezifikationen der Fehlerkorrelationen.	53
Tabelle 14	Correct rejection rates und wrong rejection rates des CFI _s mit Missspezifikationen der Fehlerkorrelationen	56
Tabelle 15	Powerraten und Alpha-Fehlerraten des two-step approachs	70
Tabelle 16	Correct rejection rates und wrong rejection rates des Δ CFI approachs	71

Abbildungsverzeichnis

Abbildung 1	Aufbau eines typischen Strukturgleichungsmodells mit dem Fokus auf der Unterscheidung zwischen Mess- bzw. Strukturmodells	3
Abbildung 2	Das Strukturmodell des Populationsmodells	23
Abbildung 3	Leichte Missspezifikationsausprägung; Pfad β_{54} auf null gesetzt	25
Abbildung 4	Moderate Missspezifikationsausprägung ; zusätzlich Pfad β_{42} auf null gesetzt	26
Abbildung 5	Schwerwiegende Missspezifikationsausprägung ; zusätzlich Pfad β_{62} auf null gesetzt	27

Zusammenfassung

Die vorliegende Arbeit beschäftigt sich mit der Identifikation von Missspezifikationen im Strukturmodell von Strukturgleichungsmodellen. Dafür wurde vorab ein für die sozialwissenschaftliche Forschung typisches Populationsstrukturgleichungsmodell aus der bestehenden Literatur abgeleitet. In weiterer Folge wurden für dieses Strukturgleichungsmodell alle Parameter wie Faktorkorrelationen und Regressionsgewichte für die durchgeführten Simulationsstudien definiert. Danach wurden unterschiedliche Ausprägungen von Missspezifikation sowohl im Struktur- als auch im Messmodell spezifiziert, um die Zuverlässigkeit der Missspezifikationsdetektion durch Modellfitmaße zu prüfen. Einerseits wurde zur Überprüfung der Modellannahmen der exakte Modellfit mittels Chi-Quadrat-Test getestet, andererseits – mittels Fit-Indizes – auch der approximative Modellfit. Als Missspezifikationsbedingungen wurden sowohl die Anzahl der auf null gesetzten Pfade im Strukturmodell als auch das Ausmaß von Fehlerkorrelationen im Messmodell systematisch variiert, ebenso wie die Stichprobengröße, die Faktorladungshöhe und die Anzahl manifester Variablen.

In der ersten Studie lag der Fokus auf dem exakten und dem approximativen Modellfit. Zum einen wurden die Modellfitmaße unter der Bedingung berechnet, dass nur Missspezifikationen im Strukturmodell vorliegen, und zum anderen auch unter der Bedingung, dass Missspezifikationen sowohl im Mess- als auch im Strukturmodell vorhanden sind. Der exakte Modellfit wurde mittels Chi-Quadrat-Test als Signifikanzüberprüfung getestet. Beim approximativen Modellfit wurden sowohl die composite als auch die structural Fit-Indizes herangezogen, um zu prüfen, ob und welche Ausprägungen der Missspezifikationen angezeigt werden können. Dabei wurden die von Hu und Bentler (1999) vorgeschlagenen Cutoff-Werte für die composite Fit-Indizes verwendet. Für die structural Fit-Indizes wurden die auf Hancock und Mueller (2011), McDonald und Ho (2002) sowie Williams und O'Boyle (2011) basierenden Cutoff-Werte benutzt. Vor allem die structural Fit-Indizes (CFIs, RMSEA-P, RMSEAs und SRMRs), die als besonders sensitiv in Hinblick auf die Detektion von Missspezifikationen im Strukturmodell gelten und eigens dafür konstruiert wurden, wurden in der vorliegenden Studie unter verschiedenen Bedingungen wie Stichprobengröße, Faktorladungshöhe

oder Anzahl der manifesten Variablen pro latenter Variable unter Variierung der Missspezifikationsausprägungen im Mess- und im Strukturmodell geprüft.

In der zweiten Studie wurden zwei weitere Ansätze der Modellüberprüfung herangezogen. Zum einen wurde der auf der Chi-Quadrat-Differenz basierende two-step approach von Anderson und Gerbing (1988), verwendet, um die Missspezifikationen im Strukturmodell zu identifizieren, und zum anderen der Δ CFI approach von Widaman (1985). Diese Ansätze haben gemeinsam, dass sie den Modellfit separat bestimmen, das heißt, es wird einerseits das gesamte Strukturgleichungsmodell, bestehend aus Mess- und Strukturmodell, auf den Modellfit hin überprüft und andererseits das Messmodell. Die Differenz der daraus resultierenden Fitmaße (Δ Chi-Quadrat oder Δ CFI) dient dann als Maß für den Modellfit des Strukturmodells, wenn zuvor sichergestellt wurde, dass das Messmodell passt.

Die Ergebnisse der ersten Studie zeigen klar, dass der exakte Modellfit mittels Chi-Quadrat-Test im Vergleich zu den composite Fit-Indizes die zuverlässigeren Ergebnisse liefert. Zudem bestätigen die Ergebnisse der vorliegenden Simulationsstudien erneut, dass die von Hu und Bentler (1999) für die composite Fit-Indizes vorgeschlagenen Cutoff-Werte nicht geeignet sind, um hinreichend gut Missspezifikationen im Strukturmodell anzuzeigen. Die structural Fit-Indizes unter Verwendung der vorgeschlagenen Cutoff-Werte laut Hancock und Mueller (2011), McDonald und Ho (2002) sowie Williams und O'Boyle (2011) konnten durchgehend die am stärksten ausgeprägte Missspezifikation im Strukturmodell entdecken. Bei den leichteren Ausprägungen der Missspezifikationen im Strukturmodell zeigte sich deutlich die Abhängigkeit der Zuverlässigkeit der structural Fit-Indizes besonders von Faktorladungshöhen und Testlänge. Somit waren die structural Fit-Indizes entgegen den theoretischen Annahmen nicht unabhängig von Gegebenheiten des Messmodells.

Die Ergebnisse der zweiten Studie verdeutlichen, dass der two-step approach von Anderson und Gerbing (1988) und der Δ CFI approach von Widaman (1985) gute Ergebnisse bringen, wenn es darum geht, zuverlässig Missspezifikationen im Strukturmodell zu entdecken. Diese beiden Ansätze können aber nur dann uneingeschränkt angewandt werden, wenn ein hinreichend gut validiertes Messmodell vorliegt.

Auf Basis dieser Befunde kann vorsichtig davon ausgegangen werden, dass in den meisten publizierten Studien die Passung des Strukturmodells inakzeptabel ist, entgegen den Aussagen auf Grundlage der composite Fit-Indizes.

Das Fazit der Analysen legt in Bezug auf die bisherige angewandte Forschung mit Strukturgleichungsmodellen die Vermutung nahe, dass publizierte Strukturgleichungsmodelle gerade im Strukturteil misspezifiziert sein könnten, obwohl die üblicherweise verwendeten composite Fit-Indizes einen mindestens hinreichenden Modellfit anzeigen. Dies wirft erhebliche Fragen bezüglich der Validität der im Strukturmodell abgebildeten inhaltlichen Theorien auf.

Als Empfehlung für die Forschungspraxis lässt sich sagen, dass nicht nur Fit-Indizes alleine herangezogen werden sollten, um sicherzustellen, dass die theoretischen Annahmen besonders im Strukturmodell durch empirische Daten ausreichend gestützt werden. Das Gewicht der Fitanalyse sollte daher wieder stärker auf den exakten Modellfit mittels Chi-Quadrat-Test (Gesamt-Chi-Quadrat-Test) verschoben werden, da er sich generell betrachtet als durchaus teststarke Methode erwies. Zudem scheinen auch die structural Fit-Indizes den composite Fit-Indizes vorzuziehen zu sein, wobei auch die structural Fit-Indizes teilweise hinter den Erwartungen bleiben. Unter Annahme eines hinreichend gut validierten Messmodell erwiesen sich auch die differenzbasierten Fitmaße (two-step approach und Δ CFI approach) als durchwegs zuverlässig. Weiterhin sollte, auf dem exakten und dem approximativen Modellfit aufbauend, der nachfolgenden Identifizierung von Misspezifikationen ein großer Raum gegeben werden, entgegen der gängigen Praxis der Bestimmung der Güte der Modellpassung mittels Fit-Indizes. Die Ergebnisse legen insgesamt auch nahe, Modellfitanalysen nur mit ausreichend großen Stichproben und reliablen Daten durchzuführen, um zu validen Ergebnissen der Modellbeurteilung zu gelangen.

Studie 1

1 Einleitung zu Strukturgleichungsmodellen

Die Datensätze in den Sozial- und Verhaltenswissenschaften haben sich in den letzten 40 Jahren stark verändert. Es gilt nun, einerseits eine große Anzahl an Daten gleichzeitig und andererseits inhaltlich sehr komplexe Überlegungen zuverlässig überprüfen zu können. Im Rahmen dieser Entwicklung stellen Strukturgleichungsmodelle (SEM) in den letzten drei Jahrzehnten ein Standardverfahren dar, um komplexe Zusammenhänge von inhaltlich postulierten Annahmen in unterschiedlichen sozialwissenschaftlichen Disziplinen empirisch überprüfen zu können, unter anderem in den Bereichen der Psychologie und der Soziologie, aber auch im Bereich der Ökonomie.

Die Relevanz von Strukturgleichungsmodellen in den Sozialwissenschaften kann schon durch eine Stichwortsuche in „Google Scholar“ veranschaulicht werden. Alleine das Wort „Strukturgleichungsmodelle“ ergibt gerundet 6200 Ergebnisse (Stand: September 2017). Bei Eingabe von „structural equation modeling“ erhält man rund 2 930 000 Treffer (Stand: September 2017). Dies verdeutlicht die Relevanz von Strukturgleichungsmodellen und spiegelt gleichzeitig die hohe Verantwortung hinsichtlich wissenschaftlich zuverlässiger Ergebnisse wider.

1.1 Allgemeiner Aufbau eines Strukturgleichungsmodells und dessen Verwendung

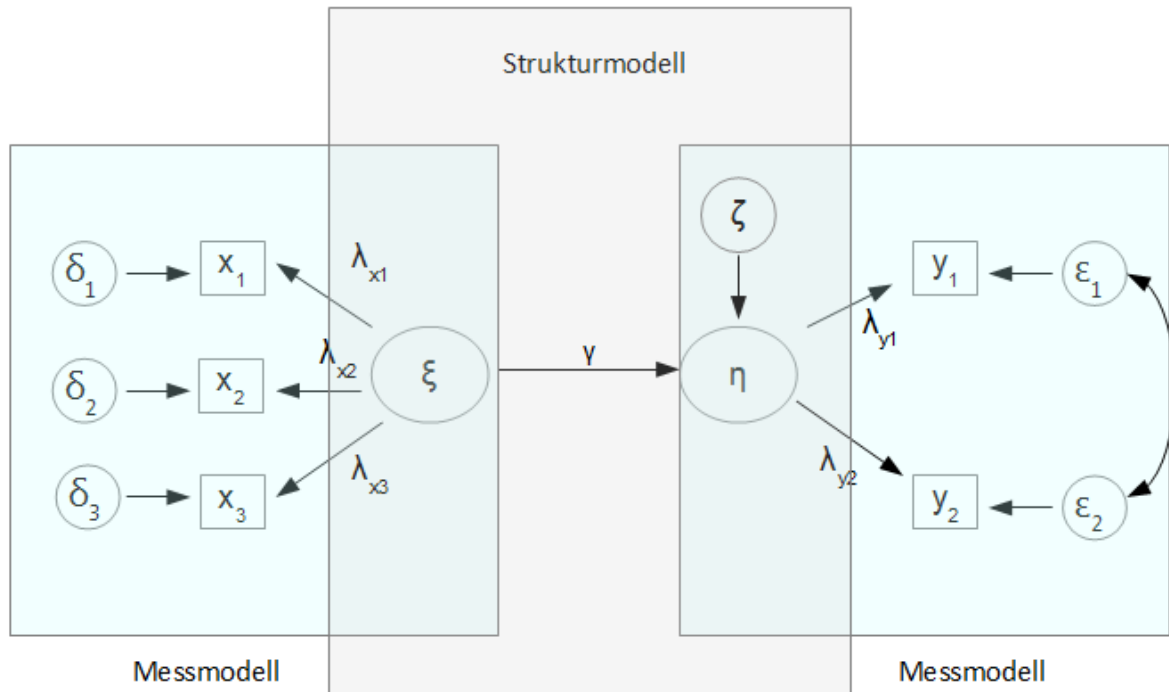
Ein Strukturgleichungsmodell bildet die Struktur bzw. den Zusammenhang von Variablen ab. In jedem angenommenen Strukturgleichungsmodell gibt es eine bestimmte Anzahl an latenten und manifesten Variablen, die eine bestimmte Zusammenhänge aufweisen. Manifeste Variablen sind dabei die gemessenen, sprich die beobachteten Variablen. Unter latenten Variablen versteht man Variablen, die theoretisch angenommen werden und die, die Kovarianzen bzw. Korrelationen zwischen manifesten Variablen reproduzieren bzw. „erklären“. Dabei stellen manifeste Variablen die Indikatoren für latente Variablen dar, die man a priori entsprechend operationalisieren muss. Um ein Beispiel zu nennen, denke man an das multidimensionale Persönlichkeitsinventar NEO-FFI von Costa und McCrae (1985). Dabei wird davon ausgegangen, dass jeweils zwölf Items, sprich zwölf manifeste Variablen, Indikatoren einer latenten Variable sind, wie etwa der latente Variable „Extraversion“.

Zwischen zwei oder mehreren latenten Variablen können gerichtete (Regressionsgewicht) oder ungerichtete Zusammenhänge (Korrelationen) bestehen. Der Pfadkoeffizient, auch Ladung genannt, drückt dabei die Stärke dieses Zusammenhangs aus.

Ein Strukturgleichungsmodell setzt sich aus zwei Teilen zusammen. Einerseits besteht es aus einem Messmodell und andererseits aus einem Strukturmodell (Anderson & Gerbing, 1982; Joreskog & Sorbom, 1981). Im Messmodell wird die Verknüpfung zwischen den manifesten und den latenten Variablen abgebildet. Das Messmodell definiert also, welche beobachteten Variablen Indikatoren für die latenten Variablen darstellen. Somit wird im Messmodell die Operationalisierung der latenten Variablen spezifiziert. Im Strukturmodell hingegen wird die Verknüpfung zwischen latenten Variablen oder zwischen latenten und manifesten Variablen, die nicht als Indikatoren dienen, abgebildet. Anders formuliert, die inhaltlichen Aussagen einer angenommenen Theorie werden im Strukturmodell abgebildet.

Abbildung 1

Aufbau eines typischen Strukturgleichungsmodells mit dem Fokus auf der Unterscheidung zwischen Mess- bzw. Strukturmodell



Notation:

\square	manifeste Variablen
\circ	latente Variablen
\leftrightarrow	Korrelation oder Kovarianz
\rightarrow	Regressionsgewicht
ξ	latente unabhängige Variable
η	latente abhängige Variable
λ	Faktorladung
ε	Fehlervariable
x und y	Indikatoren der latenten Variablen

Der Hauptvorteil der Verwendung von Strukturgleichungsmodellen liegt darin, dass latente Variablen und deren Indikatoren explizit getrennt und dadurch Messfehler schätzbar werden. Dies hat zur Folge, dass die Beziehungen zwischen latenten Variablen unbeeinflusst vom Messfehler sind, da dieser alleine Bestandteil der manifesten Variablen

ist. Somit sind dann die Beziehungen (Pfadkoeffizienten oder Korrelationen) zwischen latenten Variablen nicht vom Messfehler gemindert.

Strukturgleichungsmodelle werden unter anderem häufig zur Überprüfung von Theorieannahmen herangezogen. Dabei werden inhaltliche Annahmen in Form von Zusammenhangsstrukturen zwischen den angenommenen Variablen postuliert, die dann mittels statistischer Modellüberprüfung getestet werden. Somit wird ein A-priori-Modell auf Übereinstimmung mit empirischen Daten geprüft. Es gilt die Frage zu beantworten, ob das angenommene Modell der Prüfung standhält oder verworfen bzw. gegebenenfalls angepasst werden muss.

Die Modellprüfung nimmt also einen bedeutenden Teil bei der Strukturgleichungsmodellierung ein. Bevor die Parameterschätzungen eines Modells interpretiert werden können, muss die Passung des Modells mittels Daten sichergestellt sein. Andernfalls können aus nicht-passenden Modellen Verzerrungen von Parameterschätzungen resultieren (Yuan, Marshall & Bentler, 2003), die zu falschen Interpretationen bezüglich der theoretischen Implikationen führen können. Es stellt sich somit die zentrale Frage, wie zuverlässig die bisherige Modellprüfungspraxis in Bezug auf Aufdeckung von Modellmisspezifikationen funktioniert und wie zuverlässig neuere Methoden der Modelltestung sind.

1.2 Möglichkeiten der Modellüberprüfung

Wenn man ein hypothetisches Modell hat und seine Modellparameter nach der Schätzung der Parameter anhand empirischer Daten interpretieren möchte, muss man also zuerst überprüfen, ob das Modell das beobachtete Phänomen überhaupt hinreichend gut abbildet.

Im Sinne des Falsifikationsprinzips in wissenschaftlichen Bereichen ist es wichtig, bei der Überprüfung von komplexen inhaltlichen Annahmen mögliche Modellabweichungen zuverlässig detektieren zu können. Postulierte Modellannahmen können falsch sein, daher stellt sich die Frage, ab wann ein postuliertes Modell als richtig oder falsch gilt. Anders gefragt – wie kann überprüft werden, ob ein angenommenes Modell von den empirischen Daten hinreichend unterstützt wird?

Zur Beantwortung dieser Frage der Passung zwischen Daten und Modell existieren statistische Modellüberprüfungsverfahren. Wenn nun eine Modellüberprüfung

durchgeführt wird, sprich die Testung, ob das angenommene A-priori-Modell passt, dann werden empirisch Daten erhoben und es wird statistisch überprüft, ob diese Daten die postulierten Modellannahmen hinreichend abbilden können. Es wird also getestet, ob es eine hinreichende Passung zwischen dem beobachteten Phänomen und dem hypothetischen Modell gibt (Schermelleh-Engel & Moosbrugger, 2003).

Generell unterscheidet man zwei Arten, wie man den Modellfit überprüfen kann (Yuan, 2005; McDonald & Ho, 2002). Einerseits kann ein Chi-Quadrat-Test zur Überprüfung herangezogen werden, mit dem ein exakter Modellfit (exact data-model fit) berechnet wird. Hierdurch werden „Entweder-oder-Aussagen“ über Passung oder Nicht-Passung der Daten mit dem Modell möglich. Andererseits besteht die Möglichkeit, den Modellfit über sogenannte Fit-Indizes zu testen, wobei sich ein approximativer Modellfit (close fit) ergibt, der gegenüber dem Chi-Quadrat-Test das *Ausmaß* der Daten-Modellpassung quantifizieren soll.

Mathematisch betrachtet wird bei der Daten-Modellpassung die vom Modell implizierte Kovarianzmatrix mit der Kovarianzmatrix der beobachteten Variablen verglichen (u.a. Hoyle, 1995). Die implizierte Matrix resultiert aus den geschätzten Parameterwerten und den Modellrestriktionen. Die beobachtete Kovarianzmatrix beinhaltet die Kovarianzen zwischen den manifesten Variablen.

Beim Chi-Quadrat-Test handelt es sich um ein differenzstatistisches Maß, das beurteilt, ob das angenommene Modell hinreichend gut mittels empirisch erhobener Daten abgebildet werden kann (u.a. Hoyle, 1995).

Im Gegensatz dazu spiegeln die Fit-Indizes den „Closeness-of-fit“ (wie nahe man dem perfekten Modellfit kommt) bzw. den „Badness-of-fit“ (wie weit man vom perfekten Modellfit entfernt ist) wider (u.a. Bentler, 1990). Die Überprüfung mittels Fit-Indizes wird in der angewandten Forschung oft eingesetzt, da der Chi-Quadrat-Test zur Ermittlung eines exakten Modellfits ein strenges Verfahren darstellt, das dazu führen kann, dass bei einer großen Stichprobengröße bereits kleine Modellabweichungen angezeigt werden. Zudem zeigen die Fit-Indizes das Ausmaß der Missspezifikation an.

1.2.1 Der exakte Modellfit – der Chi-Quadrat-Test

Beim sogenannten exakten Modellfit wird ein vorher angenommenes Strukturgleichungsmodell durch einen Chi-Quadrat-Test auf Passung mit den empirisch erhobenen Daten geprüft (Schermelleh-Engel & Moosbrugger, 2003). Dabei soll die Diskrepanz zwischen der vom Modell implizierten Kovarianzmatrix und der Populationskovarianzmatrix möglichst gering sein (u.a. Hoyle, 1995).

Diese Diskrepanz F („value of the fitting function“, s. u.a. Hoyle, 1995) zwischen der beobachteten und der implizierten Kovarianzmatrix wird mit der Stichprobengröße gewichtet. Dieser gewichtete Wert stellt dann die Prüfgröße für den Chi-Quadrat-Test dar (Hoyle, 1995).

Als Nullhypothese H_0 wird dabei angenommen, dass die empirischen Daten die Annahmen des Modells abbilden können, also dass die beobachtete Kovarianzmatrix durch das Modell beschrieben bzw. reproduziert werden kann. Statistisch ausgedrückt heißt das, dass die Populationskovarianzmatrix Σ der vom Modell implizierten Matrix S entspricht.

Somit ergibt sich für die Nullhypothese:

$$H_0: \Sigma = S$$

S = implizierte Populationskovarianzmatrix

Σ = modelltheoretische Kovarianzmatrix des spezifizierten Modells in der Population

Die H_1 -Hypothese besagt hingegen, dass die empirisch erhobenen Daten die vom Modell postulierten Annahmen nicht abbilden können und somit das angenommene Modell verworfen werden muss.

Die H_1 -Hypothese besagt somit:

$$H_1: \Sigma \neq S$$

S = implizierte Populationskovarianzmatrix

Σ = modelltheoretische Kovarianzmatrix des spezifizierten Modells in der Population

Als logische Konsequenz ist ein nicht signifikantes Ergebnis wünschenswert, da dies zeigen würde, dass die Annahme gleicher Kovarianzmatrizen nicht verworfen werden muss. Im Umkehrschluss muss bei einem signifikanten Ergebnis das angenommene Modell verworfen werden.

Die Formel zur Berechnung des Chi-Quadrat-Tests lautet:

$$\chi^2(df) = (N - 1) \cdot F$$

wobei:

F ... value of the fitting function

N ... Stichprobengröße

df ... Freiheitsgrade

df ... $m - q$

m ... Anzahl bekannter Parameter

$m = [p(p + 1)]/2$

p ... Anzahl der Variablen

q ... Anzahl zu schätzender Parameter

Da es sich beim Chi-Quadrat-Test um eine Signifikanzprüfung handelt, wird für gewöhnlich ein Alpha-Niveau von 5% angenommen.

Generell beantwortet der Chi-Quadrat-Test, da er ein Inferenztest ist, ausschließlich die Frage, ob das angenommene Modell passt oder eben nicht. Aufgrund der Tatsache, dass bei einer großen Stichprobengröße der Stichprobenfehler kleiner wird, die Ergebnisse somit genauer werden und daher die Teststärke steigt, führen bereits kleine Modellabweichungen zu einer Verwerfung der H_0 . Umgekehrt können kleine Stichproben dazu führen, dass selbst starke Modellabweichungen unentdeckt bleiben und fälschlicherweise das Modell als richtig angenommen wird.

1.2.2 Der approximative Modellfit - die Fit-Indizes

Da der Chi-Quadrat-Test nur die Frage beantwortet, ob ein Modellfit vorliegt oder nicht, bleibt die Frage nach der Größe möglicher detektierter Abweichungen offen. Diese kann mittels Fit-Indizes beantwortet werden, da die Fit-Indizes das Ausmaß bzw. die Effektstärke des Missfits anzeigen (u.a. Bentler, 1990).

Generell kann man Fit-Indizes nach verschiedenen Gesichtspunkten unterscheiden (z.B. absoluter oder komparativer Fit-Index). Der Übersicht halber wird in der vorliegenden Arbeit vor allem zwischen den composite Fit-Indizes und den structural Fit-Indizes unterschieden. Hervorzuheben ist, dass die structural Fit-Indizes speziell Missspezifikationen im Strukturmodell des Strukturgleichungsmodells erkennen sollen (Hancock & Mueller, 2011). Generell werden in der Literatur häufiger die composite Fit-Indizes verwendet.

1.2.2.1 Die composite Fit-Indizes

Comparative Fit Index (CFI)

Der CFI (Bentler, 1990) gehört zu den sogenannten komparativen Fit-Indizes. Komparative bzw. inkrementelle Fit-Indizes vergleichen die relative Verbesserung des Modellfits mit dem Modellfit eines restriktiveren Nullmodells. Ein Nullmodell wird darüber definiert, dass alle Kovarianzen zwischen den manifesten Variablen auf null gesetzt werden. Konzeptuell gesprochen vergleicht der CFI die Modellpassung des hypothetischen Modells mit derjenigen des Nullmodells, indem er den auf dem Chi-Quadrat basierenden Fit (Nonzentralitätsparameter, siehe bspw. Mulaik, 2009) des hypothetischen Modells ins Verhältnis zu dem des Nullmodells setzt. Dabei wird das hypothetische Modell in der Formel als Zielmodell benannt. Der Vergleich dieser proportionalen Verbesserung zwischen dem angenommenen Modell und dem restriktiven Nullmodell kommt auch in der Formel des CFI zum Ausdruck. Dabei wird im Zähler das getestete Modell dargestellt und im Nenner das restriktive Nullmodell.

Die Formel zur Berechnung des CFI lautet:

$$CFI = 1 - \frac{\chi^2 T - df T}{\chi_0^2 - df_0}$$

wobei:

T ... Minimum der Diskrepanzfunktion

$\chi^2 T$... Chi-Quadrat-Statistik für das Zielmodell

$df T$... Freiheitsgrade für das Zielmodell

χ_0^2 ... Chi-Quadrat-Statistik für das Nullmodell

df_0 ... Freiheitsgrade für das Nullmodell

df ... $m - q$

m ... Anzahl bekannter Parameter

$m = [p(p + 1)]/2$

p ... Anzahl der Variablen

q ... Anzahl zu schätzender Parameter

Der Wertebereich des CFI liegt zwischen 0 und 1, wobei 1 bedeutet, dass ein perfekter Modellfit vorliegt.

Ist der Quotient aus dem getesteten Modell und dem restriktiven Nullmodell gleich 1, so ergibt sich über die Subtraktion der Differenzwert 0, was bedeutet, dass kein Modellfit vorliegt. Je größer der Unterschied in der Differenz zwischen dem getesteten Modell und dem restriktiven Nullmodell wird, umso kleiner wird der Wert des Bruches und desto mehr nähert sich das Ergebnis dem Wert 1.

Root Mean Squared Error Of Approximation (RMSEA)

Der RMSEA (Steiger, 1990) zählt zu den sogenannten absoluten Fit-Indizes. Durch den Chi-Quadrat-Wert im Zähler wird beim RMSEA die Abweichung zwischen der beobachteten und der implizierten Kovarianzmatrix betrachtet. Dabei gilt, dass der RMSEA umso größer ausfällt, je größer diese Diskrepanz ist. Die Formel definiert die typische Modellabweichung pro Dimension (df), in der ein Modell von den Daten abweichen kann (Mulaik, 2009).

Die Formel für den RMSEA lautet:

$$RMSEA = \sqrt{\max\left\{\frac{\chi^2 - df}{N * df}, 0\right\}}$$

wobei:

χ^2 ... Chi-Quadrat-Statistik

df ... Freiheitsgrade

df ... $m - q$

m ... Anzahl bekannter Parameter

$m = [p(p + 1)]/2$

p ... Anzahl der Variablen

q ... Anzahl zu schätzender Parameter

N ... Stichprobengröße

Die Freiheitsgrade beeinflussen sowohl den Nenner als auch den Zähler. Bei komplexeren Modellen fällt die Differenz zwischen bekannten und zu schätzenden Parametern geringer aus. Daher ist der Zähler bei komplexen Modellen größer, da in diesem Fall die Anzahl an Freiheitsgraden gering ist. Für den Nenner gilt, dass dieser bei einer niedrigen Anzahl an Freiheitsgraden kleiner wird. Die Modellkomplexität wird somit beim RMSEA über die Freiheitsgradanzahl im Nenner berücksichtigt. Je größer der Zähler wird, desto kleiner wird der Nenner, folglich wird der Wert des Bruches größer und das Ergebnis für den RMSEA steigt.

Auch hier liegt der Wertebereich zwischen 0 und 1, wobei nun der Wert 0 einen perfekten Modellfit anzeigt.

Standardized Root Mean Square Residual (SRMR)

Auch der SRMR (Jöreskog & Sörbom, 1981) zählt zu den absoluten Fit-Indizes, wobei ein Wert von 0 wieder einen perfekten Fit anzeigt. Wie in der unten angeführten Formel ersichtlich, wird beim SRMR die Wurzel aus der standardisierten durchschnittlichen Abweichung zwischen den vom Modell reproduzierten und den beobachteten Korrelationen berechnet. Die Modellkomplexität wird hierbei nicht beachtet.

Die Formel für den SRMR ist:

$$SRMR = \sqrt{\left\{ 2 \sum_{i=1}^p \sum_{j=1}^i \left[\frac{s_{ij} - \hat{\delta}_{ij}}{s_{ii} s_{jj}} \right]^2 \right\} / p(p+1)}$$

wobei:

s_{ij} ... beobachtete Kovarianzen zwischen zwei manifesten Variablen

p ... Anzahl der beobachteten Variablen

$\hat{\delta}_{ij}$... reproduzierte Kovarianzen

s_{ii} und s_{jj} ... Standardabweichung der beiden Variablen

1.2.2.2 Die structural Fit-Indizes

Da im Strukturteil des Strukturgleichungsmodells die allermeisten inhaltlich relevanten Annahmen und Hypothesen abgebildet sind, liegt darauf meistens der Fokus der angewandten Forschung. Zudem ist es ein bekanntes Problem, dass ein perfektes Messmodell einen möglichen Missfit im Strukturmodell überdecken kann (Hancock und Mueller, 2011). Die structural Fit-Indizes wurden genau aus diesem Grund entwickelt, um speziell die kausale Struktur im Strukturteil zu prüfen (Hancock und Mueller, 2011).

CFI_s - der auf den Strukturteil bezogene Comparative Fit Index

Neben dem oben beschriebenen CFI wird auch der CFI_s verwendet. Diese Form des CFI geht auf Hancock und Mueller (2011) zurück. Auch dieser Fit-Index vergleicht die proportionale Verbesserung eines angenommenen Strukturgleichungsmodells mit einem restriktiven Nullmodell, wobei es sich hier um eine Variante des CFI handelt, die sich nur auf den Strukturteil des Modells bezieht, während der CFI immer das Gesamtmodell überprüft.

Die Formel des CFI_s lautet:

$$CFI_s = 1 - \frac{\max\left[\left(\tilde{T}_{\text{model}} - df_{\text{model}}\right), 0\right]}{\max\left[\left(\tilde{T}_{\text{model}} - df_{\text{model}}\right), \left(\tilde{T}_{\text{null}} - df_{\text{null}}\right), 0\right]}$$

$$= 1 - \frac{\max\left\{\left[(N-1)\tilde{F}_{\text{model}} - df_{\text{model}}\right], 0\right\}}{\max\left\{\left[(N-1)\tilde{F}_{\text{model}} - df_{\text{model}}\right], \left[(N-1)\tilde{F}_{\text{null}} - df_{\text{null}}\right], 0\right\}}$$

$$df_{\text{null}} = \frac{K(K+1)}{2} - K = \frac{K(K-1)}{2}.$$

wobei:

\tilde{T}_{model} ... Minimum der Diskrepanzfunktion des Strukturmodells

\tilde{T}_{null} ... Minimum der Diskrepanzfunktion des restriktiven Nullmodells

df_{model} ... Freiheitsgrade des Strukturmodells

df_{null} ... Freiheitsgrade des restriktiven Nullmodells

\tilde{F}_{model} ... value of the fitting function des Strukturmodells

\tilde{F}_{null} ... value of the fitting function des restriktiven Nullmodells

N ... Stichprobengröße

K ... Anzahl der latenten Variablen ξ und der Faktoren η

Auch hierbei gilt, dass der Wert 1 einen perfekten Modellfit aufzeigt.

RMSEA-P - der auf den Strukturteil bezogene Root Mean Squared Error Of Approximation, basierend auf McDonald und Ho (2002) und Williams und O'Boyle (2011)

Zusätzlich wurde der RMSEA-P in der vorliegenden Simulationsstudie berücksichtigt. McDonald und Ho (2002) passten die Formel des RMSEA insofern an, als der RMSEA-P nun speziell den Fit des Strukturmodells prüft. Diese angepasste Formel wurde ebenso von Williams und O'Boyle (2011) verwendet. Auch hierbei wird der Unterschied zwischen dem angenommenen Modell und dem saturierten Modell bezüglich der Verbesserung des Modellfits berechnet.

Die Formel des RMSEA-P lautet:

$$\text{RMSEA-P} = \{[\chi_p^2 - df_p] / [df_p(N-1)]\}^{1/2}$$

wobei :

N ... Stichprobengröße

und χ_p^2 und df_p stehen für die Differenzen von χ^2 und df zwischen dem Gesamtmodell und dem saturierten strukturellen Teil des Modells

df ... $m-q$

m ... Anzahl bekannter Parameter

$m = [p(p+1)]/2$

p ... Anzahl der Variablen

q ... Anzahl zu schätzender Parameter

Auch hier gilt, dass der Wert 0 einen perfekten Modellfit widerspiegelt.

RMSEA_s - der auf den Strukturteil bezogene Root Mean Squared Error Of Approximation, basierend auf Hancock und Mueller (2011)

Eine weitere Variante des RMSEA, der RMSEA_s, geht auf Hancock und Mueller (2011) zurück. Auch diese Variante fokussiert den Strukturteil von Strukturgleichungsmodellen.

Die Formel des RMSEA_s ist:

$$RMSEA_s = \sqrt{\max\left[\left(\frac{\tilde{T}_{\text{model}} - df_{\text{model}}}{df_{\text{model}}(N-1)}\right), 0\right]} = \sqrt{\max\left[\left(\frac{\tilde{F}_{\text{model}}}{df_{\text{model}}} - \frac{1}{(N-1)}\right), 0\right]}$$

$$df_{\text{model}} = \frac{K(K+1)}{2} - t_{\text{model}}$$

wobei:

\tilde{T}_{model} ... Minimum der Diskrepanzfunktion des Strukturmodells

df_{model} ... Freiheitsgrade des Strukturmodells

\tilde{F}_{model} ... value of the fitting function des Strukturmodells

t_{model} ... Anzahl der freien Parameter

N ... Stichprobengröße

K ... Anzahl der latenten Variablen ξ und der Faktoren η

Auch beim RMSEA_s drückt der Wert 0 einen perfekten Modellfit aus.

Der Unterschied zwischen RMSEA-P und RMSEA_s liegt darin, dass der RMSEA-P die Passung des Strukturmodells nicht unabhängig von sogenannten „nuisance parameters“ überprüfen kann, da der Chi-Quadrat-Wert aus dem composite Modell verwendet wird. Aus der Literatur ist bekannt, dass dieser Chi-Quadrat-Wert von „nuisance parameters“ beeinflusst wird (Hancock & Mueller, 2011; cf. Heene et al., 2011; Saris & Satorra, 1988);

Saris et al.; 2009; Savalei, 2012). Unter „nuisance parameters“ versteht man Parameter, die bei der Schätzung eines Modells einen Einfluss haben, obwohl sie eigentlich nicht von Interesse sind, wie z.B. die Höhe der Faktorladungen oder die Anzahl der manifesten Variablen. Wie verschiedene Autoren zeigen konnten (Hancock & Mueller, 2011; cf. Heene et al., 2011; Saris & Satorra, 1988; Saris et al.; 2009; Savalei, 2012) nehmen allerdings diese Parameter Einfluss auf die Sensitivität, Misspezifikationen eines Modells zu entdecken. Im Gegensatz dazu soll der RMSEA_s durch die Differenz der Chi-Quadrat-Werte der Kovarianzmatrix des Modells und der Kovarianzmatrix des saturierten Modells unabhängig von diesen Faktoren sein und damit einen Indikator zur Überprüfung des Strukturmodells bei Strukturgleichungsmodellen darstellen (Hancock & Mueller, 2011).

SRMR_s - der auf den Strukturteil bezogene Standardized Root Mean Square Residual

Auch zum SRMR wurde zusätzlich der SRMR_s (Hancock und Mueller, 2011) verwendet. Er bezeichnet die Wurzel aus der standardisierten durchschnittlichen Abweichung zwischen den vom Modell reproduzierten und den beobachteten Korrelationen der *latenten* Variablen.

Die Formel des SRMR_s lautet:

$$SRMR_s = \frac{2}{K(K+1)} \sum_{i \geq j} \left[\frac{\hat{\omega}_{(sat)ij} - \hat{\omega}_{(model)ij}}{\sqrt{\hat{\omega}_{(model)ii} \hat{\omega}_{(model)jj}}} \right]^2$$

wobei:

K ... Anzahl der latenten Variablen ξ und der Faktoren η

$\hat{W}_{(sat)ij}$... beobachtete Kovarianzmatrix der latenten Variablen

$\hat{\omega}_{(model)ij}$... replizierte Kovarianzmatrix der latenten Variablen

Ein Wert von 0 drückt hierbei wieder einen perfekten Modellfit aus.

1.2.3 Cutoff-Werte bei der Verwendung von Fit-Indizes

Im Gegensatz zum exakten Modellfit handelt es sich bei den Fit-Indizes um keine Signifikanzprüfung. Hierbei finden sogenannte Cutoff-Werte ihre Anwendung, die zeigen sollen, ob ein Modell als passend oder nicht passend angesehen werden kann. Die Cutoff-Werte sind nötig, da die Fit-Indizes alleine noch nicht aussagen, was nun ein großer bzw. kleiner Missfit ist. Daher wurden Cutoff-Werte für vordefinierte Missspezifikationen definiert (Hu & Bentler, 1999). Die von Hu und Bentler (1999) ermittelten Cutoff-Werte wurden für die Fit-Indizes RMSEA (Steiger, 1998), CFI (Bentler, 1990) und SRMR (z.B. Hu & Bentler) angewandt. Die vorgeschlagenen Cutoff-Werte von Hu und Bentler (1999) konnten auch von Beauducel und Wittmann (2005) bei Verletzung der Einfachstruktur bei konfirmatorischen Faktorenanalysen bestätigt werden. Für die structural Fit-Indizes lehnten sich Hancock und Mueller (2011), McDonald und Ho (2002) und Williams und O'Boyle (2011) an die Cutoff-Werte der composite Fit-Indizes von Hu und Bentler (1999) bzw. Beauducel und Wittmann (2005) an. In der Tabelle sind die gängig angewandten Cutoff-Werte für einen passenden Modellfit angeführt.

Tabelle 1

Vorgeschlagene Cutoff-Werte der composite Fit-Indizes laut Hu und Bentler (1999) sowie Beauducel und Wittmann (2005) und der structural Fit-Indizes laut Hancock und Mueller (2011), McDonald und Ho (2002) und Williams und O'Boyle (2011)

Fit-Indizes	Cutoff-Werte
RMSEA-P/RMSEA _s /RMSEA	<.05
SRMR _s /SRMR	<.08
CFI _s /CFI	>.95

1.3 Probleme bzw. Einschränkungen bei der Verwendung von SEM speziell unter Berücksichtigung der Fit-Indizes

Die häufige Verwendung von Strukturgleichungsmodellen zur Überprüfung von komplexen Modellannahmen, bei der den Fit-Indizes eine besondere Relevanz zugesprochen wird, wurde bereits zu Beginn in der Einleitung erörtert. Nun stellt sich die Frage, wie zuverlässig komplexe Modellannahmen überprüft werden können und speziell, welche Probleme bisherige Studien dazu aufzeigen konnten.

Die am häufigsten verwendeten Cutoff-Werte gehen auf Hu und Bentler (1999) zurück. Hu und Bentler spezifizierten zwei unterschiedlich komplexe Populationsmodelle sowie Missspezifikationen in den theoretischen Modellen. Sie stellten sich die Frage, bei welchen Cutoff-Werten die nicht passenden Modelle hinreichend gut abgelehnt werden können. Die daraus berechneten Cutoff-Werte der Fit-Indizes beschränken sich somit aber auf diese von Hu und Bentler (1999) analysierten Modelle. Dennoch wurden diese Cutoff-Werte als „Goldene Regel“ verstanden, obwohl Hu und Bentler (1999) keine Empfehlung geben, sondern vielmehr Vorschläge machen wollten.

In der Vergangenheit hat sich gezeigt, dass feste Cutoff-Werte nur eine geringe Sensitivität aufweisen, Missspezifikationen zu entdecken (Breivik & Olsson, 2001; Chen, Curran, Bollen, Kirby & Paxton, 2008; Fan & Sivo, 2007; Heene, Hilbert, Draxler, Ziegler & Bühner, 2011; Heene, Hilbert, Freudenthaler & Bühner, 2012; McDonald & Ho, 2002; Saris & Satorra, 1993; Saris, Satorra & van der Veld, 2009; Tomarken & Waller, 2003; Yuan, 2005). Die Studien zeigen auch, dass die zuverlässige Detektion von Missspezifikationen von unterschiedlichen Indikatoren wie Faktorladungshöhen, Stichprobengröße und Art der Missspezifikation abhängig sind. All diese Parameter haben allerdings nichts mit der eigentlichen Größe bzw. dem eigentlichen Ausmaß der Missspezifikation selbst zu tun, beeinflussen aber dennoch maßgeblich die Sensitivität der Cutoff-Werte.

Als zusammenfassendes Ergebnis lässt sich festhalten, dass die Sensitivität der Cutoff-Werte gängiger Fit-Indizes sehr niedrig ist (u.a. Breivik & Olsson, 2001), speziell bei Modellen, die komplexer als die von Hu und Bentler (1999) sind und denen realistischere Faktorladungshöhen zugrunde liegen. Hu und Bentler (1999) gingen bei ihren Simulationsstudien von sehr hohen Faktorladungshöhen zwischen .70 und .80 aus. Die empirische Realität in Hinblick auf Faktorladungshöhen sieht aber grundlegend anders aus. So konnte Peterson (2000) metaanalytisch bei Fragebögen eine durchschnittliche

Ladungshöhe von .32 ermitteln. Ein weiteres Beispiel ist der Ladungsrange beim NEO-FFI, der laut Borkenau und Ostendorf (1993) zwischen .30 und .60 liegt.

Schlussfolgernd lässt sich sagen, dass die von Hu und Bentler (1999) in ihrer Simulationsstudie angenommenen Ladungshöhen im Vergleich zu empirischen Daten sehr hoch angesetzt wurden.

Die Unterschiede zwischen aktuellen Studien und den Modellen von Hu und Bentler (1999) bezüglich der Modellkomplexitäten einerseits und der Ladungshöhen andererseits stellen die Verwendung der vorgeschlagenen Cutoff-Werte für die gängigen Fit-Indizes stark in Frage. Es muss schlussfolgernd aufgrund der geringen Sensitivität der Cutoff-Werte der Fit-Indizes davon ausgegangen werden, dass in bisherigen Studien eine hohe Anzahl von eigentlich misspezifizierten Modellen fälschlicherweise als richtig anerkannt wurde.

Um auf dieses Problem der fälschlichen Annahme eines Modells näher eingehen zu können, muss zuerst ein terminologischer Unterschied erklärt werden. Es wird zwischen folgenden zwei Begriffen unterschieden: Modellmisspezifikation und Modell-Missfit. Unter einer Modellmisspezifikation versteht man, dass bestimmte Parameter im Modell auf einen bestimmten Wert, meist null, gesetzt werden (z.B. Pfadkoeffizienten oder Ladungshöhen), obwohl diese in der Population bestehen. Diese Parameter werden somit fälschlicherweise auf null gesetzt, auch wenn sie in der Population ungleich null sind. Im Gegensatz dazu wird ein Modell-Missfit statistisch über die Differenz zwischen der beobachteten und der vom Modell implizierten Kovarianzmatrix berechnet (u.a. Hoyle, 1995).

Ein Ziel von Strukturgleichungsmodellen ist es, kausale Misspezifikationen zuverlässig zu finden. Das Problem dabei ist, dass z.B. die fälschliche Annahme eines Pfadkoeffizienten zwischen zwei latenten Variablen zwar inhaltlich eine entscheidende Misspezifikation darstellt, sich allerdings nicht zwingend eine große Abweichung zwischen der beobachteten und der implizierten Kovarianzmatrix ergibt (Hayduk, 2014). Somit würde eine schwerwiegende kausale Misspezifikation nicht detektiert werden.

Ein weiterer kritischer Punkt ist, dass die Fit-Indizes zur Überprüfung des Gesamtmodells, bestehend aus dem Struktur- und dem Messmodell, herangezogen werden. Dabei können die Schätzung der Parameter im Strukturmodell verzerrt

(cf. Kaplan, 1988; Yuan, Marshall & Bentler, 2003) und mögliche bestehende Missspezifikationen zwischen latenten Variablen dadurch verdeckt werden. McDonald und Ho (2002) überprüften insgesamt 14 Studien, in denen Strukturgleichungsmodelle angewandt worden waren. Im Rahmen dessen stellte sich heraus, dass das Messmodell im Vergleich zum Strukturmodell überwiegend den besseren Modellfit aufwies. Durch den guten Fit des Messmodells kann der schlechte Fit des Strukturmodells überdeckt werden. Auch Tomarken und Waller (2003) konnten zeigen, dass ein hinreichend gutes Messmodell den schlecht passenden Strukturteil überlagern kann. Darüber hinaus stellten Hancock und Mueller (2011) fest, dass auch bei einem perfekten Messmodell der Modellfit im Strukturmodell von der Qualität der Indikatoren im Messmodell abhängig ist. So zeigte sich bei einer schlechteren Qualität ein besserer Modellfit als bei einer besseren Qualität (Hancock und Mueller, 2011).

Da die wesentlichen und ausschlaggebenden inhaltlichen Hypothesen und die Wirkzusammenhänge im Strukturmodell abgebildet werden, sind Missspezifikationen in diesem Teil für die eigentlich zu testende Theorie allerdings als folgenschwerer zu bewerten als mögliche Missspezifikationen im Messmodell. Dies stellt die Validität bisheriger Forschung in Frage, da ein passendes Messmodell Missspezifikationen im Strukturmodell überdecken kann, was zu inhaltlich falschen Annahmen und somit auch zu falschen abgeleiteten Konsequenzen führen kann. Dies heißt jedoch nicht, dass das Messmodell für die Theorie irrelevant ist.

Man muss einräumen, dass es unter bestimmten Umständen auch kritisch sein kann, wenn ein Messmodell nicht passt. Beispielsweise ist das der Fall, wenn die Anzahl der Faktoren für die angenommene Theorie ausschlaggebend ist. Dennoch ist es erheblicher für Theorien, dass das Strukturmodell zuverlässig geprüft werden kann, da dort über Kausalzusammenhänge psychologische Theorien abgebildet werden. Dies führt zur Betonung der Wichtigkeit einer zuverlässigen Missspezifikationsdetektion im Strukturmodell.

Das Problem, wie bereits oben beschrieben, liegt nun aber in der Abhängigkeit der Cutoff-Werte von Indikatoren, wie z.B. Stichprobengröße oder Faktorladungshöhen, und der generellen Frage der Sinnhaftigkeit von fixierten Cutoff-Werten (u.a. Heene et al., 2011).

Insgesamt lässt sich sagen, dass die Sensitivität bzw. die Power zur Detektion einer Missspezifikation von Einflussgrößen wie der Modellgröße (cf. Brevik & Olsson, 2001; Kenny & McCoach, 2003; Savalei, 2012), der Faktorladungshöhe (cf. Browne, MacCallum, Kim, Andersen & Glaser, 2002; Hancock & Mueller, 2011; Heene et al., 2011; Schonemann, 1981b) und der Verletzung der Annahme einer multivariaten Normalverteilung (Hu & Bentler, 1999; Yuan, 2005) abhängig sind.

1.4 Fragestellung

Die zentrale Fragestellung der vorliegenden Arbeit war es, zu klären, wie die structural Fit-Indizes, die sich speziell auf das Strukturmodell fokussieren, unter systematisch variierenden Bedingungen auf graduelle Missspezifikationen reagieren. Bisherige Studien prüften das Gesamtmodell als solches. In der vorliegenden Arbeit wird neben dem exakten Modellfit und den composite Fit-Indizes, speziell die Zuverlässigkeit der structural Fit-Indizes in Hinblick auf die Missspezifikationsdetektion im Strukturmodell überprüft.

Als weiterer Ansatz werden auch differenzbasierte Fitmaße auf ihre Zuverlässigkeit hin überprüft, um mögliche Alternativen zu der gängigen Methode der Überprüfung mittels Fit-Indizes zu evaluieren.

2 Methode des simulierten Strukturgleichungsmodells

Die Simulationsstudie wurde mittels R (R Core Team, 2015) unter Verwendung des Paketes `simsem`, Version .5-11 (Pornprasertmanit, Miller & Schoemann, 2015) durchgeführt.

Da es um die Frage nach der generellen Zuverlässigkeit einer möglichen Missspezifikationsdetektion geht, war es essentiell, ein möglichst typisches Strukturgleichungsmodell, wie es auch in der angewandten Forschung häufig vorkommt, als Simulationsgrundlage heranzuziehen. Dafür wurde eine Vielzahl an bisher publizierten Strukturgleichungsmodellen aus unterschiedlichen Bereichen wie der klinischen Psychologie, dem Bildungsbereich oder dem differential- und organisationspsychologischen Bereich in der Literatur gesucht und auf Gemeinsamkeit geprüft.

Im Folgenden werden die Kriterien, die bei der Auswahl eines für die Empirie typischen Strukturgleichungsmodells relevant waren, kurz erörtert. Zum Ersten wurde die Größe bzw. Komplexität des Strukturgleichungsmodells definiert, indem geprüft wurde, aus wie vielen latenten und manifesten Variablen pro latenter Variable gängige Strukturgleichungsmodelle bestehen. Zum Zweiten wurde die Komplexität gängiger Strukturgleichungsmodelle definiert, indem typischerweise angenommene Zusammenhangsstrukturen der latenten Variablen im Strukturmodell aus der Literatur abgeleitet wurden. Zuletzt wurden Parametergrößen, d.h. Faktorkorrelationen, Faktorladungshöhen und die Pfadkoeffizienten, auf Basis empirischer Befunde und unter Berücksichtigung ihrer Aussagekraft für die anwendungsbezogene bzw. anwendungsorientierte Forschung definiert.

Nach eingehender Literaturrecherche erwies sich ein repräsentatives Strukturgleichungsmodell als Modell der Wahl, das im Folgenden beschrieben wird.

2.1 Das Populationsmodell

2.1.1 Das Strukturmodell

Bei dem in dieser Arbeit verwendeten Populationsmodell handelt es sich um ein latentes doppeltes Mediationsmodell, bestehend aus insgesamt sechs latenten Variablen, davon zwei Mediationsfaktoren, drei exogenen Variablen sowie einer endogenen Variablen. Das vollständige Strukturmodell des Populationsmodells ist in Abbildung 2 ersichtlich. Bei der Annahme der standardisierten Regressionsgewichte dienten Peterson und Brown (2005) als Orientierung. Peterson und Brown (2005) untersuchten die Verteilung von insgesamt 1.504 standardisierten Regressionsgewichten, veröffentlicht in sozialwissenschaftlichen Zeitschriften. Dabei stellte sich heraus, dass rund 97 Prozent einen Range von $\pm .50$ Einheiten ($M = .06$, $SD = .21$) aufwiesen. Die standardisierten Regressionsgewichte wurden in der vorliegenden Simulation als hohe Werte angenommen, um auch theoretisch nicht triviale Missspezifikationen zu operationalisieren, und lagen zwischen $.30$ und $.57$. Die Missspezifikation schon eines einzelnen Pfades sollte dadurch bereits eine inhaltlich relevante bzw. nicht triviale Abweichung darstellen.

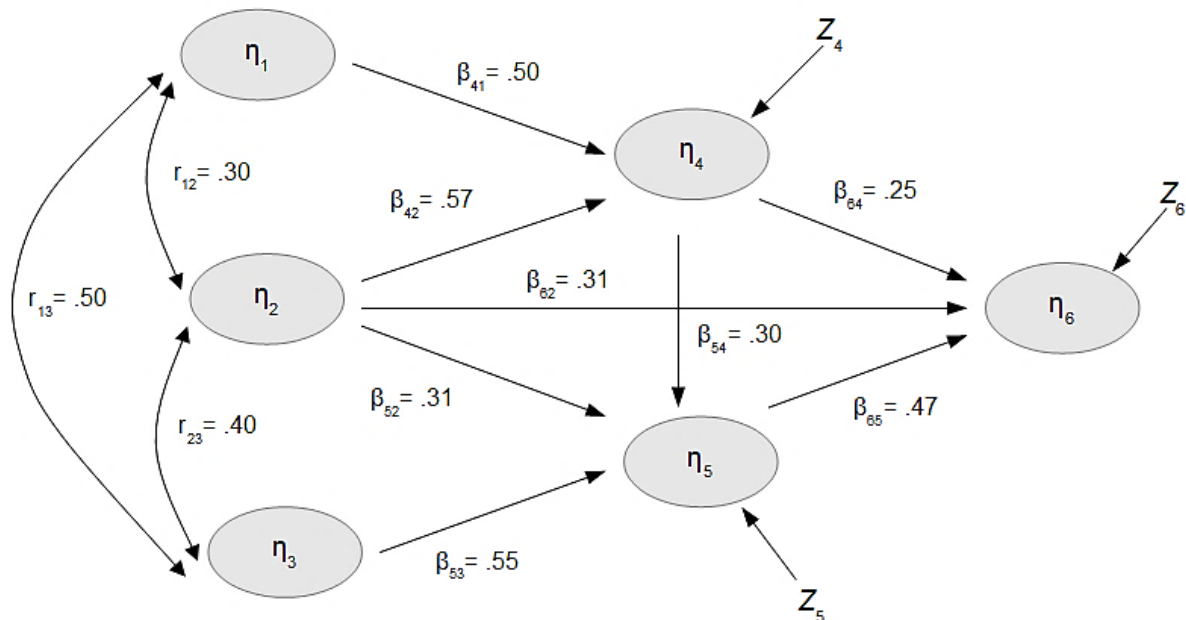
Wie in Abbildung 2 ersichtlich, wurden die standardisierten Regressionsgewichte in Anlehnung an Peterson und Brown (2005) wie folgt postuliert::

$\beta_{41} = .50$, $\beta_{42} = .57$, $\beta_{62} = .31$, $\beta_{52} = .31$, $\beta_{53} = .55$, $\beta_{64} = .25$, $\beta_{54} = .30$ und $\beta_{65} = .47$.

Die Faktorkorrelationen zwischen den drei exogenen Variablen, vergleichbar mit den Annahmen von Hu und Bentler (1999), wurden wie folgt spezifiziert: $r_{12} = .30$, $r_{23} = .40$ und $r_{13} = .50$.

Abbildung 2

Das Strukturmodell des Populationsmodells



2.1.2 Das Populationsmessmodell

Obwohl der Fokus auf der Missspezifikationsdetektion im Strukturmodell liegt, wurde auch das Ausmaß korrelierter Fehler im Messmodell variiert, um realistische Bedingungen zu erhalten, da korrelierte Fehler in der Empirie sehr häufig vorkommen. So konnten Cole, Ciesla und Steiger (2007) zeigen, dass 26.6 % bis 31.9 % von SEMs in renommierten Journalen korrelierte Fehler zulassen.

Daher wurde einerseits ein Messmodell ohne korrelierte Fehler spezifiziert, und andererseits wurden korrelierte Fehler zwischen den beobachteten Variablen in drei weiteren Populationsmodellen spezifiziert. Die paarweisen Messfehlerkorrelationen wurden aus einer Gleichverteilung in folgenden Intervallen gezogen: $\epsilon_{\rho_{ij}} \in [-.1, .1]$, $\epsilon_{\rho_{ij}} \in [-.2, .2]$ und $\epsilon_{\rho_{ij}} \in [-.3, .3]$, wobei $i \neq j$, und per Zufallsgenerator den Items der jeweils sechs Faktoren zugewiesen.

Da auch bekannt ist, dass die Modellkomplexität einen entscheidenden Einfluss auf die Sensitivität der Missspezifikationsdetektion hat (Breivik & Olsson, 2001; Heene et al., 2011; Kenny & McCoach, 2003), wurde auch die Testlänge variiert. So wurden in der Langversion die latenten Variablen mit je zehn manifesten Variablen gemessen, in der kurzen Version mit je fünf manifesten Variablen. Die Daten der manifesten Variablen wurden aus einer multivariaten Normalverteilung generiert. Obwohl Daten basierend auf der multivariaten Normalverteilung in den Sozialwissenschaften selten vorkommen (Micceri, 1989), wurden diese in der durchgeführten Simulation gewählt, um die Vergleichbarkeit mit anderen Simulationsstudien zu gewährleisten, und zwar unter Verwendung der Maximum-Likelihood-Schätzmethode (cf. Browne et al., 2002; Curran, Bollen, Paxton, Kirby & Chen, 2002; Hu & Bentler, 1999; Marsh, Balla & McDonald, 1988).

Auch die Faktorladungshöhe hat einen bedeutenden Einfluss auf die Sensitivität, Missspezifikationen zu entdecken (Heene et al., 2011; Schonemann, 1981), und wurde daher in der vorliegenden Simulation ebenfalls systematisch variiert. Hierbei wurden zwei Bedingungen unterschieden: zum einen hohe Faktorladungen mittels Pseudo-Zufallszahlenziehung aus einer Gleichverteilung in einem Intervall $\lambda_{\text{hoch}} \in [.6, .8]$, und zum anderen niedrige Faktorladungen mittels Pseudo-Zufallszahlen aus einer Gleichverteilung in einem Intervall $\lambda_{\text{niedrig}} \in [.4, .6]$.

Abschließend, unabhängig von den Annahmen des Mess- oder Strukturmodells, wurden unterschiedliche Stichprobengrößen definiert, da die Stichprobengröße einen Einfluss auf die Sensitivität, Missspezifikationen zu entdecken, hat (Chen, Curran, Bollen, Kirby & Paxton, 2008). Insgesamt wurden drei Stichprobengrößen ($N = 150$, $N = 250$ und $N = 1000$) bestimmt.

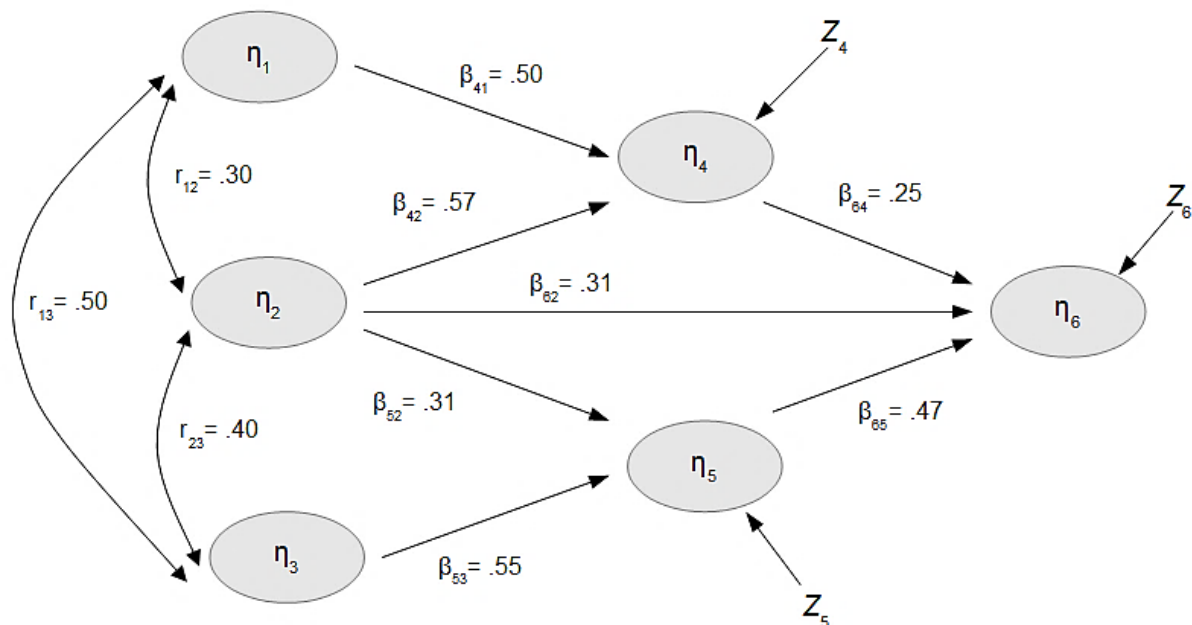
2.2 Die Missspezifikationen im Struktur- und Messmodell

Das Ausmaß der Missspezifikation im Strukturmodell wurde in drei unterschiedlichen Stufen simuliert, um einen graduellen Anstieg der Missspezifikation darzustellen. Dabei wurden im Strukturmodell sukzessive ein bis drei Pfade auf null gesetzt.

In Abbildung 3 erkennt man, dass in der leichtesten Ausprägung der Missspezifikation der Pfad β_{54} auf null gesetzt wurde.

Abbildung 3

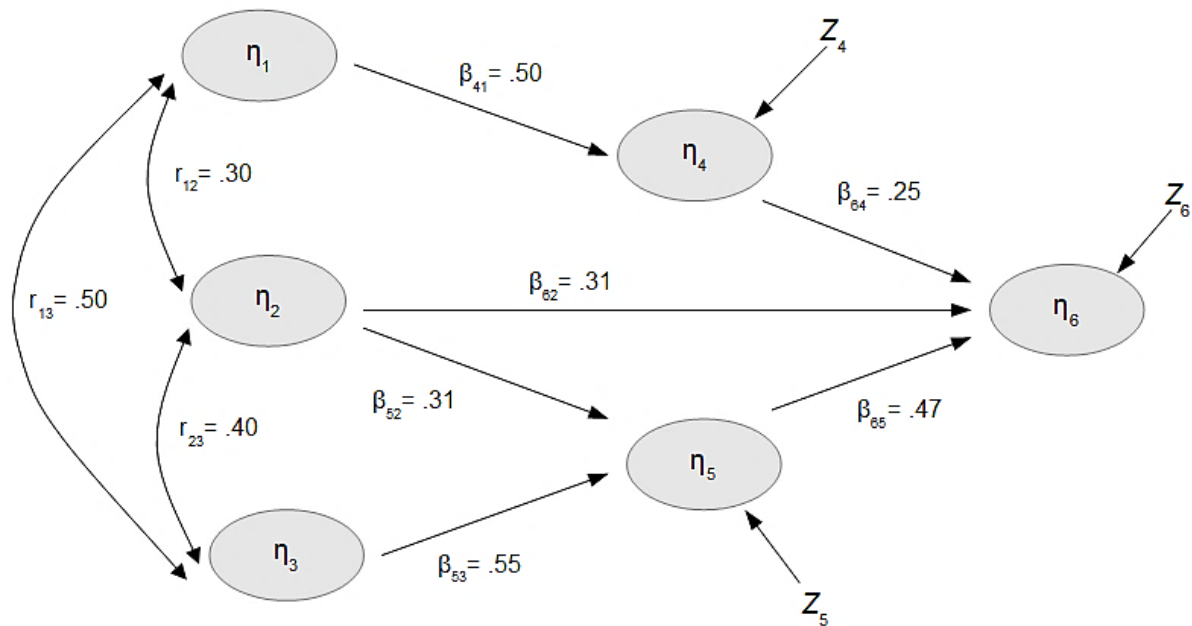
Leichte Missspezifikationausprägung; Pfad β_{54} auf null gesetzt



Bei der moderaten Ausprägung der Missspezifikation, dargestellt in Abbildung 4, sieht man, dass nun zusätzlich der Pfad β_{42} auf null gesetzt wurde. Somit sind bei der moderaten Ausprägung der Missspezifikation bereits zwei relevante Pfade auf null gesetzt.

Abbildung 4

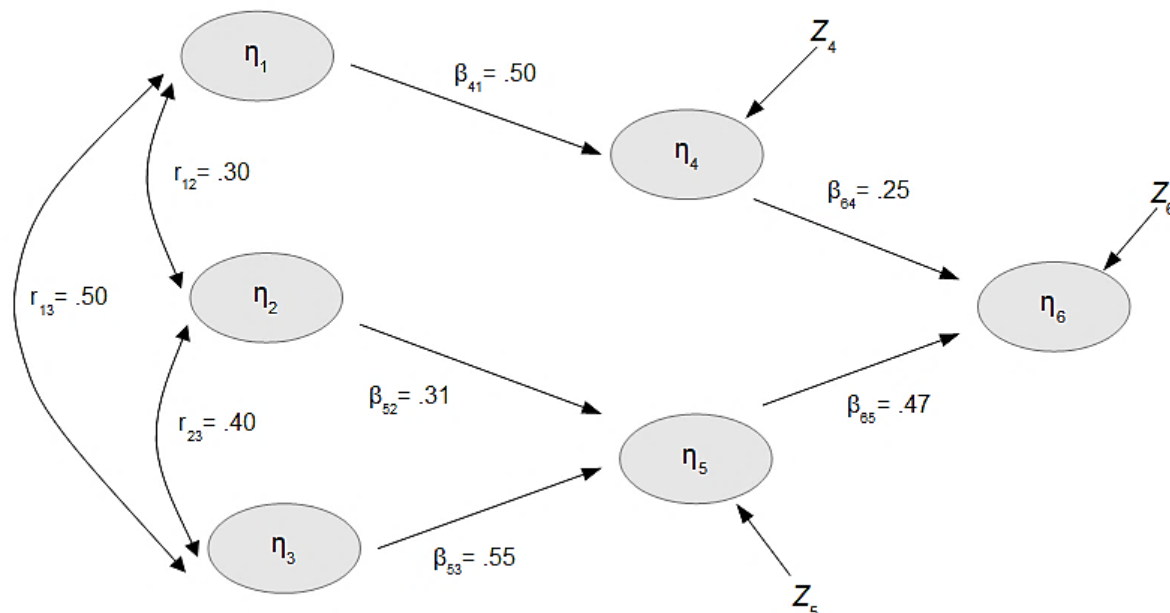
Moderate Missspezifikationausprägung; zusätzlich Pfad β_{42} auf null gesetzt



Als schwerste und somit deutlichste Missspezifikation wurde zusätzlich der Pfad β_{62} auf null gesetzt. Dieses sehr schwerwiegend misspezifizierte Modell ist in Abbildung 5 ersichtlich. Durch das Nullsetzen von insgesamt drei Pfaden weicht dieses misspezifizierte Modell maßgeblich von den inhaltlichen Annahmen des Populationsmodells ab.

Abbildung 5

Schwerwiegende Missspezifikationausprägung; zusätzlich Pfad β_{62} auf null gesetzt



Um Missspezifikationen im Messmodell zu definieren, wurden Fehlerkorrelationen zwischen den beobachteten Variablen auf null gesetzt. Insgesamt wurden vier Bedingungen der Messfehler-Misspezifikation definiert. In der ersten Bedingung wurden alle Fehlerkorrelationen im Messmodell auf null gesetzt. In den Bedingungen zwei bis vier wurden Fehlerkorrelationen zwischen den fünf bzw. zehn manifesten Variablen im Populationsmodell spezifiziert, welche zufällig aus Gleichverteilungen in folgenden Intervallen gezogen wurden: $\epsilon_{pij} \in [-.1, .1]$, $\epsilon_{pij} \in [-.2, .2]$ oder $\epsilon_{pij} \in [-.3, .3]$, wobei $i \neq j$. Bei der jeweiligen Misspezifikationsbedingung des Messmodells wurden diese dann auf null gesetzt.

Somit ergab sich ein vollständig gekreuztes 4 (Anzahl der Misspezifikationen im Messmodell) x 3 (Anzahl der Misspezifikationen im Strukturmodell) x 2 (Faktorladungshöhen) x 2 (Testlängen) x 3 (Stichprobengrößen) Simulationsdesign. Für die daraus resultierenden 144 Bedingungen wurden pro Bedingung 1000 Stichprobenreplikationen durchgeführt. Bei kleinen Stichproben und niedrigen Faktorladungshöhen sind sogenannte „improper solutions“ (nicht-konvergierende Lösungen) zu erwarten (Driel, 1978). Laut Chen et al. (2001) sind improper solutions Parameterschätzungen, die Werte angenommen haben, die außerhalb des zulässigen Parameterraums liegen. Häufig sind in diesen Fällen negative Fehlervarianzen die

Ursache für improper solutions. Problematisch hieran ist, dass sich durch improper solutions die Anzahl an analysierbaren Replikationen vermindert. Um dennoch jeweils 1000 Replikationen je Bedingung zu erhalten, wurden so lange weitere Daten simuliert, bis die geforderte Anzahl von 1000 Replikationen je Bedingung erreicht war.

Zusammenfassend kann man sagen, dass die Ausprägung der Missspezifikationen sowohl im Struktur- als auch im Messmodell systematisch variiert wurde. Für die Auswertung wurden auf der einen Seite die correct-rejection-rates sowie die Power berechnet, auf der anderen Seite die wrong-rejection-rates der Fit-Indizes und die Alpha-Fehlerraten des Chi-Quadrat-Tests.

Dabei wurden die 1000 Replikationen pro Bedingung ausgezählt, um zu ermitteln, wie oft die missspezifizierten Modelle unter Verwendung der entsprechenden Cutoff-Werte für die composite und structural Fit-Indizes abgelehnt wurden. So ergaben sich für die Fit-Indizes die correct-rejection rates. Ebenso war das Vorgehen beim exakten Modell-Test, um die Power des Chi-Quadrat-Tests zu berechnen. Weiters wurde für die passenden Modelle ausgezählt, wie oft diese fälschlicherweise abgelehnt wurden. Dadurch wurden die wrong-rejection-rates der Fit-Indizes und die Alpha-Fehlerraten des Chi-Quadrat-Tests berechnet.

Das Ziel ist, dass sich für die Fit-Indizes hohe correct-rejection-rates ergeben, da dies eine hohe Sensitivität, Missspezifikationen zu detektieren, ausdrückt. Ebenso soll die Power des Chi-Quadrat-Tests möglichst hoch sein, da nur dann Missspezifikationen aufgezeigt werden. Für die Beurteilung statistischer Eigenschaften eines Fit-Index bzw. eines inferenzstatistischen Verfahrens wie des Chi-Quadrat-Tests ist es ebenso notwendig zu wissen, wie oft richtige Modelle als falsch abgelehnt werden.

Die Wahrscheinlichkeit, dass missspezifizierte Modelle abgelehnt werden (d.h. die statistische Power), sollte in Übereinstimmung mit Cohen (1992) bei mindestens 80% liegen.

Tabelle 2
Übersicht über die Bedingungen

Art der Bedingung	Ausprägung
Fehlerkorrelationen im Messmodell	keine Fehlerkorrelation; $\epsilon \in [-.1, .1]$, $\epsilon \in [-.2, .2]$ $\epsilon \in [-.3, .3]$
Missspezifikationen im Strukturmodell	leicht (ein Pfad auf null gesetzt) moderat (zwei Pfade auf null gesetzt) schwerwiegend (drei Pfade auf null gesetzt)
Testlänge	fünf oder zehn Items pro latenter Variable
Faktorladungen	hoch $\lambda_{\text{hoch}} \in [.6, .8]$ und niedrig $\lambda_{\text{niedrig}} \in [.4, .6]$
Stichprobengröße	$N = 150$, $N = 250$, $N = 1000$

3 Ergebnisse des exakten und approximativen Modellfits

Bei der Ausführung der Resultate wird zuerst auf die Ergebnisse des exakten sowie des approximativen Modellfits eingegangen. Dabei wurden sowohl die vollkommen korrekt spezifizierten Modelle, als auch die Modelle mit Missspezifikationen im Strukturmodell ausgewertet. Zuerst werden die Ergebnisse des Chi-Quadrat-Tests erörtert, im Anschluss die Ergebnisse der composite Fit-Indizes und danach die Ergebnisse der structural Fit-Indizes. Abschließend werden die Ergebnisse unter Berücksichtigung der Missspezifikationen sowohl im Mess- als auch im Strukturmodell beschrieben.

3.1 Ergebnisse des Modellfits zum Modell ohne Fehlerkorrelationen im Messmodell

3.1.1 Die Ergebnisse des Chi-Quadrat-Tests – der exakte Modellfit

Alpha-Fehlerraten

Wie in Tabelle 3 ersichtlich, unterscheidet sich bei kleinen Stichproben ($N < 1000$) der beobachtete Alpha-Fehler mit z.B. .999 (unter der Bedingung niedrige Faktorladungen, zehn Items und $N = 150$) erheblich vom nominellen Alpha-Niveau von 5%. Der Alpha-Fehler wird in Spalte „0“ (d.h. keine Missspezifikation vorhanden) dargestellt. Die gewünschten 5% werden durchgehend nicht eingehalten. Die einzige Annäherung an das gewünschte Alpha-Niveau findet man mit .078 unter der Bedingung niedrige Faktorladungen, fünf Items und $N = 1000$. Dies bedeutet schlussfolgernd, dass das richtig spezifizierte Modell fälschlicherweise generell zu oft abgelehnt wird, wobei diesbezüglich offensichtlich erhebliche Unterschiede zwischen den Bedingungen bestehen. Der Effekt einer extrem erhöhten Alpha-Fehlerrate bei kleinen Stichproben lässt sich erklären, wenn man sich Mittelwert und Standardabweichung der Chi-Quadrat-Statistik in Abhängigkeit von der Stichprobengröße ansieht. Der Erwartungswert E einer zentral chiquadratverteilten Zufallsvariablen X ist $E(X) = df$ und ihre Standardabweichung $\sigma(X) = \sqrt{2 \cdot df}$. Die Werte für Mittelwert und Standardabweichung weichen aber deutlich von den theoretisch zu erwartenden Werten ab. Man sieht, dass hier ein positiver Bias vorliegt. Das weist auf eine nicht hinreichende Approximation der zentralen Chi-Quadrat-Verteilung bei komplexen Strukturgleichungsmodellen und gleichzeitig kleinen

Stichproben hin. Bentler und Yuan (1999) konnten ähnliche Ergebnisse aufzeigen. Dieser Effekt wird unter der Bedingung zehn Items pro latenter Variable noch verstärkt und führt zu inakzeptablen Alpha-Fehlerraten. Dies kann sogar unter der größten Stichprobenbedingung von $N = 1000$ beobachtet werden. Dieser Befund steht in Übereinstimmung mit Moshagen (2012), der zeigen konnte, dass dieser Bias-Effekt bei kleinen Stichproben in Abhängigkeit von der Größe der Kovarianzmatrix der manifesten Variablen zunimmt.

Powerraten

Wie zu erwarten und in Tabelle 3 erkennbar, nimmt die Power mit steigendem Ausmaß der Missspezifikation zu. In dieser Simulation heißt das also, dass die Power umso höher ist, je mehr Pfade fälschlicherweise (d.h. misspezifiziert) auf null gesetzt werden. Die Ergebnisse zeigen klar, dass die dritte und somit die am stärksten ausgeprägte Missspezifikation im Strukturmodell durchgehend angezeigt werden kann. Die Powerraten liegen hierbei zwischen 1 und .982. Die einzigen Ausnahmen bei der stärksten Missspezifikationsausprägung bilden die Powerraten unter der Bedingung, dass fünf Items pro latenter Variable, niedrige Faktorladungen und eine Stichprobengröße von $N \leq 250$ vorliegen. Hier sind die Powerraten .704 (unter der Bedingung $N = 150$, fünf Items und niedrige Faktorladungen) und .778 (unter der Bedingung $N = 250$, fünf Items und niedrige Faktorladungen). Es zeigt sich klar der starke Einfluss der Faktorladungshöhen auf die Zuverlässigkeit der Missspezifikationsdetektion im Strukturmodell.

Generell lässt sich weiters festhalten, dass die Power mit steigender Anzahl an manifesten Variablen ansteigt (siehe auch Saris & Satorra, 1988 für ähnliche Befunde). Bei der kürzeren Testlänge (fünf manifeste Variablen pro latenter Variable) scheint darüber hinaus die Faktorladung in Kombination mit einer großen Stichprobe entscheidend zu sein. Dies impliziert, dass die Power nur dann ausreichend ist, wenn für die Daten hohe Faktorladungen vorliegen. Das steht in Übereinstimmung mit den Befunden von Schonemann (1981b) sowie Heene, Hilbert, Draxler, Ziegler und Bühner (2011).

Tabelle 3
Alpha-Fehlerraten und Powerraten des Chi-Quadrat-Tests

N	Item- anzahl	Faktor- ladungs- höhe	Anzahl der Missspezifikationen im Strukturmodell			
			0	1	2	3
150	5	niedrig	.331 (43.314;31.093)	.378 (434.388;3.789)	.400 (437.006;3.613)	.704 (462.255;32.830)
		hoch	.355 (431.129;31.278)	.473 (442.033;31.761)	.576 (45.170;32.132)	.982 (516.198;34.649)
	10	niedrig	.999 (2027.386;71.380)	.999 (2034.388;71.409)	.999 (2039.395;71.363)	1 (2091.314;72.625)
		hoch	.998 (2026.943;71.077)	1 (2046.817;72.087)	1 (2061.653;71.955)	1 (215.667;73.034)
250	5	niedrig	.194 (414.896;3.023)	.229 (42.563;29.932)	.267 (424.449;3.285)	.778 (469.218;33.007)
		hoch	.194 (415.402;29.825)	.377 (432.539;31.044)	.509 (445.615;32.041)	1 (555.860;37.283)
	10	niedrig	.905 (1875.671;63.029)	.936 (1887.790;63.045)	.947 (1895.884;63.388)	.998 (1981.581;65.961)
		hoch	.903 (1876.805;63.190)	.963 (1909.898;64.033)	.987 (1933.931;65.104)	1 (2076.913;68.168)
1000	5	niedrig	.078 (4.784;28.508)	.216 (419.164;29.646)	.365 (433.179;3.661)	1 (61.227;41.457)
		hoch	.083 (401.060;28.437)	.771 (468.132;32.912)	.986 (519.477;36.942)	1 (953.811;55.151)
	10	niedrig	.173 (174.999;58.061)	.415 (1784.095;6.224)	.594 (1814.526;6.698)	1 (2153.156;71.288)
		hoch	.175 (1741.126;58.648)	.889 (1869.571;62.825)	.995 (1961.893;66.786)	1 (2547.044;79.068)

Ein weiteres interessantes Ergebnis ist, dass die Power bei fünf Items pro latenter Variable bei leichter und bei moderater Missspezifikation nicht ausreichend ist, *unabhängig* von der Faktorladungshöhe und der Stichprobengröße. Die einzige Ausnahme findet sich unter fünf Items, hohen Faktorladungen und $N = 1000$ mit einer Powerrate von .986. Die Stichprobengröße scheint bei der leichten bis moderaten Form der Missspezifikationsbedingung wenig bis keinen Einfluss auf die Powerraten zu haben. Als einheitliches Bild zeigt sich bei der leichten bis moderaten Missspezifikationsbedingung, dass die Faktorladungen einen entscheidenden Einfluss auf

die Powerraten haben. Dies sieht man vor allem unter der Bedingung, dass zehn Items und niedrige Faktorladungen bei einer Stichprobengröße von $N = 1000$ vorliegen. Hier kann die erste und zweite Ausprägung der Missspezifikation im Strukturmodell nicht hinreichend detektiert werden. So liegen die Powerraten für die leichte Missspezifikationsausprägung unter dieser Bedingung bei .415 und bei der moderaten Missspezifikationsausprägung bei .594.

Darüber hinaus kann gegenintuitiv beobachtet werden, dass die Power innerhalb der leichten und moderaten Missspezifikationsbedingung unter der Bedingung $N = 250$ *geringer* ist als unter der Bedingung $N = 150$. Dies scheint gegen die statistische Theorie zu stehen, wonach die Power bei gleicher Missspezifikation bei größerem N größer ausfallen sollte. Allerdings lässt sich dieses Ergebnis durch die Ergebnisse von Curran, Bollen, Paxton, Kirby und Chen (2002) erklären. Diese Autoren zeigten, dass bei kleinen gegenüber großen Stichproben und einem speziellen Typ von Missspezifikation die nonzentrale Chi-Quadrat-Verteilung nicht hinreichend approximiert werden kann und der Mittelwert und die Varianz dieser Verteilung eine starke positive Verzerrung aufweisen. Der Mittelwert der nonzentralen Chi-Quadrat-Verteilung liegt somit bei kleinen Stichproben höher als bei großen, was zu einer künstlichen Erhöhung der Power bei kleinen Stichproben führt, da der Nonzentralitätsparameter folglich höher liegt. Ein weiterer Hinweis dafür ist, dass bei der leichten und der moderaten Missspezifikationsausprägung der Mittelwert und die Standardabweichung unter der Bedingung $N = 150$ größer sind als unter der Bedingung $N = 250$. Dies impliziert, dass bei kleinen Stichproben, im vorliegenden Fall bei $N = 150$, keine hinreichende Approximation der nonzentralen Chi-Quadrat-Verteilung gewährleistet ist.

Zusammengefasst lässt sich sagen, dass die Powerraten bei leichter bis moderater Missspezifikation nur dann als hinreichend bezeichnet werden können, wenn die Daten über eine ausreichende Anzahl an manifesten Variablen in Kombination mit hohen Faktorladungen erhoben wurden. Der Einfluss der Faktorladungen zeigt sich auch bei der dritten und folglich stärksten Ausprägung der Missspezifikation im Strukturmodell. So werden Missspezifikationen im Strukturmodell zuverlässiger angezeigt, wenn hohe Faktorladungen vorliegen. Der Einfluss der Stichprobe scheint über alle Bedingungen hinweg nicht ausschlaggebend zu sein.

3.1.2 Die Ergebnisse der Fit-Indizes – der approximative Modellfit

3.1.2.1 Die composite Fit-Indizes

Ergebnisse des RMSEA < .05

Wie in Tabelle 4 ersichtlich, kann der RMSEA bei einem Cutoff-Wert von < .05 unter keiner Bedingung die Missspezifikation im Strukturmodell zuverlässig anzeigen. Die correct rejection rates liegen bei allen Bedingungen zwischen 0 und .003. Auch wenn es sich bei den Fit-Indizes um keine Teststatistik und folglich nicht um Signifikanztests handelt, lehnen wir uns bei der Interpretation der Ergebnisse an die Inferenzstatistik an und würden eine correct rejection rate von mindestens .80 (im Sinne einer Power von $\geq .80$) als hinreichend bezeichnen (Cohen, 1992). Diese wird beim RMSEA bei einem Cutoff-Wert von .05 unter keiner Bedingung erreicht.

Tabelle 4

Correct rejection rates und wrong rejection rates des RMSEA

N	Item- anzahl	Faktor- ladungs- höhe	Anzahl der Missspezifikationen im Strukturmodell			
			0	1	2	3
150	5	niedrig	0 (.022;.012)	0 (.024;.011)	0 (.024;.011)	0 (.032;.01)
	5	hoch	0 (.023;.011)	0 (.026;.011)	0 (.029;.01)	.003 (.044;.007)
	10	niedrig	0 (.036;.004)	0 (.036;.004)	0 (.036;.004)	0 (.039;.004)
	10	hoch	0 (.036;.004)	0 (.037;.004)	0 (.037;.004)	0 (.042;.003)
250	5	niedrig	0 (.013;.009)	0 (.014;.009)	0 (.015;.009)	0 (.026;.007)
	5	hoch	0 (.013;.009)	0 (.018;.009)	0 (.021;.009)	0 (.04;.005)
	10	niedrig	0 (.020;.004)	0 (.021;.004)	0 (.021;.004)	0 (.025;.003)
	10	hoch	0 (.020;.004)	0 (.022;.003)	0 (.023;.003)	0 (.03;.003)
1000	5	niedrig	0 (.004;.004)	0 (.007;.005)	0 (.009;.005)	0 (.023;.002)
	5	hoch	0 (.004;.004)	0 (.013;.003)	0 (.017;.003)	0 (.037;.002)
	10	niedrig	0 (.004;.003)	0 (.007;.003)	0 (.008;.002)	0 (.016;.001)
	10	hoch	0 (.004;.003)	0 (.010;.002)	0 (.012;.002)	0 (.022;.001)

Die Betrachtung der wrong rejection rates in Spalte „0“, die in Anlehnung an die Inferenzstatistik bei rund 5% (im Sinne eines Alpha-Fehlers) liegen sollen, zeigt klar, dass diese 5% unter keiner Bedingung eingehalten werden. Die wrong rejection rates des RMSEA liegen durchgehend bei 0. Generell muss man aber anmerken, dass im Vergleich zu den äußerst unzufriedenstellenden correct rejection rates die wrong rejection rates nicht derart ausschlaggebend sind.

Ergebnisse des SRMR < .08

Wie in Tabelle 5 erkennbar, zeigen die Ergebnisse des SRMR klar, dass die simulierte Missspezifikation im Strukturmodell unter den beiden ersten Ausprägungen (ein Pfad bzw. zwei Pfade auf null gesetzt) unter keiner Bedingung detektiert wird. Unter der leichten bis moderaten Missspezifikationsausprägung liegen die correct rejection rates durchgehend bei 0.

Die Ergebnisse unter der schwersten Ausprägung der Missspezifikation zeigen deutlich den Einfluss der Faktorladungen auf die correct rejection rates. So kann die dritte Ausprägung der Missspezifikation bei hohen Faktorladungen eher detektiert werden. Die correct rejection rates liegen zwischen .982 (unter der Bedingung $N = 1000$, zehn Items und hohe Faktorladungen) und .853 (unter der Bedingung $N = 150$, fünf Items und hohe Faktorladungen). Die einzigen Ausnahmen finden sich unter der Bedingung $N = 250$, fünf Items und hohe Faktorladungen mit einer correct rejection rate von .746 sowie unter $N = 1000$, fünf Items und hohe Faktorladungen mit einer correct rejection rate von .511. Unter dieser Bedingung zeigt sich auch, dass die correct rejection rates mit steigender Stichprobe sinken. Dies ist zwar unerwartet, steht aber in Übereinstimmung mit Heene et al. (2011) und Marsh et al. (2004).

Die wrong rejection rates in Spalte „0“ sind vergleichbar mit denen des RMSEA und liegen unter allen Bedingungen bei 0.

Tabelle 5
Correct rejection rates und wrong rejection rates des SRMR

N	Item- anzahl	Faktor- ladungs- höhe	Anzahl der Missspezifikationen im Strukturmodell			
			0	1	2	3
150	5	niedrig	0 (.060;.003)	0 (.061;.003)	0 (.061;.003)	.020 (.068;.005)
	5	hoch	0 (.048;.003)	0 (.050;.004)	0 (.052;.004)	.853 (.092;.011)
	10	niedrig	0 (.065;.002)	0 (.066;.002)	0 (.066;.002)	.155 (.075;.004)
	10	hoch	0 (.053;.003)	0 (.056;.003)	0 (.057;.003)	.986 (.102;.010)
250	5	niedrig	0 (.046;.002)	0 (.047;.002)	0 (.047;.002)	0 (.056;.005)
	5	hoch	0 (.037;.002)	0 (.040;.003)	0 (.041;.003)	.746 (.087;.01)
	10	niedrig	0 (.050;.001)	0 (.051;.001)	0 (.051;.001)	0 (.063;.004)
	10	hoch	0 (.041;.002)	0 (.044;.002)	0 (.045;.002)	.963 (.094;.008)
1000	5	niedrig	0 (.023;.001)	0 (.024;.001)	0 (.025;.001)	0 (.038;.003)
	5	hoch	0 (.019;.001)	0 (.024;.002)	0 (.026;.002)	.511 (.080;.005)
	10	niedrig	0 (.025;.001)	0 (.026;.001)	0 (.027;.001)	0 (.045;.003)
	10	hoch	0 (.020;.001)	0 (.026;.001)	0 (.028;.001)	.982 (.089;.004)

Ergebnisse des CFI > .95

Die Ergebnisse des auf dem Chi-Quadrat basierenden Comparative Fit Index erscheinen zuerst unerwartet durch das *Absinken* der correct rejection rates mit *steigender* Stichprobengröße. Dieser Effekt, sichtbar in Tabelle 6, zeigt sich über alle Missspezifikationsbedingungen hinweg und erklärt sich dadurch, dass die nonzentrale Chi-Quadrat-Verteilung bei kleinen Stichproben nicht ausreichend gut approximiert wird und einen positiven Bias des Erwartungswerts aufweist. Von daher beeinflussen nicht hinreichende Approximationen der nonzentralen Chi-Quadrat-Verteilung auch die Eigenschaften des CFI.

Schaut man sich die Ergebnisse unter der größten Stichprobengröße und folglich der Annahme einer hinreichend guten Approximation der nonzentralen Chi-Quadrat-Verteilung an, so sieht man, dass unabhängig von der Stärke der Missspezifikation, der Faktorladungshöhe und der Testlänge keine Ausprägung der Missspezifikation im Strukturmodell hinreichend detektiert werden kann. Die correct rejection rates liegen

unter der Bedingung $N = 1000$ über alle Bedingungen hinweg zwischen 0 und .128. Somit lässt sich sagen, dass der CFI auch bei großen Stichproben nicht in der Lage ist, zuverlässig Missspezifikation im Strukturmodell aufzuzeigen.

Tabelle 6
Correct rejection rates und wrong rejection rates des CFI

N	Item- anzahl	Faktor- ladungs- höhe	Anzahl der Missspezifikationen im Strukturmodell			
			0	1	2	3
150	5	niedrig	.415 (.954;.034)	.466 (.951;.034)	.496 (.949;.034)	.786 (.920;.039)
	5	hoch	.011 (.983;.013)	.021 (.979;.013)	.039 (.976;.014)	.585 (.947;.015)
	10	niedrig	.999 (.845;.033)	.999 (.841;.033)	.999 (.840;.033)	1 (.815;.033)
	10	hoch	.834 (.936;.014)	.886 (.933;.014)	.925 (.930;.014)	.996 (.913;.014)
250	5	niedrig	.074 (.981;.018)	.094 (.979;.019)	.106 (.977;.019)	.568 (.945;.024)
	5	hoch	0 (.993;.007)	0 (.989;.008)	0 (.986;.008)	.244 (.957;.010)
	10	niedrig	.609 (.944;.020)	.677 (.94;.020)	.729 (.938;.020)	.981 (.911;.020)
	10	hoch	.002 (.978;.008)	.003 (.974;.008)	.005 (.972;.008)	.347 (.953;.008)
1000	5	niedrig	0 (.997;.004)	0 (.995;.005)	0 (.993;.005)	.128 (.959;.008)
	5	hoch	0 (.999;.001)	0 (.995;.002)	0 (.992;.003)	0 (.962;.004)
	10	niedrig	0 (.996;.004)	0 (.993;.005)	0 (.990;.005)	.017 (.962;.006)
	10	hoch	0 (.998;.001)	0 (.995;.002)	0 (.992;.002)	0 (.974;.002)

Die wrong rejection rates in Spalte „0“ nähern sich vor allem bei einer großen Stichprobe und einer kürzeren Testlänge dem nominellen Alpha-Fehler von 5% an. Auffallend schlechte wrong rejection rates finden sich bei einer kleinen Stichprobe ($N = 150$) und zehn Items. Hier liegen die wrong rejection rates bei .834 für die hohe Faktorladungsbedingung und bei .999 für die niedrige Faktorladungsbedingung.

Zusammenfassend kann man sagen, dass die Ergebnisse der composite Fit-Indizes unter Verwendung der gängigen Cutoff-Werte klar zeigen, dass mögliche Missspezifikationen nicht zuverlässig detektiert werden können. Der RMSEA ist unter keiner Bedingung in der Lage, eine Missspezifikation anzuzeigen, ebenso der CFI unter der Berücksichtigung einer hinreichenden Approximation der nonzentralen Chi-Quadrat-Verteilung bei einer Stichprobengröße von $N \geq 1000$. Bei den Ergebnissen des CFI ist deutlich erkennbar, dass

die correct rejection rates bei steigender Stichprobe abnehmen, was auf eine nicht hinreichende Approximation der nonzentralen Chi-Quadrat-Verteilung hinweist. Allein der SRMR ist unter der Bedingung, dass hohe Faktorladungen vorliegen, in der Lage, die stärkste Ausprägung der Missspezifikation anzuzeigen. Bei den Ergebnissen des SRMR zeigt sich deutlich der Einfluss der Faktorladungen und zum Teil auch der Einfluss der Testlänge. So wird die dritte Ausprägung der Missspezifikation vor allem dann erkannt, wenn hohe Faktorladungen und zehn Items pro latenter Variable vorliegen, wobei der Einfluss der Faktorladungen wesentlich stärker zu bewerten ist. Insgesamt muss betont werden, dass sich die composite Fit-Indizes nicht eignen, um Missspezifikation im Strukturmodell zuverlässig aufzuzeigen.

3.1.2.2 Die structural Fit-Indizes

Ergebnisse des RMSEA-P < .05

Tabelle 7 zeigt, dass eine akzeptable correct rejection rate von $\geq .80$ bei der stärksten Ausprägung der Missspezifikation, bei der alle drei Pfade auf null gesetzt wurden, durchgehend über alle Bedingungen erreicht wird. Die correct rejection rates liegen hierbei zwischen .995 und 1. Allgemein heißt dies, dass sich die correct rejection rate erwartungsgemäß im steigenden Ausmaß der Missspezifikation erhöht.

Auch die zweite Ausprägung der Missspezifikation wird durchwegs erkannt, die einzige Ausnahme bilden dabei die Bedingungen niedrige Faktorladungen und fünf Items bei einer Stichprobe $N \leq 1000$. Unter diesen beiden Bedingungen liegen die correct rejection rates bei .419 (unter der Bedingung $N = 150$, fünf Items und niedrige Faktorladungen), .482 (unter der Bedingung $N = 250$, fünf Items und niedrige Faktorladungen) und .764 (unter der Bedingung $N = 1000$, fünf Items und niedrige Faktorladungen). Hier zeigt sich folglich bereits der Einfluss der Stichprobengröße, der Testlänge und der Faktorladungshöhe. Erwartungsgemäß steigen die correct rejection rates unter der moderaten Ausprägung der Missspezifikation mit dem Anstieg der Stichprobengröße an.

Tabelle 7
Correct rejection rates und wrong rejection rates des RMSEA-P

N	Item- anzahl	Faktor- ladungs- höhe	Anzahl der Missspezifikationen im Strukturmodell			
			0	1	2	3
150	5	niedrig	.105 (.014;.028)	.321 (.037;.041)	.419 (.048;.043)	.995 (.157;.033)
	5	hoch	.173 (.023;.033)	.842 (.105;.046)	.956 (.132;.041)	1 (.278;.031)
	10	niedrig	.146 (.020;.031)	.689 (.077;.045)	.829 (.096;.042)	1 (.237;.032)
	10	hoch	.213 (.027;.036)	.982 (.154;.041)	.999 (.188;.038)	1 (.339;.029)
250	5	niedrig	.058 (.013;.022)	.351 (.045;.034)	.482 (.056;.032)	1 (.165;.026)
	5	hoch	.104 (.019;.027)	.931 (.110;.033)	.991 (.135;.029)	1 (.280;.023)
	10	niedrig	.095 (.018;.027)	.774 (.084;.034)	.932 (.104;.030)	1 (.241;.024)
	10	hoch	.115 (.019;.027)	.999 (.157;.030)	1 (.190;.029)	1 (.336;.022)
1000	5	niedrig	0 (.008;.013)	.401 (.055;.017)	.746 (.069;.014)	1 (.171;.013)
	5	hoch	.001 (.010;.015)	1 (.114;.014)	1 (.139;.014)	1 (.280;.012)
	10	niedrig	0 (.009;.013)	.984 (.091;.015)	1 (.109;.014)	1 (.242;.012)
	10	hoch	0 (.009;.013)	1 (.159;.014)	1 (.190;.014)	1 (.339;.011)

Bei der leichten Ausprägung der Missspezifikation im Strukturmodell wird der Einfluss der Faktorladungen auf die correct rejection rates noch wesentlich deutlicher. So kann diese Ausprägung der Missspezifikation nur dann zuverlässig detektiert werden, wenn hohe Faktorladungen gegeben sind. Dieser Einfluss zeigt sich bei kleineren Stichproben ($N \leq 1000$) unabhängig von der Itemanzahl. Mit steigender Stichprobengröße lässt der Einfluss der Faktorladungshöhen auf die correct rejection rates nach.

Wichtig ist, dass sich der RMSEA-P auf das Strukturmodell des Strukturgleichungsmodells fokussiert (siehe Kapitel 1.2.2.2 Die structural Fit-Indizes). Dennoch scheint es, dass der RMSEA-P aufgrund der vorliegenden Ergebnisse nicht von sogenannten „nuisance parameters“, wie etwa der Faktorladungshöhe, unabhängig ist. Diese nuisance parameters, wie schon in der Einleitung beschrieben, beeinflussen die Eigenschaften dieses Fit-Index, Missspezifikationen im Strukturmodell anzuzeigen. Dies lässt sich aber logisch erklären, da der RMSEA-P auf der Chi-Quadrat-Differenz zwischen dem

Gesamtmodell und dem saturierten strukturellen Teil des Strukturgleichungsmodells basiert.

Die oben beschriebenen Ergebnisse zeigen klar, dass die Fähigkeit des RMSEA-P, Missspezifikationen anzuzeigen, stark von den nuisance parameters (wie Höhe der Faktorladung oder Anzahl der manifesten Variablen) des Messmodells beeinflusst wird und daher *nicht* von dessen Eigenschaften unabhängig ist.

Die wrong rejection rates des RMSEA-P in Spalte „0“ zeigen ein inkonsistentes Bild und schwanken zwischen 0 und .213. Aber auch hier zeigt sich der Einfluss der Faktorladungshöhe. So liegen die wrong rejection rates bei niedrigen Faktorladungen und vor allem bei einer kleinen Stichprobengröße ($N \leq 250$) eher bei den gewünschten 5% .

Ergebnisse des RMSEAs < .05

Die Ergebnisse des RMSEAs, die in Tabelle 8 dargestellt sind, zeigen deutlich, dass alle Ausprägungen der Missspezifikationen im Strukturmodell über alle Bedingungen hinweg zuverlässig erkannt werden. Die correct rejection rates liegen durchgehend zwischen .996 und 1. Somit wird über alle Bedingungen und über alle Ausprägungen der Missspezifikationen im Strukturmodell hinweg eine akzeptable correct rejection rate von $\geq .80$ erreicht.

Was die wrong rejection rates des RMSEAs in Spalte „0“ betrifft, so kann ein angestrebter Wert von 5% unter keiner Bedingung erreicht werden. Die wrong rejection rates sind durchgehend inakzeptabel hoch und weichen stark vom gewünschten 5%-Niveau ab. Sie liegen zwischen .268 (unter der Bedingung hohe Faktorladungen, zehn Items und $N = 1000$) und .986 (unter der Bedingung niedrige Faktorladungen, fünf Items und $N = 250$).

Tabelle 8
Correct rejection rates und wrong rejection rates des RMSEAs

<i>N</i>	Item- anzahl	Faktor- ladungs- höhe	Anzahl der Missspezifikationen im Strukturmodell			
			0	1	2	3
150	5	niedrig	.982 (.380;.190)	.996 (.439;.177)	1 (.476;.167)	1 (.552;.129)
	5	hoch	.942 (.239;.146)	1 (.369;.129)	1 (.418;.115)	1 (.518;.090)
	10	niedrig	.977 (.324;.185)	1 (.425;.164)	1 (.453;.145)	1 (.546;.112)
	10	hoch	.807 (.134;.088)	1 (.315;.078)	1 (.372;.071)	1 (.483;.052)
250	5	niedrig	.986 (.358;.197)	1 (.428;.162)	1 (.474;.157)	1 (.558;.124)
	5	hoch	.911 (.173;.109)	1 (.337;.105)	1 (.391;.095)	1 (.497;.068)
	10	niedrig	.967 (.269;.172)	1 (.386;.140)	1 (.434;.128)	1 (.533;.098)
	10	hoch	.716 (.095;.059)	1 (.298;.055)	1 (.357;.050)	1 (.474;.038)
1000	5	niedrig	.962 (.203;.125)	1 (.356;.121)	1 (.413;.113)	1 (.512;.078)
	5	hoch	.636 (.077;.042)	1 (.291;.042)	1 (.351;.037)	1 (.466;.026)
	10	niedrig	.839 (.110;.06)	1 (.307;.062)	1 (.365;.058)	1 (.477;.038)
	10	hoch	.268 (.044;.026)	1 (.284;.025)	1 (.344;.025)	1 (.461;.017)

Ergebnisse des SRMRs < .08

Die Ergebnisse, dargestellt in Tabelle 9, zeigen, dass lediglich die dritte und folglich am stärksten ausgeprägte Missspezifikation im Strukturmodell vom SRMRs bei einem Cutoff-Wert von <.08 hinreichend gut detektiert werden kann. Die correct rejection rates liegen hier zwischen .998 (unter der Bedingung niedrige Faktorladungen, fünf Items und $N = 150$) und 1 (bei allen anderen Bedingungen unter der stärksten Ausprägung der Missspezifikation).

Bei der ersten und zweiten Ausprägung der Missspezifikation wird unter keiner Bedingung ein hinreichender Wert von $\geq .80$ für die correct rejection rates erreicht. Somit ist der SRMRs ungeeignet, leichte und moderate Missspezifikationen hinreichend zuverlässig anzuzeigen, unabhängig von der Stichprobengröße, der Faktorladungshöhe und der Itemanzahl pro latenter Variable. Die correct rejection rates liegen hier durchgehend zwischen 0 und .210.

Tabelle 9
Correct rejection rates und wrong rejection rates des SRMRs

N	Item- anzahl	Faktor- ladungs- höhe	Anzahl der Missspezifikationen im Strukturmodell			
			0	1	2	3
150	5	niedrig	.011 (.035;.016)	.100 (.053;.021)	.210 (.062;.023)	.998 (.174;.035)
	5	hoch	0 (.020;.008)	.004 (.042;.013)	.026 (.051;.014)	1 (.172;.022)
	10	niedrig	0 (.025;.010)	.02 (.046;.016)	.070 (.054;.016)	1 (.175;.027)
	10	hoch	0 (.015;.007)	.001 (.040;.011)	.005 (.049;.011)	1 (.172;.019)
250	5	niedrig	.002 (.029;.012)	.042 (.049;.017)	.116 (.057;.018)	1 (.176;.029)
	5	hoch	0 (.015;.006)	.001 (.040;.011)	.003 (.049;.011)	1 (.172;.018)
	10	niedrig	0 (.020;.008)	.003 (.043;.013)	.021 (.051;.013)	1 (.174;.020)
	10	hoch	0 (.011;.005)	0 (.039;.008)	0 (.048;.009)	1 (.172;.015)
1000	5	niedrig	0 (.015;.006)	0 (.041;.01)	.004 (.050;.010)	1 (.173;.014)
	5	hoch	0 (.008;.003)	0 (.038;.005)	0 (.048;.005)	1 (.171;.009)
	10	niedrig	0 (.010;.004)	0 (.039;.006)	0 (.048;.007)	1 (.172;.010)
	10	hoch	0 (.006;.002)	0 (.038;.004)	0 (.047;.004)	1 (.171;.007)

Die wrong rejection rates des SRMRs in Spalte „0“ liegen zwischen 0 und .011. Das an die Inferenzstatistik angelehnte gewünschte Niveau von 5% wird somit durchgehend nicht eingehalten.

Ergebnisse des CFI_s > .95

Die Ergebnisse des CFI_s in Tabelle 10 zeigen, dass durchgehend alle Ausprägungen der Missspezifikationen im Strukturmodell unabhängig von der Stichprobengröße, der Faktorladungshöhe und der Itemanzahl zuverlässig angezeigt werden können. Die zwei einzigen Ausnahmen finden sich unter der leichten Ausprägung der Missspezifikation unter der Bedingung, dass hohe Faktorladungen, zehn Items und eine Stichprobengröße von $N = 150$ (correct rejection rate: .778) bzw. $N = 250$ (correct rejection rate: .775) vorliegen. Man muss jedoch festhalten, dass diese correct rejection rates gerundet den gewünschten 80% entsprechen würden.

Tabelle 10

Correct rejection rates und wrong rejection rates des CFI_s

N	Item- anzahl	Faktor- ladungs- höhe	Anzahl der Missspezifikationen im Strukturmodell			
			0	1	2	3
150	5	niedrig	.691 (.903;.080)	.874 (.850;.096)	.972 (.793;.109)	1 (.683;.102)
	5	hoch	.259 (.96;.044)	.805 (.898;.060)	.974 (.847;.069)	1 (.734;.064)
	10	niedrig	.485 (.932;.065)	.872 (.869;.079)	.976 (.823;.088)	1 (.710;.080)
	10	hoch	.034 (.986;.017)	.778 (.924;.034)	.982 (.875;.043)	1 (.759;.041)
250	5	niedrig	.548 (.918;.076)	.883 (.863;.084)	.976 (.805;.097)	1 (.694;.091)
	5	hoch	.087 (.978;.029)	.801 (.914;.047)	.980 (.865;.054)	1 (.750;.048)
	10	niedrig	.303 (.951;.058)	.862 (.890;.067)	.980 (.839;.075)	1 (.723;.069)
	10	hoch	.001 (.993;.008)	.775 (.932;.023)	.998 (.884;.030)	1 (.765;.031)
1000	5	niedrig	.150 (.971;.037)	.816 (.905;.056)	.988 (.851;.065)	1 (.737;.054)
	5	hoch	0 (.995;.005)	.806 (.935;.017)	1 (.888;.021)	1 (.770;.021)
	10	niedrig	.012 (.991;.011)	.811 (.929;.026)	.998 (.880;.033)	1 (.764;.028)
	10	hoch	0 (.998;.002)	.865 (.939;.011)	1 (.892;.015)	1 (.774;.014)

Die Ergebnisse für die leichteste Missspezifikation zeigen bei genauer Betrachtung ein unerwartetes Muster. Die correct rejection rates fallen teilweise in der Bedingung $N = 250$ im Vergleich zu $N = 150$ und steigen dann in der größten Stichprobe $N = 1000$ wieder an. Dieser Bias zeigt sich unabhängig von der Faktorladungshöhe und der Testlänge.

Dieses Phänomen erklärt sich über den u-förmigen Zusammenhang zwischen der Stichprobengröße und den correct rejection rates. Bei kleinen Stichproben kommt es zum bekannten Bias des Erwartungswertes der nonzentralen Chi-Quadrat-Verteilung, da diese bei zu kleinen Stichproben nicht hinreichend approximiert werden kann (Curran et al., 2002). Bei mittelgroßen Stichprobengrößen nimmt die Power ab, gleichzeitig wird auch der Bias des Erwartungswertes kleiner, da nun der Mittelwert der nonzentralen Chi-Quadrat-Verteilung kleiner ist.

Bei großen Stichproben wird die nonzentrale Chi-Quadrat-Verteilung dann hinreichend gut approximiert, der Erwartungswert zeigt keinen oder einen nur geringen positiven Bias im Vergleich zu kleinen Stichproben, und die Wahrscheinlichkeit, eine Missspezifikation zu entdecken, steigt bedingt durch den Anstieg der Stichprobengröße.

Ein Blick auf die wrong rejection rates in Spalte „0“ zeigt, dass ein gewünschtes Niveau von 5% noch am ehesten unter der Bedingung kleine Stichprobe, zehn Items pro latenter Variable und hohe Faktorladungen erreicht wird (wrong rejection rates .034). Ansonsten schwanken die wrong rejection rates zwischen 0 und .691.

Zusammengefasst lässt sich sagen, dass die structural Fit-Indizes durchaus in der Lage sind, Missspezifikationen im Strukturmodell zuverlässig anzuzeigen. Die dritte und somit am stärksten ausgeprägte Missspezifikation wird von allen structural Fit-Indizes hinreichend gut angezeigt. Auch die erste und zweite Ausprägung der Missspezifikation im Strukturmodell werden vom CFI_s und RMSEA_s durchgehend erkannt, beim RMSEA-P zeigt sich unter diesen Missspezifikationsbedingungen deutlich der Einfluss der Faktorladungen auf die correct rejection rates. So kann der RMSEA-P unter der leichten bis mittelgradigen Ausprägung der Missspezifikation nur dann zuverlässig detektieren, wenn die Faktorladungen hoch sind. Allein der SRMRs kann weder die erste noch die zweite Missspezifikationsausprägung im Strukturmodell hinreichend gut detektieren.

3.2 Ergebnisse des Modellfits zum Modell mit Missspezifikationen sowohl im Mess- als auch Strukturmodell

Nach der Betrachtung der Ergebnisse unter der Bedingung, dass das Messmodell perfekt passt, werden nun die Ergebnisse beschrieben, wo sowohl im Mess- als auch im Strukturmodell Missspezifikationen auftreten. Der Fokus liegt, im Sinne des Hauptthemas dieser Arbeit, auf den structural Fit-Indizes; die Eigenschaften der composite Fit-Indizes werden nicht betrachtet. Die wrong rejection rates werden nicht mehr gesondert besprochen, da diese im vorherigen Ergebnisteil ausführlich diskutiert wurden. Dennoch werden diese in der Spalte 0-0 der Vollständigkeit halber angeführt.

Ergebnisse des RMSEA-P < .05

Die Ergebnisse des RMSEA-P, dargestellt in Tabelle 11, zeigen klar, dass unter allen Bedingungen die stärkste Ausprägung der Missspezifikation im Strukturmodell unabhängig von der Stärke der korrelierten Fehler im Messmodell hinreichend zuverlässig detektiert wird. Hier liegen über alle Bedingungen hinweg die correct rejection rates nahe 1 (Bedingung 3-1 bis 3-3).

Bei der moderaten Ausprägung der Missspezifikation im Strukturmodell zeigt sich, dass unter der am stärksten ausgeprägten Fehlerkorrelationsbedingung die correct rejection rates durchgehend die gewünschten 80% erreichen. Die einzige Ausnahme findet sich unter der Bedingung 2-3 bei niedrigen Faktorladungen, fünf Items und einer Stichprobe von $N = 150$. Die correct rejection rate unter dieser Bedingung beträgt .705.

Ein ähnliches Bild zeigt sich unter der Bedingung, dass zwei Pfade im Strukturmodell und die mittleren Fehlerkorrelationen im Messmodell fälschlich auf null gesetzt wurden. Bis auf die Bedingung niedrige Faktorladungen und fünf Items bei einer Stichprobe $N \leq 250$ werden alle Missspezifikationen hinreichend gut angezeigt. Die correct rejection rates liegen zwischen .823 und 1.

Der Einfluss der Faktorladungen in Kombination mit der Testlänge auf die correct rejection rates wird bei der moderaten Ausprägung der Missspezifikation im Strukturmodell und der leichtesten Ausprägung der Fehlerkorrelation im Messmodell deutlich. So liegen die correct rejection rates über alle drei Stichprobenbedingungen

hinweg bei niedrigen Faktorladungen und fünf Items zwischen .359 und .607, was als nicht ausreichend zu bewerten ist. Zudem kann diese Missspezifikationsausprägung (2-1) mit einer correct rejection rate von .794 unter der Bedingung niedrige Faktorladungen und zehn Items und einer Stichprobengröße von $N = 150$ nicht angezeigt werden. Über alle anderen Bedingungen hinweg liegen die correct rejection rates unter der Missspezifikationsbedingung 2-1 nahe 1.

Besonders ausschlaggebend wird der Einfluss der Faktorladungen in Kombination mit der Testlänge bei Stichproben $N \leq 1000$ bei den leichten Missspezifikationsausprägungen im Strukturmodell über alle Fehlerkorrelationsausprägungen hinweg. Bei kleineren Stichproben ($N \leq 1000$) können die correct rejection rates bei den Missspezifikationsbedingungen 1-2 (ein Pfad im Strukturmodell auf null gesetzt und eine moderate Fehlerkorrelationsbedingung) und 1-3 (ein Pfad im Strukturmodell auf null gesetzt und eine schwere Fehlerkorrelationsbedingung) nur unter der Bedingung hohe Faktorladungen und zehn Items als hinreichend bezeichnet werden. Die correct rejection rates liegen hier zwischen .984 und .999. Bei einer Stichprobengröße von $N = 1000$ werden bis auf die Bedingung fünf Items und niedrige Faktorladungen alle Missspezifikationen über die Bedingungen 1-1 bis 1-3 hinweg hinreichend gut detektiert.

Unter der Bedingung, dass keine Missspezifikationen im Strukturmodell vorliegen, sondern nur Fehlerkorrelationen fälschlicherweise auf null gesetzt wurden, zeigt der RMSEA-P keine Missspezifikationen an. Die wrong rejection rates liegen zwischen .002 und .746.

Tabelle 11

Correct rejection rates and wrong rejection rates des RMSEA-P mit Misspezifikationen der Fehlerkorrelationen

N	Item-anzahl	Faktorladungshöhe	Anzahl der Misspezifikationen im Mess- und Strukturmodell												
			0-0	0-1	0-2	0-3	1-1	1-2	1-3	2-1	2-2	2-3	3-1	3-2	3-3
150	5	niedrig	.105 (.014;.028)	.111 (.016;.028)	.196 (.025;.035)	.369 (.047;.045)	.173 (.024;.033)	.153 (.022;.032)	.352 (.045;.046)	.359 (.042;.041)	.462 (.053;.042)	.705 (.080;.046)	.991 (.152;.035)	.992 (.149;.035)	.993 (.166;.037)
150	5	hoch	.173 (.023;.033)	.212 (.026;.037)	.236 (.030;.038)	.353 (.042;.043)	.878 (.112;.046)	.634 (.074;.047)	.641 (.075;.047)	.948 (.125;.041)	.947 (.124;.041)	.953 (.130;.040)	1 (.270;.030)	1 (.269;.031)	1 (.270;.031)
150	10	niedrig	.146 (.020;.031)	.142 (.020;.031)	.167 (.021;.035)	.197 (.024;.036)	.517 (.058;.042)	.456 (.052;.047)	.486 (.060;.045)	.794 (.090;.042)	.823 (.093;.041)	.887 (.102;.039)	1 (.235;.031)	1 (.235;.033)	1 (.241;.032)
150	10	hoch	.213 (.027;.036)	.255 (.032;.038)	.361 (.044;.044)	.459 (.053;.045)	.958 (.142;.044)	.984 (.147;.039)	.986 (.149;.039)	1 (.184;.037)	1 (.187;.036)	1 (.191;.034)	1 (.338;.029)	1 (.339;.029)	1 (.340;.029)
250	5	niedrig	.058 (.013;.022)	.077 (.015;.024)	.171 (.026;.031)	.463 (.053;.038)	.124 (.022;.028)	.156 (.027;.030)	.447 (.053;.035)	.362 (.047;.033)	.571 (.063;.032)	.83 (.089;.034)	1 (.156;.027)	1 (.158;.026)	.997 (.172;.026)
250	5	hoch	.104 (.019;.027)	.132 (.021;.028)	.175 (.026;.031)	.349 (.044;.037)	.835 (.091;.034)	.716 (.077;.036)	.739 (.079;.037)	.980 (.127;.031)	.991 (.126;.030)	.996 (.133;.029)	1 (.273;.024)	1 (.270;.023)	1 (.271;.024)
250	10	niedrig	.095 (.018;.027)	.102 (.02;.027)	.294 (.036;.035)	.265 (.037;.034)	.810 (.088;.036)	.616 (.068;.032)	.660 (.073;.036)	.920 (.099;.03)	.928 (.103;.028)	.977 (.105;.027)	1 (.240;.025)	1 (.239;.027)	1 (.244;.027)
250	10	hoch	.115 (.019;.027)	.175 (.027;.032)	.376 (.045;.036)	.474 (.054;.039)	.995 (.146;.031)	.996 (.149;.030)	.999 (.153;.029)	1 (.187;.028)	1 (.189;.027)	1 (.192;.027)	1 (.339;.022)	1 (.339;.022)	1 (.340;.022)
1000	5	niedrig	0 (.008;.013)	.002 (.012;.016)	.090 (.034;.021)	.746 (.072;.019)	.016 (.024;.018)	.051 (.031;.019)	.688 (.067;.017)	.607 (.063;.014)	.893 (.076;.014)	1 (.099;.014)	1 (.164;.013)	1 (.165;.013)	1 (.179;.014)
1000	5	hoch	.001 (.010;.015)	.002 (.012;.015)	.030 (.022;.019)	.235 (.046;.019)	.988 (.095;.016)	.909 (.082;.016)	.939 (.084;.015)	1 (.129;.014)	1 (.128;.014)	1 (.135;.014)	1 (.273;.011)	1 (.270;.011)	1 (.272;.011)
1000	10	niedrig	0 (.009;.013)	.049 (.035;.018)	.167 (.051;.011)	.333 (.057;.016)	.909 (.079;.015)	.905 (.081;.013)	1 (.103;.028)	1 (.111;.013)	1 (.116;.009)	1 (.103;.015)	1 (.243;.012)	1 (.243;.014)	1 (.236;.012)
1000	10	hoch	0 (.009;.013)	.031 (.025;.019)	.262 (.049;.019)	.483 (.059;.019)	1 (.149;.014)	1 (.151;.014)	1 (.156;.014)	1 (.189;.013)	1 (.190;.013)	1 (.193;.013)	1 (.339;.010)	1 (.339;.011)	1 (.339;.011)

Anmerkung. 0-0: keine Missspezifikation im Mess- oder Strukturmodell. 0-1: keine Missspezifikation im Strukturmodell, Fehlerkorrelationen im Messmodell im Intervall $[-.1; .1]$ auf null gesetzt. 0-2: keine Missspezifikation im Strukturmodell, Fehlerkorrelationen im Messmodell im Intervall $[-.2; .2]$ auf null gesetzt. 0-3: keine Missspezifikation im Strukturmodell, Fehlerkorrelationen im Messmodell im Intervall $[-.3; .3]$ auf null gesetzt. 1-1: ein Pfad im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.1; .1]$ auf null gesetzt. 1-2: ein Pfad im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.2; .2]$ auf null gesetzt. 1-3: ein Pfad im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.3; .3]$ auf null gesetzt. 2-1: zwei Pfade im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.1; .1]$ auf null gesetzt. 2-2: zwei Pfade im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.2; .2]$ auf null gesetzt. 2-3: zwei Pfade im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.3; .3]$ auf null gesetzt. 3-1: drei Pfade im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.1; .1]$ auf null gesetzt. 3-2: drei Pfade im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.2; .2]$ auf null gesetzt. 3-3: drei Pfade im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.3; .3]$ auf null gesetzt.

Ergebnisse des $RMSEA_s < .05$

Die Ergebnisse des $RMSEA_s$ in Tabelle 12 zeigen, dass die leichte, die moderate und die schwerste Ausprägung der Missspezifikation im Strukturmodell unabhängig von den Fehlerkorrelationsmissspezifikationen bei allen Simulationsbedingungen (Stichprobengröße, Itemanzahl und Faktorladungshöhe) hinreichend gut angezeigt werden können. Die correct rejection rates liegen hier zwischen .948 und 1.

Unter der Bedingung, dass nur im Messmodell Missspezifikationen vorliegen und keine Missspezifikationen im Strukturmodell vorhanden sind, zeigt sich, dass diese Missspezifikationen im Messmodell durchgehend angezeigt werden. Die einzigen Ausnahmen bilden dabei die correct rejection rates unter den Bedingungen hohe Faktorladungen, fünf Items und $N = 1000$ bei den Missspezifikationsbedingungen 0-1 und 0-2 sowie unter der Bedingung hohe Faktorladungen, zehn Items und $N = 1000$ bei der Missspezifikationsbedingung 0-1. Obwohl sich der $RMSEA_s$ als Fit-Index auf den Strukturteil eines Strukturmodells bezieht, werden hier offensichtlich Missspezifikationen im Messmodell angezeigt. Laut Hancock und Mueller (2011) sollte sich der $RMSEA_s$ aber eben von nuisance parameters unabhängig zeigen. Hier wird aber deutlich die Abhängigkeit des $RMSEA_s$ von Gegebenheiten im Messmodell sichtbar.

Tabelle 12

Correct rejection rates and wrong rejection rates des RMSEA_s mit Missspezifikationen der Fehlerkorrelationen

Anzahl der Missspezifikationen im Mess- und Strukturmodell															
<i>N</i>	Item- an- zahl	Faktor- ladungs- höhe	0-0	0-1	0-2	0-3	1-1	1-2	1-3	2-1	2-2	2-3	3-1	3-2	3-3
150	5	niedrig	.982 (.380; .190)	.989 (.403; .197)	.989 (.383; .193)	.995 (.456; .210)	.998 (.426; .172)	.997 (.382; .161)	.997 (.425; .173)	.998 (.454; .156)	1 (.451; .159)	1 (.510; .170)	1 (.560; .133)	1 (.507; .130)	1 (.537; .134)
150	5	hoch	.942 (.239; .146)	.924 (.247; .167)	.940 (.235; .143)	.946 (.253; .152)	.948 (.757; .809)	.997 (.313; .131)	.996 (.319; .134)	1 (.406; .130)	1 (.381; .115)	1 (.394; .113)	1 (.527; .106)	1 (.487; .083)	1 (.482; .085)
150	10	niedrig	.977 (.324; .185)	.978 (.376; .203)	1 (.423; .225)	.972 (.474; .238)	1 (.444; .184)	1 (.468; .207)	1 (.506; .208)	1 (.495; .157)	1 (.506; .152)	1 (.550; .161)	1 (.590; .128)	1 (.603; .118)	1 (.623; .126)
150	10	hoch	.807 (.134; .088)	.876 (.170; .111)	.927 (.228; .15)	.940 (.253; .155)	.999 (.318; .103)	1 (.344; .106)	1 (.360; .106)	1 (.384; .080)	1 (.407; .092)	1 (.426; .099)	1 (.496; .063)	1 (.510; .070)	1 (.519; .073)
250	5	niedrig	.986 (.358; .197)	.990 (.354; .188)	.998 (.379; .184)	.996 (.451; .195)	.995 (.365; .165)	.993 (.381; .186)	1 (.436; .179)	1 (.424; .149)	1 (.434; .149)	1 (.503; .150)	1 (.525; .124)	1 (.519; .131)	1 (.545; .120)
250	5	hoch	.911 (.173; .109)	.896 (.180; .124)	.919 (.175; .105)	.949 (.204; .112)	.998 (.305; .111)	.997 (.276; .105)	.994 (.282; .110)	1 (.363; .087)	1 (.350; .087)	1 (.363; .093)	1 (.479; .070)	1 (.464; .064)	1 (.463; .063)
250	10	niedrig	.967 (.269; .172)	.981 (.361; .195)	.991 (.479; .201)	1 (.497; .184)	.982 (.587; .649)	1 (.472; .166)	1 (.463; .155)	1 (.499; .156)	1 (.539; .142)	1 (.552; .196)	1 (.579; .106)	1 (.619; .121)	1 (.632; .114)
250	10	hoch	.716 (.095; .059)	.813 (.130; .083)	.929 (.188; .103)	.959 (.213; .113)	1 (.296; .063)	1 (.320; .074)	1 (.341; .086)	1 (.368; .056)	1 (.385; .066)	1 (.403; .078)	1 (.482; .040)	1 (.494; .050)	1 (.502; .054)
1000	5	niedrig	.962 (.203; .125)	.961 (.222; .139)	.998 (.297; .159)	1 (.476; .182)	.998 (.269; .132)	.998 (.288; .135)	1 (.434; .161)	1 (.344; .100)	1 (.373; .120)	1 (.497; .131)	1 (.471; .080)	1 (.458; .078)	1 (.540; .101)
1000	5	hoch	.636 (.077; .042)	.661 (.080; .042)	.793 (.095; .043)	.977 (.142; .045)	1 (.254; .044)	1 (.225; .043)	1 (.238; .042)	1 (.322; .037)	1 (.309; .035)	1 (.324; .034)	1 (.447; .025)	1 (.433; .023)	1 (.436; .023)
1000	10	niedrig	.839 (.110; .060)	1 (.401; .186)	1 (.593; .18)	1 (.742; .176)	1 (.437; .158)	1 (.653; .197)	1 (.917; .082)	1 (.519; .128)	1 (.634; .125)	1 (.314; .143)	1 (.584; .095)	1 (.667; .103)	1 (.701; .160)
1000	10	hoch	.268 (.044; .026)	.752 (.087; .039)	.975 (.143; .047)	.996 (.171; .052)	1 (.275; .028)	1 (.290; .032)	1 (.311; .034)	1 (.350; .025)	1 (.361; .028)	1 (.374; .029)	1 (.468; .017)	1 (.475; .020)	1 (.482; .021)

Anmerkung. 0-0: keine Missspezifikation im Mess- oder Strukturmodell. 0-1: keine Missspezifikation im Strukturmodell, Fehlerkorrelationen im Messmodell im Intervall $[-.1; .1]$ auf null gesetzt. 0-2: keine Missspezifikation im Strukturmodell, Fehlerkorrelationen im Messmodell im Intervall $[-.2; .2]$ auf null gesetzt. 0-3: keine Missspezifikation im Strukturmodell, Fehlerkorrelationen im Messmodell im Intervall $[-.3; .3]$ auf null gesetzt. 1-1: ein Pfad im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.1; .1]$ auf null gesetzt. 1-2: ein Pfad im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.2; .2]$ auf null gesetzt. 1-3: ein Pfad im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.3; .3]$ auf null gesetzt. 2-1: zwei Pfade im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.1; .1]$ auf null gesetzt. 2-2: zwei Pfade im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.2; .2]$ auf null gesetzt. 2-3: zwei Pfade im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.3; .3]$ auf null gesetzt. 3-1: drei Pfade im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.1; .1]$ auf null gesetzt. 3-2: drei Pfade im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.2; .2]$ auf null gesetzt. 3-3: drei Pfade im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.3; .3]$ auf null gesetzt.

Ergebnisse des $SRMR_s < .08$

Die Ergebnisse des $SRMR_s$ in Tabelle 13 zeigen, dass die stärkste Ausprägung der Missspezifikation im Strukturmodell unabhängig von der Fehlerkorrelationsmissspezifikation über alle Bedingungen hinweg hinreichend gut detektiert werden kann. So liegen die correct rejection rates zwischen .993 und 1.

Die leichte und die moderate Ausprägung der Missspezifikation im Strukturmodell können unabhängig von den Ausprägungen der Fehlerkorrelationen und über alle Simulationsbedingungen (Stichprobengröße, Itemanzahl und Faktorladungshöhe) hinweg nicht zuverlässig angezeigt werden. Die correct rejection rates liegen zwischen .287 und 0, was als nicht ausreichend zu bewerten ist.

Unter der Bedingung, dass das Strukturmodell richtig spezifiziert ist und Missspezifikationen nur im Messmodell auftreten, liegen die correct rejection rates unabhängig von der Stichprobengröße, der Itemanzahl und den Faktorladungen zwischen .081 und 0. Somit können Missspezifikationen im Messmodell unter keiner Simulationsbedingung angezeigt werden.

Tabelle 13

Correct rejection rates and wrong rejection rates des SRMRs mit Missspezifikationen der Fehlerkorrelationen

Anzahl der Missspezifikationen im Mess- und Strukturmodell															
N	Item-anzahl	Faktorladungshöhe	0-0	0-1	0-2	0-3	1-1	1-2	1-3	2-1	2-2	2-3	3-1	3-2	3-3
150	5	niedrig	.011 (.035;.016)	.014 (.036;.016)	.026 (.040;.018)	.081 (.048;.020)	.040 (.045;.018)	.049 (.044;.019)	.160 (.053;.025)	.111 (.056;.020)	.135 (.058;.019)	.287 (.070;.023)	.993 (.161;.034)	.979 (.152;.035)	.986 (.160;.039)
150	5	hoch	0 (.020;.008)	0 (.020;.009)	0 (.021;.009)	0 (.022;.009)	.071 (.055;.037)	.003 (.034;.012)	0 (.033;.013)	.018 (.048;.013)	.014 (.047;.013)	.022 (.048;.014)	1 (.173;.023)	1 (.164;.023)	1 (.164;.023)
150	10	niedrig	0 (.025;.010)	0 (.025;.011)	.008 (.026;.011)	0 (.028;.012)	.008 (.040;.014)	.007 (.039;.016)	.014 (.041;.015)	.024 (.050;.015)	.031 (.052;.015)	.070 (.054;.015)	1 (.172;.027)	1 (.168;.029)	1 (.169;.026)
150	10	hoch	0 (.015;.007)	0 (.015;.006)	0 (.016;.007)	0 (.017;.007)	0 (.038;.011)	0 (.038;.010)	0 (.037;.010)	.004 (.048;.010)	.004 (.048;.010)	.006 (.048;.010)	1 (.171;.019)	1 (.170;.019)	1 (.169;.020)
250	5	niedrig	.002 (.029;.012)	.002 (.029;.013)	.005 (.034;.014)	.021 (.042;.016)	.003 (.036;.015)	.014 (.037;.016)	.044 (.046;.016)	.037 (.048;.015)	.055 (.053;.016)	.190 (.064;.019)	.998 (.158;.028)	.998 (.154;.028)	.992 (.158;.030)
250	5	hoch	0 (.015;.006)	0 (.015;.007)	0 (.016;.007)	0 (.019;.007)	0 (.035;.010)	0 (.031;.010)	0 (.030;.010)	.003 (.045;.010)	.001 (.044;.010)	.004 (.045;.010)	1 (.167;.018)	1 (.164;.017)	1 (.164;.018)
250	10	niedrig	0 (.020;.008)	0 (.020;.009)	0 (.024;.009)	0 (.025;.008)	.024 (.047;.024)	0 (.036;.011)	0 (.037;.011)	.007 (.047;.012)	.010 (.048;.011)	0 (.048;.010)	1 (.172;.022)	1 (.168;.022)	1 (.168;.026)
250	10	hoch	0 (.011;.005)	0 (.012;.005)	0 (.013;.005)	0 (.014;.005)	0 (.037;.008)	0 (.036;.008)	0 (.036;.008)	0 (.047;.008)	0 (.047;.008)	.001 (.047;.008)	1 (.171;.015)	1 (.169;.015)	1 (.168;.015)
1000	5	niedrig	0 (.015;.006)	0 (.016;.006)	0 (.023;.008)	0 (.037;.009)	0 (.024;.009)	0 (.025;.009)	0 (.037;.009)	0 (.040;.008)	.002 (.044;.008)	.013 (.057;.009)	1 (.158;.014)	1 (.152;.015)	1 (.157;.016)
1000	5	hoch	0 (.008;.003)	0 (.008;.003)	0 (.009;.004)	0 (.013;.004)	0 (.032;.005)	0 (.028;.005)	0 (.027;.005)	0 (.043;.005)	0 (.041;.005)	0 (.042;.005)	1 (.166;.009)	1 (.163;.009)	1 (.163;.009)
1000	10	niedrig	0 (.010;.004)	0 (.014;.004)	0 (.017;.003)	0 (.018;.004)	0 (.034;.006)	0 (.033;.006)	0 (.036;.009)	0 (.044;.006)	0 (.044;.005)	0 (.044;.007)	1 (.170;.010)	1 (.164;.011)	1 (.154;.011)
1000	10	hoch	0 (.006;.002)	0 (.007;.002)	0 (.009;.002)	0 (.010;.003)	0 (.036;.004)	0 (.035;.004)	0 (.035;.004)	0 (.046;.004)	0 (.045;.004)	0 (.046;.004)	1 (.170;.007)	1 (.168;.007)	1 (.167;.007)

Anmerkung. 0-0: keine Missspezifikation im Mess- oder Strukturmodell. 0-1: keine Missspezifikation im Strukturmodell, Fehlerkorrelationen im Messmodell im Intervall $[-.1; .1]$ auf null gesetzt. 0-2: keine Missspezifikation im Strukturmodell, Fehlerkorrelationen im Messmodell im Intervall $[-.2; .2]$ auf null gesetzt. 0-3: keine Missspezifikation im Strukturmodell, Fehlerkorrelationen im Messmodell im Intervall $[-.3; .3]$ auf null gesetzt. 1-1: ein Pfad im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.1; .1]$ auf null gesetzt. 1-2: ein Pfad im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.2; .2]$ auf null gesetzt. 1-3: ein Pfad im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.3; .3]$ auf null gesetzt. 2-1: zwei Pfade im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.1; .1]$ auf null gesetzt. 2-2: zwei Pfade im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.2; .2]$ auf null gesetzt. 2-3: zwei Pfade im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.3; .3]$ auf null gesetzt. 3-1: drei Pfade im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.1; .1]$ auf null gesetzt. 3-2: drei Pfade im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.2; .2]$ auf null gesetzt. 3-3: drei Pfade im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.3; .3]$ auf null gesetzt.

Ergebnisse des CFI_s > .95

Die Ergebnisse des CFI_s zeigen, dass bei der schweren und der moderaten Ausprägung der Missspezifikation im Strukturmodell über Stichprobengröße, Faktorladungshöhe und Itemanzahl hinweg sowie unabhängig von der Missspezifikationsausprägung im Messmodell alle Missspezifikationen sehr zuverlässig angezeigt werden können. Die correct rejection rates liegen zwischen .916 und 1.

Unter der Bedingung, dass ein Pfad im Strukturmodell auf null gesetzt wurde und die Fehlerkorrelationen im Messmodell im Intervall [-.3; .3] auf null gesetzt wurden, und unter der Bedingung hohe Faktorladungen, fünf Items pro latenter Variable und über alle drei Stichprobengrößen hinweg müssen die correct rejection rates als nicht hinreichend bezeichnet werden. Die correct rejection rates liegen unter diesen Bedingungen zwischen .345 und .646. Auffallend dabei ist auch, dass die correct rejection rates mit steigender Stichprobengröße sinken. Dies scheint wiederum ein Indikator für die unzureichende Approximation der nonzentralen Chi-Quadrat-Verteilung zu sein, wobei dieser Frage in einer weiteren Simulationsstudie explizit nachgegangen werden sollte. Alle anderen Missspezifikationsbedingungen unter 1-3 werden hinreichend angezeigt.

Unter der Bedingung 1-2 (ein Pfad im Strukturmodell auf null gesetzt bei einer mittleren Ausprägung der Fehlerkorrelationsbedingung im Messmodell) zeigt sich, dass vor allem die Faktorladungshöhe eine entscheidende Rolle spielt. So zeigt sich, dass die correct rejection rates bei niedrigen Faktorladungen und zehn Items eher die gewünschten 80% erreichen. Ein ähnliches Bild zeigt sich unter der Bedingung 1-1.

Tabelle 14

Correct rejection rates and wrong rejection rates des CFI_s mit Missspezifikationen der Fehlerkorrelationen

Anzahl der Missspezifikationen im Mess- und Strukturmodell															
N	Item-anzahl	Faktorladungshöhe	0-0	0-1	0-2	0-3	1-1	1-2	1-3	2-1	2-2	2-3	3-1	3-2	3-3
150	5	niedrig	.691 (.903;.080)	.714 (.896;.083)	.695 (.896;.083)	.815 (.858;.099)	.847 (.859;.090)	.826 (.876;.083)	.893 (.845;.094)	.963 (.810;.100)	.951 (.803;.103)	.981 (.749;.118)	.998 (.687;.101)	1 (.714;.099)	1 (.675;.104)
150	5	hoch	.259 (.960;.044)	.276 (.956;.052)	.260 (.960;.043)	.313 (.954;.050)	.836 (.763;.277)	.650 (.921;.058)	.646 (.916;.061)	.959 (.856;.074)	.952 (.868;.067)	.951 (.857;.067)	1 (.731;.073)	1 (.756;.060)	1 (.756;.062)
150	10	niedrig	.485 (.932;.065)	.561 (.917;.074)	.667 (.901;.088)	.761 (.882;.098)	.839 (.870;.084)	.846 (.858;.099)	.889 (.836;.109)	.976 (.813;.091)	.969 (.802;.092)	1 (.775;.100)	1 (.699;.086)	1 (.692;.085)	1 (.666;.089)
150	10	hoch	.034 (.986;.017)	.097 (.979;.028)	.215 (.964;.044)	.285 (.957;.047)	.693 (.925;.044)	.774 (.916;.047)	.833 (.909;.047)	.988 (.874;.046)	.988 (.863;.052)	.993 (.852;.056)	1 (.759;.045)	1 (.753;.048)	1 (.747;.050)
250	5	niedrig	.548 (.918;.076)	.582 (.918;.074)	.683 (.905;.075)	.815 (.868;.090)	.735 (.894;.079)	.764 (.882;.092)	.899 (.848;.096)	.949 (.841;.086)	.966 (.824;.092)	.990 (.766;.102)	1 (.725;.089)	1 (.719;.093)	1 (.684;.094)
250	5	hoch	.087 (.978;.029)	.113 (.975;.034)	.096 (.977;.027)	.166 (.969;.033)	.647 (.928;.049)	.536 (.938;.044)	.565 (.934;.047)	.960 (.882;.048)	.952 (.887;.049)	.976 (.876;.054)	1 (.766;.049)	1 (.774;.046)	1 (.770;.046)
250	10	niedrig	.303 (.951;.058)	.531 (.926;.067)	.789 (.884;.078)	.898 (.881;.068)	.861 (.817;.226)	.920 (.863;.076)	.915 (.862;.074)	.983 (.815;.087)	1 (.790;.085)	1 (.785;.110)	1 (.713;.071)	1 (.688;.079)	1 (.676;.079)
250	10	hoch	.001 (.993;.008)	.032 (.987;.015)	.112 (.976;.025)	.171 (.97;.030)	.685 (.936;.026)	.784 (.927;.031)	.831 (.918;.038)	.996 (.883;.031)	.997 (.876;.037)	.998 (.865;.044)	1 (.769;.030)	1 (.764;.035)	1 (.758;.037)
1000	5	niedrig	.150 (.971;.037)	.207 (.965;.042)	.422 (.940;.058)	.912 (.863;.084)	.444 (.941;.055)	.537 (.931;.057)	.928 (.857;.083)	.916 (.894;.053)	.964 (.873;.066)	1 (.784;.081)	1 (.777;.052)	1 (.779;.052)	1 (.706;.068)
1000	5	hoch	0 (.995;.005)	0 (.995;.005)	.001 (.993;.006)	.005 (.986;.009)	.459 (.950;.017)	.237 (.959;.015)	.354 (.954;.016)	.995 (.905;.020)	.989 (.910;.019)	.998 (.899;.020)	1 (.787;.019)	1 (.794;.018)	1 (.789;.019)
1000	10	niedrig	.012 (.991;.011)	.618 (.916;.067)	1 (.843;.081)	1 (.779;.086)	.881 (.883;.072)	1 (.783;.095)	1 (.642;.049)	1 (.810;.072)	1 (.744;.072)	.978 (.904;.057)	1 (.719;.060)	1 (.666;.067)	1 (.642;.104)
1000	10	hoch	0 (.998;.002)	0 (.995;.004)	.003 (.987;.008)	.008 (.982;.011)	.683 (.945;.010)	.787 (.904;.012)	.920 (.932;.014)	1 (.894;.014)	1 (.889;.015)	1 (.882;.016)	1 (.779;.013)	1 (.777;.015)	1 (.773;.015)

Anmerkung. 0-0: keine Missspezifikation im Mess- oder Strukturmodell. 0-1: keine Missspezifikation im Strukturmodell, Fehlerkorrelationen im Messmodell im Intervall $[-.1; .1]$ auf null gesetzt. 0-2: keine Missspezifikation im Strukturmodell, Fehlerkorrelationen im Messmodell im Intervall $[-.2; .2]$ auf null gesetzt. 0-3: keine Missspezifikation im Strukturmodell, Fehlerkorrelationen im Messmodell im Intervall $[-.3; .3]$ auf null gesetzt. 1-1: ein Pfad im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.1; .1]$ auf null gesetzt. 1-2: ein Pfad im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.2; .2]$ auf null gesetzt. 1-3: ein Pfad im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.3; .3]$ auf null gesetzt. 2-1: zwei Pfade im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.1; .1]$ auf null gesetzt. 2-2: zwei Pfade im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.2; .2]$ auf null gesetzt. 2-3: zwei Pfade im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.3; .3]$ auf null gesetzt. 3-1: drei Pfade im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.1; .1]$ auf null gesetzt. 3-2: drei Pfade im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.2; .2]$ auf null gesetzt. 3-3: drei Pfade im Strukturmodell auf null gesetzt, Fehlerkorrelationen im Messmodell im Intervall $[-.3; .3]$ auf null gesetzt.

Unter der Bedingung, dass Missspezifikationen nur im Messmodell auftreten, zeigt sich erneut der Einfluss der Faktorladungshöhen bei der schweren Ausprägung der Fehlerkorrelationsbedingung. Bei niedrigen Faktorladungen und unabhängig von der Itemanzahl bei einer Stichprobe $N \geq 150$ werden die Missspezifikationen angezeigt. Einzige Ausnahme findet sich unter $N = 150$, zehn Items und niedrigen Faktorladungen. Die moderate Ausprägung der Missspezifikation im Messmodell wird nur unter der Bedingung niedrige Faktorladungen und zehn Items bei einer Stichprobengröße von $N = 1000$ angezeigt. Die leichteste Ausprägung der Missspezifikationsbedingung im Messmodell wird unter keiner Bedingung detektiert. Hier ist wieder festzuhalten, dass der CFI_s somit nicht unabhängig von Gegebenheiten des Messmodells ist.

Zusammengefasst lässt sich festhalten, dass die dritte Missspezifikationsausprägung im Strukturmodell über alle Simulationsbedingungen hinweg und unabhängig von der Missspezifikationsausprägung im Messmodell von allen structural Fit-Indizes zuverlässig aufgezeigt wird. Der SRMR_s kann keine weitere Missspezifikationsausprägung anzeigen. Besonders zuverlässig zeigt sich der RMSEA_s, wobei die Ergebnisse den Einfluss der Gegebenheiten des Messmodells klar aufzeigen. Ebenso zeigen der CFI_s und der RMSEA-P den deutlichen Einfluss der Faktorladungen auf die correct rejection rates und damit die Abhängigkeit von nuisance parameters.

4 Diskussion des exakten und des approximativen Modellfits

4.1 Der exakte Modellfit – der Chi-Quadrat-Test

Zum exakten Modellfit durch den Chi-Quadrat-Test ist zu sagen, dass erwartungsgemäß die Power mit dem graduellen Anstieg der Missspezifikation unter der Bedingung eines perfekten Messmodellfits zunahm. Die Power des Chi-Quadrat-Tests bei der ausgeprägtesten Missspezifikation im Strukturmodell war unter allen Bedingungen als ausreichend zu bewerten. Außerdem verbesserte sich die Power des Chi-Quadrat-Tests in Abhängigkeit von der Testlänge, bei höheren Faktorladungen und mit steigender Stichprobengröße. Die Power des Chi-Quadrat-Tests war bei leichter und bei moderater Ausprägung der Missspezifikation nur dann durchgehend hinreichend (≥ 0.80), wenn hohe Faktorladungen und zehn Items pro latenter Variable vorlagen.

Bezüglich korrekt spezifizierter Modelle lässt sich sagen, dass das nominelle Alpha-Niveau von 5% bei kleinen Stichproben ($N < 1000$) nicht eingehalten wurde. Es lag unter diesen Bedingungen stets über bzw. unter 5%. Richtig spezifizierte Modelle wurden also fälschlicherweise zu oft abgelehnt.

Es lässt sich somit festhalten, dass der Chi-Quadrat-Test durchaus zuverlässige Ergebnisse lieferte, auch im Vergleich zu den Ergebnissen der composite Fit-Indizes. Allerdings war es entscheidend, dass die Daten möglichst reliabel über eine ausreichend große Anzahl manifester Variablen pro latenter Variable erhoben wurden. Vor allem bei kleineren Stichproben von $N \leq 250$ war eine hohe Itemanzahl entscheidend. Auch die Faktorladungshöhe hatte einen Einfluss und sollte daher möglichst hoch sein. Im Vergleich zu den composite Fit-Indizes, die nachfolgend besprochen werden, war der exakte Modellfit eher in der Lage, Missspezifikationen im Strukturmodell zuverlässig aufzuzeigen und konnte somit als zuverlässiger angesehen werden. Es soll jedoch auch darauf hingewiesen werden, dass durch den hohen beobachteten Alpha-Fehler bei kleinen Stichproben richtige Modelle als falsch abgelehnt wurden, weshalb erneut betont werden muss, dass die Stichprobengröße möglichst hoch sein soll.

4.2 Der approximative Modellfit – die composite Fit-Indizes

Der RMSEA mit einem Cutoff-Wert von $<.05$ konnte unter keiner Bedingung die Missspezifikation im Strukturmodell entdecken. Auch die stärkste Missspezifikation, bei der drei Pfade im Vergleich zum Populationsmodell fälschlicherweise auf null gesetzt worden waren, wurde nicht erkannt. Das bedeutet, dass auch dieses schwerwiegend misspezifizierte Modell nicht abgelehnt wurde. Die correct rejection rates fielen für alle Bedingungen und alle Ausprägungen der Missspezifikationen im Strukturmodell mit Werten nahe null viel zu niedrig aus. Somit lässt sich sagen, dass der RMSEA das misspezifizierte Modell unter keiner Missspezifikationsbedingung ablehnte. Diese äußerst schlechte Detektion der Missspezifikation lässt sich auch auf die hohe Anzahl der manifesten Variablen, die in unserer Studie verwendet wurden, zurückführen. So konnten Kenny und McCoach (2003) zeigen, dass eine hohe Anzahl an manifesten Variablen einen negativen Einfluss auf die Sensitivität der Missspezifikationsdetektion des RMSEA hat.

Die Ergebnisse des SRMR unter Verwendung des gängigen Cutoff-Werts von $<.08$ legten dar, dass der SRMR nur bei der stärksten Ausprägung der Missspezifikation und unter der Bedingung, dass hohe Faktorladungen in Kombination mit zehn manifesten Variablen gegeben waren, in der Lage war, die Missspezifikation anzuzeigen. Wenn man das auf die angewandte Forschung überträgt, muss man klar anerkennen, dass diese Bedingungen (hohe Faktorladungen und ausreichende Anzahl manifeste Variablen) in der Forschungspraxis wohl nur äußerst selten gegeben sind. Unter allen anderen Missspezifikationsbedingungen wies der SRMR keine hinreichende Sensitivität auf, Missspezifikationen im Strukturmodell aufzuzeigen.

Auch die Ergebnisse des CFI unter der Verwendung des Cutoff-Wertes $>.95$ zeigten, dass die Sensitivität vermeintlich mit steigendem Ausmaß der Missspezifikation im Strukturmodell stieg. Bei genauer Betrachtung zeigte sich bedenklicherweise, dass die Ergebnisse des CFI erst bei einer ausreichend großen Stichprobe als hinreichend zuverlässig angesehen werden konnten, da bei kleinen Stichproben ($N \leq 1000$) die nonzentrale Chi-Quadrat-Verteilung nicht hinreichend approximiert wurde. Dieses Ergebnis steht in Übereinstimmung mit Curran et al. (2002) und Yuan (2005). Somit konnten die Ergebnisse des CFI erst bei großen Stichproben ab $N \geq 1000$ zuverlässig interpretiert werden. Allerdings zeigte sich dann auch, dass Missspezifikationen im

Strukturteil nicht hinreichend erkannt werden konnten, unabhängig von der Ausprägung der Missspezifikation und über alle Simulationsbedingungen hinweg.

Zusammengefasst lässt sich sagen, dass die composite Fit-Indizes (RMSEA, CFI, SRMR) unter Verwendung der gängigen Cutoff-Werte nach Hu und Bentler (1999) für das in dieser Studie umgesetzte Strukturgleichungsmodell ungeeignet sind, Missspezifikationen im Strukturmodell anzuzeigen. Zudem sind sie stark von nuisance parameters abhängig, d.h. von Parametern, die nichts mit dem Ausmaß der Missspezifikation an sich zu tun haben, wie beispielsweise die Faktorladungshöhe oder die Itemanzahl. Die nicht hinreichende Detektion der Missspezifikationen im Strukturmodell durch die üblicherweise angewandten composite Fit-Indizes ist besonders folgenschwer, da genau im Strukturmodell die inhaltlichen Annahmen einer psychologischen Theorie über das Zusammenspiel oder den Zusammenhang von latenten Variablen abgebildet werden. Zudem zeigt die Literatur, dass die composite Fit-Indizes besonders häufig angewandt werden. Die vorliegenden Ergebnisse werfen somit Fragen nach der Validität bereits publizierter Studien mit – laut composite Fit-Indizes – mindestens hinreichendem Modellfit auf.

4.3 Der approximative Modellfit – die structural Fit-Indizes

Der RMSEA-P konnte die Missspezifikation nur in ihrer stärksten Ausprägung über alle Bedingungen hinweg und unabhängig von der Fehlerkorrelationsbedingung erkennen. Die correct rejection rates lagen hierbei alle über .80. Bei der leichten und der moderaten Ausprägung der Missspezifikationsausprägung im Strukturmodell und unabhängig von der Fehlerkorrelationsbedingung zeigte sich deutlich der Einfluss der Faktorladungen auf die Zuverlässigkeit der Missspezifikationsdetektion. Gerade bei der leichtesten Ausprägung der Missspezifikation im Strukturmodell und über alle Missspezifikationsbedingungen im Messmodell hinweg war der kombinierte Einfluss von Faktorladungen und Testlänge klar erkennbar. Weiters waren die Ergebnisse abhängig von der Missspezifikationsbedingung der Fehlerkorrelationen im Messmodell. Obwohl der RMSEA-P Missspezifikationen im Strukturmodell anzeigen soll, zeigte er sich dennoch auch sensitiv gegenüber Missspezifikationen im Messmodell, was als kritisch zu bewerten ist. Zudem zeigte sich, wie oben beschrieben, auch eine deutliche Sensitivität gegenüber nuisance parameters wie Faktorladungshöhe und Itemanzahl, vor allem bei leichteren Ausprägungen der Missspezifikationen.

Ähnlich dem RMSEA-P erwies sich auch der RMSEA_s mit correct rejection rates von ≥ 0.80 als hinreichend sensitiv gegenüber der stärksten Ausprägung der Missspezifikation mit drei misspezifizierten Pfaden im Strukturmodell. Aber der RMSEA_s konnte auch die leichte und die moderate Ausprägung der Missspezifikationen, kombiniert im Mess- und Strukturmodell, über alle Bedingungen hinweg und unabhängig von der Faktorladungshöhe, der Testlänge und der Stichprobengröße hinreichend gut anzeigen. Als problematisch ist zu bewerten, dass sich der RMSEA_s aber sensitiv gegenüber Missspezifikationen des Messmodells zeigte, sprich, wenn Missspezifikationen ausschließlich im Messmodell simuliert wurden, obwohl er sich eigentlich auf den Strukturteil des Strukturgleichungsmodells beziehen sollte.

Der SRMR_s konnte die Missspezifikation im Strukturmodell nur in ihrer stärksten Ausprägung mit einer correct rejection rate von ≥ 0.80 hinreichend gut detektieren. Die leichte und die moderate Ausprägung der Missspezifikation im Strukturmodell wurden unter keiner Bedingung hinreichend gut erkannt, unabhängig von Stichprobengröße, Itemanzahl, Faktorladungshöhe oder Ausprägung der Missspezifikation im Messmodell. Somit wurden die falsch spezifizierten Modelle bei den beiden ersten Ausprägungen der Missspezifikation im Strukturmodell durchgehend zu selten verworfen. Ähnliche Ergebnisse zeigten sich beim Betrachten der Missspezifikationsbedingungen der Fehlerkorrelationen im Messmodell. Der SRMR_s zeigte keine Missspezifikationen im Messmodell an und kann somit als nicht sensitiv gegenüber möglichen Missspezifikationen im Messmodell angesehen werden. Dies entspricht den Erwartungen, da der SRMR_s nicht sensitiv gegenüber möglichen Missspezifikationen im Messmodell sein soll.

Der CFI_s war in der Lage, die moderate bis schwere Ausprägung der Missspezifikation im Strukturmodell unabhängig von Gegebenheiten des Messmodells (Fehlerkorrelationsmissspezifikationsbedingung, Itemanzahl, Faktorladung) hinreichend gut anzuzeigen. Dennoch zeigten die Ergebnisse des CFI_s auch, dass dieser structural Fit-Index von nicht misspezifikationsrelevanten Parametern wie der Faktorladungshöhe abhängig war, vor allem bei leichteren Ausprägungen der Missspezifikation. Die Ergebnisse, wenn man die Missspezifikationsbedingungen im Messmodell ansieht, bestärken nur die Annahme, dass der CFI_s stark von nuisance parameters abhängig ist. Zudem reagierte der CFI_s sensitiv auf Missspezifikationen im Messmodell, vor allem wenn niedrige Faktorladungen vorlagen.

Betrachtet man die wrong rejection rates sowohl für die composite als auch die structural Fit-Indizes, so zeigt sich hauptbefundlich ein sehr inkonsistentes Bild bei allen Bedingungen. Allgemein lässt sich aber sagen, dass die wrong rejection rates ein gewünschtes Niveau von 5% äußerst selten einhielten.

Das neue Ergebnis dieser Studie ist, dass auch die structural Fit-Indizes, die explizit dafür entwickelt wurden, Missspezifikationen im Strukturmodell aufzuzeigen, nicht unabhängig von den Bedingungen des Messmodells waren. Dies zeigte sich bei allen verwendeten structural Fit-Indizes (RMSEA-P, RMSEA_s, SRMR_s und CFI_s).

Zusammengefasst lässt sich festhalten, dass die Performanz der structural Fit-Indizes besonders von Faktorladungen und teilweise von der Kombination Faktorladungen und Testlänge sowie von der Stichprobengröße, also von Gegebenheiten des Messmodells (sowohl nuisance parameters als auch Missspezifikationen im Messmodell), abhängig waren. Die stärkste Ausprägung der Missspezifikation im Strukturmodell konnte durchgehend aufgezeigt werden, leichte und moderate Missspezifikationsausprägungen im Strukturmodell wurden nicht durchgehend zuverlässig angezeigt und waren stark von nuisance parameters abhängig. Allein der RMSEA_s lieferte in Hinblick auf die correct rejection rates durchgehend zufriedenstellende Ergebnisse. Obwohl die structural Fit-Indizes besonders sensitiv gegenüber Missspezifikationen im Strukturmodell sein sollten, traf dies nur auf sehr schwere Ausprägungen der Missspezifikation im Strukturmodell zu. Leichte bis moderate Missspezifikationen blieben teilweise unerkannt oder waren stark von Faktorladungen abhängig und folgedessen wurden missspezifizierte Modelle als richtig angenommen.

Studie 2

5 Einleitung zu den differenzbasierten Fitmaßen

Gerade vor dem Hintergrund der insgesamt sehr kritischen Befunde der ersten Studie zu den composite Fit-Indizes wurden zwei weitere Ansätze der Modellüberprüfung als ergänzender Teil der bisher beschriebenen Simulationsstudie angewandt. Die in der ersten Studie genannten Ergebnisse werfen ja weiterhin die Frage nach alternativen Möglichkeiten, Missspezifikationen im Strukturmodell zuverlässig zu entdecken, auf. Hierzu wurde zum einen der two-step approach, basierend auf Anderson und Gerbing (1988), zum anderen der Δ CFI approach, basierend auf Widaman (1985), durchgeführt. Neu an dieser Studie ist, dass diese beiden Ansätze der Überprüfung des Modellfits erstmals unter Variation unterschiedlicher Bedingungen wie z.B. Testlänge oder Faktorladungshöhe auf ihre Zuverlässigkeit getestet wurden.

5.1 Der two-step approach von Anderson und Gerbing (1988)

Aufgrund des Problems, dass mögliche Missspezifikationen im Strukturmodell durch einen besseren oder passenden Fit im Messmodell überdeckt werden können, schlugen Anderson und Gerbing (1988) als weiteren Ansatz der Modelltestung ergänzend zum herkömmlichen one-step approach einen sogenannten two-step approach vor. Um diesen Ansatz zu verstehen, muss zuerst der one-step approach kurz beschrieben werden. Wie bereits bekannt, besteht jedes Strukturgleichungsmodell aus einem Mess- und einem Strukturmodell (Anderson & Gerbing, 1982; Joreskog & Sorbom, 1981). Die Parameter dieser beiden Teile des Strukturgleichungsmodells werden simultan geschätzt und die Diskrepanz zwischen empirischer und implizierter Kovarianzmatrix über den Chi-Quadrat-Test statistisch getestet. Der Chi-Quadrat-Test für dieses Gesamtmodell quantifiziert somit mögliche Diskrepanzen zwischen diesen Matrizen (u.a. Hoyle, 1995), die durch Missspezifikationen im Mess- und/oder Strukturmodell auftreten können. Im Vergleich dazu wird beim two-step approach die Differenz zwischen dem Chi-Quadrat-Wert des Messmodells und dem Chi-Quadrat-Wert des Gesamtmodells, bestehend aus Mess- und Strukturmodell, berechnet. Eine signifikante Differenz spricht dann, unter der Voraussetzung eines passenden Messmodells, für einen Missfit des Strukturmodells.

Dieser Ansatz setzt also voraus, dass man ein passendes Messmodell vorliegen hat. In der Praxis impliziert dies, dass das Messmodell bei einem signifikanten Missfit so lange respezifiziert werden muss, bis ein ausreichender Messmodellfit gegeben ist. Erst im Anschluss daran wird der two-step approach, wie oben skizziert, durchgeführt (Anderson & Gerbing, 1988). Die genaue methodische Erläuterung erfolgt im Abschnitt 6.

Diese getrennte Überprüfung des Modellfits für das Mess- und das Strukturmodell führte allerdings zu starker Kritik, wie z. B. durch Fornell und Yi (1992), Hayduk (1996) oder Mulaik und Millsap (2000). Dabei wurde vor allem kritisiert, dass durch diese Trennung nicht beachtet wird, dass die Frage nach dem Fit des Messmodells für die zu prüfende Theorie des Zusammenhangs zwischen latenten Variablen im Strukturmodell durchaus theoretisch relevant ist, gerade dann, wenn es um die Frage nach der Anzahl der zugrunde liegenden Faktoren geht. Die Ermittlung der korrekten Anzahl der Faktoren hat dann selbstverständlich direkte Auswirkungen auf den Fit des Strukturmodells bzw. darauf, wie die Zusammenhänge modelliert werden (Hayduk, 1996).

5.2 Der Δ CFI approach von Widaman (1985)

Der Δ CFI approach geht zurück auf Widaman (1985). Auch dieser Ansatz trennt nach Fit des Mess- und Strukturmodells und berechnet daher getrennte CFI-Werte für das composite Modell, also das Gesamtmodell, und für das Messmodell. Obwohl der Δ CFI approach von unterschiedlichsten Autoren verwendet wurde – unter anderen von Smith, Davy & Easterling (2004) oder Scheck, Kinicki & Davy (1995) –, weiß man bis dato nichts über dessen Sensitivität in Hinblick auf die Missspezifikationsdetektion.

6 Methode der differenzbasierten Fitmaße

Das methodische Vorgehen, die Modellannahmen sowie alle durchgeführten Simulationsbedingungen und die angenommenen Missspezifikationen im Strukturmodell wurden bereits ausführlich im allgemeinen Methodenteil erörtert und werden daher an dieser Stelle nicht erneut angeführt (siehe 2 Methode des simulierten Strukturgleichungsmodells). Diese differenzbasierten Fitmaße wurden nur für die Bedingungen, dass Missspezifikationen im Strukturmodell auftreten, berechnet. Das Messmodell war korrekt spezifiziert, d.h. ohne Fehlerkorrelationen, da dies auch die Basis der Ansätze darstellt. Im Folgenden soll nur auf die beiden zusätzlichen Ansätze, den two-step approach nach Anderson und Gerbing (1988) und den Δ CFI approach von Widaman (1985), in Hinblick auf die methodische Umsetzung eingegangen werden.

6.1 Die Methode des two-step approachs von Anderson und Gerbing (1988)

Wie oben beschrieben, wird beim two-step approach nach Anderson und Gerbing (1988) der Modellfit separat berechnet. Zum einen wird das composite Modell, also das gesamte Strukturgleichungsmodell, bestehend aus Mess- und Strukturmodell, auf seinen Modellfit hin überprüft, zum anderen das Messmodell. Hierfür werden dann Chi-Quadrat-Werte gesondert berechnet. Das Messmodell ist eine Teilmenge des composite Modells, weshalb man hier von einem nested model spricht. Infolgedessen kann ein wiederum chi-quadrat-verteilter Differenzwert der beiden Chi-Quadrat-Werte wie folgt berechnet werden (siehe hierzu Steiger, Shapiro & Browne, 1985):

$$\chi_{diff}^2 = \chi_{composite\ model}^2 - \chi_{measurement\ model}^2$$

$$df_{diff} = df_{composite\ model} - df_{measurement\ model}$$

Wenn nun dieser Chi-Quadrat-Differenzwert für df_{diff} signifikant ist, unter der Voraussetzung, dass das Messmodell einen inferenzstatistisch passenden Fit aufweist, so legt dies nahe, dass es einen Missfit im Strukturmodell gibt. Dies impliziert, dass

mindestens eine Restriktion im Strukturmodell misspezifiziert ist und das Strukturmodell somit abgelehnt werden müsste.

6.2 Die Methode des ΔCFI approachs von Widaman (1985)

Der ΔCFI approach wird genauso wie der two-step approach (Anderson und Gerbing; 1988) bei nested models, sprich bei ineinander verschachtelten Modellen eingesetzt. Dabei werden CFI-Werte für das Gesamtmodell und das Messmodell separat berechnet, und es ergibt sich folgende Definition:

$$\Delta CFI = CFI_{\text{composite model}} - CFI_{\text{measurement model}}$$

Widaman (1985) schlägt dabei einen Cutoff-Wert von $\geq .01$ für die Differenz vor. Wenn die Differenz diesen Grenzwert von .01 überschreitet, bedeutet dies, dass eine Missspezifikation vorliegt und das Modell somit abgelehnt werden muss. Auch dieser Ansatz setzt voraus, dass ein passendes Messmodell vorliegt.

7 Ergebnisse der differenzbasierten Fitmaße

7.1 Die Ergebnisse des two-step approachs von Anderson und Gerbing (1988)

Die Ergebnisse des two-step approachs nach Anderson und Gerbing (1988) in Tabelle 15 zeigen, dass die dritte und somit ausgeprägteste Missspezifikation im Strukturmodell über alle Bedingungen hinweg hinreichend gut angezeigt werden kann. Die Powerraten liegen hier zwischen .975 (unter der Bedingung $N = 150$, fünf Items und niedrige Faktorladungen) und 1 bei allen anderen Bedingungen..

Die Ergebnisse bei der moderaten Ausprägung der Missspezifikation im Strukturmodell mit insgesamt zwei auf null gesetzten Pfaden zeigen, dass die Faktorladungen bei kleinen Stichproben ($N \leq 250$) entscheidend sind. Unter der kleinsten Stichprobenbedingung mit $N = 150$ können nur die Bedingungen hohe Faktorladungen und fünf Items (Powerraten: .874) bzw. zehn Items (Powerraten: .992) zuverlässig angezeigt werden. Ab einer Stichprobengröße von $N \geq 250$ zeigt sich, dass alle Missspezifikationen zuverlässig detektiert werden können. Die einzige Ausnahme findet sich erneut bei niedrigen Faktorladungen, fünf Items und $N = 250$ mit einer Powerrate von .391, was als nicht ausreichend zu bewerten ist.

Die erste und leichteste Ausprägung der Missspezifikation im Strukturmodell kann bei einer Stichprobengröße von $N = 150$ nur unter der Bedingung hohe Faktorladungen und zehn Items mit einer Powerrate von .936 angezeigt werden. Alle anderen Bedingungen bei einer Stichprobengröße von $N \geq 250$ können durchgehend detektiert werden. Die einzigen Ausnahmen findet sich erneut unter der Bedingung niedrige Faktorladungen, fünf Items und $N = 250$ mit einer Powerrate von .247, sowie unter $N = 250$, zehn Items und niedrige Faktorladungen mit einer Powerrate von .684.

Tabelle 15

Powerraten und Alpha-Fehlerraten des two-step approachs

<i>N</i>	Item- anzahl	Faktor- ladungs- höhe	Anzahl der Missspezifikationen im Strukturmodell			
			0	1	2	3
150	5	niedrig	.019 (2.975;2.293)	.133 (6.336;4.071)	.223 (8.901;5.174)	.975 (33.719;1.535)
	5	hoch	.035 (3.939;2.600)	.657 (14.670;7.091)	.874 (23.028;9.361)	1 (88.393;17.699)
	10	niedrig	.029 (3.683;2.410)	.421 (1.711;5.439)	.649 (15.660;7.026)	1 (66.798;15.594)
	10	hoch	.051 (4.161;2.843)	.936 (23.909;9.192)	.992 (38.974;12.804)	1 (127.779;2.446)
250	5	niedrig	.021 (3.258;2.354)	.247 (8.596;4.791)	.391 (12.126;5.715)	1 (55.919;14.602)
	5	hoch	.040 (4.060;2.764)	.889 (21.315;8.638)	.987 (34.528;11.602)	1 (144.205;22.818)
	10	niedrig	.041 (3.980;2.744)	.684 (15.243;7.017)	.902 (23.376;9.077)	1 (108.957;19.976)
	10	hoch	.056 (4.058;2.848)	.999 (36.642;11.617)	1 (6.917;16.233)	1 (204.259;26.129)
1000	5	niedrig	.046 (3.776;2.675)	.901 (21.804;8.970)	.994 (36.064;11.758)	1 (212.540;29.930)
	5	hoch	.074 (4.236;3.165)	1 (71.295;16.405)	1 (122.889;22.619)	1 (556.827;46.526)
	10	niedrig	.054 (3.939;2.694)	1 (47.088;13.352)	1 (77.712;18.001)	1 (416.159;39.241)
	10	hoch	.043 (3.939;2.729)	1 (132.362;22.046)	1 (224.730;32.518)	1 (81.156;51.251)

Der beobachtete Alpha-Fehler der χ^2_{diff} liegt unter allen Bedingungen durchgehend nahe beim nominellen Alpha-Niveau von .05 (siehe Spalte „0“). So liegt der beobachtete Alpha-Fehler zwischen dem Minimum .019 unter der Bedingung niedrige Faktorladungen, fünf Items und einer Stichprobengröße von $N = 150$ und dem Maximum .074 unter der Bedingung hohe Faktorladungen, fünf Items und einer Stichprobengröße von $N = 1000$. Das nominelle Alpha-Fehler-Niveau wird somit hinreichend gut eingehalten.

7.2 Die Ergebnisse des ΔCFI approaches von Widaman (1985)

Die Ergebnisse, in Tabelle 16 dargestellt, zeigen klar, dass die am stärksten ausgeprägte Missspezifikation im Strukturmodell unter allen Bedingungen hinreichend gut angezeigt werden kann. Alle correct rejection rates liegen zwischen .897 (unter der Bedingung niedrige Faktorladungen, zehn Items und $N = 150$) und 1 über alle anderen Bedingungen hinweg.

Tabelle 16

Correct rejection rates und wrong rejection rates des ΔCFI approaches

N	Item- anzahl	Faktor- ladungs- höhe	Anzahl der Missspezifikationen im Strukturmodell			
			0	1	2	3
150	5	niedrig	.658 (.051;.084)	.838 (.100;.098)	.942 (.151;.105)	.995 (.235;.099)
	5	hoch	.511 (.022;.044)	.941 (.08;.058)	.997 (.127;.066)	1 (.213;.061)
	10	niedrig	.100 (-.087;.072)	.276 (-.028;.084)	.445 (.015;.088)	.897 (.107;.082)
	10	hoch	.014 (-.049;.022)	.437 (.009;.035)	.883 (.055;.042)	1 (.154;.040)
250	5	niedrig	.785 (.062;.075)	.961 (.114;.083)	.991 (.166;.092)	1 (.251;.086)
	5	hoch	.476 (.015;.029)	.986 (.076;.047)	1 (.121;.052)	1 (.207;.045)
	10	niedrig	.251 (-.007;.061)	.698 (.049;.069)	.930(.098;.075)	1 (.188;.068)
	10	hoch	.024 (-.014;.011)	.933 (.043;.023)	.999 (.088;.029)	1 (.188;.029)
1000	5	niedrig	.649 (.026;.037)	.998 (.088;.054)	1 (.140;.062)	1 (.220;.050)
	5	hoch	.080(.004;.005)	1 (.060;.017)	1 (.104;.020)	1 (.192;.018)
	10	niedrig	.186 (.005;.011)	1 (.064;.026)	1 (.110;.032)	1 (.198;.026)
	10	hoch	0 (0;.002)	1 (.056;.010)	1 (.100;.014)	1 (.199;.013)

Die moderate Form der Missspezifikationsausprägung kann über alle Bedingungen hinweg hinreichend gut detektiert werden, mit Ausnahme der Bedingung, dass zehn Items und niedrige Faktorladungen bei einer Stichprobengröße von $N = 150$ vorliegen. Hier liegt die correct rejection rate bei .445. Unter allen anderen Bedingungen liegen die correct rejection rates zwischen .883 und 1 und sind somit ausreichend.

Die leichteste Ausprägung der Missspezifikation kann ebenso durchgehend angezeigt werden. Die einzigen Ausnahmen finden sich unter der Bedingung zehn Items, $N = 150$ und niedrige Faktorladungen (correct rejection rate: .276) sowie hohe Faktorladungen (correct rejection rate: .437) und unter niedrige Faktorladungen, zehn Items und $N = 250$ (correct rejection rate: .698)

Die wrong rejection rates in Spalte „0“, die in Anlehnung an die Inferenzstatistik bei rund 5% (im Sinne eines Alpha-Fehlers) liegen sollen, liegen zwischen dem Minimum 0 und dem Maximum .785 und somit weit außerhalb des angestrebten Bereichs. Da nur wenige Werte nahe 5% liegen, müssen die wrong rejection rates als nicht zufriedenstellend bezeichnet werden.

8 Diskussion der differenzbasierten Fitmaße

8.1 Der two-step approach von Anderson und Gerbing (1988)

Die Ergebnisse des two-step approachs nach Anderson und Gerbing (1988) zeigten insgesamt bei der am stärksten ausgeprägten Missspezifikation im Strukturmodell über alle Bedingungen hinweg zuverlässig an, dass eine Missspezifikation im Strukturmodell vorhanden und dieses somit abzulehnen war. Es zeigte sich klar, ähnlich wie beim herkömmlichen Chi-Quadrat-Test, dass mit dem graduellen Anstieg der Missspezifikation die Power der χ^2_{diff} zunahm. Zudem, ebenso ähnlich wie beim Chi-Quadrat-Test, zeigte sich deutlich die Abhängigkeit der Power von Faktorladungshöhe, Anzahl der manifesten Variablen und Stichprobengröße. Die Power war deutlich besser, wenn hohe Faktorladungen vorlagen, und auch ein Anstieg der Stichprobe begünstigte die Power. Dieses Phänomen entsteht dadurch, dass bei hohen Faktorladungen auch die Chi-Quadrat-Werte des Messmodells steigen und daher die Differenz der Chi-Quadrat-Werte geringer wird (Heene et al., 2011). Gerade bei der ersten Ausprägung der Missspezifikation waren bei kleinen Stichproben ($N \leq 250$) die Faktorladungen in Kombination mit der Itemanzahl relevant, um eine hinreichende Missspezifikationsdetektion gewährleisten zu können.

Sehr positiv zu betonen ist, dass der two-step approach das nominelle Alpha-Niveau von 5% sehr gut einhielt, was beinhaltet, dass richtig spezifizierte Modelle selten als falsch abgelehnt wurden.

Zusammengefasst lässt sich sagen, dass sich der two-step approach dann eignet, wenn man ein bereits passendes Messmodell entwickelt hat. In diesem Fall empfiehlt sich der two-step approach durchaus, um die angenommenen Restriktionen im Strukturmodell zu überprüfen. Im Vergleich zu den vorher beschriebenen composite und teilweise auch structural Fit-Indizes ist auf jeden Fall der two-step approach vorzuziehen, da dessen Ergebnisse eindeutig zuverlässiger sind, wenngleich auch hier nuisance parameters nach wie vor eine Rolle in Hinblick auf eine zuverlässige Missspezifikationsdetektion spielen.

Aufgrund der bereits in der Einleitung zu den differenzbasierten Fitmaßen genannten Kritik (cf. Fornell & Yi, 1992; Hayduk, 1996; Hayduk & Glaser, 2000) an diesem Ansatz soll

auch an dieser Stelle nochmals darauf hingewiesen werden, dass dieser Ansatz nur unter der strengen Annahme eines passenden Messmodells in der Praxis anwendbar ist. Zudem kann die Struktur des Messmodells eine hohe theoretische Relevanz haben, insbesondere dann, wenn es um die Dimensionalität von Fragebögen geht. Angenommen, ein wichtiger Bestandteil der aufgestellten Hypothese ist die Anzahl der latenten Variablen, die den Indikatoren zugrunde liegt. Dies betrifft dann das gesamte Strukturgleichungsmodell und eine Trennung scheint wenig sinnvoll, wenn es um die Überprüfung des Modellfits geht (Hayduk, 1996). Zudem führen Fornell und Yi (1992) an, dass man das Messmodell nicht unabhängig von der zugrunde liegenden Theorie, welche im Strukturmodell abgebildet ist, betrachten kann. Auch hier wird argumentiert, dass Veränderungen im Strukturmodell einen Einfluss auf die Parameter im Messmodell haben (Fornell & Yi, 1992). Diese Kritikpunkte schränken die Anwendbarkeit der beiden Methoden ein und sollten in der angewandten Forschung stark berücksichtigt werden.

8.2 Der Δ CFI approach von Widaman (1985)

Die Ergebnisse des Δ CFI approachs zeigten, dass die stärkste und die moderate Missspezifikationsausprägung bis auf eine Ausnahme ($N = 150$; niedrige Faktorladungen bei zehn Items pro latenter Variable) durchgehend zuverlässig detektiert werden konnten. Auch hier zeigte sich, dass die correct rejection rates mit steigender Ausprägung der Missspezifikationsbedingung stiegen und ebenso von den Faktorladungshöhen abhängig waren. Der Einfluss der Stichprobengröße auf die correct rejection rates zeigte sich vor allem bei der leichtesten Ausprägung der Missspezifikation im Strukturmodell. Der Einfluss der Faktorladungen verschwand zunehmend mit steigender Stichprobengröße.

Zusammengefasst lässt sich sagen, dass der Δ CFI approach durchaus zuverlässige Ergebnisse lieferte und die Missspezifikation im Strukturmodell bei hohen Faktorladungen mit hoher Wahrscheinlichkeit anzeigen konnte. Das nominelle Alpha-Niveau von 5% wurde allerdings nicht hinreichend gut eingehalten.

Auch hier trifft die Kritik am Ansatz von Anderson und Gerbing (1988) zu, sodass die Aussagekraft der separaten Fit-Beurteilung von Mess- und Strukturmodell nur bei hinreichend gut validierten und entsprechend passenden Messmodellen gegeben ist (vgl. Hayduk, 1996).

9 Allgemeine Diskussion der durchgeführten Simulationsstudien

9.1 Allgemeine Diskussion

In der vorliegenden Studie wurde der Frage nachgegangen, inwieweit klassische und neuere Ansätze der Modellmissfitbestimmung zuverlässige Ergebnisse liefern können, besonders in Bezug auf die Entdeckung von Missspezifikationen unter variierenden Bedingungen.

Zum einen wurden dafür klassische Verfahren wie der Chi-Quadrat-Test und die composite Fit-Indizes (CFI, RMSEA, SRMR) herangezogen, zum anderen strukturelle Varianten der Fit-Indizes (RMSEA-P, RMSEA_s, SRMR_s und CFI_s) sowie auch differenzbasierte Fitmaße (two-step approach nach Anderson und Gerbing (1988) und der Δ CFI approach nach Widaman (1985)). Diese Modellfitmaße wurden auf ihre Zuverlässigkeit hin geprüft, insbesondere, wie gut simulierte Missspezifikationen im Strukturteil des Strukturgleichungsmodells angezeigt werden können.

Dazu wurde ein für den sozialwissenschaftlichen Kontext typisches Strukturgleichungsmodell definiert. Stichprobengröße, Höhe der Faktorladungen sowie Testlänge wurden bei unterschiedlich schweren Ausprägungen der Missspezifikationen sowohl im Struktur- als auch im Messmodell systematisch variiert. Einerseits wurden für den Chi-Quadrat-Test bzw. für den two-step approach die Alpha-Fehlerraten und die Power berechnet, andererseits die correct und wrong rejection rates für die Fit-Indizes und den Δ CFI approach.

Die Ergebnisse zeigten, dass die herkömmlichen composite Fit-Indizes unter der Verwendung der gängigen Cutoff-Werte basierend auf Hu und Bentler (1999) nicht in der Lage waren, Missspezifikationen im Strukturmodell hinreichend zu erkennen.

Die structural Fit-Indizes konnten durchgehend die stärkste Ausprägung der Missspezifikation im Strukturmodell (insgesamt drei relevante Pfade auf null gesetzt) zuverlässig anzeigen. Bei der leichten bis moderaten Ausprägung der Missspezifikation zeigte sich bei den structural Fit-Indizes deutlich eine Abhängigkeit von nuisance parameters (u.a. Faktorladungshöhe, Anzahl der manifesten Variablen) sowie die

Abhängigkeit von Gegebenheiten des Messmodells (u. a. die Sensitivität auch auf Missspezifikationsausprägungen im Messmodell).

Der exakte Modellfit lieferte durchaus akzeptable Ergebnisse, jedoch müssen gerade bei der leichten und der moderaten Ausprägung der Missspezifikationen im Strukturmodell die Faktorladungen hoch sein und eine hohe Itemanzahl muss den latenten Variablen zugrunde liegen, um die Missspezifikation im Strukturmodell zu erkennen. Zudem sollte die Stichprobengröße möglichst groß sein, um eine zuverlässige Missspezifikationsdetektion gewährleisten zu können.

Die Ergebnisse der differenzbasierten Fitmaße (two-step approach und Δ CFI approach) waren über alle Simulationsbedingungen hinweg durchaus zufriedenstellend, allerdings unter der strengen Voraussetzung, dass das Messmodell exakt passt. Ein theoretisch und psychometrisch gut fundiertes Messmodell bietet daher die Voraussetzung, dass Missspezifikationen im Strukturmodell als solche erkannt werden.

In Hinblick auf die Fragestellung nach der Zuverlässigkeit klassischer und neuerer Modellfitmaße zeigte sich klar, dass die üblicherweise angewandten composite Fit-Indizes unter Verwendung gängiger Cutoff-Werte unzureichend waren, aber auch die structural Fit-Indizes teilweise hinter den Erwartungen blieben. Sie eigneten sich durchgehend dafür, eine starke Missspezifikationsausprägung im Strukturmodell anzuzeigen. Es zeigte sich darüber hinaus, dass, entgegen den Annahmen von McDonald und Ho (2002), Williams und O'Boyle (2011) sowie Hancock und Mueller (2011), die structural Fit-Indizes nicht unabhängig von Bedingungen des Messmodells (wie unter anderem Itemanzahl und Faktorladungshöhe) waren. Es erwies sich, dass die structural Fit-Indizes auch sensitiv auf Missspezifikationen im Messmodell reagieren. Dennoch sind die structural Fit-Indizes in jedem Fall den composite Fit-Indizes vorzuziehen.

Als vielversprechend in Hinblick auf die zuverlässige Detektion von Missspezifikationen im Strukturmodell erwiesen sich sowohl der globale Chi-Quadrat-Test als auch die differenzbasierten Fitmaße. Die differenzbasierten Fitmaße unterliegen der vorher genannten Einschränkung, dass ein exakt passendes Messmodell vorhanden sein muss, und ihre Sensitivität, Missspezifikationen zu entdecken, ist darüber hinaus abhängig von Stichprobengröße, Testlänge und Faktorladungshöhe.

9.2 Einschränkungen der durchgeführten Simulationsstudien

Die Ergebnisse der Simulationsstudien lassen sich nicht ohne Einschränkungen für andere Strukturgleichungsmodelle generalisieren. Zwar waren die Daten, die in der vorliegenden Studie verwendet wurden, kontinuierlich und multivariat, um eine Vergleichbarkeit mit anderen Simulationsstudien zu gewährleisten, allerdings liegen in der Praxis allermeistens kategoriale Daten z. B. aus Fragebögen vor. Dies schränkt die Generalisierbarkeit dieser Studie für empirische Daten ein. Zukünftige Studien sollten diesen Aspekt beachten.

Ein anderer einschränkender Punkt ist, dass in der vorliegenden Simulationsstudie davon ausgegangen wurde, dass die latenten Variablen immer mit der gleichen Anzahl an manifesten Variablen über alle latenten Variablen hinweg erhoben wurden. Zudem wurden nur zwei Bedingungen für die Testlänge simuliert (fünf oder zehn Items pro latenter Variable). Dies stellt insofern eine Einschränkung dar, als in der angewandten Forschung den latenten Variablen nicht immer die gleiche Anzahl manifester Variablen zu Grunde liegt und auch oft mehr oder weniger Items pro latenter Variable verwendet werden.

Eine weitere Einschränkung stellt die Wahl der Missspezifikation im Strukturmodell dar. In den vorliegenden Studien wurden kausale Beziehungen zwischen bestimmten latenten Variablen als in der Population nicht vorhanden angenommen, indem einzelne Pfade sukzessive auf null gesetzt wurden. Dies war eine sehr spezifische Wahl der Missspezifikationsbedingungen im Strukturmodell, die jedoch in der angewandten Forschung als durchaus realistisch und für psychologische Theorien besonders relevant zu bewerten ist. Alternativ ließen sich auch Pfade auf Werte ungleich null fixieren, etwa, wenn bestimmte Annahmen aus Vorstudien bekannt wären. Dieses Vorgehen ist aber nach Kenntnis der Autorin sehr selten in der Literatur zu finden. In der Praxis kann man davon ausgehen, dass es zu noch gravierenderen Missspezifikationen wie z. B. omitted variables im Strukturmodell kommen kann, da man die tatsächliche Datenstruktur der Population bekanntlich nicht kennt (s. u.a. Tomarken & Waller, 2003; Yuan, Marshall & Bentler, 2003). Dennoch scheint es, dass der exakte Modelltest und die structural Fit-Indizes gerade schwerwiegende Missspezifikationen im Strukturmodell zuverlässig detektieren können, wenn eine ausreichend große Stichprobe und reliable Daten zur Verfügung stehen.

Ein Punkt, der für die Annahmen der Simulationsstudien spricht, ist die Modellkomplexität. Das Modell bestand aus insgesamt sechs latenten Variablen und bildete einen doppelten Mediationseffekt ab. Die Literatur zeigt, dass so komplexe Modelle sehr häufig angenommen werden. Dennoch bleibt die Frage offen, ob die composite bzw. die structural Fit-Indizes Missspezifikationen im Strukturmodell bei einem inhaltlich einfacheren Modell besser anzeigen würden als bei einem komplexen. Hinsichtlich der Ergebnisse scheint es jedoch generalisierbar zu sein, dass der exakte Modelltest bei einer ausreichend großen Stichprobe am teststärksten ist.

Positiv hervorzuheben an der durchgeführten Simulationsstudie ist, dass nicht nur die herkömmlichen composite Fit-Indizes mit den gängigen Cutoff-Werten auf deren Sensitivität geprüft wurden, sondern der Fokus auf den structural Fit-Indizes lag. Weiters wurde auch der Einfluss der Missspezifikationen von Fehlerkorrelationen im Messmodell untersucht, wie sie in der angewandten Forschung häufig vorkommen (s. u.a. Cole, Ciesla & Steiger, 2007). Hier zeigte sich, dass der RMSEA_s und der CFI_s in komplexer Weise sensitiv auf Fehlerkorrelationsbedingungen im Messmodell reagierten, obwohl ihr Fokus auf dem Strukturteil liegen sollte. Allein der RMSEA-P und der SRMR_s zeigten keinerlei Missspezifikationen im Messmodell an.

9.3 Implikation für die Praxis

Die vorliegende Simulationsstudie zeigten einmal mehr auf, dass die Beurteilung des Modellfits mittels gängiger Cutoff-Werte (Hu & Bentler, 1999) für die häufig verwendeten composite Fit-Indizes nicht zuverlässig ist. Die übergeneralisierte Verwendung der vorgeschlagenen Cutoff-Werte von Hu und Bentler (1999) wurde bereits in den letzten 15 Jahren mehrfach untersucht und stark kritisiert (cf. Fan & Sivo, 2005, 2007; Hayduk, Cummings, Boadu, Pazderka-Robinson & Boulianne, 2007; Heene et al., 2011, 2012; Marsh et al., 2004; Saris et al., 2009; Steiger, 2007; Tomarken & Waller, 2003; Yuan, 2005). Auch diese Simulationsstudie bestätigt, dass die Modellfitüberprüfung mittels composite Fit-Indizes nicht hinreichend zuverlässig funktioniert.

Aus den Schlussfolgerungen ergeben sich direkte Implikationen für die angewandte Forschung. Es liegt aufgrund der enttäuschenden Ergebnisse der Missspezifikationsdetektion mittels composite Fit-Indizes nahe anzunehmen, dass viele der bisher publizierten Strukturgleichungsmodelle fälschlicherweise als richtig angenommen wurden, was die folgende Vermutung von McDonald und Ho (2002)

bestärkt: „*We tentatively conclude that in the majority of studies (...) the goodness of approximation of the path model may be unacceptable, contrary to the published conclusions*“.

Die weiterhin hohe Zitationshäufigkeit des Artikels von Hu und Bentler (1999) zu den Cutoff-Werten in der angewandten Forschung (laut Google Scholar 42780-mal zitiert, Stand: September 2017) wirft weitreichende Fragen zur Validität von Theorien und Hypothesen auf, die mittels Strukturgleichungsmodellen überprüft wurden.

Nun könnte man annehmen, dass die structural Fit-Indizes, die sich speziell auf eine mögliche Missspezifikation im Strukturmodell beziehen, inhaltlich falsche Annahmen bzw. Missspezifikationen im Strukturmodell besser aufzeigen können (Hancock und Mueller, 2011; McDonald und Ho, 2002). Diese Fit-Indizes wurden bislang noch nicht hinreichend unter den unterschiedlichsten Bedingungen (z.B. Testlänge, Ladungshöhen und Itemanzahl) wie in den vorliegenden Studien getestet. Die Ergebnisse zeigten allerdings, dass, obwohl der Fokus auf Missspezifikationen im Strukturmodell lag, die structural Fit-Indizes nur die am stärksten ausgeprägte Missspezifikation im Strukturmodell durchgehend zuverlässig erkennen konnten. Allein der RMSEA_s lieferte ausnahmslos zufriedenstellende Ergebnisse. Bei allen anderen structural Fit-Indizes zeigten sich insofern Einschränkungen, als die Zuverlässigkeit der Missspezifikationsdetektion stark von den Faktorladungshöhen abhängig war. Geringere, aber dennoch inhaltlich schwerwiegende Missspezifikationen, wie etwa ein fehlender gerichteter Pfadkoeffizient von erheblicher Höhe wie in diesen Studien, bleiben aller Wahrscheinlichkeit nach auch unter Verwendung der structural Fit-Indizes bei nicht ausreichend hohen Faktorladungen und einer nicht ausreichend großen Itemanzahl in der angewandten Praxis unentdeckt.

Zusammengefasst lässt sich bezüglich des auf den Fit-Indizes basierenden Ansatzes Folgendes festhalten: Auch wenn die Fit-Indizes einen ausreichend guten Modellfit bestätigten, zeigte die vorliegende Studie, dass vor allem bei einer geringen Itemanzahl und niedrigen Faktorladungen fehlspezifizierte Modelle zu selten verworfen wurden. Besonders unzufriedenstellende Ergebnisse lieferten die composite Fit-Indizes, aber auch die structural Fit-Indizes konnten nicht alle Missspezifikationsbedingungen im Strukturmodell zufriedenstellend anzeigen. Dies wiederum beinhaltet, dass angenommene kausale Zusammenhänge, die mittels empirischer Daten geprüft werden,

statistisch nicht hinreichend untermauerbar sind. Es reicht folglich nicht aus, den Modellfit nur über Fit-Indizes zu prüfen, vielmehr sollte man den Fokus wieder auf den exakten Modellfit legen und im Anschluss eventuell die structural Fit-Indizes berechnen.

Wenn ein signifikanter Unterschied mittels exakter Modellfitüberprüfung festgestellt wird, muss sich der Anwender selbstverständlich fragen, *wo* die Missspezifikation liegt. Hierzu eignet sich bislang noch am ehesten der Ansatz der „misspecification search“ von Saris, Satorra und Van der Veld (2009). Dabei wird versucht, über eine A-priori-Poweranalyse für Modifikationsindizes den Modellfit ausreichend zu verbessern. Vorab wird hierbei eine inhaltlich relevante Modellabweichung (etwa standardisierter Pfad $>.10$, Fehlerkorrelation $>.10$) spezifiziert, die bei einer Power von mindestens 80% über die Modifikationsindizes entdeckt werden kann.

Weiterhin konnten die vorliegenden Studien auch zeigen, dass, wenn man den Pfad der Modellprüfung mittels Fit-Indizes verlässt, der two-step approach von Anderson und Gerbing (1988) durchaus reliable Ergebnisse liefert. Er könnte sich in der angewandten Forschung vor allem dann als akzeptables Mittel der Modellüberprüfung erweisen, wenn bereits ein hinreichend gut validiertes Messmodell vorliegt. Dennoch sollte auch bei diesem Ansatz die Abhängigkeit der Performanz von der Anzahl der manifesten Variablen und den Faktorladungshöhen bedacht werden. Daher ist es wichtig, bei diesen Ansätzen den Fokus auf das Messmodell zu richten, da hier Einflussmöglichkeiten in Hinblick auf Itemanzahl und Stichprobengröße bestehen.

Allgemein – über die unterschiedlichen Prüfgrößen hinweg – lässt sich sagen, dass eine möglichst große Stichprobe zur Bestimmung des Modellfits herangezogen werden sollte. Eine zu kleine Stichprobe kann dazu führen, dass die statistische Grundlage einer hinreichenden Approximation sowohl der zentralen als auch der nonzentralen Chi-Quadrat-Verteilung nicht gewährleistet ist, was zu einer Verzerrung der Ergebnisse und zu fälschlichen Ablehnungen und Annahmen von Modellen bei Verwendung von auf den Chi-Quadrat basierenden Maßen führen kann. Zudem sollten möglichst hohe Faktorladungen vorliegen und die Daten möglichst über eine ausreichend große Anzahl an manifesten Variablen erhoben werden.

Aufgrund der vorliegenden Studie lässt sich sagen, dass die Ergebnisse des Chi-Quadrat-Tests als durchaus reliabel angesehen werden können. Die structural Fit-Indizes können unter bestimmten idealen Bedingungen Aufschluss über mögliche Missspezifikationen

geben, die composite Fit-Indizes allerdings sind laut der vorliegenden Studie kaum geeignet, Misspezifikationen im Strukturmodell anzuzeigen. In Übereinstimmung mit zahlreichen früheren Studien zeigt diese Studie auch, dass man sich nicht auf die gängigen Cutoff-Werte der Fit-Indizes verlassen kann, weil sie nicht für verschiedene Modellstrukturen verallgemeinerbar sind und zudem ihre Sensitivität zu sehr von Randbedingungen abhängt, die nichts mit der Misspezifikation eines Modells an sich zu tun haben. Diese Aussage trifft sowohl auf die klassischen composite Fit-Indizes als auch auf ihre strukturellen Gegenstücke zu. Insgesamt betrachtet, scheint sowohl die generalisierte Verwendbarkeit der vorgeschlagenen Cutoff-Werte als auch die überzeugende Verwendbarkeit der composite Fit-Indizes und teilweise sogar der structural Fit-Indizes in der sozialwissenschaftlichen Forschung einen Mythos darzustellen und deren Einsatz eher zu schaden als zu nutzen. Ihre weiterhin weitverbreitete Anwendung, vor allem die der composite Fit-Indizes, ist äußerst fragwürdig und sollte daher nicht weiter unterstützt werden. Womöglich liefert das folgende Zitat von Schönemann (1981a) einen Grund, warum sie trotz seit langer Zeit bestehender Gegenevidenz weiter verwendet werden:

“The uncritical acceptance of such myths is widespread, because they are convenient and ennobling. They are convenient because they promise a mechanical device for conducting research. They are ennobling because they lend the mantle of scientific legitimacy to unproven conjectures.” (S. 354)

Literaturverzeichnis

- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*(3), 411–423.
- Beauducel, A., & Wittmann, W. W. (2005). Simulation study on fit indexes in CFA based on data with slightly distorted simple structure. *Structural Equation Modeling*, *12*, 41–75. doi: 10.1207/s15328007sem1201_3.
- Bentler, P. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246.
- Bentler, P., & Yuan, K. H. (1999). Structural equation modeling with small samples: Test statistics. *Multivariate Behavioral Research*, *34*(2), 181–197. <http://doi.org/10.1207/S15327906Mb340203>
- Borkenau, P., & Ostendorf, F. (1993). NEO-Fünf-Faktoren Inventar (NEO-FFI) nach Costa und McCrae [NEO Five-Factor Personality Inventory (NEO-FFI) according to Costa and McCrae]. *Göttingen: Hogrefe*.
- Breivik, E., & Olsson, U. H. (2001). Adding variables to improve fit: The effect of model size on fit assessment in LISREL. In R. Cudeck, S. Du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future: A Festschrift in honour of Karl Jöreskog* (pp. 169–194). Chicago: Scientific Software International.
- Browne, M. W., MacCallum, R. C., Kim, C. T., Andersen, B. L., & Glaser, R. (2002). When fit indices and residuals are incompatible. *Psychol. Methods*, *7*(4), 403–421. <http://doi.org/10.1037//1082-989x.7.4.403>
- Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). *An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models*. *Sociological Methods & Research* Vol 36(4) May 2008, 462–494.
- Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in structural equation models: causes, consequences, and strategies. *Sociological Methods & Research*, *29*(4), 468–508.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.
- Cole, D. A., Ciesla, J. A., & Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods*, *12*(4), 381–398. <https://doi.org/10.1037/1082-989X.12.4.381>

- Costa, P. T., & McCrae, R. R. (1985). The NEO personality inventory.
- Curran, P. J., Bollen, K. A., Paxton, P., Kirby, J., & Chen, F. (2002). The noncentral chi-square distribution in misspecified structural equation models: Finite sample results from a Monte Carlo simulation. *Multivariate Behavioral Research*, 37(1), 1–36.
- Driel, O. P. (1978). On various causes of improper solutions in maximum likelihood factor analysis. *Psychometrika*, 43(2), 225–243. <http://doi.org/10.1007/BF02293865>
- Fan, X., & Sivo, S. A. (2005). Sensitivity of fit indices to misspecified structural or measurement model components: rationale of two-index strategy revisited. *Structural Equation Modeling*, 12(3), 343–367.
- Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, 42(3), 509–529.
- Fornell, C., & Yi, Y. (1992). Assumptions of the two-step approach to latent variable modeling. *Sociological Methods & Research*, 20(3), 291–320. <http://doi.org/10.1177/0049124192020003001>
- Hancock, G. R., & Mueller, R. O. (2011). The Reliability Paradox in Assessing Structural Relations Within Covariance Structure Models. *Educational and Psychological Measurement*, 71(2), 306–324. <http://doi.org/10.1177/0013164410384856>
- Hayduk, L. A. (1996). *LISREL issues, debates and strategies*. Baltimore, MD: Johns Hopkins University Press.
- Hayduk, L. A. (2014). Seeing perfectly fitting factor models that are causally misspecified: Understanding that close-fitting models can be worse. *Educational and Psychological Measurement*, 74(6), 905–926. <http://doi.org/10.1177/0013164414527449>
- Hayduk, L. A., Cummings, G. G., Boadu, K., Pazderka-Robinson, H., & Boulianne, S. (2007). Testing! testing! one, two, three - Testing the theory in structural equation models! *Personality and Individual Differences*, 42(5), 841–850. <http://doi.org/10.1016/j.paid.2006.10.001>
- Hayduk, L. A., & Glaser, R. (2000). Jiving the four-step, waltzing around factor analysis, and other serious fun. *Structural Equation Modeling*, 7(1), 1–35.
- Heene, M., Hilbert, S., Draxler, C., Ziegler, M., & Bühner, M. (2011). Masking misfit in confirmatory factor analysis by increasing unique variances: A cautionary note on the usefulness of cutoff values of fit indices. *Psychological Methods*, 16(3), 319–336. <http://doi.org/10.1037/a0024917>
- Heene, M., Hilbert, S., Freudenthaler, H. H., & Bühner, M. (2012). Sensitivity of SEM fit indexes with respect to violations of uncorrelated errors. *Structural Equation Modeling*, 19(1), 36–50. <http://doi.org/10.1080/10705511.2012.634710>

- Hoyle, R. H. (1995). *Structural equation modeling: Concepts, issues, and applications*. Sage.
- Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <http://doi.org/10.1080/10705519909540118>
- Joreskog, K. G., & Sorbom, D. (1981). LISREL: Analysis of linear structural relationships by the method of maximum likelihood. *Chicago: National Educational Resources*.
- Jöreskog, K. G., & Sörbom, D. (1981). *LISREL V: Analysis of linear structural relationships by maximum likelihood and least squares methods*. University of Uppsala, Department of Statistics.
- Kaplan, D. (1988). The impact of specification error on the estimation, testing, and improvement of structural equation models. *Multivariate Behavioral Research*, 23(1), 69–86. http://doi.org/10.1207/s15327906mbr2301_4
- Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling*, 10, 333–351. http://doi.org/10.1207/S15328007SEM1003_1
- Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, 103(3), 391–410. <http://doi.org/10.1037/0033-2909.103.3.391>
- Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). *In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler's (1999) Findings*. *Structural Equation Modeling* Vol 11(3) Jul 2004, 320-341.
- McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, 7(1), 64–82. <http://doi.org/10.1037/1082-989X.7.1.64>
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, 105(1), 156. <http://doi.org/10.1037/0033-2909.105.1.156>
- Moshagen, M. (2012). The model size effect in SEM: inflated goodness-of-fit statistics are due to the size of the covariance matrix. *Structural Equation Modeling: A Multidisciplinary Journal*, 19(1), 86–98. <http://doi.org/10.1080/10705511.2012.634724>
- Mulaik, S. A. (2009). *Linear Causal Modeling with Structural Equations* (1st ed.). Chapman & Hall.

- Mulaik, S. A., & Millsap, R. E. (2000). Doing the four-step right. *Structural Equation Modeling*, 7(1), 36–73. http://doi.org/10.1207/S15328007SEM0701_02
- O'Boyle, E. H. J., & Williams, L. J. (2011). Decomposing model fit: Measurement vs. theory in organizational research using latent variables. *Journal of Applied Psychology*, 96(1), 1–12. <http://doi.org/10.1037/a0020539>
- Peterson, R. A., & Brown, S. P. (2005). On the use of beta coefficients in meta-analysis. *Journal of Applied Psychology*, 90(1), 175–181. <http://doi.org/10.1037/0021-9010.90.1.175>
- Peterson, R. A. (2000). A meta analysis of variance accounted for and factor loadings in exploratory factor analysis. *Marketing Letters*, 11, 261–275. doi:10.1023/A:1008191211004
- Pornprasertmanit, S., Miller, P., & Schoemann, A. (2015). *simsem: SIMulated Structural Equation Modeling*. Retrieved from <http://CRAN.R-project.org/package=simsem>
- R Core Team. (2015). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Saris, W. E., & Satorra, A. (1988). Characteristics of structural equation models which affect the power of the likelihood ratio test. *Sociometric Research*, 2, 220–236.
- Saris, W. E., & Satorra, A. (1993). Power evaluations in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 181–204). Thousand Oaks: Sage.
- Saris, W. E., Satorra, A., & van der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications. *Structural Equation Modeling*, 16(561–582). <http://doi.org/10.1080/10705510903203433>
- Savalei, V. (2012). The relationship between root mean square error of approximation and model misspecification in confirmatory factor analysis models. *Educational and Psychological Measurement*, 72(6), 910–932.
- Scheck, C. L., Kinicki, A. J., & Davy, J. A. (1995). A longitudinal study of a multivariate model of the stress process using structural equations modeling. *Human Relations*, 48(12), 1481–1510.
- Schermelleh-Engel, K., & Moosbrugger, H. (2003). Evaluating the fit of structural equation models: test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8, 23–74.
- Schonemann, P. H. (1981a). Factorial definitions of intelligence: Dubious legacy of dogma in data analysis. In I. Borg (Ed.), *Multidimensional Data Representations: When & Why*. Ann Arbor: Mathesis Press.

- Schonemann, P. H. (1981b). Power as a function of communality in factor analysis. *Bulletin of the Psychonomic Society*, 17(1), 57–60.
- Smith, K. J., Davy, J. A., & Easterling, D. (2004). An examination of cheating and its antecedents among marketing and management majors. *Journal of Business Ethics*, 50(1), 63–80.
- Steiger, J. H. (1998). A note on multiple sample extensions of the RMSEA fit index. *Structural Equation Modeling*, 5(4), 411–419. <http://doi.org/10.1080/10705519809540115>
- Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate behavioral research*, 25(2), 173-180.
- Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, 42(5), 893–898. <http://doi.org/10.1016/j.paid.2006.09.017>
- Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika*, 50(3), 253–263. <http://doi.org/10.1007/BF02294104>
- Tomarken, A. J., & Waller, N. G. (2003). Potential problems with “well fitting” models. *Journal of Abnormal Psychology*, 112(4), 578–598. <http://doi.org/10.1037/0021-843X.112.4.578>
- Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement*, 9(1), 1–26.
- Yuan, K.-H. (2005). Fit Indices Versus Test Statistics. *Multivariate Behavioral Research*, 40(1), 115–148. http://doi.org/10.1207/s15327906mbr4001_5
- Yuan, K.-H., Marshall, L. L., & Bentler, P. (2003). Assessing the effect of model misspecifications on parameter estimates in structural equation models. *Sociological Methodology*, 33(1), 241–265. <http://doi.org/10.1111/j.0081-1750.2003.00132.x>

Danksagung

Vor einigen Jahren schrieb ich in der Danksagung meiner Diplomarbeit, dass eine solche Arbeit nicht alleine durch eine Person entsteht. Nun, ein wenig später, kann ich nur sagen, es ist auch bei einem Dissertationsvorhaben das Gleiche – alleine würde man es nicht schaffen.

Ein ganz besonderer Dank gilt meinem Dissertationsbetreuer und Mentor, Herrn Prof. Dr. Moritz Heene, der mich schon seit meiner Diplomarbeit begleitet. Ohne ihn und seine großartige Unterstützung wäre ich nie so weit gekommen und hätte auch nicht den nötigen kritischen Sinn entwickelt. Ich möchte auch Herrn Prof. Dr. Markus Bühner danken, der sich als Zweitgutachter zur Verfügung gestellt hat. Großer Dank gebührt auch Dr. Sunthud Pornprasertmanit, der uns immer mit Rat und Tat zur Seite stand.

Natürlich braucht man auch viel an privater Unterstützung, damit einem die Luft auf diesem Weg nicht ausgeht. Hier sind zuerst meine Eltern, Karin und Karl-Heinz Malirsch, zu nennen. Ihr habt an mich geglaubt und meine Bildung immer gefördert und unterstützt. Ohne euch wäre ich heute nicht dort, wo ich bin.

Meinem geliebten Mann gebührt besonderer Dank. Du hast mir beigebracht, dass schon alles gut werden wird – und du hattest wie immer recht. Aber unsere kleine Familie hat auch einen vierbeinigen Unterstützer, unseren Hund Elli, die immer mit ihrem außergewöhnlichen Radar meine „Dissertationsschwingungen“ wahrgenommen hat. Auch wenn es ungewöhnlich klingt, darf man eine weitere vierbeinige Freundin von mir nicht vergessen, Lorli, die mit ihrem schwungvollen Galopp so manchen Rückschlag weggeweht hat.

Auch meinen Großeltern und meinem Bruder möchte ich an dieser Stelle danken. Ihr habt mich all die Jahre begleitet. Besonderer Dank gilt Ruth, die verhindert hat, dass diese Dissertation „dem Genitiv sein Tod“ geworden ist. Auch Marie möchte ich herzlich danken, dafür, dass wir immer über alles lachen können. Noch ein Dankeschön an Familie Ishmurzin – ihr seid etwas ganz Besonderes. Natürlich auch ein Danke an all meine Freunde, die immer ein offenes Ohr für mich haben.