Identifikation und Kontrolle von Fragereihenfolgeneffekten. Entwicklung und Evaluation eines CFA Modells zur Bereinigung messfehlerbedingter Verzerrungen auf Indikatorebene.

Von der Fakultät für Human- und Sozialwissenschaften der
Technischen Universität Chemnitz genehmigte

Dissertation

zur Erlangung des akademischen Grades

Dr. rer. soc.

vorgelegt von

Christoph Giehl, M.A.

(Geboren am 20.12.1988 in Rockenhausen)

Erstgutachter: Prof. Dr. Jochen Mayerl

Zweitgutachter: Prof. Dr. Thomas Lachmann

Tag der Einreichung: 07.11.2022

Tag der Disputation: 26.06.2023

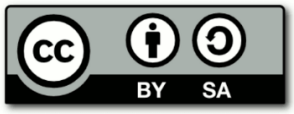
Veröffentlichungsjahr: 2023

Online: https://nbn-resolving.org/urn:nbn:de:bsz:ch1-qucosa2-864400

Diese Arbeit – ausgenommen sämtliche Abbildungen – steht unter der Creative-Commons-Lizenz

Attribution – ShareAlike 4.0 International (CC BY-SA 4.0)

https://creativecommons.org/licenses/by-sa/4.0/



Zusammenfassung

In der empirischen Forschung gehören Fragebatterien (Sammlungen verschiedener Fragen zu einem gemeinsamen Thema von Interesse) in Umfragen zu den etablierten Instrumenten der Datenerhebung. Die Antworten Befragter auf einzelne Fragen solcher Batterien müssen jedoch nicht zwangsläufig der Wahrheit entsprechen, sondern können durch sogenannte Antwortverzerrungen bewusst oder unbewusst von der eigentlich wahren Antwort abweichen. Eine der bekanntesten Antwortverzerrungen, welche darüber hinaus vielfältig in verschiedenen Studien nachgewiesen werden konnte, ist der Fragereihenfolgeneffekt (kurz: „FRE“). Von einem solchen ist die Rede, wenn die Antwort auf eine nachfolgend gestellte Frage durch die Antwort auf eine vorangegangene Frage beeinflusst wird und die Antwort auf die Folgefrage dadurch systematisch von der eigentlich wahren Antwort abweicht. Um die Entstehungsbedingungen solcher FRE zu identifizieren, wurden verschiedene etablierte Modelle und Theorien des Befragtenverhaltens und der Verhaltensforschung analysiert und hinsichtlich ihrer kleinsten gemeinsamen Nenner verglichen. Hierauf basierend zeigt sich theoretisch fundiert, dass der kognitive Modus der Informationsverarbeitung bei der Beantwortung einer Frage, die Zugänglichkeit von Informationen bezüglich des Objekts der Fragebatterie, sowie das allgemeine Framing der Befragtenrolle maßgebliche Determinanten des Befragtenverhaltens und des Auftretens von FRE darstellen. Um die durch FRE auftretenden Verzerrungen zu bereinigen und um darüber hinaus die konkrete Höhe der FRE und deren Prädiktoren zu quantifizieren, wurde nachfolgend ein CFA Modell entwickelt (das NLC-CFA Modell), welches in der Lage ist, den Einfluss der drei identifizierten Determinanten zu berücksichtigen. Durch die konkrete Berücksichtigung der Determinanten hat das NLC-CFA Modell einen erhöhten Mehrwert gegenüber einem etablierten Modell zur Bereinigung von Antwortverzerrungen (das CFA-MTMM Modell), sowohl hinsichtlich der konkreten Effizienz der Bereinigung, als auch hinsichtlich des Informationsgewinns durch die zusätzliche Quantifizierung des FRE sowie dessen Abhängigkeiten, wie eine umfassende Evaluation des Modells auf Basis mehrerer Monte-Carlo Simulationsstudien zeigt. Das NLC-CFA Modell kann darüber hinaus situativ adaptiert werden, um beliebige Antwortverzerrungen und nicht nur Verzerrungen durch den FRE zu bereinigen.

Inhaltsverzeichnis

Zusammenfassung ..... **I**
Inhaltsverzeichnis ..... **II**
Abbildungsverzeichnis ..... **IV**
Tabellenverzeichnis ..... **V**
1 Einleitung **1**
2 Theoretischer Hintergrund: Die Erklärung von Fragereihenfolgeneffekten **5**
2.1 Kognitive Theorien und Modelle des Befragtenverhaltens 6
2.1.1 Explizite, klassische Modelle des Befragtenverhaltens 6
2.1.2 Implizite Verhaltensmodelle mit Fokus auf differente Elaborationsgrade 14
2.1.3 Implizite, integrative Prozessmodelle der Einstellungs-Verhaltensforschung 28
2.1.4 Der Prozess der Antwortgenerierung: Ableitung allgemeiner Thesen 50
2.2 Fragereihenfolgeneffekte in der Umfrageforschung 61
2.2.1 Antwortverzerrungen – Eine Einführung 62
2.2.2 Die Logik von Fragereihenfolgeneffekten 66
2.2.3 Bedingungsfaktoren von Fragereihenfolgeneffekten 72
2.3 Fragereihenfolgeneffekte: Modell und Hypothesen 85
2.3.1 Zusammenfassung und Systematisierung der entwickelten Thesen 86
2.3.2 Interaktion und Gruppierung der Einflussfaktoren 92
2.3.3 Synthese, Modell und Hypothesen 98
3 Methodischer Ansatz: Statistische Kontrolle von Fragereihenfolgeneffekten **102**
3.1 Systematische Messfehler in der empirischen Forschung 103
3.1.1 Systematische Messfehler innerhalb der klassischen Testtheorie 104
3.1.2 Systematische Messfehler in Strukturgleichungsmodellen 111
3.1.3 Systematische Messfehler in Multi-Trait-Multimethod-Analysen 123
3.1.4 Implikationen für Messungen der Datenqualität 134
3.2 Modellierung von Fragereihenfolgeneffekten 139
3.2.1 Identifikation und Quantifizierung von Fragereihenfolgeneffekten 140
3.2.2 Kontrolle von Fragereihenfolgeneffekten mittels Moderatoranalyse 148
3.3 Modelle zur Identifikation und Kontrolle von Fragereihenfolgeneffekten 172
3.3.1 CFA-MTMM Modell 173
3.3.2 Nichtlinear latent kontrollierte CFA (NLC-CFA) 178
3.3.3 Gegenüberstellung der vorgeschlagenen Modelle und Ausblick 186

4 Simulationsstudie zur Evaluation der Modelle 192
4.1 Simulationsparameter und Analysesystematik 193
4.2 Methodik der Datensimulation und -Auswertung 207
4.3 Verteilungsanalysen der relevanten Variablen 212
4.4 Analyse der Verzerrungen von Faktorladungen 218
4.5 Effekt Effizienz Analyse 229
4.6 Standardfehler Präzisionsanalyse 248
4.7 Detaillierte Analyse der angepassten Benchmark Modelle 265
5 Diskussion und Zusammenfassung 273
5.1 Diskussion hinsichtlich der methodischen Implikationen 274
5.2 Diskussion hinsichtlich der theoretischen Implikationen 282
5.3 Anwendung der NLC-CFA in der Praxis 291
5.4 Gesamtzusammenfassung 304
6 Literaturverzeichnis 308
7 Anhang 329
7.1 R Syntax der Monte-Carlo Analyse mit 10000 Iterationen 329
7.2 R Syntax der angepassten Benchmark Analyse 353
Eidesstattliche Erklärungen 361
Wissenschaftlicher Lebenslauf und Veröffentlichungen 362
Danksagung 367

Abbildungsverzeichnis

Abbildung 1: Bedingungsfaktoren des Assimilationseffektes 96
Abbildung 2: Bedingungsfaktoren des Kontrasteffektes 97
Abbildung 3: Bedingungsfaktoren von Fragereihenfolgeneffekten 99
Abbildung 4: SEM Y-Notation 113
Abbildung 5: SEM X-Notation 114
Abbildung 6: Varianzzerlegung von x (und analog für y) 115
Abbildung 7: Einfaches Messmodell mit $p = 3$ und $m = 1$ (X-Notation) 117
Abbildung 8: Vereinfachte Darstellung der MTMM-Matrix 125
Abbildung 9: CFA-MTMM Modell 128
Abbildung 10: CFA-CTC(M-1) Modell mit 4 Indikatoren und einem Methodenfaktor 133
Abbildung 11: Einfaches CFA-MTMM Modell mit spezifizierter Kovarianz auf Messfehlerebene 144
Abbildung 12: CFA-Modell mit spezifiziertem Fragereihenfolgeneffekt 147
Abbildung 13: Einfaches MG-CFA Modell mit Kovarianzen auf Messfehlerebene 153
Abbildung 14: Einfaches Multigruppenmodell mit regressiver Beziehung auf Indikatorebene 155
Abbildung 15: CFA Modell mit Interaktionen auf Indikatorebene 165
Abbildung 16: MG-CFA-MTMM Modell zur Analyse von FRE 175
Abbildung 17: Strukturmodell zur Identifikation und Kontrolle von Fragereihenfolgeneffekten 182
Abbildung 18: Einfaches CFA Benchmark Modell 196
Abbildung 19: NLC-CFA Benchmark Modell 197
Abbildung 20: CFA-MTMM Benchmark Modell 198
Abbildung 21: Generisches Regressionsmodell 218
Abbildung 22: C-CFA zur Bereinigung von Akquieszenz-Effekten 300
Abbildung 23: LC-CFA zur Bereinigung des Bedürfnisses nach sozialer Anerkennung 301
Abbildung 24: Generisches NC-CFA Modell 302

Tabellenverzeichnis

Tabelle 1: Effektinterpretation bei Multigruppenanalysen 158
Tabelle 2: Bedeutung der Interaktionseffekte 181
Tabelle 3: Technische Koeffizienten der Effektberechnung je Gruppe 202
Tabelle 4: Zusammenfassung der simulierten unabhängigen Variablen 205
Tabelle 5: Zusammenfassung der abhängigen Variablen 206
Tabelle 6: Verteilung der *SV* der Faktorladungen 213
Tabelle 7: Verteilung der *SV* der NLC-CFA Effekte 215
Tabelle 8: Verteilung der *SV* Messfehlerkovarianzen 216
Tabelle 9: Verteilung der Standardfehler der Effektschätzer 217
Tabelle 10: Verzerrungen der Faktorladungen ohne Kontrolle von Fragereihenfolgeneffekten 221
Tabelle 11: Verzerrungen der Faktorladungen bei gegebener Kontrolle von Fragereihenfolgeneffekten 224
Tabelle 12: Gegenüberstellung der Faktorladungsverzerrung (unstand. Regressionskoeffizienten *b*) 227
Tabelle 13: Durchschnittliche Verzerrung der Effekte von $\boldsymbol{\beta 1}$ $(\boldsymbol{yj - 1})$ 231
Tabelle 14: Durchschnittliche Verzerrung der Effekte von $\boldsymbol{\beta 2}$ $(\boldsymbol{\xi 1})$ 233
Tabelle 15: Durchschnittliche Verzerrung der Effekte von $\boldsymbol{\beta 3}$ $(\boldsymbol{\xi 2})$ 234
Tabelle 16: Durchschnittliche Verzerrung der Effekte von $\boldsymbol{\beta 4}$ $(\boldsymbol{yj - 1\xi 1})$ 235
Tabelle 17: Durchschnittliche Verzerrung der Effekte von $\boldsymbol{\beta 5}$ $(\boldsymbol{yj - 1\xi 2})$ 237
Tabelle 18: Durchschnittliche Verzerrung der Effekte von $\boldsymbol{\beta 6}$ $(\boldsymbol{\xi 1 \xi 2})$ 238
Tabelle 19: Durchschnittliche Verzerrung der Effekte von $\boldsymbol{\beta 7}$ $(\boldsymbol{yj - 1\xi 1\xi 2})$ 240
Tabelle 20: Durchschnittliche Verzerrung der Messfehlerkovarianz in Gruppe 1 243
Tabelle 21: Durchschnittliche Verzerrung der Messfehlerkovarianz in Gruppe 2 245
Tabelle 22: Durchschnittliche Verzerrung der Messfehlerkovarianz in Gruppe 3 246
Tabelle 23: Durchschnittliche SE der Effekte von $\boldsymbol{\beta 1}$ $(\boldsymbol{SEyj - 1})$ 250
Tabelle 24: Durchschnittliche SE der Effekte von $\boldsymbol{\beta 2}$ $(\boldsymbol{SE\xi 1})$ 251
Tabelle 25: Durchschnittliche SE der Effekte von $\boldsymbol{\beta 3}$ $(\boldsymbol{SE\xi 2})$ 252
Tabelle 26: Durchschnittliche SE der Effekte von $\boldsymbol{\beta 4}$ $(\boldsymbol{SEyj - 1\xi 1})$ 254
Tabelle 27: Durchschnittliche SE der Effekte von $\boldsymbol{\beta 5}$ $(\boldsymbol{SEyj - 1\xi 2})$ 255
Tabelle 28: Durchschnittliche SE der Effekte von $\boldsymbol{\beta 6}$ $(\boldsymbol{SE\xi 1\xi 2})$ 256
Tabelle 29: Durchschnittliche SE der Effekte von $\boldsymbol{\beta 7}$ $(\boldsymbol{SEyj - 1\xi 1\xi 2})$ 258
Tabelle 30: Durchschnittlicher SE der Messfehlerkovarianz in Gruppe 1 260
Tabelle 31: Durchschnittlicher SE der Messfehlerkovarianz in Gruppe 2 262
Tabelle 32: Durchschnittlicher SE der Messfehlerkovarianz in Gruppe 3 263
Tabelle 33: Geschätzte Faktorladungen der angepassten Benchmark Analyse 267
Tabelle 34: Geschätzte Effekte der angepassten Benchmark Analyse 270
Tabelle 35: Geschätzte Effekte der angepassten Benchmark Analysen je Gruppe (Zusammenfassung) 284

1 Einleitung

In der empirischen Sozialforschung gehören standardisierte Interviews zu den etablierten Instrumenten der Datenerhebung, auch wenn die komplexen kognitiven und sozialen Prozesse, welche während eines solchen Interviews ablaufen, zwar beschrieben, aber noch nicht vollständig erklärt werden können (Maynard, Houtkoop-Steenstra, & Schaeffer, 2002) und Kritiker dem Interview sogar vorwerfen, ein fragiles Instrument zu sein, welches ungeeignet ist, reale Interaktionen abzubilden (Suchman & Jordan, 1990). Unabhängig von dieser Kritik stellt das standardisierte Interview im Speziellen, ebenso wie das Interview im Allgemeinen eine Interaktion zwischen Interviewenden und Befragten dar mit dem Ziel, den Befragten Informationen, Meinungsäußerungen oder auch Überzeugungen zu entlocken (Maccoby & Maccoby, 1976), wobei Kahn und Cannell (1968, S. 149) besondere betonen, dass das Interview als Interaktion zwischen zwei Personen eine spezielle Form einer sozialen Situation darstellt. Entsprechend ist das sozialwissenschaftliche Interview kein neutraler Prozess, woraus sich als eines der Hauptprobleme von Interviews die Reaktivität von Befragten auf spezielle (soziale) Umstände der Interviewsituation ergibt (Stocké, 2004, S. 303).

Erklärungsversuche besagter Reaktivität finden sich zum Beispiel bei Esser (1991), der das Befragtenverhalten im Rahmen der Rational Choice Theory als Versuch zur Problemlösung interpretiert. Steht für Befragte der wahre Wert einer Antwort gemäß dieser Theorie nicht im Vordergrund, z.B. weil auf das Geben einer wahren Antwort eine Sanktion durch die soziale Umwelt erfolgen könnte, ist mit systematischen Antwortverzerrungen als Folge von unter anderem Interviewereffekten, Anwesenheit dritter Personen oder allgemein sogenannter „response sets“ zu rechnen. Auch Strack und Martin (1987) diskutieren das Phänomen, dass das Beantworten von Fragen in Interviewsituationen nicht ausschließlich von z.B. individuellen Einstellungen gegenüber dem Objekt der Befragung, sondern ebenfalls von sogenannten Kontexteffekten abhängig ist, welche zu Antwortverzerrungen führen können. Solche Kontexteffekte ergeben sich beispielsweise aus dem spezifischen Inhalt einer Frage oder der Antwortskala. Aber auch eine vorangegangene Frage kann das Antwortverhalten auf eine nachfolgende Frage beeinflussen. Ist dies der Fall, ist von sogenannten Fragereihenfolgeneffekten (kurz: „FRE“) die Rede, deren Analyse zentral für die vorliegende Arbeit ist.

Interviews sind somit komplexe Situationen, die bei Befragten aufgrund sozialer und kognitiver Wirkmechanismen, welche zu Antwortverzerrungen wie FRE führen können, ein nicht immer

wahrheitskonformes Antwortverhalten provozieren. Gerade hieraus ergibt sich die große Herausforderung für die Umfrageforschung. Denn während es das Ziel ist, unverzerrte Messungen durchzuführen, wird dennoch davon ausgegangen, dass sich der konkrete Messwert aus dem wahren Wert und einem Messfehler ergibt (Moosbrugger, 2012). Problematisch ist dies, da solche Messfehler lediglich während des Datenerhebungsprozesses vermieden und danach nur noch durch statistische Verfahren kontrolliert werden können (Stocké, 2004), wodurch sowohl der Erforschung der Determinanten solcher Antwortverzerrungen, als auch der Entwicklung von statistischen Verfahren zur Kontrolle von Messfehlern eine besondere Bedeutung zukommt. Diesbezüglich soll die vorliegende Arbeit mit Fokus auf die bereits erwähnten FRE einen Beitrag leisten.

Diese Arbeit setzt sich entsprechend dieses Anspruchs die nachfolgenden Ziele:

1. Zunächst wird der kognitive Entstehungsprozess von Antworten in Umfragen auf theoretischer Ebene analysiert um zu identifizieren, wo dieser Prozess anfällig ist für Antwortverzerrungen bzw. insbesondere FRE. Hiervon sollen allgemeine Determinanten und Bedingungsfaktoren für das Auftreten von FRE abgeleitet werden. Dies mündet schließlich in der Entwicklung eines einfachen Erklärungsmodells, auf dessen Basis empirisch überprüfbare Hypothesen hinsichtlich des Auftretens von FRE formuliert werden.
2. Auf Basis des entwickelten Erklärungsmodells wird aufgezeigt, wie FRE mithilfe eines etablierten Verfahrens, der Multi-Trait-Multimethod (MTMM) Analyse, identifiziert und kontrolliert werden können. Daneben wird ein neues, allgemein anwendbares Verfahren zur Identifikation und Kontrolle von FRE vorgeschlagen und eingeführt, die nichtlinear latent kontrollierte konfirmatorische Faktoranalyse (**n**onlinear **l**atent **c**ontrolled **c**onfirmatory **f**actor **a**nalysis, kurz: NLC-CFA). Die NLC-CFA weißt dabei aufgrund der Möglichkeit zur Spezifikation der konkreten Einflussbeziehungen der Determinanten systematischer Messfehler einen höheren Grad an berücksichtigten Informationen auf als die MTMM Analyse und soll somit zu einer effizienteren Bereinigung von Verzerrungen durch FRE beitragen.
3. Analytisch wird mittels mehrerer Monte Carlo Simulationsstudien aufgezeigt, dass die Anwesenheit von FRE dort geschätzte Faktorladungen eines durch eine Fragebatterie repräsentierten latenten Konstruktes verzerren, wo diese nicht statistisch kontrolliert werden und das besagte Verzerrungen durch Anwendung des MTMM Verfahrens sowie

der NLC-CFA bereinigt werden können, um so zu effizienteren Schätzungen der Indikatoren der Konstruktvalidität und der Kompositreliabilität zu gelangen. In diesem Zusammenhang erfolgt ebenfalls eine Gegenüberstellung des etablierten MTMM Verfahrens und der NLC-CFA wobei gezeigt werden soll, dass die NLC-CFA FRE-bedingte Verzerrungen von Faktorladungen effizienter bereinigt. Es erfolgt zudem eine Evaluation der Präzision und Effizienz derjenigen geschätzten Parameter, welche den Effekt des FRE quantifizieren. Auf Basis dieser Analysen kann letztlich geschlossen werden, unter welchen Bedingungen die theoretisch abgeleiteten Hypothesen ohne Typ 1 oder Typ 2 Fehler überprüft werden können.

Die Forschungsfrage dieser Arbeit lautet somit: Wie können FRE möglichst präzise und effizient quantifiziert und kontrolliert werden und wie wirkt sich diese Kontrolle auf die Schätzungen der Faktorladungen und damit die Indikatoren der Datenqualität aus?

Um sich der Forschungsfrage und den allgemeinen Zielen dieser Arbeit anzunähern, werden in Kapitel 2 (Theoretischer Hintergrund: Die Erklärung von Fragereihenfolgeneffekten) auf theoretischer Ebene allgemeine Theorien und Modelle des Befragtenverhaltens diskutiert, um hieraus Determinanten und Bedingungsfaktoren sowie die zugrundeliegenden Mechanismen bei der Entstehung von FRE abzuleiten. Hierauf aufbauend soll ein theoretisches Erklärungsmodell zum Auftreten von FRE entwickelt werden, aus welchem sich wiederum empirisch überprüfbare Hypothesen ergeben.

In Kapitel 3 (Methodischer Ansatz: Identifikation und Kontrolle von Fragereihenfolgeneffekten) wird, hierauf aufbauend, zum einen aufgezeigt, wie FRE mittels MTMM Analyse identifiziert und kontrolliert werden können. Zum anderen wird das innovative NLC-CFA Modell zur Identifikation und Kontrolle von FRE schrittweiße entwickelt und eingeführt. Diesbezüglich wird aufgezeigt, wie sich Messfehler allgemein statistisch-mathematisch modellieren und FRE im Speziellen mittels verschiedener klassischer und etablierter Verfahren (darunter auch der MTMM Ansatz) identifizieren lassen. Daneben wird die Verwendung von (latenten) Interaktionstermen auf Indikatorebene innerhalb eines Multigruppenmodells als Möglichkeit zur Quantifizierung der FRE diskutiert. Die Kontrolle von Indikatoren eines CFA Modells mittels latenter Interaktion entspricht dabei der fundamentalen Logik der NLC-CFA. Entsprechend dieser Diskussion wird ein NLC-CFA Modell zur Identifikation und Kontrolle von FRE auf Basis des theoretischen Modells entwickelt.

In Kapitel 4 (Simulationsstudie zur Evaluation der Modelle) wird sowohl das MTMM Modell, als auch das NLC-CFA Modell innerhalb mehrerer Monte Carlo Analysen (mit insgesamt

10000 Iterationen je Analyse) mit bekannten wahren Werten analysiert. Im Detail wird 1.) analysiert, ob Faktorladungen als Indikatoren der Konstruktvalidität und der Kompositreliabilität bei unspezifizierten FRE innerhalb einfacher konfirmatorischer Faktoranalysen (kurz: CFA) verzerrt sind und inwiefern sich verschiedene Modell- und Simulationsspezifikationen (wie etwa die Fallzahl oder die Höhe der „wahren" Faktorladungen) in den einzelnen Iterationen der Simulation auf die Höhe der entsprechenden Verzerrungen auswirken. Hinsichtlich der Faktorladungen wird ebenfalls analysiert, inwiefern besagte Verzerrungen durch Schätzung des MTMM Modells sowie des NLC-CFA Modells bereinigt werden, sowie erneut, inwiefern die verbleibenden Verzerrungen von verschiedenen Modell- und Simulationsspezifikationen abhängig sind. Hinsichtlich der Effizient bei der Bereinigung der Verzerrungen werden die beiden Verfahren schließlich gegenübergestellt. Zudem werden ebenfalls 2.) die Verzerrungen der Effektparameter analysiert, welche zwecks Quantifizierung des FRE innerhalb beider Modelle spezifiziert werden. Die Verzerrungen der Effektparameter werden diesbezüglich erneut auf verschiedene Modell- und Simulationsspezifikationen zurückgeführt. 3.) erfolgt eine Analyse der geschätzten Standardfehler der geschätzten Effektparameter beider Modelle, wobei diese ebenfalls auf verschiedene Modell- und Simulationsspezifikationen innerhalb der einzelnen Iterationen der Monte Carlo Simulation zurückgeführt werden. 4.) wird zuletzt eine einfache, auf Basis der vorangegangenen Ergebnisse angepasste Monte Carlo Analyse ohne variierende Modell- und Simulationsspezifikationen innerhalb der einzelnen Iterationen auf Basis eines sogenannten „Benchmark Modells" durchgeführt, um die konkret geschätzten Parameter des innovativen NLC-CFA Verfahrens im Detail zu analysieren und mit den geschätzten Parametern des etablierten MTMM Verfahrens zu vergleichen.

Die Ergebnisse der jeweiligen Analysen werden abschließend in Kapitel 5 (Diskussion und Zusammenfassung) diskutiert. Insbesondere soll aufgezeigt werden, 1.) dass vorhandene, unspezifizierte systematische Messfehler wie der FRE Faktorladungen verzerren und dass die Verzerrung durch Spezifikation des FRE bereinigt werden kann, sowie, dass die Bereinigung durch die NLC-CFA effizienter ist als bei Schätzung des MTMM Modells; 2.) dass das NLC-CFA Modell innerhalb der simulierten Daten zu präzisen und effizienten Schätzungen der Effektparameter führt, wodurch die formulierten, theoretischen Hypothesen ebenfalls überprüfbar sind; und 3.) dass das vorgeschlagene Modell insgesamt als komplexer Anwendungsfall eines allgemeinen Frameworks zur Quantifizierung systematischer Messfehler und damit zur Bereinigung verzerrter Parameterschätzungen dienen kann, welches situativ beliebig adaptierbar ist.

2 Theoretischer Hintergrund: Die Erklärung von Fragereihenfolgeneffekten

Wissenschaftliche Befragungen und Interviews, ganz gleich ob diese nun qualitativ oder quantitativ durchgeführt werden, stellen stets besondere Herausforderungen an Befragte. So ist es auf der einen Seite das Ziel, bei Befragungen nach Möglichkeit „wahre" Antworten von Befragten zu erhalten. Auf der anderen Seite lassen sich in Befragten eine ganze Reihe von Bedürfnissen und Motiven identifizieren, welche für oder gegen eine wahrheitsgemäße Beantwortung von Fragen sprechen (Strack, 2013). Umfragesituationen lassen sich somit aufgrund ihres inhärenten Gesprächscharakters als komplexe soziale Situationen beschreiben.

Die Komplexität der Situation, in welcher Befragte sich während eines Interviews befinden, ergibt, dass diese sich in Umfragesituationen sowohl inter-, als auch intraindividuell different verhalten. Faktoren wie Müdigkeit, Motivation, Kognition oder die Situation bedingen diese Varianz (Moosbrugger & Brandt, 2020). Entsprechend kommt der Betrachtung des kognitiven Prozesses der Antwortgenerierung eine besondere Bedeutung zu, so das konkrete Antwortverhalten Befragter verstanden werden will.

Das Verständnis hierum ist alles andere als trivial, da für Befragte das Geben einer wahren Antwort kein zwangsläufiges Bedürfnis darstellt – und das unabhängig davon, ob eine Antwort nun bewusst oder unbewusst unwahr formuliert wurde. Unwahre, das heißt verzerrte Antworten wiederum resultieren dabei nicht nur in inhaltlichen Verzerrungen, sondern ebenfalls in verzerrten Messungen der Datenqualität (Hlawatsch & Krickl, 2019). Ist es somit das Ziel quantitativer und standardisierter Untersuchungen, wahre Werte und unverzerrte Messungen der Datenqualität zu erhalten, so ist es unabdingbar, Antwortverzerrungen wie den in der Einleitung hervorgehobenen Fragereihenfolgeneffekt (kurz: „FRE") sowie die kognitiven Prozesse, welche zu diesen führen, einer genaueren Analyse zu unterziehen.

Das Ziel des vorliegenden Kapitels ist es somit, den kognitiven Entstehungsprozess von Antworten in Umfragen zu analysieren (vgl. Kapitel 2.1) und zu identifizieren, wo dieser Prozess anfällig ist für Antwortverzerrungen bzw. insbesondere FRE (vgl. Kapitel 2.2). Dies mündet schließlich in der Entwicklung eines Erklärungsmodells, auf dessen Basis FRE identifiziert und kontrolliert werden können (vgl. Kapitel 2.3).

2.1 Kognitive Theorien und Modelle des Befragtenverhaltens

Standardisierte Umfragesituationen sind komplexe soziale Situationen, in welchen Befragte Stimuli durch Interviewende, Fragebögen etc. erhalten und als Reaktion auf diese Stimuli eine Form von Output generieren (Reinecke, 1991). Im besten Fall deckt sich dieser formulierte Output mit der Realität, häufig weicht dieser jedoch auch von der „Wahrheit" ab, wie die Forschung rund um Antwortverzerrungen und systematische Messfehler zeigt (Phillips, Phillips, & Aaron, 2013). Entsprechend ist es bedeutsam, dem kognitiven Prozess der Antwortgenerierung, insbesondere mit Hinblick auf die Generierung „wahrer" und „verzerrter" Antworten, im Kontext standardisierter Umfragen genauer zu beleuchten. Diesbezüglich erfolgt in diesem Kapitel ein vergleichender Überblick zu verschiedenen Modellen und Theorien.

Zu besagtem Prozess finden sich in der Sozialpsychologie sowie der Kognitionsforschung verschiedene Vorschläge, welche sich, einer eigenen Differenzierung entsprechend, im Wesentlichen Unterscheiden lassen in Modelle der Informationsverarbeitung in Umfragesituationen (vgl. z.B. Cannell, Miller, & Oksenberg, 1981; Strack & Martin, 1987; Sudman, Bradburn, & Schwarz, 1996; Tourangeau, Rips, & Rasinski, 2000) sowie in (kognitive) Verhaltensmodelle (vgl. z.B. Ajzen, 1991; Chen & Chaiken, 1999; Fazio, 1990; Fazio & Olson, 2014; Krosnick, 1991; Petty, Cacioppo, & Kasmer, 2015; Strack & Deutsch, 2012). Zu beachten ist dabei, dass Modelle der Informationsverarbeitung zumeist sequenziell verlaufen und inhärent die Generierung von Antworten in Befragungen beleuchten, während (kognitive) Verhaltensmodelle zumeist als mehr oder weniger generische Prozessmodelle dargestellt werden, welche gezeigtes Verhalten allgemein auf kognitive Prozesse (wie etwa der Einfluss einer Einstellung) zurückführen und dabei lediglich für das spezielle Verhalten „Generierung von Antworten in Befragungen" adaptiert werden. Ebenso können die entsprechenden Modelle demnach ebenfalls als explizite Modelle des Befragtenverhaltens sowie implizite Verhaltensmodelle, die sich auf Befragtenverhalten abstrahieren lassen, beschrieben werden.

2.1.1 Explizite, klassische Modelle des Befragtenverhaltens

Als explizite Modelle der Informationsverarbeitung in Umfragesituationen bzw. des Befragtenverhaltens lassen sich die klassischen sozialpsychologischen Modelle von Tourangeau, Rips und Rasinski (2000), Strack und Martin (1987), Sudman, Bradburn und Schwarz (1996) und Cannell, Miller und Oksenberg (1981) nennen, welche im Folgenden beschrieben werden.

Vier Schritte der Antwortgenerierung nach Tourangeau et al.

Der Prozess der Informationsverarbeitung und Antwortgenerierung in Umfragen verläuft nach Tourangeau und Kollegen (2000) in vier schematischen, sequenziellen Schritten, welche von Befragten für jedes Umfrageitem mehr oder weniger vollständig durchlaufen werden. So müssen Befragte Fragen 1.) zunächst verstehen, 2.) relevanten Informationen zur Urteilsbildung aus dem Gedächtnis abrufen, 3.) die abgerufenen Informationen hinsichtlich ihrer Vollständigkeit und Relevanz bewerten und 4.) eine Antwort selektieren, formulieren und geben.

Das (1) Verständnis der Frage ist nach dem Antwortmodell von Tourangeau und Kollegen (ebd.) von Syntax, Semantik und Pragmatik der Frageformulierung abhängig – so scheitert dieses Beispielsweise an uneindeutig formulierten Fragen und Antwortkategorien oder wenn das Wissen bzw. die kognitiven Fähigkeiten Befragter nicht ausreichen, um die formulierte Frage zu prozessieren. Das (2) Abrufen von für die Urteilsbildung relevanten Informationen wiederum ist zunächst abhängig von der Strategie, mit welcher Befragte Informationen abrufen, sowie von situativen Hinweisreizen. Strategien werden dabei nicht zwangsläufig für jedes Befragungsitem neu elaboriert, sondern sind mitunter zeitlich stabil. So könnte eine Strategie beispielsweise lauten, möglichst wenig Zeit in die Beantwortung von Umfragen zu investieren, so dass der Abrufprozess eher oberflächlich verläuft. Situative Hinweisreize wiederum können sich darauf auswirken, welche Informationen als relevant erachtet werden und können sich z.B. aus dem allgemeinen Thema der Befragung oder dem Inhalt vorangegangener Fragen ergeben. Weitere Faktoren, welche sich auf das Abrufen von Informationen auswirken, sind a) die zeitliche Nähe zwischen dem empirischen Auftreten eines Ereignisses und der Abfrage zu diesem Ereignis in Befragungen (je länger etwas her ist, desto schwerer fällt das Erinnern), b) die Unterscheidungskraft von Ereignissen, da sehr ähnliche Ereignisse verwechselt werden können und c) die emotionale Relevanz des Ereignisses (je höher die emotionale Relevanz, desto leichter fällt das Erinnern). Sind die Informationen abgerufen, werden diese (3) hinsichtlich einer Urteilsbildung evaluiert. So erfolgt ein kognitiver Prozess der Bewertung von Vollständigkeit und Relevanz der Informationen, wobei die aktuelle Verfügbarkeit der Informationen ein Indikator für deren Relevanz darstellt. Je leichter bzw. schneller Befragte Informationen zur Urteilsbildung erinnern, als desto relevanter werden diese bewertet. Dies impliziert ebenfalls die Subjektivität der Evaluation, da unterschiedliche Informationen für unterschiedliche Befragte in unterschiedlichen Situationen unterschiedlich verfügbar sind. Wurden die Informationen nun hinsichtlich ihrer Relevanz und Vollständigkeit evaluiert, erfolgt (4) die Selektion, Formulie-

rung und Gabe der Antwort. In diesem letzten Schritt formulieren Befragte auf Basis der vorangegangenen Evaluation eine Antwort, welche im Folgenden dann noch derart editiert wird, so dass sie zu den vorgegebenen Antwortkategorien passt.

Gemäß dieses Modelles ist die Generierung einer Antwort zunächst abhängig von Syntax, Semantik und Pragmatik oder zusammengefasst, der a) Frageformulierung, sowie darüber hinaus b) den kognitiven Fähigkeiten Befragter die benötigt werden, die Frage zu verstehen, c) der zugrundeliegenden Antwortstrategie, d) dem Vorhandensein situativer Hinweisreize, e) der zeitlichen Nähe, der Unterscheidungskraft und der emotionalen Relevanz des interessierenden Items, f) der allgemeinen Verfügbarkeit von Informationen sowie g) den vorgegebenen Antwortkategorien.

Verzerrte, d.h. unwahre Antworten werden demnach gegeben, wenn Fragen nicht verstanden werden, wenn eine oberflächliche Antwortstrategie gewählt wird welche nicht ausreicht, alle relevanten Informationen abzurufen, wenn situative Hinweisreize zur Evaluation von Relevanz und Vollständigkeit mitverwendet werden, wenn der Erinnerungsprozess oder die Evaluation relevanter Informationen aufgrund deren geringer Zugänglichkeit fehlerhaft verläuft oder wenn die Antwortkategorien nicht zu der eigentlich formulierten Antwort passen.

Vier Schritte der Antwortgenerierung nach Strack und Martin sowie Sudman et al.

Strack und Martin (1987) sowie später Sudman und Kollegen (1996) formulieren diesen Prozess nach einem sehr ähnlichen Schema. So kommt es bei der Präsentation von Fragen 1.) zur Interpretation der Frage, woraufhin 2.) eine Meinung gebildet wird. 3.) Erfolgt hiernach eine Anpassung der gebildeten Meinung an das optimale Antwortformat und 4.) eine mögliche Editierung der Antwort aufgrund individueller Motive.

Die (1) Interpretation der Frage ist auch nach Strack und Martin (1987) sowie Sudman, und Kollegen (1996) abhängig von Syntax, Semantik und Pragmatik der Frageformulierung und somit ebenfalls von den kognitiven Fähigkeiten Befragter, Fragen grundsätzlich verstehen und prozessieren zu können. Zusätzlich zu den benötigten kognitiven Fähigkeiten, um eine Frage verstehen zu können, benötigen Befragte das Bedürfnis zur Kooperation in Befragungssituationen, um eine Frage auch verstehen und folglich als generierte Antwort relevante Informationen zur Verfügung stellen zu wollen. Darüber hinaus wird der Prozess der Interpretation beschrieben als entweder passiv und automatisch (ohne bewusste Kontrolle Befragter) oder als bewusster und gesteuerter Versuch einer Problemlösung. Abschließend ist die Interpretation einer Frage abhängig von den Informationen, welchen Befragten zum Zeitpunkt der Interpreta-

tion zur Verfügung stehen. Diese verfügbaren Informationen unterscheiden sich in: a) dem Inhalt der Frage selbst; b) der Antwortskala; und c) der vorangegangenen Frage. So interpretieren Befragte Fragen im besten Fall korrekt a) auf Basis aller durch die Fragestellung bereitgestellter Informationen. Darüber hinaus können auch b) vorgegebene Antwortkategorien dabei helfen, eine Frage zu interpretieren bzw. zu verstehen, welche Art von Antworten intendiert sind. Vorangegangene Fragen (die sogenannte Kontextfrage) aktivieren c) darüber hinaus ebenfalls situativ Informationen, welche zu deren Interpretation und Antwortgenerierung benötigt werden und die zum Zeitpunkt der Interpretation einer nachfolgenden Frage (sog. Zielfrage) nach wie vor verfügbar sind. Die (2) Generierung einer Meinung auf Basis der Interpretation der Frage ist nun im Wesentlichen davon abhängig, ob in Befragten bereits eine Meinung hinsichtlich der Frage verfügbar ist, z. B. weil im Vorfeld der Umfrage bereits über die Thematik nachgedacht wurde und/oder generell eine starke Meinung zum Objekt der Befragung vorhanden ist. Ist eine solche Meinung demnach bereits vorhanden, entspricht die Generierung einer Meinung lediglich einem Abrufen der Meinung aus dem Gedächtnis. Ist eine solche Meinung jedoch nicht vorhanden, so muss besagte Meinung auf der Basis verfügbarer Informationen generiert werden. Wird eine Meinung generiert, so müssen Befragte hierfür Informationen aus dem Gedächtnis abrufen. Aufgrund impliziter zeitlicher Restriktionen in Umfragesituationen ist es für Befragte jedoch nicht praktikabel, alle verfügbaren Informationen, welche zu einer Meinungsbildung herangezogen werden können, abzurufen und zu evaluieren. Stattdessen verwenden Befragte lediglich die zugänglichsten Informationen zur Meinungsbildung, wobei besagte Zugänglichkeit wiederum davon abhängig ist, a) wie häufig die jeweiligen Informationen bereits abgerufen wurden (je häufiger Informationen abgerufen werden, desto zugänglicher sind diese) und b) wie lange der zeitliche Abstand zur letzten Aktivierung, d.h. Abrufung von Informationen ist (je länger her die letzte Aktivierung von Informationen ist, desto geringer sind diese zugänglich). Ähnlich wie bei der Interpretation der Frage können dabei auch bei der Meinungsbildung gerade dann Informationen hoch zugänglich sein, so diese bereits durch die Kontextfrage aktiviert wurden. Ist eine Meinung auf diese Art abgerufen oder gebildet, so erfolgt im nächsten Schritt (3) die Anpassung der Meinung an das vorgegebene Antwortformat. Ähnlich wie bei der Interpretation der Frage werden die vorgegebenen Antwortkategorien dazu verwendet zu antizipieren, wo Befragte sich innerhalb der durch die Forscher bereitgestellten Intervalle selbst einordnen. So kann die Reichweite einer Antwortskala beispielsweise subjektiv dazu genutzt werden, durchschnittliche Antworten abzuleiten und sich entsprechend gemäß dieser Antwortkategorien als unterdurchschnittlich, durchschnittlich oder überdurchschnittlich einzustufen. Haben Befragte sich schließlich für eine passende Antwortkategorie entschieden, so kann

es vor der Gabe der Antwort (4) zu einem finalen Editierungsprozess kommen, so die eigentliche Antwort als sozial eher unerwünscht bzw. situativ nicht akzeptabel wahrgenommen wird. Befragte können so beispielsweise geneigt sein, sich bei Selbstbewertungen gegenüber Interviewenden positiver darzustellen oder bestimmte, als unerwünscht eingeschätzte Meinungen abzuschwächen. Ob es zu einem solchen Editierungsprozess kommt ist dabei im Wesentlichen davon abhängig, als wie privat Befragte die Umfragesituation wahrnehmen.

Gemäß dieses Modelles ist die Generierung einer Antwort folglich ebenfalls abhängig von Syntax, Semantik und Pragmatik oder zusammengefasst, der a) Frageformulierung und b) den kognitiven Fähigkeiten Befragter die benötigt werden, die Frage zu verstehen, aber auch korrekt zu interpretieren, c) der Art und der Verfügbarkeit (Häufigkeit und zeitliche Nähe der Aktivierung) der Informationen, welche zur Interpretation der Frage sowie der Bildung einer Meinung verwendet werden, d) der vorgegebenen Antwortkategorien, e) der Wahrnehmung der Umfragesituation als Privat sowie f) der Wahrnehmung der eigentlichen Antwort als situativ akzeptabel. Innerhalb dieses Modells kommt der Zugänglichkeit von Informationen für die Generierung einer Antwort somit eine besondere Bedeutung zu, da ein wiederkehrender Einfluss an verschiedenen Stellen des Antwortprozesses festzustellen ist. Diesbezüglich wird insbesondere die Aktivierung von Informationen durch vorangegangene Fragen (Kontextfragen) betont, wodurch an dieser Stelle bereits ein erster Hinweis bezüglich der Erklärung von FRE gegeben ist, worauf in Kapitel 2.2 und 2.3 vertiefend eingegangen wird.

Verzerrte, d.h. unwahre Antworten werden demnach auch nach Strack und Martin (1987) sowie Sudman und Kollegen (1996) vor allem dann erwartet, wenn Fragen falsch Interpretiert werden (z.B. aufgrund eines Mangels kognitiver Fähigkeiten, aufgrund eines fehlenden Bedürfnisses zur Kooperation in Umfragen oder wenn irrelevante Informationen als Basis der Interpretation verwendet werden), wenn vorhandene Meinungen fehlerhaft aus dem Gedächtnis abgerufen werden, wenn neue Meinungen nicht auf der Basis der bestmöglichen verfügbaren Informationen gebildet werden, wenn die vorgegebenen Antwortkategorien nicht zu der für die Antwort vorgesehenen Meinung passen und wenn es aufgrund der Vermeidung des Gebens einer sozial unerwünschten Antwort, z.B. aufgrund fehlender Privatheit der Umfragesituation, zu einem finalen Editierungsprozess kommt.

Fünf Schritte der Antwortgenerierung nach Cannell et al.

Cannell und Kollegen (1981) differenzieren hingegen fünf Schritte: 1.) Die Interpretation der Frage; 2.) Die Prozessierung von Informationen; 3.) Die Bildung einer Antwort, welches auf Basis der prozessierten Informationen als adäquat erachtet wird; 4.) Die Evaluation der Antwort

hinsichtlich ihrer Kollision oder Verträglichkeit mit persönlichen Zielen; und 5.) Die Bildung der finalen Antwort. Während diese fünf Schritte den optimalen Prozess zur Bildung einer angemessenen Antwort nachzeichnen, wird ebenfalls die Möglichkeit auftretender Antwortverzerrungen berücksichtigt. So wird das Modell weiterhin ergänzt um weitere mögliche Schritte: 6.) Die Editierung der Antwort; und 7.) Die Bildung der finalen, editierten Antwort.

Der erste Schritt (1), die Interpretation der Frage, wird auch in diesem Modell beschrieben als abhängig von Syntax, Semantik und Pragmatik der Frageformulierungen, wobei Cannell, Miller, und Oksenberg (ebd.) bezüglich dieser Phase spezifisch die Möglichkeit hervorheben, dass in Befragten mehrere mögliche Interpretationen konkurrieren können. In diesem Fall bewerten und evaluieren Befragte zunächst die möglichen Interpretationen um zu entscheiden, welche akzeptiert werden soll. Dies bedeutet ebenfalls, dass Befragte Fragen auf der Basis individueller Annahmen bezüglich deren intendierter Bedeutung beantworten. Sind solche Annahmen gebildet und Fragen interpretiert, so erfolgt im nächsten Schritt (2) die Prozessierung relevanter Informationen für die Antwortgenerierung. Dieser Schritt umfasst eine Bewertung dessen, welche Informationen zur Bildung einer Antwort herangezogen werden, wobei besagte Bewertung auch in Abhängigkeit der Verfügbarkeit von Hinweisreizen und des Referenzrahmens stattfindet. Typischerweise beinhaltet dieser Prozess das Abrufen relevanter Informationen aus dem Gedächtnis sowie die Organisation besagter Informationen. Sind die Informationen prozessiert, so (3) evaluieren Befragte als nächstes die potenzielle Antwort hinsichtlich der intendierten Anforderungen der formulierten Frage. Wird die potenzielle Antwort in diesem Schritt als inadäquat hinsichtlich der gestellten Frage bewertet, so springen Befragte zurück zu Schritt zwei, um weitere, besser geeignete Informationen zur Antwortgenerierung zu prozessieren. Wird die potenzielle Antwort hingegen als adäquat hinsichtlich der gestellten Frage bewertet, so erfolgt in einem nächsten Schritt (4) eine weitere Evaluation der potenziellen Antwort, dieses Mal jedoch hinsichtlich der psychologischen Bedeutung der potenziellen Antwort mit Bezug auf persönliche Ziele der Befragten. So können potenzielle Antworten persönliche Ziele wie etwa eine möglichst positive Selbstdarstellung gefährden, während andere Ziele wie etwa die Gabe möglichst wahrer Antworten durch die potenzielle Antwort ungefährdet bleiben. Ist letzteres der Fall, so kommt es (5) schließlich zur Gabe der Antwort ohne weitere Editierung.

Über diesen optimalen Pfad der Antwortgenerierung (Schritte eins bis fünf) hinaus berücksichtigt das Modell ebenfalls Abweichungen im Prozess der Antwortgenerierung, welche schließlich zu Antwortverzerrungen führen. So existiert ein zu Schritt zwei bis vier alternativer Schritt (6) zu welchem Befragte abweichen, die die Frage nicht richtig verstehen oder interpretieren, die nicht die (kognitiven) Fähigkeiten oder die Motivation besitzen, relevante Informationen zu

prozessieren, deren Evaluation der potenziellen Antwort hinsichtlich der gestellten Frage an der Zugänglichkeit weiterer relevanter Informationen scheitert oder deren Evaluation der potenziellen Antwort hinsichtlich weiterer persönlicher Ziele einen Konflikt offenbart und eine Editierung der Antwort erfordert. Antworten in Schritt sechs werden demnach nicht mehr auf der Basis relevanter Informationen, sondern auf der Basis situativer Hinweisreize wie Eigenschaften der Interviewer, Spezifika der Frage und von vorangegangenen Fragen oder persönlicher weiterer Einstellungen, Überzeugungen, Motiven und Ziele formuliert, was schließlich (7) in der Gabe inadäquater Antworten mündet.

Erneut ist die Generierung einer Antwort gemäß diesem Modell abhängig von Syntax, Semantik und Pragmatik oder zusammengefasst, der a) Frageformulierung und b) den kognitiven Fähigkeiten Befragter die benötigt werden, die Frage zu verstehen, aber auch korrekt zu interpretieren. Hinzu kommen c) die Zugänglichkeit von für die Antwortgenerierung relevanter Informationen und d) den Zielen und Motiven Befragter für die Befragung.

Verzerrte Antworten sind gemäß diesem Modell dann zu erwarten, wenn Befragte nicht dem Prozess gemäß den Schritten eins bis fünf folgen, sondern auf den alternativen Pfad, welcher die Schritte sechs und sieben umfasst, abweichen. Dies ist der Fall, wenn Befragte Fragen nicht richtig interpretieren oder verstehen, wenn nicht alle notwendigen Informationen zur Antwortgenerierung vorliegen oder wenn keine Motivation besteht, alle relevanten Informationen abzurufen oder wenn die Gabe einer potenziellen Antwort einen Zielkonflikt innerhalb Befragter auslöst. In diesem Fall ist die Antwortgenerierung beeinflusst durch persönliche Motive und situative Hinweisreize, was schließlich zu verzerrten Antworten führt.

Aggregation generischer Determinanten des Befragtenverhaltens

Ob der Prozess der Antwortgenerierung nun nach den definierten Schritten von Tourangeau und Kollegen (2000), Strack und Martin (1987), Sudman und Kollegen (1996) und Cannell und Kollegen (1981) beschrieben wird ist mehr oder weniger generisch. So finden sich in allen beschriebenen Modellen prinzipiell die folgenden Schritte wieder, auch wenn die Sequenzierung der einzelnen Prozessschritte je Modell different sein kann: 1.) kommt es zu einer Phase des Verstehens und des Interpretierens einer Frage. 2.) werden, nachdem Inhalt, Kontext und Bedeutung der Frage (korrekt oder nicht) interpretiert wurden, benötigte Informationen abgerufen, um hieraus eine Meinung oder ein Urteil zu bilden. 3.) erfolgt die weitere Prozessierung der abgerufenen Informationen, welche eine Bewertung der Informationen hinsichtlich einer Antwortbildung beinhaltet und welche schließlich zu einer vorläufigen Formulierung der potenzi-

ellen Antwort bzw. der Bildung einer Meinung führt. 4.) erfolgt eine Anpassung der potenziellen Antwort hinsichtlich des möglichen Antwortformates bzw. die Editierung der potenziellen Antwort hinsichtlich persönlicher Motive, bevor 5.) die finale Antwort formuliert und gegeben wird.

Hinzu kommen die folgenden Annahmen bezüglich der Antwortgenerierung und des Antwortverhaltens: 1.) wird davon ausgegangen, dass Befragte unter gewissen Umständen keine optimale Antwort im Sinne einer „wahren" Antwort geben, sondern Antworten mitunter verzerrt sein können. 2.) wird davon ausgegangen, dass solche Verzerrungen unbewusst (z.B. durch fehlerhafte Interpretationen), aber ebenso auch bewusst (z.B. durch Editierung) auftreten können. 3.) werden Befragten Motive und persönliche Ziele zugeschrieben, welche unabhängig vom Kontext der eigentlichen Befragung sein können, die sich jedoch ebenfalls auf das Antwortverhalten auswirken. Und 4.) wird dem Abrufen und dem Prozessieren von Informationen entlang des Prozesses der Antwortgenerierung eine besondere Bedeutung zugemessen, da Befragte nicht alle potenziell zur Verfügung stehenden, relevanten Informationen zur Urteilsbildung verwenden, sondern es vor allem eher zugängliche Informationen sind, welche für die Generierung einer Antwort verwendet werden. Informationen werden somit unter Umständen nicht nach ihrer Relevanz, sondern nach Ihrer Zugänglichkeit bewertet und verwendet.

Kognitive Faktoren mit Einfluss auf das Antwortverhalten sind zusammenfassend demnach: a) Die grundlegenden kognitiven Fähigkeiten die benötigt werden, eine Frage verstehen und interpretieren zu können; b) die kognitive Zugänglichkeit von Informationen, welche zur Antwortgenerierung verwendet werden; und c) die zugrundeliegenden privaten Motive und Ziele Befragter, welche sich in der Wahl einer spezifischen Antwortstrategie äußern.

Hieraus ergibt sich aus kognitiver Perspektive, dass verzerrte Antworten vor allem dann zu erwarten sind, a) wenn Befragte Fragen nicht korrekt entsprechend der intendierten Bedeutung interpretieren; b) wenn nicht alle relevanten Informationen (aufgrund deren Zugänglichkeit bzw. der selektierten Antwortstrategie) zur Urteilsbildung herangezogen werden; c) wenn irrelevante oder lediglich situativ verfügbare Informationen zur Bildung einer Antwort herangezogen werden; sowie d) wenn aufgrund persönlicher Motive Befragter Zielkonflikte durch die Gabe einer „wahren" Antwort auftreten.

Bezüglich der Zugänglichkeit von Informationen, welche (neben grundlegenden kognitiven Fähigkeiten und der persönlichen Motive Befragter) den Modellen entsprechend eine herausragende Rolle in der Generierung „wahrer" Antworten einnimmt, merkt Schwarz (1991) an, dass

Befragte nicht immer alle für die Antwortgenerierung relevanten Informationen berücksichtigen, sondern den impliziten Suchprozess mitunter vorzeitig abbrechen, sobald mit subjektiv hinreichend großer Sicherheit ein Urteil gebildet werden kann. Urteile beruhen somit häufig auf denjenigen Informationen, welche Befragten zum Zeitpunkt der Urteilsbildung kognitiv am ehesten zugänglich sind. Dies können somit zum einen chronisch zugängliche Informationen sein, über welche Befragte schon viel nachgedacht haben und welche daher im Gedächtnis eher verfügbar sind (Higgins & Bargh, 1987), zum anderen aber auch Informationen, welche lediglich unter bestimmten Umständen, das heißt situativ verfügbar sind, z. B. weil Befragte diese Informationen bereits zum Beantworten einer vorangegangenen Frage aktiviert haben.

2.1.2 Implizite Verhaltensmodelle mit Fokus auf differente Elaborationsgrade

Neben diesen expliziten, klassischen Modellen des Befragtenverhaltens können ebenfalls implizite Verhaltensmodelle betrachtet werden, welche sich nicht konkret auf eine Beschreibung des Befragtenverhaltens fokussieren, sondern generischen Verhaltensmodellen entsprechen und welche sich für das spezifische Verhalten „Beantwortung von Fragen“ adaptieren lassen. Da bereits identifiziert wurde, dass Antworten in Umfragen bewusst oder unbewusst verzerrt sein können, werden diese impliziten Verhaltensmodelle hinsichtlich möglicher Erklärungen solcher verzerrten Urteilsbildungen betrachtet.

Das Satisficing-Optimizing Model

Während der Einfluss der Zugänglichkeit von Informationen sowie der persönlichen Motive auf die Urteilsbildung in Befragungssituationen später in diesem Kapitel vertiefend beleuchtet wird, finden sich in der Literatur rund um die psychologischen Prozesse des Antwortverhaltens ebenfalls Theorien mit welchen erklärt werden kann, unter welchen Umständen Befragte den impliziten Suchprozess nach relevanten Informationen vorzeitig abbrechen, hiervon am bedeutsamsten das Satisficing-Optimizing Model (Krosnick, 1991; 1999a; Krosnick & Fabrigar, 1997). Dem Modell zugrunde liegt die Theorie, dass sich kognitive Prozesse in zwei Modi der Informationsverarbeitung differenzieren lassen, das Optimizing und das Satisficing. Von erstem ist die Rede, wenn bei der Prozessierung von Informationen ein hoher kognitiver Aufwand betrieben wird, von zweitem, wenn kognitiv aufwändige Prozesse vermieden werden. Oder auf das Antwortverhalten in Umfragen adaptiert: Betreiben befragte Optimizing, so betreiben sie

einen hohen kognitiven Aufwand bei der Prozessierung von Informationen für die Antwortgenerierung, betreiben Befragte hingegen Satisficing, so folgt lediglich ein geringer kognitiver Aufwand.

Werden die bisher aufgezeigten, klassischen Theorien des Befragtenverhaltens zugrunde gelegt so stellen Krosnick und Fabrigar (1998) fest, dass ein relativ großer kognitiver Aufwand von Befragten betrieben werden muss, wenn diese tatsächlich jeden einzelnen Prozessschritt der nötig ist, um eine optimale Antwort zu generieren, für jede der gestellten Fragen durchlaufen. Da Befragungen in der Regel nicht nur aus lediglich einer einzelnen Frage bestehen, kumuliert sich der kognitive Aufwand Befragter entsprechend entlang einer Befragung, wobei persönlichen Motive vorhanden sind die darüber entscheiden, ob Befragte diesen Aufwand tatsächlich betreiben. Solche Motive sind etwa das Bedürfnis nach Selbstdarstellung, nach zwischenmenschlicher Interaktion, nach intellektuellen Herausforderungen, nach Selbst-Erkenntnis, nach persönlicher Katharsis, nach Dankbarkeit für eine erzielte Leistung oder aus Altruismus. Führen solche Motive nun dazu, dass Befragte jeden einzelnen der benötigten Schritte zur Antwortgenerierung mit hohem kognitiven Aufwand betreiben, so wird die entsprechende Strategie zur Antwortgenerierung als Optimizing bezeichnet (Krosnick, 1999a).

In der Realität gilt es jedoch als höchst unrealistisch, dass alle Befragte für jede der Fragen Optimizing betreiben, schlicht, da unterschiedliche Grade an Motivation zur Beantwortung der Fragen vorhanden sind. Manche Befragte können beispielsweise am Thema einer Befragung hoch interessiert sein und somit einen besonders hohen Grad an Motivation für das Betreiben von Optimizing mitbringen, während andere Befragte die Teilnahme an einer Befragung, etwa einer Lehrevaluation an Universitäten, eher als Pflicht wahrnehmen und deshalb nur einen geringen Grad an Motivation mitbringen. Ebenso können beispielsweise für zu Beginn hoch motivierte Befragte im Verlauf einer Befragung Ermüdungserscheinungen auftreten, so dass die Motivation für das Betreiben von Optimizing im Verlauf einer Befragung abnehmen kann. Ist dies der Fall, so wählen Befragte eine andere, dem Optimizing gegenübergestellte Antwortstrategie, das Satisficing (Krosnick, 1991).

Statt also Anstrengungen zur Generierung optimaler Antworten zu maximieren ist es möglich, dass Befragte weniger Energie investieren, in dem sie z.B. Fragen oberflächlich statt gründlich erfassen, Informationen lediglich oberflächlich abrufen und prozessieren, Meinungen mit geringerem kognitivem Aufwand bilden oder Antworten weniger sorgfältig selektieren. Die Schritte zur Generierung einer Antwort werden somit zwar vollständig, jedoch oberflächlicher

durchlaufen, als wenn Optimizing betrieben werden würde, was als schwaches Satisficing bezeichnet wird. Betreiben Befragte daneben starkes Satisficing, so überspringen sie gar Phasen der Antwortgenerierung, z.B. in dem Fragen nicht auf Basis deren Inhalt und Kontext, sondern auf Basis situativer Hinweisreize wie dem Vorhandensein einzelner Schlüsselwörter interpretiert werden, woraufhin auch das Abrufen und Prozessieren relevanter Informationen nicht auf der Basis optimaler Interpretationen verläuft, sondern dem Vorhandensein situativer Hinweisreize basiert. Neben solchen situativen Hinweisreizen basieren Antworten, die auf Basis einer Strategie des Satisficing generiert wurden, auf einfachen Antwortheuristiken, welche ohne hohen kognitiven Aufwand angewandt werden können (Krosnick, 1999a).

Als Determinanten für das Auftreten der jeweiligen Antwortstrategien nennt Krosnick (1991) a) die Komplexität der zu lösenden Aufgabe; b) die (kognitiven) Fähigkeiten Befragter, um Optimizing zu betreiben; und c) die Motivation Befragter, um Optimizing zu betreiben. Je höher also die Komplexität der Aufgabe und je geringer die Fähigkeiten und die Motivation zum Betreiben von Optimizing in Befragten vorhanden sind, desto wahrscheinlicher wird es, dass diese eine Antwortstrategie des mehr oder weniger starken Satisficing anwenden (Krosnick, Narayan, & Smith, 1996). Die Komplexität der zu lösenden Aufgabe (a) beschreibt Krosnick (1999b) weiterhin als Funktion von unter anderem dem kognitiven Aufwand der nötig ist, die Bedeutung der Frage und der Antwortoptionen korrekt zu interpretieren; dem kognitiven Aufwand der nötig ist, relevante Informationen abzurufen und zu prozessieren; der Geschwindigkeit, in der Interviewer Fragen vorlesen; und der Anwesenheit von ablenkenden Ereignissen. Die kognitiven Fähigkeiten (b) sind weiterhin für Befragte höher, die häufiger komplexe mentale Operationen prozessieren; die Übung darin haben, über Themen von Befragungen nachzudenken; und denen bereits im Vorfeld zur Umfrage mentale Bewertungen bezüglich des Themas der Befragung vorliegen. Faktoren zuletzt, die die Motivation (c) positiv beeinflussen sind das Kognitionsbedürfnis (Cacioppo, Petty, Feinstein, & Jarvis, 1996);das persönliche Interesse am Inhalt der Befragung; die Überzeugung, dass die durchgeführte Befragung sinnvolle Konsequenzen mit sich bringt; das Verhalten von Interviewern; sowie die Müdigkeit Befragter.

Gegenübergestellt können Optimizing und starkes Satisficing ebenfalls als Pole eines Kontinuums beschrieben werden, welches den Grad an Elaboration beschreibt, mit welchem eine Antwort generiert wird. Der Optimizing-Pol beinhaltet dabei das komplette und elaborierte Durchlaufen aller kognitiven Schritte, welche zur Generierung einer optimalen Antwort benötigt werden. Der Pol des starken Satisficing beinhaltet hingegen eine oberflächliche Phase der Interpre-

tation und der Antwortformulierung sowie kein adäquates Abrufen und Prozessieren von Informationen. Zwischen diesen Polen finden sich verschiedene Grade des schwachen Satisficing (Krosnick 1999a).

Zusammengefasst und mit konkretem Bezug auf das Antwortverhalten und das Auftreten von Antwortverzerrungen bedeutet dies, dass Befragte, welche jeden einzelnen benötigten Schritt zur Generierung einer Antwort in Befragungen hoch elaboriert durchlaufen, eine Antwortstrategie des Optimizing selektieren. Durchlaufen Befragte hingegen die einzelnen Prozessschritte nicht hoch elaboriert oder überspringen diese gar einzelne Schritte, so betreiben diese eine Antwortstrategie des Satisficing. Optimizing und (starkes) Satisficing können folglich ebenfalls als Pole eines kognitiven Elaborationskontinuums mit dem Optimizing als Pol höchster Elaboration und dem (starken) Satisficing als Pol geringster Elaboration betrachtet werden. Wie hoch oder gering der Grad an kognitiver Elaboration bei der Beantwortung einer Frage ausfällt ist davon abhängig, als wie Komplex sich die zu lösende Aufgabe darstellt (je Komplexer die Aufgabe, desto wahrscheinlicher betreiben Befragte Satisficing), ob Befragte über die benötigten kognitiven Fähigkeiten zum Betreiben von Optimizing verfügen (je geringer die kognitiven Fähigkeiten, desto wahrscheinlicher betreiben Befragte Satisficing) und ob Befragte über die benötigten Motivation zum Betreiben von Optimizing verfügen (je geringer die Motivation, desto wahrscheinlicher betreiben Befragte Satisficing). Befragte, die Optimizing betreiben, generieren optimale Antworten in dem Sinne, dass jeder einzelne Schritt der Antwortgenerierung elaboriert durchlaufen wurde. Solche optimalen Antworten müssen hingegen keine wahren Antworten sein, da die bisher vorgestellten Modelle ebenfalls Phasen der Antworteditierung vorsehen, welche auch bei einer Antwortstrategie des Optimizings elaboriert durchlaufen werden können, was trotz der selektierten Antwortstrategie zu Verzerrungen führen kann. Kommt es trotz Optimizing demnach zu Verzerrungen so ist davon auszugehen, dass diese von Befragten bewusst vorgenommen werden. Befragte, die Satisficing betreiben, generieren hingegen keine optimalen Antworten in dem Sinne, dass nicht jeder der Modellschritte elaboriert durchlaufen wird. Antworten basieren somit nicht zwangsläufig auf Dispositionen Befragter, sondern etwa auf Heuristiken und situativen Hinweisreizen (z. B. ausgelöst durch die vorangegangene Frage). Kommt es demnach während der Antwortgenerierung, welcher eine Strategie des Satisficing zugrunde liegt, zu Antwortverzerrungen so ist davon auszugehen, dass Antworten durch situative Hinweisreize und Antwortheuristiken verzerrt sind und Verzerrungen somit eher unbewusst auftreten.

Das Modell spontanen Prozessierens

Mit der Etablierung des Satisficing-Optimizing Models wird explizit berücksichtigt, dass Informationsverarbeitungsprozesse (wie die Generierung von Antworten in Umfragen) verschiedenen kognitiven Strategien unterliegen, welche sich, wie für dieses Modell gezeigt, entlang eines Elaborationskontinuums differenzieren lassen. Bezüglich Elaboration und kognitivem Aufwand führt Fazio (1990) an, dass der Großteil menschlichen Verhaltens in sozialen Situationen schnell und spontan verlaufen muss, wird die relative Leichtigkeit betrachtet, mit der sich Menschen in ihrer Umwelt bewegen, wobei eben in diesen spontanen Entscheidungsprozessen die heuristische Funktion von Einstellungen einsetzen, welche in ihrer Eigenschaft als Bilanzurteile eines Objektes dazu beitragen, dass Situationen schnell eingeschätzt werden können, um hierauf aufbauend Verhalten zu aktivieren, ohne in jeder Situation mit Handlungsbedarf eine elaborierte und kognitiv aufwändige Entscheidung treffen zu müssen. Diese Überlegungen basieren auf Fazios Modell spontanen Prozessierens (Fazio, 1986), welches schnelles, automatisches und spontanes Verhalten von Personen als Funktion der Wahrnehmung eines Ereignisses versteht, wobei sich besagte Wahrnehmung wiederum aus der kontextbezogenen Definition der Situation sowie der von der Einstellung abhängigen, selektiven Wahrnehmung der Situation bzw. des Einstellungsobjektes ergibt. Das Modell spontanen Prozessierens entspricht somit ebenfalls einem impliziten, (kognitiven) Verhaltensmodell, da es nicht explizit für das spezifische Verhalten „Generierung von Antworten in Umfragen“ entwickelt wurde, sondern allgemeines Prozessieren beschreibt, welches sich für diesen Spezialfall adaptieren lässt.

Im Detail postuliert Fazio (1986; 1990), dass in alltäglichen Entscheidungssituationen, in welchen das ständige elaborieren von Entscheidungen dysfunktional und unpraktikabel wäre, soziales Verhalten eine Funktion der individuellen Interpretation der gegenwärtigen Situation und der Wahrnehmung des Objektes mit Handlungsbedarf darstellt. In spontanen Entscheidungssituationen sind es entsprechend des Modells somit zum einen Einstellungen gegenüber handlungsrelevanten Objekten sowie zum anderen Interpretationen der Situationen mit Handlungsbedarf, welche Verhaltensweisen leiten. Solche Einstellungen müssen in entsprechenden Handlungssituationen nicht zwangsläufig bewusst abgerufen werden, sondern können sich ebenfalls automatisch aktivieren. Daneben ist es jedoch die Definitionen und Interpretation der entsprechenden Situationen welche darüber entscheidet, welche konkreten Handlungen selektiert werden. Besagte Definitionen können zum einen durch Hinweisreize der Situation beeinflusst werden: nehmen Individuen etwa aufsteigenden Rauch in der Nähe war, so kann die Situationen entweder als gefahrvoll interpretiert werden, wenn beobachtet wird, dass andere Individuen panisch auf den Rauch reagieren, wobei in diesem Fall die Handlung „weglaufen“ selektiert

werden könnte. Ebenso ist es möglich, dass die gleiche Situation als harmlos interpretiert wird, wenn beobachtet werden kann, dass andere Individuen gelassen auf den Rauch reagieren, woraufhin die Handlung „ruhig bleiben“ selektiert werden kann. Solche Hinweisreize müssen sich jedoch nicht zwangsläufig aus der Situation bzw. der Umgebung (wie etwa dem Verhalten anderer) ergeben, sondern können ebenfalls aus aktivierten, relevanten Konstrukten aus dem Gedächtnis stammen, da Wahrnehmungen und damit Interpretationen ebenfalls abhängig sind von individuellen Wissensstrukturen, Affekten, Werten und Erwartungen. Die Aktivierung solcher Konstrukte, auch als Priming bezeichnet, entfaltet ihren Einfluss vor allem passiv, automatisch und mitunter auch unbewusst und ist hinreichend, um Interpretationen zu beeinflussen. Entsprechend sind Einstellungen und spontanes Verhalten umso wahrscheinlicher konsistent, desto kongruenter die Definition der handlungsrelevanten Situation und die Einstellung gegenüber dem handlungsrelevanten Objekt sind.

Ob eine Einstellung gegenüber einem Objekt nun jedoch die Wahrnehmung und Interpretation eines Objektes beeinflusst, welche wiederum das Verhalten gegenüber dem Objekt in der gegebenen Situation beeinflusst ist grundsätzlich davon abhängig, ob eine entsprechende Einstellung überhaupt aktiviert bzw. aus dem Gedächtnis abgerufen wird. Solche Einstellungen werden bei Wahrnehmungen von Einstellungsobjekten entsprechend des Modells umso wahrscheinlicher aktiviert, umso höher die chronische Zugänglichkeit besagter Einstellung ist. Die chronische Zugänglichkeit von Einstellungen kann hierbei begriffen werden als die Bereitschaft von Einstellungen, in Informationsverarbeitungsprozessen verwendet zu werden (Eagly & Chaiken, 1993) bzw. die Einfachheit, mit der Informationen abgerufen werden können (Shrum & O’Guinn, 1993).

Bezüglich der Einstellungszugänglichkeit merkt Fazio (1990) an, dass Einstellungen verstanden werden können als gedächtnishafte Assoziation zwischen einem Objekt und der persönlichen Evaluation des Objektes. Entsprechend ist diesem Verständnis implizit, dass die Stärke von Einstellungen individuell variieren, da individuelle Varianzen zwischen der Stärke der Assoziation zwischen Objekten und deren Evaluation existieren. Es ist diese Stärke der Assoziation, welche die chronische Zugänglichkeit von Einstellungen determiniert und welche die Wahrscheinlichkeit bedingt, dass eine Einstellung automatisch aktiviert wird, so das Objekt einer Einstellung wahrgenommen wird.

Entsprechend dieses Modells ist spontanes Verhalten bezüglich eines Einstellungsobjektes dann determiniert von der entsprechenden Einstellung gegenüber dem Objekt, wenn die Einstellung bei Wahrnehmung des Objektes aktiviert wird, wobei die Wahrscheinlichkeit hierfür

mit zunehmender, chronischer Einstellungszugänglichkeit zunimmt. Hinzu kommt die individuelle Definition der Situation, welche wiederum durch Normen bestimmt ist. In Kombination führt die Definition der Situation sowie die durch die Aktivierung einer Einstellung bedingte Wahrnehmung des Einstellungsobjektes zu einer Interpretation eines Ereignisses mit Handlungsbedarf, wobei sich hiervon wiederum das Verhalten ableitet. Entsprechend zeigt sich, dass spontanes Verhalten keiner Elaboration bedarf, da dieses sich aus automatisch aktivierten Einstellungen und/oder normbedingten Definitionen von Situationen und gegebenenfalls Entscheidungsheuristiken ergibt.

Die chronische Einstellungszugänglichkeit dient somit in spontanen Situationen als Moderator zwischen Einstellungen bezüglich eines Objektes und der Wahrnehmung und Interpretation des entsprechenden Objektes, welche schließlich zu Verhalten führt. Je höher die chronische Zugänglichkeit von Einstellungen, desto wahrscheinlicher werden entsprechende Einstellungen aktiviert und bedingen die Wahrnehmung von Einstellungsobjekten, welche, entsprechend der normbedingten Interpretation der Situation, zu einstellungskonformen Verhalten führt. Sind Einstellungen hingegen weniger oder nicht chronisch zugänglich, so sind es eher situativ verfügbare Informationen sowie Heuristiken, welche die Wahrnehmung des Objektes bedingen und somit wahrscheinlich zu nicht-einstellungskonformem Verhalten führen. Die chronische Zugänglichkeit einer Einstellung innerhalb automatisch-spontaner Entscheidungsprozesse wird somit zu einem wichtigen Prädiktor einstellungskonformen Verhaltens.

Mit konkretem Bezug auf spontanes Antwortverhalten bedeutet dies im Kontext dieses Modells, dass Befragte bei Wahrnehmung des Inhaltes einer Frage (was in diesem Kontext dem Einstellungsobjekt entspricht) mit zunehmender Wahrscheinlichkeit eine Einstellung aktivieren, welche sich auf den Inhalt der Frage richtet, je stärker chronisch zugänglich besagte Einstellung ist. Ist die Einstellung entsprechend aktiviert, so beeinflusst diese die weitere Definition und Interpretation der Frage. Zusätzlich erfolgt eine Interpretation des Kontextes der Befragung (der Situation), welche durch zugrundeliegende Normen beeinflusst wird. Je nach Interpretation der Frage und der Befragungssituation werden dann im Folgenden mehr oder weniger kongruente Antworten generiert. Die spontane Generierung einer Antwort ist demnach abhängig von den folgenden kognitiven Determinanten: a) Der konkreten Einstellung bezüglich dem Objekt der Frage, welche die Interpretation der Frage (dem Einstellungsobjekt) beeinflusst; b) der chronischen Zugänglichkeit der Einstellung bezüglich dem Objekt der Befragung, welche die Wahrscheinlichkeit der Aktivierung der spezifischen Einstellung beeinflusst; und c) zugrundeliegenden Normen, welche die Definition und Interpretation der Befragungssituation determinieren.

Antwortverzerrungen sind entsprechend dieses Modells in spontanen Entscheidungssituationen zu erwarten, wenn relevante Einstellungen bezüglich dem Objekt der Befragung chronisch nicht zugänglich sind und folglich nicht aktiviert werden, wenn keine kongruenten Einstellungen bezüglich dem Objekt der Befragung existieren oder wenn zugrundeliegende Normen zu Interpretationen der Situation führen, welche eine kongruente Einstellungs-Verhaltensrelation verhindern.

Theory of reasoned action (TRA) und theory of planned behavior (TPB)

Demgegenüber existieren ebenfalls implizite (kognitive) Verhaltenstheorien, welche das elaborierte, d.h. wohl überlegte und nicht-spontane Prozessieren beleuchten, unter diesen am populärsten die Theory of reasoned action (Ajzen & Fishbein, 1980) und hierauf basierend die Theory of planned behavior (Ajzen, 1991; Fishbein & Ajzen, 2010). Diese zuletzt genannte Theorie besagt im Wesentlichen, dass die Beziehung zwischen Einstellung und geplantem bzw. elaboriertem Verhalten von einer weiteren Variable, der Verhaltensintention interveniert wird. Die Einstellung übt somit in elaborierten, nicht-spontanen Entscheidungssituationen keinen direkten Effekt auf das Verhalten aus, welches stattdessen durch eine Verhaltensintention sowie gegebenenfalls der Verhaltenskontrolle beeinflusst wird. Die Verhaltensintention wiederum ist bedingt durch die Einstellung gegenüber diesem Verhalten, einer subjektiv wahrgenommenen sozialen Norm sowie der Verhaltenskontrolle. Diese drei Prädiktoren der Verhaltensintention wiederum werden von bestimmten „Beliefs“ (Überzeugungen) beeinflusst. Der Begriff der Verhaltensintention bezieht sich somit auf motivationale Aspekte, welche die Absicht umfassen, wie viel Aufwand und Anstrengung für die Durchführung eines Verhaltens aufgebracht wird (Ajzen, 1991; Fishbein & Ajzen, 2010).

Grundsätzlich handelt es sich somit bei der TPB, äquivalent zum Modell spontanen Prozessierens ebenfalls um eine Theorie, welche ein gezeigtes Verhalten auf eine zugrundeliegende Einstellung zurückführt. In Abgrenzung impliziert die TPB jedoch keine zugrundeliegenden, spontanen Kognitionsprozesse, sondern beleuchtet Verhalten innerhalb überlegter und elaborierter Kognitionsprozesse, so dass die beiden Modelle nicht als konkurrierend, sondern sich ergänzend betrachtet werden können. Ajzen (1991) postuliert diesbezüglich, dass geplantes, das heißt wohl überlegtes und elaboriertes Verhalten als Funktion einer Handlungsintention, einer Einstellung gegenüber dem Verhalten, subjektiven Normen und der wahrgenommenen Verhaltenskontrolle definiert werden kann.

Gemäß dem Modell handelt es sich bei der Handlungsintention um den stärksten Prädiktor für ein gezeigtes Verhalten, wobei der Intention eine motivationale Komponente zukommt, da sie

als Indikator für die Bereitschaft eines Individuums wahrgenommen werden kann, ein spezifisches Verhalten auszuführen. Darüber hinaus postuliert das Modell jedoch, dass die Intention, ein Verhalten auszuführen, keine hinreichende Bedingung für das Ausführen eines spezifischen Verhaltens darstellt, da der Zusammenhang zwischen der Intention und dem gezeigten Verhalten weiterhin durch die sogenannte wahrgenommene Verhaltenskontrolle bedingt wird. Die wahrgenommene Verhaltenskontrolle kann verstanden werden als Grad, mit welchem Individuen davon überzeugt sind, ein spezifisches Verhalten kontrollieren, sprich, entsprechend eines subjektiv wünschenswerten Ergebnisses erfolgreich ausführen zu können. Alternativ kann sie verstanden werden als Erwartung bezüglich der Mühelosigkeit, mit der ein Verhalten erfolgreich umgesetzt werden kann. Dies bedeutet, dass die Wahrscheinlichkeit, ein spezifisches Verhalten zu zeigen mit der Überzeugung zunimmt, dass das entsprechende Verhalten erfolgreich ausgeführt werden kann und zu einem gewünschten Ergebnis führt, so eine kongruente Intention besteht. Das Konzept der wahrgenommenen Verhaltenskontrolle findet sich in der TPB (Ajzen, 1991) in Abgrenzung zur TRA (Ajzen & Fishbein, 1980) und basiert auf der Self-Efficacy Theory nach Bandura (1977). Demnach basiert die wahrgenommene Verhaltenskontrolle auf der sogenannten Selbstwirksamkeitserwartung, wobei diese verstanden werden kann als die Überzeugung, dass das zur Erzielung der gewünschten Ergebnisse erforderliche Verhalten erfolgreich ausgeführt werden kann (Bandura, 1994). Gemäß der TPB (Ajzen, 1991; Fishbein & Ajzen, 2010) wird ein spezifisches Verhalten entsprechend umso wahrscheinlicher gezeigt, wenn Individuen die Intention gebildet haben, das entsprechende Verhalten auszuführen und wenn Individuen subjektiv davon überzeugt sind, dass sie das für das gewünschte Ergebnis benötigte Verhalten erfolgreich bzw. mit Leichtigkeit ausführen können.

Die Intention, ein spezifisches Verhalten zu zeigen wiederum steht gemäß dem Modell in Abhängigkeit von der zugrundeliegenden Einstellung gegenüber dem entsprechenden Verhalten, subjektiv vorhandenen Normen und ebenfalls von der wahrgenommenen Verhaltenskontrolle, wobei zwischen diesen drei genannten Determinanten ebenfalls wechselseitige, assoziative Verknüpfungen definiert sind.

Einstellungen gegenüber dem gezeigten Verhalten können gemäß dem Modell verstanden werden als individuellen Evaluationen des Verhaltens und formen sich auf der Basis zugrundeliegender, verfügbarer Überzeugungen (Beliefs) gegenüber dem Verhalten, wobei solche Beliefs wiederum definiert werden als subjektiv wahrgenommene Wahrscheinlichkeit, dass das Verhalten zu einem gewissen Ergebnis führt. Im speziellen werden vorangegangene Ergebnisse von Handlungen evaluiert und nehmen damit Einfluss auf die subjektive Wahrnehmung davon, ob ein Verhalten mehr oder weniger wahrscheinlich zu diesem Ergebnis führt, wodurch sich als

Folge dieser Evaluation Beliefs bilden und festigen. Beispielhaft für den Prozess der Antwortgenerierung in Umfragen bedeutet dies, dass Individuen dann eine positive Einstellung gegenüber dem Verhalten „wahrheitsgemäßes Beantworten von Umfragen“ haben, wenn zugrundeliegende Beliefs z. B. lauten, dass das wahrheitsgemäße Beantworten von Umfragen zu besserer wissenschaftlicher Forschung führt, wobei ein solcher Belief sich wiederum dadurch bildet, dass Individuen das Ergebnis „Förderung der Wissenschaft“ in der Vergangenheit bereits als Konsequenz der Handlung „wahrheitsgemäßes Beantworten von Umfragen“ wahrgenommen haben. Ist dies der Fall, so trägt eine solche positive Einstellung mit zunehmender Stärke dazu bei, dass eine entsprechende Intention gebildet wird, das Verhalten, auf welches sich die Einstellung bezieht, zu zeigen.

Neben der Einstellung gegenüber einem Verhalten nehmen subjektiv vorhandene Normen gemäß dem Modell ebenfalls Einfluss auf die Bildung einer Verhaltensintention. Solche Normen können verstanden werden als individuelle Wahrnehmung und Bewertung des entsprechenden Verhaltens, welche wiederum beeinflusst wird von der subjektiven Wahrnehmung der Bewertung des Verhaltens durch dem Individuum nahestehende andere Personen wie z.B. Eltern, Ehepartner, Familienmitglieder etc. Dies bedeutet, dass Individuen bei der Bildung einer Intention nicht nur die eigene Evaluation dessen berücksichtigen, ob das jeweilige Verhalten zu einem spezifischen Ergebnis führt, sondern diese ebenfalls eine mentale Bewertung davon vornehmen, wie andere nahestehende Personen es bewerten würden, wenn das Individuum das entsprechende Verhalten zeigt, sprich, wie akzeptiert bzw. erwünscht das jeweilige Verhalten gegeben der Situation ist.

Zuletzt ist es ebenfalls die bereits weiter oben beschriebene wahrgenommene Verhaltenskontrolle, welche die Intention, ein Verhalten zu zeigen, beeinflusst. Zweifeln Individuen etwa an ihren Fähigkeiten, ein spezifisches Verhalten derart umzusetzen, so dass es zu einem gewünschten Ergebnis führt, so verringert dies die Wahrscheinlichkeit, eine entsprechende Intention bezüglich des Verhaltens auszubilden, selbst wenn eine positive Einstellung und kohärente subjektive Normen vorhanden sind.

Das Modell zusammenfassend wird überlegtes Verhalten von verhaltensrelevanten Beliefs, normativen Beliefs und Kontroll-Beliefs geleitet. Verhaltensrelevanten Beliefs formen eine Einstellung gegenüber dem Verhalten, welche die Wahrnehmung des Verhaltens subjektiv als positiv und wünschenswert oder negativ und abzulehnend gegeben einer subjektiv wünschenswerten Verhaltenskonsequenz darstellt. Normative Beliefs resultieren in subjektiv wahrgenommenen Normen, Kontroll-Beliefs wiederum in der wahrgenommenen Verhaltenskontrolle. In

Kombination führen die drei genannten Komponenten zur Bildung einer mehr oder weniger stark ausgeprägten Verhaltensintention. Diese in Kombination mit der Überzeugung, dass das Verhalten tatsächlich so umgesetzt werden kann, dass es zu einem gewünschten Ergebnis führt, aktiviert schließlich das spezifische Verhalten.

Da es sich bei der TPB somit um ein implizites Verhaltensmodell handelt, welches verschiedene Handlungen generisch auf vorhandene Einstellungen, subjektive Normen und die wahrgenommene Verhaltenskontrolle zurückführt, können damit bezüglich des Antwortverhaltens in Umfragen verschiedene, überlegte und elaborierte Antwortstrategien wie etwa zum einen das Verhalten (1) „wahrheitsgemäßes Beantworten von Fragen als Konsequenz elaborierter Prozesse“ und zum anderen das Verhalten (2) „verzerrtes Beantworten von Fragen als Konsequenz elaborierter Prozesse“ beleuchtet werden.

Die Handlung (1) „wahrheitsgemäßes Beantworten von Fragen“ wird innerhalb elaborierter Prozesse demnach unter den folgenden Bedingungen ausgeführt: Befragte benötigen zunächst a) eine positive Einstellung gegenüber dem wahrheitsgemäßen Beantworten von Fragen in Umfragen, wobei sich eine solche Einstellung durch zugrundeliegende Beliefs wie etwa der Überzeugung bildet, dass dieses Verhalten zu einem subjektiv wünschenswerten Ergebnis wie etwa der Förderung wissenschaftlicher Forschung führt. Entsprechend in der Vergangenheit erlebte Wahrnehmungen tragen wiederum zur Bildung solcher Beliefs bei. Darüber hinaus ist b) eine kohärente subjektive Norm entscheidend, sprich: Befragte erleben das wahrheitsgemäße Beantworten nicht nur selbst als positiv, sondern erwarten subjektiv, dass ihnen nahestehende Personen die Ausführung dieses Verhaltens (nicht die Verhaltenskonsequenz) ebenfalls als positiv bewerten, sprich, dass das Verhalten wünschenswert und akzeptiert ist. Zusätzlich ist c) die wahrgenommene Verhaltenskontrolle von Relevanz: Damit Befragte in Umfragen wahrheitskonform Antworten ist die Überzeugung entscheidend, dass diese über alle erforderlichen Fähigkeiten verfügen, um mit ihren Antworten zu einem antizipierten Ziel beizutragen, wobei es sich bei diesem Ziel in diesem Beispiel um die Förderung der wissenschaftlichen Forschung handelt. Je positiver demnach die Einstellung gegenüber dem wahrheitsgemäßen Beantworten von Umfragen ausfällt (sprich: je tiefer die Überzeugung, dass das entsprechende Verhalten grundsätzlich zur Förderung wissenschaftlicher Forschung beiträgt), umso stärker die subjektive Wahrnehmung, dass das wahrheitsgemäße Beantworten auch durch andere als positiv und wünschenswert bewertet wird und umso eher Befragte davon überzeugt sind, dass sie mit ihren wahren Antworten individuell auch tatsächlich zu einer Förderung der wissenschaftlichen Forschung beitragen und sie über alle benötigten Fähigkeiten zum wahren Antworten verfügen,

umso wahrscheinlicher wird d) die Intention gebildet, das Verhalten auch tatsächlich auszuführen. Da die Intention zum wahrheitsgemäßen Beantworten demnach ebenfalls bei stark positiver Einstellung, einer kohärent subjektiven Norm, jedoch mit geringer wahrgenommener Verhaltenskontrolle gebildet werden kann, ist die Intention alleine keine hinreichende Bedingung für das wahrheitsgemäße und elaborierte Antworten. Vielmehr hat die wahrgenommene Verhaltenskontrolle an dieser Stelle einen erneuten Einfluss. Bei gegebener Intention wird das entsprechende Verhalten lediglich dann ausgeführt, wenn tatsächlich die Überzeugung vorhanden ist, dass alle benötigten Fähigkeiten vorhanden sind, um durch das wahrheitsgemäße Beantworten von Fragen zur Förderung wissenschaftlicher Forschung beizutragen.

Zu der Bildung der Intention für die Handlung (2) „verzerrtes Beantworten von Fragen“ in elaborierten Prozessen kommt es demnach, wenn a) keine positiven Einstellungen bzw. Überzeugungen vorhanden sind, dass das wahrheitsgemäße Beantworten von Umfragen zu einer Förderung wissenschaftlicher Forschung (oder zu anderen wünschenswerten Zielen) beiträgt oder wenn gar eine positive Einstellung gegenüber dem verzerrten Antworten besteht, weil dies z. B. zu dem subjektiv wünschenswerten Ziel führt, eine Umfrage mit so wenig aufwand wie möglich abzuschließen (etwa, weil unter allen Teilnehmenden ein Gewinn verlost wird) oder sich selbst durch die gegebenen Antworten in einem besonders positiven Licht darzustellen. Weiterhin führt b) das Fehlen einer subjektiven Norm, das wahrheitsgemäße Antworten von nahestehenden Menschen als positiv bewertet werden bzw. das Vorhandensein der Norm, dass nahestehende Menschen verzerrtes Antworten als positiv bewerten zur Bildung der Intention, verzerrte Antworten zu geben. Auch die Überzeugung, dass c) die eigenen Fähigkeiten nicht ausreichen, um mit wahren Antworten zu einer Förderung wissenschaftlicher Forschung beizutragen oder gar das Vorhandensein der Überzeugung, dass alle benötigten Fähigkeiten vorhanden sind, um durch verzerrte Antworten Umfragen möglichst ohne Anstrengung zu beenden (und damit z.B. einen Preis zu gewinnen) trägt dazu bei, verzerrte Antworten zu intendieren. Es zeigt sich somit, dass elaborierte und verzerrte Antworten zum einen als Konsequenz einer Handlungsstrategie auftreten können, welche zwar wahres Antworten intendiert, für die entsprechende Umsetzung jedoch nicht alle Bedingungen erfüllt sind. Zum anderen können elaborierte und verzerrte Antworten auch die Konsequenz einer Handlungsstrategie sein, welche bewusste Verzerrungen intendiert.

Aggregation generischer Determinanten des Befragtenverhaltens

Mit Einführung des Satisficing-Optimizing Modells, dem Modell spontanen Prozessierens und der TPB ist somit im Kontext der Antwortgenerierung in Umfragen die Möglichkeit gegeben,

über die klassischen, sequenziellen Modelle der Informationsprozessierung hinaus das Antwortverhalten vertiefend auf Basis unterschiedlicher, nach dem Elaborationsgrad differenzierter Informationsverarbeitungsprozesse zu beleuchten. Die genannten Modelle sind dabei insofern als sich ergänzend zu betrachten, als dass das Satisficing-Optimizing Modell ein Erklärungsversuch dafür leistet, unter welchen Bedingungen der Prozess der Antwortgenerierung gering oder nicht elaboriert stattfindet (Satisficing) und unter welchen Bedingungen besagter Prozess elaboriert stattfindet (Optimizing). Das Modell spontanen Prozessierens wiederum beleuchtet, welche kognitiven Prozesse der Antwortgenerierung bei geringer Elaboration wirken, die TPB wiederum beleuchtet, welche kognitiven Prozesse der Antwortgenerierung bei hoher Elaboration wirken. Demnach sind es zunächst 1.) die Schwierigkeit der Aufgabe; 2.) die kognitiven Fähigkeiten Befragter; und 3.) die Motivation der Befragten welche den Grad an Elaboration bedingen: Je weniger komplex die zu lösende Aufgabe, je höher die kognitiven Fähigkeiten und je höher die Motivation zur Elaboration, desto wahrscheinlicher beantworten Befragte Fragen elaboriert (d.h. sie betreiben Optimizing). Werden Antworten in der Folge auf der Basis gering oder nicht elaborierter kognitiver Prozesse gebildet, so ist es 4.) die Einstellung gegenüber dem Objekt der Befragung, welche die weitere Definition und Interpretation der Frage bedingt. Diese Einstellung wirkt umso stärker, 5.) je chronisch Zugänglicher diese ist. Das Vorhandensein sowie die chronische Zugänglichkeit einer Einstellung ist während des spontanen Prozessierens jedoch nicht hinreichend für die Formulierung einer einstellungskohärenten Antwort. Vielmehr ist es zuletzt 6.) die Definition der Situation, welche auf Hinweisreizen oder weiteren zugrundeliegenden Einstellungen und Normen basieren kann, welche schließlich in Abgleich mit den zugrundeliegenden Einstellungen gegenüber dem Objekt der Befragung zu kongruenten Antworten führt. Werden Antworten hingegen auf der Basis elaborierter kognitiver Prozesse gebildet, so ist es 7.) die Einstellung gegenüber eines spezifischen Antwortverhaltens respektive ihrer Eignung zum Erreichen einer subjektiv wünschenswerte Konsequenz des Verhaltens, welche sich auf der Basis zugrundeliegender Beliefs bildet; 8.) die subjektive Wahrnehmung sozialer Normen bezüglich des spezifischen Antwortverhaltens; sowie 9.) die wahrgenommene Verhaltenskontrolle in dem Sinne, dass Befragte davon überzeugt sind, über alle benötigten Fähigkeiten zu verfügen, um mit dem spezifischen Antwortverhalten zu einem subjektiv wünschenswerten Ergebnis zu gelangen; welche 10.) zur Bildung einer mehr oder weniger kohärenten Handlungsintention führt. Sind Intention und wahrgenommene Verhaltenskontrolle im Sinne des spezifischen Verhaltens und der subjektiv wünschenswerten Konsequenz kohärent, so folgt die Handlung „Generierung einer Antwort in Umfragen“ gemäß dem zugrundeliegenden Handlungsziel.

Innerhalb der beschriebenen Modelle finden sich somit die folgenden Annahmen bezüglich der Antwortgenerierung: 1.) bedingt der Elaborationsgrad des kognitiven Informationsverarbeitungsprozesses den kognitiven Pfad, welcher zur Generierung einer Antwort in Umfragen führt. 2.) sind es je Prozesspfad verschiedene Einstellungen, Wahrnehmungen und Normen auf der Basis zugrundeliegender Beliefs, welche mehr oder weniger handlungsweisend sind. 3.) bedingt die Interpretation dessen, welches Verhalten in der jeweiligen Situation angebracht bzw. erwünscht ist ebenfalls das Verhalten. Innerhalb des Modells des spontanen Prozessierens wirkt dies auf die Definition und Interpretation der Situation, innerhalb der TPB wirkt dies in Form normativer Beliefs, welche die Wahrnehmung subjektiver Normen bedingen. Und 4.) wird die heuristische Funktion situativer Hinweisreize im Informationsverarbeitungsprozess für Fälle betont, in welchen individuelle Dispositionen nicht als Handlungsleitend zur Verfügung stehen oder aus sonstigen Gründen nicht aktiviert oder verwendet werden.

Kognitive Faktoren mit Einfluss auf das Antwortverhalten sind demnach a) die Motivation zum elaborierten Antworten; b) die kognitiven Fähigkeiten; c) die Einstellung gegenüber dem Objekt der Befragung; d) die chronische Zugänglichkeit der Einstellung gegenüber dem Objekt der Befragung; e) die Einstellung gegenüber einem spezifischen Antwortverhalten respektive der antizipierten Verhaltenskonsequenz; f) die wahrgenommene Verhaltenskontrolle; und g) die Wahrnehmung dessen, welches Verhalten als sozial wünschenswert und angemessen erscheint, wobei sich besagte Wahrnehmung auf der Basis zugrundeliegender Dispositionen wie etwa normativen Beliefs bildet.

Hieraus ergibt sich aus kognitiver Perspektive, dass verzerrte Antworten je nach zugrundeliegendem Informationsverarbeitungsprozess, differenziert anhand des Elaborationsgrades, in folgenden Fällen auftreten: Beantworten Befragte Fragen gering- oder nicht-elaboriert, so sind Antwortverzerrungen zu erwarten, wenn Antworten nicht auf der Basis relevanter Einstellungen, sondern auf der Basis von einfachen Antwortheuristiken und situativen Hinweisreizen formuliert werden. Dies ist dann der Fall, wenn keine potenziell relevanten Einstellungen hinsichtlich dem Objekt der Befragung chronisch zugänglich sind oder wenn die Wahrnehmung eines in der Situation angemessenen und akzeptablen Verhaltens zu einer Interpretation der Situation führt, welche nicht kohärent ist mit der Einstellung gegenüber dem Objekt der Befragung. Beantworten Befragte Fragen hingegen elaboriert, so ist die Einstellung gegenüber dem selektierten Antwortverhalten, welche auf der subjektiv als wünschenswert wahrgenommenen Handlungskonsequenz basiert, von Relevanz. Intendieren Befragte das geben wahrer Antworten, so kommt es innerhalb elaborierter Informationsverarbeitungsprozesse gemäß der Modelle trotz-

dem dann zu Antwortverzerrungen, wenn subjektiv wahrgenommene Normen situativ das Geben einer wahren Antwort verhindern und/oder wenn Befragte nicht davon überzeugt sind in der Lage zu sein, wahre Antworten auf eine Art und Weise zu geben, um damit eine subjektiv wünschenswerte Handlungskonsequenz zu erreichen. Zudem können Befragte auch generell die Gabe verzerrter Antworten intendieren, so dies der Erwartung gemäß ebenfalls zu einer subjektiv wünschenswerten Handlungskonsequenz führt.

2.1.3 Implizite, integrative Prozessmodelle der Einstellungs-Verhaltensforschung

Durch die Integration der Annahmen des Satisficing-Optimizing Modells, des Modells spontanen Prozessierens und der TPB ist es wie gezeigt möglich, die Pfade der verschiedenen Informationsverarbeitungsprozesse nachzuzeichnen und für unterschiedliche Antwortstrategien zu adaptieren. Diesbezüglich finden sich in der Einstellungs-Verhaltensforschung ebenfalls Modelle, welche diese Integration verschiedener Prozesspfade bereits leisten und somit zu sparsameren Modellen mit geringerer Komplexität beitragen.

Solche integrativen Modelle, welche die beiden idealtypischen Modi der Informationsverarbeitung verbinden, sind in der Sozialpsychologie sowie der Kognitionsforschung äußerst beliebt und werden als duale Prozessmodelle bezeichnet (vgl. z.B. Chaiken & Trope, 1999; Kahneman, 2011; Mayerl, 2009; Smith & DeCoster, 2000). Diese haben gemein, dass als archetypische Modi der Informationsverarbeitung, wenn auch unter verschiedenen Bezeichnungen[1], das automatisch-spontane und das überlegt-kontrollierte Prozessieren unterschieden wird. Dabei zeichnen sich automatisch-spontane Prozesse durch einen geringen kognitiven Aufwand bzw. einen geringen Elaborationsgrad aus und sind nicht intentional gesteuert, während überlegt-kontrollierte Prozesse einen erhöhten kognitiven Aufwand bzw. Elaborationsgrad bedürfen, auf dem bewussten Durchdenken von Informationen basieren und allgemein weniger durch Heuristiken beeinflusst werden als automatisch-spontane Prozesse (Bargh, 1989). Hinzu kommen sogenannte Framing-Modelle (vgl. z.B. Esser, 2001), welche über den Einfluss von Einstellungen hinaus die subjektiv-situationelle Wirkung von Normen und weiteren Dispositionen auf das Verhalten im Sinne sogenannter Frames berücksichtigen. Es folgt entsprechend eine Übersicht

[1] Die beiden Modi der Informationsverarbeitung werden in verschiedenen Modellen unterschiedlich benannt und auch verschieden definiert: So benennt z.B. das *heuristic-systematic model* nach Chaiken (1980) statt des automatisch-spontanen und des überlegt-kontrollierten den heuristischen und den systematischen Modus; Fazio (1986) differenziert die Prozesse nach *attitude based* und *attribute based*, bzw. in seinem MODE-Modell (Fazio, 1990) nach *theory driven* und *data driven*; Esser (1996) unterscheidet automatisch und rational; Strack und Deutsch (2004) impulsiv und reflektiv etc.

über gängige duale Prozess- und Framing-Modelle der Einstellungs-Verhaltensforschung sowie deren Adaption für den spezifischen Prozess der Antwortgenerierung.

Das MODE-Modell

Zunächst soll das MODE-Modell (***M**otivation and **O**pportunity as **De**terminants*) beleuchtet werden, ein Modell zur Betrachtung der Einstellungs-Verhaltensrelation nach Fazio (1990), welches durch seine besondere Betonung moderierender Faktoren innerhalb der Einstellungs-Verhaltensrelation zu Geltung gelangt ist. Diesem ist zunächst inhärent, dass der Prozess der Informationsverarbeitung, welcher schließlich zu einem spezifischen Verhalten führt, einem von zwei qualitativ differenten kognitiven Pfaden unterliegt: Einem automatisch-spontanen Pfad, der als Theoriegetrieben und in Abhängigkeit automatisch aktivierter Einstellungen bezüglich dem Einstellungsobjekt beschrieben wird und einem geplant-kontrollierten Pfad, der als Datengetrieben und hoch elaboriert bezüglich der Determinanten der Verhaltenseinstellung sowie der Verhaltenskonsequenzen beschrieben wird. Während der automatisch-spontane Prozesspfad im Wesentlichen dem im vorangegangenen Kapitel beschriebenen Modell spontanen Prozessierens entspricht (Fazio, 1986) und gering elaboriert verläuft, basiert der geplant-kontrollierte Prozesspfad auf Postulaten der TRA (Ajzen & Fishbein, 1980) sowie der TPB (Ajzen, 1991; Fishbein & Ajzen, 2010) und verläuft somit hoch elaboriert. Integriert werden die beiden Modelle durch eine zusätzliche Erklärungskomponente, unter welchen Bedingungen welcher der Prozesspfade selektiert bzw. aktiviert wird, der Motivation sowie der Möglichkeit (Fazio, 1990).

Da der automatisch-spontanen Pfad demnach gering elaboriert verläuft, während der geplant-kontrollierte Pfad hoch elaboriert verläuft, müssen gemäß des Modells gewisse Faktoren erfüllt sein, damit Individuen den gesteigerten kognitiven Aufwand zum elaborieren in Kauf nehmen, statt Entscheidungen lediglich gering elaboriert und spontan zu treffen. Als solcher Faktor stellt sich zum einen die Motivation von Individuen zum elaborieren dar, wobei Individuen laut dem Modell dann motiviert sind, wenn die Verhaltenskonsequenz von besonderer Bedeutung ist, oder in anderen Worten, wenn sogenannte „Angst vor Invalidität“ vorliegt. Das Konzept der „Angst vor Invalidität“ basiert auf der Lay Epistemic Theory nach Kruglanski und Freund (1983) und beschreibt die Angst von Individuen, mit ihrem Verhalten eine ungünstige oder unerwünschte Verhaltenskonsequenz zu erzeugen, was in der Folge dazu führt, dass Individuen, desto höher deren „Angst vor Fehlern“ ist, umso motivierter sind, ihr Verhalten im Vorfeld zu elaborieren, um unerwünschte Verhaltenskonsequenzen zu vermeiden. Ohne besagte Angst, so Fazio (1990) weiter, besteht für Individuen nur ein geringer Grund, den gesteigerten kognitiven

Aufwand der Elaboration in Kauf zu nehmen, so dass diese stattdessen automatisch-spontan handeln. Die Motivation zum elaborierten Prozessieren ist laut MODE-Modell jedoch nicht hinreichend, um elaboriertes Prozessieren zu determinieren, da durchaus Situationen denkbar sind, in welchen Individuen zwar motiviert zum elaborieren sind, hierzu jedoch nicht die Möglichkeit besteht, z. B. weil eine schnelle und spontane Reaktion gefordert ist. Entsprechend ist es ebenfalls die Möglichkeit zum elaborierten Prozessieren, welche einen Einfluss auf die Aktivierung einer der beiden Prozesspfade nimmt. Individuen aktivieren den elaborierten Prozesspfad demnach lediglich, wenn sowohl die Motivation, als auch die Möglichkeit zum elaborieren besteht. In allen anderen Fällen werden Individuen den automatisch-spontanen Prozesspfad selektieren bzw. aktivieren.

Zusammenfassend basiert Verhalten, welches einem geplant-kontrollierten Elaborationsprozess entspricht, auf der spezifischen Einstellung gegenüber dem geplanten Verhalten (NICHT der Einstellung gegenüber dem Objekt der Befragung, auf welches sich das geplante Verhalten bezieht) sowie vorhandenen, subjektiven Normen und der wahrgenommenen Verhaltenskontrolle. Führen Einstellungen, konsistente soziale Normen und wahrgenommene Verhaltenskontrolle entsprechend zur Bildung einer kohärenten Verhaltensintention, so zeigt sich geplantes Verhalten gemäß der zugrundeliegenden Verhaltenseinstellung. Ist dies hingegen nicht der Fall, so prozessieren Individuen Einstellungen gegenüber alternativen Verhaltensweisen auf der Basis verfügbarer Informationen („Daten“) und gewichten diese entsprechend zugrundeliegender Normen und der jeweiligen wahrgenommenen Verhaltenskontrolle. Zwischen allen Handlungsalternativen wird schließlich diejenige selektiert, welche, gegeben der Gewichtung, am wahrscheinlichsten zu einem subjektiv wünschenswerten Ergebnis führt.

Verhalten, welches auf einem automatisch-spontanen Elaborationsprozess basiert, ist entsprechend abhängig von der zugrundeliegenden Einstellung gegenüber dem Objekt der Befragung (NICHT der Einstellung gegenüber einem spezifischen Verhalten) sowie der chronischen Zugänglichkeit einer solchen Einstellung. Ist eine solche Einstellung vorhanden und wird diese in automatisch-spontanen Informationsverarbeitungsprozessen aufgrund ihrer chronischen Zugänglichkeit automatisch aktiviert, so erfolgt das Verhalten unter Berücksichtigung weiterer, kohärenter situativer Normen entsprechend Einstellungsbasiert („Theoriebasiert“). Sind Einstellungen hingegen nicht vorhanden oder chronisch nicht zugänglich, so basiert automatisch-spontanes Verhalten auf einfachen Entscheidungsheuristiken und situativen Hinweisreizen, welche wiederum durch die Interpretation der Situation, basierend auf situationell verfügbaren Normen und Informationen, beeinflusst werden kann.

Darüber hinaus postuliert das MODE-Modell, dass automatisch-spontane Prozesse ebenfalls in geplant-kontrollierten Verhaltenssituationen eine Rolle spielen und dass geplant-kontrollierte Prozesse Einfluss auf automatisch-spontane Prozesse nehmen können. Entsprechend nennt Fazio (1990) als automatischen Prozess, welche innerhalb geplant-kontrollierter Prozesse stattfinden eine mögliche automatische Aktivierung der Einstellung bezüglich dem Verhalten statt der erneuten Evaluation der Einstellung auf Basis zugrundeliegender Beliefs, falls eine solche Verhaltenseinstellung bereits gebildet wurde, diese chronisch zugänglich ist und die Situation, in welcher das Verhalten erforderlich ist einer vergangenen Situation ähnelt, in welcher das spezifische Verhalten bereits zu einem subjektiv wünschenswerten Ergebnis geführt hat. Zusätzlich zur automatischen Aktivierung von Einstellungen ist es ebenfalls möglich, dass innerhalb der Neuprozessierung und Evaluation von Verhaltenseinstellungen automatisch-spontane Prozesse involviert sind, wenn diese auf der Basis automatisch aktivierter Beliefs gebildet werden und/oder wenn eine Evaluation der zu erwartenden Verhaltenskonsequenz automatisch aktiviert wird und somit Einfluss auf die Formierung einer Verhaltenseinstellung nimmt, wobei eine solche automatische Aktivierung Abhängig ist von der chronischen Zugänglichkeit besagter Beliefs und Evaluationen. Zusätzlich kann auch eine chronisch zugängliche Einstellung gegenüber dem Objekt der Befragung (NICHT der Verhaltenseinstellung) automatisch aktiviert werden und dadurch die Evaluation der Verhaltenseinstellung beeinflussen, in dem diese die Wahrnehmung einer subjektiv wünschenswerten Verhaltenskonsequenz beeinflusst. Demgegenüber stehen kontrollierte Prozesse, welche ebenfalls in automatisch-spontanen Entscheidungssituationen Einfluss nehmen. Dies ist darauf zurück zu führen, dass automatisch-spontane Prozesse auf einer spezifischen Objekteinstellung sowie deren chronischer Zugänglichkeit basieren. Solche Einstellungen sind entsprechend dann chronisch zugänglich, wenn diese z. B. in der Vergangenheit häufig bewusst und elaboriert prozessiert wurden. Entsprechend führt die geplant-kontrollierte Evaluation von Einstellungen gegenüber dem Objekt der Befragung zu einer erhöhten Zugänglichkeit, so dass eine solche Einstellung in automatisch-spontanen Entscheidungssituationen eher aktiviert wird und entsprechend Einfluss auf das Verhalten nimmt.

Soll nun das konkret gezeigte Antwortverhalten von Individuen gemäß diesem Modell erklärt werden, so gilt es zunächst zu ermitteln, welche zugrundeliegende Motivation vorhanden ist, da sich hieraus die kognitiven Pfade für das Antwortverhalten sowie die subjektiv wünschenswerte Handlungskonsequenz ergibt. Zur Illustration sei beispielhaft die Motivation gewählt, Fragen so exakt wie möglich zu beantworten, was zu der (ebenfalls beispielhaft gewählten) subjektiv wünschenswerten Verhaltenskonsequenz führt, durch die eigene Teilnahme an Befragungen zu einer Förderung der Wissenschaften beizutragen. Dies führt zunächst dazu, dass sich

als „Angst vor Invalidität“ die Angst bilden kann, falsche Angaben in der Befragung zu machen, wodurch sich die Motivation ergibt, Fragen besonders elaboriert zu beantworten. Die Möglichkeit zum Elaborieren sei beispielsweise dann gegeben, wenn kein subjektiver Zeitdruck herrscht oder wenn keine Anweisung erfolgt, alle Fragen so schnell wie möglich zu beantworten. Um das Beispiel nicht überkomplex zu gestalten, gilt die Möglichkeit zum Elaborieren für diesen Fall als gegeben, wodurch bei vorhandener Angst vor Fehlern der geplant-kontrollierte Prozesspfad aktiviert wird. Fehlt die Angst vor Invalidität hingegen, z. b. weil Befragte ein großes Interesse am Thema der Befragung haben und bereits im Vorfeld davon überzeugt sind, korrekte Antworten auch ohne Elaboration geben zu können, so fehlt entsprechend die Motivation zum Elaborieren, was zur Aktivierung eines automatisch-spontanen Antwortprozesses führt, welcher dennoch nicht in einem Zielkonflikt mit der subjektiv wünschenswerten Verhaltenskonsequenz „durch wahre Antworten zur wissenschaftlichen Forschung beitragen“ stehen muss wenn Befragte davon überzeugt sind, auch mit spontanen Antworten zu einer Förderung der Wissenschaften beitragen zu können.

Prozessieren Befragte Antworten entsprechend automatisch-spontan, so ist die gegebene Antwort dann kohärent mit der individuellen Einstellung bezüglich dem Objekt der Befragung (dem Thema der Befragung bzw. dem Inhalt der konkreten Frage), so diese Einstellung chronisch zugänglich ist, wobei zugrundeliegende Beliefs sowie die Interpretation der Situation Einfluss nehmen auf die Selektion und Aktivierung einer im Kontext der Fragestellung als relevant wahrgenommenen Einstellung, was in der Konsequenz zu „wahren“ Antworten führt. Erwähnt sei diesbezüglich erneut, dass die chronische Zugänglichkeit von Einstellungen ebenfalls durch im Vorfeld stattfindende geplant-kontrollierte Evaluationen der entsprechenden Einstellung bedingt sein kann. Sind solche Einstellungen hingegen nicht chronisch zugänglich, so basiert eine automatisch-spontan prozessierte Antwort auf Antwortheuristiken oder situativen Hinweisreizen, welche als Stellvertreter für chronisch zugängliche, für den Inhalt der Fragestellung relevante Einstellungen dienen. Ebenfalls sei erwähnt, dass Einstellungen, welche im Kontext vorangegangener Fragen aktiviert wurden, zum Zeitpunkt der Antwortprozessierung nachfolgender Fragen ebenfalls noch situativ zugänglich sind und bei einem Fehlen relevanter, chronisch zugänglicher Einstellungen stattdessen als Einstellungsgrundlage für das Antwortverhalten dienen können. Basieren automatisch-spontane Antworten entsprechend auf Antwortheuristiken, situativen Hinweisreizen oder alternativen zugänglichen, aber weniger relevanten Einstellungen, so ist mit einer Antwortverzerrung zu rechnen.

Prozessieren Befragte Antworten entsprechend des Modells hingegen geplant-kontrolliert, so zeigt sich das konkrete Antwortverhalten in Abhängigkeit von der Einstellung gegenüber dem

spezifischen Antwortverhalten (z. B. „Führt mein Geben wahrer Antworten tatsächlich zu einer Förderung der Wissenschaften?“), sozialen Normen (z. B. „wie erwünscht ist es, wahre Antworten zu geben?“) und der Handlungskontrolle (z. B. „Verfüge Ich über alle notwendigen Fähigkeiten, um wahre Antworten geben zu können, so dass diese zur Förderung der Wissenschaften beitragen?“). Sind alle genannten Faktoren kohärent positiv, so resultiert dies in dem Verhalten „Geben wahrer Antworten“. Sind Befragte hingegen nicht überzeugt, dass ihre wahren Antworten zu einer Förderung der Wissenschaften beitragen, dass das geben einer wahren Antwort im entsprechenden Kontext erwünscht ist oder dass sie tatsächlich über alle benötigten Fähigkeiten verfügen, so werden, gegeben subjektiv wünschenswerter Verhaltenskonsequenzen, alternative Verhaltenseinstellungen evaluiert und entsprechend zugrundeliegender Normen und der jeweiligen wahrgenommenen Verhaltenskontrolle gewichtet. So könnte die Gabe einer wahren Antwort beispielsweise daran scheitern, dass die entsprechende Antwort subjektiv wahrgenommenen sozialen Normen widerspricht. Bei der Evaluation möglicher Handlungsalternativen könnten Befragte dann zu dem Schluss kommen, dass eine minimale Verzerrung der wahren Antwort hin zu einem sozial wünschenswerteren Ergebnis keine sonderlich großen Auswirkungen auf den eigenen Beitrag zur wissenschaftlichen Forschung hat, so dass als alternative Antwortstrategie „Verzerrung der Antwort“ selektiert wird, so dies kohärent mit der Handlungskontrolle ist. Die Gabe einer verzerrten Antwort ist demnach die Konsequenz. An dieser Stelle sei ebenfalls vermerkt, dass auch innerhalb dieses geplant-kontrollierten Antwortprozesses automatisch-spontane Faktoren Einfluss nehmen. So kann eine automatisch aktivierte Einstellung gegenüber dem Inhalt der Frage die Verhaltenseinstellung in dem Sinne beeinflussen, dass sich beispielsweise ein besonders hohes Interesse am Thema der Befragung in einem entsprechenden, wahrheitsorientierten Antwortverhalten niederschlägt. Auch weitere, automatisch aktivierte Beliefs können einen entsprechenden Einfluss auf die subjektiv wünschenswerte Verhaltenskonsequenz, die Verhaltenseinstellung, die wahrgenommene soziale Norm sowie die Handlungskontrolle nehmen.

Kognitive Faktoren mit Einfluss auf das Antwortverhalten sind demnach: a) die zugrundeliegende Motivation hinsichtlich einer spezifischen Antwortstrategie, welche sich aus der Angst vor Invalidität ergibt und im Kontext der subjektiv wünschenswerten Verhaltenskonsequenz evaluiert wird; b) kognitive Faktoren, welche die Möglichkeit bedingen, die spezifische Antwortstrategie anzuwenden; c) die Einstellung gegenüber dem Objekt der Befragung; d) die chronische Zugänglichkeit der Einstellung gegenüber dem Objekt der Befragung; e) die Einstellung gegenüber dem selektierten Antwortverhalten; f) subjektiv wahrgenommene Normen

bezüglich des selektierten Antwortverhaltens; g) die wahrgenommene Verhaltenskontrolle hinsichtlich des selektierten Antwortverhaltens; und h) zugrundeliegende Beliefs, welche wiederum die subjektiv wünschenswerte Verhaltenskonsequenz, die zugrundeliegende Motivation, Einstellungen und deren chronische Zugänglichkeit, wahrgenommene soziale Normen und die wahrgenommene Verhaltenskontrolle bedingen.

Verzerrte, automatisch-spontan generierte Antworten sind entsprechend zu erwarten, wenn eine Antwort nicht auf der Basis einer relevanten Einstellung bezüglich dem Objekt der Befragung generiert wird, sondern auf der Basis von Antwortheuristiken, situativen Hinweisreizen oder alternativen, für den Kontext der spezifischen Frage weniger relevanten Einstellungen. Dies ist vor allem dann der Fall, wenn keine relevanten Einstellungen hinsichtlich dem Objekt der Befragung chronisch zugänglich sind oder wenn zugrundeliegende Beliefs und weitere Dispositionen zu einer Interpretation der Situation und damit zu einer automatischen Aktivierung von alternativen, für den Kontext der Befragung weniger relevanten Einstellungen führt. Verzerrte, geplant-kontrolliert generierte Antworten sind zu erwarten, wenn das selektierte Antwortverhalten respektive der subjektiv wünschenswerten Verhaltenskonsequenz keine wahren Antworten intendiert bzw. wenn situativ selektierte, alternative Antwortstrategien eine bewusste Verzerrung der formierten Antworten vorsehen. Alternative Antwortstrategien, welche mögliche Verzerrungen intendieren, werden evaluiert und selektiert, so die Verhaltenseinstellung, soziale Normen und Handlungskontrolle nicht kohärent sind.

Das Reflective-Impulsive Model (RIM)

Während das MODE-Modell somit zwei qualitative Modi der Informationsprozessierung postuliert, wobei entweder der eine oder der andere Modus aktiviert ist (auch wenn Einflüsse der jeweils anderen Prozesspfade wirken können) und durch seine besondere Betonung moderierender Faktoren in der Vorhersage individuellen Verhaltens Beachtung findet, verfolgt das sogenannte **R**eflective-**I**mpulsive-**M**odel (RIM) nach Strack und Deutsch (2004) den Ansatz, die beiden Modi geringer und hoher Elaboration nicht als getrennt, sondern als integriert, interagierend, interdependent und unter Umständen parallel verlaufend darzustellen, wodurch auch dieses Modell zu Popularität gelangt ist, insbesondere, da neurologische Studien Evidenzen für solche parallel statt sequenziell verlaufenden kognitive Prozesse liefern (Krishna & Strack, 2017). Diesbezüglich wird das RIM im Folgenden beschrieben und auf das spezifische Verhalten „Beantwortung von Fragen in Umfragen“ angewandt.

Grundsätzlich unterscheidet das RIM, ebenso wie das zuvor erörterte MODE-Modell, zwei Systeme der Informationsprozessierung: Ein reflexives System, welches gemäß der bisher verwendeten Terminologie als hoch elaboriert bezeichnet werden kann und welches mit der Bewusstseinswahrnehmung einer aktiven Prozessierung einhergeht, sowie ein impulsives System, welches entsprechend als nicht-elaboriert zu verstehen ist und häufig unbewusst prozessiert wird, auch wenn eine mehr oder weniger stark ausgeprägte Wahrnehmung eines spezifischen Gefühls vorhanden sein kann, welches mit dem zu prozessierenden Element assoziiert ist.

Das reflexive System wird hierbei als rationales, regelbasiertes System bezeichnet, welches Informationen in der Tendenz intraindividuell langsam prozessiert und hierfür auf Ressourcen des Arbeitsgedächtnisses zurückgreift. Innerhalb des reflexiven Systems werden Informationen auf Basis von Syllogismen prozessiert, in dem bezüglich dem Objekt der Befragung Konzepte und Relationen zwischen den Konzepten aktiviert werden, welchen nachfolgend ein Wahrheitswert zugeordnet wird (Strack & Deutsch, 2004). Zur Illustration nennen Krishna und Strack (2017) diesbezüglich das Beispiel, dass bei Exposition des Objektes „Salat“ das Konzept „Gesund“ sowie die Relation „ist“ aktiviert werden kann, für welches der Wahrheitswert „Wahr“ prozessiert wird, so dass die Objektevaluation des Salates in der wahren Aussage „Salat ist gesund“ resultiert. Relationen sind dabei nicht auf das hier exemplarisch gewählte „ist“/“ist nicht“ beschränkt, sondern umfassen generelle Relationskonzepte wie etwa „impliziert“/“impliziert nicht“, „führt zu“/“führt nicht zu“ oder „ist eine Menge von“/“ist keine Menge von“. Ebenfalls sind Syllogismen nicht darauf reduziert, den Wahrheitswert je eines Subjektes und eines Prädikates zu prozessieren. Stattdessen können diese, insbesondere bei der Prozessierung neuer Informationen bzw. neuen Wissens zunehmend komplex konstruiert werden. Diesbezüglich sei erneut das Beispiel von Krishna und Strack (ebd.) aufgegriffen, innerhalb welchem die Aussage „Salat ist gesund“ als erste Prämisse des Syllogismus um die zweite Prämisse „gesund ist gut“ ergänzt wird, woraus die (im Sinne der Logik) wahre Konklusion „Salat ist gut“ folgt. Dadurch, dass innerhalb des reflexiven Systems Aussagen Wahrheitswerte zugeordnet werden ist dieses in der Lage, Informationen regelbasiert zu prozessieren, was die Konsistenz der Repräsentationen maximiert (Gawronski & Strack, 2004). Eine Sonderstellung innerhalb des RIM (Strack & Deutsch, 2004) nimmt die Operation von Negationen ein, welche gemäß dem Modell lediglich innerhalb des reflexiven Systems prozessiert werden können, so diesem hinreichende Ressourcen des Arbeitsgedächtnisses zur Verfügung stehen, welche hierfür gemäß dem Modell benötigt werden. Daneben ist das reflexive System gemäß RIM jedoch in seinen Möglichkeiten zur Informationsprozessierungen durch Randbedingungen limitiert, unter welchen am relevan-

testen die Arbeitsgedächtniskapazität sowie „Arousal“ sind. Bezüglich der Arbeitsgedächtniskapazität wird angenommen, dass die Aktivierung von Konzepten und Relationen sowie deren Prozessierung Kapazitäten des Arbeitsgedächtnisses beansprucht, wodurch eine individuelle Limitation bezüglich der Komplexität reflexiver Operationen in dem Sinne gegeben ist, als dass die für die Operation benötigten Konzepte, Relationen und Evaluationen nicht die Kapazität des Arbeitsgedächtnisses überschreiten können. „Arousal“ auf der anderen Seite wirkt, gemäß dem Yerkes-Dodson Gesetz (Yerkes & Dodson, 1908) nichtlinear auf die Leistung des reflexiven Systems in dem Sinne, dass ein mittleres Maß „Arousal“ am förderlichsten für Operationen des reflexiven Systems sind, ein hohes Maß an „Arousal“ die Komplexität sozialer Urteile reduziert und ein geringes Maß an „Arousal“ mit einer geringeren Fertigkeit zum reflexiven Prozessieren assoziiert ist.

Zusätzlich zum reflexiven System postuliert das RIM (Strack & Deutsch, 2004) ein sogenanntes impulsives System, welches als assoziativ, kontinuierlich und schnell arbeitend sowie kaum kognitiv fordernd beschrieben wird. Das impulsive System greift dabei auf das Langzeitgedächtnis zurück und hat somit theoretisch, im Gegensatz zum reflexiven System, welches auf das Arbeitsgedächtnis angewiesen ist, unlimitierte Kapazitäten. Die Funktionsweise des impulsiven Systems entspricht dabei der Prozessierung eines assoziativen Netzwerkes in dem Sinne, dass zwischen Elementen rigide assoziative Verknüpfungen mit variierender stärke existieren, wobei es bei Aktivierung eines Elementes zur Aktivierung weiterer Elemente in Abhängigkeit der Stärke ihrer Verknüpfungen mit dem ursprünglich aktivierten Element innerhalb des Netzwerkes kommt. Elemente repräsentieren diesbezüglich nicht ausschließlich Objekte, Wissensstrukturen, Einstellungen, Überzeugungen etc., sondern können ebenfalls als Gefühle bzw. Emotionen auftreten, welche wiederum mit weiteren Elementen verknüpft sind, wodurch die automatische Aktivierung einer Emotion bei Exposition eines gewissen Elementes erklärt wird. Die zeitliche Nähe sowie die Häufigkeit der Aktivierung von Elementen erhöht weiterhin die kognitive Zugänglichkeit des Elementes sowie die Wahrscheinlichkeit, dass das Element innerhalb impulsiver Operationen auch künftig aktiviert wird. Die Verknüpfungen zwischen den Elementen basieren diesbezüglich auf den Prinzipien der Kontiguität und der Ähnlichkeit, so dass diese entstehen (oder wenn bereits vorhanden, sich weiter verstärken), wenn mit verschiedenen Elementen assoziierte Stimuli zeitlich oder räumlich dicht aufeinandertreffen, die relative Rigidität von Assoziationen berücksichtigend. Diesbezüglich ist es ebenfalls möglich, dass Assoziationen zwischen Elementen durch reflexive Operationen gebildet werden, z. B. wenn diese häufig im gleichen Kontext prozessiert werden, so dass auch Verbindungen zwischen Elementen entstehen können, welche in der empirischen Realität weder zeitlich, noch räumlich

aufeinandertreffen. Bilden sich durch solche reflexiven Prozesse assoziative Verknüpfungen innerhalb des impulsiven Systems, so können diese leichter und schneller prozessiert werden, auch wenn sich hierdurch die Bedeutung von Elementen verändern kann, wenn reflexives Prozessieren nicht möglich ist. Reflektieren Individuen beispielsweise „zu viel Zucker ist nicht gut“ (Als Konklusion eines Syllogismus, welcher das Element „Gesundheit“ enthält) häufig genug, so bildet sich hieraus eine assoziative Verbindung zwischen „Zucker“ und „gut“ via dem Element „gesund“. Individuen werden somit, wenn sie reflexiv prozessieren, zu dem Schluss kommen, dass Zucker nicht gut ist. Durch die entstandene assoziative Verknüpfung kann das Element „Zucker“ nun ebenfalls impulsiv prozessiert werden. Da Negationen, wie bereits erwähnt, jedoch auf Reflexion angewiesen sind, könnte innerhalb des gleichen Individuums eine rein impulsive Prozessierung des Elementes „Zucker“ zu dem Ergebnis „Zucker ist gut“ führen, so eine reflexive Prozessierung verhindert wird.

Über diese Beschreibung des reflexiven und des impulsiven Systems hinaus leistet das RIM ebenfalls eine Beschreibung der Interaktion beider Systeme. Diesbezüglich sei erneut erwähnt, dass das impulsive System kontinuierlich aktiv arbeitet, ohne dass dies mit bewusst-kognitiven Anstrengungen verbunden ist, während das reflexive System ebenfalls inaktiv sein kann. Die Konsequenz hieraus ist, dass das reflexive System, wenn aktiv, nicht anstelle des impulsiven Systems, sondern zusätzlich zu diesem, das heißt parallel operiert. Dies bedeutet, dass die beiden Systeme bei aktiver Prozessierung lediglich für reflexive Operationen interagieren, da das reflexive System zum Zeitpunkt rein impulsiver Operationen nicht aktiv ist, auch wenn sich vergangene reflexive Prozesse qua Bildung und Festigung assoziativer Verknüpfungen auf aktuelle, rein impulsive Prozesse auswirken könne. Entsprechend übertragbar entstammen Elemente des Arbeitsgedächtnisses, welche gemäß dem Modell reflexiv prozessiert werden, dem Langzeitgedächtnis, welches mit dem impulsiven System assoziiert ist, so dass reflexives Operieren stets von vorangegangenen impulsiven Prozessen abhängig ist. Reflexive und impulsive Operationen sind somit bedingt durch Wahrnehmungen eines Stimulus, wobei die Intensität und die Höhe der dem Stimulus zugeteilten Aufmerksamkeit darüber bestimmen, ob ein Stimulus impulsiv prozessiert werden kann oder zusätzlich reflektiert werden soll oder muss. Operieren Individuen entsprechend reflexiv, so wurden bereits assoziative Elemente des impulsiven Systems bei Wahrnehmung prozessrelevanter Objekte aktiviert. Krishna und Strack (2017) führen diesbezüglich das Beispiel aus, dass wenn Individuen die Frage reflektieren, was sie in einer Cafeteria essen wollen, sich z.B. durch Perzeption der Speisekarte, durch das impulsive System gesteuert, automatisch assoziative Elemente aktivieren. Der Anblick eines Kuchens könnte so

beispielsweise zur Aktivierung der Assoziation „Kuchen ist lecker“ führen, was wiederum, zusätzlich zur Wahrnehmung weiterer Elemente wie etwa dem Aussehen der Speisen weitere Assoziationen aktivieren kann. Es ist dann das reflexive System, welches die aktivierten Elemente und Assoziationen weiter prozessiert, wodurch sich erneut weitere Elemente und Assoziationen automatisch aktivieren können, so dass z.B. die syllogistischen Konklusionen „Kuchen ist ungesund“ und „Salat ist gesund“ resultieren. Diese reflexive Prozessierung der impulsiv aktivierten Elemente führt folglich zur impulsiven Aktivierung weiterer Elemente, welche erneut reflektiert werden usw., wobei sich dieser Prozess solange wiederholt, bis schließlich eine Entscheidung abgeleitet wird. Zwischen allen Entscheidungsalternativen wird diesbezüglich diejenige selektiert, welche am konsistentesten mit den zuvor reflexiv prozessierten Aussagen ist, so dass beispielsweise „Gesundheit ist gut“, „Geschmack ist gut“ und „Gesundheit ist wichtiger als Geschmack“ zur konsistenten Entscheidung führt, in einer Cafeteria einen Salat statt einen Kuchen zu essen. Entsprechend sind reflexive Operationen stets durch die automatischen Aktivierungsmuster assoziativer Elemente des impulsiven Systems bedingt und limitiert – es werden zunächst diejenigen Elemente innerhalb reflexiver Operationen berücksichtigt, welche impulsiv und automatisch aktiviert werden.

Strack und Deutsch (2004) führen weiterhin aus, dass wenn die durch das impulsive System aktivierten Assoziationen valide für und konsistent mit reflexiven Operationen sind, weitere Elemente, welche relevant für die aktuelle reflexive Operation sind, zugänglicher werden, wodurch sich der kognitive Aufwand, der zum Reflektieren nötig ist, reduziert. Entsprechend entstehen in Individuen positive Affekte und die Entscheidungsfindung fällt leichter, wenn impulsiv aktivierte Assoziationen konsistent mit reflexiv prozessierten Operationen sind, was als synergistischer Prozess beschrieben wird. Dies wird erneut durch das Cafeteriabeispiel von Krishna und Strack (2017) illustriert, in welchem sich das zuvor beschriebene Individuum natürlicherweise und mit einem guten Gefühl für den gesunden Salat entscheidet, wenn dieser das einzige leckere Gericht auf der Speisekarte ist. Demgegenüber entstehen in Individuen gemäß RIM konflikthafte Gefühle, wenn impulsiv aktivierte Assoziationsmuster in Widerspruch stehen zu konsistent reflektierten Konklusionen. In diesem Fall erfordert es zusätzlichen kognitiven Aufwand, um gänzlich neue bzw. andere Assoziationsmuster zu aktivieren, welche bei weiterer reflexiver Prozessierung schließlich zu konsistenten Konklusionen führen, was als antagonistischer Prozess beschrieben wird. Dies führt schließlich zu der Konsequenz, dass Wahrheitswerte von Konklusionen höher sind für synergistische Prozesse und geringer sind für antagonistische Prozesse, so das reflexive System Konklusionen nicht bewusst um Konsequenzen synergistischer Effekte editiert. Ob ein Prozess synergistisch oder antagonistisch ist, was sich

auch dadurch beschreiben lässt, wie flüssig ein Prozess verläuft (flüssiger für synergistisch, weniger flüssig für antagonistisch), bedingt schließlich die Sicherheit eines Individuums bezüglich der prozessierten, syllogistischen Konklusion sowie wie sehr ein Individuum hinter einer getroffenen Entscheidung steht. Verhalten auf Basis reflexiver Prozesse ist schließlich eine Konsequenz eines Entscheidungsprozesses, in welchen Wissen über den Wert und die Eintrittswahrscheinlichkeit verschiedener Handlungskonsequenzen einfließen, welche entsprechend gewichtet werden und somit zur Bildung von Präferenzen und Intentionen führen. Verhalten auf Basis rein impulsiver Prozesse basieren hingegen auf automatisch aktivierten Assoziationsnetzwerken hinsichtlich verschiedener Verhaltensschemata, welche durch Wahrnehmungen ebenso wie durch vorausgegangene Reflektionen gesteuert sind.

Auf das Antwortverhalten von Individuen in Umfragen abstrahiert bedeutet dies zunächst, dass zwischen einem rein impulsiven und zusätzlich einem reflexiven Antwortverhalten differenziert werden muss. Beantworten Befragte gemäß RIM Fragen rein impulsiv, so aktivieren sich in ihnen entsprechend der Exposition mit dem Inhalt der Frage gewisse Assoziationen. Diese betreffen zum einen Elemente, welche mit dem konkreten Inhalt der Frage assoziiert sind, zum anderen können aber auch weitere Elemente automatisch aktiviert werden welche sich aus zusätzlichen, frageunabhängigen Stimuli wie Eigenschaften der Befragungssituation oder der Tatsache an sich, dass das Individuum sich in einer Befragungssituation befindet, ergeben. Für den Inhalt der Frage relevante Assoziationen aktivieren sich entsprechend bei Perzeption einer Frage automatisch gemäß ihrer Zugänglichkeit in dem Sinne, dass wenn besonders starke Assoziationen bezüglich dem Inhalt der Frage vorliegen und zugänglich sind, sich diese automatisch innerhalb eines Assoziationsnetzwerkes aktivieren, wobei das Antwortverhalten schließlich auf diesen kontextrelevanten Assoziationen basiert, was in einer Antwort resultiert, welche kohärent mit individuellen Dispositionen ist. Sind Assoziationen bezüglich des Inhaltes der Frage hingegen nicht verfügbar oder zugänglich, so basiert impulsives Antwortverhalten auf für den Inhalt der Frage weniger relevanten bis irrelevanten Assoziationen, was zu verzerrten Antworten führt. Impulsiv werden in Befragungssituationen jedoch nicht nur Assoziationen zum konkreten Inhalt der Frage aktiviert (so diese vorhanden sind), da auch die Wahrnehmung situativer Faktoren Assoziationen aktivieren kann, welche Handlungsleitend sind. So können beispielsweise ebenfalls Assoziationen, welche zu Befragungen im Allgemeinen (unabhängig von deren Inhalt) existieren (wie etwa dem Eindruck, dass Befragungen „sowieso nichts bringen“ oder die gegenteilige Assoziation, dass Befragungen sehr wichtig sind), aktiviert werden, welche schließlich zur Aktivierung eines spezifischen Verhaltensschemas führt. Es ist somit zu

erwarten, dass impulsiv prozessierte Antworten in Umfragen dann „wahr“ sind, wenn Assoziationen zum Inhalt der Frage verfügbar sind, wenn positive Assoziationen gegenüber der Teilnahme an Umfragen im Allgemeinen (bzw. gegenüber weitern, für die Befragungssituation potenziell relevanten Faktoren) vorliegen und wenn beide Assoziationsnetzwerke kohärent aktiviert und prozessiert werden. Ist dies nicht der Fall, so kann für rein impulsiv prozessierte Antworten mit der ein oder anderen Form von Antwortverzerrung gerechnet werden.

Das Fehlen oder die geringe Zugänglichkeit solcher Assoziationen kann (zusätzlich zur wahrgenommenen Komplexität des zu prozessierenden Stimulus) ebenfalls dazu führen, dass die Frage nicht nur rein impulsiv, sondern zusätzlich reflexiv prozessiert wird, wobei auch hier das jeweils aktivierte Assoziationsnetzwerk, nun jedoch in Kombination mit der Prozessierung der Assoziationen innerhalb der syllogistischen Logik und antizipierter Erwartungswerte der Handlungskonsequenzen über das finale Antwortverhalten entscheidet. So ist bei vorhandenen Assoziationen bezüglich des Inhaltes einer Frage gemeinsam mit positiven Assoziationen bezüglich der Teilnahme an Befragungen, so diese valide für und konsistent mit reflexiven Operationen sind, für reflexive, synergistische Prozesse ein wahres Antwortverhalten zu erwarten. Demgegenüber kann es für antagonistische Operationen zu bewussten Antwortverzerrungen kommen, so die Reflexion zu einer Editierung einer formal vorgesehenen Antwort führt, z. B. weil eine automatisch aktivierte negative Assoziation von Umfragen in der syllogistischen Konklusion resultiert, dass wahre Antworten einen geringeren Wert hinsichtlich einer antizipierten Verhaltenskonsequenz aufweisen als verzerrte Antworten. Aber auch eine zu geringe Zugänglichkeit von Assoziationen hinsichtlich dem Inhalt der Frage kann unter der Bedingung reflexiver, antagonistischer Prozesse zu verzerrten Antworten führen, wenn die reflexive Operation darin besteht, für die Befragungssituation neue Meinungen zu prozessieren, diese aber auf automatisch aktivierten, für den Inhalt der Frage jedoch irrelevanten Assoziationen beruht.

Kognitive Faktoren mit Einfluss auf das Antwortverhalten gemäß RIM sind somit a) die Form, Validität und Zugänglichkeit der Assoziationen bzw. des assoziativen Netzwerkes bezüglich dem konkreten Inhalt der Frage; b) die Form, Validität und Zugänglichkeit der Assoziationen bzw. des assoziativen Netzwerkes bezüglich weiterer Faktoren der Umfragesituation bzw. bezüglich der Teilnahme an Umfragen selbst; und c) der Prozesspfad (rein impulsiv oder zusätzlich reflexiv) welcher darüber entscheidet, innerhalb welcher der Modi (unter Berücksichtigung der jeweiligen Limitationen) eine Antwort prozessiert wird. Die weiteren genannten Modellfaktoren wie etwa die Arbeitsgedächtniskapazität werden an dieser Stelle nicht genannt, da diese vor allem den kognitiven Aufwand reflexiver Operationen bedingen, wobei keine Aussage darüber getroffen werden kann, dass besonders elaborierte Antworten einen höheren

Wahrheitsgehalt aufweisen. Dieser ist gemäß dem Modell vor allem von der Zugänglichkeit und Validität spezifischer Assoziationen abhängig und für zusätzlich reflektierte Antworten höher für synergistische als für antagonistische Operationen.

Zusammenfassend kommt es gemäß des Modells entsprechend dann zu Antwortverzerrungen, wenn keine validen Assoziationen bezüglich dem konkreten Inhalt der Frage vorliegen oder zugänglich sind und Antworten stattdessen auf der Basis irrelevanter Assoziationen prozessiert werden, wenn aktivierte Assoziationen bezüglich der Umfragesituation Handlungsschemata aktivieren, welche die Verzerrung von Antworten implizieren und wenn die Generierung von Antworten antagonistischen Operationen in dem Sinne entsprechen, als dass impulsiv aktivierte Assoziationen nicht kohärent sind mit reflexiv prozessierten Syllogismen.

Das Modell der Frame-Selektion (MFS)

Bezüglich des Prozesses der Antwortgenerierung ist den bisher beschriebenen Modellen nun also zum einen die Bedeutung der Differenzierung verschiedener Modi der Informationsprozessierung (vgl. z.B. das Satisficing-Optimizing Model nach Krosnick, 1991; das MODE-Modell nach Fazio, 1990) und zum anderen die Bedeutung der Zugänglichkeit von Einstellungen, Informationen, Assoziationen etc. zu entnehmen (vgl. z.B. die Theory of planned behavior nach Ajzen, 1991; das MODE-Modell nach Fazio, 1991; das RIM nach Strack und Deutsch, 2004). Über diese generischen Verhaltensfaktoren hinaus implizieren die jeweiligen Modelle im Kontext des Befragtenverhaltens jedoch auch, dass spezifische Wahrnehmungen und Einstellungen gegenüber der Befragungssituation ebenfalls einen Einfluss auf das konkrete Antwortverhalten nehmen. Um diesen Punkt abschließend genauer zu erörtern wird ein weiteres Modell beleuchtet, welches ebenfalls zur Erklärung von Antwortverhalten Verwendung findet und innerhalb dessen Stocké (2004) den Einfluss verschiedener Befragtenrollen modelliert: Die Frame-Selektionstheorie (Esser, 2001), welche zu Beginn zwar aufgrund einer Reihe von Ungenauigkeiten und Unvollständigkeiten in der Kritik stand, nun jedoch in einer vollständigen und in sich konsistent formulierten Fassung vorliegt (Kroneberg, Die Definition der Situation und die variable Rationalität der Akteure. Ein allgemeines Modell des Handelns, 2005) und ebenfalls von Esser bestätigt wurde (Esser, 2006a; Esser, 2006b). Zudem wurden in der Vergangenheit eine ganze Reihe sogenannter Framing-Modelle formuliert, um die Überlegungen dualer Prozessmodelle, vor allem automatisch-spontane und habitualisierte Handlungen (also der Modi eingeschränkter Rationalität) in die Erklärungslogik der Rational Choice Theory zu integrieren. Im deutschen

Sprachraum durchgesetzt hat sich jedoch vor allem Essers bereits zuvor erwähntes Modell (Mayerl, 2009, S. 156). Die wichtigsten beiden Grundannahmen der verschiedenen Framing-Modelle lauten nach Skarbek-Kozietulska, Preisendörfer und Wolter (2012, S. 10) dabei:

1. Um Situationen im Alltag zu bewältigen, greifen Akteure auf bestimmte, zur Interpretation der Situation passende Skripte zurück, wobei besagte Definition der Situation dem Prozess des Framings entspricht.
2. Dem Framingprozess voraus geht die Entscheidung, ob sowohl das Framing, als auch die nachfolgende Handlung in einem automatisch-spontanen oder einem überlegt-kontrollierten Modus ablaufen, wobei der überlegt-kontrollierte Modus dem Handeln gemäß der Rational Choice Theory entspricht.

Entsprechend finden sich die beiden Grundgedanken ebenfalls in dem von Esser (2001) formulierten und von Kroneberg (2005) ergänzten Modell der Frame-Selektion (MFS), welches im Folgenden zusammenfassend beschrieben wird (vgl. Esser, 2001; Esser, 2006a; Esser, 2006b; Kroneberg, 2005; Kroneberg, 2007; Kroneberg, 2011).

Grundsätzlich handelt es sich beim MFS um eine Handlungstheorie, welche Handeln auf eine Vorstrukturierung durch kognitiv-emotional verankerte Schemata zurückführt und von einer variablen Rationalität der Akteure ausgeht. Dem Modell inne liegt die Annahme, dass dem Handeln jeweils eine Selektion eines Frames, eines Skriptes und einer Handlungsalternative vorausgeht, wobei alle drei Selektionsprozesse jeweils entweder innerhalb eines automatisch-spontanen oder eines reflexiv-kalkulierenden Modus stattfinden. Werden Frames, Skripte oder Handlungen automatisch-spontan selektiert, so erfolgt die Selektion unhinterfragt hinsichtlich bestimmter Alternativen, beispielsweise auf Basis von Gewohnheiten oder Normen. Erfolgt hingegen eine reflexiv-kalkulierte Selektion, so geht dieser ein systematischer Vergleich verschiedener Alternativen voraus, woraus die bewusste Reflexion der Alternativen und schließlich eine Selektion erfolgt. Selektionen innerhalb dieses Modus können somit auf Basis von Entscheidungsregeln formalisiert und innerhalb des Rational Choice Ansatzes eingebettet werden. Beispielsweise können Individuen innerhalb dieses Selektionsprozesses diejenige Alternative wählen, welche für diese subjektiv hinsichtlich einer wünschenswerten Handlungskonsequenz möglichst optimal ist oder aber welche hinsichtlich der Wahrnehmung der sozialen Situation am angemessensten erscheint. Die Frame-Selektion entspricht diesbezüglich der individuellen und situativen Definition der Situation, wie sie als Einflussfaktor innerhalb der vorangegangenen Modelle beschrieben wurde, bezieht sich hierin auf Situationsmerkmale oder Situationstypen und grenzt damit für das Individuum ein, welche Situation inklusive daraus

folgender Handlungsschemata vorliegt. Die Skript-Selektion entspricht der darauffolgenden Aktivierung eines Handlungsprogramms, hat somit unmittelbaren Handlungsbezug und wird beeinflusst von Normen, Routinen, (kulturellen) Verhaltensdispositionen und (genetisch) fixierten Verhaltensprogrammen. Die Handlungs-Selektion zuletzt entspricht schließlich der hierauf folgenden Selektion einer Handlungsalternative. Frames und Skripte verweisen somit in dem Sinne auf soziale Vorstrukturierungen des Handelns, als dass sie im Wesentlichen das Ergebnis individueller Lern- und Sozialisationsprozesse sind und häufig durch Interaktionsprozesse der handelnden Individuen ausgelöst werden, innerhalb welcher diese die Bedeutung der aktuellen Situation durch das Senden und Empfangen von für die Situation relevanten Symbolen definieren.

Die zugrundeliegenden Regeln reflexiv-kalkulierter Selektionen von Frames, Skripten und Handlungen können vereinfacht gemäß der SEU-Theorie (SEU = **S**ubjective **E**xpected **U**tility) dargestellt werden (Kroneberg 2007), wonach Selektionen erfordern, dass für alle relevanten Alternativen *m = 1, ..., M* entsprechend relevante Erwartungen und Bewertungen identifiziert werden, für welche schließlich ein subjektives SEU-Gewicht auf Basis eines subjektiven Nutzens einer Handlungskonsequenz U_m sowie der subjektiv wahrgenommenen Eintrittswahrscheinlichkeit der Handlungskonsequenz bei Auftreten der Handlung p_m wie folgt prozessiert wird:

$$SEU(.) = \sum p_m U_m \qquad \textbf{Formel 1}$$

$SEU(.)$ entspricht hierbei entweder $SEU(F_i)$, $SEU(S_j|F_i)$ oder $SEU(A_k|S_j)$, je nachdem, ob das SEU-Gewicht für ein Frame F_i (mit *i* als Laufindex aller Frames), ein Skript S_j (mit *j* als Laufindex aller Skripte) oder eine Handlungsalternative A_k (mit *k* als Laufindex aller Handlungsalternativen) kalkuliert wird.

Innerhalb des automatisch-spontanen Modus wird hingegen derjenige Frame selektiert, welcher über den höchsten Match verfügt, sprich, unmittelbar am kohärentesten mit der gegebenen Situation ist. Das Selektionsgewicht eines Frames $G(F_i)$ entspricht somit dem Match m_i, welcher sich wie folgt darstellen lässt:

$$G(F_i) = m_i = a_i v_i o_i \quad \textbf{Formel 2}$$

Diesbezüglich beschreibt a_i (∈ [0,1]) die Stärke der mentalen Zugänglichkeit des Frames, o_i (∈ [0,1]) den Grad des Vorliegens von für die Identifikation der Situation signifikanten Stimuli in der aktuellen Situation und v_i (∈ [0,1]) die Stärke der individuellen Assoziation zwischen Frame und Stimuli.

Die Skriptselektion im automatisch-spontanen Modus wiederum ist davon abhängig, welches der Skripte S_j mental am stärksten mit dem zuvor aktivierten Frame F_i assoziiert ist, was durch die Zugänglichkeit $a_{j|i}$ (∈ [0,1]) dargestellt wird, sowie, welches der Skripte S_j generell am stärksten verankert ist, was durch die generelle Zugänglichkeit a_j (∈ [0,1]) dargestellt wird. Da die Aktivierung von Skripten somit umso unwahrscheinlicher ist, je unsicherer Individuen bezüglich des Matchs m_i des selektierten Frames F_i mit der jeweiligen Situation sind, ergibt sich das Selektionsgewicht des Skriptes S_j in Abhängigkeit der Definition einer Situation F_i wie folgt:

$$G(S_j|F_i) = a_j a_{j|i} m_i \quad \textbf{Formel 3}$$

Die Selektion einer Handlung im automatisch-spontanen Modus schließlich ist zusätzlich davon abhängig, mit welchem Grad das aktivierte Skript S_j die jeweilige Handlungswahl bedingt, was durch den Parameter $a_{k|j}$ (∈ [0,1]) dargestellt wird. Das Selektionsgewicht der Handlungsalternative A_k lässt sich entsprechend wie folgt darstellen:

$$G(A_k|S_j) = a_{k|j} G(S_j|F_i) \quad \textbf{Formel 4}$$

Darüber hinaus leistet das MFS ebenfalls eine Beschreibung der Bedingungen, unter welchen Individuen Frames, Skripte und Handlungsalternativen automatisch-spontan oder reflexiv-kalkulierend selektieren, wodurch es zur Erklärung variabler Rationalität und heuristischen Handelns beiträgt. Demnach sind Selektionen im reflexiv-kalkulierenden Modus umso wahrscheinlicher, a) je höher die Opportunitäten und die Motivation zum bewussten Reflektieren sind; b) je geringer der Aufwand für eine bewusste Reflexion ist; und c) je geringer die Aktivierung

eines für die Situation relevanten mentalen Schemas ist, welches auch ohne bewusste Reflexion adäquat Handlungsweisend ist. Die entsprechenden Werte der genannten Variablen ergeben sich aus der direkten Wahrnehmung von Situationsmerkmalen und der automatischen Aktivierung von Gedächtnisstrukturen und Dispositionen, welche sich Großteils aus früheren Erfahrungen ergeben. Aktiviert wird entsprechend jeweils derjenige Modus, welcher im Sinne der genannten Variablen optimal ist, wobei es sich hierbei um einen vorbewussten Prozess handelt, der keiner kontrollierten Aufmerksamkeitszuweisung seitens der Individuen unterliegt. Die Aktivierung eines Modus ergibt sich entsprechend aus den wahrgenommenen Reflexionsopportunitäten p ($\in [0,1]$), der Motivation zur Reflexion U (≥ 0), den Reflexionskosten C (≥ 0) sowie dem Aktivierungsgewicht der am stärksten zugänglichen Alternative, d.h. je nach Selektion eines Frames, eines Skriptes oder einer Handlung $G(F_i)$, $G(S_j|F_i)$ oder $G(A_k|S_j)$. Individuen aktivieren entsprechend dann den reflexiv-kalkulierenden Modus, wenn das Aktivierungsgewicht der automatisch-spontan zugänglichen Alternative zu gering ist, wobei sich der Schwellenwert, unterhalb welchem stattdessen der reflexiv-kalkulierenden Modus aktiviert wird, wie folgt darstellen lässt:

$$G(.) < 1 - C/pU \qquad \textbf{Formel 5}$$

(Wobei $G(.)$ hierbei entweder $G(F_i)$, $G(S_j|F_i)$ oder $G(A_k|S_j)$ entspricht).

Aus automatisch-spontan oder reflexiv-kalkulierend selektierten Frames, welche wiederum spontane oder kalkulierende Skripte beinhalten, lassen sich vier Handlungstypen identifizieren, welche jedoch um zwei Varianten ergänzt werden müssen die eintreten, wenn einem spontanen oder kalkulierenden Frame kein Skript zugeordnet ist. Im Falle kalkulierenden Prozessierens handeln Individuen dann rational und selektieren diejenige Alternative mit der höchsten subjektiven Nutzenerwartung unter Gültigkeit des zuvor selektierten Frames gemäß der Rational Choice Theory. Im Falle spontanen Prozessierens ist das Handeln von Akteuren bei fehlenden Skripten eine reine Zufallsreaktion (Esser, 2001).

Auf das Antwortverhalten in Umfragen adaptiert bedeutet dies, dass Befragte zunächst (automatisch-spontan oder reflexiv-kalkulierend) einen Frame selektieren, welcher deren individueller Definition und Bewertung der Umfragesituation entspricht. Hierauf folgend selektieren Befragte (automatisch-spontan oder reflexiv-kalkulierend) ein Handlungsskript, welches als valide und akzeptabel im Kontext des selektierten Frames wahrgenommen wird und ein konkretes

Handlungsprogramm für die wahrgenommene Umfragesituation enthält. Zuletzt erfolgt die (automatisch-spontane oder reflexiv-kalkulierende) Selektion der konkreten Handlungsalternative welche schließlich beschreibt, wie sich Befragte konkret in Umfragen verhalten und in welchem der beiden Modi die konkrete Antwort prozessiert wird. Zur Illustration sei diesbezüglich angenommen, dass Befragte Umfragen (bzw. eine Umfragesituation) entweder positiv bewerten und tendenziell durch wahre Antworten unterstützen (unterstützender Frame), was in einer spezifischen Umfragesituation ebenfalls durch ein konkretes Interesse am Thema der Befragung bedingt sein kann, oder Umfragen und deren Unterstützung indifferent oder gar ablehnend gegenüberstehen, was in einer erlebten Indifferenz gegenüber der Gabe wahrer Antworten resultiert (indifferenter Frame). Diese beiden Pole der Wahrnehmung und Bewertung von Umfragen entsprechen zwei möglichen, beispielhaften Frames, welche für Befragte in konkreten Umfragesituationen als Alternativen für die Selektion zur Verfügung stehen.

Wird die Frameselektion automatisch-spontan prozessiert, so wird entsprechend des Modells derjenige der beiden beispielhaften Frames selektiert, welcher für Befragte am zugänglichsten ist, welcher durch situativ vorhandene Stimuli am deutlichsten betont wird und für welchen die stärksten individuellen Assoziationen zwischen Frame und Stimuli vorliegen. Dies bedeutet, dass Befragte dann automatisch-spontan einen unterstützenden Frame selektieren, wenn die generelle individuelle Unterstützung von Umfragen oder die spezifische Unterstützung von Umfragen zu einem spezifischen Thema mental präsent und zugänglich ist, wenn situative Hinweisreize kohärent mit der Unterstützung der Umfragen sind und wenn individuelle Assoziation zwischen einem unterstützenden Frame und kohärenten Hinweisreizen vorhanden sind. Der indifferente Frame wird hingegen dann automatisch-spontan selektiert, wenn kein unterstützender Frame zugänglich ist oder wenn gar ein ablehnender/indifferenter Frame höher zugänglich ist als ein unterstützender Frame, wenn situativ vorhandene Stimuli eher eine Ablehnung/Indifferenz implizieren (z.B. durch die Wahrnehmung, dass die Umfrage keinen wissenschaftlichen Standards entspricht oder bewusste Täuschungen seitens der Entwickler der Umfrage vorliegen) und wenn keine Assoziationen zwischen Stimuli und unterstützendem Frame bzw. wenn Assoziationen zwischen Stimuli und ablehnendem Frame vorliegen. Im reflexiv-kalkulierenden Modus hingegen selektieren Befragte denjenigen Frame, welcher für diese zwischen allen prozessierten Framealternativen das größte Produkt zwischen subjektivem Nutzen einer Handlungskonsequenz und der subjektiv wahrgenommenen Eintrittswahrscheinlichkeit der Handlungskonsequenz aufweist. Dies bedeutet, dass der unterstützende Frame dann reflexiv-kalkulierend selektiert wird, wenn der subjektive Nutzen eines unterstützenden Handelns (z.B. kön-

nen Befragte davon überzeugt sein, dass sie durch eine Unterstützung der Umfrage zum wissenschaftlichen Erkenntnisgewinn beitragen) und die subjektive Wahrnehmung der Eintrittswahrscheinlichkeit (z.B. können Befragte davon überzeugt sein, dass das die Umfrage durchführende Institut seriös arbeitet und die Studienerkenntnisse sorgfältig aufbereitet der Öffentlichkeit zur Verfügung stellt) höher sind als für alternative Frames. Ein indifferenter Frame wird hingegen dann reflexiv-kalkulierend selektiert, wenn der subjektive Nutzen eines indifferenten Handelns (z.B. können Befragte kein Interesse an der Umfrage selbst, sondern an einer möglichen Belohnung für die Teilnahme an der Umfrage haben, wodurch es vor allem attraktiv ist, Fragen möglichst schnell und oberflächlich statt gründlich und ernsthaft zu beantworten) und die subjektiv wahrgenommene Eintrittswahrscheinlichkeit einer Handlungskonsequenz (z.B. können Befragte der Meinung sein, dass das die Umfrage durchführende Institut nicht seriös ist und mögliche Erkenntnisse generell nicht wissenschaftlich verwertet werden) höher sind als für alternative Frames. Sowohl automatisch-spontan, als auch reflexiv-kalkulierend selektierte Frames können somit entweder zu einer kooperativen Unterstützung von Umfragen oder einer tendenziell indifferenten oder ablehnenden Haltung gegenüber Umfragen führen, je nach den kognitiv individuell zugrundeliegenden generellen Einstellungen, Bewertungen und Erfahrungen bzw. Dispositionen bezüglich Umfragen. Situative Faktoren mit weiterem Einfluss auf die Selektion eines Frames können diesbezüglich als weitere Randbedingungen genannt werden, welche selbst nicht kognitiver Natur sind, kognitive Prozesse jedoch beeinflussen.

Je nachdem, ob nun ein unterstützender oder ein indifferenter Frame selektiert ist, kommt es im nächsten Schritt zur Selektion eines Skriptes, welches innerhalb des spezifisch selektierten Frames eingebettet ist. Wurde ein unterstützender Frame selektiert so ist es als Beispiel plausibel, dass als Handlungsskript „intendiertes wahrheitsgemäßes Beantworten von Fragen“ selektiert wird. Für automatisch-spontane Selektionen ist dies dann der Fall, wenn das wahrheitsgemäße Beantworten von Fragen generell und innerhalb des unterstützenden Frames am stärksten zugänglich ist. Für reflexiv-kalkulierende Selektionen ist dies dann der Fall, wenn der wahrgenommene Nutzen und die Wahrnehmung der Eintrittswahrscheinlichkeit einer antizipierten Handlungskonsequenz (z.B. „zum wissenschaftlichen Erkenntnisgewinn beitragen“) am höchsten zwischen allen Skriptalternativen ist. Wurde hingegen der indifferente Frame selektiert, so stehen als Handlungsskripte beispielsweise eine lediglich oberflächliche Beantwortung von Fragen oder gar die bewusst verzerrte Beantwortung von Fragen zur Verfügung. Werden Skripte automatisch-spontan prozessiert, so wird diejenige Alternative selektiert, welche sowohl innerhalb des indifferenten Frames, als auch generell am stärksten verfügbar ist, was z.

B. durch Gewohnheiten in der jeweiligen Situation, sprich, dem eigenen Handeln in vergangenen, vergleichbaren Umfragesituationen bedingt ist. Neigen Befragte entsprechend generell eher dazu, Fragen oberflächlich zu beantworten, so wird dieses Skript statt der Alternative selektiert. Neigen Befragte hingegen dazu, Fragen nicht wahrheitsgemäß zu beantworten (z.B. um eine besonders positive Selbstdarstellung zu leisten), so werden sie eher dieses Skript selektieren. Zugrundeliegende generelle Einstellungen, Bewertungen und Erfahrungen bzw. Dispositionen bezüglich Umfragen bzw. der konkret vorliegenden Umfrage (was das Thema und den Inhalt der konkret vorliegenden Umfrage berücksichtigt) bedingen somit nicht nur die Selektion von Frames, sondern ebenfalls von Handlungsskripten in Umfragen: Sind generelle Einstellungen, Bewertungen und Erfahrungen bzw. Dispositionen bezüglich Umfragen bzw. der konkret vorliegenden Umfrage eher positiv, so wird eher ein unterstützender Frame und innerhalb dessen eher Handlungsskripte, welchen das wahrheitsgemäße Beantworten von Fragen inhärent ist, selektiert. Sind Einstellungen, Bewertungen und Erfahrungen bzw. Dispositionen hingegen eher negativ bzw. indifferent, so entscheiden situative Faktoren und insbesondere kognitiv verfügbare Handlungsgewohnheiten in vergleichbaren, durch Ablehnung oder Indifferenz geprägten Umfragesituationen die Selektion eines Handlungsskriptes. Antworten können unter diesen Voraussetzungen dennoch verschiedene Grade an Elaboration aufweisen welche schließlich darüber Aufschluss geben, ob beispielsweise eine oberflächliche Beantwortung von Fragen vorliegt (welche durch Heuristiken und kognitive Verzerrungen geprägt sein können) oder Fragen trotz Elaboration bewusst verzerrt wurden (was einer willentlichen Editierung einer eigentlich wahren Antwort entspricht).

Innerhalb der Skripte stehen schließlich jeweils eingebettete Handlungsalternativen zur Verfügung, welche entsprechend ebenfalls automatisch-spontan oder reflexiv-kalkulierend ausgeführt werden. Wird die eingebettete Handlung des Skriptes „wahrheitsgemäßes Beantworten der Fragen“ automatisch-spontan ausgeführt, so ist es plausibel, dass auch eine wahrheitsgemäße Beantwortung intendiert wird. Auch für eine reflexiv-kalkulierende Ausführung der Handlung innerhalb des Skriptes werden wahre Antworten intendiert, so diese jeweils über den höchsten wahrgenommenen Nutzen und die höchste assoziierte Eintrittswahrscheinlichkeit der Handlungskonsequenz, bedingt durch aktivierte Dispositionen der zugrundeliegenden Frames verfügen, was ebenfalls als plausibel angenommen werden kann. Das Skript „oberflächliche Beantwortung der Frage“ impliziert beispielsweise eine automatisch-spontane Ausführung der Handlung, was, so die Handlung innerhalb des Skriptes am verfügbarsten ist, zu einer entsprechend schnellen und oberflächlichen Beantwortung von Fragen führt. Das Skript „verzerrte Be-

antwortung von Fragen“ zuletzt impliziert eine reflexiv-kalkulierende Ausführung der Handlung und wird aktiviert, so diese jeweils über den höchsten wahrgenommenen Nutzen und assoziierte Eintrittswahrscheinlichkeit der Handlungskonsequenz verfügt, was schließlich zu einer elaborierten, aber unter Umständen verzerrten Beantwortung von Fragen führt.

Gemäß des MFS lassen sich somit aus dem obigen Beispiel drei exemplarische Handlungstypen für die Teilnahme an Umfragen ableiten, welche 1.) in einem intendierten wahrheitsgemäßen Beantworten der Fragen; 2.) in einer oberflächlichen Beantwortung von Fragen; und 3) in einer bewusst verzerrten Beantwortung von Fragen resultieren. Auf kognitiver Ebene wird dann entsprechend das wahrheitsgemäße Beantworten von Fragen intendiert, wenn ein Frame selektiert wird, welcher Umfragen unterstützt und innerhalb dessen ein Skript selektiert wird, welche die Unterstützung von Umfragen durch das wahrheitsgemäße Beantworten der Fragen definiert. Relevant hierbei bleibt, dass Antworten unter Umständen dennoch verzerrt sein können, auch wenn wahre Antworten intendiert sind. Eine oberflächliche (und schnelle) Beantwortung von Fragen wird hingegen dann selektiert, wenn ein Frame selektiert wurde, welches sich durch eine ablehnende oder indifferente Haltung gegenüber Umfragen äußert und ein Skript selektiert ist, welches schnelles und oberflächliches Beantworten als akzeptable Handlung innerhalb des Frames definiert, z. B. weil das entsprechende Verhalten auch in vergangenen Umfragesituationen subjektiv erfolgreich angewandt wurde. Eine bewusst verzerrte (und elaborierte) Beantwortung von Fragen wird schließlich dann selektiert, wenn ebenfalls ein Frame selektiert wurde, welches sich durch eine ablehnende oder indifferente Haltung gegenüber Umfragen äußert und ein Skript selektiert ist, welches bewusst verzerrtes Beantworten als akzeptable Handlung innerhalb des Frames definiert, z. B. weil eine positive Verzerrung der eigenen Selbstdarstellung im höchsten SEU-Wert resultiert. Die aktivierten Handlungsschemata innerhalb von Umfragen sind somit sowohl für die Frame-, also auch für die Skript- sowie die Handlungsselektion abhängig von individuellen Dispositionen bezüglich Umfragen, was Faktoren wie generelle Einstellungen gegenüber Umfragen, Einstellungen gegenüber dem konkreten Thema einer Umfrage, die Zugänglichkeit bestimmter Faktoren mit Relevanz für die Umfrage, aber auch vorausgegangene Handlungen in vergangenen Umfragen umfasst.

Zusammenfassend ist gemäß der Ableitung des MFS dann mit Antwortverzerrungen in Umfragen zu rechnen, wenn Befragte bei der Beantwortung von Fragen a) einen indifferenten oder gar ablehnenden Frame gegenüber Umfragen selektieren und b) wenn Befragte zwar einen unterstützenden Frame selektieren, weitere kognitive Faktoren wie etwa die Zugänglichkeit von Einstellungen bezüglich dem Inhalt der Frage jedoch dazu führen, dass Fragen nicht auf der Basis von Dispositionen, sondern möglichen situativen Faktoren beantwortet werden.

2.1.4 Der Prozess der Antwortgenerierung: Ableitung allgemeiner Thesen

Wie in den vergangenen Kapiteln gezeigt, lässt sich der Prozess der Antwortgenerierung in Umfragen auf vielfältige Weise modellhaft darstellen. Trotz aller Unterschiede der diesbezüglich vorgestellten Modelle sind diesen dennoch mehr oder weniger generische bzw. allgemeingültige Faktoren und Annahmen gemein, die im Folgenden in Form von Thesen der Antwortgenerierung zusammengefasst werden, welche sich theoretisch fundiert aus den vorgestellten Modellen ableiten lassen und als kleinster gemeinsamer Nenner dieser den aktuellen Konsens in der Umfrageforschung illustrieren. Dies trägt schließlich in Kapitel 2.3 dazu bei, ein Erklärungsmodell des Auftretens von FRE abzuleiten.

Hier stellt sich zunächst die Frage, auf welchen Faktoren das Antwortverhalten in Umfragen generell basiert. Tourangeau und Kollegen (2000) verweisen diesbezüglich in ihrem Modell der Antwortgenerierung auf Faktoren, welche sich aus der Umfragesituation selbst ergeben. Hierzu gehören die Frageformulierung, weitere situative Hinweisreize und die vorgegebenen Antwortkategorien. Weiterhin existieren individuelle Faktoren mit Einfluss auf konkrete Antworten, welche durch die Umfragesituation und die gestellten Fragen bedingt sind. Hierzu gehört die gewählte Antwortstrategie, welche je nach Umfrage und Frage inter- und intraindividuell variieren kann, die zeitliche Nähe, Unterscheidungskraft und emotionale Relevanz des interessierenden Items sowie die allgemeine Zugänglichkeit derjenigen Informationen, welche für die Generierung einer Antwort im Kontext einer spezifischen Frage benötigt werden. Als von der Umfragesituation unabhängiger, individueller Faktor werden schließlich die kognitiven Fähigkeiten Befragter genannt. Auch bei Strack und Martin (1987) sowie Sudman und Kollegen (1996) findet sich ein ähnliches Schema: Als situative Faktoren werden ebenfalls die Frageformulierung sowie die vorgegebenen Antwortkategorien genannt, während als individuelle Faktoren, welche abhängig sind von situativen Faktoren, die Art und Zugänglichkeit (Häufigkeit und zeitliche Nähe der Aktivierung) der Informationen, welche zur Interpretation der Frage sowie der Bildung einer Meinung verwendet werden, die Wahrnehmung der Umfragesituation als Privat sowie die Wahrnehmung der eigentlichen Antwort als sozial unerwünscht genannt werden. Die kognitiven Fähigkeiten Befragter die benötigt werden, die Frage zu verstehen, aber auch korrekt zu interpretieren, wird schließlich ebenfalls als von der Umfragesituation unabhängiger, individueller Faktor genannt. Cannell et al. (1981) schließlich nennen ebenfalls die Frageformulierung als situativer Faktor, die kognitiven Fähigkeiten als individueller Faktor so-

wie die Zugänglichkeit von für die Antwortgenerierung relevanter Informationen und schließlich die Ziele Befragter für die Befragung als individuelle, von der Umfragesituation abhängige Faktoren.

Auf Basis dieser klassischen Modelle der Antwortgenerierung lassen sich entsprechend Faktoren identifizieren, welche konkret formulierte Antworten bedingen. Neben den allgemeinen situativen Faktoren und den prinzipiellen kognitiven Fähigkeiten sind dies auf kognitiver Ebene vor allem Definitionen, Interpretationen und Bewertungen der bzw. Einstellungen gegenüber der konkreten Umfragesituation und der Frage. Offensichtlich ist es alles andere als trivial, wie Befragte diese bewerten, ergibt sich hieraus doch der Kontext der Situation und damit die Bewertung der Relevanz von Informationen, welche schließlich zur Bildung einer Antwort verwendet werden. Erneut sei an dieser Stelle auf Schwarz (1991) verwiesen, wonach Befragte den impliziten Suchprozess nach für die Formulierung einer Antwort relevanten Informationen mitunter vorzeitig abbrechen, sobald mit subjektiv hinreichend großer Sicherheit ein Urteil gebildet werden kann, wodurch Antworten zum einen auf chronisch zugänglichen Informationen, zum anderen aber auch auf Informationen, welche lediglich unter bestimmten Umständen, das heißt situativ zugänglich sind, basieren. Aber auch die Bewertung der Situation als privat bzw. als risikobehaftet hinsichtlich der Gabe sozial potenziell unerwünschter Antworten zeichnet einen deutlichen Einfluss auf das Antwortverhalten. Eine solche Definition und Bewertung der Situation ergibt sich diesbezüglich aus situativen Hinweisreizen und individuellen Dispositionen. Schließlich zeigt sich, dass die Ziele Befragter für die Befragung sowie die gewählte Antwortstrategie auf die Formulierung einer Antwort einwirken, was als Motiv für die Befragung zusammengefasst werden kann, wobei es gemäß der Modelle nicht ersichtlich ist, ob ein solches Motiv durch die konkrete Umfragesituation oder situative Hinweisreize bedingt ist, als stabiles Persönlichkeitsmerkmal vorliegt o.Ä. Generisch betrachtet basieren Antworten gemäß dieser Modelle somit auf kognitiver Ebene auf chronisch oder situativ zugänglichen Informationen, Definition und Bewertung der Umfragesituation und der Frage sowie den Motiven Befragter für die Umfrage.

Darüber hinaus werden die hier genannten Faktoren ebenfalls durch die vorgestellten impliziten Verhaltensmodelle unterstützt. So postuliert beispielsweise auch Fazio (1986; 1990), dass soziales Verhalten eine Funktion der individuellen Interpretation der gegenwärtigen Situation und der Wahrnehmung des Objektes mit Handlungsbedarf darstellt, wodurch die Bewertung der Situation erneut betont wird. Und auch innerhalb der TPB (Ajzen, 1991) wird wohl überlegtes und elaboriertes Verhalten als Funktion einer Handlungsintention (welche einer motivationalen

Komponente entspricht), einer Einstellung gegenüber dem Verhalten (bedingt durch die konkreten Handlungsziele), subjektiven Normen (bedingt durch die Definition der sozialen Situation) und der wahrgenommenen Verhaltenskontrolle (bedingt durch die Zielerwartung) bedingt, was generisch ebenfalls zum einen als Definitionen und Bewertung der Situation und zum anderen in Form zugrundeliegender Ziele und Motive zur Teilnahme an Umfragen zusammengefasst werden kann.

Innerhalb des Satisficing-Optimizing-Models werden als relevante Faktoren mit Einfluss auf das Antwortverhalten zum Zeitpunkt der Urteilsbildung kognitiv verfügbare Informationen genannt, aus welchen schließlich eine konkrete Antwort prozessiert wird. Solche Informationen können zum einen gemäß ihrer Stärke und Zugänglichkeit aus dem Gedächtnis abgerufen werden, zum anderen können solche Informationen jedoch auch durch situative Hinweisreize wie etwa vorangegangene Fragen aktiviert werden und somit zugänglich sein (Krosnick, 1991; Krosnick, 1999a; Krosnick & Fabrigar, 1997; Schwarz, 1991). Auch Das Modell spontanen Prozessierens (Fazio, 1986; 1990) und ebenfalls das MODE-Modell (Fazio, 1990) berücksichtigt den Einfluss kognitiv chronischer Zugänglichkeit, wobei diese sich innerhalb des Modells auf die Zugänglichkeit von Einstellungen bezieht, wodurch der Einstellungszugänglichkeit eine moderierende Funktion zwischen einer Einstellung und einem Verhalten zukommt. Sind gemäß den beiden Modellen in spontanen Handlungssituationen keine handlungsrelevanten Einstellungen zugänglich, so basiert Handeln hingegen auf einfachen Heuristiken und situativ verfügbaren Hinweisreizen. Der Einfluss chronischer Zugänglichkeit wird ebenfalls durch das RIM (Krishna & Strack, 2017) betont, welches die Zugänglichkeit von Assoziationen innerhalb des impulsiven Systems als maßgeblichen Faktor für deren Aktivierung modelliert, wodurch ein Priming für die weitere Prozessierung von Antworten stattfindet. Und auch innerhalb des MFS zuletzt (vgl. Esser, 2001; Esser, 2006a; Esser, 2006b; Kroneberg, 2005; Kroneberg, 2007; Kroneberg, 2011) wird der Einfluss der kognitiven Zugänglichkeit von Informationen, Assoziationen etc. dahingehend als Relevant für den konkreten Handlungsprozess, sprich die Beantwortung von Fragen, modelliert, als dass diese die konkrete Frame- und Skriptselektion bedingen. Antworten in Umfragen basieren somit vereinfacht auf zum Zeitpunkt der Urteilsbildung zugänglichen Informationen, auch wenn solche Informationen keine oder keine hohe Relevanz für die konkret vorliegende Frage haben müssen.

Empirische Evidenzen dafür, dass die Definition der Situation Einfluss auf das Antwortverhalten nimmt gibt, sind zahlreich. Wird die Umfragesituation beispielsweise ebenfalls als soziale Situation begriffen, so zeigt sich der Einfluss der Definition der sozialen Umgebung in Form

sozial erwünschten Antwortverhaltens, worunter allgemein die Tendenz von Befragten verstanden wird, sich selbst in einem der Wahrnehmung nach besseren Licht darzustellen, als dies akkurat wäre (Paulhus, 2002). Andersen und Mayerl (2019) beispielsweise verorten sozial erwünschtes Antwortverhalten innerhalb dualer Prozessmodelle und finden Evidenzen, dass Befragte Antworten sowohl automatisch, als auch überlegt verzerren, sprich, der Einfluss der Situation unabhängig vom kognitiven Prozess der Antwortgenerierung wirkt. Daneben geht auch aus einem Literatur Review von Kumpal (2013) deutlich hervor, dass es neben individuellen Dispositionen vor allem Eigenschaften der Situation sind (Fragebogendesign, Eigenschaften der Interviewer, weitere Faktoren der Umfragesituation) welche bedingen, ob Befragte Antworten verzerren. Weitere Studien, die den Einfluss des Fragebogendesigns (vgl. z.B. Ayidiya & McClendon, 1990; Borger, Sikkel, & Hox, 2004; Weijters, Cabooter, & Schillewaert 2010; Yan & Tourangeau, 2008), der Eigenschaften der Interviewer (vgl. z.B. Campanelli & O'Muircheartaigh, 1999; Jäckle, Lynn, Sinibaldi, & Tipping, 2011; Samples, Woods, Davis, Rhodes, Shahane, Kaslow, 2012) und weiterer Faktoren der Umfragesituation (vgl. z.B. Aquilina, 1994; Kiesler & Sproull, 1986; Schwarz, Strack, Hippler, & Bishop, 1991), also der allgemeinen Umfragesituation auf das Antwortverhalten aufzeigen, sind zahlreich.

Auch für den Einfluss der zugrundeliegenden Motive bzw. der Motivation zur Partizipation an Befragungen auf das Antwortverhalten finden sich zahlreiche Evidenzen: So zeigt sich zum Beispiel grundsätzlich der Einfluss des Interesses am Thema einer Befragung auf Nonresponse-Raten für postalische Befragungen (Wijnen, Vermeir, & Van Kenhovea, 2007) sowie für Web Surveys (Fan & Yan, 2010; Keusch, 2012). Darüber hinaus modelliert Stocké (2006) die generelle Einstellung gegenüber Umfragen als motivationale Komponente der Beantwortung von Umfragen und findet auch diesbezüglich einen Einfluss auf Nonresponse-Raten. Groves, Cialdini und Couper (1992) wiederum liefern Evidenz, wie die Einstellung gegenüber Umfragen auf die grundsätzliche Bereitschaft, an Umfragen teilzunehmen einwirkt. Über die allgemeine Teilnahmebereitschaft und Nonresponse-Raten hinaus belegen Rogelberg, Fisher, Maynard, Hakel und Horvath (2001) den Einfluss der Einstellung gegenüber Umfragen auf das Befolgen von Anweisungen, das nachträglich erteilte Einverständnis zur Teilnahme an Follow-Up Befragungen, das Einverständnis zur Teilnahme an anderen Befragungen, die Reaktionsrate auf die Anfrage zur Teilnahme an Befragungen und ebenfalls auf Nonresponse-Raten.

Der Einfluss der chronischen Zugänglichkeit von Informationen, Einstellungen, Assoziationen etc. auf das (Antwort-)Verhalten zuletzt ist ebenfalls zahlreich erforscht. So analysieren Andersen und Mayerl (2017) beispielsweise die Rolle der Zugänglichkeit von Einstellungen bezüglich des Objekts der Befragung für das Auftreten sozial erwünschten Antwortverhaltens. Und

auch für den Question-Behaviour Effekt, welcher beschreibt, dass Befragte, die zu einem Verhalten befragt werden, dieses Verhalten anpassen, gibt es Befunde, dass dieser durch die Zugänglichkeit von Einstellungen beeinflusst wird (Wilding, et al., 2016). Ebenfalls lässt sich der Einfluss der chronischen Zugänglichkeit von Einstellungen auf das Auftreten von Zustimmungstendenzen (Mayerl, 2013), Fragereihenfolgeneffekte (Chen, 2010), das Antwortverhalten bei dichotomen Antwortoptionen (Bar-Hillel, Peer, & Acquisti, 2014), die Evaluation des Objektes der Befragung (Young & Fazio, 2013), sowie schließlich die Konstruktvalidität von Einstellungsskalen (LeBel, 2010) zeigen.

Basierend auf den hier diskutierten Theorien und Evidenzen lässt sich schließlich bezüglich des Antwortverhaltens in Umfragen die folgende These ableiten:

> *These 1: Antworten in Umfragen basieren auf chronisch und situativ zugänglichen Informationen und werden beeinflusst durch Definitionen und Bewertungen der Umfragesituation und der Frage sowie den Zielen bzw. der Motivation Befragter für die Umfrage*

Nach der Diskussion generischer Faktoren mit Einfluss auf das Antwortverhalten soll als nächstes der konkrete Prozess der Antwortgenerierung beleuchtet werden. Diesbezüglich formulieren Tourangeau und Kollegen (2000), Strack und Martin (1987), Sudman und Kollegen (1996) sowie Cannell und Kollegen (1981) innerhalb ihrer jeweiligen Modelle, dass Fragen zunächst auf kognitiver Ebene verstanden werden müssen, sprich: Allgemein ein kognitiver Verarbeitungsprozess hinsichtlich der Prozessierung der Frage stattfindet. Daneben erfolgt zusätzlich ein kognitiver Prozess der Antwortgenerierung, welcher qua der in den Modellen vorgesehenen Editierung der Antwort unterschiedlich elaboriert stattfindet. Neben einer solchen zusätzlichen Elaboration beschreibt Schwartz (1991) ebenfalls eine verminderte Elaboration, welche immer dann auftritt, wenn Befragte nicht jeden der Schritte der Antwortgenerierung vollständig durchlaufen. Entsprechend dieses variablen Grades an Elaboration werden zahlreiche Implikationen zugrundeliegender kognitiver Verarbeitungsprozesse diskutiert. So werden innerhalb des Satisficing-Optimizing Modells (Krosnick, 1991; Krosnick, 1999a; Krosnick & Fabrigar, 1997) das Satisficing als Pol geringer Elaboration und das Optimizing als Pol hoher Elaboration entlang eines Elaborations-Kontinuums differenziert. Auch mit der Postulierung verschiedener Verhaltensmodelle explizit für spontanes, unelaboriertes Prozessieren (Fazio, 1986; Fazio, 1990) sowie für geplantes, elaboriertes Verhalten (Ajzen, 1991) wird den differenten kognitiven Strategien, welche mit verschiedenen Graden an Elaboration einhergehen, Rechnung getragen. Die Integration verschieden elaborierter kognitiver Prozesse wird ebenfalls durch die vorgestellten

dualen Prozessmodelle wie das MODE-Modell (Fazio, 1990) und das RIM (Krishna & Strack, 2017) geleistet. Und auch innerhalb des MFS wird postuliert, dass Individuen Frames, Skripte und konkrete Handlungen jeweils innerhalb eines automatisch-spontanen (unelaborierten) oder eines reflexiv-kalkulierenden (elaborierten) Modus selektieren (vgl. Esser, 2001; Esser, 2006a; Esser, 2006b; Kroneberg, 2005; Kroneberg, 2007; Kroneberg, 2011). Für einen solchen variablen Grad an Antwortelaboration innerhalb der zuvor beschriebenen Modelle finden sich ebenfalls zahlreiche empirische Belege (vgl. z.B. Andersen & Mayerl, 2019; Mayerl, 2009; Mayerl, 2013; Mayerl & Giehl, 2018), so dass als zweite allgemeingültige These zusammengefasst werden kann:

> *These 2: Die Generierung von Antworten in Umfragen hat einen individuell und situativ variablen Grad an Elaboration, welche durch den kognitiven Aufwand beschrieben ist der aufgebracht wird, um eine Antwort zu generieren. Je nach Grad der Elaboration unterliegt der Prozess der Antwortgenerierung verschiedenen kognitiven Prozessen.*

Über die Determinanten und den kognitiven Prozess der Antwortgenerierung hinaus beschreiben bzw. implizieren die in diesem Kapitel erörterten Modelle ebenfalls die Möglichkeit, dass Antworten in Umfragen unwahr, verzerrt, editiert etc. sein können, sprich: Der gegebene Antwort-Wert nicht dem wahren Wert entspricht. So beschreiben Tourangeau und Kollegen (2000) beispielsweise zum einen die Notwendigkeit der Editierung einer Antwort bezüglich der Passung hinsichtlich eines vorgegebenen Antwortformates, was quasi einer bewusst-notwendigen Verzerrung entspricht. Zum anderen können verzerrte, d.h. unwahre Antworten gemäß des Modells ebenfalls unbewusst gegeben werden, wenn die Frage nicht korrekt verstanden wurde, wenn nicht alle relevanten Informationen zur Antwortgenerierung herangezogen werden, wenn situative Hinweisreize zur Evaluation von Relevanz und Vollständigkeit mitverwendet werden oder wenn der Erinnerungsprozess oder die Evaluation relevanter Informationen aufgrund geringer Zugänglichkeit fehlerhaft verläuft, was ebenfalls innerhalb der Modelle von Strack und Martin (1987) sowie Sudman und Kollegen (1996) und im Falle der Antwortgenerierung auf Basis nicht aller relevanter Informationen bzw. situativ verfügbarer Informationen innerhalb des Satisficing*s* (Krosnick, 1991; Krosnick, 1999a; Krosnick & Fabrigar, 1997) beschrieben wird. Die Modelle nach Strack und Martin (1987) sowie Sudman und Kollegen (1996) sehen neben solchen unbewussten Verzerrungen und der notwendigen, durch die Antwortkategorie bedingten Editierung ebenfalls eine bewusste Verzerrung einer formal korrekten Antwort in dem Fall vor, wenn diese als sozial unerwünscht wahrgenommen wird und Befragte sich durch die Editierung ihrer Antwort in einem besseren Licht darstellen können (sozial erwünschtes

Antwortverhalten), so die Umfragesituation nicht als privat wahrgenommen wird. Cannell und Kollegen (1981) beschreiben diese zuletzt genannte bewusste Verzerrung allgemeiner in dem Sinne, dass Befragte eine formal vorgesehene Antwort unter Umständen editieren, so das Geben der vorgesehenen Antwort mit persönlichen Zielen, Motiven etc. kollidiert.

Dass Verzerrungen somit sowohl bewusst, als auch unbewusst auftreten können, wird weiterhin durch die hier vorgestellten impliziten Verhaltensmodelle unterstützt. So ist es nach Fazio (1986; 1990) sowie nach Ajzen (1991) und Fishbein & Ajzen (2010) die chronische Einstellungszugänglichkeit welche darüber entscheidet, ob alle für die Urteilsbildung relevanten Einstellungen/Informationen berücksichtigt werden oder Antworten unter Umständen ebenfalls auf situativ verfügbaren Informationen und Heuristiken basieren und somit potenziell verzerrt sind. Dies gilt ebenfalls innerhalb des MFS (vgl. Esser, 2001; Esser, 2006a; Esser, 2006b; Kroneberg, 2005; Kroneberg, 2007; Kroneberg, 2011) sowie des impulsiven Systems des RIM (Krishna & Strack, 2017), wobei das erste Modell im Kern nicht die Zugänglichkeit von Einstellungen, sondern von Informationen und das zweite Modell von Assoziationen im Fokus hat. Bewusste Verzerrungen treten in den besagten Modellen hingegen nicht zwangsläufig bei geringer chronischer Zugänglichkeit von Einstellungen, Informationen oder Assoziationen auf, sondern sind das Ergebnis eines zweckrationalen und elaborierten Prozesses in dem Sinne, dass Antworten bewusst editiert werden, so formal vorgesehene Antworten mit vorhandenen Zielen und Motiven Befragter kollidieren, was innerhalb des MFS z. B. dann gegeben sein kann, wenn Befragte einen indifferenten oder gar ablehnenden Frame gegenüber Umfragen selektieren.

Empirische Belege für das allgemeine Vorhandensein von Antwortverzerrungen sind nicht nur zahlreich – mit der Methoden- und Umfrageforschung beschäftigt sich sogar ein gesamter Forschungszweig mit dieser Thematik (vgl. z.B. Bogner & Landrock, 2014), so dass als dritte These geschlossen werden kann:

> *These 3: Gegebene Antworten in Umfragen sind nicht ausschließlich wahr, sondern können auch in dem Sinne der Unwahrheit entsprechen, als dass sie nicht kohärent mit individuellen Dispositionen bzw. der empirisch überprüfbaren Realität sind. Antworten in Umfragen verfügen somit über einen variablen Grad an Verzerrung, wobei Verzerrungen unbewusst oder bewusst eintreten können. Unbewusste Verzerrungen resultieren, wenn Antworten nicht auf der Basis optimaler, sondern situativ verfügbarer Informationen und Assoziationen oder Heuristiken basieren. Bewusste Verzerrungen resultieren, wenn vorgesehene Antworten aufgrund der Kollision mit persönlichen Motiven editiert werden.*

Wird entsprechend angenommen, dass Verzerrungen sowohl bewusst, als auch unbewusst – oder wie innerhalb des RIM beschreibbar, reflexiv oder impulsiv, sprich unelaboriert und spontan oder elaboriert und überlegt auftreten können, so kommt den Bedingungsfaktoren des Grades der Antwortelaboration in der Umfrageforschung eine besondere Bedeutung zu. So wird innerhalb der expliziten, klassischen Modelle des Befragtenverhaltens hinsichtlich variabler Elaboration beschrieben, dass Befragte nicht zwangsläufig alle relevanten Informationen für die Antwortgenerierung prozessieren (Schwarz, 1991), was entsprechend einem verminderten Grad an Antwortelaboration entspricht und wie bereits beschrieben als Satisficing bezeichnet werden kann (Krosnick, 1991). Nach Cannell et al. (1981) ist dies beispielsweise darin begründet, dass entweder nicht alle relevanten Informationen vorliegen oder keine Motivation besteht, alle relevanten Informationen abzurufen. Als Determinanten für das Auftreten einer Antwortstrategie des Optimizing nennt Krosnick (1991) im Speziellen: a) die Komplexität der zu lösenden Aufgabe; b) die (kognitiven) Fähigkeiten Befragter, um Optimizing zu betreiben; und c) die Motivation Befragter, um Optimizing zu betreiben. Auch Fazio (1990) beschreibt innerhalb seines MODE-Modells variable Grade an Elaboration entlang eines Kontinuums mit einem automatisch-spontanen Pfad als idealtypischer Pol geringer Elaboration und einem geplant-kontrollierten Pfad als idealtypischer Pol hoher Elaboration, wobei sowohl die Motivation, als auch die Möglichkeit zum elaborierten Prozessieren als notwendige und hinreichende Bedingung für dieses gilt. Der Faktor „Möglichkeit“ ist in diesem Sinne relativ allgemein gefasst und kann sich z. B. auf vorhandenen Zeitdruck (keine Zeit zum elaborieren) oder fehlende Ressourcen (es fehlt z.B. Wissen, um zu elaborieren) beziehen, während sich der Faktor „Motivation“ vor allem auf die bereits zuvor beschriebene „Angst vor Invalidität“ bezieht. Auch innerhalb des diskutierten MFS (vgl. Esser, 2001; Esser, 2006a; Esser, 2006b; Kroneberg, 2005; Kroneberg, 2007; Kroneberg, 2011) ist eine Erklärungskomponente gegeben, unter welchen Bedingungen Individuen Frames, Skripte und Handlungsalternativen automatisch-spontan oder reflexiv-kalkulierend selektieren. Demnach sind Selektionen im reflexiv-kalkulierenden Modus umso wahrscheinlicher, a) je höher die Opportunitäten und die Motivation zum bewussten Reflektieren sind; b) je geringer der Aufwand für eine bewusste Reflexion ist; und c) je geringer die Aktivierung eines für die Situation relevanten mentalen Schemas ist, welches auch ohne bewusste Reflexion adäquat Handlungsweisend ist.

Die hier genannten Faktoren elaborierten Prozessierens lassen sich, unterschiedliche Niveaus an Verallgemeinerung berücksichtigend, im weitesten Sinne entsprechend Mayerls (2009) generischem, dualen Prozessmodell der Einstellungs-Verhaltensforschung, welches eine Aggre-

gation der Schnittmengen gängiger dualer Prozessmodelle darstellt, als Interaktion der allgemeinen Motivation sowie der allgemeine Möglichkeit zum Elaborieren benennen, wobei es besagte Interaktion ist, welche als notwendige und hinreichende Bedingung für elaboriertes Prozessieren angesehen werden kann. Nur wenn Befragte motiviert sind, eine elaborierte Antwort zu prozessieren und hierzu die Möglichkeit haben, werden sie elaborieren. In allen anderen Fällen bleibt die Antwort unelaboriert prozessiert.

Die Motivation und Möglichkeit als Bedingungsfaktoren für Elaboration bzw. für überlegtes Handeln ist im Kontext der Forschung rund um duale Prozessmodelle der Einstellungs-Verhaltensforschung vielfach untersucht und bestätigt (vgl. z.B. Krosnick et al., 1996; Mayerl, 2009; Mayerl, 2013; Schuette & Fazio, 1995), so dass als vierte These abgeleitet werden kann:

> *These 4: Dem Prozess der Antwortgenerierung liegt ein variabler Grad an Elaboration zugrunde, welcher sich in differenten kognitiven Prozessen äußert. Die Antwortelaboration ist bedingt durch die individuelle und situative Motivation und Möglichkeit zum Elaborieren, wobei es sich hierbei um eine interaktive Beziehung handelt. Nur wenn die Motivation und die Möglichkeit gegeben ist, werden Antworten elaboriert prozessiert. In allen anderen Fällen verläuft der Prozess der Antwortgenerierung unelaboriert.*

Über den kognitiven Prozesspfad hinaus, innerhalb welchem Befragte Antworten prozessieren, wird in den zuvor beschriebenen Modellen aufgeführt, dass von Seiten der Befragten an vielen Punkten der Antwortgenerierung bewusste und unbewusste Vorannahmen (unter welchen sich in einem weiteren Sinne ebenfalls Beliefs und Primings etc. subsummieren lassen) herrschen, welche sich beispielsweise auf die Interpretation der Frage (Ajzen, 1991; Cannell et al., 1981; Fishbein & Ajzen, 2010; Strack & Martin, 1987; Sudman et al., 1996), die Bewertung von Informationen als Relevant für die Urteilsbildung (Cannell et al., 1981; Krosnick, 1991; Tourangeau et al., 2000), die Bewertung von Handlungsalternativen als situativ angemessen (Ajzen, 1991; Fazio, 1986; Fazio, 1990; Fishbein & Ajzen, 2010) sowie auf die in These 1 bereits beschriebenen Ziele und Motive Befragter für die Befragung (Ajzen, 1991; Cannell et al., 1981; Fazio, 1986; Fazio, 1990; Fishbein & Ajzen, 2010) bzw. allgemeiner die Definitionen und Bewertungen der Umfragesituation und der Frage auswirken.

Solche Vorannahmen werden durch das MFS (vgl. Esser, 2001; Esser, 2006a; Esser, 2006b; Kroneberg, 2005; Kroneberg, 2007; Kroneberg, 2011) innerhalb verhaltenswirksamer Frames modelliert und gebündelt, die in kognitiven Entscheidungssituationen, welche innerhalb des

Modells idealtypisch als sequenziell beschrieben werden, als erstes, sprich noch vor Handlungsskript und konkreter Handlungsalternative aktiviert werden. Automatisch-spontan selektierte Frames werden diesbezüglich unhinterfragt auf Basis von beispielsweise Gewohnheiten und (situativ wirksamen) Normen aktiviert, während der reflexiv-kalkulierten Selektion eines Frames ein systematischer Vergleich verschiedener Alternativen vorausgeht. Die individuelle Frame-Selektion entspricht somit der individuellen und situativen Definition der Situation, bezieht sich hierin auf Situationsmerkmale oder Situationstypen und grenzt damit für das Individuum ein, welche Situation inklusive daraus folgender angemessener Handlungsschemata vorliegt. Entsprechend dieser Eingrenzung von Frames im Kontext des Befragtenverhaltens lassen sich idealtypisch zwei voneinander abgrenzbare Frames definieren, welche sich entweder durch eine positive Bewertung und Unterstützung (durch wahre Antworten) von Befragungen (unterstützender Frame) oder durch eine indifferente oder gar ablehnende Haltung und fehlende Unterstützung (durch eine indifferente Einstellung gegenüber der Gabe wahrer Antworten) gegenüber Befragungen auszeichnen (indifferenter Frame). Es kann somit plausibel angenommen werden, dass der selektierte Frame mit der in These 1 beschriebenen Definition und Bewertung der Umfragesituation sowie den Zielen und Motiven Befragter für die Befragung assoziiert ist.

Empirisch wurde eine solche Abgrenzung idealtypischer Frames in Befragungssituationen von Stocké (2004, S. 306) als kooperative Befragtenrolle (unterstützender Frame) auf der einen Seite und konformistische Befragtenrolle (indifferenter Frame) auf der anderen Seite modelliert und hinsichtlich der Wirkung auf das Befragtenverhalten untersucht. Kooperative Befragte zeichnen sich diesbezüglich durch eine tendenziell bedingungslose Unterstützung der Befragung qua Gabe wahrer Antworten aus und werden mit einer positiven generalisierten Einstellung gegenüber Umfragen assoziiert bzw. durch diese approximiert. Konformistische Befragte hingegen zeichnen sich durch ein an der sozialen Realität der Interviewsituation und der Situationsdefinition orientierten Befragtenverhalten aus, welches an möglichst vorteilhaften Bewertungsreaktionen orientiert und mit einer indifferenten bis negativen generalisierten Einstellung gegenüber Umfragen assoziiert ist. Entsprechende Evidenzen der Verhaltenswirksamkeit der verschiedenen Frames bzw. Rollen im Kontext von Umfragesituationen finden sich z.B. hinsichtlich des Auftretens sozial erwünschten Antwortverhaltens (Stocké, 2004), von Item Nonresponse (Stocké, 2006; Stocké & Becker, 2004) sowie der generellen Teilnahmebereitschaft an Umfragen (Heerwegh & Loosveldt, 2009). Der Einfluss der generalisierten Einstellung von Personen gegenüber Umfragen auf das Verhalten von Personen in (potentiellen) Befragungssituationen ist ebenfalls umfassend erforscht. Nach Goyder (1986) beispielsweise neigen Personen mit einer positiven Einstellung gegenüber Umfragen stärker zur spontanen Teilnahme an

Umfragen als Personen mit einer negativen Einstellung. Äquivalent hierzu erweisen sich nach Erbslöh und Koch (1988) sowie Stinchcombe, Jones und Sheatsley (1981) Personen, die die Teilnahme an einer Befragung zunächst verweigert haben und erst in einem zweiten Versuch zur Teilnahme bewegt werden können, als kritischer gegenüber Umfragen eingestellt als Personen, die sofort und spontan an Befragungen teilnehmen. Auch zeigt sich, dass sich Teilnehmer einer postalischen Befragung eher an Fragebogenanweisungen halten, Fragebögen schneller ausgefüllt zurückschicken und eine geringere Item-Nonresponse aufweisen, wenn sie Umfragen im Allgemeinen eher positiv bewerten (Rogelberg et. al., 2001). Entsprechend kann als weitere These formuliert werden:

> *These 5: Das Befragtenverhalten sowie konkret formulierte Antworten in Umfragen stehen in Abhängigkeit von der individuell zugrundeliegenden Befragtenrolle, wobei die Selektion und Aktivierung einer entsprechenden Rolle dem Prozess des Framings entspricht. Idealtypisch lassen sich qualitativ eine kooperative, unterstützende Befragtenrolle sowie eine der Befragung indifferent oder ablehnend gegenüberstehende Befragtenrolle differenzieren.*

Werden diese allgemeingültigen Thesen der Antwortgenerierung nun generisch verdichtet, so ergibt sich hieraus als vereinfachte Beschreibung des Befragtenverhaltens, dass Befragte innerhalb einer Umfragesituation eine spezifische Befragtenrolle (unterstützender/kooperativer Frame oder indifferenter/konformistischer Frame) einnehmen, welche Ziele und Motive Befragter für die Befragung, die Relevanzbewertung von Informationen (Einstellungen, Assoziationen etc.) während der Urteilsbildung, die selektierte Antwortstrategie etc. bedingen. Bei Wahrnehmung einer konkreten Frage entscheidet sich weiterhin der Grad an Elaboration, mit welcher die Antwort prozessiert wird, bedingt sowohl durch die Motivation, als auch die Möglichkeit zum elaborieren. An dieser Stelle wird nicht angenommen, dass die Befragtenrolle einen Einfluss auf die Motivation zur Elaboration aufweist da es nicht plausibel ist, dass kooperative Befragte mit einer festen Meinung zu einem Thema diese zwangsläufig in Umfragesituationen stets neu überdenken. Spontan prozessierte Antworten sind diesbezüglich im Wesentlichen abhängig von den zum Zeitpunkt der Antwortgenerierung zugänglichen Informationen, wobei diese relevant für die zugrundeliegende Fragestellung (kognitiv chronisch zugänglich), jedoch auch lediglich situativ aktiviert (situativ akut zugänglich) und irrelevant für die zugrundeliegende Fragestellung sein können. Auch elaboriert formulierte Antworten basieren zunächst auf kognitiv, jedoch auch auf situativ zugänglichen Informationen, auch wenn Befragte die Möglichkeit haben, auf Basis einer Evaluation der Relevanz der Informationen weitere, für die Urteilsbildung benötigte Informationen zu aktivieren oder zu elaborieren, so sie motiviert

zum elaborieren sind und diesbezüglich die Möglichkeit haben. Für elaboriert prozessierte Antworten ist es ebenfalls möglich, dass diese zusätzlich editiert, d.h. bewusst verzerrt werden, so die vorgesehene Antwort mit Zielen und Motiven Befragter für die Befragung kollidieren.

Entsprechend dieser Betrachtung stellt sich im Kontext dieser Arbeit die Frage, wie diese Prozesse auf das Auftreten von FRE einwirken. Entsprechend wird die Logik dieser spezifischen Form von Antwortverzerrung im nachfolgenden Kapitel beleuchtet.

2.2 Fragereihenfolgeneffekte in der Umfrageforschung

Die Generierung von Antworten in Umfragen ist, wie im vorangegangenen Kapitel diskutiert, nicht ausschließlich von z. B. individuellen Einstellungen gegenüber dem Objekt der Befragung, sondern ebenfalls von kognitiven und situativen Faktoren abhängig, welche zu Antwortverzerrungen führen können. Solche Faktoren ergeben sich beispielsweise aus dem spezifischen Inhalt einer Frage, der präsentierten Antwortskala oder gar der vorangegangenen Frage selbst, wie im vorangegangenen Kapitel aufgezeigt. Hat die vorangegangene Frage einen solchen Einfluss auf eine nachfolgende Frage, so wird dies als Fragereihenfolgeneffekt (kurz: „FRE“) bezeichnet, welcher im Fokus dieser Arbeit steht. Entsprechend ist es das Ziel dieses Kapitels, die zugrundeliegenden Mechanismen und Determinanten des Auftretens von FRE zu erörtern, um diese nachfolgend im Kontext des Prozesses der Antwortgenerierung zwecks Modell- und Hypothesenbildung einzubetten.

Diesbezüglich wird im Folgenden eine allgemeine Einführung in den Themenkomplex der systematischen Messfehler und der Antwortverzerrungen auf theoretischer Ebene geleistet (vgl. Kapitel 2.2.1), auf welcher aufbauend die Logik von FRE diskutiert wird (vgl. Kapitel 2.2.2). Entsprechend dieser zugrundeliegenden Logik werden schließlich als mögliche, durch Kapitel 2.1 fundierte Einflussfaktoren für FRE der Grad an Antwortelaboration, die kognitive Zugänglichkeit von Einstellungen, Informationen, Assoziationen etc. sowie die selektierte Befragtenrolle (vgl. Kapitel 2.2.3) aufgezeigt, woraus schließlich in Kapitel 2.3 die Entwicklung eines Modells zur Erklärung von FRE sowie die Ableitung von Hypothesen bezüglich des Auftretens von FRE hervorgeht.

2.2.1 Antwortverzerrungen – Eine Einführung

Wie bereits beschrieben sind sozialwissenschaftliche Interviews keine neutralen Situationen, sondern Prozesse, die unter bestimmten Umständen gewisse Reaktivitäten bei Befragten hervorrufen können und welche dazu führen, dass besagte Befragte nicht wahrheitsgemäß (ob bewusst oder unbewusst) antworten (Stocké, 2004). Weichen auf diese Weise die empirisch beobachteten Messwerte systematisch von den wahren Werten ab, wird in einem weiteren Sinne von Antwortverzerrungen, systematischen Messfehlern oder im englischen von response effects oder response bias gesprochen (Bogner & Landrock, 2014). Solche systematischen Messfehler bzw. Antwortverzerrungen können, den gängigen Lehrbüchern zufolge, allgemein in Interviewer-, Befragten- und Instrumenteneffekte (Schnell, Hill, & Esser, 2018; Groves, 1989) sowie ergänzend in Situationseffekte (Eirmbter & Jacob, 1996) unterteilt werden. Quer dazu kann eine weitere Verortung von Antwortverzerrungen in Anlehnung an die im vorangegangenen Kapitel diskutierten Modelle des Befragtenverhaltens in bewusstes und elaboriertes Editieren sowie in unbewusstes, automatisches Verzerren stattfinden. Da eine solche Zuordnung jedoch einer genaueren Betrachtung der zugrundeliegenden kognitiven Prozesse einer Antwortverzerrung Bedarf, wird diese Unterscheidung lediglich spezifisch für FRE geleistet, während die allgemeine Betrachtungsebene der Effekte in diesem Kapitel zwecks Einführung in die Thematik vorerst auf Basis der Lehrbuch-Systematisierung erfolgt.

Instrumentenbedingte Antwortverzerrungen

Beginnend mit den instrumentenbedingten Effekten werden unter diesen alle jene Verzerrungen diskutiert, welche durch Eigenschaften des Instrumentes der Datenerhebung selbst, also dessen Design, Layout, Form oder Inhalt, hervorgerufen werden. Zu den Eigenschaften des Instrumentes, welche einen Einfluss auf das Antwortverhalten von Befragten haben können, gehören unter anderem die Verwendung bestimmter Formulierungen bzw. Frageninhalte, die Position der Fragen im Fragebogen, die Anzahl und Reihenfolge der vorgegebenen Antwortkategorien etc. (Bradburn, 1983; Groves, 1989). In der Regel sind solche instrumentenbedingte Antwortverzerrungen jedoch nicht unabhängig von Eigenschaften der Befragten oder genauer, der Tendenz von Befragten, Fragen nicht nach deren Inhalt, sondern auf der Grundlage bestimmter Heuristiken, situativer Hinweisreize oder anderer kognitiver Prozesse zu beantworten. Genaugenommen lassen sich demnach entsprechende Instrumenteneffekte nicht unabhängig von Befragteneffekten analysieren (Eirmbter & Jacob, 1996, S. 91).

Ein Beispiel für Frageninhalte, die das Antwortverhalten nicht unabhängig von Dispositionen der Befragten beeinflussen, ergibt sich im Kontext der Betrachtung des Effektes sozialer Erwünschtheit. Dieser beschreibt die Tendenz von Befragten, in Befragungen dem zu entsprechen, was die Interviewer oder andere beteiligte Personen vermeintlich von Befragten erwarten, wobei diese Erwartungsbildung vor allem auf sozialen Normen basiert. Dieser Effekt tritt unter anderem bei Fragen zu Themen auf, welche von befragten Personen als sensibel gewertet werden wie z.B. Fragen zum Sexualverhalten, zur Einstellung gegenüber Abtreibung oder zu Drogenmissbrauch (Kreuter, Presser, & Tourangeau, 2008). Aber nicht nur der Inhalt bestimmter Fragen, sondern auch die Reihenfolge, in der solche Fragen gestellt werden, kann einen Einfluss auf die Interpretation von Fragen haben und Antworten verzerren, wenn vorangehende Fragen sich auf den Kontext nachfolgender Fragen auswirken, sprich, ein FRE vorliegt (Krosnick & Presser, 2010; Tourangeau et al., 2000). Auf FRE, welche zentral für diese Arbeit sind, wird in dem nachfolgenden Kapitel noch vertiefend eingegangen. Die Reihenfolge der Antwortvorgaben in einer Liste oder Ratingskala zuletzt kann einen Einfluss darauf haben, welche Antwortkategorie Befragte bevorzugt selektieren. So wählen Befragte bei visueller Präsentation von Antwortkategorien (wie etwa bei Onlineumfragen) eher die zuerst präsentierten Alternativen, während Befragte bei mündlich vorgebrachten Nennungen der Antwortalternativen (wie etwa bei Telefoninterviews) eher aus den zuletzt genannten Alternativen wählen (Bishop & Smith, 2001; Krosnick, 1991).

Situationsbedingte Antwortverzerrungen

Sind es hingegen spezifische Eigenschaften der Interviewsituation, welche zu einer Verzerrung der Antworten von Befragten führen, wird von situationsbedingten Antwortverzerrungen gesprochen. Hierzu gehören unter anderem Eigenschaften der Interviewer sowie die Anwesenheit dritter Personen (Bogner & Landrock, 2014). So beeinflussen zum einen Eigenschaften von Interviewenden vor allem dann das Antwortverhalten von Befragten, wenn Charakteristika, welche den Interviewenden gut erkennbar zugeschrieben werden können (wie etwa Geschlecht oder Ethnie), Gegenstand der Befragung sind (Groves, 1989). Zum anderen können aber auch Verhaltensweisen der Interviewenden zu Antwortverzerrungen führen (Mangione, Fowler, & Louis, 1992). Als besonders anfällig für Interviewereffekte gelten dabei schwierige bzw. heikle Fragen, Meinungsfragen und offene Fragen, wobei wiederum Merkmale wie das Geschlecht und die Bildung der Interviewer die Beantwortung schwieriger bzw. heikler Fragen insofern beeinflussen können, als dass z.B. die Frage nach dem Einkommen eher bei weiblichen Interviewern und Interviewern mit höherer Bildung verweigert wird (Haunberger, 2006).

Mit dem Geschlecht und der Bildung der Interviewer werden zwei verschiedene Typen von Interviewereigenschaften angesprochen: Sichtbare und nicht sichtbare Eigenschaften. Sichtbare Eigenschaften werden von Befragten direkt erfasst und können einen unmittelbaren Einfluss auf die Interviewsituation haben. Zu diesen gehören unter anderem Merkmale wie das Geschlecht, die Ethnie, das Auftreten, die Körpersprache und das verbale Verhalten. Nicht sichtbare Eigenschaften hingegen treten während des Prozesses des Interviews in der Interaktion zwischen Interviewenden und Befragten auf und beeinflussen Befragte eher indirekt, in dem sie das Verhalten der Interviewer prägen. Hierzu gehören stabile Persönlichkeitsmerkmale der Interviewer wie deren Extrovertiertheit, aber auch rollenbezogene Eigenschaften der Interviewer wie deren Einstellung gegenüber dem Objekt der Befragung oder den Befragten selbst (Glantz & Michael, 2014).

Hinsichtlich der sichtbaren Eigenschaften zeigen Untersuchungen, dass die ethnische Herkunft etwa kaum Auswirkungen auf die Teilnahmebereitschaft, hingegen aber auf das Antwortverhalten selbst aufweist (Schaeffer, Dykema, & Maynard, 2010; Weisberg, 2005). Ähnliche Befunde zeigen sich hinsichtlich des Geschlechts der Befragten (Huddy, et al., 1997). Aus diesen Befunden leitet sich die sogenannte Matching-Hypothese ab welche besagt, dass Interviewer, die die gleiche ethnische Herkunft bzw. das gleiche Geschlecht wie Befragte haben, validere Antworten von diesen erhalten (Schaeffer et al., 2010). Gleichzeitig kann ein Matching auch dazu führen, dass Befragte ihre Antworten an die wahrgenommenen sozialen Erwartungen der Interviewer anpassen (Glantz & Michael, 2014).

Neben Eigenschaften des Interviewers kann auch die Anwesenheit dritter Personen das Antwortverhalten von Befragten beeinflussen, wobei dritte Personen z. B. Partner, Kinder, andere Verwandte, aber auch Freunde und Nachbarn sein können (Hartmann, 1994).Hinsichtlich solcher Anwesenheitseffekte zeigen Studien, dass sich die Anwesenheit einer dritten Person, z.B. des Ehepartners, positiv auf die Beantwortung von Faktenfragen auswirkt, da diese kontrollieren können, ob eine zutreffende Aussage gemacht wurde (Lander, 2000) bzw. den Befragten helfen können, sich zu erinnern (Hartmann, 1994). Daneben werden jedoch auch Auswirkungen auf die Validität der erhobenen Daten durch veränderte Verhaltenserwartungen aller Anwesenden diskutiert, z.B. da die Anwesenheit dritter Effekte sozialer Erwünschtheit erzeugen kann. So sind z.B. deutliche Unterschiede im Antwortverhalten von Befragten festzustellen, je nachdem, ob ein Ehepartner anwesend ist oder nicht (Lander, 2000).

Befragtenbedingte Verzerrungen

Zu den befragtenbedingten Verzerrungen gehören alle jene Effekte, welche auf spezifische Tendenzen der Befragten zurückzuführen sind. Hierzu gehören zum Beispiel Tendenzen zur Mitte, Milde und Härte (Greenleaf, 1992), die Tendenz zur Zustimmung (Krosnick, 1999b) oder eine Verzerrung der Antwort aufgrund des Bedürfnisses nach sozialer Anerkennung (Kreuter et al., 2008).

Von einer Tendenz zur Mitte ist dann die Rede, wenn Befragte unabhängig vom Frageninhalt oder davon, ob dies tatsächlich ihrer Einstellung entspricht, systematisch, vor allem in schriftlichen, standardisierten Befragungen die mittlere Antwortkategorie einer Ratingskala wählen. Äquivalent hierzu wird mit der Tendenz zur Milde bzw. Härte die Neigung von Befragten beschrieben, unabhängig vom Inhalt der Fragen die jeweilige Extremantwort einer Ratingskala anzugeben (Bogner & Landrock, 2014). Diesbezüglich zeigen Studien, dass die Tendenz zu solchen Extremantworten über die Zeit bzw. Länge einer Befragung stabil ist und dass besonders ältere Menschen sowie Menschen mit geringer Bildung und Einkommen hierzu neigen (Greenleaf, 1992).

Bei der sogenannten Zustimmungstendenz, auch als Akquieszenzeffekt bezeichnet, stimmen Befragte Aussagen unabhängig von deren Inhalt zu (Krosnick, 1999b), was systematische Verzerrungen der erhobenen Daten nach sich ziehen kann (Billiet & Davidov, 2008). Formal betrachtet kann Akquieszenz auch eine instrumentenbedingte Verzerrung darstellen, da sie sich aus dem Stimulus selbst ergeben kann: so erhöhen z.B. widersprüchlich formulierte Fragen (McBride & Moran, 1967) oder Fragen, die aus sehr vielen Wörtern bestehen (Trott & Jackson, 1967) die Wahrscheinlichkeit des Auftretens von Akquieszenzeffekten. Hingegen zeigt sich ebenfalls, dass Akquieszenz zum einen rein aus Höflichkeit entstehen kann, da es angenehmer ist, zuzustimmen als zu widersprechen (Krosnick, 1999b), zum anderen aber auch die Folge einer Antwortstrategie sein kann, in welcher mehr oder weniger bewusst auf anstrengende, kognitive Prozesse verzichtet wird (Bogner & Landrock, 2014). Empirische Studien zeigen zudem, dass Befragte mit geringer Intelligenz bzw. kognitiven Fähigkeiten (Krosnick, 1999b) oder geringer formaler Bildung (Rammstedt, Goldberg, & Borg, 2010) besonders zur Akquieszenz neigen.

Als Antwortverzerrung aufgrund des Bedürfnisses nach sozialer Anerkennung, auch als sozial erwünschtes Antwortverhalten beschrieben, wird die Tendenz von Befragten verstanden, eine überwiegend positive Beschreibung der eigenen Person abzugeben (Paulhus, 2002), was durch

eine übertriebene Nennung von als erwünscht wahrgenommenem Verhalten oder der untertriebenen Nennung von als unerwünscht wahrgenommenem Verhalten geschehen kann (Kreuter et al., 2008), wie im Kontext der instrumentenbedingten Antwortverzerrung bereits beschrieben. Persönlichkeitstheoretisch kann das Auftreten von sozial erwünschtem Antwortverhalten auf das in Befragten unterschiedlich stark ausgeprägte Bedürfnis nach sozialer Anerkennung zurückgeführt werden, wohingegen ein kognitiver Erklärungsansatz das Auftreten von sozial erwünschtem Antwortverhalten auf das Vermeiden von Konsequenzen durch z.B. den Interviewer zurück führt (Stocké, 2004).

Somit ist insgesamt ersichtlich, dass Kontext, soziale Situation, Befragtentendenzen u.v.m. einen umfangreichen und lediglich schwer zu überblickenden Einfluss auf das Antwortverhalten von Befragten haben und somit zu Antwortverzerrungen führen können. Deren Erforschung kommt entsprechend eine erhebliche Bedeutung zu, weshalb eine detaillierte Diskussion der Determinanten und Kontextfaktoren der für diese Arbeit zentralen FRE folgt.

2.2.2 Die Logik von Fragereihenfolgeneffekten

Die Bildung einer Meinung, einer Antwort oder eines Urteils in Umfragen ist stets von gewissen Kontextfaktoren und kognitiven Prozessen abhängig, wie in Kapitel 2.1 gezeigt. Ein solcher Kontextfaktor, welcher sich auf die Bildung einer Antwort auswirken kann, ist die vorangegangene Frage (Krosnick & Presser, 2010; Strack & Martin, 1987; Tourangeau et al., 2000). Hat eine solche vorangegangene Frage (die Kontextfrage) einen Einfluss auf das Antwortverhalten einer nachfolgenden Frage (die Zielfrage), so ist von einem FRE die Rede (Strack & Martin, 1987). Zur Entstehung solcher FRE wird davon ausgegangen, dass bei Prozessierung der Kontextfrage bestimmte Informationen aktiviert werden welche, da kognitiv nun hoch Zugänglichkeit, ebenfalls bei der Urteilsbildung und der Antwortgenerierung der Zielfrage verwendet werden. Denn um eine Frage während eines Interviews beantworten zu können, müssen gewisse Informationen abgerufen werden, wie in den Modellen zur Erklärung des Befragtenverhaltens im vorangegangenen Kapitel beschrieben. Diese abgerufenen Informationen sind jedoch auch noch für nachfolgende Fragen schnell verfügbar, bilden den Kontext (mit), in dem eine Antwort generiert wird und können das Antwortverhalten somit beeinflussen (Bogner & Landrock, 2014).

Entsprechend ist eine Beachtung solcher FRE während des Designs von Fragebögen, aber auch während der empirischen Auswertung statistischer Ergebnisse alles andere als trivial. So finden

Brookes, Chalmers, Avery, Coulman und Blazeby (2018) beispielsweiße FRE bereits innerhalb des Designprozesses von Delphi-Studien: Während der Entwicklung einer Krebsstudie bewerteten Experten Items zur klinischen Diagnose als essenzieller, wenn diese in einem Fragebogen auf Items folgten, welche sich an Symptomen orientierten, die von Patienten selbst berichtet wurden, als bei umgekehrter Reihenfolge. Cunningham, Kushnir und McCambridge (2016) fanden solche Effekte ebenfalls in Gesundheitsstudien, in welchen die Prävalenz von Befragten, das Rauchen aufzugeben höher war, wenn zuvor nach dem Interesse an kostenlosen Nikotinpflastern gefragt wurde. Und auch Garbarski, Schaeffer und Dykema (2015) stellten fest, dass die Selbsteinschätzung des Gesundheitszustandes höher mit spezifischen Gesundheitsfaktoren korrelieren, wenn die Selbsteinschätzung als nachfolgende Zielfrage statt als Kontextfrage präsentiert wird. Aber auch außerhalb der Gesundheitsforschung sind FRE zu beobachten: so konnte beispielsweise in einer Studie zu Mobbing an amerikanischen Middle Schools festgestellt werden, dass die Angaben, bereits Opfer von Mobbing gewesen zu sein, durch Rotation der Fragereihenfolge um bis zu 45% erhöht sein können (Huang & Cornell, 2016). Und auch in Studien zur Sterbehilfe können immer wieder FRE identifiziert werden (Magelssen, Supphellen, Nortvedt, & Materstvedt, 2016). Johnson, Villanti, Glasser, Pearson und Delnevo (2019) zeigen sogar auf, dass nicht nur die Rotation von Fragen selbst, sondern auch die Rotation von Fragetypen zu FRE führen können. Es zeigt sich somit, dass solche FRE eine gängige Begleiterscheinung innerhalb von Umfragen sind. Die Effekte erweisen sich sogar als derart stabil, dass sie bestehen bleiben, wenn Befragte nacheinander beide Variationen der Fragereihenfolge präsentiert bekommen (Bard & Weinstein, 2017).

Ein möglicher Erklärungsansatz für dieses Phänomen findet sich bei genauerer Betrachtung des kognitiven Prozesses der Antwortgenerierung, innerhalb welchem Befragte nicht zwangsläufig alle für die Beantwortung relevanten Informationen (in Kapitel 2.1 ist, je nach vorgestelltem Modell ebenfalls die Rede von Einstellungen und Assoziationen - um Komplexität zu reduzieren, werden diese jedoch im Folgenden vereinfachend unter „Informationen“ zusammengefasst) berücksichtigen, sondern den impliziten Such- und Erinnerungsprozess nach und den Evaluationsprozess von relevanten Informationen vorzeitig abbrechen, sobald mit subjektiv hinreichend großer Sicherheit ein Urteil gebildet werden kann, so der Prozess der Antwortgenerierung generell elaboriert verläuft. Auch unelaborierte, d.h. spontan prozessierte Antworten basieren diesbezüglich nicht zwangsläufig auf allen notwendigen und relevanten Informationen, sondern unter Umständen auf denjenigen Informationen, welche zum Zeitpunkt der Antwortgenerierung zugänglich sind (vgl. Kapitel 2.1, insb. 2.1.4). Solche Informationen können zum einen kognitiv chronisch zugängliche Informationen sein, über welche Befragte z. B. schon

viel nachgedacht haben und welche daher im Gedächtnis schneller verfügbar sind (Higgins & Bargh, 1987), zum anderen aber auch Informationen, welche lediglich unter bestimmten Umständen, das heißt situativ und akut zugänglich sind, z. B. weil Befragte diese Informationen bereits zum Beantworten einer vorangegangenen Frage aktiviert haben. Diese zuletzt beschriebene situative Zugänglichkeit von Informationen gilt als die Grundlage von FRE, wobei solche Effekte umso ausgeprägter ausfallen, je weniger alternative, für die Urteilsbildung relevante Informationen kognitiv zugänglich sind. Dies bedeutet in der Konsequenz, dass Befragte, die bereits viel über ein Thema nachgedacht haben, denen bestimmte Informationen zu diesem Thema also chronisch kognitiv zugänglich sind, weniger wahrscheinlich FRE unterliegen und zudem, dass solche Effekte umso geringer ausfallen, je mehr Fragen im Vorfeld gestellt wurden, welche alternative Informationen aktivieren (Schwarz, 1991).

FRE werden zudem in sogenannte Assimilations- sowie Kontrasteffekte differenziert. Diese werden im Folgenden ebenfalls detailliert beschrieben.

Assimilationseffekte

Ein FRE wird dann als Assimilationseffekt bezeichnet, wenn das Urteil bezüglich der Zielfrage in die gleiche Richtung verzerrt ist, wie das Urteil bezüglich der Kontextfrage ausfällt (Strack & Martin, 1987). Ein Assimilationseffekt wird erwartet, wenn die Kontextfrage bestimmte Informationen aktiviert, welche sich dann stärker auf das Antwortverhalten der Zielfrage auswirken, als sie es ohne die Aktivierung getan hätten und sich das Urteil der Zielfrage dadurch an das Urteil der Kontextfrage annähert, wobei mit zunehmender chronischer Zugänglichkeit situationsunabhängiger Informationen die situationell aktivierten Informationen an Relevanz für das Antwortverhalten verlieren (Stocké, 2002, S. 29). Ein prominentes Beispiel für einen solchen Assimilationseffekt stammt von Strack, Martin und Schwarz (1987, S. 12ff): In einer Studie unter Studierenden der Universität von Illinois wurde die Reihenfolge nach der Zufriedenheit mit „Dating“ und der Frage nach der Zufriedenheit mit „Life in general“ experimentell variiert. Wurde die Dating-Frage vor der Life-in-general-Frage gestellt, korrelierte das Antwortverhalten signifikant höher als bei umgekehrter Reihenfolge.

Erklärt werden kann dieses Phänomen anhand der bereits aufgegriffenen Theorie kognitiver Zugänglichkeit, welche davon ausgeht, dass unter anderem Urteile vor allem auf Basis derjenigen Informationen gebildet werden, welche zum Zeitpunkt der Urteilsbildung kognitiv präsent bzw. zugänglich sind. Diesbezüglich sind es vor allem die Häufigkeit der Nutzung und die zeitliche Distanz der letztmaligen Nutzung einer solchen Information, welche die kognitive Zugänglichkeit bedingen (Eagly & Chaiken, 1993; Fazio, 1990; Mayerl, 2009; Shrum & O’Guinn,

1993; Strack & Martin, 1987; Wyer & Srull, 1986). Mit Bezug auf das Auftreten eines Assimilationseffektes bedeutet dies, dass eine Kontextfrage bestimmte Informationen aktivieren kann, welche aufgrund der geringen zeitlichen Distanz zum Moment der Antwortgenerierung der Zielfrage kognitiv hoch zugänglich sind, was die Wahrscheinlichkeit erhöht, dass sie einen Einfluss auf die Urteilsbildung der Zielfrage haben können (Strack et al., 1987). Da es sich bei dem Assimilationseffekt somit um eine unbewusste Verzerrung von Urteilen auf Basis kognitiv zugänglicher Informationen handelt, kann dieser im Kontext der einleitend in Kapitel 2.2.1 vorgeschlagenen Terminologie als unbewusste und automatisch auftretende Antwortverzerrung (in Abgrenzung zur bewussten, elaborierten Editierung) beschrieben werden.

Empirische Studien hierzu zeigen zudem, dass der Grad der Spezifität von Kontextfrage und Zielfrage ebenfalls Auswirkungen auf das Auftreten von Assimilationseffekten hat. In dem oben erwähnten Experiment von Strack et al. (1987) variierten diese, wie beschrieben, die Reihenfolge, in welcher nach einem spezifischen Frageninhalt (der Zufriedenheit mit der Häufigkeit von Verabredungen) und nach einem allgemeinen Frageninhalt (der allgemeinen Lebenszufriedenheit) gefragt wurde. Für das Design mit der spezifischen Frage als Kontextfrage und der allgemeinen Frage als Zielfrage wurde eine starke Korrelation gefunden. Wurden die beiden Fragen hingegen in umgekehrter Reihenfolge gestellt, war der Zusammenhang nur sehr schwach. Auch konnten Schwarz, Strack und Mai (1991) einen Assimilationseffekt zwischen der Frage nach der Ehezufriedenheit (spezifische Kontextfrage) und der allgemeinen Lebenszufriedenheit (allgemeine Zielfrage) nachweisen. Dies begründen Schwarz und Kollegen damit, dass die Zufriedenheit in der Ehe als Teilmenge in das Urteil nach der Zufriedenheit mit dem Leben im Allgemeinen mit einfließt und sich die beiden Urteile folglich annähern. Strack et al. (1987) formulieren diese Beobachtung allgemeiner, in dem sie die These aufstellen, dass die kontextbedingte Aktivierung einer spezifischen Information dazu führt, dass diese Information bei der Generierung eines Urteils über ein allgemeines Einstellungsobjekt als besonders relevant, da kognitiv zugänglich und präsent, betrachtet wird und somit als Basis der Urteilsbildung dient. Befragte wurden folglich zunächst durch ihr eigenes Urteil daran erinnert, wie glücklich (oder unglücklich) sie in ihrer Ehe sind. Diese Ehezufriedenheit spielt nachfolgend eine größere Rolle bei der Einschätzung der Lebenszufriedenheit als zuvor nicht aktivierte Aspekte wie etwa die Arbeitszufriedenheit.

Über diese Befunde hinaus führt eine spezifisch formulierte Kontextfrage und eine allgemein formulierte Zielfrage bei geringer Zugänglichkeit alternativer Informationen zur Urteilsbildung jedoch nicht automatisch zu einem Assimilationseffekt. Unter bestimmten Bedingungen kann für die gleiche Abfolge der Fragestellung ebenfalls ein sogenannter Kontrasteffekt auftreten,

welcher sich nicht durch eine erhöhte Korrelation zwischen Kontext- und Zielfrage, sondern im Gegenteil durch eine verringerte Korrelation auszeichnet (Strack et al., 1987). Solche Kontrasteffekte werden im Folgenden erörtert.

Kontrasteffekte

Ein FRE wird dann als Kontrasteffekt bezeichnet, wenn das Urteil bezüglich der Zielfrage in die gegenläufige Richtung abweicht als das Urteil bezüglich der Kontextfrage ausfällt (Strack & Martin, 1987). Ein Kontrasteffekt wird somit erwartet, wenn Befragte eine aktivierte und kognitiv zugängliche Information gerade nicht in ein Urteil mit einbeziehen, sondern diese stattdessen aus dem Urteil ausschließen. Auf kognitiver Ebene kann diesbezüglich vermutet werden, dass Befragte Informationen subtrahieren, um eine Wiederholung zu vermeiden (Schwarz & Bless, 1992), was einem bewussten und elaborierten Editieren entspricht. Eine mögliche Erklärung für diesen Effekt liegt in den Normen einer Konversation begründet: In einer natürlichen Gesprächssituation, welche sich durch Fragen und Antworten der Gesprächsteilnehmer auszeichnet und somit nicht unähnlich einer Befragungssituation ist, gilt eine kooperative Norm des Fragens nach neuen Informationen und des Gebens dieser neuen Informationen, sprich eine Norm des Vermeidens von Redundanzen (Clark, 1985). Die durch die Kontextfrage aktivierten, spezifische Informationen werden somit gerade nicht zur Urteilsbildung der Zielfrage herangezogen, wie es für den Assimilationseffekt der Fall ist. Stattdessen dienen diese als Grundlage zur Evaluation dessen, über welche Informationen die fragende Person bereits verfügt und welche Informationen gegeben werden sollten, um Redundanzen zu vermeiden (Strack & Martin, 1987).

Ausschlaggebend für das Auftreten eines Kontrasteffektes ist somit, dass Befragte die hier beschriebene konversationale Norm akzeptieren bzw. sich als kooperative Konversationspartner verstehen. In diesem Sinne sind Befragte dann kooperativ, wenn sie informativ sind, sprich, neue Informationen zu demjenigen Wissen hinzufügen, über welches die interviewende Person bereits verfügt. Denn grundsätzlich ist der Erfolg einer jeden Konversation davon abhängig, dass die Gesprächsteilnehmer sich kooperativ verhalten, wofür vor allem die Bemühung ausschlaggebend ist, sich informativ zu verhalten (Strack et al., 1987). Was jedoch als informativ gilt, ist wiederum kontextabhängig: Wird eine Person beispielsweise in ihrer Heimat danach gefragt, wo sie Urlaub machen wird, ist die Antwort „Griechenland“ durchaus informativ. Wird die gleiche Person dies jedoch bei Ankunft in Griechenland am Flughafen gefragt, wäre die gegebene Antwort nicht informativ, da sich die Frage wahrscheinlich eher auf eine Stadt oder ein konkretes Hotel beziehen wird (Beispiel nach Strack & Martin, 1987). Ein Kontrasteffekt

ist somit, auch wenn er anderen, kognitiven Prozessen als ein Assimilationseffekt unterliegt, ebenfalls ein Kontexteffekt und somit von spezifischen Kontextbedingungen abhängig.

In Befragungssituationen werden kooperative Befragte demnach äquivalent zu einer fortlaufenden Konversation evaluieren, über welche Informationen die Interviewer bereits verfügen und hierüber den intendierten Informationsgewinn einer Frage antizipieren (Hilton & Slugoski, 1986). Anschließend kann die bereits kommunizierte Information, um Redundanz zu vermeiden, aus der nachfolgenden Antwort subtrahiert werden (Schuman & Presser, Questions and answers in attitude surveys, 1981).

Abgrenzung der Determinanten von Assimilations- und Kontrasteffekten

Problematisch ist an dieser Stelle die Abgrenzung derjenigen Kontextfaktoren, welche zu Kontrasteffekten führen gegenüber denjenigen, welche zu Assimilationseffekten führen. Denn eine spezifisch formulierte Kontextfrage kann in einer allgemein formulierten Zielfrage ebenfalls einen Kontrasteffekt auslösen, wie Schuman, Presser und Ludwig (Schuman, Presser, & Ludwig, 1981) in einer Studie über Abtreibung zeigen. Hier reduziert sich die Zustimmung zu Abtreibung im Allgemeinen (Zielfrage), wenn zuvor die Frage gestellt wurde, ob Abtreibung erlaubt sein sollte, wenn das Kind mit einer Behinderung zur Welt kommen würde (spezifisch formulierte Kontextfrage). Dieses Ergebnis lässt sich zwar dadurch erklären, dass Befragte dem speziellen Fall zustimmen und diese Zustimmung dann im allgemeineren Fall subtrahieren, um Redundanz zu vermeiden. Jedoch zeigt es auch, dass der Grad an Spezifikation von Kontext- und Zielfrage nicht ausreicht, um das Auftreten von Assimilations- und Kontrasteffekten differenziert zu bestimmen.

Stattdessen scheint der Grad an Elaboration bei der Prozessierung der zugrundeliegenden Antwort ein vielversprechendes Abgrenzungskriterium zwischen Assimilations- und Kontrasteffekt zu sein, welches zwar nicht als hinreichende, aber doch zumindest als notwendige Bedingung für die Differenzierung angesehen werden kann. Wie diskutiert handelt es sich bei einem Assimilationseffekt um eine unbewusste, spontane Verzerrung, während der Kontrasteffekt eine bewusste und elaborierte Editierung beschreibt, wodurch die beiden Effekte den in Kapitel 2.1 beschriebenen, qualitativen kognitiven Prozesspfaden automatisch-spontanen und elaborierten Prozessierens zugeordnet werden können.

Aus der bisherigen Diskussion der FRE geht hervor, dass diese, auch wenn sie als situationsspezifische Kontexteffekte diskutiert werden, nicht unabhängig sind von diversen kognitiven Prozessen und Dispositionen. So wurde erstens diskutiert, dass Assimilationseffekte dann auf-

treten, wenn die Kontextfrage Informationen oder Einstellungen aktiviert, welche zur Urteilsbildung der nachfolgenden Zielfrage verwendet werden. Stocké (2002) verweist hierzu jedoch limitierend darauf, dass die durch die Kontextfrage situativ aktivierten Informationen nicht zwingend die einzigen verfügbaren Informationen zur Urteilsbildung sein müssen, schließlich verfügen Befragte durchaus auch über von der Kontextfrage unabhängige Informationen, welche zur Urteilsbildung verwendet werden und dem wahren Wert entsprechen würden. Die situationell aktivierten Informationen werden somit mit zunehmender Zugänglichkeit der situationsunabhängigen Informationen an Bedeutung für die Urteilsbildung verlieren. Demzufolge kann davon ausgegangen werden, dass der Grad der kognitiven Zugänglichkeit situationsunabhängiger und für die Frage relevanter Informationen einen Einfluss auf das Auftreten von Assimilationseffekten aufweist. Zweitens wurde diskutiert, dass das Befolgen einer kooperativen Konversationsnorm damit assoziiert ist, ob Befragte Informationen aus einem Urteil subtrahieren, sprich, ein Kontrasteffekt auftritt (Strack & Martin, 1987). Somit erscheint auch der Kooperationsgrad als möglicher Einflussfaktor für das Auftreten von FRE. Drittens wurde festgehalten, dass ein Assimilationseffekt einem kognitiv eher unaufwändigen Prozess entsprich, während ein Kontrasteffekt kognitiv durchaus anspruchsvoller zu sein scheint (Mayerl, 2013). Somit scheint auch zwischen dem Elaborationsgrad des der Urteilsbildung zugrundeliegenden kognitiven Prozesses und dem Auftreten von Assimilations- und Kontrasteffekten ein Zusammenhang zu bestehen.

Diese drei möglichen Einflussfaktoren, der Elaborationsgrad des kognitiven Verarbeitungsprozesses, der Grad der kognitiven Zugänglichkeit von Informationen sowie der Kooperationsgrad der Befragten werden im Folgenden im Detail diskutiert.

2.2.3 Bedingungsfaktoren von Fragereihenfolgeneffekten

Im vorangegangenen Kapitel wurden der Elaborationsgrad des kognitiven Verarbeitungsprozesses, der Grad der kognitiven Zugänglichkeit von Informationen sowie der Kooperationsgrad der Befragten als mögliche Einflussfaktoren für das Auftreten von Assimilations- und Kontrasteffekten aufgezeigt. Diesbezüglich werden im Folgenden die genauen Einflussbeziehungen dieser Faktoren diskutiert.

Modus der Informationsverarbeitung und Antwortelaboration

Es ist illusorisch anzunehmen, dass Befragte in Interviewsituationen stets hoch motiviert sind und jede gegebene Antwort maximal elaborieren. Stattdessen ist es wahrscheinlicher, dass der

Antwortprozess mehr oder weniger häufig oberflächlich verläuft, dass nicht alle potenziell verfügbaren Informationen, Einstellungen, Überzeugungen etc. hinsichtlich ihrer Beitragsgüte zur Urteilsbildung evaluiert werden und dass Urteile stattdessen auf der Basis der gerade verfügbaren Informationen gebildet werden (Tourangeau & Rasinski, 1988), wie in Kapitel 2.1, insbesondere 2.1.4 dargestellt. Ein solcher oberflächlicher und unelaborierter Prozess der Antwortgenerierung kann, je nach Modell, durch welches dieser beschrieben wird, als das Überspringen einzelner Schritte des Prozesses der Antwortgeneriung innerhalb der expliziten, klassischen Modelle (Cannell et al., 1981; Strack & Martin, 1987; Sudman et al., 1996; Tourangeau et al., 2000), als Satisficing innerhalb des Satisficing-Optimizing Models (Krosnick, 1991; Krosnick & Fabrigar, 1997), als automatisch-spontaner Prozessmodus innerhalb des MODE-Modells (Fazio, 1990) bzw. des Modells der Frame Selektion (Strack & Deutsch, 2004) oder als impulsiver Prozessmodus innerhalb des Reflective-Impulsive Model (Kroneberg, 2011) beschrieben werden, auch wenn Satisficing, der automatisch-spontane Prozessmodus innerhalb des MODE-Modells, das automatisch-spontane Selektieren innerhalb des Modells der Frame Selektion und das impulsive System innerhalb des Reflective-Impulsive Model nicht als synonym zu verstehen und inhaltlich aufgrund verschiedener Modellannahmen voneinander abzugrenzen sind (vgl. Kapitel 2.1).

Insofern kann gemäß der für das Befragtenverhalten abgeleiteten These 2 (vgl. Kapitel 2.1.4) zusammengefasst werden, dass Befragte während des Antwortprozesses bzw. der Urteilsbildung einen variablen Grad an Antwortelaboration entlang eines Kontinuums aufweisen. Im Kontext der Diskussion systematischer Messfehler und Antwortverzerrungen, insbesondere der FRE, ist dies insofern relevant, als dass empirische Studien zeigen, dass unterschiedliche Arten von Antwortverzerrungen mit den verschiedenen Niveaus der Antwortelaboration assoziiert sind (vgl. z.B. Mayerl, 2013).

Hinsichtlich der Operationalisierung des Modus der Informationsverarbeitung bzw. des Grades an Antwortelaboration wurden die beiden idealtypischen Modi der Informationsverarbeitung, das automatisch-spontane Prozessieren und das geplante Verhalten als idealtypische und qualitative Pole eines quantitativen Elaborationskontinuums diskutiert. Das automatische Prozessieren wurde dabei als Pol geringer Elaboration, das geplanten Prozessieren als Pol hoher Elaboration benannt. Es erscheint somit vielversprechend, sich dem qualitativen Konzept der Antwortmodi über den quantifizierbaren Grad der Antwortelaboration anzunähern. Zur Messung des Elaborationsgrades wurden verschiedene Möglichkeiten vorgeschlagen: Die direkte Messung, z.B. über Selbstreport-Angaben der Befragten (Kahlor, Dunwoody, Griffin, Neuwirth, &

Giese, 2003; Petty & Cacioppo, 1986; Tormala, Petty, & Brinol, 2002), die Messung über Bestimmungsfaktoren der Elaboration, z.B. über die Persönlichkeitsskala für das Kognitionsbedürfnis (Keller, Bohner, & Erb, 2000; Reimer, 2003) und die Operationalisierung über mögliche Konsequenzen des Elaborationsgrades, auch wenn dies formal betrachtet einem kausalen Fehlschluss entspricht (Eagly & Chaiken, 1993). Die direkte Messung kann jedoch problematisch sein, da viele kognitive Prozesse nicht bewusst zugänglich sind (Petty & Cacioppo, 1986, S. 136f). Auch die Messung über Bestimmungsfaktoren ist dann problematisch, wenn statistische Modelle zur Erklärung des Elaborationsgrades spezifiziert werden, welche wiederum in sich selbst fehlerhaft sein können (Mayerl, 2009, S. 129). Eine etablierte Möglichkeit der Messung des Elaborationsgrades liegt hingegen in der Verwendung von Antwortreaktionszeiten, da diese als nicht reaktive Maße nicht bewusst verzerrt werden (Mayerl, 2013). Hierbei zu beachten ist, dass die Antwortreaktionszeiten bereinigt werden sollten (Mayerl & Urban, 2008), in dem 1.) Fälle ausgeschlossen werden, welche mehr als die zweifache Standardabweichung vom Median der Reaktionszeit abweichen (Outlier-Bereinigung); 2.) ein Residualindex mittels OLS-Regression und unabhängigen Variablen berechnet wird für die angenommen werden kann, dass Befragte eine Antwort nicht lange elaborieren müssen (z.B. Frage nach dem Geburtsjahr oder dem Geschlecht) und 3.) der Residualindex standardisiert wurde[2]. Dieser bereinigte, standardisierte Residualindex der Antwortreaktionszeit kann nachfolgend als Proxy-Variable des Grades an Antwortelaboration verwendet werden. Via Median-Split des Residual-Index können darüber hinaus wieder die qualitativen Gruppen der Befragten mit geringer und mit hoher Elaboration differenziert werden.

Der Assimilationseffekt wurde bereits als unbewusst und automatisch auftretende Antwortverzerrungen mit geringem kognitivem Aufwand beschrieben (vgl. Kapitel 2.2), wobei es relevant zu bemerken ist, dass es sich bei dem Elaborationsgrad nicht um eine hinreichende, sondern lediglich um eine notwendige Bedingung für das Auftreten von Assimilationseffekten handelt. Gering elaborierte Antwort führen nicht deterministisch zu einem Assimilationseffekt, während Antworten, für welche Assimilationseffekte auftreten, gering elaboriert sind. Bezogen auf den Prozess der Antwortgenerierung (vgl. Kapitel 2.1) bedeutet dies, dass Befragte, für welche Assimilationseffekte auftreten, je nach Betrachtungsebene nicht jeden der kognitiven Schritte der Antwortgenerierung vollständig durchlaufen, eine Antwortstrategie des Satisficing selektieren, eine Antwort im automatisch-spontanen Modus prozessieren bzw. eine Antwort lediglich impulsiv, ohne Hinzunahme des reflexiven Systems generieren. Zusammengefasst bedeutet dies

[2] Für eine theoretische Fundierung und genaue Beschreibung der Bereinigung von Antwortreaktionszeiten, siehe Mayerl & Urban, 2008.

für das Auftreten von Assimilationseffekten im Kontext des kognitiven Prozesses der Antwortgenerierung:

> Verfügen Befragte nicht über die Motivation oder die Möglichkeit, eine Antwort zu elaborieren, so tendieren sie zu einem Informationsverarbeitungsprozess, der durch einen geringen Elaborationsgrad der gegebenen Antwort gekennzeichnet ist. Erfolgt während eines solchen Prozesses der Antwortgenerierung eine Beeinflussung der Antwort durch Informationen, welche durch die Kontextfrage aktiviert wurden, so dass sich das Urteil der Zielfrage dem Urteil der Kontextfrage annähert, liegt ein Assimilationseffekt vor. Ein geringer Grad an Antwortelaboration stellt eine notwendige, jedoch keine hinreichende Bedingung für das Auftreten von Assimilationseffekten dar.

Der Kontrasteffekt hingegen wurde als Antwortverzerrungen mit hohem kognitivem Aufwand bzw. als bewusstes und elaboriertes Editieren einer formal vorgesehenen Antwort beschrieben (vgl. Kapitel 2.2.2), wobei auch bezüglich dieses Effektes zu beachten ist, dass der Elaborationsgrad eine notwendige und keine hinreichende Bedingung für das Auftreten von Kontrasteffekten darstellt. Das Elaborieren einer Antwort führt nicht deterministisch zu einem Kontrasteffekt, während Antworten, für welche Kontrasteffekte auftreten, stets elaboriert sind. Bezogen auf den Prozess der Antwortgenerierung (vgl. Kapitel 2.1) bedeutet dies, dass Befragte, für welche Kontrasteffekte auftreten, eine Antwortstrategie des Optimizing selektieren, eine Antwort im geplant-kontrollierten Modus prozessieren, reflexiv-kalkulierend prozessieren bzw. eine Antwort zusätzlich zum impulsiven System ebenfalls im reflexiven Systems prozessieren und dass innerhalb des Antwortprozesses eine Anpassung der generierten Antwort weg von ihrem „wahren“ Wert stattfindet, welche sich bei Kontrasteffekten durch das subtrahieren von Informationen äußert, welche zuvor bei der Beantwortung der Kontextfrage aktiviert und verwendet wurden, um eine Wiederholung zu vermeiden (Schwarz & Bless, 1992). Zusammengefasst bedeutet dies für das Auftreten von Kontrasteffekten im Kontext des kognitiven Prozesses der Antwortgenerierung:

> Verfügen Befragte sowohl über die Motivation, als auch die Möglichkeit, eine Antwort zu elaborieren, so tendieren sie zu einem Informationsverarbeitungsprozess, der durch einen hohen Elaborationsgrad der gegebenen Antworten gekennzeichnet ist. Erfolgt während eines solchen Prozesses der Antwortgenerierung eine Editierung der Antwort, so dass die durch die Kontextfrage aktivierte und berichtete Information subtrahiert wird, liegt ein Kontrasteffekt vor. Ein hoher Grad an Antwortelaboration stellt eine notwendige, jedoch keine hinreichende Bedingung für das Auftreten von Kontrasteffekten dar.

Es wurde somit diskutiert, dass der Grad der Antwortelaboration bzw. der Modus der Informationsverarbeitung assoziiert ist mit dem Auftreten von Assimilations- und Kontrasteffekten. Assimilationseffekte sind demnach bedingt durch eine Antwortstrategie, welche sich durch einen geringen Elaborationsgrad ausdrückt, während Kontrasteffekte einen erhöhten Elaborationsgrad bedürfen.

Zugänglichkeit von Informationen bezüglich des Objekts der Befragung

Mit Bezug auf den Modus der Informationsverarbeitung, in welchem Personen Entscheidungen treffen, führt Fazio (1990) darüber hinaus an, dass der Großteil menschlichen Verhaltens in sozialen Situationen schnell und spontan verlaufen muss, wird die relative Leichtigkeit betrachtet, mit der sich Menschen in ihrer Umwelt bewegen. Für Mayerl (2009) setzt eben in diesen spontanen Entscheidungsprozessen die heuristische Funktion von Einstellungen ein, welche in ihrer Eigenschaft als Bilanzurteile eines Objektes dazu beitragen, dass Situationen schnell eingeschätzt werden können, um hierauf aufbauend Verhalten zu aktivieren, ohne in jeder Situation mit Handlungsbedarf eine elaborierte und kognitiv aufwändige Entscheidung treffen zu müssen. Diese Überlegungen basieren auf Fazios Modell spontanen Prozessierens (Fazio, 1986), welches schnelles, automatisches und spontanes Verhalten von Personen als Funktion der Wahrnehmung eines Ereignisses versteht, wobei sich besagte Wahrnehmung wiederum aus der kontextbezogenen Definition der Situation sowie der von der Einstellung abhängigen, selektiven Wahrnehmung der Situation bzw. des Einstellungsobjektes ergibt. Das Verhalten ist demnach insofern einstellungsabhängig, als dass die Einstellung die selektive Wahrnehmung von Einstellungsobjekten auf einer unbewussten Ebene beeinflusst (Mayerl, 2009), wobei das Ausmaß, mit dem eine Einstellung kognitive Wahrnehmungsprozesse steuert, in hohem Maße die Stärke des Zusammenhangs zwischen Einstellungen und Verhalten bedingt (Fazio, 1986).

Demnach kommt es dann zu einem einstellungskonformen Verhalten, wenn eine Einstellung automatisch, in spontanen Prozessen aktiviert wird, diese Einstellung die Wahrnehmung des handlungsbezogenen Objektes beeinflusst und wenn diese so beeinflusste Wahrnehmung der Situation mit der zuvor erwähnten kontextbezogenen Definition der Situation übereinstimmt (Mayerl, 2009, S. 50; nach Fazio, 1986, S. 211 und Fazio, 1990, S. 80). Weiterhin führt Mayerl aus, dass je kognitiv zugänglicher solche Einstellungen von Befragten sind, desto wahrscheinlicher diese auch in Handlungssituationen aktiviert werden, wobei diese chronische Zugänglichkeit von Einstellungen begriffen werden kann als die Bereitschaft von Einstellungen, in Informationsverarbeitungsprozessen verwendet zu werden (Eagly & Chaiken, 1993) bzw. die Einfachheit, mit der Informationen abgerufen werden können (Shrum & O'Guinn, 1993).

Die Operationalisierung der kognitiven Zugänglichkeit von Einstellungen bzw. Informationen gegenüber dem Objekt der Befragung folgt der Forschung rund um die sogenannte Einstellungsstärke. Die Idee der Einstellungsstärke als Determinante des Antwortverhaltens wurde in vielen kognitiven Prozessmodellen aus dem Moderatoransatz der Konsistenztheorien der Einstellungs-Verhaltensforschung, wie etwa dem Modell spontanen Prozessierens (Fazio, 1986) übernommen (Mayerl, 2009). Dabei ist die Annahme, dass je stärker eine Einstellung ist, desto eher ist diese auch kognitiv hoch zugänglich und desto größer wirkt ihr Einfluss in Informationsverarbeitungsprozessen (Krosnick & Petty, 1995). Da die Einstellungsstärke als latentes Konstrukt verstanden wird, wurde in der Literatur eine ganze Reihe verschiedener Bestimmungsfaktoren und Indikatoren diskutiert, darunter unter anderem die direkten Erfahrungen mit dem Objekt, die Häufigkeit des Nachdenkens über das Objekt oder das Wiederholen von Einstellungsäußerungen gegenüber dem Objekt (vgl. z.B. Cooke & Sheeran, 2004; Fazio, 1986; Krosnick & Petty, 1995). Darüber hinaus werden, aufgrund der Schwierigkeit, die Einstellungsstärke direkt und in Abgrenzung zur Einstellung an sich zu erfassen, verschiedene, von der Einstellung besser abgrenzbare Einstellungsmaße diskutiert, darunter die Extremität, die Sicherheit und die Wichtigkeit der Einstellung, das Wissen über das Einstellungsobjekt sowie die Zugänglichkeit der Einstellung (Bassili & Krosnick, 2000; Krosnick & Schumann, 1988), wobei insbesondere die (chronische) Einstellungszugänglichkeit in den vergangenen Kapiteln immer wieder Erwähnung fand.

Die chronische Zugänglichkeit einer Einstellung innerhalb automatisch-spontaner Entscheidungsprozesse wird somit zu einem wichtigen Prädiktor einstellungskonformen Verhaltens. Werden diese Befunde darüber hinaus mit den Postulaten des Reflexiv-Impulsive Model (Strack & Deutsch, 2004) bzw. des Modells der Frame Selektion (Kroneberg, 2011) in Einklang gebracht, so lassen sich Einstellungen, ebenso wie Assoziationen, in einem weiteren Sinne generisch als mentale Repräsentationen bzw. allgemeiner, als Informationen bündeln, welche einem (situativ) variablen Grad an kognitiver, chronischer Zugänglichkeit unterliegen.

Nun wurde in Kapitel 2.2.2 erörtert, dass insbesondere der Assimilationseffekt dadurch bedingt ist, dass während des Prozesses der Antwortgenerierung nicht alle relevanten bzw. optimalen Informationen abgerufen und verwendet werden. Stattdessen werden situativ verfügbare bzw. akut zugängliche Informationen, welche durch die vorangegangene Frage aktiviert wurden, für die Antwortgenerierung verwendet, wodurch sich das Urteil bezüglich der Zielfrage dem Urteil bezüglich der Kontextfrage annähert. Zudem tritt ein solcher Assimilationseffekt, wie zuvor diskutiert, in schnellen, spontanen und unelaborierten Informationsverarbeitungsprozessen auf.

Es lässt sich mit Bezug auf Assimilationseffekte im speziellen, aber auch für andere Antwortverzerrungen im Allgemeinen nach Mayerl (2009) festhalten, dass während des spontanen Prozessierens bei fehlenden, zugänglichen und relevanten Informationen vor allem situative Hinweisreize die Wahrnehmung und folglich auch das Verhalten beeinflussen.

Für den Assimilationseffekt ergibt sich somit, dass zusätzlich zum Elaborationsgrad der Antwortgenerierung als notwendige Bedingung die Zugänglichkeit von Informationen welche sich beispielsweise daraus ergibt, wie häufig eine befragte Person bereits über ein gewisses Thema nachgedacht hat, insofern einen Einfluss auf das Antwortverhalten hat, als dass mit zunehmender kognitiver chronischer Zugänglichkeit relevanter Informationen die situationell akut zugänglichen Informationen (z. B. aktiviert durch die Kontextfrage) an Relevanz für das Antwortverhalten verlieren und das Auftreten von Assimilationseffekten somit unwahrscheinlicher wird (Stocké, 2002). Es lässt sich schließen:

> Innerhalb eines durch einen geringen Elaborationsgrad geprägten Informationsverarbeitungsprozess wird das Auftreten von Assimilationseffekten umso wahrscheinlicher, je geringer kognitiv chronisch zugänglich diejenigen Informationen sind, auf deren Basis Befragte eine unverzerrte Antwort generieren würden und je mehr die Antwort stattdessen auf situativ durch die Kontextfrage aktivierten Informationen basieren.

Demgegenüber wird der Kontrasteffekt beschrieben als zum einen Abhängig vom Vorliegen einer kooperativen Konversationsnorm bzw. Befragtenrolle und zum anderen als notwendig bedingt durch einen elaborierten Modus der Informationsverarbeitung (vgl. Kapitel 2.2.2). Während Modelle wie das MODE-Modell (Fazio, 1990) oder die TPB (Ajzen, 1991) die kognitive Zugänglichkeit von Informationen als Bedingungsfaktor innerhalb elaborierte kognitive Prozesse nur am Rande betonen stellt sich dennoch die Frage, welche Informationen zum Zeitpunkt der elaborierten Antwortgenerierung einer Zielfrage durch die Kontextfrage aktiviert wurden und verfügbar sind, da es eben diese Informationen sind, welche bewusst subtrahiert werden. Die kognitive Zugänglichkeit von Informationen spielt somit ebenfalls für Kontrasteffekte eine Rolle, auch wenn diese innerhalb eines elaborierten Prozesses der Antwortgenerierung auftreten.

Plausibel wird dies unter Zuhilfenahme des RIM (Strack & Deutsch, 2004) welches postuliert, dass die beiden kognitiven Systeme, das impulsive (automatisch-spontane, unelaborierte) und das reflexive (überlegt-kontrollierte, elaborierte) System, nicht voneinander abgegrenzt, sondern integriert operieren: Kognitive Prozesse verlaufen demnach impulsiv oder zusätzlich

hierzu reflexiv, indem die impulsiv automatisch aktivierten Informationen (innerhalb des Modells ist von Assoziationen die rede) reflexiv prozessiert werden. Entsprechend basieren demnach auch elaborierte Prozesse auf automatisch aktivierten Informationen, wobei es vor allem kognitiv chronisch zugängliche, aber auch situativ akut zugängliche Informationen sind, welche automatisch aktiviert werden. Individuen können diese automatisch aktivierten Informationen innerhalb elaborierter Prozesse zudem hinsichtlich ihrer Relevanz (und ebenso hinsichtlich ihrer individuellen Motive für die Befragung) evaluieren, um subjektiv besser geeignete Informationen zu aktivieren. Situativ verfügbare Informationen können so in elaborierten Prozessen zugunsten relevanterer, kognitiv zugänglicher Informationen verworfen werden, wodurch die kognitive Zugänglichkeit von Informationen auch für elaborierte Prozesse der Antwortgenerierung eine Rolle spielt.

Diesbezüglich zeigt Martin (1986), dass situativ aktivierte Informationen einer Kontextfrage nicht ausschließlich zu Assimilationseffekten, sondern unter bestimmten Bedingungen ebenfalls zu Kontrasteffekten führen. Dies liegt daran, dass Befragte in Kontexten, in denen es für sie offensichtlich ist, dass eine bestimmte Information, z. B. durch eine vorangegangene Frage, aktiviert wurde, diese Informationen bewusst nicht verwenden, um Redundanzen zu vermeiden (so eine wie in Kapitel 2.2.2 beschriebene kooperative Konversationsnorm eingenommen wurde). Weiterhin kann argumentiert werden, dass wenn Befragte eine solche situative Aktivierung von Informationen nicht bemerken, z. B. weil die eigenen Einstellungen zum Objekt der Befragung kognitiv hoch zugänglich sind und primär eher diese als die situativ aktivierten Informationen zur Urteilsbildung verwendet werden, ein Kontrasteffekt nicht auftreten kann. So geht unter anderem auch Stocké (2002) davon aus, dass mit zunehmender kognitiver chronischer Zugänglichkeit der Informationen gegenüber dem Objekt der Befragung das Auftreten von Kontrasteffekten unwahrscheinlicher wird. Es lässt sich demnach schließen:

> Innerhalb eines durch einen hohen Elaborationsgrad geprägten Informationsverarbeitungsprozess wird das Auftreten von Kontrasteffekten umso wahrscheinlicher, je weniger kognitiv chronisch zugänglich diejenigen Informationen sind, auf deren Basis Befragte eine unverzerrte Antwort generieren würden und je eher Antworten stattdessen auf situativ durch die Kontextfrage aktivierten Informationen basieren.

Es zeigt sich somit, dass Assimilationseffekte gerade innerhalb unelaborierter Prozesse bei geringer Zugänglichkeit von für das Objekt der Befragung relevanten Informationen auftreten sollten, da das Handeln hier einfachen Heuristiken und situativen Hinweisreizen folgt, zu wel-

chen die situative Aktivierung einer Information durch eine Kontextfrage gehört. Kontrasteffekte auf der anderen Seite unterliegen zwar einer von Assimilationseffekten unterschiedlichen Prozesslogik und sind notwendig durch eine erhöhte Antwortelaboration bedingt, sollten jedoch ebenfalls mit abnehmender Zugänglichkeit von für das Objekt der Befragung relevanten Informationen zunehmend auftreten.

Kooperationsgrad und Framing der Befragtenrolle

Zusätzlich zum Grad der Antwortelaboration sowie der kognitiven Zugänglichkeit von Informationen wurde als dritter Einflussfaktor für das Auftreten von FRE, insbesondere hinsichtlich des Kontrasteffektes, in Kapitel 2.2.2 das Befolgen einer kooperativen Konversationsnorm bzw. das Einnehmen einer kooperativen Befragtenrolle diskutiert und innerhalb des Modells der Frame Selektion (vgl. Kapitel 2.1.3 und 2.1.4) als Selektion eines unterstützenden Frames in Abgrenzung eines konformistischen bzw. indifferenten Frames eingebettet. Diesbezüglich beschreibt Clark (1985) allgemein, dass eine natürliche Gesprächssituation, welche sich durch Fragen und Antworten der Gesprächsteilnehmer auszeichnet, durch eine kooperative Norm des Fragens nach neuen Informationen und des Gebens dieser neuen Informationen, sprich einer Norm des Vermeidens von Redundanzen, geprägt ist. Da eine solche natürliche Gesprächssituation (durch das Stellen von Fragen und Geben von Antworten) einer Befragungssituation in sozialwissenschaftlichen Interviews nicht unähnlich ist, kann davon ausgegangen werden, dass Befragte, welche solch einer kooperativen Norm folgen, ebenfalls bestrebt sind, Redundanzen zu vermeiden. Um dies zu gewährleisten, kann aus Basis des Befragungskontextes, insbesondere der bereits gegebenen Antworten (implizit: der bereits gestellten Fragen) evaluiert werden, über welche Informationen die fragende Person bereits verfügt und welche Informationen nachfolgend gegeben werden sollten, um Redundanzen zu vermeiden (Strack & Martin, 1987). Im Falle von Kontrasteffekten erfolgt eben dies durch Subtraktion der bereits kommunizierten Information (Schuman & Presser, 1981).

Diese Strategie der Vermeidung von Redundanz ist jedoch nicht mit einer Antwortstrategie des Verringerns oder Vermeidens von Aufwand, wie etwa beim Satisficing der Fall, gleichzusetzen, da sie dem Wunsch Befragter entspringt, sich möglichst Informativ und damit unterstützend zu verhalten und unter Umständen sogar mit einem kognitiven Mehraufwand durch Editierung einer möglichen Antwort einhergeht. Folglich sind kooperative Befragte informative Befragte, da sie bemüht sind, neue Informationen zu demjenigen Wissen hinzuzufügen, über welches der Gesprächspartner, hier die interviewende Person, bereits verfügt (Strack et al., 1987).

Bezüglich des Befolgens einer kooperativen Norm postuliert Stocké (2004), dass für Befragte zwei gesellschaftliche Rollen zur Verfügung stehen, welche als Grundlage der Selektion verschiedener Antwortstrategien dienen: Eine kooperative und eine konformistische Befragtenrolle. Dabei streben Befragte mit kooperativer Rolle eine bedingungslose Unterstützung der Befragung durch möglichst korrekte Antworten an, während sich Befragte mit konformistischer Rolle gemäß der Rational Choice Theory verhalten und dadurch mitunter Effekten wie der Verzerrung durch soziale Erwünschtheit unterliegen können. Die Selektion einer der Rollen entspricht dabei der Selektion eines Frames gemäß dem Modell der Frameselektion (vgl. Kapitel 3.1.3). Ob eine kooperative bzw. eine konformistische Befragtenrolle selektiert wird, ist laut Stocké (ebd.) abhängig von der generalisierten Einstellung gegenüber Umfragen, wobei diese in Anlehnung an Mayerl (2009) und Stocké und Langfeld (2003) verstanden werden kann als eine individuelle, mentale und bilanzierende Bewertung von Umfragen im Allgemeinen. Demnach tendieren Befragte mit generell positiver Einstellung gegenüber Umfragen zu einem Framing einer kooperativen Rolle, während Befragte mit generell negativer Einstellung eher zu einem Framing einer konformistischen Rolle tendieren, welche aus Gründen der Konsistenz nachfolgend als indifferente Rolle bzw. Frame bezeichnet wird. Die (quantitative) allgemeine Einstellung gegenüber Umfragen kann somit als Approximation der Selektion bzw. des Framings der (qualitativen) Befragtenrolle dienen.

Eine Möglichkeit, sich der Operationalisierung der generalisierten Einstellung gegenüber Umfragen als Approximation der selektierten Befragtenrolle zu nähern, besteht in der Betrachtung der Nutzenabwägung von Umfragen. So wird die allgemeine Akzeptanz von Umfragen beispielsweise erhöht, wenn die Befragten die Durchführung der Befragung als sinnvolle und akkurate Unternehmung einschätzen (Roper, 1986). Auch steigt die Akzeptanz vom Umfragen und deren allgemeine Nutzenwahrnehmung, wenn Befragungen sowie deren Ergebnisse als relevant, vor allem im gesellschaftlichen Sinne, erachtet werden (Bradburn, 1979) oder wenn Befragte selbst gerne Umfrageergebnisse in der Presse oder entsprechenden Publikationen verfolgen (Stocké & Langfeldt, 2003). Zusätzlich zur Nutzenabwägung liegen Hinweise darauf vor, dass die Erfahrungen und hier vor allem die wahrgenommene Belastung aus vergangenen Umfragen auf die generalisierte Einstellung gegenüber Umfragen einwirken. Konnte zum Beispiel ein persönlicher Nutzen aus der Teilnahme an einer vergangenen Befragung gezogen werden, wobei hierunter auch subsumiert werden kann, ob die Teilnahme Spaß gemacht hat und als angenehm und abwechslungsreich empfunden wurde, so wirkt sich dies positiv auf die generalisierte Einstellung gegenüber Umfragen aus (Porst, 1998). Die wahrgenommene Belastung

während der Teilnahme an einer vergangenen Umfrage steht hingegen in einem negativen Zusammenhang mit der Umfrageeinstellung (Bradburn, 1979; Sharp & Frankel, 1983). Die allgemeine Erfahrung mit vergangenen Umfragen wiederum operationalisieren Stocké und Langfeld (2003) nach Porst (1998) und Bradburn (1979) anhand von drei Indikatoren: 1.) die wahrgenommene Belastung durch die Teilnahme an einer vorangegangenen Umfrage, wobei der zeitliche Aufwand und die notwenige Anstrengung berücksichtigt werden; 2.) der wahrgenommene Umfragenutzen, wobei die Bewertung der Erfahrung als Interessant sowie der empfundene Spaß berücksichtigt werden; 3.) das erfahrene Frageverständnis, wobei unklare Formulierungen sowie das Nicht-Erkennen eines Sinnes berücksichtigt werden. Hinsichtlich der Operationalisierung der generalisierten Einstellung gegenüber Umfragen sind somit verschiedene Indikatoren aufgezeigt, welche berücksichtigt werden sollten. Diese lauten wie folgt:

1. Die Einschätzung der Befragung als sinnvolle und akkurate Unternehmung (Roper, 1986)
2. Die Relevanz der Befragung, vor allem im gesellschaftlichen Sinne (Bradburn, 1979)
3. Das Individuelles Interesse an Umfrageergebnissen (Stocké & Langfeldt, 2003) sowie der persönliche Nutzen aus der Teilnahme an einer Befragung (Porst, 1998), zusammengefasst als der allgemeine persönliche Nutzen
4. Die wahrgenommene Belastung während der Teilnahme an einer (vergangenen) Umfrage (Bradburn, 1979; Sharp & Frankel, 1983).

Auf Basis dieser Erkenntnisse formulieren Stocké und Becker (2004) eine Skala zur generalisierten Einstellung gegenüber Umfragen, deren 16 Indikatoren den vier hier genannten Subskalen zugeordnet werden können.

Diesen zunächst allgemeinen Zusammenhang zwischen der Einstellung gegenüber Umfragen und einem sich hieraus ergebenen, spezifischen Verhalten führen Stocké und Langfeld (2003, S. 59) auf die „wahrgenommenen Kosten durch Umfragen“, also die Belastung, welche bei Befragten subjektiv durch die Befragungssituation entsteht, zurück. Es scheint somit, dass Befragte mit positiver Einstellung gegenüber Umfragen besonders motiviert bzw. kooperativ sind und demzufolge den Normen einer kooperativen Rolle entsprechen wollen, welche sich unter anderem durch das Geben neuer Informationen und das Vermeiden von Redundanzen auszeichnet, wobei empirische Belege für den konkreten Einfluss der Einstellung gegenüber Umfragen bzw. der selektierten Befragtenrolle auf das Verhalten von Personen in (potentiellen) Befragungssituationen zahlreich sind (vgl. Kapitel 2.1.4).

Entsprechend erscheint es kontraintuitiv, dass gerade diejenigen Befragten einem Kontrasteffekt unterliegen sollen, welche eine besonders positive, generalisierte Einstellung gegenüber Umfragen aufweisen und demnach die Intention haben, wahrheitskonform zu antworten. Diesbezüglich erörtert Stocké (2006), dass Befragte mit einer positiven, generalisierten Einstellung gegenüber Umfragen eine spezifische Umfragesituation derart framen, dass die uneingeschränkte Unterstützung eben jener Umfrage zum höchsten Ziel wird. Entsprechend werden weder kognitive oder psychologische, noch soziale Kosten Befragte davon abhalten, Fragen zu beantworten. Demnach ist es weniger die Intention, die spezifische Umfrage uneingeschränkt zu unterstützen, welche zu Kontrasteffekten führt. Vielmehr liegt deren Auftreten zugrunde, dass Befragte um jeden Preis neue Informationen geben möchten. Dies schließt den kognitiven Aufwand, bereits gegebene Informationen zu subtrahieren, mit ein. Da es sich hierbei um einen kognitiv aufwändigen, bewussten und elaborierten Prozess handelt, welcher einen erhöhten Grad an Antwortelaboration bedingt (Mayerl, 2013), lässt sich schlussfolgern:

> Innerhalb eines durch einen hohen Elaborationsgrad geprägten Informationsverarbeitungsprozess wird das Auftreten von Kontrasteffekten umso wahrscheinlicher, je kooperativer die selektierte Befragtenrolle ist.

Offen bleibt bei dieser Ausführung, ob und wie sich das Framing der Befragtenrolle auf das Auftreten von Assimilationseffekten auswirkt, der als unelaborierte und unbewusst auftretende Verzerrung dargestellt wurde. Wird diesbezüglich gemäß Stocké (2004) das in Kapitel 2.1.3 vorgestellte Modell der Frame Selektion (vgl. Esser, 2001; Esser, 2006a; Esser, 2006b; Kroneberg, 2005; Kroneberg, 2007; Kroneberg, 2011) bemüht so geht hieraus hervor, dass die zugrundeliegende Befragtenrolle, aktiviert durch den initial selektierten Frame, ebenfalls in automatisch-spontan verlaufenden Informationsverarbeitungsprozessen wirken, welche als notwendige Bedingung für das Auftreten von Assimilationseffekten geschildert wurden. Diesbezüglich wurde beschrieben, dass die Selektion einer Handlung im automatisch-spontanen Modus davon abhängig ist, mit welchem Grad das aktivierte Skript S_j die jeweilige Handlungswahl bedingt (vgl. Formel 4), während die Skriptselektion im automatisch-spontanen Modus wiederum davon abhängig ist, welches der Skripte S_j mental am stärksten mit dem zuvor aktivierten Frame F_i assoziiert ist, was durch die Zugänglichkeit $a_{j|i}$ ($\in [0,1]$) dargestellt wird, sowie welches der Skripte S_j generell am stärksten verankert ist, was durch die generelle Verfügbarkeit a_j ($\in [0,1]$) dargestellt wird (vgl. Formel 3).

Die zur Selektion stehenden Skripte (und hierin verschachtelt ebenfalls die Handlungsalternativen) werden somit durch die selektierten Frames bedingt. Umgekehrt beeinflussen die aktivierten Frames für automatisch-spontane Selektionen die Zugänglichkeit der Skripte und Handlungsalternativen sowie entsprechend in einem weiteren Sinne die zum Zeitpunkt der automatisch-spontanen Selektion wirkenden und dem Frame inhärenten Normen, Einstellungen, Assoziationen etc. oder zusammengefasst, welche konkreten Information in der Handlungssituation akut-situativ zugänglich sind. Eine solche akute Zugänglichkeit von Informationen ist diesbezüglich abzugrenzen von der bereits diskutierten chronischen Zugänglichkeit von Informationen. Der kognitiv chronischen Zugänglichkeit von Informationen ist eine relative zeitliche Stabilität inhärent. Es handelt sich hierbei um kognitiv generell verfügbare Informationen, die auch z.B. in Form eines assoziativen Netzwerkes wie durch das Reflective-Impulsive Model (Strack & Deutsch, 2004) beschrieben innerhalb des Langzeitgedächtnisses wirkt. Akut zugängliche Informationen hingegen sind vor allem situativ verfügbar und wirksam und können z. B. auf Basis situativer Hinweisreize wie etwa der Kontextfrage oder aber durch den selektierten Frame aktiviert werden.

Im Kontext des Beantwortens von Fragen in Umfragen bedeutet dies, dass wenn Befragte bereits viel über ein Thema nachgedacht oder mit anderen Menschen über das Thema diskutiert haben etc., die Informationen zu diesem Thema kognitiv chronisch zugänglich sind. In der konkreten Befragungssituation wirken für automatisch-spontan prozessierte Antworten nun hingegen ebenfalls die in der Situation akut zugänglichsten Informationen, welche nicht zwangsläufig deckungsgleich mit relevanten chronisch zugänglichen Informationen sein müssen. Welche Informationen in der konkreten Handlungssituation am zugänglichsten sind und somit in den Prozess der Antwortgenerierung einfließen, wird entsprechend dem Modell der Frame Selektion durch den selektierten Frame beschrieben. Wurde ein kooperativer Frame bzw. eine kooperative Befragtenrolle selektiert so kann davon ausgegangen werden, dass innerhalb dieses Frames diejenigen Informationen akut am zugänglichsten sind und für die Generierung einer Antwort verwendet werden, welche innerhalb des Frames die größte Relevanz für eine bedingungslose Unterstützung der Befragung durch möglichst korrekte Antworten haben. In diesem Fall entsprechen die akut zugänglichen und für die Antwortgenerierung verwendeten Informationen mit hoher Wahrscheinlichkeit den für die Antwortgenerierung relevanten, kognitiv chronisch zugänglichen Informationen. Wurde hingegen eine indifferente Befragtenrolle selektiert so ist es denkbar, dass trotz chronischer Zugänglichkeit relevanter Informationen diese dennoch nicht für die Urteilsbildung verwendet werden, weil die Frage z.B. initial schon nicht korrekt

gelesen und interpretiert wurde (vgl. Kapitel 2.1.1), wodurch situativ zugängliche Informationen (wie diejenigen, welche durch die Kontextfrage aktiviert wurden) in einem höheren Maß zugänglich sind als relevante chronisch zugängliche Informationen, wodurch letztere nicht (oder vermindert) in die Urteilsbildung einfließen. Für Befragte mit indifferenter Rolle wird es demnach umso wahrscheinlicher, dass sie situativ zugängliche Informationen trotz des Vorhandenseins relevanter, chronisch zugänglicher Informationen in die Antwortgenerierung mit einfließen lassen, wodurch das Auftreten von Assimilationseffekten wahrscheinlicher wird. Durch die Differenzierung zwischen situativ akut und kognitiv chronisch zugänglichen Informationen lässt sich somit ebenfalls ein Einfluss der selektierten Befragtenrolle auf das Auftreten von Assimilationseffekten theoretisch fundiert modellieren, welcher als unelaborierte und spontan auftretende Verzerrung klassifiziert wurde. Entsprechend lässt sich zusammenfassen:

> Innerhalb eines durch einen geringen Elaborationsgrad geprägten Informationsverarbeitungsprozess wird das Auftreten von Assimilationseffekten umso wahrscheinlicher, je indifferenter die selektierte Befragtenrolle ist.

Es lässt sich somit theoretisch schließen, dass sowohl die selektierte Befragtenrolle, als auch die chronische Zugänglichkeit von für die Antwortgenerierung relevanten Informationen einen Einfluss auf das Auftreten von Assimilations- und Kontrasteffekten haben, welche zudem durch den Grad an Antwortelaboration notwendig bedingt sind. Entsprechend dieser Diskussion lassen sich nachfolgend sowohl ein theoretisch fundiertes Modell zur Erklärung und Analyse von FRE entwickeln, als auch empirisch überprüfbare Hypothesen ableiten.

2.3 Fragereihenfolgeneffekte: Modell und Hypothesen

Es ist das Ziel des vorliegenden zweiten Kapitels, den kognitiven Prozess der Antwortgenerierung in Umfragen zu analysieren und zu identifizieren, wo dieser Prozess anfällig ist für das Auftreten von FRE oder spezifischer, von Assimilations- und Kontrasteffekten, um hierauf aufbauend ein theoretisch fundiertes Erklärungsmodell, auf dessen Basis FRE identifiziert und kontrolliert werden können, zu entwickeln. Diesbezüglich werden im Folgenden die in Kapitel 2.1 aus den Theorien des Befragtenverhaltens abgeleiteten, allgemeinen Thesen der Antwortgenerierung mit den in Kapitel 2.2 diskutierten Bedingungsfaktoren von FRE verschränkt und systematisiert, um hiervon ebenfalls allgemeingültige Thesen bezüglich des Auftretens von FRE abzuleiten. Bezüglich dieser Thesen erfolgt anschließend eine Betrachtung der Interaktion der einzelnen Einflussfaktoren untereinander, welche zur Bildung von daraus resultierenden,

idealtypischen Befragtengruppen genutzt wird. Schließlich wird aus diesen gebündelten Erkenntnissen ein einfaches Erklärungsmodell für FRE gebildet, aus welchem empirisch überprüfbare Hypothesen und notwendige Bedingungen für das Auftreten von FRE abgeleitet werden.

2.3.1 Zusammenfassung und Systematisierung der entwickelten Thesen

Soll nun also ein theoretisch fundiertes Erklärungsmodell für das Auftreten von FRE abgeleitet werden, so ist es unabdingbar, dieses innerhalb gängiger Theorien des Befragtenverhaltens einzubetten. Entsprechend werden nachfolgend die in Kapitel 2.1.4 bereits abgeleiteten Thesen der Antwortgenerierung zusammengefasst, um hierauf basierend Thesen für das Auftreten von FRE abzuleiten.

Zusammenfassung der Thesen des Befragtenverhaltens

Innerhalb einer ersten Verdichtung der in Kapitel 2.1.1 bis 2.1.3 diskutierten expliziten Modelle des Befragtenverhaltens (Cannell et al., 1981; Strack & Martin, 1987; Sudman et al., 1996; Tourangeau et al., 2000) bzw. der impliziten, kognitiven Prozessmodelle und der dualen Prozessmodelle (Ajzen ,1991; Esser, 2001; Esser, 2006a; Esser, 2006b; Fazio, 1986; Fazio, 1990; Fishbein & Ajzen, 2010; Kroneberg, 2005; Kroneberg, 2007; Kroneberg, 2011; Krosnick, 1991; Krosnick, 1999a; Krosnick & Fabrigar, 1997; Strack & Deutsch, 2004) konnte zusammengefasst werden, dass diesen trotz aller Unterschiede dennoch mehr oder weniger generische bzw. allgemeingültige Faktoren und Annahmen gemein sind. Neben allgemeinen situativen Faktoren und den kognitiven Fähigkeiten Befragter sind dies auf kognitiver Ebene vor allem Definitionen, Interpretationen und Bewertungen der bzw. Einstellungen gegenüber der konkreten Umfragesituation (woraus sich ebenfalls individuelle Ziele und Motive für die Befragung ergeben) sowie die Frage selbst, da sich hieraus der Kontext der Situation und damit die Bewertung der Relevanz von Informationen, welche schließlich zur Bildung einer Antwort verwendet werden, ergibt. Zusammengefasst bedeutet dies:

> *These 1: Antworten in Umfragen basieren auf chronisch und situativ zugänglichen Informationen und werden beeinflusst durch Definitionen und Bewertungen der Umfragesituation und der Frage sowie den Zielen bzw. der Motivation Befragter für die Umfrage*

Über diese grundsätzlichen Einflussfaktoren der Antwortgenerierung hinaus postulieren die vorgestellten Modelle, hier insbesondere das Satisficing-Optimizing Model (Krosnick, 1991;

Krosnick, 1999a; Krosnick & Fabrigar, 1997), das Modell spontanen Prozessierens (Fazio, 1986; Fazio, 1990), die TPB (Ajzen, 1991; Fishbein & Ajzen, 2010), das MODE Modell (Fazio, 1990), das RIM (Strack & Deutsch, 2004) sowie das MFS (Esser, 2001; Esser, 2006a; Esser, 2006b; Kroneberg, 2005; Kroneberg, 2007; Kroneberg, 2011), dass kognitive Informationsverarbeitungsprozesse einen variablen Grad an Elaboration aufweisen und sich folglich in ihren kognitiven Mechanismen je Elaborationsgrad unterscheiden. Entsprechend wurde abgeleitet:

> *These 2: Die Generierung von Antworten in Umfragen hat einen individuell und situativ variablen Grad an Elaboration, welche durch den kognitiven Aufwand beschrieben wird der aufgebracht wird, um eine Antwort zu generieren. Je nach Grad der Elaboration unterliegt der Prozess der Antwortgenerierung verschiedenen kognitiven Prozessen.*

Weiterhin und sehr ausschlaggebend für den vorliegenden Kontext ist das Auftreten von Antwortverzerrungen innerhalb des Prozesses der Antwortgenerierung, also der Verzerrung der gegebenen Antwort weg von ihrem „wahren" Wert. Von besonderer Relevanz ist hierbei, dass innerhalb der beschriebenen Modelle sowohl unbewusst und automatisch auftretende Verzerrungen, als auch Verzerrungen auf Basis bewusst-elaborierter Editierung vorgesehen sind. Entsprechend gilt:

> *These 3: Gegebene Antworten in Umfragen sind nicht ausschließlich wahr, sondern können auch in dem Sinne der Unwahrheit entsprechen, als dass sie nicht kohärent mit individuellen Dispositionen bzw. der empirisch überprüfbaren Realität sind. Antworten in Umfragen verfügen somit über einen variablen Grad an Verzerrung, wobei Verzerrungen unbewusst oder bewusst eintreten können. Unbewusste Verzerrungen resultieren, wenn Antworten nicht auf der Basis optimaler, sondern situativ verfügbarer Informationen und Assoziationen oder Heuristiken basieren. Bewusste Verzerrungen resultieren, wenn vorgesehene Antworten aufgrund der Kollision mit persönlichen Motiven editiert werden.*

Dass Verzerrungen zum einen automatisch, unbewusst und unelaboriert, zum anderen aber auch bewusst und elaboriert auftreten ist innerhalb der Erklärungslogik dualer Prozessmodelle gut darzustellen. Zwar stellt der Grad an Antwortelaboration für den FRE, wie in Kapitel 2.2 diskutiert, keine hinreichende, sondern lediglich eine notwendige Bedingung dar, dennoch sind die Bedingungsfaktoren elaborierten Prozessierens und Handelns alles andere als trivial. Entsprechend gilt:

These 4: Dem Prozess der Antwortgenerierung liegt ein variabler Grad an Elaboration zugrunde, welcher sich in differenten kognitiven Prozessen äußert. Die Antwortelaboration ist bedingt durch die individuelle und situative Motivation und Möglichkeit zum Elaborieren, wobei es sich hierbei um eine interaktive Beziehung handelt. Nur wenn die Motivation und die Möglichkeit gegeben ist, werden Antworten elaboriert prozessiert. In allen anderen Fällen verläuft der Prozess der Antwortgenerierung unelaboriert.

Zuletzt wurde aufgezeigt, dass sich die in These 1 aufgegriffenen und aus verschiedenen Modellen des Befragtenverhaltens abgeleiteten Bedingungsfaktoren „Definitionen und Bewertungen der Umfragesituation und der Frage“ sowie die „Ziele bzw. Motive Befragter für die Umfrage“ gemäß dem Modell der Frame Selektion (vgl. Esser, 2001; Esser, 2006a; Esser, 2006b; Kroneberg, 2005; Kroneberg, 2007; Kroneberg, 2011) aus dem individuellen und situativen Framing einer spezifischen Befragtenrolle ergeben. Entsprechend kann festgehalten werden:

These 5: Das Befragtenverhalten sowie konkret formulierte Antworten in Umfragen stehen in Abhängigkeit von der individuell zugrundeliegenden Befragtenrolle, wobei die Selektion und Aktivierung einer entsprechenden Rolle dem Prozess des Framings entspricht. Idealtypisch lassen sich qualitativ eine kooperative, unterstützende Befragtenrolle sowie eine der Befragung indifferent oder ablehnend gegenüberstehende Befragtenrolle differenzieren.

Ableitung der These über das Auftreten eines Assimilationseffektes

Über diese in Kapitel 2.1 abgeleiteten, allgemeingültigen Thesen des Befragtenverhaltens hinaus wurden in Kapitel 2.2.3 theoriegeleitet Annahmen bezüglich des Auftretens von Assimilations- und Kontrasteffekten formuliert, welche mit den hier zusammengefassten Thesen in Abgleich gebracht werden, um hieraus ebenfalls Thesen des Auftretens von Assimilations- und Kontrasteffekten zu entwickeln. So wurde bezüglich des Assimilationseffektes postuliert:

Verfügen Befragte nicht über die Motivation oder die Möglichkeit, eine Antwort zu elaborieren, so tendieren sie zu einem Informationsverarbeitungsprozess, der durch einen geringen Elaborationsgrad der gegebenen Antwort gekennzeichnet ist. Erfolgt während eines solchen Prozesses der Antwortgenerierung eine Beeinflussung der Antwort durch Informationen, welche durch die Kontextfrage aktiviert wurden, so dass sich das Urteil der Zielfrage dem Urteil der Kontextfrage annähert, liegt ein Assimilationseffekt vor. Ein geringer Grad an Antwortelaboration stellt eine notwendige, jedoch keine hinreichende Bedingung für das Auftreten von Assimilationseffekten dar.

Entsprechend These 1 wird das aufgezeigte Postulat insofern bestätigt, als dass durch Assimilation geprägt Antworten auf situativ verfügbaren Informationen, aktiviert durch die Kontextfrage, basieren. Der variable Grad an Antwortelaboration (These 2), die Bedingungsfaktoren für den Elaborationsgrad (These 4) sowie die Notwendigkeit eines gering bzw. nicht elaborierten Informationsverarbeitungsprozesses für das Auftreten der Assimilation (These 3) finden sich ebenfalls wieder. Bezüglich These 5 wird innerhalb des Postulates keine Aussage getroffen, so dass diese nicht verletzt wird. Weiterhin wurde bezüglich des Assimilationseffektes postuliert:

> Innerhalb eines durch einen geringen Elaborationsgrad geprägten Informationsverarbeitungsprozess wird das Auftreten von Assimilationseffekten umso wahrscheinlicher, je geringer kognitiv chronisch zugänglich diejenigen Informationen sind, auf deren Basis Befragte eine unverzerrte Antwort generieren würden und je mehr die Antwort stattdessen auf situativ durch die Kontextfrage aktivierten Informationen basieren.

Durch die Berücksichtigung der Wirkung situativ akut zugänglicher Informationen, z.B. aktiviert durch die Kontextfrage, in Abgrenzung zu kognitiv chronisch zugänglichen Informationen innerhalb eines gering elaborierten Informationsverarbeitungsprozesses ist auch in diesem Postulat kein Verstoß gegen die zuvor abgeleiteten Thesen ersichtlich. Insbesondere die Aussage in These 3: „Unbewusste Verzerrungen resultieren, wenn Antworten nicht auf der Basis optimaler, sondern situativ zugänglicher Informationen und Assoziationen oder Heuristiken basieren“ unterstützt das obige Postulat zur Gänze. Zuletzt wurde bezüglich des Auftretens von Assimilationseffekten formuliert:

> Innerhalb eines durch einen geringen Elaborationsgrad geprägten Informationsverarbeitungsprozess wird das Auftreten von Assimilationseffekten umso wahrscheinlicher, je indifferenter die selektierte Befragtenrolle ist.

Wird These 5 in dem Sinne akzeptiert, dass Befragte situativ verschiedene Befragtenrollen einnehmen, welche sich idealtypisch durch eine Unterstützung der Umfrage oder einer indifferenten Haltung gegenüber der Umfrage äußern und folglich eine spezifische Einstellung gegenüber Umfragen bilden, dass die Selektion einer solchen Rolle dem Prozess des Framings entspricht und dass solche Frames gemäß These 1 auf die Zugänglichkeit und Relevanzbewertung von Informationen einwirken, so ist auch dieses Postulat in dem Sinne plausibel, als dass der Umfrage indifferent gegenüberstehende Befragte andere Informationen als Relevant für die Situation framen als unterstützende Befragte. In diesem Sinne ist es plausibel, dass indifferente Be-

fragte vor allem diejenigen Informationen als Relevant bewerten, welche ohne hohen kognitiven Aufwand verwendet werden können, wobei es sich hierbei vor allem um situativ zugängliche Informationen handelt wie jene, welche durch die Kontextfrage aktiviert wurden.

Werden diese Postulate nun verdichtet, so kann für den Assimilationseffekt als weitere These formuliert werden:

> *These 6: Assimilationseffekte sind notwendig durch einen geringen Grad an Antwortelaboration bedingt und treten auf, wenn situativ durch die Kontextfrage aktivierte Informationen in den Prozess der Antwortgenerierung einfließen. Dies ist der Fall, wenn solche Informationen akut in einem höheren Maße zugänglich sind als alternative, jedoch für die Beantwortung der Frage relevantere und kognitiv chronisch zugängliche Informationen. Die situative Zugänglichkeit von Informationen wird durch die eingenommene Befragtenrolle bedingt. Für kooperative Befragte sind die für die wahre Beantwortung der Frage benötigten und chronisch verfügbaren Informationen hoch relevant und entsprechend akut zugänglich, für indifferente Befragte sind jene Informationen relevant und zugänglich welche benötigt werden, um Fragen entsprechend eigener Motive zu beantworten.*

Ableitung der These über das Auftreten eines Kontrasteffektes

Demgegenüber wurden, ebenfalls theoriegeleitet, Postulate für das Auftreten von Kontrasteffekten formuliert, welche im Folgenden ebenfalls mit den formulierten Thesen der Antwortgenerierung abgeglichen werden. Entsprechend lautet das erste Postulat:

> Verfügen Befragte sowohl über die Motivation, als auch die Möglichkeit, eine Antwort zu elaborieren, so tendieren sie zu einem Informationsverarbeitungsprozess, der durch einen hohen Elaborationsgrad der gegebenen Antworten gekennzeichnet ist. Erfolgt während eines solchen Prozesses der Antwortgenerierung eine Editierung der Antwort, so dass die durch die Kontextfrage aktivierte und berichtete Information subtrahiert wird, liegt ein Kontrasteffekt vor. Ein hoher Grad an Antwortelaboration stellt eine notwendige, jedoch keine hinreichende Bedingung für das Auftreten von Kontrasteffekten dar.

Entsprechend These 1 wird das aufgezeigte Postulat ebenfalls insofern bestätigt, als dass innerhalb von durch Kontrasteffekte geprägten Antworten jene Informationen subtrahiert werden, welche je nach Wahrnehmung und Interpretation der Situation bereits gegeben wurden, das persönliche Motiv „sich informativ verhalten“ vorausgesetzt. Darüber hinaus ist das Postulat ebenfalls Konform mit dem variablen Grad an Antwortelaboration (These 2), den Bedingungsfaktoren für den Elaborationsgrad (These 4) sowie der Notwendigkeit eines hoch elaborierten

Informationsverarbeitungsprozesses für das Auftreten des Kontrasteffektes (These 3). Zusätzlich wurde bezüglich des Kontrasteffektes postuliert:

> Innerhalb eines durch einen hohen Elaborationsgrad geprägten Informationsverarbeitungsprozess wird das Auftreten von Kontrasteffekten umso wahrscheinlicher, je weniger kognitiv chronisch zugänglich diejenigen Informationen sind, auf deren Basis Befragte eine unverzerrte Antwort generieren würden und je eher Antworten stattdessen auf situativ durch die Kontextfrage aktivierten Informationen basieren.

Durch die Berücksichtigung der Wirkung von durch die Kontextfrage aktivierten, situativ akut zugänglichen Informationen in Abgrenzung zu kognitiv chronisch zugänglichen Informationen ist auch in diesem Postulat kein Verstoß gegen die zuvor abgeleiteten Thesen ersichtlich so berücksichtigt wird, dass die Wahrnehmung und Verfügbarkeit von Informationen, wie z.B. durch das Reflective-Impulsive Model (Strack & Deutsch, 2004) beschrieben, auch für elaborierte Informationsverarbeitungsprozesse von Relevanz sind. Insbesondere die Aussage in These 3: „Bewusste Verzerrungen resultieren, wenn vorgesehene Antworten aufgrund der Kollision mit persönlichen Motiven editiert werden“ unterstützt das obige Postulat zur Gänze in dem Sinne, dass bei eintretendem Kontrasteffekt Befragte vorgesehene Antworten bewusst editieren (um die bereits gegebenen Informationen zu subtrahieren), um einer Kollision mit dem persönlichen Motiv „neue Informationen zu den bereits gegebenen Informationen hinzufügen und Redundanz vermeiden“ zu vermeiden. Zuletzt wurde bezüglich des Auftretens von Kontrasteffekten formuliert:

> Innerhalb eines durch einen hohen Elaborationsgrad geprägten Informationsverarbeitungsprozess wird das Auftreten von Kontrasteffekten umso wahrscheinlicher, je kooperativer die selektierte Befragtenrolle ist.

Wird, analog zur Diskussion der Assimilationseffekte, These 5 in dem Sinne akzeptiert, dass Befragte situativ verschiedene Befragtenrollen einnehmen, welchen unterschiedliche Motive bzw. Ziele für die Befragung inhärent sind, so ist auch dieses Postulat in dem Sinne plausibel, als dass kooperative Befragte vor allem informative Befragte sind, welche sich durch das Bedürfnis nach der Gabe neuer Informationen und dem Vermeiden von Redundanzen auszeichnen, sprich Informationen, welche in der vorangegangenen Frage bereits gegeben wurden, für die nachfolgende Frage subtrahieren. Entsprechend treten Kontrasteffekte umso wahrscheinlicher auf, desto kooperativer Befragte sind. Werden diese Postulate nun verdichtet, so kann auch für den Kontrasteffekt formuliert werden:

These 7: Kontrasteffekte sind notwendig durch einen hohen Grad an Antwortelaboration bedingt und treten auf, wenn situativ durch die Kontextfrage aktivierte Informationen bewusst aus dem Prozess der Antwortgenerierung ausgeschlossen werden. Dies ist der Fall, wenn solche situativ aktivierten Informationen akut in einem Höheren Maße zugänglich sind als alternative, jedoch für die Beantwortung der Frage relevantere und kognitiv chronisch zugängliche Informationen und Befragte aufgrund ihrer selektierten Befragtenrolle über das Motiv verfügen, Redundanzen zu vermeiden, was zu einer bewussten Subtraktion der situativ und akut zugänglichen Informationen führt.

Assimilations- und Kontrasteffekte treten somit, gemäß der hier formulierten Thesen, in Abhängigkeit der selektierten Befragtenrolle und der Zugänglichkeit von für die Beantwortung der Frage relevanten Informationen auf und sind notwendig bedingt durch den Grad an Antwortelaboration. Darüber hinaus ist jedoch hinsichtlich der Thesen noch unberücksichtigt, wie diese Faktoren in Beziehung zueinanderstehen, ob interaktive Einflüsse auf FRE herrschen und wie verschiedene Gruppen von Befragten zu modellieren sind. Dies wird nachfolgend ebenfalls beleuchtet, bevor aus den abgeleiteten Befunden zusammen mit den hier formulierten Thesen ein Modell zur Erklärung von FRE abgeleitet wird.

2.3.2 Interaktion und Gruppierung der Einflussfaktoren

Gemäß den zuvor zusammengefassten Thesen ist das Auftreten von FRE abhängig von der selektierten Befragtenrolle und der Zugänglichkeit von Informationen mit Relevanz für das Objekt der Befragung, sowie notwendig bedingt durch den Grad an Antwortelaboration. In diesem Kapitel wird entsprechend erörtert, wie die Interaktion dieser Einflussfaktoren auf das Auftreten von FRE einwirkt.

Definition idealtypischer Befragtengruppen

Werden die quantitativen Kontinuen der Antwortelaboration und der chronischen Zugänglichkeit von Information mit Relevanz für das Objekt der Befragung qualitativ hinsichtlich der Extremwerte an den jeweiligen Polen idealtypisch dichotomisiert (geringe Elaboration vs. hohe Elaboration und geringe Zugänglichkeit vs. hohe Zugänglichkeit) und wird das Framing der Befragtenrolle ebenfalls als qualitativ dichotom idealtypisch (kooperative vs. indifferente Rolle) statt quantitativ entlang eines Kontinuums (vgl. Kapitel 2.2.3) betrachtet, so ergeben sich

hieraus zunächst acht verschiedene, idealtypische Gruppen von Befragten hinsichtlich des Auftretens von FRE, welche den Polen der jeweilig zugrundeliegenden Skalen entsprechen:

1. Befragte mit einem geringen Grad an Antwortelaboration, geringer Zugänglichkeit und einer indifferenten Befragtenrolle
2. Befragte mit einem geringen Grad an Antwortelaboration, geringer Zugänglichkeit und einer kooperativen Befragtenrolle
3. Befragte mit einem geringen Grad an Antwortelaboration, hoher Zugänglichkeit und einer indifferenten Befragtenrolle
4. Befragte mit einem geringen Grad an Antwortelaboration, hoher Zugänglichkeit und einer kooperativen Befragtenrolle
5. Befragte mit einem hohen Grad an Antwortelaboration, geringer Zugänglichkeit und einer indifferenten Befragtenrolle
6. Befragte mit einem hohen Grad an Antwortelaboration, geringer Zugänglichkeit und einer kooperativen Befragtenrolle
7. Befragte mit einem hohen Grad an Antwortelaboration, hoher Zugänglichkeit und einer indifferenten Befragtenrolle
8. Befragte mit einem hohen Grad an Antwortelaboration, hoher Zugänglichkeit und einer kooperativen Befragtenrolle

Da der Grad an Antwortelaboration jedoch notwendig bedingt, ob entweder ein Assimilations- oder ein Kontrasteffekt vorliegt und für diese beiden Effekte jeweils verschiedene kognitive Mechanismen zugrunde liegen, können die interaktiven Beziehungen der hier genannten Bedingungsfaktoren spezifisch für den Assimilations- als auch für den Kontrasteffekt separat voneinander analysiert werden, wodurch der Grad an Antwortelaboration aus der Betrachtung exkludiert werden kann, da dieser für Assimilationseffekte als notwendig gering und für Kontrasteffekten als notwendig hoch angenommen werden kann. Sowohl für Assimilations-, als auch für Kontrasteffekte reduziert sich somit die Betrachtung auf die Interaktion der nunmehr als Determinanten zu betrachtende Zugänglichkeit von Informationen sowie das Framing der Befragtenrolle. Es ergeben sich entsprechend, erneut idealtypisch, die folgenden qualitativen Gruppen sowohl für die Betrachtung von Assimilations- als auch von Kontrasteffekten:

1. Befragte mit geringer Zugänglichkeit und einer indifferenten Befragtenrolle
2. Befragte mit geringer Zugänglichkeit und einer kooperativen Befragtenrolle
3. Befragte mit hoher Zugänglichkeit und einer indifferenten Befragtenrolle
4. Befragte mit hoher Zugänglichkeit und einer kooperativen Befragtenrolle

Es bleibt dabei zu berücksichtigen, dass diese qualitative Gruppendarstellung lediglich der Übersicht und einer vereinfachten Diskussion dient, da sowohl der Grad an Zugänglichkeit, als auch das Framing der Befragtenrolle einem quantitativen Kontinuum mit den hier beschriebenen Gruppeneigenschaften als jeweilige Pole folgt.

Entsprechend gilt es nun zu erörtern, inwiefern sowohl Assimilations- (bei geringer Elaboration), als auch Kontrasteffekte (bei erhöhter Elaboration) innerhalb dieser Gruppen auftreten, wobei hinsichtlich der hieraus abzuleitenden Modelle ebenfalls zu berücksichtigen ist, dass sowohl Assimilations-, als auch Kontrasteffekte in einer variablen Stärke auftreten können. Beschreibt die Assimilation eine Annäherung des Urteils einer Zielfrage an das Urteil der Kontextfrage, so beschreibt der Grad der Annäherung die Stärke der Assimilation. Beschreibt demgegenüber der Kontrast eine Entfernung des Urteils der Zielfrage vom Urteil der Kontextfrage, so beschreibt der Grad der Entfernung die Stärke des Kontrastes. Assimilations- und Kontrasteffekte können entsprechend quantitativ in unterschiedlicher Stärke ausfallen und entsprechend theoretisch qualitativ in starke und schwache Assimilations- und Kontrasteffekte differenziert werden.

Assimilationseffekte innerhalb der idealtypischen Gruppen

Beginnend mit der Betrachtung des Assimilationseffektes kann für diesen auf Basis der vorangegangenen Erörterungen zusammengefasst werden, dass er umso stärker ausfällt, je geringer die chronische Zugänglichkeit von Informationen mit Relevanz für das Objekt der Befragung und je eher eine indifferente Befragtenrolle vorliegt. Ein geringer Grad an Antwortelaboration wird zudem für das Auftreten von Assimilationseffekten vorausgesetzt und wird entsprechend im nachfolgenden Abschnitt nicht mehr gesondert aufgeführt.

Für Befragte (1) mit geringer Zugänglichkeit und einer indifferenten Befragtenrolle bedeutet dies, dass für diese ein starker Assimilationseffekt erwartet werden kann, da sowohl eine indifferente Befragtenrolle, als auch eine geringe Zugänglichkeit von Informationen mit Relevanz für die Frage vorliegen. Befragte werden Fragen entsprechend auf dem zum Zeitpunkt der Urteilsbildung situativ zugänglichen Informationen generieren, auch da die Verwendung dieser hinsichtlich der indifferenten Befragtenrolle den geringsten kognitiven Aufwand erfordert.

Für Befragte (2) mit geringer Zugänglichkeit und einer kooperativen Befragtenrolle bedeutet dies, dass ein schwacher Assimilationseffekt erwartet werden kann. Die Befragtenrolle bedingt in diesem Fall zwar die Aktivierung aller relevanter Informationen für eine informative und wahrheitskonforme Generierung einer Antwort, jedoch sind diejenigen Informationen mit

höchster Relevanz für die Antwortgenerierung chronisch nicht oder wenig zugänglich. Antworten basieren somit sowohl auf weniger relevanten, als auch auf situativ zugänglichen Informationen wie jenen, welche durch die Kontextfrage aktiviert wurden. Assimilationseffekte können somit trotz kooperativer Befragtenrolle erwartet werden, auch wenn diese im Gegensatz zur ersten Befragtengruppe schwächer ausfallen sollten, da Befragte dieser zweiten Gruppe Antworten zumindest nicht ausschließlich auf situativ verfügbaren Informationen generieren, wie dies für die erste Gruppe der Fall ist.

Für Befragte (3) mit hoher Zugänglichkeit und indifferenter Befragtenrolle bedeutet dies ebenfalls, dass ein schwacher Assimilationseffekt erwartet werden kann. Zwar sind in dieser Gruppe Informationen mit Relevanz für das Objekt der Befragung kognitiv vorhanden, jedoch nimmt die aktivierte Befragtenrolle Einfluss auf die situative Zugänglichkeit kognitiv vorhandener Information und deren Relevanzbewertung. Das bedeutet, dass auch wenn innerhalb Befragter Informationen mit Relevanz für das Objekt der Befragung vorliegen, diese situativ nicht in den Prozess der Urteilsbildung mit einfließen müssen, da auf Basis der Befragtenrolle alternativ verfügbare Informationen eine höhere Relevanzbewertung erhalten und somit zugänglicher sein können. Soll aufgrund der Befragtenrolle etwa kognitiver Aufwand vermieden werden, so werden situativ durch die Kontextfrage aktivierte Informationen sicher als relevanter bewertet als chronisch verfügbare Informationen, schlicht, weil diese bereits aktiviert sind und bereits akut in Urteilsbildungsprozessen verwendet wurden. Entsprechend basieren Antworten innerhalb dieser Befragtengruppe nicht ausschließlich auf relevanten, chronisch zugänglichen Informationen, sondern ebenfalls auf situativ zugänglichen Informationen, wodurch ein schwacher Assimilationseffekt auftreten kann.

Für Befragte (4) mit hoher Zugänglichkeit und positiver Befragtenrolle zuletzt ist kein Assimilationseffekt zu erwarten. Informationen mit Relevanz für die Antwortgenerierung sind chronisch zugänglich und werden aufgrund der selektierten Befragtenrolle ebenfalls aktiviert und verwendet, so dass Antworten auf solchen Informationen und nicht auf situativ aktivierten Informationen basieren.

Mit Bezug auf den Assimilationseffekt lässt sich somit festhalten, dass die beiden Determinanten der Informationszugänglichkeit und der Befragtenrolle additiv und nicht interaktiv wirken. So resultiert eine geringe Zugänglichkeit in jedem der diskutierten Fälle (d.h. ebenfalls bei konstanter Befragtenrolle) in einem Assimilationseffekt, während eine indifferente Befragtenrolle ebenfalls in jedem der diskutierten Fälle (d.h. ebenfalls bei konstanter Zugänglichkeit) zu einem Assimilationseffekt führt. Während der Assimilationseffekt bei geringer Zugänglichkeit und

kooperativer Rolle ebenso wie bei hoher Zugänglichkeit und indifferenter Rolle lediglich schwach ausfällt, sollte er, werden die beiden Konditionen „geringe Zugänglichkeit“ und „indifferente Rolle“ addiert, stark ausfallen und ansonsten, d.h. bei Fehlen der Konditionen, nicht auftreten. Dies wird in der folgenden Abbildung verdeutlicht:

Abbildung 1: Bedingungsfaktoren des Assimilationseffektes



Kontrasteffekte innerhalb der idealtypischen Gruppen

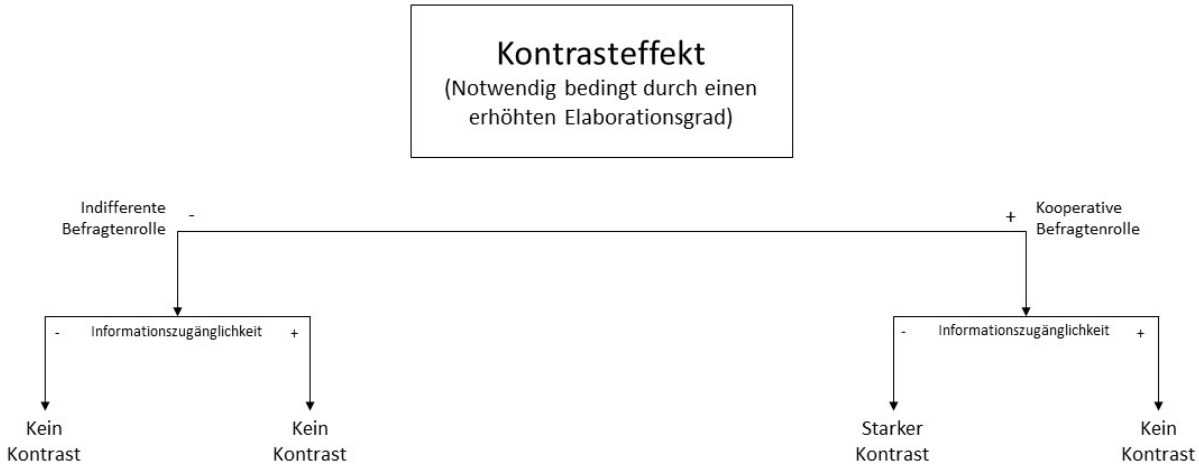
Entsprechend kann für den Kontrasteffekt auf Basis der vorangegangenen Erörterungen zusammengefasst werden, dass dieser umso stärker auftritt, je geringer die chronische Zugänglichkeit von Informationen mit Relevanz für das Objekt der Befragung ausfällt und wenn eine indifferente Befragtenrolle vorliegt. Ein hoher Grad an Antwortelaboration wird zudem für das Auftreten von Kontrasteffekten vorausgesetzt und wird entsprechend im nachfolgenden Abschnitt ebenfalls nicht gesondert aufgeführt.

Für Befragte (1) mit geringer Zugänglichkeit und einer indifferenten Befragtenrolle ebenso wie für Befragte (3) mit hoher Zugänglichkeit und einer indifferenten Befragtenrolle bedeutet dies, dass für diese kein Kontrasteffekt erwartet werden kann, da es generell einer kooperativen Befragtenrolle bedarf, um bewusst den kognitiven Mehraufwand einer Editierung der Antwort in Form der Subtraktion redundanter Informationen in Kauf zu nehmen. Ob Informationen zugänglich sind oder nicht ist bei indifferenter Befragtenrolle für das Auftreten von Kontrasteffekten entsprechend irrelevant, auch wenn dies für das Auftreten anderer Antwortverzerrungen wohl nicht trivial ist. Da diese jedoch nicht im Fokus dieser Arbeit stehen, kann dieser Punkt an dieser Stelle ignoriert werden.

Für Befragte (2) mit geringer Zugänglichkeit und einer kooperativen Befragtenrolle bedeutet dies, dass ein Kontrasteffekt erwartet werden kann. Das Fehlen chronisch zugänglicher Informationen mit Relevanz für das Objekt der Befragung impliziert, dass zum Zeitpunkt der Urteilsbildung situativ akut zugängliche Informationen wirken, wie sie etwa durch die Kontextfrage aktiviert werden. Das Vorhandensein einer kooperativen Befragtenrolle impliziert weiter, dass eben jene situativ verfügbaren Informationen bewusst aus der Urteilsbildung subtrahiert werden, um Redundanz zu vermeiden.

Für Befragte (4) mit hoher Zugänglichkeit und kooperativer Befragtenrolle zuletzt wird wiederum kein Kontrasteffekt erwartet. Zwar impliziert die Befragtenrolle das bewusste Subtrahieren redundanter Informationen. Jedoch ist es auf aufgrund der chronischen Zugänglichkeit von für die Urteilsbildung relevanten Informationen nicht plausibel, dass in einer solchen elaborierten und durch Unterstützung der Umfrage geprägten Antwortsituation situativ verfügbare Informationen wie jene, welche durch die Kontextfrage aktiviert werden, mit in die Antwortgenerierung einfließen. Antworten basieren somit nicht auf durch die Kontextfrage aktivierten Informationen, weshalb auch keine Subtraktion dieser Informationen vorgenommen werden muss, um Redundanz zu vermeiden.

Abbildung 2: Bedingungsfaktoren des Kontrasteffektes



Mit Bezug auf den Kontrasteffekt lässt sich somit festhalten, dass die beiden Determinanten der Informationszugänglichkeit und der Befragtenrolle ausschließlich interaktiv und nicht additiv wirken. Ein Kontrasteffekt tritt entsprechend dieser Diskussion ausschließlich auf, wenn eine kooperative Befragtenrolle vorliegt und wenn für die Urteilsbildung relevante Informationen nicht chronisch vorliegen. Weiterhin ist der Kontrasteffekt umso stärker, desto geringer die

chronische Zugänglichkeit und desto höher die Kooperationsbereitschaft als Konsequenz der selektierten Befragtenrolle. Dies wird in Abbildung 2 zusammengefasst.

2.3.3 Synthese, Modell und Hypothesen

Bezüglich der Diskussion von FRE wurden für das Auftreten dieser Theorien beschrieben, Evidenzen aufgezeigt und Postulate und Thesen aufgestellt, welche im vorangegangenen Kapitel verdichtet wurden.

So lassen sich FRE wie beschrieben grundsätzlich in Assimilations- und Kontrasteffekte unterscheiden, wobei der Assimilationseffekt eine Annäherung des Urteils einer Zielfrage an das Urteil einer Kontextfrage und der Kontrasteffekt eine Entfernung des Urteils einer Zielfrage von dem Urteil einer Kontextfrage darstellt.

Es wurde diskutiert, dass beide Effekte verschiedenen kognitiven Prozessen unterliegen, insbesondere, dass der Assimilationseffekt notwendig durch einen geringen Grad an Antwortelaboration und der Kontrasteffekt notwendig durch einen erhöhten Grad an Antwortelaboration bedingt ist. Darüber hinaus wurden die chronische Zugänglichkeit von Informationen mit Relevanz für das Objekt der Befragung sowie die selektierte Befragtenrolle als Determinanten für das Auftreten der jeweiligen Effekte beschrieben.

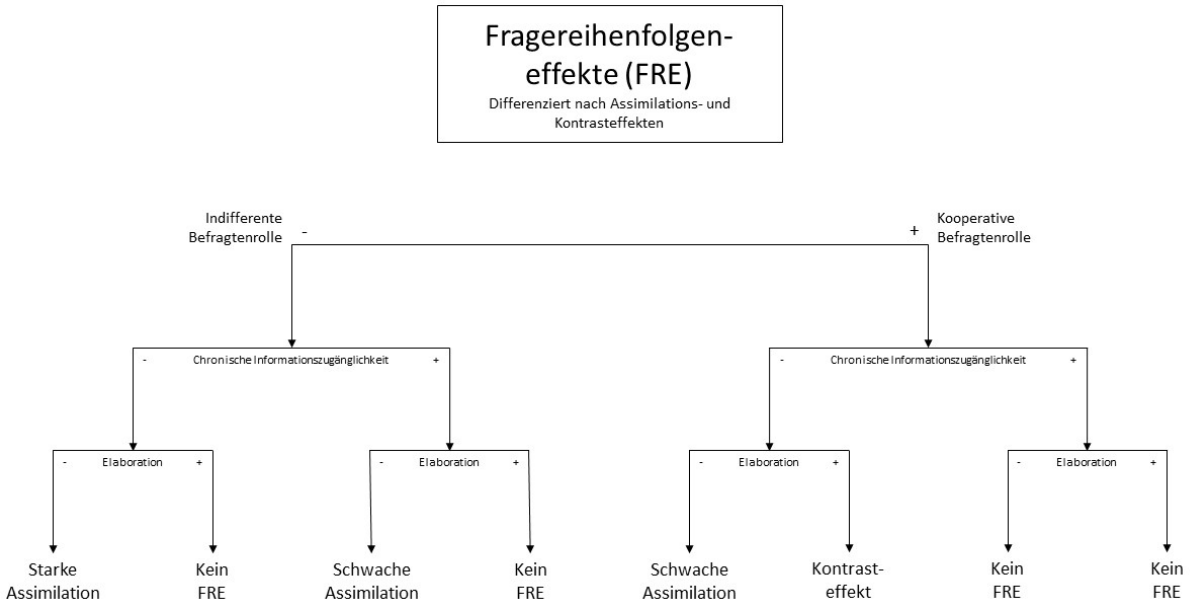
Assimilationseffekte treten demnach umso stärker auf, desto weniger Befragten relevante Informationen chronisch zugänglich sind sowie wenn Befragte eine indifferente Befragtenrolle selektieren. Kontrasteffekte treten entsprechend auf, wenn Befragte eine kooperative Befragtenrolle selektieren und ihnen ebenfalls keine relevanten Informationen chronisch verfügbar sind. Der Einfluss der Informationszugänglichkeit und der Befragtenrolle auf den Assimilationseffekt ist diesbezüglich additiv, der Einfluss auf den Kontrasteffekt interaktiv.

Entsprechend dieser Verdichtung kann nun theoriegeleitet ein vereinfachtes Gesamtmodell zur Erklärung des Auftretens von FRE, welches sowohl Assimilations-, als auch Kontrasteffekte umfasst, entwickelt und hieraus empirisch überprüfbare Hypothesen abgeleitet werden.

Befragte werden demnach bei eintretender Befragungssituation gemäß der sequenziellen Prozessbeschreibung (in Anlehnung an die Modellierung kausaler Pfade) des Modells der Frame Selektion (vgl. Kapitel 2.1.3 bzw. Esser, 2001; Esser, 2006a; Esser, 2006b; Kroneberg, 2005; Kroneberg, 2007; Kroneberg, 2011) zunächst eine Befragtenrolle selektieren, welche gemäß Stocké (2004) kooperativ bzw. konformistisch oder, wie in dieser Arbeit konsistent formuliert,

als kooperativ oder indifferent beschrieben werden kann. Die aktivierte Befragtenrolle beeinflusst nachfolgend entsprechend eines Primings (Esser, 2001; Fazio, 1990; Krishna & Strack, 2017; Kroneberg, 2011) die Interpretation von Fragen und der Befragungssituation, die Wahrnehmung von Normen und akzeptablem Verhalten innerhalb der Situation, die Relevanzbewertung chronischer und akut zugänglicher Informationen für die Urteilsbildung, persönliche Motive für die Teilnahme an der Befragung etc. Nicht von der Befragtenrolle abhängig, jedoch sequenziell mit nachfolgendem Einfluss für das Auftreten von FRE ist die kognitive chronische Zugänglichkeit von Informationen mit Relevanz für die Prozessierung einer Antwort. Und schließlich entscheiden Motivation und Möglichkeit zum elaborierten Antworten (vgl. Kapitel 2.1.4) darüber, ob ein elaborierter, bewusster und überlegter kognitiver Prozess der Antwortgenerierung oder ein unelaborierter, unbewusster und spontaner kognitiver Prozess der Antwortgenerierung selektiert wird, was als notwendige Bedingung für das Auftreten von FRE formuliert wurde. Entsprechend lässt sich das Auftreten von FRE (Assimilations- und Kontrasteffekten) gemäß der folgenden Abbildung darstellen:

Abbildung 3: Bedingungsfaktoren von Fragereihenfolgeneffekten



Theoretisch fundierte, empirisch überprüfbare Hypothesen sowie notwendige Bedingungen lassen sich entsprechend der in Kapitel 2.3.1 formulierten These 6 zum Auftreten von Assimilationseffekten sowie der formulierten These 7 zum Auftreten von Kontrasteffekten sowie den in

Kapitel 2.3.2 zusammengefassten Befunden bezüglich der Interaktion und Gruppierung der Einflussfaktoren ableiten.

Gilt zunächst die notwendige Abhängigkeit der Effekte von der zugrundeliegenden Antwortelaboration, so lässt sich ableiten:

Notwendige Bedingung 1: Geringe Elaboration ist eine notwendige Bedingung für einen Assimilationseffekt.

Notwendige Bedingung 2: Hohe Elaboration ist eine notwendige Bedingung für einen Kontrasteffekt.

Gilt weiterhin bezüglich des Auftretens von Assimilationseffekten These 6:

> *„Assimilationseffekte sind notwendig durch einen geringen Grad an Antwortelaboration bedingt und treten auf, wenn situativ durch die Kontextfrage aktivierte Informationen in den Prozess der Antwortgenerierung einfließen. Dies ist der Fall, wenn solche Informationen akut in einem höheren Maße zugänglich sind als alternative, jedoch für die Beantwortung der Frage relevantere und kognitiv chronisch zugängliche Informationen. Die situative Zugänglichkeit von Informationen wird durch die eingenommene Befragtenrolle bedingt. Für kooperative Befragte sind die für die wahre Beantwortung der Frage benötigten und chronisch verfügbaren Informationen hoch relevant und entsprechend akut zugänglich, für indifferente Befragte sind jene Informationen relevant und zugänglich welche benötigt werden, um Fragen entsprechend eigener Motive zu beantworten“*

und wird darüber hinaus angenommen, dass der Einfluss der Befragtenrolle sowie der Einfluss der chronischen Zugänglichkeit von Informationen additiv statt interaktiv wirkt, so können die folgenden Hypothesen abgeleitet werden:

H_{1a}: Desto geringer die kognitive chronische Zugänglichkeit von Information mit Relevanz für das Objekt der Befragung bei geringer Antwortelaboration, desto mehr nähert sich die Antwort auf eine Zielfrage der Antwort auf die Kontextfrage an.

H_{1b}: Bei indifferenter Befragtenrolle und geringer Antwortelaboration nähert sich die Antwort auf eine Zielfrage der Antwort auf die Kontextfrage an.

Gilt zuletzt für Kontrasteffekte These 7:

> *„Kontrasteffekte sind notwendig durch einen hohen Grad an Antwortelaboration bedingt und treten auf, wenn situativ durch die Kontextfrage aktivierte Informationen bewusst aus dem Prozess der Antwortgenerierung ausgeschlossen werden. Dies ist der Fall, wenn solche situativ aktivierten Informationen akut in einem Höheren Maße zugänglich sind als alternative, jedoch für die Beantwortung der Frage relevantere und kognitiv chronisch zugängliche Informationen und Befragte aufgrund ihrer selektierten Befragtenrolle über das Motiv verfügen, Redundanzen zu vermeiden, was zu einer bewussten Subtraktion der situativ und akut zugänglichen Informationen führt“*

und wird darüber hinaus entsprechend der Diskussion in Kapitel 2.3.2 angenommen, dass der Einfluss der Befragtenrolle sowie der Einfluss der chronischen Zugänglichkeit von Informationen interaktiv statt additiv wirkt, so kann die folgende Hypothese abgeleitet werden:

H_2: Bei kooperativer Befragtenrolle und hoher Antwortelaboration entfernt sich die Antwort auf eine Zielfrage von der Antwort auf die Kontextfrage, desto geringer die kognitive chronische Zugänglichkeit von Information mit Relevanz für das Objekt der Befragung.

Entsprechend ist es das Ziel, eine Möglichkeit zu entwickeln, diese drei theoretisch fundierten Hypothesen sowie die beiden notwendigen Bedingungen empirisch zu testen. Diesbezüglich werden im folgenden Kapitel statistische Verfahren zur Identifikation, Quantifizierung und Kontrolle von FRE präsentiert und diskutiert.

3 Methodischer Ansatz: Statistische Kontrolle von Fragereihenfolgeneffekten

Über die theoretische Diskussion von Fragereihenfolgeneffekten (kurz: „FRE“) hinaus hat deren Auftreten ebenfalls eine Reihe praktischer statistischer und messtheoretischer Implikationen, welche dazu führen, dass empirische Messwerte von den wahren Werten abweichen, wodurch z.B. Messungen der Datenqualität (der Validität und der Reliabilität) verzerrt sind.

So wurde im vergangenen Kapitel bereits unter anderem Diskutiert, dass bei vorhandenen FRE Korrelationen aufeinanderfolgender Items unterschiedlich ausfallen, je nachdem, welches Item innerhalb eines Fragebogens bzw. einer Fragebatterie zuerst präsentiert wird (Bard & Weinstein, 2016; Cunningham et al., 2016; Garbarski et al., 2014; Huang & Cornell, 2015; Magelssen, 2016). Aus dieser Beobachtung ergibt sich zwangsläufig, dass durch das Auftreten von FRE eine Verzerrung der gemessenen Items auf statistischer Ebene stattfindet. Eine Entfernung weg von den wahren Werten und eine geringere Effizienz der empirischen Messungen ist hierdurch implizit. Darüber hinaus zeigen Mayerl und Giehl (2018), dass Messungen der Datenqualität dort positiv verzerrt sind, wo FRE auftreten. Dies erzeugt für Fragebatterien, welche besonders anfällig gegenüber solchen Effekten sind, künstlich erhöhte Messungen der Reliabilität (z.B. Cronbach‘s Alpha) und der Validität (z.B. Faktorladungen).

Die Konsequenzen solcher Verzerrungen sind weitreichend. Folglich kommt der Identifikation, aber auch der Kontrolle von FRE auf statistischer Ebene eine besondere Bedeutung zu. Hierzu soll das vorliegende Kapitel einen Beitrag leisten, in dem ein statistisches Verfahren zur Identifikation und Kontrolle eben dieser Effekte schrittweiße eingeführt und diskutiert wird, die nichtlinear latent kontrollierte konfirmatorische Faktoranalyse (**n**onlinear **l**atent **c**ontrolled **c**onfirmatory **f**actor **a**nalysis, kurz: NLC-CFA). Berücksichtigt werden hierbei insbesondere die Prädiktoren „Modus der Informationsverarbeitung“, „chronische Zugänglichkeit von Informationen bezüglich des Objekts der Befragung“ und „Framing der Befragtenrolle“, welche bereits im vergangenen Kapitel identifiziert und diskutiert wurden. Die besagten Prädiktoren können innerhalb der NLC-CFA darüber hinaus als generisch betrachtet werden, so FRE auf andere Variablen zurückgeführt werden sollen (bezüglich des generischen Charakters der NLC-CFA, siehe Kapitel 5.3). Parallel hierzu wird ein etablierter Ansatz zur Kontrolle von systematischen Messfehlern, die Multi-Trait-Multimethod (MTMM) Analyse diskutiert und aufgezeigt, wie hierauf basierend FRE ebenfalls identifiziert und kontrolliert werden können. Dies hat zum Ziel,

die beiden Verfahren in der nachfolgenden Analyse gegenüberstellend zu vergleichen und hinsichtlich ihrer Effizienz zur Identifikation und Kontrolle von FRE bzw. zur Bereinigung des Einflusses systematischer Messfehler bewerten zu können. Entsprechend ist es das Ziel dieses Kapitels, die NLC-CFA einzuführen sowie je ein MTMM- sowie ein NLC-CFA Modell zu entwickeln, mit welchen FRE identifiziert und kontrolliert werden können.

Hierzu erfolgt zunächst als Einführung in den Themenkomplex eine generelle Diskussion systematischer Messfehler und deren Auswirkungen auf Messungen der Datenqualität im Kontext verschiedener methodischer Paradigmen und Verfahren (Kapitel 3.1) bevor aufgezeigt wird, wie FRE im Speziellen auf Basis verschiedener Verfahren identifiziert und kontrolliert werden können (Kapitel 3.2). In Kapitel 3.2 werden dabei insbesondere die notwendigen Bausteine der NLC-CFA schrittweiße eingeführt. In Kapitel 3.3 wird die NLC-CFA schließlich zusammenfassend diskutiert. Zudem werden die genauen MTMM- und NLC-CFA Modelle vorgestellt, welche in der nachfolgenden Analyse untersucht werden.

3.1 Systematische Messfehler in der empirischen Forschung

FRE gehören entsprechend messtheoretischer Logik zu den systematischen Messfehlern (Bogner & Landrock, 2014). Diesen inhärent ist, dass sie, wenn nicht bereits während des Datenerhebungsprozesses vermieden, lediglich durch statistische Verfahren kontrolliert werden können (Stocké, 2004), wodurch deren Identifikation und Kontrolle, wie bereits beschrieben, eine besondere Bedeutung zukommt. In diesem Kapitel wird daher grundlegend diskutiert, wie (systematische) Messfehler innerhalb unterschiedlicher Paradigmen und Methoden behandelt (und quantifiziert) werden und welche Konsequenzen sich hieraus für Messungen der Datenqualität ergeben. Hierzu werden zunächst Gütekriterien und Möglichkeiten zur Quantifizierung von Messfehlern innerhalb der klassischen Testtheorie als Grundlage aller weiterer Betrachtungen diskutiert (Kapitel 3.1.1). Hierauf aufbauend erfolgt eine Diskussion des Potenzials zur Modellierung von Messfehlern in Strukturgleichungsmodellen (SEM) (Kapitel 3.1.2), einem statistischen Modellierungsverfahren, welches unter anderem in der Einstellungs- und Verhaltensforschung breite Anwendung findet und als Kombination von konfirmatorischer Faktoranalyse (CFA) und Pfadanalyse betrachtet werden kann (Hox & Bechger, 1999). Dabei ist es ebenfalls das Ziel dieses Kapitels, die mathematische Notation einzuführen, welche in der weiteren Arbeit, insbesondere bei der Einführung der NLC-CFA, verwendet wird. In Kapitel 3.1.3 wird weiterhin aufgezeigt, wie systematische Messfehler innerhalb von CFA und SEM auf Basis des

etablierten MTMM Ansatzes modelliert werden können, da diesem Verfahren ebenfalls eine zentrale Bedeutung in dieser Arbeit zukommt, so der Anspruch besteht, dass das einzuführende NLC-CFA Modell zu einer stärkeren Bereinigung von durch FRE bedingte Verzerrungen führen soll als das etablierte MTMM Verfahren. Um die Notwendigkeit der Berücksichtigung systematischer Messfehler vertiefend aufzuzeigen, schließt das Kapitel mit einer Diskussion der Auswirkungen einer fehlenden Behandlung systematischer Messfehler auf Messungen der Datenqualität ab (Kapitel 3.1.4).

3.1.1 Systematische Messfehler innerhalb der klassischen Testtheorie

Allgemein handelt es sich bei der klassischen Testtheorie um eine Theorie der Konstruktion und Interpretation von Testverfahren, welche als Basis vieler aktueller Mess- und Diagnoseinstrumente dient (Moosbrugger, 2012) und auf die Arbeit von Gulliksen (1950) und Lord & Novick (1968) zurückgeht. Die Kerntheorie lautet, dass sich Messwerte einer Person innerhalb eines Items aus zwei Komponenten zusammensetzen: einem „wahren“ Wert und einem Messfehler. Da es die Intention der Umfrageforschung ist, solche wahren Werte zu erheben, ist die Feststellung der Ausprägung des Messfehlers nötig, wofür es testtheoretischer Grundannahmen bedarf, welche als Axiome formuliert sind.

Bei der klassischen Testtheorie handelt es sich somit im Wesentlichen um eine Messfehler-Theorie. Die formulierten Axiome beinhalten alle nötigen Überlegungen, um bei der Messung von Merkmalen mehrerer Befragter auf den wahren Wert schließen zu können (Moosbrugger, 2012, S. 104). Die drei Axiome, zwei Zusatzannahmen sowie deren Implikationen werden im Folgenden nach Moosbrugger (2012) vorgestellt:

$$\tau_{vi} = E(x_{vi}) \qquad \textbf{Formel 6}$$

Das erste Axiom der klassischen Testtheorie (vgl. Formel 6), das sogenannte Existenzaxiom, besagt, dass der wahre Wert τ eines Befragten *v* in Item *i* als Erwartungswert der Messung x_{vi} existiert. Es wird somit implizit die Annahme getroffen, dass jede Messung x_{vi} zumindest einen Anteil des wahren Wertes τ_{vi} enthält, welcher bei ausreichend hoher Fallzahl identifiziert werden kann.

$$x_{vi} = \tau_{vi} + \varepsilon_{vi}\ mit\ E(\varepsilon_{vi}) = 0 \qquad \textbf{Formel 7}$$

Das sogenannte Verknüpfungsaxiom (vgl. Formel 7) beschreibt jede Messung x_{vi} als additive Funktion des wahren Wertes τ_{vi} sowie des zufälligen Messfehlers ε_{vi}. Eine Verbindung der beiden Axiome 1 und 2 führt weiterhin zu dem Schluss, dass der Zufallsfehler ε_{vi} über einen Erwartungswert von 0 verfügt. Der wahre Wert τ_{vi} ergibt sich somit aus der Differenz des Messwertes x_{vi} sowie des Messfehlers ε_{vi}.

$$Corr(\tau_{vi}, \varepsilon_{vi}) = 0 \qquad \textbf{Formel 8}$$

Das Unabhängigkeitsaxiom (vgl. Formel 8) beschreibt, dass keine Korrelation zwischen dem Messfehler ε_{vi} und dem wahren Wert τ_{vi} für beliebige Personen und Items existiert. Dies impliziert eine Unabhängigkeit des Messfehlers ε_{vi} vom wahren Wert τ_{vi}.

$$Corr(\varepsilon_{vi}, \varepsilon_{vj}) = 0 \qquad \textbf{Formel 9}$$

Zusätzlich zur Unkorreliertheit zwischen dem Messfehler ε_{vi} und dem wahren Wert τ_{vi} wird die Zusatzannahme getroffen, dass die Messfehler ε zweier Messungen der Items *i* und *j* der Person *v* unabhängig, das heißt unkorreliert sind (vgl. Formel 9).

$$Corr(\varepsilon_{vi}, \varepsilon_{wi}) = 0 \qquad \textbf{Formel 10}$$

Die letzte getroffene Zusatzannahme (vgl. Formel 10) besagt, dass die Messfehler ε eines Items *i* zwischen verschiedenen Personen *v* und *w* unkorreliert sind.

Es ist somit das Ziel, welches sich aus den oben beschriebenen Axiomen ergibt, den wahren Wert τ_{vi} zu schätzen. Problematisch ist dies, da τ_{vi} selbst nicht beobachtbar ist, sondern sich aus der Differenz des beobachtbaren Wertes x_{vi} und dem Messfehler ε_{vi} ergibt (vgl. Verknüpfungsaxiom in Formel 7), wobei letzterer Ebenfalls nicht direkt beobachtbar ist. Es stellt sich

somit die Frage, wie auf der Basis eines beobachteten Wertes x_{vi} der wahre Wert τ_{vi} geschätzt werden kann.

Eine Lösung hierfür ergibt sich aus der Annahme, dass der Erwartungswert der Messfehler ε_{vi} gleich 0 ist. Gemäß des Verknüpfungsaxiom (vgl. Formel 7) gilt, dass wenn eine Messung x_{vi} unendlich häufig durchgeführt wird und hiervon ein Mittelwert gebildet werden würde, der Mittelwert dieser Messung x_{vi} identisch mit τ_{vi} wäre, da ε_{vi} für dieses Beispiel gleich 0 ist (die Messfehler heben sich gegenseitig auf)[3]. Eine sehr häufige Messung von x_{vi} wäre demnach noch immer annährend identisch mit τ_{vi}, da ε_{vi} annährend 0 wäre.

Jedoch ist eine solche Messung nicht möglich, da eine sehr häufige Wiederholung einer Messung mit der gleichen Person und dem gleichen Messinstrument zu einem Lerneffekt in der Anwendung des Messinstrumentes führen würde, was die Zufälligkeit der Messfehlerverteilung verletzt.

Eine Lösung für dieses Problem ergibt sich daraus, dass für einen Befragten *v* nicht immer wieder das gleiche Item *i* gemessen wird, sondern *m* verschiedene Items, welche jedoch alle das gleiche Persönlichkeitsmerkmal erfassen. Durch eine Kombination aller dieser Messungen x_{vi} zu einem einzelnen Testwert x_v (vgl. Formel 11) neutralisieren sich theoretisch auch hier die Messfehler gegenseitig.

$$x_v = \sum_{i=1}^{m} x_{vi} \qquad \textbf{Formel 11}$$

Aus Axiom 1 (vgl. Formel 6) ist nun bekannt, dass der Erwartungswert der Messung x_{vi} dem wahren Wert τ_{vi} entspricht. Zudem lässt sich zeigen, dass die Summe der Erwartungswerte der Messungen x_{vi} dem Erwartungswert des Testwertes x_v entspricht (vgl. Formel 12).

$$E(x_v) = E\left(\sum_{i=1}^{m} x_{vi}\right) = \sum_{i=1}^{m} E(x_{vi}) = \sum_{i=1}^{m} \tau_{vi} = \tau_v \qquad \textbf{Formel 12}$$

[3] Dies bedingt jedoch, dass ε_{vi} tatsächlich auch ein Zufallsfehler und unabhängig von weiteren Messungen (vgl. Formel 8) und Messfehlern (vgl. Formel 9) ist. Hierauf wird später in diesem Kapitel sowie in Kapitel 3.1.2 eingegangen.

Wird nun davon ausgegangen, dass der Erwartungswert des Testwerts x_v dem wahren Wert τ_v entspricht, dann kann die Messwertsumme x_v als Punktschätzung $\hat{\tau}_v$ des wahren Wertes τ_v einer Person *v* verwendet werden (vgl. Formel 13).

$$x_v = \hat{\tau}_v \qquad \textbf{Formel 13}$$

Aufgrund dessen, dass es sich bei $\hat{\tau}_v$ um eine Schätzung auf Basis empirischer Daten handelt, besteht jedoch Unsicherheit über dessen Übereinstimmung mit dem wahren Wert τ_v. Daher ist es ratsam, zusätzlich einen Standardmessfehler zu konstruieren, mit dessen Hilfe ein Konfidenzintervall um $\hat{\tau}_v$ gebildet werden kann. Hierfür muss jedoch zunächst die Testwertvarianz in den Anteil wahrer Varianz und den Anteil der Fehlervarianz auf Basis des Verknüpfungsaxioms (vgl. Formel 7) zerlegt werden. Hierbei ist zu beachten, dass sich die Varianz einer Summe von Variablen aus der Summe der Varianzen der einzelnen Variablen plus der zweifachen Summe der Kovarianz beider Variablen ergibt (vgl. Formel 14).

$$Var(x) = Var(\tau + \varepsilon) = Var(\tau) + Var(\varepsilon) + 2Cov(\tau, \varepsilon) \qquad \textbf{Formel 14}$$

Wird nun das Unabhängigkeitsaxiom (vgl. Formel 8) berücksichtigt, nach welchem die Korrelation zwischen wahrem Wert und Messfehler gleich 0 ergibt, so muss auch $Cov(\tau, \varepsilon)$ gleich 0 sein. Hieraus ergibt sich:

$$Var(x) = Var(\tau) + Var(\varepsilon) \qquad \textbf{Formel 15}$$

Es zeigt sich somit, dass sich die Varianz von x in eine „wahre“ Varianz von τ, welche sich aus den unterschiedlichen Merkmalsausprägungen der verschiedenen befragten Personen ergibt, sowie der Fehlervarianz $Var(\varepsilon)$, welche auf zufällige Messfehler bei den befragten Personen zurück zu führen ist, zerlegt werden kann.

Um nun die Varianzen $Var(\tau)$ und $Var(\varepsilon)$ zu schätzen, kann die Covarianz der Testwertvariablen x_p und x_q zweier beliebiger Tests *p* und *q* betrachtet und aufgrund der Unkorreliertheit

der Messfehler mit den wahren Werten sowie zwischen den Messfehlern wie folgt aufgelöst werden:

$$\begin{aligned} Cov(x_p, x_q) &= Cov(\tau_p + \varepsilon_p, \tau_q + \varepsilon_q) \\ &= Cov(\tau_p, \tau_q) + Cov(\tau_p, \varepsilon_q) + Cov(\varepsilon_p, \tau_q) + Cov(\varepsilon_p, \varepsilon_q) \\ &= Cov(\tau_p, \tau_q) \end{aligned} \qquad \textbf{Formel 16}$$

Damit ist gezeigt, dass die Kovarianz zweier Messungen x_p und x_q gleich der Kovarianz der wahren Werte τ_p und τ_q ist, da auch für Formel 16 gilt, dass die Kovarianz zwischen wahren Werten und Messfehlern sowie lediglich zwischen Messfehlern 0 beträgt. Werden mit x_p und x_q nun Messungen zweier Testwertvariablen mit τ-äquivalenten Tests durchgeführt, so lässt sich unter der Bedingung $\tau_p = \tau_q = \tau$ die Covarianz von x_p und x_q darstellen als wahre Varianz τ (vgl. Formel 17).

$$Cov(x_p, x_q) = Cov(\tau_p, \tau_q) = Cov(\tau, \tau) = Var(\tau) \qquad \textbf{Formel 17}$$

Die Fehlervarianz $Var(\varepsilon)$ ergibt sich demnach schließlich auf Basis von Formel 15 aus der Differenz der Varianz von x und der bekannten wahren Varianz von τ:

$$Var(\varepsilon) = Var(x) - Var(\tau) = Var(x) - Cov(x_p, x_q) \qquad \textbf{Formel 18}$$

Als Zwischenfazit dieser auf Basis von Moosbrugger (2012) geführten Erörterung lässt sich somit festhalten, dass sich jede Messung eines Merkmals x aus einem wahren Wert τ und einem Messfehler ε ergibt. Wird das Merkmal *x* nun auf der Basis mehrerer Messwerte bestimmt, so kann eine Punktschätzung des wahren Wertes von τ bestimmt werden, der Messfehler ε ergibt sich folglich aus der Differenz des Messwertes *x* und des wahren Wertes τ. Ähnliches gilt für die Varianz von τ und ε: Wird *x* durch mehrere Messwerte bestimmt, so ergibt sich die wahre Varianz von τ aus der Kovarianz der verschiedenen Messwerte. Die Fehlervarianz von ε ergibt sich im Folgenden aus der Differenz der Varianz von *x* und der wahren Varianz von τ.

Auf Basis dieser Schlussfolgerungen ist nun die Möglichkeit gegeben, die Güte eines Tests zu quantifizieren, in dem die wahre Varianz von τ mit der gemessenen Varianz von *x* ins Verhältnis gesetzt wird. Definitorisch entsprich dieses Verhältnis der sogenannten Reliabilität, das zentrale Gütekriterium der klassischen Testtheorie.

Hartig und Jude (2007, S. 21f) verstehen unter der Reliabilität die Messgenauigkeit einer Erhebung, wobei sie diese genauer definieren als den Anteil der Varianz an der beobachteten Testwertvarianz, welche tatsächlich auf Unterschiede im untersuchten Merkmal zurückzuführen ist. Krebs und Menold (2014, S. 427) verstehen unter der Reliabilität eines Messinstrumentes das „Ausmaß, in dem wiederholte Messungen eines Einstellungsobjektes zu gleichen Werten führen", wobei die Messwiederholungen unabhängig voneinander sein müssen. Diekmann (2007) definiert Reliabilität simpler als das Maß für die Reproduzierbarkeit von Messergebnissen (für eine weiterführende Diskussion der Reliabilität, siehe Kapitel 3.1.4). Nach der Definition der klassischen Testtheorie ergibt sich die Reliabilität als Anteil der Varianz des wahren Wertes τ an der Varianz des beobachteten Testwertes von *x* (Moosbrugger, 2012, S. 110) und kann wie folgt dargestellt werden:

$$Rel = \frac{Var(\tau)}{Var(x)} \qquad \textbf{Formel 19}$$

Der sich ergebende Reliabilitätskoeffizient nimmt folglich einen Wertebereich von 0 bis 1 an, wobei ein Wert von 1 bedeutet, dass die Testwertvarianz ausschließlich aus wahrer Varianz besteht, während ein Wert von 0 bedeutet, dass die Testwertvarianz ausschließlich aus Fehlervarianz besteht. Ein Test ist demnach umso genauer, je größer der wahre Varianzanteil an der gesamten Varianz ist (ebd.).

Aus Kapitel 2 und der Diskussion rund um systematisch auftretende Messfehler wie dem FRE geht nun jedoch als fundamentale Kritik an der klassischen Testtheorie hervor, dass Messfehler nicht zwangsläufig zufällig sind (sondern ebenfalls nicht-zufällig, das heißt systematisch auftreten), wodurch de facto ein Verstoß gegen $E(\varepsilon_{vi}) = 0$ (Verknüpfungsaxiom, vgl. Formel 7) aufgezeigt ist. Die Implikation hiervon ist folgenreich: Erstens führt die Summierung der Messungen x_{vi} zu einem einzelnen Testwert x_v potentiell zu einem verzerrten Punktschätzer $\hat{\tau}_v$, da sich die Messfehler ε_{vi} dort nicht zwangsläufig neutralisieren, wo diese nicht zufällig sind (vgl. Formel 20).

$$x_v = \sum_{i=1}^{m} x_{vi} \neq \hat{\tau}_v \qquad \textbf{Formel 20}$$

Hieraus ergibt sich, in Anlehnung an das Unabhängigkeitsaxiom (vgl. Formel 8) und unter der Berücksichtigung von FRE, eine potentielle Abhängigkeit der Messfehler einer Messung *j* von der Messung *i* (vgl. Formel 21).

$$Corr(x_{vi}, \varepsilon_{vj}) \neq 0 \qquad \textbf{Formel 21}$$

Gilt nun, dass sich der Messwert *x* aus dem wahren Wert τ und einem (nicht zwangsläufig zufälligen) Messfehler ε zusammensetzt, so folgt aus Formel 21, dass der Messfehler ε_{vj} ebenfalls nicht unabhängig sein muss vom Messfehler einer vorangegangenen Messung ε_{vi}, wodurch die Zusatzannahme der Unabhängigkeit der Messfehler untereinander (vgl. Formel 9) ebenfalls dort verletzt ist, wo nicht-zufällige, sprich, systematische Messfehler auftreten, dargestellt in Formel 22.

$$Corr(\varepsilon_{vi}, \varepsilon_{vj}) \neq 0 \qquad \textbf{Formel 22}$$

Dies impliziert weiter, dass die Varianz der Messung *x* nicht mehr, wie in Formel 14 und Formel 15 beschrieben, effektiv nach $Var(x) = Var(\tau) + Var(\varepsilon)$ aufgelöst werden kann mit der Folge, dass auch die Schätzung der Varianz von τ und ε verzerrt ist. In letzter Konsequenz bedeutet dies, dass die Berechnung der Reliabilität (vgl. Formel 19), dem zentralen Gütekriterium der klassischen Testtheorie, verzerrt ist, wo Messfehler nicht zufällig sind.

Zusammenfassend bedeutet dies, dass Punktschätzungen von Testwerten, deren Varianz, sowie Berechnungen der Reliabilität, welche auf Verfahren beruhen, die den Annahmen der klassischen Testtheorie folgen, dort verzerrt sind, wo systematische Messfehler wie etwa FRE auftreten. Um folglich zu unverzerrten Schätzungen zu gelangen, wird ein Verfahren benötigt, welches das Auftreten systematischer Messfehler und insbesondere von FRE durch eine Be-

rücksichtigung von $E(\varepsilon_{vi}) \neq 0$ würdigt, wodurch diese, wie es dem Ziel dieser Arbeit entspricht, identifiziert und kontrolliert werden können. Ein solches Verfahren ist in der sogenannten Strukturgleichungsmodellierung auf Basis latenter Variablenmodelle gegeben, auf welches im Folgenden eingegangen wird.

3.1.2 Systematische Messfehler in Strukturgleichungsmodellen

Allgemein betrachtet wird mit der Strukturgleichungsmodellierung (kurz: SEM) eine Technik beschrieben, mit welcher Kovarianzen bzw. Regressions- oder Pfadkoeffizienten zwischen latenten Konstrukten und/oder manifesten Variablen statistisch modelliert werden können, wodurch diese als Kombination von konfirmatorischer Faktoranalyse (kurz: CFA) und Regressions- bzw. Pfadanalyse betrachtet werden kann (Hox & Bechger, 1999). So wird typischerweise zunächst eine CFA durchgeführt, um ein Messmodell zu bestimmen, bevor Pfade, Korrelationen/Kovarianzen und/oder regressive Beziehungen zwischen den einzelnen Komponenten modelliert werden, sprich, das Strukturmodell spezifiziert wird. Im Kontext der bisher diskutierten Messfehler bietet SEM über diese Kombination verschiedener Techniken hinaus den Vorteil, dass mit ihr sogenannte messfehlerbereinigte Analysen durchführt werden können. Dies bedeutet, dass die Varianz eines jeden Items bei der Schätzung von Einflussbeziehungen zerlegt wird in einen Anteil valider Varianz, einen Anteil systematischer Fehlervarianz und einen Anteil zufälliger Fehlervarianz. Während die valide Varianz und die systematische Fehlervarianz durch Gleichungssysteme modelliert und geschätzt werden, geht die zufällige Fehlervarianzen als Residuum aus der Schätzung hervor (Urban & Mayerl, Strukturgleichungsmodellierung. Ein Ratgeber für die Praxis, 2014). Entsprechend wird das SEM Verfahren im Folgenden im Kontext der Modellierung von Messfehlern aufgrund seiner zentralen Bedeutung für diese Arbeit vorgestellt, auch, um die mathematische Notation einzuführen, welche weiterhin verwendet wird. Es schließt zudem eine Diskussion verschiedener Schätz-Algorithmen des Verfahrens an, welche insbesondere für die Analysen im nächsten Abschnitt dieser Arbeit relevant sind.

Basale Logik und Notation von SEM

Die grundlegende Logik von SEM basiert auf dem sogenannten „Common Factor Model“ (Thurstone, 1947), welches postuliert, dass jeder manifeste Indikator eines Messinstrumentes als lineare Funktion eines oder mehrerer „Common Factors“ und einem „Unique Factor“ dargestellt werden kann, wobei die valide Varianz auf den Common Factor zurückgeht, während

der Unique Factor in die systematische Fehlervarianz sowie die zufällige Fehlervarianz zerlegt wird (Brown, Confirmatory Factor Analysis for Applied Research, 2006). Entsprechend wird die fundamentale Gleichung des Common Factor Model (sowie weitere SEM-Spezifika bis einschl. Formel 25) nach Brown (ebd.) in Formel 23 dargestellt:

$$y_j = \lambda_{j1}\eta_1 + \lambda_{j2}\eta_2 + \cdots + \lambda_{jm}\eta_m + \varepsilon_j \qquad \textbf{Formel 23}$$

y_j repräsentiert die Messung von Indikator *j* aus insgesamt *p* Indikatoren, λ_{jm} steht für die Faktorladung von Indikator *j* auf das *k*-te latente Konstrukt η, wobei *m* die Gesamtzahl aller latenter Konstrukte symbolisiert. ε_j zuletzt stellt die Unique Variance des Indikators *j* dar, in ihm bündeln sich die systematische sowie die zufällige Fehlervarianz. Diese Gleichung kann somit, um den Zusammenhang zwischen dem *p* x 1 Vektor $\boldsymbol{y}$, dem 1 x *m* Vektor $\boldsymbol{\eta}$, dem *p* x 1 Vektor $\boldsymbol{\varepsilon}$ sowie der *p* x *m* Faktorladungsmatrix $\boldsymbol{\Lambda_y}$ darzustellen, wie folgt vereinfacht werden:

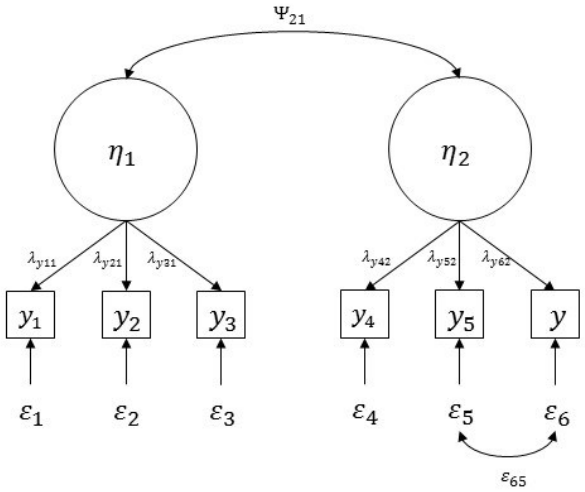
$$\boldsymbol{y} = \boldsymbol{\Lambda_y}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \qquad \textbf{Formel 24}$$

Zudem kann, da SEM mittels Matrix-Algebra berechnet werden, die in Formel 24 dargestellte Gleichung mittels Matrixschreibweise wie folgt dargestellt werden:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda_y}\boldsymbol{\Psi}\boldsymbol{\Lambda_y'} + \boldsymbol{\Theta_\varepsilon} \qquad \textbf{Formel 25}$$

$\boldsymbol{\Sigma}$ entspricht hier der symmetrischen *p* x *p* Kovarianz-Matrix der *p* Indikatoren *y*. $\boldsymbol{\Lambda_y}$ stellt erneut die *p* x *m* Matrix der Faktorladungen λ dar. $\boldsymbol{\Psi}$ steht für die symmetrische *m* x *m* Korrelations-Matrix der *m* latenten Konstrukte η und $\boldsymbol{\Theta_\varepsilon}$ symbolisiert die diagonale *p* x *p* Matrix der Unique Variances von ε.

Abbildung 4: SEM Y-Notation



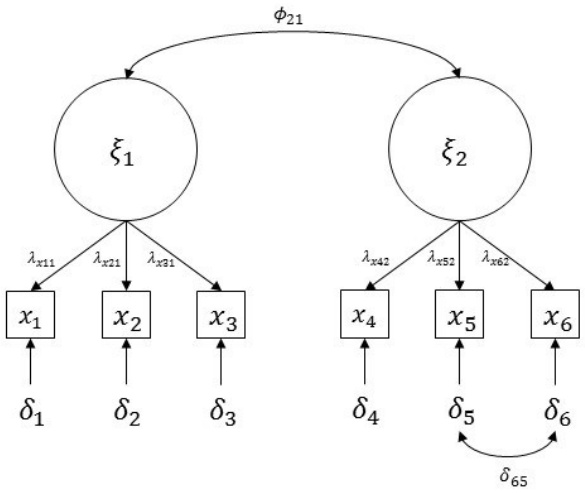
Name	Parameter	Matrix	Typ	Beschreibung
Lambda-Y	λ_y	Λ_y	Regression	Faktorladungen
Theta epsilon	ε	Θ_ε	(Co)Varianz	(Co)Varianz der Messfehler
Psi	ψ	Ψ	(Co)Varianz	(Co)Varianz der Faktoren
Tau-Y	τ_y		Mittelwert Vektor	Indikator Konstante
Alpha	α		Mittelwert Vektor	Latenter Mittelwert
Eta	η		Vektor	Name endogener Variablen

Quelle: Brown, 2006, S. 56

Zusätzlich gilt zu berücksichtigen, dass die hier beschriebene Schreibweise der sogenannten Y-Notation für endogene latente Konstrukte entspricht (vgl. Abbildung 4). Darüber hinaus findet sich ebenfalls die X-Notation für exogene latente Konstrukte. Da in den folgenden Kapiteln beide Schreibweisen Anwendung finden, wird die X-Notation im Folgenden (vgl. Abbildung 5) ebenfalls tabellarisch inklusive Abbildung zur vereinfachten Darstellung (nach Brown, 2006, S. 55f) aufgezeigt.

Abbildung 4 und Abbildung 5 zeigen ein Faktorenmodell mit *m* = 2 latenten Konstrukten, welche mittels jeweils drei Indikatoren ($p = 6$) gemessen werden. Beide latente Konstrukte verfügen jeweils über eine Varianz, sowie eine Kovarianz zwischen einander. Zu beachten ist in dieser Abbildung die Kovarianz zwischen den Unique Variances 5 und 6. Denn während ε_5 (δ_5) und ε_6 (δ_6) hier die Anteile zufälliger Fehlervarianz der Indikatoren 5 und 6 darstellen, wird mit der Korrelation der Fehlervarianzen ε_{65} (δ_{65}) deren Kovarianz als systematische Fehlervarianz spezifiziert (vgl. Kapitel 3.1.3).

Abbildung 5: SEM X-Notation



Name	Parameter	Matrix	Typ	Beschreibung
Lambda-X	λ_x	Λ_x	Regression	Faktorladungen
Theta delta	δ	Θ_δ	(Co)Varianz	(Co)Varianz der Messfehler
Phi	ϕ	Φ	(Co)Varianz	(Co)Varianz der Faktoren
Tau-X	τ_x		Mittelwert Vektor	Indikator Konstante
Kappa	κ		Mittelwert Vektor	Latenter Mittelwert
Xi	ξ		Vektor	Name exogener Variablen

Quelle: Brown 2006, S. 55

Varianzdekomposition in SEM

Es zeigt sich also, dass während sich Messungen des Merkmals bzw. der Variable x_j (die folgende Ausführung gilt ebenfalls analog für y_j) gemäß klassischer Testtheorie in einen wahren Wert τ sowie einen Messfehler ε zerlegen lassen (vgl. Formel 7 und Abbildung 6.1), sich Messungen der Variable x_j innerhalb der SEM Logik in einen Anteil Common Variance $\lambda_{jm}\xi_m$ und einen Anteil Unique Variance δ_j zerlegen (vgl. Abbildung 6.2).

Da es sich bei der Common Variance per Definition um diejenige valide Varianz handelt, welche auf das zu untersuchende Merkmal von Interesse zurückgeht (den zugrundeliegenden Common Factor), kann dieser im Sinne der klassischen Testtheorie als wahrer Wert τ einer Messung x_j identifiziert werden. Die Unique Variance δ_j innerhalb von SEM würde demnach dem Messfehler ε des Verknüpfungsaxioms der klassischen Testtheorie entsprechen.

Abbildung 6: Varianzzerlegung von x (und analog für y)

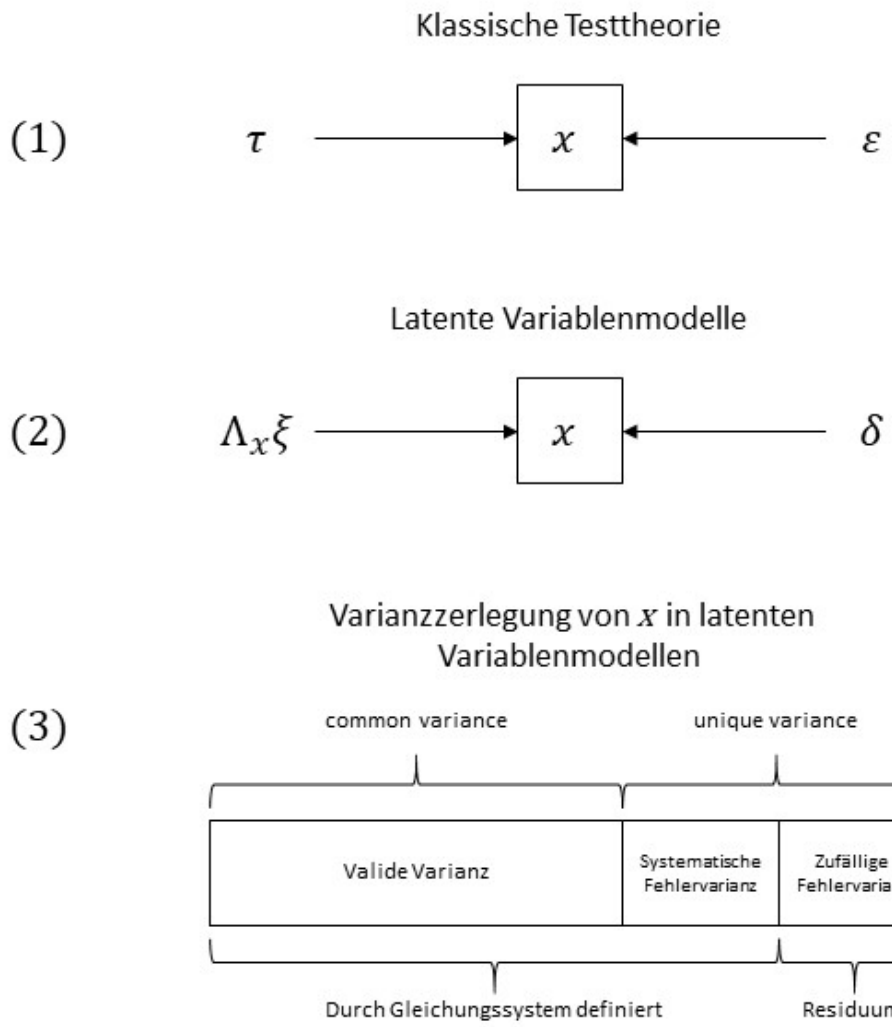


Abbildung 6.3 adaptiert nach Kline, 2016

Standardmäßig wird bei der Spezifikation von Messmodellen zwar angenommen, dass Fehlervarianzen ε_j zufällig sind, da in Anlehnung an die klassische Testtheorie die Annahme zugrunde liegt, dass alle beobachtbaren Beziehungen zwischen zwei beliebigen Indikatoren alleine auf deren gemeinsame Abhängigkeit von dem latenten Konstrukt zurückgehen (Reinecke & Pöge, 2010). Die abgebildete Kovarianz ε_{65} (vgl. Abbildung 4) bzw. δ_{65} (vgl. Abbildung 5) kann dennoch unter der Annahme spezifiziert werden, dass ein Teil der Kovarianz zwischen Indikator fünf und sechs über deren gemeinsame Abhängigkeit vom latenten Konstrukt hinaus auf andere, systematische Ursachen zurückgeführt wird, was als Anteil systematischer Fehlervarianz innerhalb von x_5 und x_6 dargestellt werden kann (vgl. Abbildung 6.3). Die Spezifikation

einer Kovarianz auf Messfehlerebene ist somit gerechtfertigt, wenn das Vorhandensein von unter anderem Methodeneffekten (Revers oder ähnlich formulierte Testitems; Anforderungscharakteristika; Akquieszenz; soziale Erwünschtheit; FRE etc.) innerhalb von x_j begründet angenommen werden kann (vgl. z.B. Brown, 2003; Brown, 2006; Marsh, 1996; Kapitel 3.1.3). Hieraus geht hervor, dass wenn FRE als systematisch auftretende Methodeneffekte definiert werden, diese als Kovarianzen auf Messfehlerebene modelliert werden können (eine vertiefende Diskussion dieser Möglichkeit erfolgt in Kapitel 3.1.3).

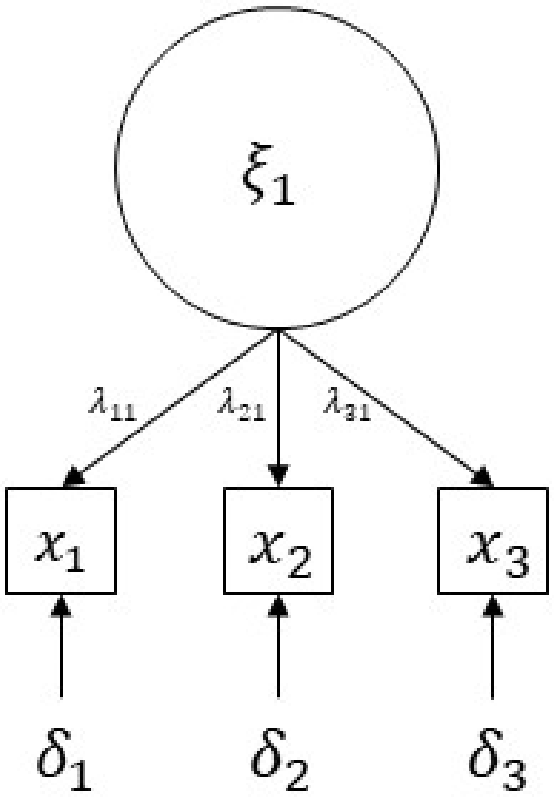
Durch die Spezifikation von Methodeneffekten bzw. Kovarianzen auf Messfehlerebene wird derjenige Anteil der Varianz, der auf die zufällige Fehlervarianz (und nicht auf die systematische Kovarianz zwischen Items) zurückgeht, tatsächlich als zufällig spezifiziert, da der systematische, nicht-zufällige Anteil der Fehlervarianz als Kovarianz separat modelliert ist. Die Annahme $E(\varepsilon_{vi}) = 0$ des Verknüpfungsaxioms (vgl. Formel 7) der klassischen Testtheorie bleibt somit unter der Bedingung erfüllt, dass systematische Messfehler spezifisch modelliert sind. Entgegen der Kritik an der klassischen Testtheorie (vgl. Kapitel 3.1.1) bleiben Punktschätzungen, Standardfehler und Reliabilitätsmessungen unter Berücksichtigung der Spezifikation solcher Kovarianzen entsprechend effizient.

Schätzung von SEM

Da der Logik, Formalisierung und Schätzung systematischer Messfehler mittels SEM somit eine besondere Bedeutung zukommt, wird diese im Folgenden vertiefend diskutiert, beginnend mit einer allgemeinen Übersicht zum Verfahren der Schätzung von SEM, welches Anhand des in Abbildung 7 aufgezeigten CFA Messmodells mit einem latentem Konstrukt (m = 1) und drei Indikatoren (p = 3) beispielhaft gemäß der X-Notation dargestellt wird. Für eine vereinfachte Übersicht des Verfahrens und der mathematischen Herleitungen siehe Reinecke und Pöge (2010), für eine vertiefende Übersicht siehe Bollen (1989).

In einem ersten Schritt zur Schätzung des Messmodells werden die dem Modell zugrundeliegenden Messgleichungen in Anlehnung an Formel 23 dargestellt, wobei (ohne weitere Spezifikation) $E(\delta_j) = 0$ (für alle j = 1, 2, 3) sowie $Cov(\delta_1, \delta_2) = Cov(\delta_1, \delta_3) = Cov(\delta_2, \delta_3) = 0$ in Anlehnung an Formel 7 und Formel 8 der klassischen Testtheorie Formel 26 gilt.

Abbildung 7: Einfaches Messmodell mit *p* = 3 und *m* = 1 (X-Notation)



$$x_1 = \lambda_{11}\xi_1 + \delta_1$$
$$x_2 = \lambda_{21}\xi_1 + \delta_2$$
$$x_3 = \lambda_{31}\xi_1 + \delta_3$$

Formel 26

Weiterhin existiert bei Schätzung des Messmodells mit $p = 3$ Indikatoren die empirische Kovarianzmatrix **S**, welche für dieses Beispiel aus sechs Elementen besteht (eine Varianz je Indikator $\sigma_{x_j}^2$ sowie drei Kovarianzen σ_{x_j,x_k} für die Beziehungen der drei Indikatoren untereinander) und die empirisch beobachteten Varianzen von und Kovarianzen zwischen x_1, x_2 und x_3 enthält:

$$\boldsymbol{S} = \begin{bmatrix} \sigma_{x_1}^2 & & \\ \sigma_{x_2,x_1} & \sigma_{x_2}^2 & \\ \sigma_{x_3,x_1} & \sigma_{x_3,x_2} & \sigma_{x_3}^2 \end{bmatrix}$$

Formel 27

Auf Basis dieser empirischen Kovarianzmatrix **S** sind drei Faktorladungen λ_{j1}, drei Messfehler δ_j sowie die Varianz der latenten Variable ξ_1, welche als ϕ_{11} bezeichnet wird, zu Schätzen[4]. Es liegen somit sieben zu schätzende Parameter vor, während **S** lediglich sechs Elemente enthält, weshalb innerhalb der SEM-Logik entweder die Varianz der latenten Variable ϕ_{11} oder eine der Faktorladungen λ_{j1} auf 1 fixiert wird. Wird für dieses Beispiel $\phi_{11} = 1$ gesetzt und das Gleichungssystem in Formel 26 nach Varianz und Kovarianz aufgelöst, so kann jedes Element der Kovarianzmatrix entsprechend des Gleichungssystems aus Formel 26 in seine Parameter zerlegt werden, so dass lediglich sechs zu schätzende Parameter verbleiben:

$$\sigma_{x_1}^2 = \lambda_{11}^2 + \sigma_{\delta_1}^2$$
$$\sigma_{x_2}^2 = \lambda_{21}^2 + \sigma_{\delta_2}^2$$
$$\sigma_{x_3}^2 = \lambda_{31}^2 + \sigma_{\delta_3}^2$$
$$\sigma_{x_2,x_1} = \lambda_{11}\lambda_{21}$$
$$\sigma_{x_2,x_3} = \lambda_{21}\lambda_{31}$$
$$\sigma_{x_1,x_3} = \lambda_{11}\lambda_{31}$$

Formel 28

Auf Basis dieser Schätzformeln wird nachfolgend die geschätzte Kovarianzmatrix $\boldsymbol{\Sigma}$ erstellt (vgl. Formel 29), auf deren Basis die drei Faktorladungen λ_{j1} sowie drei Messfehler δ_j bei überidentifizierten Modellen iterativ durch Einsetzen von Startwerten für die Parameter so geschätzt werden, so dass die Residual-Kovarianzmatrix $\boldsymbol{S - \Sigma}$ minimiert wird. Für gerade identifizierte Modelle (wie das vorliegende) können die Parameter hingegen algebraisch auf Basis von $\boldsymbol{S}$ bestimmt werden, die Differenz $\boldsymbol{S - \Sigma}$ ergibt dann stets 0. Dabei gilt, dass je Größer die

[4] Grundsätzlich gilt, dass die zu schätzenden Parameter nur dann ermittelt werden können, wenn deren Anzahl gleich oder kleiner der Anzahl der Elemente innerhalb der empirischen Kovarianzmatrix ist. Ist die Anzahl kleiner, so ist die Rede von überidentifizierten Modellen, wobei die Differenz aus verfügbaren Informationen und zu schätzenden Parametern die Freiheitsgrade des Gesamtmodells darstellen. Ist die Anzahl gleich, so ist das Modell gerade identifiziert. Ist die Anzahl größer, so ist das Modell unteridentifiziert und kann nicht geschätzt werden. Die Anzahl der Elemente innerhalb der empirischen Kovarianzmatrix kann mithilfe der Gleichung (p*(p+1))/2 ermittelt werden. Für die Anzahl der zu schätzender Parameter werden alle durch das jeweilige Modell spezifizierten Faktorladungen, Messfehler δ_j bzw, ε_j, Varianzen der latenten Konstrukte und Kovarianzen/Effekte zwischen latenten und/oder manifesten Variablen addiert und hiervon die Anzahl an latenten Konstrukten subtrahiert (Brown, 2006). Für das beispielhafte Modell in Abbildung 7 ergeben sich für p = 3 Indikatoren somit (3*4)/2 = 6 Elemente innerhalb der empirischen Kovarianzmatrix (drei Varianzen und drei Kovarianzen, vgl. Formel 27) und 6 zu schätzende Parameter (3 Faktorladungen + 3 Messfehler + 1 Varianz des latenten Konstruktes – 1 latentes Konstrukt, vgl. Formel 28), wodurch das Modell gerade identifiziert ist (FG = 6 – 6 = 0).

Differenz von $\boldsymbol{S}$ - $\boldsymbol{\Sigma}$ (bei überidentifizierten Modellen) ist, desto eher wird das Modell von den Daten in $\boldsymbol{S}$ wiederlegt.

$$\boldsymbol{\Sigma} = \begin{bmatrix} \hat{\sigma}_{x_1}^2 & & \\ \hat{\sigma}_{x_2,x_1} & \hat{\sigma}_{x_2}^2 & \\ \hat{\sigma}_{x_3,x_1} & \hat{\sigma}_{x_3,x_2} & \hat{\sigma}_{x_3}^2 \end{bmatrix} = \begin{bmatrix} \lambda_{11}^2 + \sigma_{\delta_1}^2 & & \\ \lambda_{11}\lambda_{21} & \lambda_{21}^2 + \sigma_{\delta_2}^2 & \\ \lambda_{21}\lambda_{31} & \lambda_{11}\lambda_{31} & \lambda_{31}^2 + \sigma_{\delta_3}^2 \end{bmatrix} \qquad \textbf{Formel 29}$$

Bei der geschätzten Kovarianzmatrix $\boldsymbol{\Sigma}$ handelt es sich demnach um eine Funktion der Parameter des zu schätzenden Strukturgleichungsmodells, wobei besagte Parameter innerhalb des Parametervektors $\boldsymbol{\Theta}$ zusammengefasst werden. Matrix $\boldsymbol{\Sigma}(\boldsymbol{\Theta})$ enthält somit alle zu schätzenden Parameter des jeweiligen Strukturgleichungsmodells. Für eine vertiefende Herleitung und Beschreibung der Auflösung der Matrix, siehe Bollen (1989).

SEM Schätzfunktionen

Wie schon beschrieben ist es nach Definition der Kovarianzmatrix $\boldsymbol{\Sigma}(\boldsymbol{\Theta})$ das Ziel, die Parameter $\boldsymbol{\Theta}$ qua Minimierung der Diskrepanzfunktion $\boldsymbol{S}$ - $\boldsymbol{\Sigma}$ zu ermitteln, wobei besagte Parameter durch iteratives Einsetzen von Startwerten ermittelt werden, da für das Gleichungssystem überidentifizierter Modelle keine analytisch eindeutige Lösung existiert. Die Höhe der Diskrepanz zwischen $\boldsymbol{S}$ und $\boldsymbol{\Sigma}$ ist somit zugleich ein Indikator dafür, wie gut ein postuliertes Modell durch die gegebenen empirischen Daten unterstützt wird.

Zur Quantifizierung der Diskrepanz stehen verschiedene Schätzfunktionen zur Verfügung, welche für unterschiedliche Bedingungen hinsichtlich der Datenstruktur unterschiedlich geeignet sind, die Diskrepanz zu minimieren und zu einer effizienten Schätzung der Parameter zu gelangen. Dem Verständnis hierum kommt im analytischen Teil dieser Arbeit (vgl. Kapitel 4 und insbesondere Kapitel 4.3) eine hohe Bedeutung zu, weshalb die wesentlichen Schätzfunktionen von CFA bzw. SEM im Folgenden vorgestellt und diskutiert werden.

Häufig verwendete Schätzfunktionen sind hierbei die Maximum-Likelihood(ML)-Funktion (vgl. Formel 30.1), die Unweighted-Least-Square(ULS)-Funktion (vgl. Formel 30.2), die Generalized-Least-Square(GLS)-Funktion (vgl. Formel 30.3) sowie die Weighted-Least-Square(WLS)-Funktion (vgl. Formel 30.4).

$$(1)\ F_{ML} = log\|\boldsymbol{\Sigma}(\boldsymbol{\Theta})\| + tr\left(\boldsymbol{S}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Theta})\right) - \log\|\boldsymbol{S}\| - (p+q)$$

$$(2)\ F_{ULS} = \frac{1}{2}tr\left(\boldsymbol{S} - \boldsymbol{\Sigma}(\boldsymbol{\Theta})\right)^{\mathbf{2}}$$

Formel 30

$$(3)\ F_{GLS} = \frac{1}{2}tr[(\boldsymbol{S} - \boldsymbol{\Sigma}(\boldsymbol{\Theta}))\boldsymbol{W}^{-1}]^2$$

$$(4)\ F_{WLS} = [\boldsymbol{s} - \boldsymbol{\sigma}(\boldsymbol{\Theta})]'\boldsymbol{W}^{-1}[\boldsymbol{s} - \boldsymbol{\sigma}(\boldsymbol{\Theta})]$$

$\|\boldsymbol{S}\|$ beschreibt hier die Determinanten der empirischen Kovarianzmatrix, $\|\boldsymbol{\Sigma}(\boldsymbol{\Theta})\|$ die Determinanten der zu schätzenden, modellimplizierten Kovarianzmatrix, $tr(.)$ die Spur einer Matrix, p und q die Anzahl der gemessenen Variablen x und y und damit die Größe der empirischen Kovarianzmatrix, $\boldsymbol{W}$ die Gewichtungsmatrix, $\boldsymbol{s}$ den Vektor der $(p + q)*(p + q + 1)/2$ Elemente der empirischen Kovarianzmatrix $\boldsymbol{S}$ und $\boldsymbol{\sigma}(\boldsymbol{\Theta})$ den Vektor der korrespondierenden Elemente der modellimplizierten Kovarianzmatrix $\boldsymbol{\Sigma}(\boldsymbol{\Theta})$.

Während die Vorteile der ML-Funktion in ihrer asymptotischen Konsistenz und Effizienz, vor allem bei großen Stichproben, in welchen zugrundeliegende Daten annährungsweise Normalverteilung erreichen liegt, werden Schätzer weniger Effizient bei geringen Stichprobengrößen und nicht-normalverteilten Daten. Auf Basis der ML-Funktion geschätzte Parameter sind zudem skaleninvariant, d.h. eine Skalentransformation der manifesten Variablen hat keine Auswirkungen auf den Wert der geschätzten Parameter. Wird die ML-Funktion für Modelle angewandt, welche auf (annährend) multinormalverteilen Variablen basieren, so folgt das Verhältnis der geschätzten Parameter und ihrer Standardfehler näherungsweise einer z-Verteilung (Reinecke, 2005, S. 122; Reinecke & Pöge, 2010, S. 789). Zudem sollte die ML-Funktion lediglich für kontinuierliche manifeste Variablen verwendet werden, da Variablen mit lediglich einer geringen Anzahl an Antwortkategorien gegen die Normalverteilungsannahme verstoßen (Lubke & Muthén, 2004). Nach Finney & DiStefano (2008) führt die Verwendung der ML-Funktion für Modelle auf Basis manifester Variablen mit erhöhten Schiefe- und Kurtosiswerten zwar zu relativ akkuraten Schätzungen der Parameter, jedoch zu verzerrten Standardfehlern und einer verzerrten χ^2-Statistik. Um dennoch von den Vorteilen einer ML-Schätzung bei nicht-normalverteilten Daten profitieren zu können, wird die Verwendung der robusten ML-Funktion (MLR) vorgeschlagen. MLR korrigiert die Verzerrung der Standardfehler, welche aus nicht-normalverteilten Variablen resultiert, und produziert alternativ die Satorra-Bentler χ^2-Statistik,

welche die herkömmliche χ^2-Statistik um den Einfluss der Verzerrung durch nicht-normalverteilte Daten korrigiert (Satorra & Bentler, Ensuring positiveness of the scaled difference chi-square test statistic, 2010). Aufgrund einer erhöhten Toleranz von MLR gegenüber Verstößen der Normalverteilungsannahme wird diese Funktion in der Praxis ebenfalls für geordnet-kategoriale manifeste Variablen verwendet (vgl. z.B. Li, 2016a).

Demgegenüber basiert die ULS-Funktion nicht auf entsprechend strengen Multinormalverteilungsannahmen und führt auch bei nicht-normalverteilten Daten zu konsistenten Schätzern, auch wenn im Gegensatz zur ML-Funktion nicht die effizientesten Schätzer ermittelt werden können. Extrem schiefe Verteilungen bzw. Verteilungen mit sehr hohen Kurtosiswerten können die geschätzten Parameter, Standardfehler und die χ^2-Statistik dennoch verzerren. Eine Auflösung der Multinormalverteilungsannahmen basiert hierbei auf der Annahme, dass alle Elemente der Matrix $\boldsymbol{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta})$ über die gleiche Streuung verfügen, wodurch eine weitere Schwachstelle der ULS-Funktion genannt ist, da eine ungleiche Streuung zu verzerrten Schätzungen führt. Zudem führt eine Veränderung der Skalierung der manifesten Variablen zu einer Veränderung der geschätzten Parameterwerte (keine Skaleninvarianz), weshalb allgemein empfohlen wird, die ULS-Funktion nur für Modelle anzuwenden, welche auf Basis einer Korrelationsmatrix geschätzt werden (Reinecke & Pöge, 2010). Wird die ULS-Funktion für geordnet-kategoriale manifeste Variablen auf Basis einer polychorischen Korrelationsmatrix verwendet, so zeigt eine Simulationsstudie von Li (2016b), dass ULS zu akkurateren Schätzungen der Parameter führt als ML, während MLR zu akkurateren Schätzungen robuster Standardfehler führt als ULS und dass MLR weniger anfällig ist für Typ 1 Fehler als ULS.

Bei der GLS-Funktion handelt es sich um eine Verallgemeinerung der ULS-Funktion, welche durch eine Gewichtungsmatrix $\boldsymbol{W}$ ergänzt wurde, so dass die Annahme der Streuungsgleichheit mit den damit assoziierten Problemen aufgehoben wird. GLS-Schätzer sind hierdurch, wie auch ML-Schätzer und im Gegensatz zu ULS-Schätzern Skaleninvariant. Wie bei der ML- und der ULS-Funktion können für die GLS-Funktion die geschätzten Parameter, Standardfehler und die χ^2-Statistik bei extrem schiefen Verteilungen bzw. Verteilungen mit sehr hohen Kurtosiswerten verzerrt sein (Reinecke & Pöge, 2010). Nach Kline (2016) liegt der Vorteil von GLS gegenüber ML vor allem darin, weniger Zeitaufwändig bei der Schätzung der Modelle zu sein und dabei weniger Computer-Arbeitsspeicher zu benötigen. Da dieser Vorteil bei modernen Computersystemen jedoch in der Regel zu vernachlässigen ist, ist ML(R) in den meisten Fällen gegenüber GLS und ULS zu präferieren (Kline, 2016, S. 256).

In der WLS-Funktion schließlich werden Verteilungsannahmen aufgehoben, wodurch asymptotisch effiziente Parametervektoren $\boldsymbol{\theta}$ geschätzt werden können. Informationen bezüglich Schiefe und Kurtosis der Daten werden durch die Gewichtungsmatrix $\boldsymbol{W}$ bereitgestellt, so dass alle Informationen über die Verteilungen der manifesten Variablen bei der Modellschätzung berücksichtigt und optimal ausgenutzt werden. Zu berücksichtigen bleibt hierbei jedoch die von der Größe des Modells abhängige Stichprobengröße, da sich hieraus die stabile Schätzung der asymptotischen Varianz-/Kovarianzmatrix ergibt. Über die Stichprobengröße hinaus steigt bei großen Modellen die Größe der asymptotischen Varianz-/Kovarianzmatrix überproportional an, was mitunter zu sehr vielen zusätzlich zu schätzenden Varianzen und Kovarianzen führt. Bei nur 10 manifesten Variablen handelt es sich hierbei bereits um 3080 asymptotische Varianzen und Kovarianzen (Reinecke & Pöge, 2010, S. 790f). Es ist somit zu berücksichtigen, dass die Schätzung großer Modelle mithilfe der WLS-Funktion, auch bei modernen Computersystemen, mitunter einige Zeit in Anspruch nimmt. Als Alternative zu WLS wurde WLSMV (WLS with means and variances adjusted) vorgeschlagen (Muthén & Muthén, 2007), wobei es sich bei der WLSMV-Funktion um eine mathematisch simplere Form der WLS-Funktion handelt, welche die Anzahl notwendiger Berechnungen reduziert, wobei eine Reihe von Simulationsstudien die Überlegenheit von WLSMV gegenüber WLS demonstrieren, was insbesondere bei Schätzungen von Modellen auf Basis kategorialer manifester Variablen gilt (für einen umfassenden Überblick, siehe Li, 2016a).

Wird nun also nach Kline (2016) angenommen, dass ML in den meisten Fällen ULS und GLS überlegen ist, so lässt sich für die konkrete Auswahl einer geeigneten Schätzfunktion für den analytischen Teil dieser Arbeit folgendes zusammenfassen: ML ist insbesondere geeignet für kontinuierliche und normalverteile manifeste Variablen. Für kategoriale Variablen ist WLSMV zu empfehlen, da hier alle Verteilungsannahmen aufgehoben sind und eine Reihe von Simulationsstudien die Effizienz der Funktion für eben jene Variablen aufzeigt. Für nicht-normalverteilte kontinuierliche Variablen kann ebenfalls WLSMV verwendet werden, wobei zu berücksichtigen ist, dass WLSMV (auch wenn es sich um eine mathematisch betrachtet vereinfachte Form von WLS handelt) bei großen Modellen dennoch mehr Rechenzeit benötigt. Für große Modelle auf Basis kontinuierlicher und nicht-normalverteilter manifester Variablen ist entsprechend MLR zu empfehlen.

Es zeigt sich somit, dass systematische Messfehler innerhalb der SEM-Logik auf Basis von Struktur- bzw. Messgleichungen formal modelliert und somit innerhalb der zu schätzenden Parametermatrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ dargestellt werden können, wodurch diese nicht als Residualvarianzen, sondern als zu schätzende Parameter behandelt und quantifiziert werden. Entsprechend folgt

eine genauere Betrachtung der Formalisierung und Schätzung systematischer Messfehler innerhalb von SEM, wobei hierbei auf die Logik von MTMM Analysen, einem weiteren, für diese Arbeit relevanten Verfahren, zurückgegriffen wird.

3.1.3 Systematische Messfehler in Multi-Trait-Multimethod-Analysen

Als vielversprechend und etabliert hinsichtlich der Modellierung und Quantifizierung systematischer Messfehler erweist sich die sogenannte Multi-Trait-Multimethod(MTMM)-Analyse, mit welcher initial eine Gruppe von Verfahren bezeichnet wurde, die der Bestimmung der Konstruktvalidität dienen (für eine vertiefende Diskussion der Validität im Allgemeinen und der Konstruktvalidität im Speziellen, siehe Kapitel 3.1.4). Den Verfahren liegt dabei zugrunde, dass für Messungen eines Merkmals angenommen wird, dass dessen Varianz nicht lediglich auf das konkret zu messende Merkmal und einen zufälligen Messfehler, sondern ebenfalls auf die bei der Erhebung der Daten verwendete Methodik zurückgeführt werden kann. Insbesondere bei Anwendung des MTMM Ansatzes innerhalb von CFA bzw. SEM ist es somit möglich, systematische Messfehler zu schätzen und zu quantifizieren, um somit messfehlerbereinigte Schätzungen von Merkmalen zu erreichen (Schermelleh-Engel & Schweizer, Multitraid-Multimethod-Analysen, 2012). Aufgrund eben dieser Möglichkeit, Messfehler innerhalb von CFA und SEM zu quantifizieren und Schätzungen um diese zu bereinigen, kommt der MTMM Analyse eine zentrale Bedeutung in dieser Arbeit zu. Entsprechend wird das Verfahren im Folgenden vorgestellt, um hierauf basierend in Kapitel 3.2 und 3.3 ein CFA Modell auf Basis der MTMM Logik abzuleiten, mit welchem FRE-bedingte Verzerrungen bereinigt werden können. Die Effizienz zur Bereinigung von Messfehlern dieses MTMM Modells wird dann im analytischen Teil dieser Arbeit mit der Effizienz der einzuführenden NLC-CFA verglichen.

Grundannahmen der MTMM Analyse

Die Annahme, dass die Varianz einer Messung nicht lediglich auf das konkret zu messende Merkmal und einen zufälligen Messfehler, sondern ebenfalls auf die bei der Erhebung der Daten verwendete Methodik zurückgeführt werden kann, lässt sich auf die Arbeit von Campbell und Fiske (1959) zurückführen, welche neben der konvergenten Validität die Bedeutung der diskriminanten Validität für die Konstruktvalidität diskutieren. Demnach liegt dann konvergente Validität vor, wenn mit verschiedenen Messinstrumenten durchgeführte Erhebungen eines Merkmals hoch miteinander korrelieren, während diskriminante Validität dann vorliegt, wenn Messungen eines Merkmals nur gering oder nicht mit den Messungen unterschiedlicher

Merkmale korrelieren. Problematisch an dieser Betrachtung ist nun, dass selbst wenn verschiedene Messinstrumente das gleiche Merkmal messen, die Korrelation zwischen beiden Messungen dann negativ verzerrt sein kann, wenn diese mit unterschiedlichen Erhebungsinstrumenten gemessen wurden, da die Erhebungsmethode über das eigentliche Merkmal hinaus ebenfalls Einfluss auf die Messung nimmt. Für eigentlich konvergent valide Messinstrumente ergibt sich hieraus unter Umständen eine negative Verzerrung der Korrelation, wenn die verwendete Methode zwischen den Erhebungsinstrumenten verschieden ist, diese aber einen systematischen Einfluss auf die Messung haben. Für eigentlich diskriminant valide Messinstrumente ergibt sich hieraus entsprechend unter Umständen eine positive Verzerrung der Korrelation, wenn die verwendete Methode zwischen den Erhebungsinstrumenten gleich ist und diese einen systematischen Einfluss auf die Messung nimmt. Campbell und Fisk postulieren entsprechend, dass sich jede Messung aus dem systematischen Einfluss des Traits (des Merkmals), dem systematischen Einfluss der Methode (dem Methodeneffekt) und einem unsystematischen, das heißt zufälligen Messfehler zusammensetzt.

Klassische (korrelationsbasierte) MTMM Analyse

Verfügen Messungen des gleichen Konstruktes mit unterschiedlichen Methoden entsprechend über eine erhöhte Korrelation, so wird konvergente Validität angenommen. Verfügen Messungen unterschiedlicher Konstrukte innerhalb einer Methode, aber auch zwischen Methoden darüber hinaus über eine geringe Korrelation, so wird diskriminante Validität angenommen (Schermelleh-Engel & Schweizer, 2012). Um dies zu überprüfen, wurde von Campbell und Fiske (1959) die Analyse mittels MTMM-Matrix vorgeschlagen, eine systematische Korrelationsmatrix, in welcher die Korrelationen aller möglichen Kombinationen von Merkmalen und Methoden dargestellt ist (dies bedeutet ebenfalls, dass mindestens zwei oder besser drei Merkmale und mindestens zwei oder besser drei Methoden gemessen werden müssen). Abbildung 8 zeigt diesbezüglich eine beispielhafte MTMM-Matrix, welche aus Gründen der Übersichtlichkeit mit je zwei Merkmalen und zwei Methoden dargestellt ist.

Abbildung 8: Vereinfachte Darstellung der MTMM-Matrix

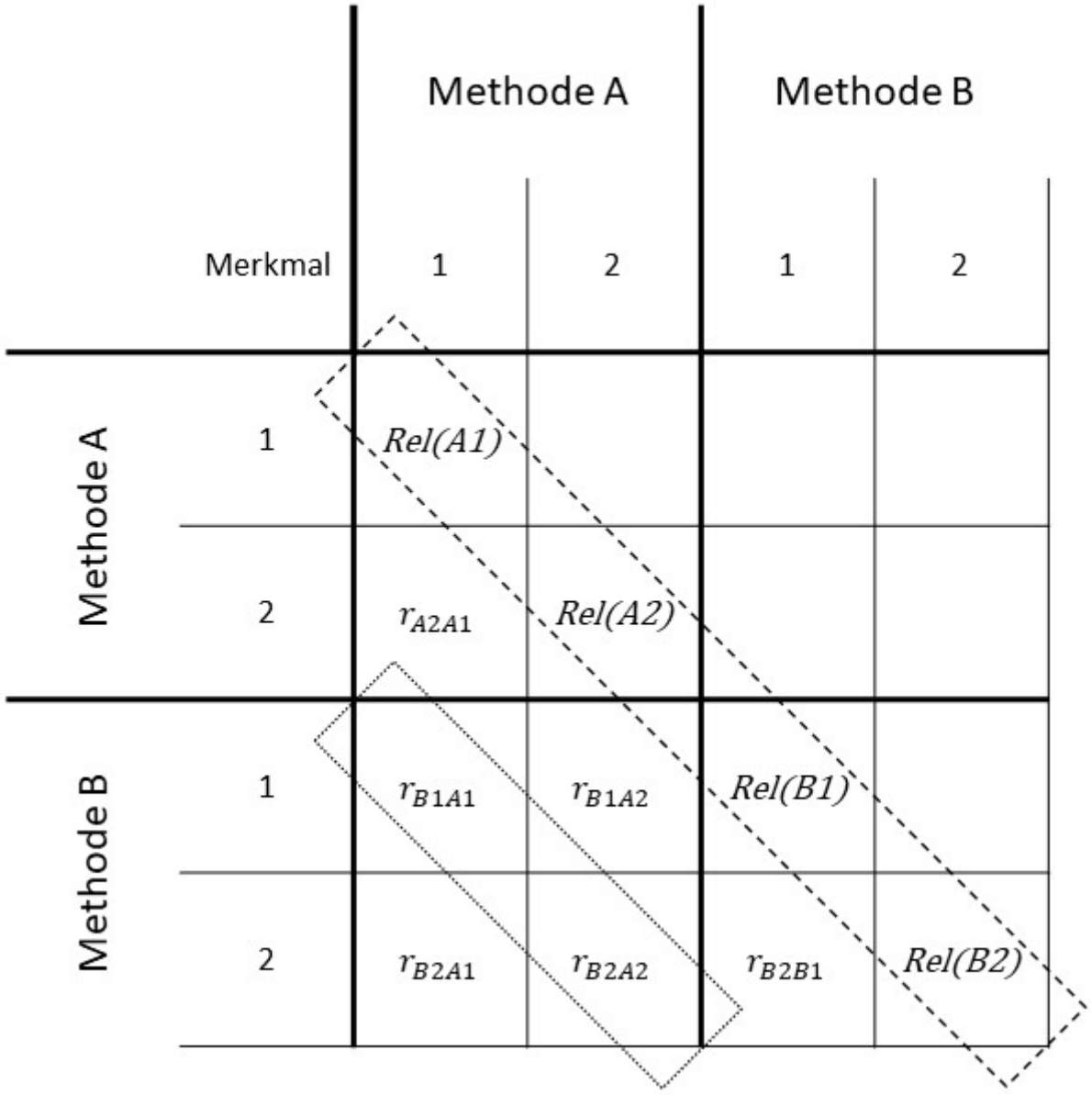


Abbildung nach Schermelleh-Engel & Schweizer, 2012

Zur Auswertung der MTMM-Matrix (nach Schermelleh-Engel & Schweizer, 2012) ist zunächst zu sagen, dass zwischen Monomethod-Blöcken und Heteromethod-Blöcken unterschieden wird. Monomethod-Blöcke erhalten demnach die Korrelationskoeffizienten derjenigen Merkmale, welche mit der gleichen Methode gemessen wurden, während Heteromethod-Blöcke die Korrelationskoeffizienten zwischen den untersuchten Merkmalen enthalten, welche mit verschiedenen Methoden gemessen wurden. Die Monomethod-Blöcke in Abbildung 8 finden sich entsprechend dort, wo sich Methode A mit Methode A ($Rel(A1); Rel(A2); r_{A2A1}$) und Methode

B mit Methode B (*Rel(B1); Rel(B2);* r_{B2B1}) schneiden. Heteromethod-Blöcke finden sich hingegen dort, wo sich Methode A mit Methode B schneidet (r_{B1A1}; r_{B1A2}; r_{B2A1}; r_{B2A2}). Darüber hinaus werden bezüglich der Interpretation vier Arten von Koeffizienten unterschieden: (1) Monotrait-Monomethod Koeffizienten, welche in der gestrichelten Hauptdiagonale zu finden sind und welche als Reliabilitätskoeffizienten innerhalb der korrelationsbasierten MTMM-Analyse betrachtet werden (z.B. *Rel(B1)*); (2) Monotrait-Heteromethod Koeffizienten, welche in der gepunkteten Nebendiagonale zu finden sind und welche die konvergente Validität innerhalb der korrelationsbasierten MTMM-Analyse quantifizieren (z.B. r_{B1A1}); (3) Heterotrait-Monomethod Koeffizienten, welche innerhalb der Monomethod-Blöcke unter- oder oberhalb der Hauptdiagonalen zu finden sind, die die Korrelation zwischen unterschiedlichen Merkmalen, die mit der gleichen Methode gemessen wurden, darstellen und als Indikator für die diskriminante Validität gelten (z.B. r_{A2A1}); sowie (4) Heterotrait-Heteromethod Koeffizienten, welche innerhalb der Heteromethod-Blöcke unter- oder überhalb der Nebendiagonale zu finden sind (z.B. r_{B2A1}), die die Korrelation zwischen unterschiedlichen Merkmalen, die jeweils mit unterschiedlichen Methoden gemessen wurden, darstellen und welche als Koeffizienten der diskriminanten Validität, welche um den Einfluss der Methode bereinigt wurden, interpretiert werden können. Entsprechend ist es der nächste Schritt, die Validität und Reliabilität über einen Vergleich der Korrelationen zwischen den Messvariablen zu analysieren. Reliabilität liegt demnach dann vor, wenn die Monotrait-Monomethod Koeffizienten der Hauptdiagonalen hoch sind. Konvergente Validität liegt dann vor, wenn die Monotrait-Heteromethod Koeffizienten der Nebendiagonalen hoch sind. Diskriminante Validität zuletzt liegt vor, wenn sowohl die Heterotrait-Monomethod Koeffizienten, als auch die Heterotrait-Heteromethod Koeffizienten möglichst gering sind.

Dieser klassische Ansatz der korrelationsbasierten MTMM-Analyse steht jedoch in der Kritik (Schermelleh-Engel & Schweizer, 2012). So setzt die Anwendung erstens streng genommen voraus, dass alle Merkmale mit der gleichen Zuverlässigkeit gemessen werden, da sich ein systematischer Unterschied in der Reliabilität der Methodik insofern auf die Auswertung auswirkt, als dass die Methode mit der höheren Reliabilität höhere Korrelationen und die Methode mit der geringeren Reliabilität geringere Korrelationen hervorruft. Die Vergleichbarkeit der Koeffizienten ist somit bei unterschiedlicher Reliabilität der Methoden streng genommen nicht gegeben – auch, da das Verfahren zur Messung der Reliabilität herangezogen wird, die Messung selbst aber durch eben jene Reliabilität beeinflusst ist. Zweitens werden innerhalb der Analyse im Wesentlichen viele Einzelvergleiche von Korrelationskoeffizienten vorgenommen, deren

Auswertung und Bewertung subjektiv sind. Und drittens basiert die Auswertung der Korrelationsmatrix auf Koeffizienten, welche auf manifesten Variablen beruhen, obwohl sich die Interpretation der Ergebnisse auf die zugrundeliegenden latenten Merkmale und Methoden bezieht. Um diesen Problemen zu begegnen, wurde die konfirmatorische MTMM-Analyse vorgeschlagen, welche im Folgenden als weiterer Ansatz vorgestellt wird, insbesondere auch, da sich Methodeneffekte, sprich systematische Messfehler, hiermit konkret modellieren und quantifizieren lassen.

Konfirmatorische MTMM Analyse

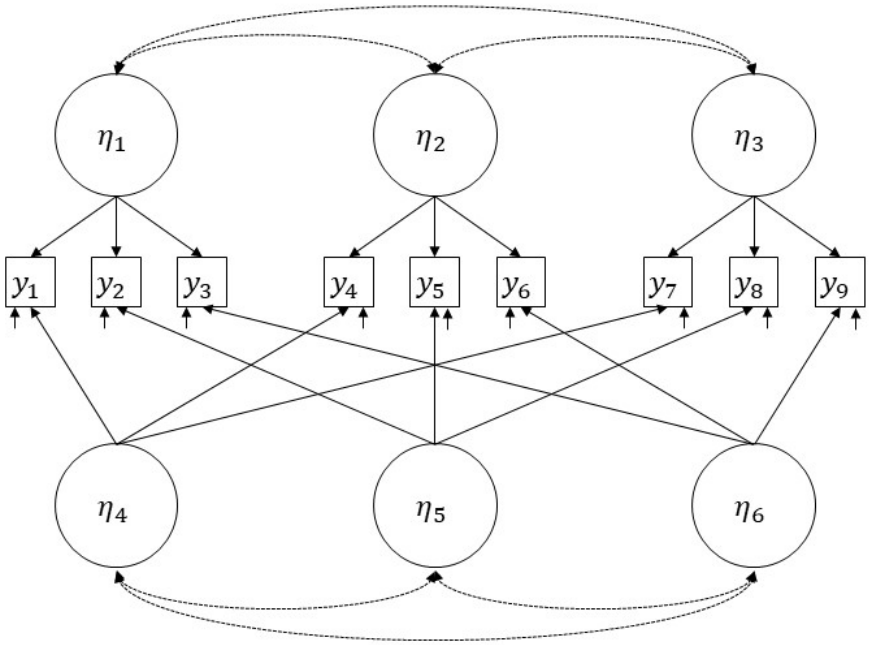
Mithilfe der konfirmatorischen Faktoranalyse (CFA, vgl. Kapitel 3.1.2) ist es grundsätzlich möglich, Messmodelle auf Basis von Korrelations- bzw. Kovarianzmatrizen zu schätzen (Jöreskog, Statistical analysis of sets of congeneric tests, 1971). Da die MTMM-Analyse auf den bereits vorgestellten MTMM-Matrizen beruht, kann diese somit ebenfalls auf Basis einer Schätzung mittels CFA durchgeführt werden (sog. CFA-MTMM). Dies ist insofern sinnvoll, als dass sowohl die Merkmale, als auch die Methoden als latente Variablen geschätzt werden können wodurch es möglich ist, die Varianz einer Messung *x* in den Anteil des Merkmals, der Methode (bzw. des systematischen Messfehlers) und des zufälligen Messfehlers zu zerlegen, wodurch es zu einer bereinigten Schätzung der Faktorladungen kommt (Schätzung des Merkmals, die um den Einfluss der Methode bereinigt ist), welche dann als Indikatoren für die Validität herangezogen werden. Darüber hinaus lassen sich weitere Annahmen wie etwa die Eindimensionalität der Merkmalsfaktoren oder die (Un)Korreliertheit von Merkmalen und Methoden überprüfen. Daneben besteht die Möglichkeit der Extraktion der Methodenvarianz durch schlichte Korrelation der Messfehler innerhalb der gleichen Methode statt der Modellierung von Methodenfaktoren wenn angenommen wird, dass der Einfluss der Methode auf die manifesten Variablen uneinheitlich ist oder die Methoden nicht eindimensional sind (Schermelleh-Engel & Schweizer, 2012). Zur Verdeutlichung ist in Abbildung 9.1 ein einfaches Messmodell mit 9 Indikatoren, 3 Merkmalsfaktoren und 3 Methodenfaktoren dargestellt, während in Abbildung 9.2 ein einfaches Messmodell mit 9 Indikatoren, 3 Merkmalsfaktoren und der Korrelation der Messfehler innerhalb einer Methode dargestellt.

Exemplarisch sei hierbei angenommen, dass die Indikatoren y_1 bis y_3 verschiedene Messungen des Merkmals η_1 darstellen, y_4 bis y_6 Messungen des Merkmals η_2 darstellen und Merkmal η_3 durch die Indikatoren y_7 bis y_9 spezifiziert wird. Eine mögliche, aber nicht zwingend zu spezifizierende Korrelation zwischen den latenten Faktoren wird durch einen gestrichelten Doppel-

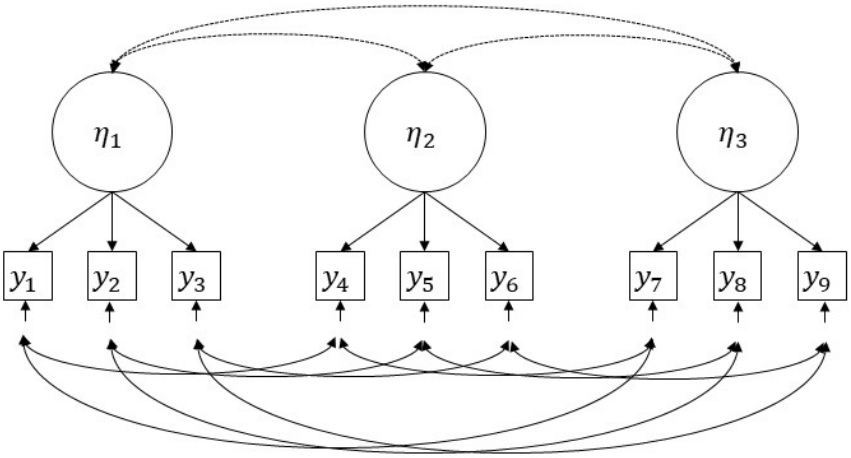
pfeil dargestellt. Weiterhin wird das Modell mit spezifizierten Methodenfaktoren (vgl. Abbildung 9.1) um die Methodenfaktoren η_4 bis η_6 ergänzt, für welche ebenfalls eine Korrelation spezifiziert werden kann. Innerhalb des Modells wird entsprechend davon ausgegangen, dass y_1, y_4 und y_7 mit Methode 1, y_2, y_5 und y_8 mit Methode 2 und y_3, y_6 und y_9 mit Methode 3 gemessen werden.

Abbildung 9: CFA-MTMM Modell

(1)



(2)



Demgegenüber steht das Modell ohne spezifizierte Methodenfaktoren (vgl. Abbildung 9.2), in welchem stattdessen Kovarianzen zwischen den Messfehlern der Items y_1, y_4 und y_7 zur Varianzextraktion von Methode 1, der Items y_2, y_5 und y_8 zur Varianzextraktion von Methode 2 und der Items y_3, y_6 und y_9 zur Varianzextraktion von Methode 3 spezifiziert sind.

Die dargestellten Modelle lassen sich weiterhin entsprechend Matrix-Schreibweise in ihre jeweiligen Struktur- bzw. Messgleichungen zerlegen, wobei Korrelationen auf Faktorebene entsprechend den Darstellungen angenommen werden. In Formel 31 ist diesbezüglich das Gleichungssystem für das Modell mit spezifizierten Methodeneffekten sowie Korrelationen zwischen den latenten Merkmalsfaktoren sowie den Methodenfaktoren und in Formel 32 das Gleichungssystem für das Modell ohne Methodeneffekte dargestellt:

$$(1)\quad \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 & 0 & \lambda_{14} & 0 & 0 \\ \lambda_{21} & 0 & 0 & 0 & \lambda_{25} & 0 \\ \lambda_{31} & 0 & 0 & 0 & 0 & \lambda_{36} \\ 0 & \lambda_{42} & 0 & \lambda_{34} & 0 & 0 \\ 0 & \lambda_{52} & 0 & 0 & \lambda_{55} & 0 \\ 0 & \lambda_{62} & 0 & 0 & 0 & \lambda_{66} \\ 0 & 0 & \lambda_{73} & \lambda_{74} & 0 & 0 \\ 0 & 0 & \lambda_{83} & 0 & \lambda_{85} & 0 \\ 0 & 0 & \lambda_{93} & 0 & 0 & \lambda_{96} \end{bmatrix} * \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \\ \eta_6 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \end{bmatrix}$$

$$(2)\quad \boldsymbol{\Psi} = \begin{bmatrix} \psi_{11} & & & & & \\ \psi_{21} & \psi_{22} & & & & \\ \psi_{31} & \psi_{32} & \psi_{33} & & & \\ 0 & 0 & 0 & \psi_{44} & & \\ 0 & 0 & 0 & \psi_{54} & \psi_{55} & \\ 0 & 0 & 0 & \psi_{64} & \psi_{65} & \psi_{66} \end{bmatrix}$$

Formel 31

$$(3)\quad \boldsymbol{\Theta}_{\varepsilon} = \begin{bmatrix} \theta_{\varepsilon_1} & & & & & & & & \\ 0 & \theta_{\varepsilon_2} & & & & & & & \\ 0 & 0 & \theta_{\varepsilon_3} & & & & & & \\ 0 & 0 & 0 & \theta_{\varepsilon_4} & & & & & \\ 0 & 0 & 0 & 0 & \theta_{\varepsilon_5} & & & & \\ 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_6} & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_7} & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_8} & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_9} \end{bmatrix}$$

Innerhalb der CFA-MTMM Logik ist es bei gegebenem Gleichungssystem nun möglich, Faktorladungen λ_y, Messfehler ε_j (sowie Kovarianzen auf Messfehlerebene) und latente Varianzen und Kovarianzen ψ_k gemäß dem beschriebenen Verfahren in Kapitel 3.1.2 zu schätzen. Für die Interpretation des CFA-MTMM-Modells hinsichtlich der Validität gilt, dass sich konvergente Validität durch hohe Faktorladungen auf den Merkmalsfaktor und diskriminante Validität durch eine geringe Korrelation der Merkmalsfaktoren untereinander auszeichnet. Der Einfluss der Methode wiederum äußert sich durch hohe Faktorladungen der Messungen auf den Methodenfaktor (so dieser spezifiziert ist) bzw. durch hohe Kovarianzen zwischen den Messfehlern (so keine Methodenfaktoren spezifiziert sind).

$$(1)\quad \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 & 0 \\ \lambda_{21} & 0 & 0 \\ \lambda_{31} & 0 & 0 \\ 0 & \lambda_{42} & 0 \\ 0 & \lambda_{52} & 0 \\ 0 & \lambda_{62} & 0 \\ 0 & 0 & \lambda_{73} \\ 0 & 0 & \lambda_{83} \\ 0 & 0 & \lambda_{93} \end{bmatrix} * \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \end{bmatrix}$$

$$(2)\quad \boldsymbol{\Psi} = \begin{bmatrix} \psi_{11} & & \\ \psi_{21} & \psi_{22} & \\ \psi_{31} & \psi_{32} & \psi_{33} \end{bmatrix}$$

Formel 32

$$(3)\quad \boldsymbol{\Theta}_{\varepsilon} = \begin{bmatrix} \theta_{\varepsilon_1} & & & & & & & & \\ 0 & \theta_{\varepsilon_2} & & & & & & & \\ 0 & 0 & \theta_{\varepsilon_3} & & & & & & \\ \theta_{\varepsilon_{41}} & 0 & 0 & \theta_{\varepsilon_4} & & & & & \\ 0 & \theta_{\varepsilon_{52}} & 0 & 0 & \theta_{\varepsilon_5} & & & & \\ 0 & 0 & \theta_{\varepsilon_{63}} & 0 & 0 & \theta_{\varepsilon_6} & & & \\ \theta_{\varepsilon_{71}} & 0 & 0 & \theta_{\varepsilon_{74}} & 0 & 0 & \theta_{\varepsilon_7} & & \\ 0 & \theta_{\varepsilon_{82}} & 0 & 0 & \theta_{\varepsilon_{85}} & 0 & 0 & \theta_{\varepsilon_8} & \\ 0 & 0 & \theta_{\varepsilon_{93}} & 0 & 0 & \theta_{\varepsilon_{96}} & 0 & 0 & \theta_{\varepsilon_9} \end{bmatrix}$$

Der Logik von CFA-MTMM-Analysen folgend ist es somit wie bereits beschrieben möglich, die Varianz einer Messung auf den Einfluss des Merkmals, der Methode (bzw. des systematischen Messfehlers) und des zufälligen Messfehlers zurückzuführen und zeitgleich den Anteil

der systematischen Fehlervarianz zu quantifizieren. Innerhalb von Modellen mit spezifizierten Methodenfaktoren stellt sich die systematische Fehlervarianz als Varianz des latenten Methodenkonstruktes dar, während innerhalb von Modellen ohne spezifizierten Methodenfaktor die systematische Fehlervarianz durch die Kovarianz auf Messfehlerebene aus der Gesamtvarianz heraus partialisiert wird. Je nach spezifiziertem Modell und Datenlage kann es hierbei jedoch zu Schätzfehlern kommen, welche sich z.B. durch negative Fehlervarianzen oder standardisierte Faktorladungen größer 1 äußern. Ebenfalls problematisch ist die allgemeine Identifikation von Modellen, wenn die Anzahl zu schätzender Parameter aufgrund der zusätzlichen zu schätzenden Methodenfaktoren und Faktorladungen bzw. Kovarianzen größer ist als die Anzahl der zur Verfügung stehenden Informationen (Schermelleh-Engel & Schweizer, 2012). Nach Marsh und Bailey (1991) treten solche Schätzprobleme jedoch überwiegend bei Modellen mit spezifiziertem Methodenfaktor und korreliertem Merkmalsfaktor auf, während Modelle ohne spezifizierte Methodenfaktoren und mit spezifizierten Kovarianzen auf Messfehlerebene (bei korrelierten Merkmalsfaktoren) überwiegend zu korrekten Lösungen führen. Weiterhin sind solche Schätzprobleme nach Marsh und Grayson (1995) vor allem dann wahrscheinlich, wenn relativ wenige Merkmalsfaktoren innerhalb des Modells spezifiziert werden und wenn die Stichprobe klein ist. Um Schätzproblemen bei spezifizierten Methodenfaktoren zu begegnen, schlägt Eid (2000) die Verwendung von CTC(M-1)-Modellen (CTCM = correlated trait, correlated method) vor, gemäß deren Logik ein Methodenfaktor weniger spezifiziert wird, als durch die MTMM-Logik vorgesehen, was die Identifikationsprobleme dieser Modelle beseitigt. Die Effekte der spezifizierten Methodenfaktoren sind in diesem Fall nicht mehr absolut, sondern relativ zum unspezifizierten Methodenfaktor zu interpretieren. Darüber hinaus sind solche Modelle nicht mehr symmetrisch was bedeutet, dass die Modellgüte verschieden je nach unspezifizierter Methode ist.

Weitere CFA-MTMM Modelle

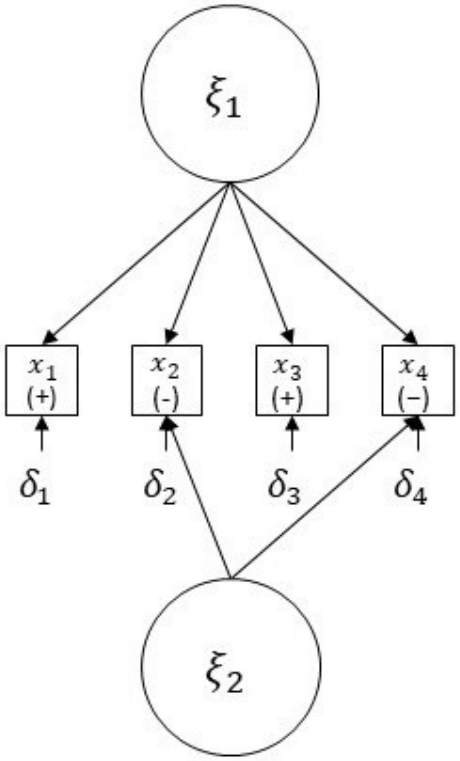
Die Möglichkeiten von CFA-MTMM zur Spezifikation der Methodenvarianz sind derart vielfältig, dass in der Praxis eine Fülle verschiedener Modelle für die Analyse komplexer Konstellationen an Traits und Methoden vorgeschlagen wurden (für eine detaillierte Übersicht, siehe Koch, Holtmann, Bohn, & Eid, 2017). Mit Hinblick auf eine genauere Betrachtung der Methodeneffekte befinden sich darunter ebenfalls Modelle zur psychometrischen Bestimmung der latenten Methodeneffekte (Geiser, Eid, West, Lischetzke, & Nussbeck, 2012) sowie zur Erklärung von latenten Methodeneffekten durch sowohl manifeste, als auch durch latente Variablen (Koch et al., 2017) bei vorhandenen Multimethod-Daten. Koch, Kelava und Eid (2018) schla-

gen darüber hinaus ein nonlinear latent difference (NL-LD) Modell sowie ein nonlinear correlated trait-correlated method-minus-one (NL-CTC[M – 1]) Modell für die Analyse moderierter Methodeneffekte zur Erklärung von Methodenfaktoren auf Basis latenter Interaktionen vor. Den Verfahren ist dabei gemein, dass Methodenfaktoren gemäß Abbildung 9Abbildung 9: CFA-MTMM Modell.1 extrahiert werden und dass der Untersuchung mehrere Methoden zugrunde liegen.

In der Praxis lassen sich die MTMM-Annahmen rund um den Einfluss der zugrundeliegenden Methode auf eine Messung ebenfalls für die Spezifikation von Modellen adaptieren, deren Methodik nicht im herkömmlichen Sinne des Erhebungsverfahrens variiert wurde. Als Beispiel sei hier die Variation positiv und negativ formulierter Zustimmungsitems innerhalb der gleichen Fragebatterie genannt, wie es beispielsweise innerhalb der Skala zur generalisierten Einstellung gegenüber Umfragen von Stocké und Becker (2004) der Fall ist. In diesem Beispiel lautet eine Frageformulierung: „die meisten Umfragen sind seriös und werden verantwortungsvoll durchgeführt“, eine Zustimmung impliziert entsprechend eine positive Einstellung gegenüber Umfragen. Eine weitere Frageformulierung lautet hingegen: „Durch Umfragen wird man nur von wichtigeren Dingen abgehalten“, was eine negative Einstellung gegenüber Umfragen impliziert. Im Sinne der MTMM-Logik befinden sich somit innerhalb der gleichen Merkmalsmessung (die generalisierte Einstellung gegenüber Umfragen) zwei verschiedene Erhebungsmethoden: Eine Erhebung auf Basis positiv formulierter Items und eine Erhebung auf Basis negativ formulierter Items. Bezüglich solcher Skalen zeigt sich, dass eine eindimensional modellierte Faktorstruktur häufig keine gute Repräsentation der Daten darstellt, auch wenn diese theoriegeleitet sinnvoller wär (Marsh, 1986), da positiv und negativ formulierte Items in der Tendenz auf verschiedenen Faktoren laden (Greenberger, Chen, Dmitrieva, & Farruggia, 2003), was die Modellierung eines zugrundeliegenden latenten Methodenfaktors entsprechend CTC(M-1)-Logik zur Extraktion der Methodenvarianz der negativ (oder positiv) formulierten Items nötig macht (Brown, 2003; Best & Mayerl, 2013; Mayerl & Giehl, 2018; (Tomás, Oliver, Galiana, Sancho, & Lila, 2013). Als Beispiel modellieren Mayerl & Giehl (2018) einen Merkmalsfaktor zur Einstellung gegenüber gesundheitsbewusster Ernährung auf Basis von 4 Indikatorvariablen x_1 bis x_4, wobei x_1 und x_3 positiv formuliert sind, während x_2 und x_4 negativ formuliert sind. Die Methodenvarianz durch die negative Itemformulierung in x_2 und x_4 wird durch die Modellierung eines Methodenfaktors gemäß CTC(M-1)-Logik extrahiert. Die empirischen Ergebnisse der Analyse zeigen diesbezüglich einen deutlichen Vorteil des Modells, in welchem die Methodenvarianz extrahiert wird versus dem Modell, in welchem dies nicht geschieht. Abbildung

10 zeigt diesbezüglich das spezifizierte Model, Formel 33 das zugrundeliegende Gleichungssystem, wobei aus Gründen der vollständigen Illustration in Abgrenzung zum vorangegangenen Beispielmodell in Abbildung 9 und dem Gleichungssystem in Formel 31 und Formel 32 die X-Notation gewählt wurde.

Abbildung 10: CFA-CTC(M-1) Modell mit 4 Indikatoren und einem Methodenfaktor



$$(1)\quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} & 0 \\ \lambda_{41} & \lambda_{42} \end{bmatrix} * \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{bmatrix}$$

$$(2)\quad \mathbf{\Phi} = \begin{bmatrix} \phi_{11} & \\ 0 & \phi_{22} \end{bmatrix}$$

Formel 33

$$(3)\quad \mathbf{\Theta}_\delta = \begin{bmatrix} \theta_{\delta_1} & & & \\ 0 & \theta_{\delta_2} & & \\ 0 & 0 & \theta_{\delta_3} & \\ 0 & 0 & 0 & \theta_{\delta_4} \end{bmatrix}$$

Es zeigt sich somit grundsätzlich, dass mittels CFA-MTMM systematische Messfehler durch Spezifikation eines oder mehrerer Methodenfaktoren, aber ebenso durch die Spezifikation von Kovarianzen auf Messfehlerebene modelliert und ebenso quantifiziert werden können. Darüber hinaus wurden bereits Modelle vorgeschlagen, mit welchen spezifizierte Methodenfaktoren durch zugrundeliegende, exogene (latente) Kovariate (nichtlinear) erklärt werden können. Den bisher vorgeschlagenen Modellen liegt dabei zugrunde, dass die zugrundeliegenden Daten auf Basis unterschiedlicher Methoden erhoben wurden. Das Vorhandensein mehrerer Traits bzw. Merkmale, wie klassisch für die korrelationsbasierte MTMM Analyse vorgesehen, ist dabei für CFA-MTMM nicht notwendig, wie das obige Beispiel zeigt.

Wurden die manifesten Variablen jedoch mit der gleichen Methode gemessen, liegen den Daten also gerade nicht mehrere Methoden zugrunde, so können die bisher diskutierten Modelle nicht angewandt werden, um die Varianz dieser einen Methode zu extrahieren, um somit zu bereinigten Messungen des Merkmals zu gelangen. Weiterhin sind die genannten Modelle lediglich dazu geeignet, Methodenvarianz bzw. systematische Fehlervarianz bei Spezifikation von Methodenfaktoren durch zugrundeliegende Kovariate zu erklären. Können solche Methodenfaktoren nicht spezifiziert werden (z.B. da dies in einer negativen Zahl an Freiheitsgraden innerhalb des Modells führen würde), wird die systematische Fehlervarianz also auf Basis von Kovarianzen auf Messfehlerebene spezifiziert, so können die bisher diskutierten Modelle ebenfalls nicht verwendet werden, um die extrahierte Methodenvarianz auf zugrundeliegende Kovariate zurückzuführen und somit auch nicht zur Erklärung und Kontrolle von FRE verwendet werden. Dies liegt insbesondere daran, dass der methodische Einfluss auf das Auftreten von FRE auf jede einzelne Zielfrage einer Fragebatterie wirkt (vgl. Kapitel 2.3). Der methodische Einfluss wirkt somit innerhalb eines einzelnen Items und nicht zwischen den Items, wie durch die bisher diskutierten Modelle angenommen. Dies muss bei der später folgenden Spezifizierung von Modellen zur Identifikation und Kontrolle von FRE berücksichtigt werden.

3.1.4 Implikationen für Messungen der Datenqualität

Das Auftreten systematischer Messfehler wie FRE hat weitreichende messtheoretische Implikationen, welche vor allem darauf zurückzuführen sind, dass Kovarianzen zwischen Items, deren Messung durch systematische Messfehler beeinflusst wurden, ebenfalls systematisch ver-

zerrt sind, wodurch statistische Maßzahlen und geschätzte Parameter, je nach Präsenz der jeweiligen Messfehler, nicht mehr präzise und effizient wahre Werte wiederspiegeln (vgl. z.B. Bard & Weinstein, 2016; Cunningham et al., 2016; Garbarski et al., 2014; Huang & Cornell, 2015; Magelssen, 2016). Die Implikationen sind sogar derart weitreichend, dass Messungen der Datenqualität durch auftretende systematische Messfehler künstlich erhöht sein können (vgl. z.B. Mayerl & Giehl, 2018). In Anbetracht dessen hat die Kontrolle von systematischen Messfehlern im Allgemeinen und von FRE im Speziellen durch Spezifikation und Extraktion der entsprechenden systematischen Messfehlervarianz und die dadurch geschehende Bereinigung der geschätzten Common Variance über die reine Identifikation von FRE und den entsprechenden Prädiktoren hinaus den Mehrwert, dass Messungen der Datenqualität via Berücksichtigung zusätzlicher Informationen bei der Schätzung der Unique Variances potentiell effizienter erfolgen können. Diesbezüglich wird nachfolgend abschließend diskutiert, inwiefern Messungen der Reliabilität und der Validität eine potentielle Verzerrung bei Vorhandensein systematischer Messfehler erfahren und wie eine statistische Kontrolle der systematischen Messfehler dazu beitragen kann, diese Verzerrung (zumindest teilweiße) zu bereinigen, um die Notwendigkeit der Berücksichtigung systematischer Messfehler innerhalb statistischer Modelle zu betonen.

Messungen der bereits diskutierten Reliabilität gehören, ebenso wie die Validität und die Objektivität, zu den klassischen Gütekriterien in der empirischen Sozialforschung (z.B. Atteslander, 2010; Diekmann, 2007; Döring & Bortz, 2016; Friedrichs, 1990; Schnell et al., 2018). Dabei ist es, gerade bei der Konstruktion und Verwendung von Skalen und der Bildung latenter Konstrukte gängige Praxis, die Qualität dieser auf Basis von Messungen der Validität und der Reliabilität zu überprüfen (vgl. z.B. Fischer & Lück, 2014; Mohr & Müller, 2014; Schupp & Gerlitz, 2014; Wingerter, 2014). Entsprechend befindet sich die empirische Sozialforschung gegenwärtig in der Position, in der Skalen entwickelt und verwendet werden, welche mittels Verfahren auf ihre Güte getestet wurden, welche selbst wiederum anfällig sind gegenüber Verzerrungen durch systematische Messfehler.

Messungen der Reliabilität

Inwiefern die Reliabilität grundsätzlich anfällig ist für Verzerrungen durch systematische Messfehler wurde bereits in Kapitel 3.1.1 aufgezeigt, weshalb an dieser Stelle auf eine weitere Diskussion der fundamentalen Logik der Reliabilität verzichtet wird. Stattdessen wird im Folgenden aufgezeigt, wie die spezifische Messung der Reliabilität am Beispiel von Cronbach‘s Alpha (Diekmann, 2007) und Jöreskog's Rho (Jöreskog, 1971) durch das Vorhandensein systematischer Messfehler verzerrt werden.

Wie bereits beschrieben, wird unter der Reliabilität einer Messung deren Messgenauigkeit verstanden, wobei diese im Sinne der klassischen Testtheorie quantifiziert werden kann als Anteil der Varianz des wahren Wertes τ an der Varianz des beobachteten Testwertes von *x* (Moosbrugger, 2012). Der Vollständigkeit wegen sei erwähnt, dass nach Diekmann (2007) drei Methoden zur Bestimmung der Reliabilität angewandt werden können, welche sich allesamt gültig aus den zuvor diskutierten Axiomen und Ableitungen der klassischen Testtheorie ergeben: 1.) die Paralleltest-Methode, bei der eine Messung mit zwei vergleichbaren Messinstrumenten durchgeführt wird, wobei die Korrelation der beiden so entstandenen Messwerte die Paralleltest-Reliabilität angibt; 2.) die Test-Retest-Methode, bei der ein Messinstrument nach einem Zeitintervall erneut angewandt wird und die Korrelation der Messwerte zu den beiden Zeitpunkten der Test-Retest-Reliabilität entspricht sowie zuletzt 3.) die Methode der Testhalbierung (Split-half-Reliabilität), bei der ein Testinstrument mit multiplen Indikatoren in zwei Hälften aufgeteilt wird, wobei die Korrelation zwischen den beiden Testhälften die Split-half-Reliabilität angibt.

Problematisch an der letztgenannten Methode, welche in der Forschungspraxis aufgrund seiner ressourcenschonenden Anwendungsmöglichkeit (ein Paralleltest oder Retest ist nicht nötig) häufig Anwendung findet ist jedoch, dass je nach Aufteilung der Indikatoren unterschiedliche Reliabilitätsschätzungen resultieren können. Behoben werden soll dieses Problem mithilfe des Koeffizienten Cronbach‘s Alpha, welcher als Mittelwert aller möglicher Split-Half-Korrelationen definiert ist und somit als Maß für die interne Konsistenz von Antworten gilt (Krebs & Menold, 2014). Cronbach‘s Alpha wird nach Diekmann (2007) gemäß Formel 34 ermittelt, wobei *N* für die Anzahl an Items, $\bar{c}$ für den Durchschnitt aller Kovarianzen zwischen den Items der aktuellen Personenstichprobe, $\bar{v}$ für die durchschnittliche Varianz jedes Items, σ_x^2 für die Varianz des beobachteten Gesamtscores und $\sigma_{y_i}^2$ für die Varianz der Komponenten *i* für die aktuelle Stichprobe von Personen steht.

$$\alpha = \frac{N\bar{c}}{(\bar{v} + (N-1) * \bar{c})} = \frac{N}{N-1}\left(1 - \frac{\sum_{i=1}^{N} \sigma_{y_i}^2}{\sigma_x^2}\right) \; mit \; x = \sum_{i=1}^{N} y_i \qquad \textbf{Formel 34}$$

Entsprechend Formel 34 ist ersichtlich, dass die Höhe von Cronbach‘s Alpha in Abhängigkeit steht vom Durchschnitt aller Kovarianzen zwischen den Items einer Messung *x*, welche, sowohl gemäß der Logik der klassischen Testtheorie (vgl. Kapitel 3.1.1), als auch der SEM-Logik (vgl.

Kapitel 3.1.2) einen Messfehler ε enthält, welcher, wie ebenfalls zuvor diskutiert, systematische Messfehler wie FRE beinhalten kann. Dass bedeutet gemäß der Diskussion der fundamentalen Logik von FRE (vgl. Kapitel 2.2), dass Cronbach‘s Alpha positiv verzerrt ist, wo Assimilationseffekte auftreten, da diese zu einer Erhöhung der Kovarianz aufeinanderfolgender Items führen und dort negativ verzerrt ist, wo Kontrasteffekte auftreten, da diese zu einer Verringerung der Kovarianz aufeinanderfolgender Items führen (vgl. z.B. Bard & Weinstein, 2016; Cunningham et al., 2016; Garbarski et al., 2014; Huang & Cornell, 2015; Krosnick & Presser, 2010; Magelssen, 2016; Mayerl & Giehl, 2018; Strack & Martin, 1986; Strack & Martin, 1987; Strack et al., 1987; Tourangeau et al., 2000). Hinzu kommt als Besonderheit von Cronbach‘s Alpha im Kontext von CFA und SEM, welches diesbezüglich auch gelegentlich tau-äquivalente Reliabilität genannt wird, dass bei der Berechnung von gleichen Faktorladungen aller Indikatoren ausgegangen wird, was in der Realität in der Regel nicht der Fall ist, wodurch die Berechnung eine weitere Verzerrung erfährt (Lienert & Raatz, Testaufbau und Testanalyse, 1994). Als Alternative zu Cronbach‘s Alpha wird daher, insbesondere bei Analysen mittels CFA und SEM, Jöreskog's Rho vorgeschlagen, welches auch als kongenerische Reliabilität oder Komposit-Reliabilität bezeichnet wird (Cho, 2016).

Der Vorteil der kongenerischen Reliabilität, welche zunächst von Jöreskog (1971) vorgeschlagen wurde, besteht, insbesondere bei der Verwendung von CFA und SEM darin, dass Faktorladungen der Indikatoren innerhalb eines Messmodells nicht homogen sein müssen, wodurch die systematische Unterschätzung der Reliabilität durch tau-äquivalente Messungen wie etwa durch Cronbach‘s Alpha verhindert werden. Cho (2016) schlägt diesbezüglich eine vereinfachte Berechnung der kongenerischen Reliabilität ρ_c vor (vgl. Formel 35), welche unterschiedliche Faktorladungen explizit berücksichtigt. p sei diesbezüglich die Anzahl der Indikatoren des Messmodells, λ_j die Faktorladung des Indikator j und σ_X^2 die Varianz des Testergebnisses:

$$\rho_c = \frac{\left(\sum_{j=1}^{p} \lambda_j\right)^2}{\sigma_X^2} \qquad \textbf{Formel 35}$$

Nun zeigen Mayerl & Giehl (2018) jedoch, dass nicht nur die Kovarianzen, sondern auch die Faktorladungen einer Messung innerhalb von Strukturgleichungsmodellen dort verzerrt sind, wo FRE auftreten. Assimilation führt demnach theoretisch zu erhöhter Kovarianz zwischen aufeinanderfolgenden Items, welche bei der Schätzung von SEM ohne Spezifikation des FRE

als Common Variance statt als Unique Variance geschätzt wird, was die geschätzten Faktorladungen theoretisch erhöht, während Kontrasteffekte die Faktorladungen gemäß der gleichen Logik theoretisch verringern. So führt die unspezifizierte Anwesenheit von FRE nicht nur bei Cronbach‘s Alpha, sondern auch bei Berechnungen der kongenerischen Reliabilität wie beispielsweise Jöreskog’s Rho zu verzerrten und damit unzuverlässigen Messungen der Reliabilität.

Messungen der Validität

Über eine Verzerrung der Reliabilität hinaus führen durch FRE verzerrte Faktorladungen auch zu verzerrten und damit unzuverlässigen Messungen der Validität. Messick (1989, S. 13) definiert Validität als "an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment" und verortet diese somit in einem testtheoretischen Zusammenhang. Lienert (1969, S. 16) definiert etwas allgemeiner und beschreibt, dass „die Validität eines Testes […] den Grad der Genauigkeit an [gibt], mit dem dieser Test dasjenige Persönlichkeitsmerkmal oder diejenige Verhaltensweise, das (die) er messen soll oder zu messen vorgibt, tatsächlich misst". Auch Bühner (2004, S. 36) hält seine Definition möglichst kurz und beschreibt mit der Validität das Ausmaß, mit dem ein Messinstrument dasjenige Phänomen misst, welches es auch messen soll. Somit geht als Schnittstelle dieser Definitionen hervor, dass die Validität eines Messinstrumentes allgemein den Umfang beschreibt, mit dem dasjenige Merkmal einer Untersuchungseinheit erfasst wird, welches auch erfasst werden soll.

In diesem Zusammenhang unterscheidet Diekmann (2007) drei verschiedene Formen der Validität: 1.) die Inhaltsvalidität, welche vorliegt, wenn eine Auswahl an Items die zu messende Eigenschaft in hohem Maße repräsentieren; 2.) die Kriteriumsvalidität, welche angibt, in welchem Maße die mit dem Messinstrument erzielten Ergebnisse mit anderen relevanten Merkmalen korrelieren sowie 3.) die Konstruktvalidität, mit welcher der Fokus auf die Brauchbarkeit des Messinstrumentes für die Entwicklung von Theorien gelegt wird. Krebs & Menold (2014, S. 430ff) kritisieren an dieser Typisierung jedoch, dass für die Beurteilung der Inhaltsvalidität keine objektiven Kriterien existieren, da mit der Forderung nach Inhaltsvalidität vor allem eine intersubjektiv nachvollziehbare Operationalisierung von Konstrukten gemeint ist. Auch die Kriteriumsvalidität kann nur schwer überprüft werden, da in der Praxis häufig keine hinreichend genau definierten, externen und von der Messung unabhängige Kriterien existieren, um diese zu validieren. In der Praxis häufig Anwendung findet hingegen eine Überprüfung der

Konstruktvalidität, wie sie ebenfalls im Kontext der MTMM-Analyse in Kapitel 3.1.3 beschrieben wurde. Diese wird, zur Wiederholung, unterschieden in die konvergente oder konvergierende Validität sowie die divergente oder diskriminante Validität, wobei konvergente Validität dann vorliegt, wenn eine Ähnlichkeit zwischen verschiedenen Items, welche ein bestimmtes Merkmal messen sollen, sowohl behauptet, als auch beobachtet wird. Diskriminante Validität liegt hingegen dann vor, wenn kein Zusammenhang zwischen Items, welche verschiedene Merkmale messen sollen besteht (ebd.).

Entsprechend dieses definitorischen Rahmens ist es gängige Praxis, die Validität anhand der Konstruktvalidität, differenziert nach der konvergenten und der diskriminanten Validität mittels CFA zu überprüfen (vgl. z.B. Allen & Meyer, 1996; Bagozzi, Yi, & Phillips, 1991; Barnett et al., 2016; Cronbach & Meehl, 1955; Krebs & Menold, 2014; Mazurek et al., 2020; Williams, Scalco, & Simms, 2018). Dabei werden latente Konstrukte als konvergent valide betrachtet, wenn deren Indikatoren über eine signifikante Faktorladung und eine standardisierte Faktorladung von > 0,5 verfügen bzw. wenn der AVE („**A**verage **V**ariance **E**xtracted“, der Mittelwert der Quadrate der standardisierten Faktorladungen aller Indikatoren) des gesamten Konstruktes > 0,5 (weniger streng: > 0,4) beträgt. Als diskriminant valide gelten Konstrukte hingegen, wenn Indikatoren eines Konstruktes über keine signifikanten Kreuzladungen mit anderen Konstrukten verfügen (Kline, 2016; Urban & Mayerl, 2014). Somit ist es ersichtlich, dass auch Messungen der Konstruktvalidität für Verzerrungen durch systematische Messfehler wie FRE anfällig sind, da diese die Höhe der Faktorladungen beeinflussen (Mayerl & Giehl, 2018), wie bereits im Kontext der kongenerischen Reliabilität diskutiert.

Es bedarf somit eines Verfahrens zur Kontrolle systematischer Messfehler wie FRE, um zu unverzerrten Schätzungen von Faktorladungen als Indikator der kongenerischen Reliabilität und der Konstruktvalidität zu gelangen, soll die Güte der Schätzung eines Merkmals korrekt, d. h. unverzerrt eingeschätzt werden können. Entsprechende CFA Modelle, welche eine solche Kontrolle speziell für FRE ermöglichen, werden entsprechend im Folgenden abgeleitet und diskutiert.

3.2 Modellierung von Fragereihenfolgeneffekten

Mittels CFA bzw. SEM ist es, wie zuvor beschrieben möglich, systematische Messfehler zu modellieren, indem entweder Kovarianzen auf Messfehlerebene zwischen denjenigen Items, für welche Methodeneffekte erwartet werden, spezifiziert werden oder indem Methodeneffekte

als latente Konstrukte mit Faktorladungen auf diejenigen Items, für welche Methodeneffekte erwartet werden, spezifiziert werden. Während die Spezifikation und Modellierung von Messfehlern somit messfehlerbereinigte Schätzungen von Merkmalen ermöglichen, ist durch die Extraktion der Merkmalsvarianz im Umkehrschluss eine „konstrukt-“ oder „merkmalsbereinigte“ Quantifizierung der Messfehler gegeben. Da diese Modellierung und Quantifizierung eine fundamentale Grundlage zur Analyse besagter Messfehler darstellt, wird im Folgenden erörtert, wie auf Basis dieser Möglichkeit FRE sowohl identifiziert und quantifiziert (vgl. Kapitel 3.2.1), als auch mittels integrierter Moderatoranalyse (vgl. Kapitel 3.2.2) kontrolliert und auf zugrundeliegende Faktoren zurückgeführt werden können. In Kapitel 3.2.2 wird insbesondere die NLC-CFA schrittweise eingeführt. Hierauf basierend erfolgt schließlich die Adaption eines CFA-MTMM Modells sowie des NLC-CFA Modells (vgl. Kapitel 3.3), welche im analytischen Teil dieser Arbeit evaluiert und gegenübergestellt werden.

3.2.1 Identifikation und Quantifizierung von Fragereihenfolgeneffekten

Im Folgenden werden verschiedene methodische Ansätze diskutiert, mit welchen FRE grundsätzlich identifiziert bzw. quantifiziert werden können. Diesbezüglich werden zunächst die klassischen, etablierten Ansätze aufgezeigt und deren Schwachstellen diskutiert, um die Notwendigkeit der Verwendung innovativer Verfahren aufzuzeigen. Hierauf folgt die Ableitung eines einfachen CFA-MTMM Modells, auf dessen Basis im Allgemeinen systematische Messfehler zwischen Kontext- und Zielfrage (welche als FRE angenommen werden) identifiziert werden können. Abschließend erfolgt eine Adaption dieses Modells unter Verwendung regressiver Beziehungen statt der Spezifikation von Kovarianzen zwischen Kontext- und Zielfrage in Abgrenzung an das klassische Vorgehen in CFA-MTMM Analysen. Das einfache CFA-MTMM Modell sowie die CFA unter Verwendung regressiver Beziehungen sind die Grundlage des CFA-MTMM Modells sowie des NLC-CFA Modells zur Identifikation und Kontrolle von FRE, welche im weiteren Verlauf von Kapitel 3.2 sowie in Kapitel 3.3 eingeführt werden.

3.2.1.1 Klassische Ansätze

Ein klassischer Ansatz zur Identifikation von FRE besteht darin, dass die Fragereihenfolge zweier Items einer Testbatterie experimentell variiert wird, so dass innerhalb der Experimentalbedingung die Zielfrage unmittelbar auf die Kontextfrage folgt, während innerhalb der Kontrollbedingung zunächst die Zielfrage und darauffolgend die Kontextfrage präsentiert wird.

Sollte der Stimulus durch die Kontextfrage das Antwortverhalten auf die Zielfrage beeinflussen, so ist dies nur unter der Experimentalbedingung der Fall, da innerhalb der Kontrollbedingung der Stimulus erst nach Beantwortung der Zielfrage erfolgt. Unterscheiden sich in der Folge die Korrelationen von Kontext- und Zielfrage zwischen den beiden Gruppen systematisch, so gilt ein FRE als identifiziert (vgl. z.B. Bard & Weinstein, 2016; Cunningham et al., 2016; Garbarski et al., 2014; Huang & Cornell, 2015; Magelssen, 2016; Stark et al., 2018; Strack & Martin, 1987).

Der Logik dieses Vorgehens in Anlehnung an das Verknüpfungsaxiom der klassischen Testtheorie (vgl. Formel 7, Kapitel 3.1.1) liegt zugrunde, dass wenn sich jede Messung *x* (und *y*) zusammensetzt aus dem wahren Wert (τ) und einem Messfehler (ε), dann entspricht die Korrelation *r* der Messungen *x* und *y* einer Funktion aus den wahren Werten sowie der Messfehler. Beeinflussen systematische Messfehler dann die Messung von *x* (und *y*) qua deren Messfehler ε, so wird auch die Korrelation *r* zwischen beiden Variablen durch diese beeinflusst, da diese eine Funktion unter anderem der Messfehler ist.

Für die Kontrollbedingung (Treatmentstatus 0) gilt nun, da das Treatment „Kontextfrage“ bei Beantwortung der Zielfrage nicht erfolgt ist, sowie unter Berücksichtigung des Unabhängigkeitsaxioms (vgl. Formel 8, Kapitel 3.1.1) sowie der Unkorreliertheit von Messfehlern zwischen Items (vgl. Formel 9, Kapitel 3.1.1) für zufällige Messfehler (da keine systematischen Messfehler erwartet werden), dass Die Korrelation zwischen den Messungen y und x keine Kovarianz beinhaltet, welche auf systematische Messfehler zurückgeht. Für die Experimentalbedingung (Treatmentstatus 1) gilt dies jedoch nicht. Da hier FRE, sprich systematische Messfehler erwartet werden, verändert sich der Anteil an (geteilter) Fehlervarianz innerhalb der Messungen y und x, was sich ebenfalls in der Korrelation zwischen beiden Messungen niederschlägt. Innerhalb dieses klassischen Ansatzes ist ein FRE somit identifiziert, wenn gilt:

$$r_{yx}^{0} \neq r_{yx}^{1} \qquad \textbf{Formel 36}$$

Aber auch wenn dieses relativ simple Verfahren dazu geeignet ist, die Existenz eines systematischen Messfehlers aufzuzeigen, welcher auf eine Variation der Fragereihenfolge zurückgeführt werden kann (was schließlich als FRE angenommen wird), ergibt sich hieraus keine Möglichkeit, diesen statistisch zu kontrollieren, da die Höhe des systematischen Messfehlers nicht geschätzt und quantifiziert wird. Zwar könnte die Differenz beider Korrelationskoeffizienten

hierüber Aufschluss geben, jedoch gilt zu bedenken, dass auch innerhalb der Kontrollgruppe FRE auftreten können, da sich Kontext- und Zielfrage lediglich verkehren und so auch der Stimulus durch die Zielfrage einen Einfluss auf die Beantwortung der Kontextfrage ausüben kann.

Eine Lösung für das zuletzt genannte Problem bietet die Verwendung computergestützter Befragungen dann, wenn die Reihenfolge von Kontext- und Zielfrage einer Fragebatterie innerhalb der Kontrollbedingung nicht lediglich vertauscht, sondern über alle Befragte der Kontrollgruppe und alle Items innerhalb der Fragebatterie komplett randomisiert ist. Bei ausreichend hoher Fallzahl sollte so der FRE aus der Berechnung eines Korrelationskoeffizienten für die Kontrollgruppe auspartialisiert werden, da die Zielfrage zufällig relativ zur Kontrollfrage präsentiert wird. Die Differenz der Korrelationskoeffizienten zwischen Experimental- und Kontrollgruppe kann somit, zumindest auf der Aggregatsebene, Aufschluss über die Höhe des Effekts geben, da in der Kontrollgruppe keine Reihenfolgeneffekte zu erwarten sind (vgl. z.B. Mayerl & Giehl, 2018; Mayerl, Andersen, & Giehl, 2019).

3.2.1.2 CFA-MTMM Ansatz

Bei Anwendung des klassischen Ansatzes verbleibt die Problematik, dass die Korrelation zwischen Kontext- und Zielfrage sowohl auf den Anteil der Kovarianz, welche auf ein gemeinsames zugrundeliegendes Merkmal von Interesse, als auch auf den Anteil der Kovarianz, welche auf einen systematischen Messfehler (den FRE) zurückgeht, wodurch der eigentliche Effekt der Fragereihenfolge auf die Beantwortung der Zielfrage nicht separat spezifiziert und damit ebenfalls nicht identifiziert, quantifiziert und auch nicht kontrolliert wird. Eine Lösung hierfür bietet die Verwendung von CFA-MTMM Modellen (vgl. Kapitel 3.1.3), in welchen die Varianz jeder Messung *y* in den Anteil der Varianz, welche auf ein gemeinsames zugrundeliegendes Merkmal, den Anteil systematischer Fehlervarianz und den Anteil zufälliger Fehlervarianz zerlegt wird. Da FRE jeweils eine Dyade an aufeinanderfolgenden Items betrifft (die Kontext- und Zielfrage), können diese, wird die MTMM-Logik adaptiert, durch die Spezifikation von Kovarianzen auf Messfehlerebene $\varepsilon_{j+1,j}$ (bzw. $\delta_{j+1,j}$ nach X-Notation, vgl. Abbildung 4 und 5) für aufeinanderfolgende Items bzw. Indikatoren einer Fragebatterie zur Schätzung eines zugrundeliegenden latenten Merkmals modelliert werden. Wird also innerhalb eines CFA Modells zwischen ε_j, dem Messfehler der Kontextfrage, und ε_{j+1}, dem Messfehler der Zielfrage, eine Kovarianz $\varepsilon_{j+1,j}$ modelliert, so ist mit dieser der gemeinsame Varianzanteil der Messung *j* und *j+1*

quantifiziert, welcher nicht auf den gemeinsamen Varianzanteil des zugrundeliegenden Merkmals bzw. latenten Konstruktes zurückgeht und welcher folglich die systematische Messfehlervarianz, in diesem Fall des FRE, darstellt (vgl. Kapitel 3.1.2 und 3.1.3).

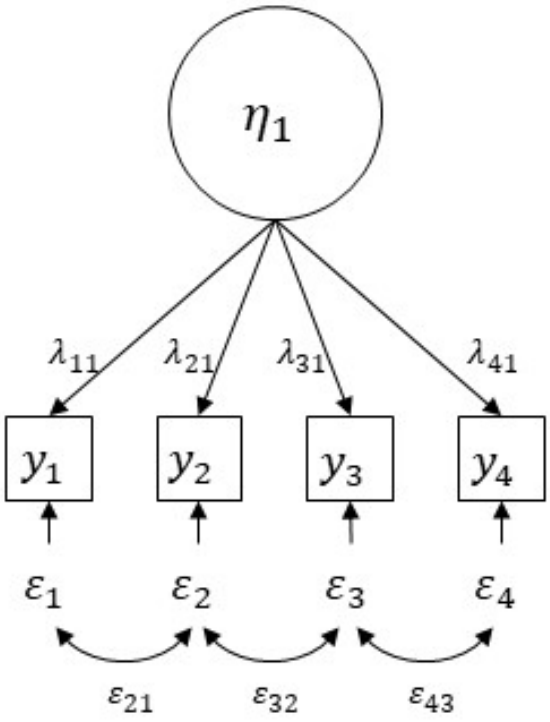
Wird diese Modellierung nun mit dem weiter oben beschriebenen experimentellen Design kombiniert, in welchem die Experimentalgruppe eine fixierte Fragereihenfolge (Treatmentstatus 1: Die Zielfrage *j+1* folgt auf die Kontextfrage *j*) und die Kontrollgruppe eine randomisierte Fragereihenfolge (Treatmentstatus 0: Die Zielfrage *j+1* wird komplett randomisiert relativ zur Kontextfrage *j* innerhalb der Fragebatterie präsentiert) erhält, so sollte innerhalb der Experimentalbedingung $\varepsilon_{j+1,j}$ die Fehlervarianz des FRE extrahieren, während innerhalb der Kontrollbedingung $\varepsilon_{j+1,j}$ keine Fehlervarianz eines Reihenfolgeneffektes beinhaltet (vgl. Mayerl & Giehl, 2018; Kapitel 3.1.2, Kapitel 3.1.3 und Kapitel 3.2.2.1). Für dieses Vorgehen gilt ein systematischer Messfehler, welcher systematisch durch die Variation der Fragereihenfolge bedingt wird (wobei plausibel angenommen werden kann, dass es sich hierbei um einen FRE handelt) als identifiziert (und zusätzlich durch $\varepsilon_{j+1,j}$ als quantifiziert), wenn gilt:

$$\varepsilon^{0}_{j+1,j} \neq \varepsilon^{1}_{j+1,j} \qquad \textbf{Formel 37}$$

Abbildung 11 zeigt diesbezüglich ein latentes Konstrukt (als zu schätzendes Merkmal von Interesse für die Befragung) mit 4 Indikatorvariablen und spezifizierten FRE zwischen allen aufeinanderfolgenden Kontext-Zielfragen-Dyaden von y_1 bis y_4 durch Spezifikation von Kovarianz auf Messfehlerebene in Anlehnung an den MTMM-Ansatz (vgl. Kapitel 3.1.3). Zu beachten ist diesbezüglich, dass solche Effekte grundsätzlich zwischen jeder Kontext-Zielfragen-Dyade modelliert werden können, für welche solche Effekte zu erwarten sind, wobei an dieser Stelle vereinfacht angenommen wird, dass FRE zwischen allen aufeinanderfolgenden Items auftreten. Die Anzahl spezifizierbarer Kovarianzen bzw. FRE ist in diesem Sinne im Wesentlichen durch die Anzahl der zur Verfügung stehenden Informationen in der Varianz-Kovarianzmatrix $\boldsymbol{S}$ (vgl. Kapitel 3.1.2) und entsprechend der übrig bleibenden Freiheitsgrade, respektive der Bevorzugung sparsamer Modelle limitiert, um eine Unteridentifikation des Modells zu vermeiden (Brown, 2006). Formel 38 zeigt zusätzlich das modellspezifische Gleichungssystem auf. Zwecks vereinfachter Einführung des Modells erfolgt die Darstellung lediglich unter Annahme

einer vorhandenen Gruppe statt zweier vorhandener Gruppen entsprechend des weiter oben beschriebenen experimentellen Designs. Die Modellspezifikation der CFA-MTMM mit mehreren Gruppen erfolgt in Kapitel 3.2.2.

Abbildung 11: Einfaches CFA-MTMM Modell mit spezifizierter Kovarianz auf Messfehlerebene



$$(1)\quad \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \\ \lambda_{41} \end{bmatrix} * \eta_1 + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

Formel 38

$$(2)\quad \boldsymbol{\Theta}_\varepsilon = \begin{bmatrix} \theta_{\varepsilon_1} & & & \\ \theta_{\varepsilon_{21}} & \theta_{\varepsilon_2} & & \\ 0 & \theta_{\varepsilon_{32}} & \theta_{\varepsilon_3} & \\ 0 & 0 & \theta_{\varepsilon_{42}} & \theta_{\varepsilon_4} \end{bmatrix}$$

Der Verfügbarkeit von Freiheitsgraden und der Bevorzugung sparsamer Modelle ist es ebenfalls geschuldet, dass die Spezifikation des systematischen Messfehlers auf Basis einer Kovarianz zwischen den Messfehlern von Kontext- und Zielfrage und nicht auf Basis eines latenten

Methodenfaktors (wie ebenfalls in CFA-MTMM Modellen üblich, vgl. Kapitel 3.1.3) mit je einer Faktorladung auf die Kontext- und die Zielfrage erfolgt, denn: Die Spezifikation eines solchen Methodenfaktors bedingt die Schätzung von zwei zusätzlichen Parametern (entweder von zwei Faktorladungen, wenn die Varianz des Faktors auf 1 fixiert wird, oder von der Varianz des Faktors und von einer Faktorladung, wenn die andere Faktorladung auf 1 fixiert wird) je Kontext-Zielfragen-Dyade. Die Spezifikation einer Kovarianz zwischen den Messfehlern erfordert hingegen lediglich die Schätzung eines zusätzlichen Parameters je Dyade, namentlich der Kovarianz. Bei Schätzung eines Methodenfaktors statt einer Kovarianz wäre es zwar ebenfalls möglich, beide Faktorladungen auf 1 zu fixieren und lediglich die Varianz des Faktors frei zu schätzen bzw. die Varianz des Methodenfaktors auf 1 zu fixieren und die beiden Faktorladungen als gleich zu schätzen, was summarisch die gleiche Anzahl zu schätzender Parameter wie bei Schätzung einer Kovarianz ergibt. Algebraisch wäre in diesem Fall die Varianz des Methodenfaktors bzw. die geschätzte Faktorladung allerdings direkt in die geschätzte Kovarianz überführbar, so dass sich aus einem solchen Vorgehen kein Vorteil ergibt, die Modellspezifikation jedoch komplexer wäre. Es bleibt somit bei der Bevorzugung von Kovarianzen auf Messfehlerebene zur Spezifikation des FRE innerhalb des einfachen CFA-MTMM Modells.

3.2.1.3 Regressive Beziehungen in CFA Modellen

Wird davon ausgegangen, dass FRE eine Beeinflussung des Antwortverhaltens der Zielfrage y_{j+1} durch das Antwortverhalten der Kontextfrage y_j implizieren und dass die Messung der Zielfrage darüber hinaus durch das zu schätzende Merkmal η bedingt ist sowie einen zufälligen Messfehler ε beinhaltet, so kann davon ausgegangen werden, dass die Messung der Zielfrage y_{j+1} bei Anwesenheit von FRE eine Funktion des zu schätzenden Merkmals η, der Kontextfrage y_j und eines zufälligen Messfehlers ε ist:

$$y_{j+1} = f(\eta, y_j, \varepsilon) \qquad \textbf{Formel 39}$$

Ein entsprechender funktionaler Zusammenhang kann durch regressive Beziehungen auf der Ebene der Indikatoren statt der zu schätzenden latenten Merkmale innerhalb von SEM modelliert werden, um den konkreten kausalen Einfluss des Antwortverhaltens der Kontextfrage auf die Zielfrage zu würdigen, was eher der Logik von FRE entspricht als die Modellierung von

FRE als Kovarianz auf Messfehlerebene, wie für CFA-MTMM Modelle vorgeschlagen (vgl. Kapitel 3.2.1.2).

Die Spezifikation von FRE erfolgt auf Indikatorebene entsprechend der Logik der seriellen Mediation, so diese zwischen allen Kontext-Zielfragen-Dyaden der Indikatoren eines latenten Konstruktes spezifiziert werden , wobei das spezifizierte latente Konstrukt den Anteil an Common Variance des zu schätzenden Merkmals extrahiert (vgl. Abbildung 12Abbildung 12: CFA-Modell mit spezifiziertem Fragereihenfolgeneffekt)[5]. Ein solches Vorgehen für die Spezifikation von FRE innerhalb von Mess- bzw. Strukturgleichungsmodellen hat den bereits erwähnten Vorteil, dass eine regressive Beziehung tatsächlich einen kausalen Zusammenhang zwischen Kontext- und Zielfrage impliziert (das Antwortverhalten auf die Kontextfrage bedingt das Antwortverhalten der Zielfrage). Zum anderen quantifiziert eine Kovarianz auf Messfehlerebene $\varepsilon_{j+1,j}$ (wir für CFA-MTMM vorgeschlagen) lediglich die Kovarianz der faktorbereinigten Messfehler ε_j und ε_{j+1}, während für FRE die Messung der Zielfrage y_{j+1} nicht lediglich von dem Messfehler der Kontextfrage ε_j, sondern von dem allgemeinen Antwortverhalten auf die Kontextfrage y_j (qua der hierdurch aktivierten Informationen, vgl. Kapitel 2.3) abhängig ist. Der systematische Messfehler der Zielfrage ε_{j+1} und somit die gesamte Messung der Zielfrage y_{j+1} wird somit durch die Messung der Kontextfrage y_j und nicht lediglich durch den quantifizierten Messfehleranteil der Kontextfrage ε_j bestimmt, wodurch eine Kovarianz auf Messfehlerebene eine unzureichende Quantifizierung eines FRE darstellt. Wird der Einfluss des FRE auf das Antwortverhalten von *y* also durch eine regressive Beziehung dargestellt, so gilt für die Messung der Kontextfrage y_j (vgl. Formel 40.1) sowie der Zielfrage y_{j+1} (vgl. Formel 40.2) entsprechend der in Kapitel 3.1 eingeführten Notation:

$$(1)\quad y_j = \lambda_{j,1}\eta_1 + \varepsilon_j$$

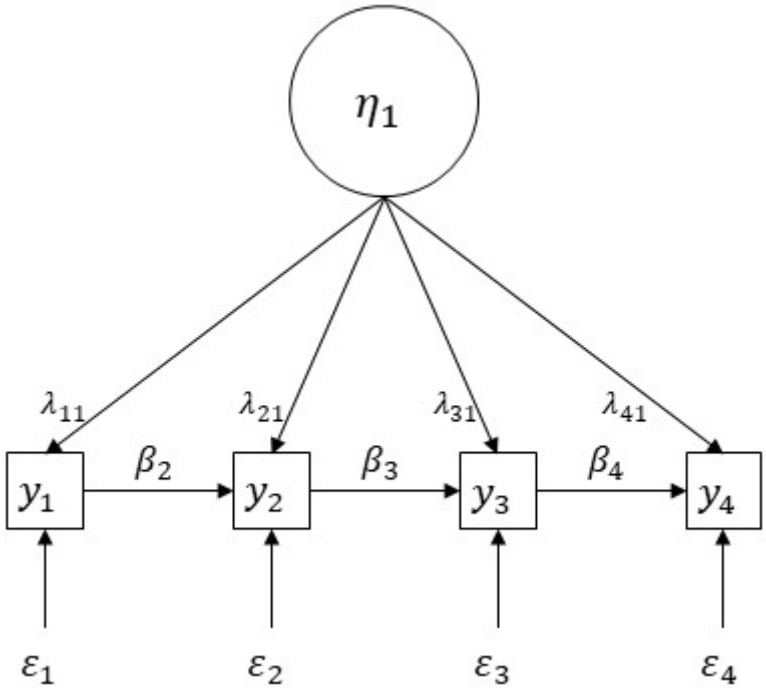
$$(2)\quad y_{j+1} = \lambda_{j+1,1}\eta_1 + \beta_{j+1}y_j + \varepsilon_{j+1}$$

Formel 40

[5] CFA Modelle, welche serielle Mediation für autoregressive latente Variablenmodelle ermöglichen, wurden bereits vorgeschlagen (Sivo, 2001; Biesanz, 2012). Für die vorgeschlagenen Modelle wird die autoregressive Beziehung bzw. serielle Mediation jedoch auf der Ebene der latenten Variablen statt auf der Ebene der Indikatoren modelliert, so dass diese für das vorliegende Vorhaben keine Anwendung finden und somit unberücksichtigt bleiben.

mit β_{j+1} als Quantifizierung des faktorbereinigten FRE zwischen den Messungen der Kontextfrage y_j und der Zielfrage y_{j+1}. Dieses Formelsystem wird in Abbildung 12 in ein einfaches CFA Modell mit spezifizierten FRE auf Indikatorebene überführt. Das Beispielmodell ist erneut mit vier Indikatoren y_1 bis y_4 und einem latenten Konstrukt spezifiziert. FRE sind zwischen jeder Kontext-Zielfragen-Dyade spezifiziert, wobei gilt, dass solche Effekte zwischen jeder Dyade spezifiziert werden können, wo dies theoriegeleitet sinnvoll ist. Es bleibt lediglich auch hier die Tendenz zu sparsamen Modellen und die allgemeine Identifizierbarkeit von Modellen zu bedenken. In Formel 41 ist zudem das Gleichungssystem zur Schätzung des entsprechenden Modells dargestellt.

Abbildung 12: CFA-Modell mit spezifiziertem Fragereihenfolgeneffekt



Entsprechend ist bei Anwendung eines solchen Vorgehens mit β_{j+1} der faktorbereinigte FRE der Messung *j*+1 quantifiziert. Unter Berücksichtigung eines experimentellen Designs mit Fixierung der Fragereihenfolge unter der Experimentalbedingung und einer kompletten Randomisierung der Fragereihenfolge aller Items unter der Kontrollbedingung, wie in Kapitel 3.2.1.1 beschrieben, gilt ein FRE schließlich als identifiziert, wenn Formel 42 gilt.

$$(1)\quad \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \\ \lambda_{41} \end{bmatrix} * \eta_1 + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \beta_2 & 0 & 0 \\ 0 & 0 & \beta_3 & 0 \\ 0 & 0 & 0 & \beta_4 \end{bmatrix} * \begin{bmatrix} 0 \\ y_1 \\ y_2 \\ y_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

Formel 41

$$(2)\quad \boldsymbol{\Theta}_{\varepsilon} = \begin{bmatrix} \theta_{\varepsilon_1} & & & \\ 0 & \theta_{\varepsilon_2} & & \\ 0 & 0 & \theta_{\varepsilon_3} & \\ 0 & 0 & 0 & \theta_{\varepsilon_4} \end{bmatrix}$$

$$\beta_{j+1}^{0} \neq \beta_{j+1}^{1}$$

Formel 42

Über die Identifikation von FRE hinaus ist es zentral für diese Arbeit und die zu entwickelnden Modelle, ebenfalls Möglichkeiten abzuleiten, mit welchen FRE auf zugrundeliegende Merkmale rückgeführt, sprich, kontrolliert werden können. Eine Diskussion dieser Möglichkeiten erfolgt im folgenden Kapitel.

3.2.2 Kontrolle von Fragereihenfolgeneffekten mittels Moderatoranalyse

Wird der FRE, wie im vorangegangenen Kapitel beschrieben, als Kovarianz bzw. regressive Beziehung zwischen in Befragungen aufeinanderfolgenden Items einer Fragebatterie zur Schätzung eines latenten Merkmals von Interesse modelliert, so ergibt sich als Möglichkeit, diesen zu kontrollieren und dessen Auftreten auf zugrundeliegende Determinanten zurück zu führen, die Moderatoranalyse. Ziel der Moderatoranalyse ist es zu analysieren, ob Effekte zwischen den Merkmalen *x* und *y* stabil sind oder ob sich diese hinsichtlich eines weiteren Merkmals *z* unterscheiden. Die Moderatoranalyse lässt sich grundlegend unterscheiden in die sogenannte Multigruppenanalyse, sowie die Interaktionsanalyse (Urban & Mayerl, 2018). Es ist folglich das Ziel dieses Kapitels Möglichkeiten aufzuzeigen, wie FRE mittels der beiden Verfahren analysiert und kontrolliert werden können sowie Vor- und Nachteile der jeweiligen Verfahren aufzuzeigen.

Diesbezüglich wird im Folgenden zunächst eine Einführung in die Multigruppenanalyse gegeben und aufgezeigt, wie FRE auf Basis dieser Logik sowohl für CFA-MTMM Modelle durch

Kovarianzen auf Messfehlerebene, als auch für CFA Modelle durch regressive Beziehungen auf Indikatorebene auf zugrundeliegende Moderatoren zurückgeführt und dadurch kontrolliert werden können. Die Beschreibung dieser Methode und der entsprechenden Modelle ist dabei zentral für die konkrete Herleitung desjenigen CFA-MTMM Modells in Kapitel 3.3, welches im analytischen Teil dieser Arbeit verwendet wird. Eine Diskussion der Nachteile dieses Verfahrens zeigt hingegen ebenfalls auf, unter welchen Bedingungen es effektiver ist, FRE stattdessen mittels des einzuführenden NLC-CFA Modell zu identifizieren und zu kontrollieren, welche eine Kombination von Multigruppen- und Interaktionsanalyse darstellt (vgl. Kapitel 3.2.2.1).

Entsprechend erfolgt anschließend eine Einführung in die Interaktionsanalyse, wobei insbesondere aufgezeigt wird, wie FRE mithilfe dieser innerhalb von CFA Modellen auf manifeste Moderatoren zurückgeführt werden können, was ein zentraler Zwischenschritt in der Einführung der NLC-CFA darstellt (vgl. Kapitel 3.2.2.2).

Über die Verwendung manifester Moderatoren hinaus wird schließlich die Verwendung von latenten Interaktionstermen diskutiert, was als letzter methodischer Baustein für die Einführung der NLC-CFA angesehen werden kann. Es folgt entsprechend zuletzt die Ableitung eines analytischen NLC-CFA Modells zur Identifikation und Kontrolle von FRE (vgl. Kapitel 3.2.2.3), welches in Kapitel 3.3 für die im theoretischen Teil dieser Arbeit identifizierten Moderatoren (Modus der Informationsverarbeitung, Zugänglichkeit von Informationen bezüglich dem, Objekt der Befragung, Framing der Befragtenrolle) konkretisiert und schließlich im analytischen Teil dieser Arbeit untersucht und dem CFA-MTMM Modell gegenübergestellt wird.

3.2.2.1 Multigruppenanalyse

In der quantitativen Methodenforschung ist es gängige Praxis, systematische Messfehler mittels Multigruppenanalysen zu identifizieren. Dies ist primär darin begründet, dass den verschiedenen Identifikationsverfahren experimentelle Designs zugrunde liegen. So wird innerhalb der Experimentalgruppe ein spezifischer Stimulus präsentiert, welcher in der Kontrollgruppe fehlt und welcher das Auftreten der Messfehler „provozieren“ soll, wie es beispielsweise hinsichtlich der experimentellen Variation von Fragereihenfolgen in Kapitel 3.2.1 beschrieben wurde. Mithilfe von Multigruppenanalysen und der experimentellen Variation von methodenrelevanten Stimuli können beispielsweise Akquieszenzeffekte (vgl. z.B. Kam & Meyer, 2015; Mayerl & Giehl, 2018), die bereits beschriebenen FRE (vgl. z.B. Mayerl & Giehl, 2018; Schreier, Biethahn, & Drewes, 2018) oder allgemeine Antwortverzerrungen, welche zu einer Verzerrung von

Messungen der Datenqualität führen (vgl. z.B. Mayerl & Giehl, 2018; Silber, Danner, & Rammstedt, 2018; Verdam, Oort, & Sprangers, 2017) identifiziert werden.

Darüber hinaus besteht jedoch auch die Möglichkeit, über die experimentellen Konditionen hinaus zusätzliche Gruppen mittels möglicher Moderatorvariablen wie etwa der Nationalität (Lee, Schwarz, & Goldstein, 2014; Stark et al., 2018) oder der Antwortreaktionszeit (Mayerl et al., 2019) zu spezifizieren, so dass in solchen Analysen (in der Regel) z x q Gruppen analysiert und verglichen werden, wobei für dieses Beispiel z für die Anzahl an Experimental- und Kontrollgruppen steht und q die Anzahl an Merkmalsausprägungen innerhalb der weiteren Moderatorvariablen repräsentiert. Darüber hinaus können zusätzliche Moderatorvariablen spezifiziert werden, wobei die Anzahl an zu analysierenden Gruppen durch die Hinzunahme weiterer Moderatorvariablen bzw. zu differenzierenden Merkmalsausprägungen weiter ansteigt, wodurch die Komplexität der Analyse zunimmt. Beispielsweise würde die Analyse von drei dichotomen Moderatorvariablen bereits die Modellierung und Analyse von acht verschiedenen Gruppen (2 x 2 x 2 = 8) bedingen. Die Anzahl möglicher Moderatorvariablen ist somit, je nach Verfahren, welches der Multigruppenanalyse zugrunde liegt, durch die mindestens benötigte Fallzahl innerhalb der einzelnen Gruppen sowie die zunehmende Komplexität der Analyse limitiert.

Folglich ist die Multigruppenanalyse vor allem bei Vorliegen diskret skalierter Moderatorvariablen mit einer geringen Anzahl an Merkmalsausprägungen geeignet, da dies zu einer überschaubaren Anzahl an miteinander zu vergleichenden Gruppen führt. Gelegentlich wird sie aber auch für kontinuierlich skalierte Moderatorvariablen wie etwa Antwortlatenzen verwendet (vgl. z.B. Mayerl et al., 2019), wobei solche, häufig mittels Median-Split, in Gruppenvariablen rekodiert werden, so dass eine Extremgruppenanalyse durchgeführt werden kann (Urban & Mayerl, 2018), was insgesamt, qua einer Verringerung der maximalen Gruppenanzahl, die Komplexität der Analysen reduziert[6]. Für Multigruppenanalysen mithilfe von CFA/SEM ist es darüber hinaus möglich, nicht nur kontinuierliche, sondern ebenfalls latente Variablen zur Gruppierung zu verwenden. Für solche Fälle hat sich ein mehrstufiges Verfahren bewährt, in welchem zunächst die latente Moderatorvariable geschätzt und dann mittels Mediansplit in eine Gruppierungsvariable rekodiert wird. Anschließend kann besagte Variable zur Gruppierung innerhalb einer Multigruppen CFA- oder SEM-Analyse verwendet werden (vgl. z.B. Mayerl, 2009; Mayerl, 2013).

[6] Zu beachten ist zudem, dass die maximale Anzahl möglicher Gruppen durch die Fallzahl der Stichprobe limitiert ist (Urban & Mayerl, 2018).

Der Logik der Multigruppenanalyse folgend können mit dieser Moderatoreffekte (am Beispiel einer Regressionsanalyse) über den Vergleich bzw. die Differenz der Einflussstärke, Einflussrichtung oder der Signifikanz eines unstandardisierten Effekts *b* zwischen den Personengruppen einer Moderatorvariable *z* identifiziert werden (Urban & Mayerl, 2018). Ein Moderatoreffekt gilt demnach als Identifiziert, wenn nach Formel 42 sinngemäß gilt: $\beta_j^0 \neq \beta_j^1$.

Multigruppen(MG)-CFA Analysen

Sollen systematische Messfehler und insbesondere FRE innerhalb von Multigruppen(MG)-CFA Modellen entweder durch Kovarianzen auf Messfehlerebene oder durch regressive Beziehungen auf Indikatorebene identifiziert und kontrolliert werden, gilt es insbesondere, die sogenannte Messinvarianz der Modelle zu beachten. Als Messinvarianz wird innerhalb von CFA (und SEM) der Grad bezeichnet, mit welchem bestimmte Parameter des Modells zwischen den Gruppen theoriegeleitet als identisch angenommen und somit als gleich geschätzt werden. Dabei gilt zu beachten, dass je nach zu analysierenden Schätzern unterschiedliche Grade an Messinvarianz erreicht werden müssen, so die entsprechenden Schätzer zwischen den Gruppen verglichen werden sollen. Für die vergleichende Analyse von Kovarianzen bzw. regressiven Beziehungen auf der Ebene der latenten Konstrukte ist es entsprechend eine Bedingung, dass mindestens metrische Invarianz vorliegen muss, sprich, dass zusätzlich zur identischen Modellierung der Messmodelle (bezeichnet als konfigurale Invarianz) alle Faktorladungen zwischen Items und den jeweiligen Konstrukten zwischen den Gruppen identisch sein müssen. Dies ist insofern notwendig, als dass nur hierdurch eine gleiche, theoretische Skalierung der latenten Konstrukte zwischen den Gruppen gewährleistet werden kann und die konkrete Höhe der geschätzten Kovarianzen und regressiven Beziehungen von der theoretischen Skalierung der latenten Faktoren abhängig ist. Wäre die theoretische Skalierung der latenten Konstrukte zwischen den Gruppen verschieden, so könnten unterschiedlich ausgeprägte Kovarianzen bzw. regressive Beziehungen entsprechend nicht mehr eindeutig auf die Ausprägungen der zur Gruppierung verwendeten Moderatorvariable zurückgeführt werden, sondern ebenfalls aus unterschiedlichen Skalierungen der latenten Konstrukte resultieren (Brown, 2006).

Die Spezifikation metrischer Invarianz für MG-CFA Modelle, auf deren Basis systematische Messfehler durch Kovarianzen oder regressive Beziehungen auf Indikatorebene analysiert werden sollen, ist jedoch ebenfalls sinnvoll. Zum einen ist es schlicht plausibel, dass (über ihre Faktorladungen) valide und reliabel geschätzte latente Merkmale unabhängig sein sollten vom Einfluss systematischer Messfehler und sich die Schätzungen folglich nicht zwischen verschie-

denen Gruppen, welche verschiedene Einflüsse systematischer Messfehler repräsentieren, unterscheiden sollten, so die Varianz der systematischen Messfehler gesondert extrahiert wird. Zum anderen führt die Spezifikation gleicher Faktorladungen (und somit gleicher Schätzmodelle der latenten Merkmale) zwischen solchen Gruppen gerade dazu, dass Unterschiede in den Varianzen und Kovarianzen der manifesten Indikatoren klarer als (systematische) Messfehler innerhalb der Gruppen identifiziert werden können. Als Gleich geschätzte Faktorladungen zwischen solchen Gruppen (in welchen der Einfluss systematischer Messfehler zwischen den Gruppen unterschiedlich wirkt) sollten entsprechend geringer durch systematische Messfehler verzerrt werden, da der entsprechende Einfluss durch Kovarianzen bzw. regressive Beziehungen spezifisch berücksichtigt wird und somit aus den Faktorladungen auspartialisiert werden kann. Dies bedingt dann jedoch, dass die Messungen der manifesten Indikatoren zwischen den Gruppen vergleichbar sind, was in der Regel gegeben ist, so die Skala der Indikatoren einer Fragebatterie nicht ebenfalls experimentell variiert wurde. Mit anderen Worten: Werden Faktorladungen sowie die Varianz des latenten Konstruktes in einem MG-CFA Modell zwischen den Gruppen als gleich geschätzt, in welchem die Gruppierungsvariable den Einfluss eines systematischen Messfehlers widerspiegelt, und werden lediglich die Messfehler der einzelnen Indikatoren sowie die Messfehlerkovarianzen und regressiven Beziehungen zwischen den Indikatoren frei geschätzt, so müssen sich zwischen den Gruppen verschiedene Einflüsse systematischer Messfehler zwangsläufig in eben diesen frei geschätzten Parametern manifestieren.

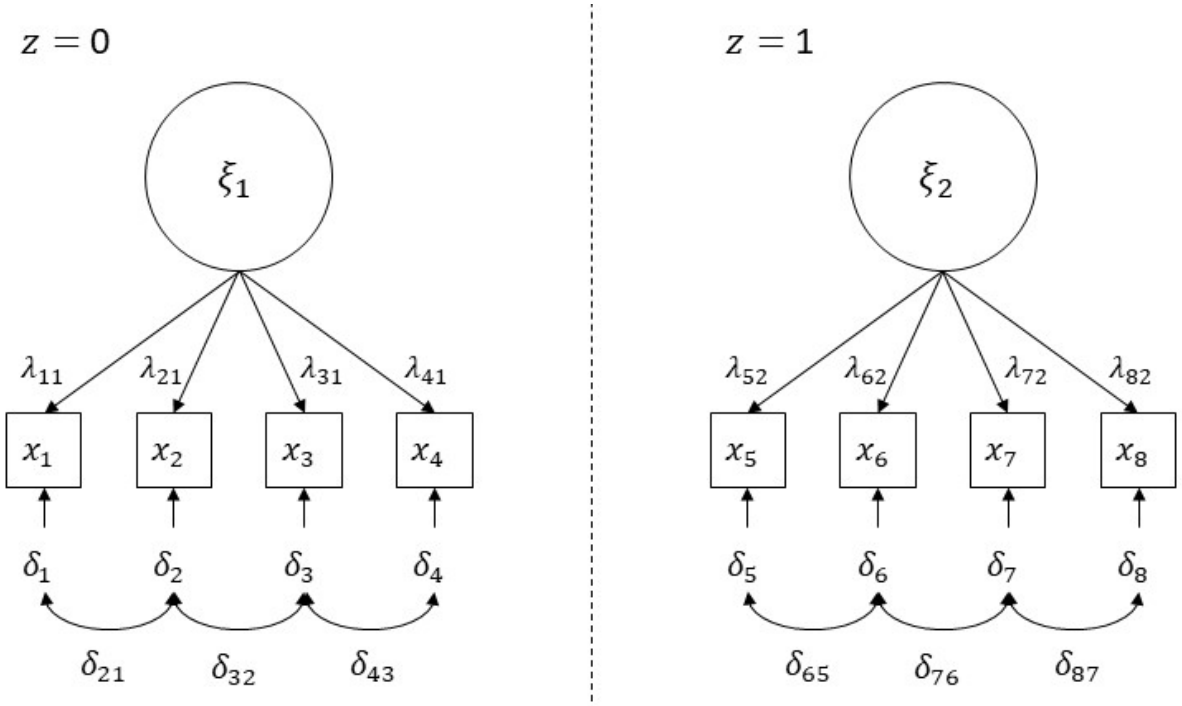
Herausfordernd ist diese Bedingung bei der Analyse von FRE, da Faktorladungen in jenen Gruppen, in denen besagte Effekte auftreten, aufgrund der messfehlerbedingt erhöhten Kovarianz zwischen aufeinanderfolgenden Items verzerrt sein können, wodurch verhindert werden kann, dass metrische Invarianz bei eigentlich akzeptablem Model Fit erreicht wird (Mayerl & Giehl, 2018), auch wenn diese Problematik durch die Einführung partieller Invarianz (Byrne, Shavelson, & Muthén, 1989; Hsiao & Lai, 2018; Lai, Richardson, & Mak, 2019; Liu & Thompson, 2022; Pokropek, Davidov, & Schmidt, 2019) bzw. approximierter Invarianz (Cieciuch, Davidov, Algesheimer, & Schmidt, 2017; Muthén & Asparouhov, 2013; Pokropek et al., 2019) abgemildert werden kann. Dennoch ist hierdurch impliziert, dass das erreichen metrischer Invarianz mit zunehmender Anzahl an zu vergleichenden Gruppen ebenfalls zunehmend herausfordernd und komplex wird.

MG-CFA-MTMM Modell

Hinsichtlich der Entwicklung eines CFA-MTMM Modells zur Identifikation und Kontrolle von FRE ist es somit möglich, innerhalb eines MG-CFA Modells mit metrischer Invarianz durch

Spezifikation von Kovarianzen auf Messfehlerebene entsprechend der in Kapitel 3.2.1.2 diskutierten CFA-MTMM Logik systematische Messfehler zu quantifizieren, welche dann als FRE identifiziert werden können, wenn die Gruppierung auf eine experimentelle Variation der Fragereihenfolge gemäß Kapitel 3.2.1.1 zurückgeht. In Abbildung 13 ist diesbezüglich ein beispielhaftes Multigruppenmodell entsprechend der CFA-MTMM Logik mit latentem Merkmal $\xi_1 = \xi_2 = \xi$, der Gruppierungsvariable z, acht Indikatoren x_j $(j = 1, \dots, 8)$, wobei x_1 bis x_4 zur Schätzung von ξ_1 in Gruppe 0 und x_5 bis x_8 zur Schätzung von ξ_2 in Gruppe 1 verwendet werden und mit durch Kovarianzen spezifizierten und zwischen den Gruppen frei geschätzten FRE zwischen allen aufeinanderfolgenden Messfehlern der Items (δ_{21}, δ_{32} und δ_{43} für die Effekte in Gruppe 0, δ_{65}, δ_{76} und δ_{87} für die Effekte in Gruppe 1) dargestellt. Formel 43 zeigt das zugehörige Gleichungssystem nach X-Notation.

Abbildung 13: Einfaches MG-CFA Modell mit Kovarianzen auf Messfehlerebene



$$(1)\quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ \lambda_{41} & 0 \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \\ 0 & \lambda_{72} \\ 0 & \lambda_{82} \end{bmatrix} * \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \end{bmatrix}$$

$$(2)\quad \begin{bmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \\ \lambda_{41} \end{bmatrix} = \begin{bmatrix} \lambda_{52} \\ \lambda_{62} \\ \lambda_{72} \\ \lambda_{82} \end{bmatrix}$$

Formel 43

$$(3)\quad \boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\phi_{11}} & \\ 0 & \boldsymbol{\phi_{22}} \end{bmatrix}$$

$$(4)\quad \boldsymbol{\phi_{11}} = \boldsymbol{\phi_{22}}$$

$$(5)\quad \boldsymbol{\Theta}_{\delta} = \begin{bmatrix} \theta_{\delta_1} & & & & & & & \\ \theta_{\delta_{21}} & \theta_{\delta_2} & & & & & & \\ 0 & \theta_{\delta_{32}} & \theta_{\delta_3} & & & & & \\ 0 & 0 & \theta_{\delta_{43}} & \theta_{\delta_4} & & & & \\ 0 & 0 & 0 & 0 & \theta_{\delta_5} & & & \\ 0 & 0 & 0 & 0 & \theta_{\delta_{65}} & \theta_{\delta_6} & & \\ 0 & 0 & 0 & 0 & 0 & \theta_{\delta_{76}} & \theta_{\delta_7} & \\ 0 & 0 & 0 & 0 & 0 & 0 & \theta_{\delta_{87}} & \theta_{\delta_8} \end{bmatrix}$$

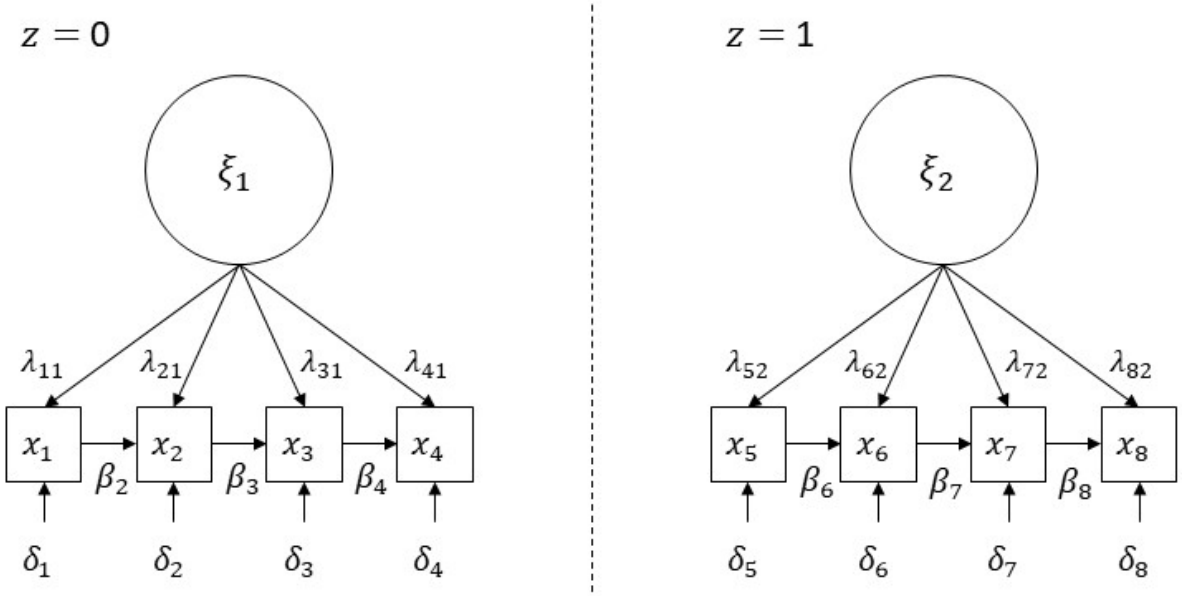
Bei der Spezifikation solcher Multigruppenmodelle mittels Gleichungssystemen ist zu beachten, dass keine Kovarianzen/Faktorladungen/Effekte zwischen Konstrukten und/oder Indikatoren modelliert werden können, welche sich nicht in der gleichen Gruppe befinden, da Parameter zwischen verschiedenen Gruppen aufgrund der hier fehlenden Daten nicht geschätzt werden können. Entsprechend muss die Kovarianz ϕ_{21} zwischen den latenten Konstrukten ξ_2 und ξ_1 auf 0 fixiert werden (vgl. Formel 43.3). Ebenso ist es nicht möglich, eine Kovarianz zwischen den Indikatoren x_5 und x_4 zu modellieren (vgl. Formel 43.5.). Je nach gewünschtem Invarianz-

Niveau der Multigruppenanalyse müssen zuletzt zu schätzende Parameter zwischen den Gruppen gleichgesetzt werden. Bei gewünschter metrischer Invarianz handelt es sich bei diesen Parametern um die Faktorladungen des Messmodells (vgl. Formel 43.2). Zudem sei erneut darauf hingewiesen, dass die Wahl von vier Indikatorvariablen lediglich dem Beispiel dient. Die konkrete Mindestzahl an Indikatorvariablen eines zu schätzenden Modells ist durch die Anzahl an Freiheitsgraden des Modells bedingt, während zusätzliche Indikatoren theoretisch grundsätzlich möglich sind, da eine zunehmende Anzahl an Indikatoren lediglich die Komplexität des Beispiels erhöht.

MG-CFA Modell mit regressiven Beziehungen auf Indikatorebene

Auf Basis der Diskussion bezüglich der Verwendung regressiver Beziehungen auf Indikatorebene statt von Kovarianzen auf Messfehlerebene für die Quantifizierung der FRE aus Kapitel 3.2.1.3 lässt sich das zuvor beschriebene Modell entsprechend Abbildung 14 und Formel 44 erneut beispielhaft mit vier Indikatorvariablen je Gruppe adaptieren:

Abbildung 14: Einfaches Multigruppenmodell mit regressiver Beziehung auf Indikatorebene



Das Gleichungssystem in Formel 44 wurde, mit dem Gleichungssystem in Formel 43 vergleichend, um die q x q (mit q als Anzahl der Indikatoren innerhalb der zu schätzenden latenten Konstrukte) diagonale Effektmatrix **B** der regressiven Effekte β_j ergänzt, wobei $\mathbf{B}_{11}$ sowie $\mathbf{B}_{55}$ entsprechend den Wert 0 enthält, da das erste Item x_1 sowie x_5 der zu schätzenden Merkmale

ξ_1 und ξ_2 über keine vorangegangenen Kontextfragen verfügen und entsprechend durch keinen Fragereihenfolgeneffekt beeinflusst werden (vgl. Formel 44.1). Darüber hinaus enthält die Residual-Kovarianzmatrix $\boldsymbol{\Theta}_{\delta}$ nun lediglich noch die zufälligen Messfehler in der Diagonalen und nicht mehr die Messfehler-Kovarianzen (vgl. Formel 44.5), da die systematischen Messfehler in **B** geschätzt und damit quantifiziert werden.

$$(1)\quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ \lambda_{41} & 0 \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \\ 0 & \lambda_{72} \\ 0 & \lambda_{82} \end{bmatrix} * \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \beta_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \beta_6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \beta_7 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \beta_8 \end{bmatrix} * \begin{bmatrix} 0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \end{bmatrix}$$

$$(2)\quad \begin{bmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \\ \lambda_{41} \end{bmatrix} = \begin{bmatrix} \lambda_{52} \\ \lambda_{62} \\ \lambda_{72} \\ \lambda_{82} \end{bmatrix}$$

Formel 44

$$(3)\quad \boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\phi_{11}} & \\ 0 & \boldsymbol{\phi_{22}} \end{bmatrix}$$

$$(4)\quad \boldsymbol{\phi_{11}} = \boldsymbol{\phi_{22}}$$

$$(5)\quad \boldsymbol{\Theta}_{\delta} = \begin{bmatrix} \theta_{\delta_1} & & & & & & & \\ 0 & \theta_{\delta_2} & & & & & & \\ 0 & 0 & \theta_{\delta_3} & & & & & \\ 0 & 0 & 0 & \theta_{\delta_4} & & & & \\ 0 & 0 & 0 & 0 & \theta_{\delta_5} & & & \\ 0 & 0 & 0 & 0 & 0 & \theta_{\delta_6} & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \theta_{\delta_7} & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_{\delta_8} \end{bmatrix}$$

Entsprechend der zuvor diskutierten Logik von FRE sollten diese für die zuvor beschriebenen Beispiele lediglich unter der Bedingung einer fixierten Fragereihenfolge auftreten, wie sie hier

für Gruppe 1 (Experimentalgruppe) gegeben ist, während unter der Bedingung einer randomisierten Fragereihenfolge in Gruppe 0 (Kontrollgruppe) keine FRE messbar sein sollten. Als „hartes" Kriterium für den Nachweis von FRE gilt entsprechend, dass in Gruppe 0 die modellierten Kovarianzen (δ_{21}, δ_{32} und δ_{43}) bzw. regressiven Beziehungen (β_2, β_3 und β_4) einen Wert von 0 annehmen, was keinen Effekt impliziert, während in Gruppe 1 die modellierten Kovarianzen (δ_{65}, δ_{76} und δ_{87}) bzw. regressiven Beziehungen (β_6, β_7 und β_8) ungleich 0 sein sollten, da diese die in der Experimentalbedingung vorhandenen FRE quantifizieren.

Da ein solches Ergebnis allerdings lediglich dann beobachtbar ist, wenn systematische Messfehler zwischen zwei aufeinanderfolgenden Items ausschließlich auf die Variation der Fragereihenfolge (fixierte Reihenfolge vs. randomisierte Reihenfolge) zurückgeführt werden können und es sich bei der Messung nicht um ein statistisches Artefakt handelt, kann als weiteres, „weiches" Kriterium für den Nachweis von FRE formuliert werden, dass die Kovarianzen zweier aufeinanderfolgender Items zwischen den Gruppen zumindest different sind, d.h.: $\delta_{21} \neq \delta_{65};\ \delta_{32} \neq \delta_{76};\ \delta_{43} \neq \delta_{87}$; bzw. dass die regressiven Beziehungen zweiter aufeinanderfolgender Items zwischen den Gruppen different sind, d.h.: $\beta_2 \neq \beta_6;\ \beta_3 \neq \beta_7;\ \beta_4 \neq \beta_8$. Ist dies der Fall, so ist gezeigt, dass systematische Messfehler zwischen zwei aufeinanderfolgenden Items kausal auf die Variation der Fragereihenfolge zurückgeführt werden können.

Identifikation von FRE mittels MG-CFA

Entsprechend dieser Logik können FRE nicht nur auf die Variation der Fragereihenfolge, sondern ebenfalls auf die Variation weiterer Moderatorvariablen zurückgeführt werden. Hierzu müssen, wie zuvor beschrieben, lediglich weitere, zu analysierende Gruppen spezifiziert werden. Diesbezüglich kann als Beispiel die in Kapitel 2 als Einflussfaktor diskutierte Antwortelaboration herangezogen werden, welche, entsprechend ihrer Pole, in eine Gruppe „geringe Elaboration" sowie eine Gruppe „hohe Elaboration" dichotomisiert werden kann. Werden diese beiden Gruppen mit den beiden Gruppen der Experimentalkondition (fixierte vs. randomisierte Reihenfolge) gekreuzt, so ergeben sich hieraus vier zu analysierende Gruppen (2 Ausprägungen auf der Moderatorvariable mal 2 Ausprägungen auf der Gruppierungsvaribale) entsprechend der folgenden Klassifikation: 1) Befragte der Experimentalbedingung mit geringer Antwortelaboration; 2) Befragte der Experimentalbedingung mit hoher Antwortelaboration; 3) Befragte der Kontrollbedingung mit geringer Antwortelaboration; 4) Befragte der Kontrollbedingung mit hoher Antwortelaboration (vgl. Tabelle 1). Die zugehörigen Modelle und Gleichungssysteme lassen sich dabei nach Abbildung 13 und Formel 43 für Modelle unter Verwendung von Kovarianzen auf Messfehlerebene und nach Abbildung 14 und Formel 44 für Modelle unter

Verwendung regressiver Beziehungen auf Indikatorebene adaptieren. Zu beachten bleibt, dass solche Analysen nicht auf die Modellierung von vier Gruppen auf Basis zweier dichotomer Variablen beschränkt sind, sondern theoretisch beliebig viele Gruppen modelliert werden können. Die Anzahl der Gruppen ist hierbei im Wesentlichen durch die mindestens benötigte Fallzahl innerhalb der Gruppen, sowie durch die Zunahme der Komplexität der Analyse bei einer zunehmenden Anzahl an Gruppen limitiert.

Tabelle 1: Effektinterpretation bei Multigruppenanalysen

	Experimentalbedingung (Fixiert)	**Kontrollbedingung (Randomisiert)**	**Differenz der Kovarianzen/Effekte**
Geringe Elaboration	Gruppe 1	Gruppe 3	Fragereihenfolgeneffekt bei geringer Elaboration
Hohe Elaboration	Gruppe 2	Gruppe 4	Fragereihenfolgeneffekt bei hoher Elaboration
Differenz der Kovarianzen/Effekte	Einfluss der Antwortelaboration auf Fragereihenfolgeneffekt	Einfluss der Antwortelaboration auf sonstige systematische Messfehler	

Gemäß der zuvor beschriebenen „weichen“ Kriterien ist ein FRE bei einer solchen 4-Gruppen-Analyse ebenfalls identifiziert, wenn sich die Kovarianzen bzw. Regressionseffekte zwischen Experimental- und Kontrollbedingung (Vergleich von Gruppe 1 mit Gruppe 3 und von Gruppe 2 mit Gruppe 4) systematisch unterscheiden. Ein Einfluss der Moderatorvariable, in diesem Fall der Antwortelaboration, auf den FRE gilt in diesem Sinne als identifiziert, wenn sich die Kovarianzen bzw. Regressionseffekte zwischen der Gruppe mit geringer Antwortelaboration und der Gruppe mit hoher Antwortelaboration innerhalb der Experimentalbedingung systematisch unterscheiden (Vergleich von Gruppe 1 mit Gruppe 2). Interpretativ quantifiziert die Differenz der Kovarianzen bzw. Regressionseffekte zwischen Gruppe 1 und Gruppe 3 entsprechend den merkmalsbereinigten FRE bei geringer Antwortelaboration, während die Differenz der Kovarianzen bzw. Regressionseffekte zwischen Gruppe 2 und Gruppe 4 den merkmalsbereinigten FRE bei hoher Antwortelaboration quantifiziert. Die Differenz der Kovarianzen bzw. Regressionseffekte zwischen Gruppe 1 und 2 zuletzt indiziert den Einfluss der Antwortelaboration auf den FRE in dem Sinne, als dass die Kovarianz bzw. der Regressionseffekt in Gruppe 1 den Extremwert des FRE bei durchschnittlich geringer Elaboration und die Kovarianz bzw. der Regressionseffekt in Gruppe 2 den Extremwert des FRE bei durchschnittlich hoher Elaboration

quantifiziert. Die Differenz der beiden Kovarianzen bzw. Effekte gibt entsprechend Auskunft über die Veränderung des FRE, wenn von Gruppe 1 „geringe Elaboration“ in Gruppe 2 „hohe Elaboration“ gewechselt wird.

Zusammenfassend wurden somit aufgezeigt, wie mithilfe der Multigruppenanalyse innerhalb von CFA Modellen FRE sowie deren Determinanten auf Basis von sowohl Kovarianzen auf Messfehlerebene, als auch von Regressionseffekten auf Indikatorebene identifiziert werden können. Die hier vorgestellten Multigruppenmodelle, innerhalb welcher die Experimental- und Kontrollgruppe mit weiteren Gruppierungsvariablen verschränkt werden, welche als Moderator dienen (z.B. die dichotomisierte Antwortelaboration) und welche den FRE als Kovarianz auf Messfehlerebene quantifizieren, entsprechen dabei im Wesentlichen bereits dem CFA-MTMM Modell, welches im analytischen Teil dieser Arbeit untersucht und der NLC-CFA gegenübergestellt wird. Entsprechend erfolgt in Kapitel 3.3.1 eine abschließende Konkretisierung des entsprechend zu verwendenden CFA-MTMM Modell. Weiterhin wurde aufgezeigt, dass es sich bei MG-CFA um ein weit verbreitetes Verfahren zur Identifikation systematischer Messfehler handelt, welches jedoch auch mit einer Reihe von Limitationen einhergeht. So kann die Extremgruppenbildung beispielsweise erstens die Dichotomisierung oder Merkmalsreduktion von Variablen bedingen, was mit einem Informationsverlust einhergeht. Weiterhin erhöht zweitens jeder weitere, zu berücksichtigende Moderator bzw. jede weitere, zu berücksichtigende Merkmalsausprägung die Anzahl zu modellierender Gruppen, was die Komplexität der Analyse mitunter deutlich erhöht. Drittens erschwert eine erhöhte Anzahl an Gruppen das Erreichen metrischer Invarianz. Und viertens gilt zu beachten, dass sich mit jeder weiteren Gruppe die Fallzahlen innerhalb der modellierten Gruppen theoretisch reduzieren, worauf diverse Analyseverfahren sensibel reagieren.

Insofern erscheint die Verwendung der Multigruppenanalyse zur Identifikation von FRE sowie deren Determinanten vor allem bei einer hohen Fallzahl bzw. einer geringen Anzahl an zu vergleichenden Gruppen als sinnvoll. Zwecks Einführung der NLC-CFA, welche die Verwendung mehrerer, kontinuierlicher und latenter Moderatoren ermöglichen soll, wird im Folgenden entsprechend die Interaktionsanalyse als weitere Möglichkeit für die Moderatoranalyse zur Identifikation und Kontrolle von FRE diskutiert, wobei in Kapitel 3.3.2 eine Verschränkung der beiden Verfahren als abschließender Baustein des NLC-CFA Modells zur Identifikation und Kontrolle von FRE vorgeschlagen wird.

3.2.2.2 Interaktionsanalyse mit manifesten Moderatoren

Im Vergleich zu Multigruppenanalysen besteht der wesentliche Vorteil der Interaktionsanalyse darin, dass Stärke, Richtung und Signifikanz der Moderatoreffekte direkt auf Basis der Interaktionseffekte identifiziert werden können, und nicht erst auf Basis der Gruppendifferenzen berechnet werden müssen (Urban & Mayerl, 2018). Zudem ist die Analyse nicht auf die Kategorien von (dichotomisierten) Moderatorvariablen beschränkt sondern kann direkt auf kontinuierliche Moderatoren angewandt werden, was ebenfalls bedeutet, dass Effekte auch für Fälle geschätzt werden können, welche sich zwischen den Extremgruppen befinden, sowie, dass kein Vergleich zwischen Gruppen stattfindet, welcher mit einer Reduktion der Fallzahl, einer Erhöhung der Analysekomplexität und der zunehmenden Schwierigkeit des Erreichens metrischer Invarianz einhergeht.

Basale Logik der Identifikation von FRE mittels Interaktionsanalyse

Gemäß Hayes (2013) bedeutet dies bezüglich der Analyse von FRE, dass dieser als Haupteffekt b_1 (vgl. Formel 45) einer Regression der Zielfrage *y* auf die Kontextfrage *x* dargestellt ist. Mittels Interaktionsanalyse wird nun die Abhängigkeit dieses Effektes von z.B. der Zugänglichkeit von Informationen *w* untersucht, welche in diesem Beispiel als Moderator dient. Hierzu wird der Interaktionsterm „*x*w*“ gebildet, welcher in die Regressionsfunktion mit eingeht (vgl. Interaktionsterm „*xw*“ in Formel 45.2 gegenüber Formel 45.1). Der Interaktionseffekt b_3 gibt in der Folge die Veränderung des Haupteffektes b_1 (FRE) bei Variation der Zugänglichkeit *w* an. Weiterhin repräsentiert i_1 in der hier nach Hayes (ebd.) zitierten Schreibweise die Konstante der Regression und e_Y das Residuum. Eine vollständige Interpretation einer Interaktion gemäß Formel 45.2 lautet entsprechend, dass i_1 als Konstante den Wert von *y* angibt, wenn *x* und *w* den Wert 0 annehmen; b_1 repräsentiert den Effekt von *x* auf *y*, wenn *w* den Wert 0 annimmt; b_2 repräsentiert den Effekt von *w* auf *y*, wenn *x* den Wert 0 annimmt; b_3 zuletzt repräsentiert die Veränderung in b_1, wenn *w* um eine empirische Einheit steigt (bzw. die Veränderung in b_2, wenn *x* um eine empirische Einheit steigt).

Zu beachten ist, dass zwecks exakter Identifikation von FRE auch das Vorgehen mittels Interaktionsanalyse im besten Fall eine experimentelle Variation der Fragereihenfolge bedingt, so dass, auch wenn Interaktionsterme die Basis der Analyse bilden, ein Gruppenvergleich zwischen Versuchsgruppe (fixierte Fragereihenfolge) und Kontrollgruppe (randomisierte Fragereihenfolge) stattfindet. Insofern handelt es sich bei dem hier diskutierten Vorgehen zur Identifikation und Kontrolle von FRE im eigentlichen Sinne um eine Interaktionsanalyse, welche in-

nerhalb einer Multigruppenanalyse eingebettet ist. Der Vorteil gegenüber der reinen Multigruppenanalyse besteht insofern darin, dass mittels Interaktionsanalyse lediglich die beiden Gruppen gebildet und verglichen werden, welche auf die Zuordnung zu den beiden experimentellen Bedingungen zurückgehen. Eine Extremgruppenbildung für weitere Moderatorvariablen sowie die exponentielle Erhöhung der zu vergleichenden Gruppen je weiterer Moderatorvariable ist für die hier beschriebene Form der Interaktionsanalyse entsprechend nicht notwendig, da die Interaktionsterme der interessierenden Moderatorvariablen potentiell innerhalb des gleichen Modells geschätzt werden können. Es ist somit ersichtlich, dass sich die Anzahl von simultan geschätzten Moderatoreffekten nicht in einer exponentiellen Zunahme der zu vergleichenden Gruppen, sondern lediglich in der Zunahme von zu modellierenden Interaktionstermen äußert.

$$(1)\quad y = i_1 + b_1 x + b_2 w + e_Y$$
$$(2)\quad y = i_1 + b_1 x + b_2 w + b_3 xw + e_Y$$

Formel 45

$$(1)\quad \beta_x^{Z=1} \neq \beta_x^{Z=0}$$
$$(2)\quad |\beta_{xm}^{Z=1}| > 0$$

Formel 46

Es kann somit bereits, erneut gemäß Formel 42, festgehalten werden, dass ein FRE erstens dadurch identifiziert ist, wenn der Haupteffekt der Kontextfrage *x* auf die Zielfrage *y* für fixierte Fragereihenfolgen (*Z=0*) vom jeweiligen Haupteffekt für randomisierte Fragereihenfolgen (*Z=1*) abweicht (vgl. Formel 46.1). Zweitens kann festgehalten werden, dass das Vorhandensein eines Interaktionseffektes (vgl. Formel 46.2) indiziert, dass die Ausprägung des Haupteffektes (sprich: der FRE) in Abhängigkeit steht zur Ausprägung einer spezifischen Moderatorvariable *w*. In anderen Worten: Bei Gültigkeit der Bedingung in Formel 46.1 gilt der FRE als identifiziert, bei Gültigkeit der Bedingung in Formel 46.2 variiert dieser systematisch bei Variation der Moderatorvariable. Der Reihenfolgeneffekt ist somit sowohl identifiziert, als auch kontrolliert.

Simultane Berücksichtigung mehrerer Moderatoren innerhalb des gleichen Modells

Gemäß der theoretischen Diskussion von FRE in Kapitel 2 wurden für diesen drei Einflussfaktoren diskutiert, welche zusätzlich unter bestimmten Bedingungen einen interaktiven Einfluss

aufweisen. Während die Antwortelaboration als notwendige Bedingung diskutiert wurde, welche zusätzlich qualitativ-dichotom in Gruppen geringer und hoher Elaboration unterteilbar ist, werden die Zugänglichkeit von Informationen gegenüber dem Objekt der Befragung, sowie das Framing der Befragtenrolle als Prädiktoren entlang eines quantitativen Kontinuums betrachtet. Entsprechend bedarf es eines Verfahrens, mehrere Moderatorvariablen im Optimalfall simultan zu analysieren. Da eine solche simultane Analyse für Interaktionsanalysen nicht wie bei der Multigruppenanalyse mit zusätzlich zu konstruierenden Gruppen einhergeht, sondern, wie bereits beschrieben, die Modellierung zusätzlicher Interaktionsterme bedingt, wird in einer Verschränkung der beiden Verfahren eine große Chance für die Identifikation und Kontrolle von FRE gesehen. So können erstens auf Basis der dichotomen bzw. dichotomisierten Variablen „experimentale Bedingung“ und „Antwortelaboration“ vier Gruppen definiert werden (2 x 2), da es theoriegeleitet genügt, die Elaboration qualitativ zu behandeln, ein 4-Gruppen-Design relativ überschaubar hinsichtlich der Analysekomplexität ist und die Anzahl an zu modellierenden Interaktionstermen relativ gering bleibt (max. vier statt bis zu elf Interaktionsterme bei zwei statt drei simultan geschätzten Moderatoren). In einem zweiten Schritt werden dann die benötigten Interaktionsterme für die Variablen „Zugänglichkeit“ und „Framing“ mit der Kontextfrage modelliert und jeweils innerhalb der vier Gruppen analysiert (vgl. Kapitel 3.3 für eine detaillierte Beschreibung des Vorgehens).

Zu beachten bleibt, dass auch innerhalb einer simultanen Analyse aller Moderatorvariablen verschiedene Möglichkeiten bestehen, Interaktionsterme abzubilden (Hayes, 2013). So können, beispielhaft für zwei simultan geschätzte Moderatoren, 1.) lediglich die Interaktionsterme der Moderatoren mit der unabhängigen Variable (vgl. Formel 47.1) oder 2.) auch die Interaktionsterme der Moderatoren miteinander (vgl. Formel 47.2) konstruiert werden, was als moderierte Moderation zu verstehen ist. Die Benennung der Parameter folgt Formel 45, wobei die Variable *m* die zweite Moderatorvariable repräsentiert:

$$(1)\quad y = i_1 + b_1x + b_2m + b_3w + b_4xm + b_5xw + e_Y$$

$$(2)\quad y = i_1 + b_1x + b_2m + b_3w + b_4xm + b_5xw + b_6mw + b_7xmw + e_Y$$

Formel 47

Werden lediglich die Interaktionsterme der Moderatoren mit der unabhängigen Variable gebildet (vgl. Formel 47.1), so gibt b_1 den Effekt der Kontext- auf die Zielfrage, sprich den FRE an, wenn die Moderatoren *m* und *w* gleich 0 sind. In b_2 findet sich der konditionale Effekt von *m*

für *x* gleich 0 bei konstantem *w*, sowie, äquivalent, in b_3 der konditionale Effekt von *w* für *x* gleich 0 bei konstantem *m*. b_4 schließlich quantifiziert die konditionale Veränderung des Effektes von *x* auf *y* (b_1), wenn *m* sich um eine empirische Einheit verändert und *w* konstant bleibt, während b_5, äquivalent, die konditionale Veränderung des Effektes von *y* auf *x* (b_1) quantifiziert, wenn *w* sich um eine empirische Einheit verändert und *m* konstant bleibt (Hayes, 2013).

Somit wäre potenziell die Möglichkeit gegeben, mehrere Moderatoren simultan und konditional zu analysieren. Unberücksichtigt bleibt bei dieser Modellierung jedoch, dass auch die beiden Moderatoren *m* und *w* miteinander in Interaktion stehen können, weshalb, zwecks Vollständigkeit der Analyse, Formel 47.2 zur Modellierung der Interaktionseffekte zu präferieren ist, da hierdurch auch potenzielle, interaktive Abhängigkeiten zwischen den Moderatoren, also moderierte Moderationen, quantifiziert werden können. Gemäß Formel 47.2 entspricht b_1 somit dem Effekt von *x* auf *y* für *m* und *w* gleich 0. b_2 gibt den Effekt von *m* auf *y* für *w* und *x* gleich 0 an, während b_3 den Effekt von *w* auf *y* für *m* und *x* gleich 0 quantifiziert. In b_4 zeigt sich die Veränderung in b_1, wenn *m* sich um eine empirische Einheit verändert und *w* gleich 0 ist, während b_5 äquivalent die Veränderung in b_1 angibt, wenn *w* sich um eine empirische Einheit verändert und *m* gleich 0 ist. b_6 steht für die Veränderung in *y*, wenn sich das Produkt von *m* und *w* um eine empirische Einheit verändert und *x* gleich 0 ist, wodurch die Interaktion zwischen den beiden Moderatorvariablen modelliert wird (moderierte Moderation). b_7 zuletzt zeigt auf, wie sich b_1 verändert, wenn sich das Produkt von *m* und *w* um eine empirische Einheit verändert (Hayes, 2013).

Interaktionsanalyse auf Indikatorebene von CFA Modellen

Da FRE darüber hinaus innerhalb von CFA modelliert werden sollen stellt sich zusätzlich die Frage, wie Interaktionstherme innerhalb dieser Logik zu spezifizieren sind, um die im Kontext der MTMM-Analyse (vgl. Kapitel 3.1.3) diskutierten faktorbereingten Methodeneffekte zu quantifizieren. Diesbezüglich kann auf Basis der MTMM-Logik spezifiziert werden, dass die Messung eines Items y_j in Abhängigkeit steht von einem latent zugrundeliegenden Merkmal η_k, welches auf Basis von y_j geschätzt werden soll, dem Einfluss der Methode, sowie einem zufälligen Messfehler ε_j. Der Einfluss der Methode ist in diesem Sinne definiert als FRE. Dies bedeutet, dass sich der Methodeneffekt zerlegen lässt in den FRE, also den direkten Effekt der Kontextfrage y_j auf die Zielfrage y_{j+1}, sowie die zugehörigen Interaktionseffekte der spezifizierten Moderatorvariablen x_1 (Zugänglichkeit von Informationen gegenüber dem Objekt der Befragung) und x_2 (Framing der Befragtenrolle). Die Einbettung dieser Spezifikation innerhalb

der zuvor beschriebenen Multigruppenanalyse wird im Detail in Kapitel 3.3 diskutiert. Ebenfalls wird zum gegenwärtigen Zeitpunkt der Betrachtung die Verwendung manifester Moderatorvariablen angenommen. Eine Diskussion bezüglich der Verwendung latenter Moderatorvariablen erfolgt in Kapitel 3.2.2.3. Für Item y_{j+1}, das im Folgenden als Zielfrage verstanden wird und welches auf die Kontextfrage y_j folgt, lässt sich somit eine der folgenden Gleichungen in Anlehnung an Formel 47 definieren, je nachdem, ob die moderierte Moderation zwischen x_1 und x_2 spezifiziert wird (vgl. Formel 48.2) oder diese nicht spezifiziert wird (vgl. Formel 48.1).

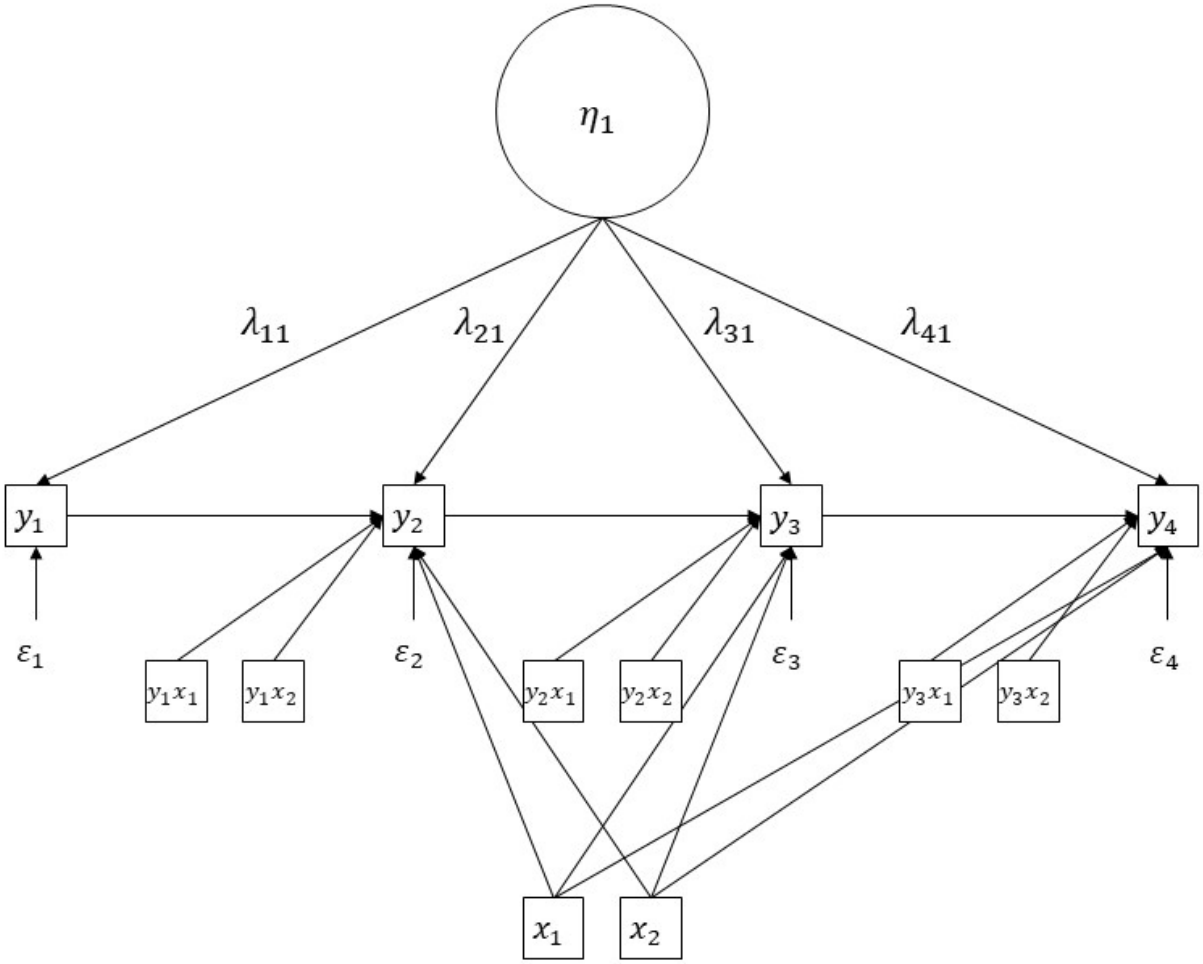
$$(1)\quad y_{j+1} = \lambda_{j+1,k}\eta_k + b_1 y_j + b_2 x_1 + b_3 x_2 + b_4 y_j x_1 + b_5 y_j x_2 + \varepsilon_{j+1}$$

$$(2)\quad y_{j+1} = \lambda_{j+1,k}\eta_k + b_1 y_j + b_2 x_1 + b_3 x_2 + b_4 y_j x_1 + b_5 y_j x_2 + b_6 x_1 x_2 + b_7 y_j x_1 x_2 + \varepsilon_{j+1}$$

Formel 48

Bei Betrachtung der Gleichung in Formel 48 wird ersichtlich, dass für diese eine große Zahl an Parametern (eine Faktorladung und fünf bzw. sieben Regressionskoeffizienten) zu schätzen sind. Zu beachten ist dabei, dass die Anzahl modellierbarer Parameter stets limitiert ist durch die Anzahl der zur Verfügung stehenden Informationen der empirischen Varianz-Kovarianzmatrix $\boldsymbol{S}$ (vgl. Kapitel 3.1.2). Folglich kann es bei der Modellierung nötig sein, dass Interaktionseffekte entweder nicht zwischen jeder Kontext-Zielfragen-Dyade spezifiziert werden oder alternativ, dass die Interaktionseffekte zwischen allen Dyaden auf den gleichen Wert fixiert werden (vgl. Kapitel 4). Zwecks Reduktion der Komplexität wird in der folgenden Abbildung das entsprechende Interaktionsmodell ohne moderierte Moderation (entsprechend Formel 48.1), ohne mehrere Gruppen und ohne die Abbildung von Regressionskoeffizienten, dafür mit Interaktionen zwischen jeder Dyade aufgezeigt. Das zugehörige Gleichungssystem (inklusive der Gleichsetzung der Regressionskoeffizienten) ist in Formel 49 dargestellt.

Abbildung 15: CFA Modell mit Interaktionen auf Indikatorebene



$$(1)\quad \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \\ \lambda_{41} \end{bmatrix} * \eta_1 + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \beta_1 & 0 & 0 & \beta_2 & \beta_3 \\ 0 & \beta_1 & 0 & \beta_2 & \beta_3 \\ 0 & 0 & \beta_1 & \beta_2 & \beta_3 \end{bmatrix} * \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ x_1 \\ x_2 \end{bmatrix}$$

$$+ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \beta_4 & \beta_5 & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_4 & \beta_5 & 0 & 0 \\ 0 & 0 & 0 & 0 & \beta_4 & \beta_5 \end{bmatrix} * \begin{bmatrix} y_1x_1 \\ y_1x_2 \\ y_2x_1 \\ y_2x_2 \\ y_3x_1 \\ y_3x_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

Formel 49

$$(2)\quad \boldsymbol{\Theta}_\varepsilon = \begin{bmatrix} \theta_{\varepsilon_{11}} & & & \\ 0 & \theta_{\varepsilon_{22}} & & \\ 0 & 0 & \theta_{\varepsilon_{33}} & \\ 0 & 0 & 0 & \theta_{\varepsilon_{44}} \end{bmatrix}$$

Entsprechend ist bei einer solchen Modellierung ein FRE grundsätzlich durch β_1 (vgl. Formel 48.1) spezifiziert, da dieser den direkten, faktorbereinigten Effekt zwischen Kontext- und Zielfrage zwischen jeder der möglichen Kontext-Zielfragen-Dyaden darstellt. Zu beachten ist, dass dies aufgrund der Gleichsetzung der Effekte der Fall ist – werden diese nicht gleichgesetzt, so ergibt sich der FRE aus den separaten Regressionskoeffizienten der entsprechenden Haupteffekte zwischen den Kontext- und Zielfragen. Darüber hinaus gilt es zu beachten, dass der Effekt β_1, je nach gewähltem Nullpunkt von Multiplikator und Multiplikand der Interaktion ebenfalls den Wert 0 annehmen kann, selbst wenn ein FRE vorliegt (Hayes, 2013). Entsprechend ist bei der Spezifikation solcher Modelle zu beachten, mit welchem Nullpunkt Variablen kodiert sind. Gemäß der hier gewählten Spezifikation und nach Formel 48.1 entsprechen die Effekte β_2 und β_3 den Haupteffekte der Moderatorvariablen auf die Zielfrage, wodurch diese bei Analyse von FRE inhaltlich unbedeutsam sind, da hierfür lediglich der Haupteffekt β_1 sowie die Interaktionseffekte β_4 und β_5 relevant sind. Nach der zuvor aufgezeigten Interpretationshilfe von Interaktionseffekten quantifiziert β_4 entsprechend die Veränderung des FRE, wenn sich x_1 (Zugänglichkeit von Informationen gegenüber dem Objekt der Befragung) um eine empirische Einheit verändert und x_2 (Framing der Befragtenrolle) konstant ist, während β_5 die Veränderung des FRE quantifiziert, wenn sich x_2 um eine empirische Einheit verändert und x_1 konstant bleibt.

Es ist somit diskutiert, dass eine Kombination von Multigruppen- und Interaktionsanalyse mit nichtlinearen Effekten zur Kontrolle der Messfehler von Indikatoren durch manifeste, exogene Variablen innerhalb von CFA Modellen die Möglichkeit bietet, FRE innerhalb von Fragebatterien aufeinanderfolgender Items zu quantifizieren und zu kontrollieren. Die Berücksichtigung mehrerer Gruppen innerhalb eines Multigruppenmodells ist dabei insofern notwendig, um den FRE zum einen durch die Variation der Experimentalbedingung (fixierte vs. randomisierte Fragereihenfolge) eindeutig zu identifizieren, und zum anderen um die Wirkung weiterer binärer Moderatoren (wie etwa der dichotomisierten Antwortelaboration) zu analysieren, so dass die Anzahl notwendig zu spezifizierender und damit zu analysierender Interaktionsterme nicht zu groß wird, was insgesamt die Komplexität der Auswertung verringert. Dies soll nun noch um die Verwendung latenter, exogener Moderatorvariablen ergänzt werden, wodurch die nichtlinear latent kontrollierte CFA (NLC-CFA) vollständig eingeführt ist.

3.2.2.3 Latente Interaktionsanalyse

Eine Besonderheit, welche sich durch die Identifikation und Kontrolle von FRE mittels Interaktionsanalyse auf Indikatorebene von CFA Modellen auf Basis der in Kapitel 2 diskutierten

theoretischen Modelle ergibt, ist in der Verwendung latenter Konstrukte zur Bildung der Interaktionsterme begründet. Dies ist insofern relevant, als dass es sich bei dem Framing der Befragtenrolle (mit dem Pol „indifferent“ und dem Pol „unterstützend“) sowie der Zugänglichkeit von Informationen gegenüber dem Objekt der Befragung (mit dem Pol „nicht zugänglich“ und dem Pol „hoch zugänglich“) nicht nur um quantitative Kontinuen handelt, sondern diese ebenfalls als latente Konstrukte gebildet werden (Mayerl, 2009; Stocké, 2006).

Hinsichtlich der Interaktionen auf Basis latenter Konstrukte kann grundsätzlich festgehalten werden, dass in Strukturgleichungen endogene (latente) Variablen in der Regel als lineare Funktion exogener (latenter) Variablen dargestellt werden. Bei Vorhandensein einer (latenten) Interaktion wird der lineare Effekt innerhalb der Strukturgleichung nun um einen Interaktionseffekt ergänzt, wobei die Interaktion als Produkt exogener Variablen in das Gleichungssystem eingeht. Dies führt im Allgemeinen dazu, dass latente Interaktionsmodelle aufgrund der Berücksichtigung von Produkten exogener Variablen nichtlineare strukturelle Beziehungen beinhalten, was mit weitreichenden Konsequenzen für die Verteilung von Variablen einhergeht (Jöreskog & Yang, 1996). So weicht beispielsweise die multivariate Verteilung der Indikatoren der endogenen Variable substanziell von der Normalverteilung ab, je größer die Kovarianzen der latenten Variablen der Strukturgleichung sind, welche in den nichtlinearen Produkttermen berücksichtigt werden (Moosbrugger, Schermelleh-Engel, & Klein, 1997), was insgesamt zu Schätzproblemen des Strukturgleichungsmodells führt.

Zur Verwendung latenter Interaktionen wurden bereits eine Reihe von Verfahren diskutiert, darunter 1.) der Constrained Approach (Kenny & Judd, 1984), 2.) der Unconstrained Approach (Marsh, Wen, & Hau, 2004), 3.) der Residual Centering Approach (Little, Bovaird, & Widaman, On the Merits of Orthogonalizing Powered and Product Terms: Implications for Modeling Interactions Among Latent Variables, 2006), 4.) der 2-Stage Approach (Ping, 1996), sowie 5.) der LMS(„**L**atent **M**oderated **S**EM“)-Approach (Klein & Moosbrugger, 2000). Die Verfahren 1) bis 4) haben diesbezüglich gemein, dass sie normalverteilte Indikatorvariablen, unabhängige Messfehler und eindimensionale latente Variablen bedingen sowie, dass latente Interaktionen händisch auf Indikatorebene und/oder in mehreren, händisch durchzuführenden Schritten zu bilden sind (Moosbrugger et al., 1997). Für den Constrained Approach (1) bedeutet dies, dass wenn beispielsweise die latente Variable ξ_1 mit den Indikatoren x_1 und x_2 und die latente Variable ξ_2 mit den Indikatoren x_3 und x_4 gebildet wird, die Interaktion zwischen ξ_1 und ξ_2 nicht als Produkt der beiden latenten Variablen berechnet, sondern als eigenständige latente Interaktionsvariable ξ_3 auf Basis aller möglicher Produkte der Indikatoren x_1 und x_2 mit x_3 und x_4 geschätzt werden (für diesen Fall wird ξ_3 entsprechend mit x_1x_3, x_1x_4, x_2x_3 und

x_2x_4 geschätzt). Hinzu kommt die nichtlineare Fixierung der Fehlervarianz der an der Interaktion beteiligten Indikatoren sowie die Fixierung der Mittelwerte der latenten Variablen (für eine exakte Beschreibung der Fixierungen, siehe Kenny & Judd, 1984). Innerhalb des Unconstrained Approach (2) werden Interaktionen ebenfalls auf Indikatorebene gebildet und auf Basis dieser geschätzt. Gegenüber dem Constrained Approach werden hingegen lediglich die Indikatoren um ihren Mittelwert zentriert, die Mittelwerte der latenten Haupteffekte auf 0 fixiert und die Mittelwerte der latenten Produktvariablen mit der Kovarianz der beiden Haupteffekt-Variablen gleichgesetzt. Der Residual Centering Approach (3) hingegen folgt dem Ansatz, dass als Indikatoren einer latenten Interaktionsvariable Residuen verwendet werden, welche aus einer Regression jeder Interaktiosvariable auf alle Indikatoren gebildet werden (für eine detaillierte Beschreibung, siehe Little et al., 2006). Innerhalb des 2-Stage Approach (4) wird das Gesamtmodell zunächst in einem ersten Schritt ohne Interaktionen geschätzt, um Faktorladungen, Kovarianzen und Fehlervarianzen zu schätzen. In einem zweiten Schritt wird das Modell dann inklusive latenter Interaktion geschätzt, wobei Faktorladungen und Fehlervarianzen der latenten Interaktionsvariable auf Basis der zuvor geschätzten Parameter fixiert werden (für eine detaillierte Beschreibung, siehe Ping, 1996). Es zeigt sich somit für die vier bisher beschriebenen Verfahre neben den einleitend formulierten Bedingungen (normalverteilte Indikatorvariablen, unabhängige Messfehler und eindimensionale latente Variablen), dass diese ebenfalls komplex in der Anwendung sind, wobei der Unconstrained-Approach die geringste Anzahl an Schritten und Spezifikationen bedarf.

Formel 50

$$(1)\quad \eta = \alpha + \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1\xi_2 + \zeta$$

$$(2)\quad \eta = \alpha + \mathbf{\Gamma}\xi + \xi'\mathbf{\Omega}\xi + \zeta$$

$$(3)\quad \mathbf{\Omega} = \begin{bmatrix} 0 & \omega_{1,2} & \cdots & \omega_{1,n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \omega_{n-1,n} \\ 0 & \cdots & \cdots & 0 \end{bmatrix}$$

$$(4)\quad \eta = \alpha + [\gamma_1 \ \gamma_2]\begin{bmatrix}\xi_1 \\ \xi_2\end{bmatrix} + [\xi_1 \ \xi_2]\begin{bmatrix}0 & \omega_{12} \\ 0 & 0\end{bmatrix}\begin{bmatrix}\xi_1 \\ \xi_2\end{bmatrix} + \zeta$$

Um dieses Komplexitätsproblem zu lösen, wurde LMS als weiteres Verfahren vorgeschlagen (5). Dieses hat den Vorteil, dass Interaktionsindikatoren nicht per Hand berechnet oder sonstig modelliert werden müssen, stattdessen werden Interaktionsparameter auf Basis der verfügbaren Verteilungsinformationen aller Indikatoren geschätzt. Das Verfahren basiert somit auf der Analyse der multivariaten Verteilung des gemeinsamen Indikator-Vektors und berücksichtigt dabei die durch die latenten Interaktionseffekte implizierte Nicht-Normalität. Darüber hinaus hat sich LMS bereits in einer Reihe von Simulationsstudien als effizient erwiesen (vgl. z.B. Barendse, Oort, & Garst, 2010; Kelava et al., 2011; Klein & Muthén, 2008; Maslowsky, Jager, & Hemken, 2015).

Die Schätzung der Parameter mittels LMS basiert grundsätzlich auf dem generellen Interaktionsmodell für latente Variablenmodelle (vgl. Formel 50.1), welches von Kenny und Judd (1984) vorgeschlagen und von Klein und Moosbrugger (2000) um eine quadratische Form der latenten Variablenvektoren ξ ergänzt wurde, um die simultane Spezifikation verschiedener Interaktionen zu ermöglichen (vgl. Formel 50.2), wobei Parameter qua Berechnung von Pseudovariablen geschätzt werden. η Entspricht hierbei der (1 x 1) latenten endogenen Variable, α ist die (1 x 1) Konstante, ξ ist der (n x 1) Vektor der latenten exogenen Variablen (mit n als Gesamtzahl aller latenter, exogener Variablen), $\boldsymbol{\Gamma}$ ist eine (1 x n) Koeffizient-Matrix des Einflusses von ξ auf η (analog zum Regressionskoeffizienten b, welcher innerhalb der „regulären“ Regressionsanalyse mit manifesten Variablen den Einfluss von x auf y spezifiziert) und ζ ist eine (1 x 1) Messfehlervariable mit $E(\zeta) = 0$ und $Cov(\zeta, \xi') = 0$. Für die quadratische Ergänzung des Terms wurde schließlich $\boldsymbol{\Omega}$ eingeführt, welches einer (n x n) Koeffizientenmatrix entspricht, die den Einfluss ω_{ij} von $\xi_i \xi_j$ (mit $i < j$) auf η spezifiziert (mit $\omega_{ij} = 0$ für $i \geq j$) und als obere dreieckige Matrix mit 0 in der Diagonale beschrieben werden kann (vgl. Formel 50.3). Dies resultiert für ein beispielhaftes, latentes Interaktionsmodell mit zwei exogenen Variablen ξ_1 und ξ_2 in der Strukturgleichung in Formel 50.4.

Weiterhin werden (unter Berücksichtigung der X- und Y-Notation, vgl. Kapitel 3.1.2) innerhalb des LMS-Verfahrens bezüglich der Indikatoren und Residuen die folgenden Annahmen getroffen:

1. $\boldsymbol{x}$ (als manifester Indikatorvektor von ξ) ist multivariat normalverteilt;
2. δ, ε sind multivariat normalverteilt und untereinander unabhängig mit $E(\delta) = 0$, $E(\varepsilon) = 0$, $Cov(\delta, \xi') = 0$, $Cov(\varepsilon, \xi') = 0$;
3. ζ ist normalverteilt mit $E(\zeta) = 0$, $Cov(\zeta, \xi') = 0$, $Cov(\zeta, \delta') = 0$, $Cov(\zeta, \varepsilon') = 0$.

4. Über die Verteilung von $\boldsymbol{y}$ wird darüber hinaus, im Gegensatz zu den anderen Verfahren, keine Annahme getroffen, da die Annahme einer multivariaten Normalverteilung von $\boldsymbol{y}$ aufgrund der quadratischen Form $\xi' \boldsymbol{\Omega} \xi$ innerhalb des Schätzterms von η nicht erfüllt ist.

Das LMS-Verfahren berücksichtigt somit explizit die Verteilungscharakteristik des nicht normalverteilten Indikatorvektors $\boldsymbol{y}$ innerhalb latenter Variablenmodelle und führt somit gemäß entsprechender Simulationsstudien zu unverzerrten und effizienteren Parameterschätzungen als die alternativen Verfahren (Klein & Moosbrugger, 2000).

Nun ist LMS zum gegenwärtigen Zeitpunkt zwar immer häufiger in gängiger Statistiksoftware wie z.B. MPlus implementiert (Vandenberg, 2006), für den Fall, dass es das nicht ist, wird jedoch ebenfalls eine effektive und effiziente Lösung für die Schätzung latenter Interaktionen innerhalb der einzuführenden NLC-CFA benötigt. Diesbezüglich merken Steinmetz, Davidov und Schmidt (2011) an, dass dem Constrained Approach aufgrund seiner vielen benötigten Modellspezifikationen, welche relativ komplex in der Anwendung sein können, eine höhere Chance für Modell-Fehlspezifikationen inhärent ist. Ähnlich bergen auch der Residual Centering Approach und der 2-Stage Approach aufgrund ihrer mehrstufigen Anwendung ein erhöhtes Risiko für Schätzfehler unter anwendungsorientierten Nutzern. Der Unconstrained Approach hingegen weißt den Vorteil auf, dass er dadurch, dass fast alle Constraints aufgehoben wurden, für anwendungsorientierte Nutzer einfach anzuwenden ist sowie daraus folgend, dass keine Constraints auf Basis multivariater Normalverteilungsannahmen der latenten Variablen abgeleitet werden, welche für nichtlineare Modelle generell ungültig sind (Marsh et al., 2004). Darüber hinaus zeigen Simulationsstudien, dass der Unconstrained Approach hinsichtlich des Typ 1 Fehlers und verzerrter Parameterschätzungen mit traditionellen Ansätzen vergleichbar ist (Marsh et al., 2004; Marsh, Wen, & Hau, 2006), dass der Ansatz algebraisch integrierbar ist (Marsh, et al., 2007) sowie, dass Schätzer weniger verzerrt sind und dass Modellschätzungen, auch bei kleiner Fallzahl, eher konvergieren als innerhalb des Constrained Approaches (Marsh et al., 2004). Für nicht-normalverteilte Indikatoren der exogenen latenten Variablen zeigen Simulationen, welche das LMS-Verfahren, den Constrained Approach und den Unconstrained Approach vergleichen, dass Parameterschätzung bei LMS zwischen allen Verfahren unter der Bedingung geringer nicht-normalität am effizientesten sind. Hingegen führt sowohl LMS, als auch der Constrained Approach bei hoher nicht-normalität zu verzerrten Schätzern, während der Unconstrained Approach mit Maximum-Likelihood-Schätzer unverzerrte latente Interaktionseffekte bei einem akzeptablen Typ 1 Fehler schätzt (Steinmetz et al., 2011). Auch eine weitere Simulationsstudie, welche LMS und den Unconstrained Approach für normalverteilte Daten, jedoch für eine hohe und ungleiche Anzahl an Produktindikatoren vergleicht kommt zu

dem Ergebnis, dass beide Lösungen zu unverzerrten Parameterschätzungen und, im Fall des Unconstrained Approach bei erhöhter Fallzahl, zu unverzerrten Standardfehlern sowie akzeptablen Typ 1 Fehlern führen (Jackman, Leite, & Cochrane, 2011). Aufgrund dieser Vergleichbarkeit, der gegebenen Effizienz sowie der Praxistauglichkeit wird bei gegenwärtigem Stand der Forschung entsprechend der Unconstrained Approach als Alternative zu LMS vorgeschlagen, wo LMS nicht in dem verwendeten Statistik-Paket implementiert ist. Entsprechend wird im folgenden analytischen Teil dieser Arbeit der Unconstrained Approach zur Schätzung der latenten Interaktionen innerhalb der NLC-CFA verwendet, da die Studie aufgrund ihrer Komplexität innerhalb der Statistiksoftware „R“ (R Core Team, 2021) auf Basis des Packages „lavaan“ (Rosseel, et al., 2022) programmiert werden musste und LMS zum aktuellen Zeitpunkt nicht in lavaan implementiert ist.

Eine generelle Besonderheit, welche zugleich eine der größten Hindernisse bei der effizienten Schätzung nichtlinearer Modellparameter darstellt, besteht in der inhärenten Nicht-Normalverteilung der nichtlinearen Terme (Moosbrugger, Schermelleh-Engel, Kelava, A., & Klein, 2009), was bei LMS beispielsweise unter Berücksichtigung der Normalverteilungsannahme für *x*-Indikatoren bei Verletzung zu Konvergenzproblemen oder verzerrten Parameterschätzungen führen kann (Cham, West, Ma, & Aiken, 2012; Klein & Moosbrugger, 2000; Maslowsky et al., 2015). Um dem zu entgehen, wird innerhalb von LMS die Heteroskedastizität der Indikatoren utilisiert, während bei Verwendung andere Verfahren wie dem Unconstrained Approach die Robustheit gegenüber Verstößen der Normalverteilungsannahme untersucht und berücksichtigt werden sollte (Moosbrugger et al., 2009).

Zuletzt ist bei der Schätzung latenter, nichtlinearer Modelle das Problem der Multikollinearität innerhalb der Produkt-Variablen zu beachten, da eine erhöhte Korrelation der Produkt-Indikatoren zwar zu einer höheren Reliabilität des Interaktionsterms führt, gleichzeitig mit zunehmender Korrelation aber auch Schätzprobleme auftreten: So sind etwa Standardfehler umso höher, je stärker die Multikollinearität zwischen den Produktvariablen ist, was die Power der gesamten Analyse verringert. Um diesem Problem zu begegnen wird die Mittelwertzentrierung der Produktindikatoren vorgeschlagen (Moosbrugger et al., 2009), wobei Mayerl und Urban darauf hinweisen, dass Multikollinearität hierdurch faktisch nicht eliminiert wird und es sich hierbei lediglich um eine scheinbare Verringerung durch eine angepasste Messung handelt (Mayerl & Urban, 2019). Bei der Schätzung von Modellen ist es somit empfehlenswert, die Korrelationen zwischen den latenten Produktvariablen zu analysieren und gegebenenfalls, falls diese zu hoch ausfällt, auf die Modellierung latenter Interaktionen zu verzichten und eine Moderatoranalyse

stattdessen beispielweise auf Basis einer Multigruppenanalyse (vgl. Kapitel 3.2.2.1) durchzuführen.

Somit ist ebenfalls diskutiert, wie latente Variablen als Faktoren innerhalb von Interaktionsmodellen berücksichtigt werden können. Das in Kapitel 3.2.2.2 abgeleitete Modell kann entsprechend statt mit Interaktionen auf Basis manifester, exogener Variablen ebenfalls auf Basis latenter Interaktionen, entweder unter Verwendung von LMS oder alternativ des Unconstrained Approach geschätzt werden (wobei im analytischen Teil dieser Arbeit wie bereits zuvor beschrieben der Unconstrained Approach verwendet wird), wodurch die Bausteine der NLC-CFA vollständig eingeführt sind. Im Einzelnen handelt es sich bei der NLC-CFA somit um ein CFA Modell, innerhalb dessen die manifesten Indikatoren eines zu schätzenden Merkmals nichtlinear auf exogene, latente Konstrukte (welche mit dem Auftreten systematischer Messfehler assoziiert sind) zurückgeführt werden, um den Varianzanteil der manifesten Indikatoren zu kontrollieren, welcher durch eben jene systematischen Messfehler bedingt ist, so dass die Schätzung des eigentlichen latenten Merkmals von Interesse via der geschätzten Faktorladungen nicht weiter (oder in geringerem Maße) durch den Anteil systematischer Fehlervarianz verzerrt wird. Eine vertiefende Diskussion der NLC-CFA sowie die Ableitung eines konkreten NLC-CFA Modells zur Kontrolle von FRE erfolgt in Kapitel 3.3.2.

3.3 Modelle zur Identifikation und Kontrolle von Fragereihenfolgeneffekten

In Kapitel 2 wurde diskutiert, dass die Antwortelaboration, das Framing der Befragtenrolle, sowie die kognitiv chronische Zugänglichkeit von Informationen gegenüber dem Objekt der Befragung mögliche Determinanten für FRE darstellen. Um diese Abhängigkeit mittels Moderatoranalyse zu testen, existiert eine Vielzahl an Möglichkeiten, FRE statistisch zu quantifizieren und zu kontrollieren, wie in Kapitel 3.2 gezeigt. Als besonders vielversprechend wurde die Möglichkeit diskutiert, FRE innerhalb von CFA Modellen zu quantifizieren, da durch die Extraktion desjenigen Varianzanteils, welcher auf ein gemeinsames zugrundeliegendes Merkmal von Kontext- und Zielfrage zurückgeht (der sog. Common Variance), eine Schätzung des systematischen Messfehlers (hier: des FRE) möglich ist, welche um den Einfluss des zugrundeliegenden Merkmals bereinigt ist. Der merkmalsbereinigte systematische Messfehler kann dann entsprechend der Logik von CFA-MTMM Modellen als Kovarianz auf der Ebene der Messfehler oder entsprechend der Logik der eingeführten NLC-CFA als regressive Beziehung auf der

Ebene der Indikatoren eines latenten Konstruktes spezifiziert werden (vgl. Kapitel 3.2). Für die Identifikation von FRE empfiehlt es sich dabei für beide Verfahren, die jeweiligen Modelle innerhalb einer Multigruppenanalyse mit metrischer Invarianz zu testen, wobei Experimentalgruppen (mit fixierter oder randomisierter Fragereihenfolge) zur Gruppierung verwendet werden (vgl. Kapitel 3.2.2.1). Für beide Verfahren ist es weiterhin möglich, das Experimentalgruppendesign um weitere Gruppen zu erweitern, so der FRE auf Basis einer Multigruppen-Moderatoranalyse auf den Einfluss weiterer, diskreter Variablen zurückgeführt werden soll. Innerhalb der NLC-CFA wird der FRE zusätzlich mittels (latenter) Interaktionsanalyse auf weitere, kontinuierliche latente Variablen zurückgeführt. Die Rückführung des FRE auf weitere Variablen entspricht dabei in beiden Modellen der Kontrolle des FRE.

In Anlehnung an die Herleitung eines möglichen CFA-MTMM Modells zur Identifikation und Kontrolle von FRE in Kapitel 3.2.2.1 wird im Folgenden zum einen das konkrete CFA-MTMM Modell vorgestellt, welches für die Identifikation und Kontrolle von FRE vorgeschlagen wird und zum anderen Diskutiert, wie das entsprechende Modell spezifiziert werden kann und im folgenden analytischen Teil dieser Arbeit spezifiziert wird (vgl. Kapitel 3.3.1). Darüber hinaus erfolgt eine vertiefende Diskussion und Zusammenfassung der einzuführenden NLC-CFA, welche nicht nur der Identifikation und Kontrolle von FRE dienen, sondern allgemein als Framework für die Bereinigung von Indikatoren eines latenten Konstruktes in CFA Modellen betrachtet werden kann. Auf Basis dieser zusammenfassenden Diskussion wird schließlich das konkrete NLC-CFA Modell vorgestellt, mit welchem FRE identifiziert und kontrolliert werden können und ebenfalls aufgezeigt, wie das entsprechende Modell spezifiziert werden kann und im analytischen Teil dieser Arbeit spezifiziert wird (vgl. Kapitel 3.3.2). Da die beiden Modellvarianten im analytischen Teil dieser Arbeit nicht isoliert evaluiert, sondern vergleichend gegenübergestellt werden, erfolgt zudem eine vergleichende Diskussion der beiden Modelle bevor zuletzt aufgezeigt wird, wie die beiden Varianten genutzt werden können, um die in Kapitel 2.3.3 theoretisch abgeleiteten Hypothesen über und die Bedingungen für das Auftreten von FRE zu testen (vgl. Kapitel 3.3.3).

3.3.1 CFA-MTMM Modell

Hinsichtlich eines konkreten CFA-MTMM Modells zur Identifikation und Kontrolle von FRE wurde bereits in Kapitel 3.2 (und hier insbesondere in Kapitel 3.2.2.1) ein CFA Modell vorgeschlagen, innerhalb welchem Kovarianzen auf Messfehlerebene zwischen den Items aufeinanderfolgender Kontext- und Zielfragen spezifiziert sind. Darüber hinaus wurde diskutiert, ein

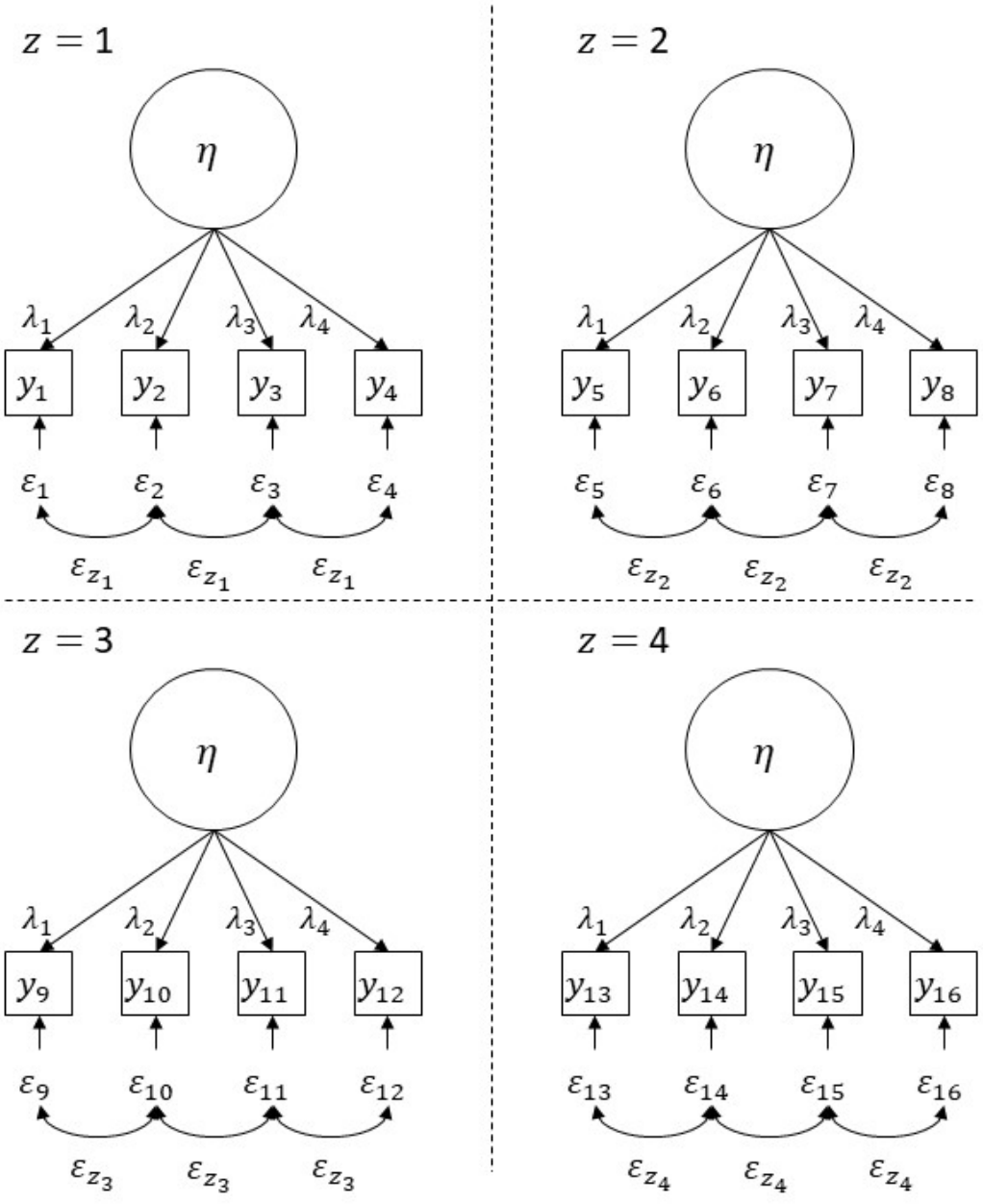
solches Modell innerhalb eines Multigruppendesigns mit metrischer Invarianz einzubetten, wobei die Gruppen entsprechend einem experimentellen Design definiert sind mit randomisierter Fragereihenfolge aller Indikatoren des zu schätzenden latenten Merkmals von Interesse für die Kontrollgruppe und einer fixierten Fragereihenfolge aller Indikatoren für die Versuchsgruppe. In einem solchen Modell quantifiziert die Kovarianz auf Messfehlerebene den merkmalsbereinigten systematischen Messfehler zwischen Kontext- und Zielfrage, welcher auf eine experimentelle Variation der Fragereihenfolge zurückgeht und somit als FRE angenommen werden kann. In diesem Sinne sind FRE dann identifiziert, wenn die Messfehlerkovarianz zwischen den beiden Gruppen verschieden ist („weiches Kriterium“) bzw. wenn in der Versuchsgruppe eine Kovarianz ungleich 0 und in der Kontrollgruppe eine Kovarianz gleich 0 identifiziert wird („hartes Kriterium“). Das entsprechende Modell wurde in Abbildung 13 aufgezeigt und wird durch das Gleichungssystem in Formel 43 definiert.

Weiterhin wurde diskutiert, dass der FRE auf Basis solcher MG-CFA-MTMM Modelle auf weitere Einflussfaktoren zurückgeführt werden kann, indem zusätzlich zur Versuchs- und Kontrollgruppe weitere Gruppen innerhalb des Designs berücksichtigt werden. Als Variable mit Einfluss auf den FRE wurde hier der Grad an Antwortelaboration diskutiert, welcher sich theoriegeleitet in eine geringe Elaboration und eine erhöhte Elaboration dichotomisieren lässt (vgl. Kapitel 2), woraus sich insgesamt ein 4-Gruppen Design ergibt: 1) Befragte der Experimentalbedingung mit geringer Antwortelaboration; 2) Befragte der Experimentalbedingung mit hoher Antwortelaboration; 3) Befragte der Kontrollbedingung mit geringer Antwortelaboration; 4) Befragte der Kontrollbedingung mit hoher Antwortelaboration.

Zwar wurden darüber hinaus ebenfalls die Zugänglichkeit von Informationen über das Objekt der Befragung sowie das Framing der Befragtenrolle als weitere Einflussfaktoren für den FRE diskutiert – deren Berücksichtigung innerhalb des zuvor beschriebenen CFA-MTMM Modells wird jedoch durch eine Reihe von Gründen limitiert. Zum einen handelt es sich bei den beiden genannten Einflussfaktoren um kontinuierliche Schätzungen auf Basis latenter Konstrukte. Solche Konstrukte können zwar in einem mehrstufigen Verfahren ebenfalls in diskrete Kategorien unterteilt bzw. dichotomisiert werden, jedoch geht dies mit einem Informationsverlust der entsprechenden Daten einher. Wird dies in Kauf genommen und wird beispielsweise die Informationszugänglichkeit dennoch dichotomisiert, so wird aus dem oben beschriebenen 4-Gruppen Design bereits ein 8-Gruppen Design, da die genannten vier Gruppen zusätzlich mit den beiden Gruppen der Informationszugänglichkeit (geringe Zugänglichkeit vs. hohe Zugänglichkeit) gekreuzt werden. Eine weitere Berücksichtigung des Framings der Befragtenrolle als dichotomi-

sierte Messung würde entsprechend in einem 16-Gruppen Design resultieren. Zu berücksichtigen ist dahingehend, dass eine erhöhte Anzahl an Gruppen das Erreichen metrischer Invarianz erschwert und dass sich mit jeder weiteren Gruppe die Fallzahlen innerhalb der modellierten Gruppen theoretisch reduzieren, wodurch die Power des Verfahrens abnimmt. Die Analyse einer derart hohen Anzahl an Gruppen ist somit zwar theoretisch möglich, in der Praxis jedoch äußerst komplex und fehleranfällig, so dass davon abgeraten wird (vgl. Kapitel 3.2.2.1).

Abbildung 16: MG-CFA-MTMM Modell zur Analyse von FRE



$$(1)\quad \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ y_{10} \\ y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{15} \\ y_{16} \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} * \eta + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \\ \varepsilon_{10} \\ \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{15} \\ \varepsilon_{16} \end{bmatrix}$$

Formel 51

$$(2)\quad \psi = 1$$

(3) $\boldsymbol{\Theta_\varepsilon}$

$$= \begin{bmatrix} \theta_{\varepsilon_1} & & & & & & & & & & & & & & & \\ \theta_{\varepsilon_{z_1}} & \theta_{\varepsilon_2} & & & & & & & & & & & & & & \\ 0 & \theta_{\varepsilon_{z_1}} & \theta_{\varepsilon_3} & & & & & & & & & & & & & \\ 0 & 0 & \theta_{\varepsilon_{z_1}} & \theta_{\varepsilon_4} & & & & & & & & & & & & \\ 0 & 0 & 0 & 0 & \theta_{\varepsilon_5} & & & & & & & & & & & \\ 0 & 0 & 0 & 0 & \theta_{\varepsilon_{z_2}} & \theta_{\varepsilon_6} & & & & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_{z_2}} & \theta_{\varepsilon_7} & & & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_{z_2}} & \theta_{\varepsilon_8} & & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_9} & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_{z_3}} & \theta_{\varepsilon_{10}} & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_{z_3}} & \theta_{\varepsilon_{11}} & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_{z_3}} & \theta_{\varepsilon_{12}} & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_{13}} & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_{z_4}} & \theta_{\varepsilon_{14}} & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_{z_4}} & \theta_{\varepsilon_{15}} & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_{z_4}} & \theta_{\varepsilon_{16}} \end{bmatrix}$$

Als sinnvoller Kompromiss zwischen Modellkomplexität, Fehleranfälligkeit und Informationsgewinn wird daher argumentiert, das CFA-MTMM Modell zur Identifikation und Kontrolle von FRE innerhalb des weiter oben beschriebenen 4-Gruppen Designs zu schätzen, wobei sich die Antwortelaboration als theoriegeleitet dichotome Variable als Moderator anbietet. Soll auf Basis des CFA-MTMM Modells ebenfalls der Einfluss der Informationszugänglichkeit oder

des Framings der Befragtenrolle auf den FRE analysiert werden so empfiehlt es sich entsprechend, dies in eigenständigen Analysen, ebenfalls innerhalb eines 4-Gruppen Designs unter Verwendung dichotomisierter Variablen zu untersuchen. Die sich ergebenden vier Gruppen bei dichotomisierter Informationszugänglichkeit lauten dann theoretisch: 1) Befragte der Experimentalbedingung mit geringer Informationszugänglichkeit; 2) Befragte der Experimentalbedingung mit hoher Informationszugänglichkeit; 3) Befragte der Kontrollbedingung mit geringer Informationszugänglichkeit; 4) Befragte der Kontrollbedingung mit hoher Informationszugänglichkeit. Analog lauten die sich ergebenden vier Gruppen bei dichotomisiertem Framing der Befragtenrolle: 1) Befragte der Experimentalbedingung mit indifferenter Befragtenrolle; 2) Befragte der Experimentalbedingung mit kooperativer Befragtenrolle; 3) Befragte der Kontrollbedingung mit indifferenter Befragtenrolle; 4) Befragte der Kontrollbedingung mit kooperativer Befragtenrolle. Die Kovarianzen der einzelnen Gruppen können dann, je nach verwendetem Moderator, analog zu Tabelle 1 interpretiert werden.

Wird ein solches CFA-MTMM Modell zur Identifikation und Kontrolle innerhalb eines 4-Gruppen Designs wie zuvor beschrieben spezifiziert, so ist konkret darauf zu achten, metrische Invarianz durch eine Gleichsetzung der Faktorladung zwischen den Gruppen zu etablieren. Die Kovarianzen auf Messfehlerebene wiederum werden zwischen den Gruppen frei geschätzt, da diese den unterschiedlichen Einfluss der zur Gruppierung verwendeten Variablen auf den FRE quantifizieren. Um eine Unteridentifikation des Modells zu vermeiden, können die Kovarianzen weiterhin innerhalb der Gruppen als gleich geschätzt werden so begründet angenommen werden kann, dass der FRE zwischen allen Kontext-Zielfragen Dyaden gleich wirkt. Ein entsprechendes Modell mit vier Gruppen z, einem latenten Merkmal von Interesse η, vier Indikatorvariablen y je Gruppe, mit zwischen den Gruppen als gleich geschätzten Faktorladungen λ und mit zwischen den Gruppen als verschieden und innerhalb der Gruppen als gleich geschätzten Kovarianzen ε_z ist in der folgenden Abbildung 16 aufgezeigt und durch das Gleichungssystem in Formel 51 definiert.

Bei dem hier beschriebenen CFA-MTMM Modell handelt es sich abschließend um dasjenige Modell, welches im analytischen Teil dieser Arbeit hinsichtlich seiner Effizienz und Präzision bei der Quantifizierung von FRE und entsprechend der Bereinigung von durch FRE verursachte Verzerrungen untersucht und dem im nachfolgenden Kapitel vorgestellten NLC-CFA Modell gegenübergestellt wird.

3.3.2 Nichtlinear latent kontrollierte CFA (NLC-CFA)

In Kapitel 3.2 wurden umfassend die Möglichkeiten diskutiert, bei der Schätzung latenter Merkmale die zur Schätzung verwendeten Indikatoren um Verzerrungen durch systematische Messfehler zu bereinigen. Die Bereinigung der Indikatoren erfolgt dabei auf Basis einer Kontrolle durch weitere exogene Variablen, welche systematische Messfehler bedingen, indem die potenziell von Verzerrungen betroffenen Indikatoren durch regressive Beziehungen auf die zusätzlichen exogenen Variablen zurückgeführt werden. Exogene Variablen, welche systematische Messfehler bedingen, werden theoriegeleitet ausgewählt und können beispielsweise der Grad an Antwortelaboration, die Zugänglichkeit von Informationen über das Objekt der Befragung oder das Framing der Befragtenrolle sein (vgl. Kapitel 2). Durch diese zusätzliche Spezifikation von Regressionen auf Indikatorebene erfolgt eine Dekomposition der Varianz der Indikatoren in denjenigen Varianzanteil, welcher durch systematische Messfehler bedingt ist, denjenigen Varianzanteil, welcher durch das zu schätzende Merkmal von Interesse bedingt ist, sowie den zufälligen Varianzanteil. Erfolgt diese Dekomposition bei vorhandenen systematischen Messfehlern nicht, so kann eine verzerrte Schätzung des Varianzanteils resultieren, der durch das zu schätzende Merkmal von Interesse bedingt ist, wenn mehrere Indikatoren durch den gleichen systematischen Messfehler beeinflusst werden. Verzerrte Schätzungen der Faktorladungen und der Varianz des zu schätzenden Merkmals von Interesse (so diese nicht fixiert wurde) sind die Folge (vgl. Kapitel 3.1.4).

Neben der Kontrolle der Indikatoren durch lineare Regressionen wurde ebenfalls die Möglichkeit aufgezeigt, nichtlineare Beziehungen in Form von Interaktionen bei der Bereinigung der systematischen Messfehler zu verwenden (vgl. Kapitel 3.2.2.2), wobei auch die Möglichkeit zur Spezifikation von Interaktionen mit latenten, exogenen Variablen gegeben ist (vgl. Kapitel 3.2.2.3). Werden also innerhalb eines CFA Modells die Indikatoren des zu schätzenden latenten Konstruktes zusätzlich mittels nichtlinearer Beziehungen auf latente exogene Variablen zurückgeführt, so wird eine „**n**onlinear **l**atend **c**ontrolled CFA“ (kurz: NLC-CFA) geschätzt. Die NLC-CFA kann somit allgemein als Framework zur Bereinigung von Indikatoren eines latenten Konstruktes betrachtet werden, welche durch systematische Messfehler verzerrt sind (vgl. Kapitel 5.3).

Neben dieser Darstellung als allgemeines Framework wurde die NLC-CFA insbesondere hergeleitet, um Schätzungen eines latenten Merkmals bzw. der Faktorladungen um Verzerrungen durch FRE zu kontrollieren und den FRE dabei auf den Grad an Antwortelaboration, die Zu-

gänglichkeit von Informationen über das Objekt der Befragung und das Framing der Befragtenrolle zurückzuführen. Entsprechend folgt die konkrete Beschreibung des entsprechenden NLC-CFA Modells.

Für eine der drei Moderatorvariablen, die Antwortelaboration, wurde in diesem Sinne bereits diskutiert, dass für diese eine qualitative Differenzierung in „geringe Elaboration“ und „hohe Elaboration“ hinreichend ist, wodurch die Antwortelaboration als dichotome Moderatorvariable verstanden werden kann, welche gemäß Kapitel 3.2.2.1 gut zur Gruppenbildung innerhalb von Multigruppenanalysen geeignet ist. Für das Framing der Befragtenrolle sowie die Zugänglichkeit der Informationen gegenüber dem Objekt der Befragung wurde hingegen diskutiert, dass es sich hierbei um quantitative Kontinuen handelt (vgl. Kapitel 2; Mayerl, 2009; Stocké, 2006), wodurch diese besser dazu geeignet sind, innerhalb von Interaktionstermen als Moderator verwendet zu werden (vgl. Kapitel 3.2.2.2). Wird darüber hinaus berücksichtigt, dass FRE durch experimentelle Designs mit einer fixierten Fragereihenfolge in der Versuchsgruppe und einer randomisierten Fragereihenfolge in der Kontrollgruppe identifizierbar sind, so ergibt sich aus der Verschränkung dieses Designs mit der dichotomisierten Antwortelaboration ein 4-Gruppen-Design, innerhalb welchem eine NLC-CFA zur Identifikation und Kontrolle von FRE geschätzt werden kann (zwei Experimentalkonditionen, Versuchsgruppe mit fixierter Fragereihenfolge und Kontrollgruppe mit randomisierter Fragereihenfolge, mal zwei Elaborationskonditionen, geringe Elaboration und hohe Elaboration). Entsprechend sei z die Gruppe, wobei die Gruppen wie folgt definiert sind: 1.) Befragte mit fixierter Fragereihenfolge und einer geringen Antwortelaboration; 2.) Befragte mit fixierter Fragereihenfolge und einer hohen Antwortelaboration; 3.) Befragte mit randomisierter Fragereihenfolge und einer geringen Antwortelaboration; und 4.) Befragte mit randomisierter Fragereihenfolge und einer hohen Antwortelaboration. Relevant ist dabei zu betonen, dass bei der Analyse metrische Invarianz zwischen den Gruppen etabliert werden muss (vgl. Kapitel 3.2.2.1).

Innerhalb der Gruppen gilt es nun, die entsprechenden Interaktionsterme für die Zugänglichkeit von Informationen ξ_1 (geschätzt als latentes Merkmal) und das Framing der Befragtenrolle ξ_2 (ebenfalls geschätzt als latentes Merkmal) zu berechnen, welche den FRE zwischen Kontextfrage y_j und der Zielfrage y_{j+1} (mit $j = 1, \dots, p$ und p als Anzahl derjenigen Indikatorvariablen innerhalb der Fragebatterie bzw. des Konstruktes, innerhalb dem FRE zu erwarten sind) moderieren. Diesbezüglich wurde in Kapitel 2 diskutiert, dass die Wirkung des Moderatoreffektes sowohl für die Zugänglichkeit ξ_1, als auch das Framing der Befragtenrolle ξ_2 auf den FRE verschieden sind, je nachdem, ob ein Assimilationseffekt zu erwarten ist, welcher notwendig durch einen geringen Grad an Antwortelaboration bedingt ist, oder ob ein Kontrasteffekt zu

erwarten ist, welcher notwendig durch einen hohen Grad an Antwortelaboration bedingt ist. So wurde postuliert, dass die Moderatorvariablen bei geringer Elaboration separat additiv wirken, wodurch es für diesen Fall ausreicht, jeweils einen Interaktionsterm je Moderatorvariable zu bilden. Bei hoher Elaboration wurde hingegen postuliert, dass die Moderatorvariablen nicht separat, sondern ausschließlich in Interaktion miteinander wirken, wodurch für diesen Fall moderierte Moderatorvariablen zu bilden sind. Entsprechend werden, um Komplexität zu reduzieren und die Analysen zwischen den Gruppen vergleichbar zu halten, für alle vier Gruppen moderierte Moderatorvariablen gebildet, um beide Fälle, die erwartet werden können (moderierte Moderation liegt nicht vor vs. moderierte Moderation liegt vor), abbilden und empirisch überprüfen zu können. Sollen ξ_1 und ξ_2 demnach den Effekt von y_j auf y_{j+1} mittels Interaktion moderieren und hierbei ebenfalls moderierte Moderationen berücksichtigen, so sind die folgenden unabhängigen Variablen und Interaktionsterme mit Einfluss auf die Zielfrage y_{j+1} zu berücksichtigen: y_j, ξ_1, ξ_2, $y_j\xi_1$, $y_j\xi_2$, $\xi_1\xi_2$ und $y_j\xi_1\xi_2$.

Für jede beliebige Zielfrage y_{j+1}, welche mittels latenter Variablenmodellierung um den Einfluss des zu messenden Merkmals, abgebildet durch die latente Variable η (entsprechend Y-Notation) bereinigt wurde, gilt mit β als Effektkoeffizient entsprechend in allen vier Gruppen:

$$y_{j+1} = \lambda_j\eta + \beta_1 y_j + \beta_2\xi_1 + \beta_3\xi_2 + \beta_4 y_j\xi_1 + \beta_5 y_j\xi_2 + \beta_6\xi_1\xi_2 + \beta_7 y_j\xi_1\xi_2 + \varepsilon_{j+1} \quad \textbf{Formel 52}$$

wobei die einzelnen Effekte, adaptiert nach Hayes (2013) entsprechend Tabelle 2 zu interpretieren sind. Relevant bei der Darstellung und Interpretation der Effekte ist diesbezüglich, dass zwar alle möglichen Interaktionen spezifiziert werden müssen, um zu überprüfen, ob moderierte Moderation vorliegt – hingegen jedoch nicht alle Effekte eine inhaltliche Bedeutung hinsichtlich der Interpretation von FRE haben. Die Effekte mit inhaltlicher Bedeutung hinsichtlich des FRE sind in diesem Sinne β_1 (quantifiziert den FRE bei $\xi_1 = \xi_2 = 0$), β_4 (quantifiziert die Veränderung des FRE β_1, wenn sich ξ_1 um eine empirische Einheit verändert und $\xi_2 = 0$ ist), β_5 (quantifiziert die Veränderung des FRE β_1, wenn sich ξ_2 um eine empirische Einheit verändert und $\xi_1 = 0$ ist), sowie β_7 (quantifiziert die Veränderung des FRE β_1, wenn sich das Produkt von ξ_1 und ξ_2 um eine empirische Einheit verändern). Die Effekte β_2, β_3 und β_6 haben hinsichtlich der Identifikation der Einflussfaktoren des hier modellierten FRE keine inhaltliche

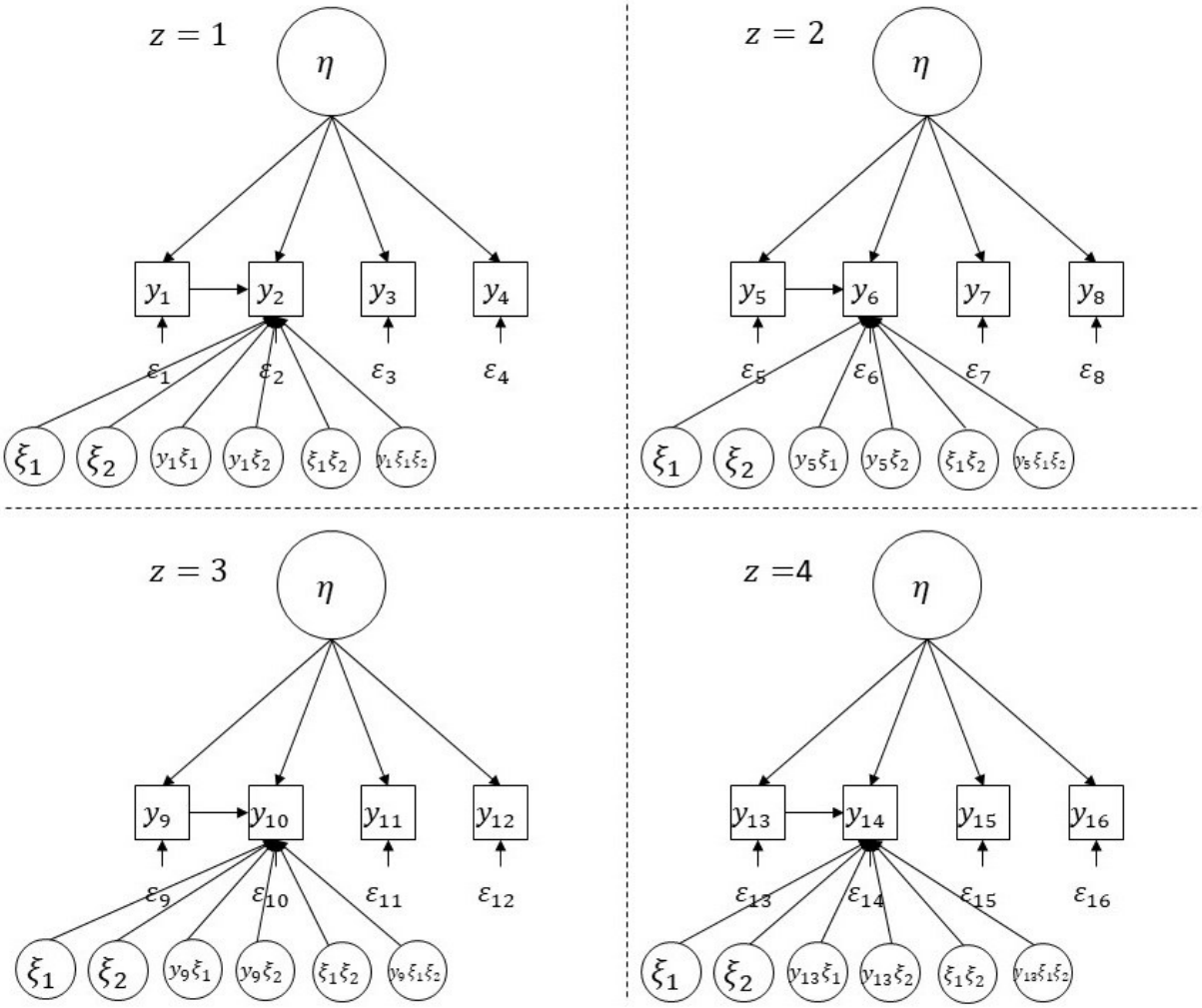
Bedeutung und sind lediglich aufgrund ihres Kontrolleffektes zwecks Spezifikation der moderierten Moderation zu modellieren.

Tabelle 2: Bedeutung der Interaktionseffekte

Koeffizient	Spezifiziert Effekt von	Bedeutung
$\boldsymbol{\beta_1}$	y_j	Effekt von y_j auf y_{j+1} für $\xi_1 = \xi_2 = 0$
$\boldsymbol{\beta_2}$	ξ_1	Effekt von ξ_1 auf y_{j+1} für $y_j = \xi_2 = 0$
$\boldsymbol{\beta_3}$	ξ_2	Effekt von ξ_2 auf y_{j+1} für $y_j = \xi_1 = 0$
$\boldsymbol{\beta_4}$	$y_j\xi_1$	Veränderung in β_1, wenn ξ_1 sich um eine empirische Einheit verändert (für $\xi_2 = 0$)
$\boldsymbol{\beta_5}$	$y_j\xi_2$	Veränderung in β_1, wenn ξ_2 sich um eine empirische Einheit verändert (für $\xi_1 = 0$)
$\boldsymbol{\beta_6}$	$\xi_1\xi_2$	Effekt auf y_{j+1}, wenn sich ξ_1 und ξ_2 je um eine empirische Einheit verändert (für $y_j = 0$)
$\boldsymbol{\beta_7}$	$y_j\xi_1\xi_2$	Veränderung in β_1, wenn sich ξ_1 und ξ_2 je um eine empirische Einheit verändern

Bei der Spezifikation der Interaktionsterme ist insbesondere zu berücksichtigen, dass es sich bei zwei der drei besagten Interaktionsvariablen um latente Variablen handelt, wobei für latente Interaktionsmodelle besondere Bedingungen zu berücksichtigen sind. Entsprechend wurde das LMS-Verfahren bzw. wo dies nicht implementiert, alternativ der Unconstrained Approach zur Schätzung der Interaktionsparameter vorgeschlagen, da Simulationsstudien zeigen, dass beide Verfahren in ihrer Effizienz relativ vergleichbar sowie einfach in der Anwendung sind. Dabei sei bereits an dieser Stelle erwähnt, dass für die Analysen in dieser Arbeit der Unconstrained Approach verwendet wird, da LMS in der verwendeten Statistiksoftware nicht implementiert ist (vgl. Kapitel 3.2.2.3).

Abbildung 17: Strukturmodell zur Identifikation und Kontrolle von Fragereihenfolgeneffekten



Entsprechend dieser Spezifikationen ist in Abbildung 17 das NLC-CFA Modell zur Identifikation und Kontrolle von FRE auf Basis regressiver Beziehungen zwischen Kontext- und Zielfrage auf Indikatorebene sowie latenten Interaktionseffekten für eine beispielhafte Fragebatterie aus $p = 4$ Items je Gruppe (wobei p hier theoretisch keine Obergrenze hat und die Untergrenze lediglich durch eine positive Menge an Freiheitsgraden im Gesamtmodell limitiert wird, siehe Brown, 2006) innerhalb einer Multigruppenanalyse mit vier Gruppen vereinfacht dargestellt, wobei der FRE sowie die Interaktionseffekte aus Gründen der Übersichtlichkeit lediglich zwischen den ersten beiden Indikatorvariablen dargestellt sind. Da FRE jedoch grundsätzlich zwischen allen aufeinanderfolgenden Items einer Fragebatterie auftreten können, sind in dem zugehörigen Gleichungssystem (vgl. Formel 53) regressive Beziehungen sowie Interaktionen zwischen allen aufeinanderfolgenden Items der Fragebatterie berücksichtigt. In Abbildung 17 ist zudem die latente Eigenschaft der Moderatorvariablen berücksichtigt, wobei die Indikatorvariablen jener latenten Moderatoren aufgrund der Übersichtlichkeit in der Abbildung nicht

dargestellt sind. Innerhalb des Gleichungssystems in Formel 53 wird an dieser Stelle entsprechend von drei Indikatorvariablen je latentem Moderator ausgegangen. Zur Reduktion der Komplexität ist das Gleichungssystem zur Definition von y_j für die vier Gruppen in den Formel 53.1 bis Formel 53.4 separat definiert. Entsprechend stellt η das zu schätzende Merkmal von Interesse dar, welches aufgrund metrischer Invarianz und einer in Formel 53 spezifizierten Fixierung der Varianz auf 1 zwischen allen Gruppen als gleich geschätzt wird. η ist für $z = 1$ durch y_1 bis y_4 gemessen, für $z = 2$ durch y_5 bis y_8 gemessen, für $z = 3$ durch y_9 bis y_{12} gemessen und für $z = 4$ durch y_{13} bis y_{16} gemessen. Bei den Items y_1, y_5, y_9 und y_{13}, den Items y_2, y_6, y_{10} und y_{14}, den Items y_3, y_7, y_{11} und y_{15}, sowie den Items y_4, y_8, y_{12} und y_{16} handelt es sich somit theoretisch um Messungen desselben Merkmals aus verschiedenen Gruppen, welche folglich unterschiedlich durch FRE beeinflusst sind. y_1 (respektive y_5, y_9 und y_{13}) ist folglich Kontextfrage für die Zielfrage y_2 (respektive y_6, y_{10} und y_{14}), welche wiederum gleichzeitig Kontextfrage für die Zielfrage y_3 (respektive y_7, y_{11} und y_{15}) etc. darstellt. e_1 bis e_4 seien die zufälligen Messfehler bzw. Residuen der jeweiligen Fragen (respektive der weiteren Residuen in den anderen Gruppen). ξ_1 repräsentiert die Informationszugänglichkeit als latenter Moderator, welche, ebenfalls beispielhaft, gemessen wird durch die Items x_1 bis x_3 und für alle Gruppen als gleich angenommen wird. Äquivalent stellt ξ_2 das Framing der Befragtenrolle als Moderator dar, welcher durch die Items x_4 bis x_6 gemessen und ebenfalls für alle Gruppen als gleich angenommen wird. Zuletzt werden die Haupt- sowie alle entsprechend zu spezifizierenden Interaktionseffekte als gleich zwischen aufeinanderfolgenden Items, jedoch als verschieden zwischen den Gruppen geschätzt, um die Anzahl zu schätzender Parameter und damit die Modellkomplexität zu verringern. Dies ist insofern zulässig, als dass für FRE keine theoretische Fundierung dahingehend existiert, ob diese im Verlauf einer Itembatterie zwischen einzelnen Kontext-Zielfragen-Dyaden abnehmen, zunehmen oder gleichbleiben (vgl. Kapitel 2). Soll dies ebenfalls getestet werden, so sind alle regressiven Effekte innerhalb des Modells frei zu schätzen.

$$(1)\quad \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} \eta + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \beta_1 & 0 & 0 & \beta_2 & \beta_3 \\ 0 & \beta_1 & 0 & \beta_2 & \beta_3 \\ 0 & 0 & \beta_1 & \beta_2 & \beta_3 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \beta_4 & \beta_5 & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_4 & \beta_5 & 0 & 0 \\ 0 & 0 & 0 & 0 & \beta_4 & \beta_5 \end{bmatrix} \begin{bmatrix} y_1\xi_1 \\ y_1\xi_2 \\ y_2\xi_1 \\ y_2\xi_2 \\ y_3\xi_1 \\ y_3\xi_2 \end{bmatrix}$$
$$+ \begin{bmatrix} 0 & 0 & 0 & 0 \\ \beta_6 & \beta_7 & 0 & 0 \\ \beta_6 & 0 & \beta_7 & 0 \\ \beta_6 & 0 & 0 & \beta_7 \end{bmatrix} \begin{bmatrix} \xi_1\xi_2 \\ y_1\xi_1\xi_2 \\ y_2\xi_1\xi_2 \\ y_3\xi_1\xi_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

$$(2)\quad \begin{bmatrix} y_5 \\ y_6 \\ y_7 \\ y_8 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} \eta + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \beta_8 & 0 & 0 & \beta_9 & \beta_{10} \\ 0 & \beta_8 & 0 & \beta_9 & \beta_{10} \\ 0 & 0 & \beta_8 & \beta_9 & \beta_{10} \end{bmatrix} \begin{bmatrix} y_5 \\ y_6 \\ y_7 \\ \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \beta_{11} & \beta_{12} & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_{11} & \beta_{12} & 0 & 0 \\ 0 & 0 & 0 & 0 & \beta_{11} & \beta_{12} \end{bmatrix} \begin{bmatrix} y_5\xi_1 \\ y_5\xi_2 \\ y_6\xi_1 \\ y_6\xi_2 \\ y_7\xi_1 \\ y_7\xi_2 \end{bmatrix}$$
$$+ \begin{bmatrix} 0 & 0 & 0 & 0 \\ \beta_{13} & \beta_{14} & 0 & 0 \\ \beta_{13} & 0 & \beta_{14} & 0 \\ \beta_{13} & 0 & 0 & \beta_{14} \end{bmatrix} \begin{bmatrix} \xi_1\xi_2 \\ y_5\xi_1\xi_2 \\ y_6\xi_1\xi_2 \\ y_7\xi_1\xi_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \end{bmatrix}$$

Formel 53

$$(3)\quad \begin{bmatrix} y_9 \\ y_{10} \\ y_{11} \\ y_{12} \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} \eta + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \beta_{15} & 0 & 0 & \beta_{16} & \beta_{17} \\ 0 & \beta_{15} & 0 & \beta_{16} & \beta_{17} \\ 0 & 0 & \beta_{15} & \beta_{16} & \beta_{17} \end{bmatrix} \begin{bmatrix} y_9 \\ y_{10} \\ y_{11} \\ \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \beta_{18} & \beta_{19} & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_{18} & \beta_{19} & 0 & 0 \\ 0 & 0 & 0 & 0 & \beta_{18} & \beta_{19} \end{bmatrix} \begin{bmatrix} y_9\xi_1 \\ y_9\xi_2 \\ y_{10}\xi_1 \\ y_{10}\xi_2 \\ y_{11}\xi_1 \\ y_{11}\xi_2 \end{bmatrix}$$
$$+ \begin{bmatrix} 0 & 0 & 0 & 0 \\ \beta_{20} & \beta_{21} & 0 & 0 \\ \beta_{20} & 0 & \beta_{21} & 0 \\ \beta_{20} & 0 & 0 & \beta_{21} \end{bmatrix} \begin{bmatrix} \xi_1\xi_2 \\ y_9\xi_1\xi_2 \\ y_{10}\xi_1\xi_2 \\ y_{11}\xi_1\xi_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \end{bmatrix}$$

$$(4)\quad \begin{bmatrix} y_{13} \\ y_{14} \\ y_{15} \\ y_{16} \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} \eta + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \beta_{22} & 0 & 0 & \beta_{23} & \beta_{24} \\ 0 & \beta_{22} & 0 & \beta_{23} & \beta_{24} \\ 0 & 0 & \beta_{22} & \beta_{23} & \beta_{24} \end{bmatrix} \begin{bmatrix} y_{13} \\ y_{14} \\ y_{15} \\ \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \beta_{25} & \beta_{26} & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_{25} & \beta_{26} & 0 & 0 \\ 0 & 0 & 0 & 0 & \beta_{25} & \beta_{26} \end{bmatrix} \begin{bmatrix} y_{13}\xi_1 \\ y_{13}\xi_2 \\ y_{14}\xi_1 \\ y_{14}\xi_2 \\ y_{15}\xi_1 \\ y_{15}\xi_2 \end{bmatrix}$$
$$+ \begin{bmatrix} 0 & 0 & 0 & 0 \\ \beta_{27} & \beta_{28} & 0 & 0 \\ \beta_{27} & 0 & \beta_{28} & 0 \\ \beta_{27} & 0 & 0 & \beta_{28} \end{bmatrix} \begin{bmatrix} \xi_1\xi_2 \\ y_{13}\xi_1\xi_2 \\ y_{14}\xi_1\xi_2 \\ y_{15}\xi_1\xi_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{15} \\ \varepsilon_{16} \end{bmatrix}$$

$$(5)\quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} \lambda_{x_1 1} & 0 \\ \lambda_{x_2 1} & 0 \\ \lambda_{x_3 1} & \lambda_{x_4 2} \\ 0 & \lambda_{x_5 2} \\ 0 & \lambda_{x_4 2} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{bmatrix}$$

$$(6)\quad \boldsymbol{\Psi} = \begin{bmatrix} 1 & & & \\ 0 & 1 & & \\ 0 & 0 & 1 & \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$(7)\quad \boldsymbol{\Phi} = \begin{bmatrix} 1 & \\ 0 & 1 \end{bmatrix}$$

$$(8)\quad \boldsymbol{\Theta}_{\varepsilon}$$

Formel 53

$$= \begin{bmatrix} \theta_{\varepsilon_{11}} & & & & & & & & & & & & & & & \\ 0 & \theta_{\varepsilon_{22}} & & & & & & & & & & & & & & \\ 0 & 0 & \theta_{\varepsilon_{33}} & & & & & & & & & & & & & \\ 0 & 0 & 0 & \theta_{\varepsilon_{44}} & & & & & & & & & & & & \\ 0 & 0 & 0 & 0 & \theta_{\varepsilon_{55}} & & & & & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_{66}} & & & & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_{77}} & & & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_{88}} & & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_{99}} & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_{1010}} & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_{1111}} & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_{1212}} & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_{1313}} & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_{1414}} & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_{1515}} & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_{1616}} \end{bmatrix}$$

$$(9)\quad \boldsymbol{\Theta}_{\delta} = \begin{bmatrix} \theta_{\delta_{11}} & & & & & \\ 0 & \theta_{\delta_{22}} & & & & \\ 0 & 0 & \theta_{\delta_{33}} & & & \\ 0 & 0 & 0 & \theta_{\delta_{44}} & & \\ 0 & 0 & 0 & 0 & \theta_{\delta_{55}} & \\ 0 & 0 & 0 & 0 & 0 & \theta_{\delta_{66}} \end{bmatrix}$$

Um das vorgeschlagene NLC-CFA Modell zur Identifikation und Kontrolle von FRE zu testen wird, in Anlehnung an das Vorgehen klassischer SEM-Analysen (Urban & Mayerl, 2014) ein

mehrstufiges Verfahren vorgeschlagen, welches innerhalb eines experimentellen Designs und folglich einer Multigruppenanalyse (mit weiteren moderatorbedingten Gruppen) eingebettet ist, so dass zwecks Identifikation der Effekte Versuchsgruppe (fixierte Fragereihenfolge) und Kontrollgruppe (randomisierte Fragereihenfolge) gegenübergestellt werden können. Die Schritte dieses mehrstufigen Verfahrens lauten:

1. Spezifikation der Messmodelle mittels CFA unter Berücksichtigung metrischer Invarianz innerhalb einer Multigruppenanalyse.
2. Erweiterung des Messmodells um regressive Beziehungen (ohne Verwendung von Interaktionstermen) auf Indikatorebene zur Identifikation von FRE innerhalb einer Multigruppenanalyse zur Identifikation des Modell-Fit (vgl. Kapitel 5.3 für eine Erörterung des Model Fit innerhalb der NLC-CFA).
3. Erweiterung des Strukturmodells um Interaktionsterme zur Identifikation der Moderatoreffekte inklusive Durchführung eines log-likelihood-ratio Test zur Ableitung des Model Fit (vgl. Kapitel 5.3).

Bei dem hier beschriebenen NLC-CFA Modell handelt es sich entsprechend abschließend um dasjenige Modell, welches im analytischen Teil dieser Arbeit hinsichtlich seiner Effizienz und Präzision bei der Quantifizierung von FRE und entsprechend der Bereinigung von durch FRE verursachte Verzerrungen untersucht und dem im vorangegangenen Kapitel vorgestellten CFA-MTMM Modell gegenübergestellt wird.

3.3.3 Gegenüberstellung der vorgeschlagenen Modelle und Ausblick

Zum Abschluss der methodischen Betrachtung von FRE wurden zwei CFA Modelle vorgeschlagen, mit welchen FRE identifiziert und kontrolliert werden können und welche nachfolgend vergleichend und zusammenfassend diskutiert werden, um deren Möglichkeiten zur Identifikation und Kontrolle von FRE hinsichtlich möglicher Hypothesentests aufzuzeigen. Diesbezüglich wurden bisher auf inhaltlicher Ebene mögliche kognitive Prozesse diskutiert, welche zur Entstehung besagter Effekte führen können (vgl. Kapitel 2). Im Kontext des Auftretens von FRE wurden insbesondere der Modus der Informationsverarbeitung bzw. der Grad an Antwortelaboration, der Grad an Zugänglichkeit von Informationen gegenüber dem Objekt der Befragung, sowie das Framing der Befragtenrolle als mögliche Prädiktoren diskutiert (vgl. Kapitel 2.2). Aus der Diskussion dieser Einflussfaktoren geht insbesondere hervor, dass es im Kontext der Betrachtung von FRE genügt, den Elaborationsgrad qualitativ in eine geringe sowie eine

hohe Elaboration zu differenzieren, während die Zugänglichkeit sowie die Kooperationsbereitschaft, welche sich aus dem Framing der Befragtenrolle ergibt, einem quantitativen Kontinuum entsprechen. Hierauf aufbauend konnte ein vereinfachtes Modell des Auftretens von FRE (vgl. Kapitel 2.3.3) sowie empirisch überprüfbare Hypothesen und notwendige Bedingungen abgeleitet werden, welche wie folgt lauten:

Notwendige Bedingung 1: Geringe Elaboration ist eine notwendige Bedingung für einen Assimilationseffekt.

Notwendige Bedingung 2: Hohe Elaboration ist eine notwendige Bedingung für einen Kontrasteffekt.

H_{1a}: Desto geringer die kognitive chronische Zugänglichkeit von Information mit Relevanz für das Objekt der Befragung bei geringer Antwortelaboration, desto mehr nähert sich die Antwort auf eine Zielfrage der Antwort auf die Kontextfrage an.

H_{1b}: Bei indifferenter Befragtenrolle und geringer Antwortelaboration nähert sich die Antwort auf eine Zielfrage der Antwort auf die Kontextfrage an.

H_2: Bei kooperativer Befragtenrolle und hoher Antwortelaboration entfernt sich die Antwort auf eine Zielfrage von der Antwort auf die Kontextfrage, desto geringer die kognitive chronische Zugänglichkeit von Information mit Relevanz für das Objekt der Befragung.

Bei den beiden in Kapitel 3 vorgeschlagenen Modellen zur Identifikation und Kontrolle von FRE sowie zur Bereinigung von durch diesen verursachte Verzerrungen handelt es sich um ein CFA-MTMM Modell sowie um das neu eingeführte NLC-CFA Modell, welche im Folgenden zusammenfassend und vergleichend, auch hinsichtlich ihrer jeweiligen Möglichkeiten zur Überprüfung der oben aufgeführten Bedingungen und Hypothesen, erörtert werden.

Das erste vorgeschlagene Modell basiert auf einer Adaption der MTMM-Logik (vgl. Kapitel 3.1.3) als Multigruppenmodell (vgl. Kapitel 3.2.2.1) und wird im Kontext dieser Arbeit als CFA-MTMM Modell bezeichnet. Kapitel 3.3.1 zusammenfassend werden innerhalb dieses CFA Modells Kovarianzen auf Messfehlerebene zwischen aufeinanderfolgenden Indikatoren (den Kontext-Zielfragen-Dyaden) eines beliebigen latenten Konstruktes utilisiert, um systematische Messfehler zu quantifizieren, welche zwischen solchen Dyaden auftreten. Weiterhin wird besagtes CFA Modell mit Kovarianzen auf Messfehlerebene innerhalb einer Multigruppenanalyse eingebettet, wobei Faktorladungen und die Varianz des latenten Konstruktes zwischen den Gruppen als gleich geschätzt werden, während die systematischen Messfehler zwischen den

Gruppen als verschieden geschätzt werden. Zur Gruppierung wurde ein 4-Gruppen Design vorgeschlagen, wobei sich die jeweiligen Gruppen aus einer Verschränkung des experimentellen Designs (fixierte Fragereihenfolge als Versuchsgruppe und randomisierte Fragereihenfolge als Kontrollgruppe) mit dem dichotomisierten Grad an Antwortelaboration (geringe Elaboration und hohe Elaboration) als Moderatorvariable ergeben. Sind die geschätzten Kovarianzen zwischen den Messfehlern in der Folge verschieden zwischen den Gruppen, so ist dies, je nach Ausprägung der Kovarianzen innerhalb der einzelnen Gruppen, auf die Variation der Fragereihenfolge sowie der Antwortelaboration zurückzuführen. Tabelle 1 in Kapitel 3.2.2.1 zeigt diesbezüglich eine vereinfachte Interpretationshilfe der jeweiligen Gruppeneffekte auf wobei vereinfacht formuliert werden kann, dass FRE dann identifiziert sind, wenn sich die Messfehlerkovarianz zwischen den Versuchs- und Kontrollgruppen unterscheiden, während der Einfluss der Antwortelaboration dadurch identifiziert werden kann, wenn sich die Messfehlerkovarianz zwischen den Gruppen mit geringer und hoher Evaluation unterscheiden.

Demgegenüber wurde ein NLC-CFA Modell eingeführt und diskutiert, welches FRE ebenfalls auf Indikatorebene innerhalb einer Multigruppenanalyse mit als gleich geschätzten Faktorladungen und als gleich geschätzter Varianz eines beliebigen latenten Konstruktes von Interesse quantifiziert, wobei das NLC-CFA Modell keine Messfehlerkovarianzen wie das CFA-MTMM Modell, sondern regressive Beziehungen zwischen aufeinanderfolgenden Items (den Kontext-Zielfragen-Dyaden) utilisiert. Diese regressiven Beziehungen werden dabei zwischen den Gruppen als verschieden geschätzt. Der Vorteil bei der Verwendung regressiver Beziehungen gegenüber Kovarianzen ist hierbei, dass der entsprechende regressive Effekt, welcher den FRE quantifiziert, mittels Interaktionsanalyse durch Spezifikation nichtlinearer Effekte auf zusätzliche moderierende latente Einflussfaktoren zurückgeführt werden kann, ohne dass hierdurch eine Spezifikation weiterer Gruppen notwendig ist, wie es etwa bei einer Moderatoranalyse auf Basis einer Multigruppenanalyse notwendig ist (vgl. Kapitel 3.3.2). Bei den vorgeschlagenen latenten Moderatorvariablen wurde entsprechend der theoretischen Diskussion in Kapitel 2 der Grad an Informationszugänglichkeit sowie das Framing der Befragtenrolle diskutiert, was insgesamt in sieben interpretierbaren Haupt- und Interaktionseffekten resultiert (vgl. Tabelle 2 in Kapitel 3.3.2). Würden die beiden Moderatoren hingegen dichotomisiert und würden die Moderatoren innerhalb einer Multigruppenanalyse analysiert werden, so würde dies zum einen mit einem Informationsverlust als Folge der Dichotomisierung einhergehen, zum anderen würde dies die Spezifikation von (mindestens) 16 statt vier Gruppen notwendig machen (zwei Expe-

rimentalbedingungen mal zwei Extremgruppen an Antwortelaboration mal zwei Extremgruppen an Informationszugänglichkeit mal zwei Extremgruppen an Frames der Befragtenrolle), was zu einer deutlich komplexeren und damit fehleranfälligeren Analyse führen würde.

Das vorgeschlagene CFA-MTMM Modell mit dem NLC-CFA Modell vergleichend ergibt sich somit, dass innerhalb der NLC-CFA eine einfachere Möglichkeit besteht, mehr zusätzliche, kontinuierliche und latente exogene Variablen bei der Erklärung der FRE zu berücksichtigen, als dies durch das CFA-MTMM Modell gegeben ist. In der Folge können innerhalb des NLC-CFA Modell mehr relevante Informationen bei der Modellschätzung berücksichtigt werden, wodurch es ebenfalls zu einer effizienteren Schätzung der Faktorladungen des entsprechenden latenten Konstruktes kommen sollte, sodass das latente Konstrukt von Interesse selbst schließlich effizienter geschätzt wird. Insgesamt ist innerhalb des NLC-CFA Modells somit eine effizientere Schätzung der Modellparameter als im CFA-MTMM Modell zu erwarten. Die Schätzung eines NLC-CFA Modells geht über dies hinaus mit einem deutlichen Informationsgewinn gegenüber dem CFA-MTMM Modell einher, da innerhalb des NLC-CFA Modell die genauen Einflusspfade der zugrundeliegenden systematischen Messfehler aufgrund der Spezifikation der nichtlinearen Effekte analysiert werden können.

Werden FRE entsprechend der hier vorgeschlagenen Modelle quantifiziert, so stellt sich abschließend die Frage, wie die entsprechenden Effekte hinsichtlich der formulierten Hypothesen und notwendigen Bedingungen zu interpretieren sind und unter welchen Bedingungen welche der Hypothesen und notwendigen Bedingungen verworfen werden können. Diesbezüglich erfolgt eine abschließende Übersicht, wobei die vertiefende Diskussion in Kapitel 5.2 geführt wird:

Notwendige Bedingung 1: (Geringe Elaboration ist eine notwendige Bedingung für einen Assimilationseffekt): *Notwendige Bedingung 1* greift die in Kapitel 2.2 diskutierte Theorie auf, dass Assimilationseffekte einen geringen Grad an Antwortelaboration bedingen, wobei sich besagte Effekte durch eine Erhöhung der Kovarianz zwischen Kontext- und Zielfrage in der Versuchsgruppe gegenüber der Kontrollgruppe äußert. *Notwendige Bedingung 1* impliziert somit, dass innerhalb der Versuchsgruppe mit geringer Antwortelaboration eine positive Verzerrung der Kovarianz von Kontext- und Zielfrage relativ zur Kontrollgruppe mit geringer Antwortelaboration stattfindet. Die Bedingung kann entsprechend durch beide Modelle verworfen werden, wenn innerhalb der Versuchsgruppe mit geringer Antwortelaboration keine positive Verzerrung der Kovarianz zwischen Kontext- und Zielfrage gegenüber der Kontrollgruppe identifiziert werden kann. Innerhalb des CFA-MTMM Modells

erfolgt die entsprechende Identifikation durch einen Vergleich der Messfehlerkovarianzen zwischen den Gruppen, innerhalb des NLC-CFA Modells durch einen Vergleich der Haupt- und Interaktionseffekte zwischen den Gruppen.

Notwendige Bedingung 2: (Hohe Elaboration ist eine notwendige Bedingung für einen Kontrasteffekt): *Notwendige Bedingung 2* basiert auf der ebenfalls in Kapitel 2.2 diskutierten Theorie, dass Kontrasteffekte einen hohem Grad an Antwortelaboration bedingen, wobei sich besagte Effekte durch eine Verringerung der Kovarianz zwischen Kontext- und Zielfrage in der Versuchsgruppe gegenüber der Kontrollgruppe äußert. *Notwendige Bedingung 2* wird entsprechend verworfen, wenn innerhalb der Versuchsgruppe mit hoher Antwortelaboration keine negative Verzerrung der Kovarianz zwischen Kontext- und Zielfrage gegenüber der Kontrollgruppe identifiziert werden kann. Erneut erfolgt die entsprechende Identifikation innerhalb des CFA-MTMM Modells durch einen Vergleich der Messfehlerkovarianzen zwischen den Gruppen und innerhalb des NLC-CFA Modells durch einen Vergleich der Haupt- und Interaktionseffekte zwischen den Gruppen.

H_{1a} (Desto geringer die kognitive chronische Zugänglichkeit von Information mit Relevanz für das Objekt der Befragung bei geringer Antwortelaboration, desto mehr nähert sich die Antwort auf eine Zielfrage der Antwort auf die Kontextfrage an): H_{1a} basiert auf der Theorie kognitiver Zugänglichkeit (vgl. Kapitel 2.1 und 2.2) und geht davon aus, dass sich diese in dem Sinne auf das Auftreten von Assimilationseffekten auswirkt, als dass durch die Kontextfrage situativ aktivierte Informationen bei gering elaborierten Antworten einen umso geringeren Einfluss auf die Antwort der Zielfrage haben, desto zugänglicher Informationen mit Relevanz für das Objekt der Befragung sind. H_{1a} wird entsprechend innerhalb des NLC-CFA Modells verworfen, wenn der Interaktionseffekt, welcher den Einfluss der Zugänglichkeit auf den FRE repräsentiert (Entspricht β_4 gemäß der Systematik in Tabelle 2) innerhalb der Versuchsgruppe mit geringer Antwortelaboration nicht signifikant bzw. positiv ist (da ein negativer Effekt erwartet wird: je geringer die Zugänglichkeit, desto positiver die Verzerrung) bzw. sich nicht signifikant von dem analogen Effekt in der Kontrollgruppe unterscheidet. Die Überprüfung der Hypothese innerhalb des vorgeschlagenen CFA-MTMM Modells ist hingegen ohne Anwendung eines 8-Gruppen Designs mit dichotomisierter Informationszugänglichkeit nicht möglich (vgl. Kapitel 3.2.2.1 und 3.3.1).

H_{1b} (Bei indifferenter Befragtenrolle und geringer Antwortelaboration nähert sich die Antwort auf eine Zielfrage der Antwort auf die Kontextfrage an): H_{1b} basiert auf einer Adaption der Theorie der Frame Selektion (vgl. Kapitel 2.1 und 2.2) und geht davon aus, dass sich die

Befragtenrolle in dem Sinne auf das Auftreten von Assimilationseffekten auswirkt, als dass die Befragtenrolle bedingt, welche Informationen zum Zeitpunkt der Urteilsbildung zugänglich sind und als relevant bewertet werden. H_{1b} wird entsprechend durch das NLC-CFA Modell verworfen, wenn der Interaktionseffekt, welcher den Einfluss der Befragtenrolle auf den FRE repräsentiert (β_5 gemäß der Systematik in Tabelle 2) innerhalb der Versuchsgruppe mit geringer Antwortelaboration nicht signifikant bzw. positiv ist (da ein negativer Effekt erwartet wird: je negativer die Befragtenrolle, desto positiver die Verzerrung) bzw. sich nicht signifikant von dem analogen Effekt in der Kontrollgruppe unterscheidet. Die Überprüfung der Hypothese innerhalb des vorgeschlagenen CFA-MTMM Modells ist erneut ohne Anwendung eines 8-Gruppen Designs, in diesem Fall jedoch mit dichotomisiertem Framing der Befragtenrolle nicht möglich (vgl. Kapitel 3.2.2.1 und 3.3.1).

H_2 (Bei kooperativer Befragtenrolle und hoher Antwortelaboration entfernt sich die Antwort auf eine Zielfrage von der Antwort auf die Kontextfrage, desto geringer die kognitive chronische Zugänglichkeit von Information mit Relevanz für das Objekt der Befragung): H_2 zuletzt basiert auf der Überlegung eines interaktiven Effektes zwischen Zugänglichkeit und Befragtenrolle bei erhöhter Antwortelaboration (vgl. Kapitel 2.3) und geht davon aus, dass vor allem kooperative Befragte mit gering Zugänglichkeit dazu neigen, Informationen bewusst aus einem Urteil zu subtrahieren, was in der Folge zu einem Kontrasteffekt führt. H_2 wird entsprechend durch das NLC-CFA Modell verworfen, wenn der Interaktionseffekt, welcher den interaktiven Einfluss der Befragtenrolle und der Zugänglichkeit auf den FRE repräsentiert (β_7 gemäß der Systematik in Tabelle 2), innerhalb der Versuchsgruppe mit erhöhter Antwortelaboration nicht signifikant bzw. negativ ist (da ein positiver Effekt erwartet wird: je positiver die Befragtenrolle und je geringer die Zugänglichkeit, desto negativer die Verzerrung) bzw. sich nicht signifikant von dem analogen Effekt in der hoch elaborierten Kontrollgruppe unterscheidet. Auch die Überprüfung dieser Hypothese ist innerhalb des vorgeschlagenen CFA-MTMM Modells ohne weitere Anpassung nicht möglich, wobei ein 16-Gruppen Designs mit dichotomisierter Informationszugänglichkeit und dichotomisiertem Framing der Befragtenrolle angewandt werden müsste (vgl. Kapitel 3.2.2.1 und 3.3.1).

Entsprechend dieser Systematik ist es nachfolgend das Ziel, die beiden hier vorgeschlagenen Modelle mittels Simulationsanalysen zu evaluieren und zu vergleichen, um eine zuverlässige Lösung zur Identifikation und Kontrolle von FRE zu etablieren.

4 Simulationsstudie zur Evaluation der Modelle

Ziel dieser Arbeit ist es nicht nur, auf theoretisch-inhaltlicher Ebene mögliche Determinanten von FRE zu identifizieren, sondern ebenfalls Modelle abzuleiten und zu diskutieren, mit welchen besagte Determinanten auf statistischer Ebene identifiziert und kontrolliert werden können, um so zu messfehlerbereinigten Schätzungen zu gelangen. Diesbezüglich wurde ein Modell basierend auf der etablierten MTMM Logik (CFA-MTMM Modell) sowie ein innovatives CFA Modell (NLC-CFA Modell) vorgeschlagen (vgl. Kapitel 3). Es soll zudem gezeigt werden, dass das NLC-CFA Modell zu effizienteren Schätzungen relevanter Modellparameter führt, als das etablierte CFA-MTMM Modell, wenn die zur Schätzung verwendeten Daten mit FRE simuliert sind. Entsprechend verfolgt die nachfolgende Simulationsstudie das Ziel, die generelle Tauglichkeit des NLC-CFA Modells zur Identifikation und Kontrolle von FRE zu überprüfen und ebenso dem CFA-MTMM Modell gegenüberzustellen. Zu diesem Zweck werden Datensätze, welche FRE gemäß der in Kapitel 2 diskutierten Logik beinhalten, auf Basis variabler Modell- und Simulationsspezifikationen (Fallzahl, Anzahl an Indikatoren, approximierter Anteil der Common Variance, Anteil der systematischen Fehlervarianz innerhalb der gesamten Fehlervarianz) in mehrfachen Monte-Carlo Iterationen simuliert, wobei das NLC-CFA Modell sowie das CFA-MTMM Modell je Iteration geschätzt wird. Über einen Vergleich der geschätzten Parameter mit den bekannten wahren Werten können die Modelle schließlich hinsichtlich ihrer Präzision (Schätzung mit geringen Standardfehlern) und Effizienz (unverzerrte Schätzung der Parameter) evaluiert und verglichen werden. Innerhalb der Simulationsstudie wird somit in den folgenden Schritten vorgegangen:

1. Initial werden einfache CFA-Modelle ohne spezifizierte FRE für Datensätze mit und ohne simulierte FRE geschätzt, um die Auswirkungen vorhandener, unspezifizierter FRE auf die Faktorladungen als Indikator der Konstrukt-Validität und der Komposit-Reliabilität zu überprüfen. Die absoluten Verzerrungen, berechnet als Standard-Verzerrung (vgl. Kapitel 4.1 und 4.2) der Faktorladungen dieser Modelle ohne Spezifikation des FRE werden mit den Standard-Verzerrungen der Faktorladungen bei gegebener Spezifikation des FRE, sowohl auf Basis des NLC-CFA Modells, als auch auf Basis des CFA-MTMM Modells verglichen und auf die variierenden Modell- und Simulationsspezifikationen zurückgeführt, um die Abhängigkeiten der Standard-Verzerrungen zu analysieren. Ein Vergleich der Verzerrungen gibt zudem Aufschluss darüber, welches

der Modelle zu einer effizienteren Schätzung der Faktorladungen aufgrund der Messfehlerbereinigung führt (Analyse der Verzerrungen von Faktorladungen, Kapitel 4.4).

2. Auf Basis der geschätzten Effekte der Modelle mit spezifiziertem FRE (Regressionskoeffizienten für die NLC-CFA und Messfehlerkovarianzen für die CFA-MTMM) werden Standard-Verzerrungen der Effektschätzungen berechnet (vgl. Kapitel 4.1 und 4.2). Die Standard-Verzerrungen werden je Iteration auf die variierenden Modell- und Simulationsspezifikationen zurückgeführt um zu analysieren, wie sich Modell- und Simulationsspezifikationen auf Standard-Verzerrungen in den geschätzten Effekten auswirken. Die Schätzungen werden diesbezüglich als effizient bezeichnet, wenn Standard-Verzerrungen lediglich gering ausfallen. Es erfolgt zudem eine vergleichende Diskussion hinsichtlich der allgemeinen Effizienz des NLC-CFA Modells und des CFA-MTMM Modells (Effekt Effizienz Analyse, vgl. Kapitel 4.5).
3. Geschätzte Standardfehler der Effekte der Modelle mit spezifiziertem FRE werden ebenfalls je Iteration auf die Modell- und Simulationsspezifikationen zurückgeführt um somit zu analysieren, wie sich diese auf die Höhe der geschätzten Standardfehler auswirken. Fallen Standardfehler lediglich sehr klein aus, wird von einer präzisen Schätzung ausgegangen. Auch hier erfolgt eine vergleichende Diskussion zwischen dem NLC-CFA Modell und dem CFA-MTMM Modell (Standardfehler Präzisionsanalyse, vgl. Kapitel 4.6).
4. Das vorgeschlagene NLC-CFA Modell sowie das CFA-MTMM Modell werden mit zuvor festgelegter Modell- und Simulationsspezifikation für alle Iterationen (d.h. ohne variierende Modell- und Simulationsspezifikationen zwischen den Iterationen) zwecks detaillierter Analyse der relevanten Parameter geschätzt. Dabei erfolgt erneut eine vergleichende Diskussion der beiden Modelle. Die Analyseergebnisse tragen schließlich ebenfalls zu der Einschätzung bei, ob die NLC-CFA tauglich für die Identifikation und Kontrolle von FRE sowie das Testen der in Kapitel 2.3.3 vorgeschlagenen Hypothesen ist oder ob ein etabliertes CFA-MTMM Modell hierzu besser geeignet ist (Detaillierte Analyse der angepassten Benchmark Modelle, vgl. Kapitel 4.7).

4.1 Simulationsparameter und Analysesystematik

Ziel der vorliegenden Simulationsanalyse (Analyse der Verzerrungen von Faktorladungen, Effekt Effizienz Analyse und Standardfehler Präzisionsanalyse, für die detaillierte Analyse der

angepassten Benchmark Modelle siehe Kapitel 4.7) ist es, die Verzerrungen der Faktorladungen und der Effektkoeffizienten sowie die Höhe der Standardfehler unterschiedlicher CFA Modelle auf jeweils zufällig variierende Modell- und Simulationsspezifikationen zurück zu führen und die Abhängigkeit der Verzerrungen und der Standardfehler von den jeweiligen Spezifikationen zwischen den Modellen zu vergleichen. Bei den zu vergleichenden Modellen handelt es sich um ein CFA-MTMM Modell sowie ein NLC-CFA Modell (vgl. Kapitel 3.3) sowie ein einfaches CFA Modell ohne weitere Spezifikationen oder Kontrolle des FRE als Kontrollmodell, alle Modelle entsprechen Multigruppenmodellen. Da dies eine komplexe Datensimulation und -Auswertung bedingt, werden die hierfür notwendigen Schritte im Folgenden einzeln und detailliert behandelt.

Iterationen *i*, Gruppen *z* und Fallzahl *n*

Um Verzerrungen und Standardfehler der geschätzten Parameter je Modell zu ermitteln, werden, über alle Iterationen *i* (bei insgesamt 10000 Iterationen) hinweg Daten für verschiedene Gruppen *z* simuliert. Der Index *z* repräsentiert die Befragtengruppen analog zur methodischen Diskussion in Kapitel 3 mit 1: fixierte Fragereihenfolge und geringe Antwortelaboration (FRE als Assimilationseffekt simuliert); 2: fixierte Fragereihenfolge und hohe Antwortelaboration (FRE als Kontrasteffekt simuliert); und 3: randomisierte Fragereihenfolge (ohne FRE simuliert), wobei in Gruppe 3 nicht zwischen geringer und hoher Antwortelaboration differenziert wird, da für Gruppen mit randomisierter Reihenfolge kein FRE zu erwarten ist, wodurch der Grad an Elaboration theoriebedingt keinen weiteren Einfluss nimmt (vgl. Kapitel 2). Alle Daten werden mit einer Fallzahl von n_i, welche sich zwischen den Iterationen *i*, jedoch nicht zwischen den Modellgruppen *z* unterscheidet, simuliert.

Simulierte Parameter und Indikatoren *j* der latenten Variablen

Die Simulation jeder Iteration *i* umschließt je eine latente Variable η_i und je q_i manifeste Indikatoren y_{izj} (mit *j = 1,…,q),* wobei y_{izj} für die einzelnen Messungen einer beliebigen Fragebatterie je Gruppe *z* stehen, mit welcher des Merkmal η_i gemessen werden soll und innerhalb welcher FRE zu identifizieren und analysieren sind; je zwei latente Variablen ξ_{i1} (repräsentiert die Zugänglichkeit von Informationen) und ξ_{i2} (repräsentiert das Framing der Befragtenrolle) mit je drei manifesten Indikatoren x_{i1} bis x_{i3} als Indikatoren von ξ_{i1} und x_{i4} bis x_{i6} als Indikatoren von ξ_{i2}. $\eta_i, \xi_{i1}, \xi_{i2}$ sowie x_{i1} bis x_{i6} unterscheiden sich zwischen den Iterationen *i*, jedoch nicht zwischen den Modellgruppen *z*, während sich y_{izj} zwischen den Iterationen *i* und den Modellgruppen *z* unterscheidet, da der hier wirkende Einfluss des FRE zwischen den Gruppen

verschieden ist. Zusätzlich werden alle Indikatoren x_{i1} bis x_{i6} derart simuliert, dass sie theoretisch einem standardisierten Wertebereich ($\mu = 0;\ \sigma^2 = 1$) entsprechen und über gleiche Faktorladungen verfügen, während für y_{izj} die Bedingung $\mu = 0$ ebenfalls zutrifft, σ^2 jedoch zwischen den Iterationen aufgrund der unterschiedlich simulierten systematischen Messfehler lediglich (geringfügig) um den Wert 1 schwankt. Für y_{izj} wurden weiterhin gleiche, wahre, unstandardisierte Faktorladungen zwischen Indikatoren und Gruppen, jedoch nicht zwischen Iterationen simuliert, wodurch die Interpretation der Ergebnisse vereinfacht ist, da jeder Indikator *y* innerhalb einer Iteration über die gleiche, wahre Faktorladung verfügt.

Simulierte Effekte zwischen den Gruppen

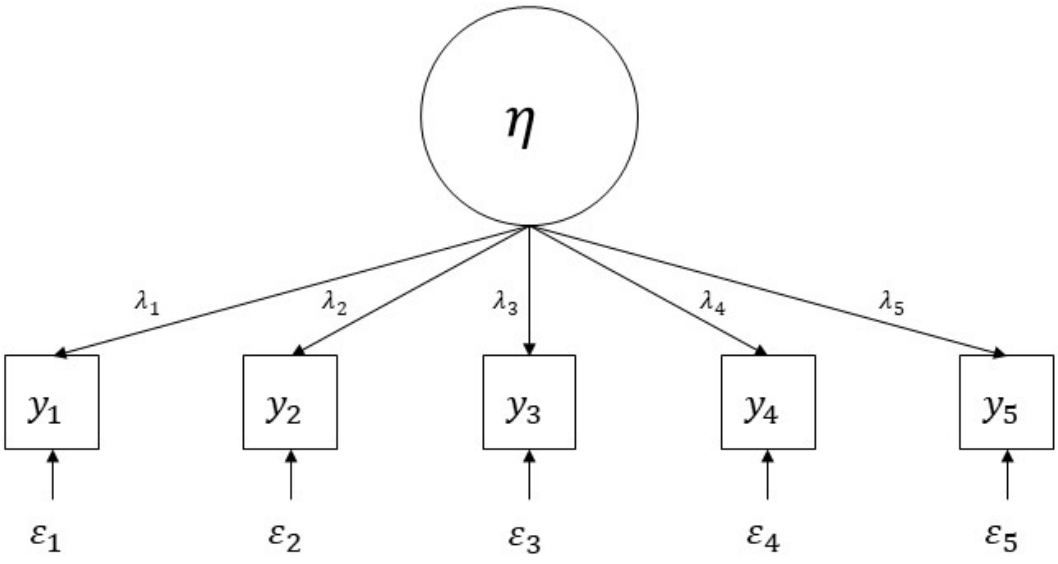
Die Indikatoren y_{izj} sind derart simuliert, dass in der Gruppen mit „fixierter Fragereihenfolge und geringer Elaboration“ (z = 1) eine negative Abhängigkeit jeder Zielfrage y_{j+1} von der jeweiligen Kontextfrage y_j in Interaktion mit ξ_{i1} ($y_j\xi_1$) und mit ξ_{i2} ($y_j\xi_2$) existiert und alle anderen Haupt- und Interaktionseffekte keinen Effekt aufweisen, während in der Gruppe mit „fixierter Fragereihenfolge und hoher Elaboration“ (z = 2) eine positive Abhängigkeit jeder Zielfrage von der jeweiligen Kontextfrage in Interaktion mit ξ_{i1} und ξ_{i2} ($y_j\xi_1\xi_2$) existiert und alle anderen Haupt- und Interaktionseffekte keinen Effekt aufweisen (vgl. Kapitel 3.3). Für Gruppe 3 mit randomisierter Reihenfolge wird hingegen kein FRE erwartet, entsprechend werden keine Abhängigkeiten der Zielfrage von der Kontextfrage oder den exogenen Variablen ξ_{i1} und ξ_{i2} simuliert. Alle vorhandenen Effekte sind somit konditional in Abhängigkeit von ξ_{i1} und ξ_{i2} simuliert, wobei ξ_{i1} und ξ_{i2} (wie weiter oben beschrieben) mit $\mu = 0$ und $\sigma^2 = 1$ simuliert sind und die Effekte von $y_j\xi_1$, $y_j\xi_2$ und $y_j\xi_1\xi_2$ auf die Zielfrage y_{j+1} für $\xi_{i1} = 0$ bzw. $\xi_{i2} = 0$ einen wahren Wert von 0 aufweisen.

Zu schätzende Modelle innerhalb der einzelnen Iterationen

Innerhalb jeder einzelnen Iteration werden die beiden in Kapitel 3 beschriebenen Modelle (CFA-MTMM und NLC-CFA) als Multigruppenmodelle über alle Gruppen *z* hinweg geschätzt. Die beiden Modelle werden somit pro Iteration und Gruppe auf Basis der gleichen, simulierten Daten geschätzt. Beide Modelle werden in allen Iterationen unter Berücksichtigung metrischer Invarianz zwischen den Gruppen und mit als gleich geschätzten Effekten zwischen Kontext-Zielfragen-Dyaden innerhalb einer Gruppe, jedoch mit frei geschätzten Effekten zwischen den Gruppen und Iterationen, auf Basis des ML-Algorithmus geschätzt. ML wird verwendet, da alle

Variablen kontinuierlich und zumindest annährend normalverteilt simuliert werden (vgl. Kapitel 3.2.2.1 und 4.2). Die Effektschätzung erfolgt für das CFA-MTMM Modell auf Basis der Messfehlerkovarianzen und für das NLC-CFA Modell auf Basis regressiver Beziehungen auf Indikatorebene. Für die NLC-CFA wird der Unconstrained Approach für die Schätzung der latenten Interaktionseffekte verwendet, da die Simulation mit dem Statistikprogramm „R“ (R Core Team, 2021) und der Library „lavaan“ (Rosseel et al., 2022) programmiert ist und LMS dort für derart komplexe Modelle nicht zur Verfügung steht (vgl. Kapitel 3.3.2). Zusätzlich zur Schätzung dieser Modelle wird auf Basis der gleichen Daten in jeder einzelnen Iteration ein einfaches Multigruppenmodell ohne Berücksichtigung metrischer Invarianz, ohne Spezifikation der FRE (weder durch Messfehlerkovarianzen, noch durch regressive Beziehungen) und ohne sonstige, für einfache konfirmatorische Faktoranalysen unübliche Modellspezifikationen auf Basis des ML-Algorithmus frei geschätzt, um Verzerrungen von Faktorladungen bei nicht spezifizierten FRE abzuleiten. Abbildung 18 zeigt dieses einfache CFA Modell mit $q = 5$ Indikatoren auf, wobei aufgrund erhöhter Übersichtlichkeit auf die Darstellung mehrerer Gruppen z verzichtet wurde. Demgegenüber sind in Abbildung 19 und Abbildung 20 die zu schätzenden NLC-CFA und CFA-MTMM Modelle, erneut mit $q = 5$ Indikatoren und ohne die Darstellung mehrerer Gruppen z aufgezeigt. Aufgrund der variablen Anzahl an Indikatoren zwischen den Iterationen i werden Modelle mit $q = 5$ Indikatoren im Folgenden als Benchmark Modelle bezeichnet, um diese einfach von Modellen mit $q = 3$ oder $q = 4$ Indikatoren abgrenzen zu können.

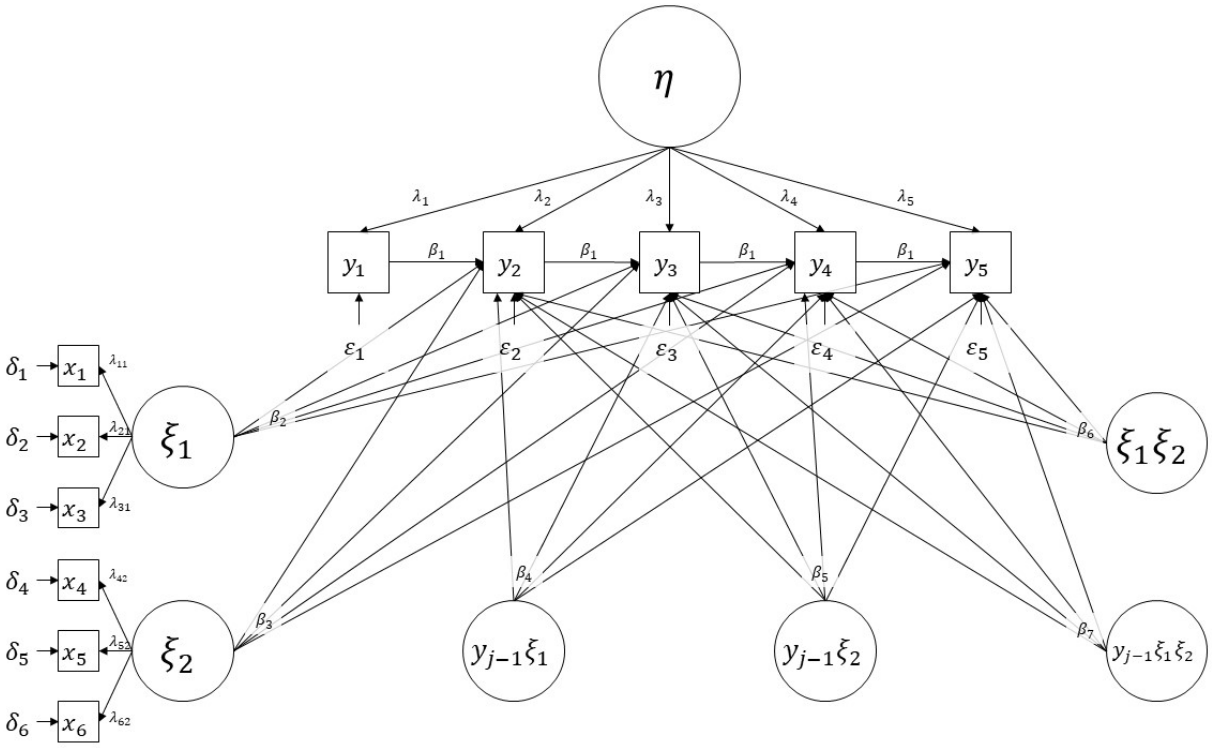
Abbildung 18: Einfaches CFA Benchmark Modell



Variierende Modell- und Simulationsspezifikationen

Zu den Modell- und Simulationsspezifikationen, welche zwischen den einzelnen Iterationen *i* variiert werden, gehört die Fallzahl *n*, die Modellkomplexität, indiziert durch die Anzahl an Indikatoren *q*, der Anteil der FRE-bedingten Fehlervarianz an der gesamten Fehlervarianz $\frac{\sigma_{sys}^2}{\sigma_{sys}^2+\sigma_{rand}^2}$ und der approximierte Anteil der Common Variance an der gesamten Varianz, was aufgrund der Simulationsspezifikation (vgl. Kapitel 4.2) der „wahren", simulierten, unstandardisierten Faktorladung λ_{sim} entspricht.

Abbildung 19: NLC-CFA Benchmark Modell

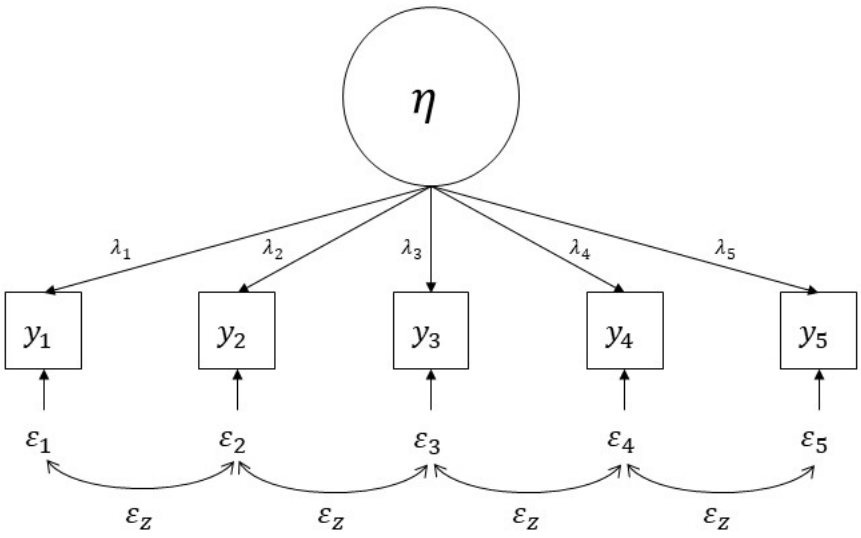


Variierende simulierte Fallzahl

Hinsichtlich der Fallzahl ist grundsätzlich festzuhalten, dass sowohl Schätzalgorithmen wie ML, als auch Verfahren zur Schätzung nichtlinearer Modelle wie etwa LMS oder der Unconstrained Approach in ihrer Effizienz abhängig sind von einer ausreichenden Fallzahl. So können zu geringe Fallzahlen beispielsweise zu fehlerhaften Schätzungen der Standardfehler, zu ineffizienten Schätzungen der Koeffizienten oder gar generell zu Konvergenzproblemen führen (Brown, 2006; Jackman et al., 2011; Klein & Moosbrugger, 2000; Marsh et al., 2004; Marsh et

al., 2006; Steinmetz et al., 2011). Um die Auswirkungen einer zu geringen Fallzahl auf die Identifizierbarkeit von FRE sowie deren Determinanten bzw. allgemein die Güte der Schätzungen innerhalb der vorgeschlagenen Modelle zu untersuchen, wird die Fallzahl zwischen den Iterationen zufällig variiert. Diesbezüglich wird für jede Iteration *i* ein neues *n* per Pseudo-Zufallsgenerator zwischen 200 und 5000 mit zurücklegen gezogen, was als Reichweite von *n* einer in sozialwissenschaftlichen Untersuchungen üblichen mittleren bis hohen Fallzahl (Buchner, Erdfelder, & Faul, 1996) entspricht. Innerhalb einer Iteration *i* sind die Fallzahlen zwischen den Gruppen *z* somit identisch, während sie sich zwischen den Iterationen unterscheiden.

Abbildung 20: CFA-MTMM Benchmark Modell



Variierende simulierte Anzahl an Indikatoren

Neben der Fallzahl *n* hat ebenfalls die gesamte Modellkomplexität einen Einfluss auf die Konvergenz von Modellen und die effiziente Schätzung von Effekten, da diese die Anzahl zu schätzender Parameter, somit die Freiheitsgrade und entsprechend ebenfalls die Größe der empirischen und zu schätzenden Kovarianzmatrix bedingen (Brown, 2006; Reinecke & Pöge, 2010). Um entsprechend die Abhängigkeiten der Schätzungen von der Modellkomplexität zu analysieren, wird die Anzahl an Indikatoren *q* innerhalb des zur Analyse stehenden latenten Konstruktes η_i variiert, so dass diese sich zwischen den Iterationen *i* unterscheiden, jedoch innerhalb der Gruppen *z* identisch sind. *q* entspricht einer ganzen Zufallszahl zwischen 3 und 5 mit

zurücklegen. Zu beachten ist, dass die latenten Moderatorvariablen ξ_{i1} und ξ_{i2} über alle Iterationen *i* hinweg konstant über je drei Indikatoren verfügen.

Variierende simulierte Faktorladungen

Weiterhin werden die simulierten unstandardisierten Faktorladungen λ_{sim} als Indikatoren der Konstruktvalidität (Kline, 2016; Schermelleh-Engel & Schweizer, 2012; Urban & Mayerl, 2014) und der Komposit-Reliabilität (Cho 2016) herangezogen und innerhalb der einzelnen Iterationen variiert. Die simulierten Faktorladungen λ_{sim} können dabei (unter der Bedingung einer Fixierung der Varianz des latenten Konstruktes auf 1 und einer freien Schätzung der Faktorladungen) als Approximation des relativen Anteils an Common Variance an der gesamten Varianz[7] bzw. als Annäherung an die standardisierte Faktorladung interpretiert werden, da die Varianz der Indikatoren in den einzelnen Iterationen (geringfügig) um den Wert 1 schwankt. λ_{sim} wird als identisch zwischen den Indikatoren *j* und den Modellgruppen *z* und als verschieden zwischen den Iterationen *i* simuliert. Die Reichweite von λ_{sim} liegt zwischen 0,3 und 0,9. Der unstandardisierte Faktorladungsbereich von 0,3 bis 0,9 entspricht somit einem Bereich geringer bis hoher Varianzausschöpfung des Indikators durch das latente Konstrukt (Brown, 2006). Der generierte Wert von λ_{sim} ist innerhalb des Intervalls von 0,3 bis 0,9 gleichförmig zufällig für alle Iterationen.

Variierender simulierter Anteil der Messfehler und der FRE

Da mit einem zunehmenden Anteil an Common Variance derjenige Varianzanteil absolut geringer wird, welcher auf Messfehler zurückgeht, da die Summe aller Varianzanteile innerhalb der Indikatoren je Iteration geringfügig um den Wert 1 schwankt (vgl. Kapitel 4.2), muss der absolute FRE in Summe umso geringer sein, je höher die Faktorladungen sind, da der relative Anteil der systematischen Fehlervarianz umso kleiner wird, desto größer der relative Anteil der Common Variance ist. Der relative Anteil der gesamten Fehlervarianz (sowohl der zufälligen, als auch der systematischen) an der gesamten Varianz wird bei der vorliegenden Modellspezifikation demnach durch $1 - \lambda_{sim}$ approximiert.

[7] y_j wird mit $y_j = \lambda\eta + (1-\lambda)\varepsilon_j$ simuliert, wobei ε_j für Gruppe 1 und Gruppe 2 den Anteil des systematischen Messfehlers enthällt, wodurch die simulierte Faktorladung gleichzeitig dem Anteil der Common Variance, also $Var(\eta)$ approximiert, da η mit einer Varianz von 1 simuliert wurde, die Varianz von ε_j aufgrund unterschiedlich stark ausgeprägter FRE innerhalb der einzelnen Iterationen geringfügig um den Wert 1 schwankt und $Cov(\eta, \varepsilon_j) = 0$ gilt. Wäre die Varianz von ε_j in jeder Iteration konstant 1, dann würde die Faktorladung exakt dem Anteil an Common Variance entsprechen. Da dies nicht zutrifft und die Varianz von ε_j in jeder Iteration zufällig geringfügig um den Wert 1 schwankt, handelt es sich bei der Faktorladung lediglich um eine Approximation des Anteils an Common Variance, vgl. Kapitel 4.2

Entsprechend dieser Betrachtung lässt sich ebenfalls die konkrete Höhe des FRE innerhalb der einzelnen Iterationen variieren, indem der relative Anteil simulierter Fehlervarianz weiter in den relativen Anteil an zufälliger Fehlervarianz und an systematischer Fehlervarianz zerlegt wird, wobei innerhalb des Anteils der systematischen Fehlervarianz die den FRE bedingenden Interaktionsterme simuliert werden. Zu beachten ist hierbei jedoch, dass FRE lediglich für die Gruppen $z = 1$ und $z = 2$ simuliert werden und für $z = 3$ entsprechend kein Anteil systematischer Fehlervarianz vorliegt. Der relative Anteil der systematischen Fehlervarianz an der gesamten, also der zufälligen und der systematischen Fehlervarianz $\frac{\sigma_{sys}^2}{\sigma_{sys}^2+\sigma_{rand}^2}$ ist zwischen den Indikatoren j und den Gruppen z (mit Ausnahme von $z = 3$) identisch und variiert zwischen den Iterationen i, wobei auch wenn die Anteile der systematischen Fehlervarianzen zwischen den Gruppen $z = 1$ und $z = 2$ identisch sind, sich die Ausprägung des eigentlichen FRE zwischen den Gruppen unterscheidet (vgl. Kapitel 2.3). Die Reichweite des relativen Anteils systematischer Fehlervarianz innerhalb der gesamten Fehlervarianz liegt zwischen 10% (entspricht einem sehr geringen Effektanteil) und 90% (entspricht einem sehr hohen Effektanteil) und ist gleichförmig zufällig für alle Iterationen.

Metaanalyse der einzelnen Modellschätzungen

Für alle Iterationen werden variierende Spezifikationen wie die Fallzahl n, die Anzahl an Indikatoren q etc. sowie die auf Basis der simulierten Daten geschätzten Faktorladungen, Effekte und Standardfehler der einzelnen Modelle in einem separaten Datensatz $\boldsymbol{D}$ gespeichert, wobei die Spalten den gespeicherten Parametern und Maßzahlen und die Zeilen der Iteration der Simulation i entsprechen. Auf Basis dieses Datensatzes, welcher alle relevanten Daten der einzelnen Modellschätzungen je Iteration enthält, können schließlich auf Basis eines geeigneten Regressionsverfahrens (vgl. Kapitel 4.3) die geschätzten Verzerrungen und Standardfehler auf die Simulations- und Modellspezifikation zurückgeführt werden.

Diesbezüglich werden die Variablen des Datensatzes $\boldsymbol{D}$, welche die Simulations- und Modellspezifikation repräsentieren und als unabhängige Variablen in den Regressionsmodellen dienen derart rekodiert, dass die Konstante a je Regression die entsprechende Verzerrung bzw. den Standardfehler angibt, welche bzw. welcher für ein Modell mit einer Fallzahl n von 1500, fünf Indikatoren q innerhalb der Fragebatterie, innerhalb welcher FRE erwartet werden, einem relativen Anteil der systematischen Fehlervarianz, gegeben durch den FRE, innerhalb der gesamten Fehlervarianz von 50% ($\frac{\sigma_{sys}^2}{\sigma_{sys}^2+\sigma_{rand}^2} = 0{,}5$), sowie einer unstandardisierten Faktorladung λ_{sim} von 0,6 über alle fünf Indikatoren hinweg (was als Approximation des relativen Anteils an

Common Variance von 60% bzw. der standardisierten Faktorladung interpretiert werden kann) zu erwarten ist, was insgesamt einem Modell mit durchschnittlicher bis hoher Fallzahl, einer ausreichenden Datenqualität und einem erhöhten FRE entspricht. Neben der Anzahl an Indikatoren (q = 5) gehört n = 1500, $\lambda_{sim} = 0{,}6$ und $\frac{\sigma_{sys}^2}{\sigma_{sys}^2+\sigma_{rand}^2} = 0{,}5$ ebenfalls zur Definition der weiter oben eingeführten „Benchmark Modelle“.

Berechnung der Verzerrung der geschätzten Faktorladungen

Um zu testen, wie sich Simulations- und Modellspezifikationen auf die Schätzung und Verzerrung von Modellparametern auswirken, müssen auf Basis der Schätzer aus Datensatz ***D*** abhängige Variablen für die an die Datensimulation anschließende Regression berechnet werden. Gemäß den einleitend formulierten Fragestellungen handelt es sich hierbei zunächst um die Verzerrung der geschätzten Faktorladungen λ_{est}, da diese insbesondere als Indikatoren der Datenqualität, namentlich der Konstrukt-Validität und der Komposit-Reliabilität dienen (vgl. Kapitel 3.1.4). Grundsätzlich wird, um Verzerrungen aller Art zu quantifizieren, der wahre Wert benötigt, um die Differenz zum geschätzten Wert zu berechnen. Diesbezüglich entspricht der wahre Wert der Faktorladungen der Indikatoren von η je Iteration *i* der simulierten Faktorladung λ_{sim_i}, wobei bei der Modellschätzung zu berücksichtigen ist, dass die Varianz von η auf 1 fixiert und Faktorladungen frei geschätzt werden müssen, damit die simulierte Faktorladung potenziell mit dem korrekten Wert geschätzt werden kann. Die Verzerrung der einzelnen Faktorladungen je Iteration *i*, Gruppe *z* und Indikator *j* wird in Anlehnung an die Formel zur Berechnung der Standardabweichung als sogenannte Standard-Verzerrung (*SV*) entsprechend Formel 54 berechnet. Dabei kann die *SV* als Wurzel der durchschnittlichen quadrierten Abweichung der geschätzten Faktorladungen $\lambda_{est_{izj}}$ zur wahren Faktorladung λ_{sim_i} innerhalb von Gruppe *z* und Iteration *i* interpretiert werden, was insgesamt den Vorteil hat, dass sich eventuell vorhandene positive und negative Verzerrungen innerhalb einer Gruppe nicht im Durchschnitt ausgleichen. Die *SV* wird zuletzt mit 100 multipliziert, um die Interpretation der Ergebnisse zu vereinfachen.

$$\mathrm{SV}_{\lambda_{iz}} = \sqrt{\frac{\sum_{j=1}^{q}(\lambda_{est_{izj}} - \lambda_{sim_i})^2}{q}} * 100 \qquad \textbf{Formel 54}$$

Berechnung der Verzerrungen der geschätzten Regressionseffekte

Weiterhin lassen sich die Verzerrungen der geschätzten Regressionseffekte des NLC-CFA Modells ebenfalls als *SV* auf Basis der Differenz der wahren Effekte β_{sim} und der geschätzten Effekte β_{est} quantifizieren. Diesbezüglich kann der simulierte, wahre *b*-the Effekt $\beta_{sim_{izb}}$ (bei insgesamt sieben Haupt- und Interaktionseffekten, $b = 1, \ldots, 7$, vgl. Tabelle 3 und Kapitel 3.3) je Iteration und Gruppe durch Multiplikation der reversen Faktorladung $1 - \lambda_{sim_i}$ (als Quantifizierung des Anteils der gesamten Fehlervarianz) mit dem relativen Anteil der systematischen Fehlervarianz an der gesamten Fehlervarianz $(\frac{\sigma_{sys}^2}{\sigma_{sys}^2+\sigma_{rand}^2})_i$ sowie einem modellgruppenspezifischen, technischen Simulationskoeffizienten $\beta_{tech_{zb}}$ (mit $\beta_{tech_{zb}} = -0{,}5$ für $z = 1$ und $b = 4$ sowie $b = 5$; $\beta_{tech_{zb}} = 1$ für $z = 2$ und $b = 7$; und $\beta_{tech_{zb}} = 0$ für alle anderen Fälle, vgl. Tabelle 3 und Kapitel 4.2) multipliziert werden, um die wahren Effekte je Iteration und Gruppe $\beta_{sim_{izb}}$ zu quantifizieren. Die einzelnen Effekte sind demzufolge entsprechend als Anteile der Zielfragen zu interpretieren, welche durch die entsprechenden Faktoren oder Moderatoren bedingt sind. Die Berechnung der *SV* erfolgt erneut in Anlehnung an die Formel zur Berechnung der Standardabweichung, jedoch wird für die *SV* der Effekte die Verzerrung im Durchschnitt über alle Gruppen und nicht innerhalb der Gruppen gemäß Formel 56 gemittelt. Formel 55 zeigt diesbezüglich die Berechnung des wahren *b*-the Effektes $\beta_{sim_{izb}}$. Die *SV* wird zudem ebenfalls mit 100 multipliziert, um die Interpretation der Ergebnisse zu vereinfachen.

Tabelle 3: Technische Koeffizienten der Effektberechnung je Gruppe

b	1	2	3	4	5	6	7
Effekt von	y_{j-1}	ξ_1	ξ_2	$y_{j-1}\xi_1$	$y_{j-1}\xi_2$	$\xi_1\xi_2$	$y_{j-1}\xi_1\xi_2$
$\beta_{tech_{z=1}}$	0	0	0	**-0,5**	**-0,5**	0	0
$\beta_{tech_{z=2}}$	0	0	0	0	0	0	**1**
$\beta_{tech_{z=3}}$	0	0	0	0	0	0	0

$$\beta_{sim_{izb}} = (1 - \lambda_{sim_i})(\frac{\sigma_{sys}^2}{\sigma_{sys}^2 + \sigma_{rand}^2})_i \beta_{tech_{zb}} \qquad \textbf{Formel 55}$$

$$\mathrm{SV}_{\beta_{ib}} = \sqrt{\frac{\sum_{z=1}^{3}(\beta_{est_{izb}} - \beta_{sim_{izb}})^2}{z}} * 100 \qquad \textbf{Formel 56}$$

Berechnung der Verzerrungen der geschätzten Messfehlerkovarianzen

Die Verzerrungen der geschätzten Messfehlerkovarianzen der CFA-MTMM Modelle lassen sich theoretisch analog zu den Verzerrungen der geschätzten Regressionseffekte berechnen, wobei hierfür die Bestimmung der wahren Messfehlerkovarianz $\varepsilon_{sim_{iz}}$ notwendig ist. Problematisch ist hieran für dieses Beispiel jedoch, dass auf Basis der simulierten Daten keine „wahre" Messfehlerkovarianz existiert, da systematische Kovarianzen zwischen Indikatoren, welche unabhängig sind vom Einfluss des latenten Konstruktes, als nichtlineare Beziehungen zwischen den Indikatoren und nicht lediglich als Kovarianz zwischen den Residuen der Indikatoren simuliert sind. Die Spezifikation von Messfehlerkovarianzen, wie für das CFA-MTMM Modell beschrieben, entspricht demnach generell einer Fehlspezifikation der jeweiligen Modelle.

Wird jedoch berücksichtigt, dass FRE (egal, ob es sich hierbei um Assimilations- oder Kontrasteffekte handelt) konditional in Abhängigkeit der exogenen Moderatoren ξ_1 und ξ_2 simuliert wurden und auf Basis der vorliegenden Simulationssystematik generell nur dann auftreten, wenn ξ_1 und ξ_2 von ihrem simulierten Mittelwert 0 abweichen (vgl. Kapitel 4.2, insbesondere Schritt 5 und 6), so führt eine Konstanthaltung von ξ_1 und ξ_2 um ihren Mittelwert dazu, dass die quantifizierte Höhe des FRE ebenfalls 0 betragen muss. Wird weiterhin berücksichtigt, dass der Einfluss von ξ_1 und ξ_2 innerhalb der CFA-MTMM nicht spezifiziert wird, ξ_1 und ξ_2 also in der Schätzung Konstant bleiben und nur der FRE für die konditionale Bedingung $\xi_1 = \xi_2 = 0$ geschätzt werden kann, so muss der durch die Messfehlerkovarianz quantifizierte FRE ebenfalls 0 betragen. Eine Abweichung der geschätzten Messfehlerkovarianz je Gruppe von 0 entspricht also bei der vorliegenden Datensimulation der Verzerrung der zu erwartenden Messfehlerkovarianz, da diese bei Konstantem ξ_1 und ξ_2 den Wert 0 annehmen muss.

Entsprechend erfolgt die Berechnung der *SV* der Messfehlerkovarianz als Wurzel der quadrierten, tatsächlichen Messfehlerkovarianz je Gruppe ε_{iz} (da dies automatisch die Abweichung von 0 impliziert) gemäß Formel 57, wobei die *SV* zwecks vereinfachter Interpretation erneut mit 100 multipliziert wird.

$$\mathrm{SV}_{\varepsilon_{iz}} = \sqrt{\varepsilon_{iz}{}^2} * 100 \qquad \textbf{Formel 57}$$

Berechnung der (durchschnittlich) geschätzten Standardfehler

Zuletzt soll die Höhe der Standardfehler der Effektschätzer ebenfalls auf die variierenden Modellspezifikationen sowie die Simulationsparameter zurückgeführt werden. Im Allgemeinen können Standardfehler in SEM als Messungen der Präzision der Schätzer in dem Sinne verstanden werden, als dass je geringer die Standardfehler, desto höher die Präzision der Schätzer, desto höher die Effizienz des Verfahrens und desto geringer der Typ 2 Fehler ausfällt (Moosbrugger et al., 2009). Eine solche Überprüfung der Effizienz ist über die Analyse der Effektverzerrungen hinaus für die vorliegende Arbeit insofern relevant, als dass es sich bei dem definierten NLC-CFA Modell um ein nichtlineares Modell handelt, für welches herkömmliche Model Fit Maße wie etwa Chi², CFI, RMSEA oder SRMR (Brown, 2006; Browne & Cudeck, 1992; Kline, 2016; Little, 2013; Reinecke & Pöge, 2010; Urban & Mayerl, 2014), welche für gewöhnlich zur Modellevaluation verwendet werden, nicht zuverlässig interpretiert werden können (Klein & Moosbrugger, 2000). Dies liegt daran, dass diese im Wesentlichen auf Evaluationen der Kovarianz-Matrix basieren, während Kovarianzen keine Informationen über Nichtlinearität in den Daten beinhalten, wodurch die Evaluation des Model Fit bei nichtlinearen Modellen ein ungelöstes Problem der Methodenforschung bleibt (Schermelleh-Engel, Kerwer, & Klein, 2014). Zur Lösung dieses Problems werden zwar bereits Vorschläge diskutiert, welche jedoch noch weiterer Analysen und Evaluationen bedürfen (Pfleger, 2019) und in der vorliegenden Simulation aufgrund der enormen Komplexität bei 10000 Iterationen nicht berücksichtigt werden können. Eine Analyse der Standardfehler entspricht somit einer Annäherung an eine Analyse der Präzision der Modellschätzung, wo Model Fit Maße nicht zuverlässig interpretierbar sind. Entsprechend werden für die vorliegende NLC-CFA Analyse die einzelnen Standardfehler SE_{izb} je Effektschätzer b über alle Iterationen i und Gruppen z extrahiert. Anschließend wird für SE_{izb} der Mittelwert über alle Gruppen z berechnet, um zu einem durchschnittlichen Standardfehler je Iteration und Effekt zu gelangen, wobei der durchschnittliche SE mit 100 multipliziert wird, um die Interpretation der Ergebnisse zu vereinfachen (vgl. Formel 58). Für die Standardfehler der Messfehlerkovarianzen der CFA-MTMM Analyse werden analog zu den zuvor beschriebenen Messfehlerkovarianzen die einzelnen, gruppenspezifischen Standardfehler SE_{iz} als abhängige Variable in der Standardfehler Präzisionsanalyse verwendet.

$$\overline{\mathrm{SE}}_{ib} = 100 * \frac{1}{3} \sum_{z=1}^{3} SE_{izb} \qquad \textbf{Formel 58}$$

Zusammenfassung

Sind alle Koeffizienten berechnet, werden schließlich die Ausprägungen von $\mathrm{SV}_{\lambda_{iz}}$, $\mathrm{SV}_{\beta_{ib}}$, $\mathrm{SV}_{\varepsilon_{iz}}$, $\overline{\mathrm{SE}}_{ib}$ und SE_{iz} mittels multipler Regression auf die Ausprägungen von n_i, q_i, λ_{sim_i}, $(\frac{\sigma_{sys}^2}{\sigma_{sys}^2+\sigma_{rand}^2})_i$ sowie die Interaktion von λ_{sim_i} und $(\frac{\sigma_{sys}^2}{\sigma_{sys}^2+\sigma_{rand}^2})_i$ (da sich die absolute Höhe des FRE wie beschrieben multiplikativ aus den beiden Faktoren ergibt) zurückgeführt.

Tabelle 4: Zusammenfassung der simulierten unabhängigen Variablen

Unabhängige Variable	Bedeutung	Intervall der simulierten Werte	Wert innerhalb des Benchmark Modell
$\boldsymbol{n_i}$	Fallzahl innerhalb der Gruppen	200 - 5000	1500
$\boldsymbol{q_i}$	Anzahl an Indikatorvariablen des latenten Konstruktes η	3 - 5	5
$\boldsymbol{\lambda_{sim_i}}$	Wahre, unstandardisierte Faktorladung; approximierte/r standardisierten Faktorladung bzw. Anteil an Common Variance	0,3 - 0,9	0,6
$(\frac{\boldsymbol{\sigma_{sys}^2}}{\boldsymbol{\sigma_{sys}^2+\sigma_{rand}^2}})_i$	Anteil der systematischen Fehlervarianz an der gesamten Fehlervarianz	0,1 - 0,9	0,5

Diesbezüglich sind in Tabelle 4 die unabhängigen Variablen der einzelnen Regressionsmodelle zusammengefasst, welche auf die variierten Parameter der zugrundeliegenden Simulation zurückgehen. Die Spalte „Wert innerhalb des Benchmark Modells“ beschreibt hier, um welchen

Wert die jeweilige unabhängige Variable vor Durchführung der Regressionsanalysen zentriert wurden und folglich, für welche Simulationsparameter der Effekt in der Regressionskonstante *a* interpretierbar ist. Konstante *a* gibt in den einzelnen Regressionsanalysen also den geschätzten Wert der abhängigen Variable bei einer Fallzahl von 1500, bei fünf Indikatorvariablen, bei einer wahren Faktorladung von 0,6 und bei einem Anteil an systematischer Fehlervarianz an der gesamten Fehlervarianz von 50% an.

Tabelle 5: Zusammenfassung der abhängigen Variablen

Abhängige Variable	Bedeutung	Berechnung
$\mathbf{SV}_{\lambda_{iz}}$	Standard-Verzerrung der Faktorladungen über alle Indikatoren, unterschieden nach den verschiedenen Gruppen	Vgl. Formel 54
$\mathbf{SV}_{\beta_{ib}}$	Standard-Verzerrung der einzelnen Regressionseffekte des NLC-CFA Modells zwischen den verschiedenen Gruppen	Vgl. Formel 55 und Formel 56
$\mathbf{SV}_{\varepsilon_{iz}}$	Geschätzte Messfehlerkovarianz des CFA-MTMM Modells, unterschieden nach den verschiedenen Gruppen	Vgl. Formel 57
$\overline{\mathbf{SE}}_{ib}$	Durchschnittlicher Standardfehler der einzelnen Regressionseffekte des NLC-CFA Modells zwischen den verschiedenen Gruppen	Vgl. Formel 58
SE_{iz}	Geschätzter Standardfehler der Messfehlerkovarianz des CFA-MTMM Modells, unterschieden nach den verschiedenen Gruppen	Direkt aus den geschätzten Modellen extrahiert

Daneben sind in Tabelle 5 die abhängigen Variablen zusammengefasst, für welche die jeweiligen Regressionsanalysen mit den entsprechenden unabhängigen Variablen durchgeführt werden. Im Einzelnen handelt es sich hierbei um Analysen zu den Verzerrungen der Faktorladungen, zu den Verzerrungen der Regressionseffekte der NLC-CFA Modelle, zu den Verzerrungen

der Messfehlerkovarianzen der CFA-MTMM Modelle, zu den durchschnittlichen Standardfehlern der NLC-CFA Modelle sowie zu den geschätzten Standardfehlern der Messfehlerkovarianzen der CFA-MTMM Modelle. Entsprechend erfolgt nun die Beschreibung der Simulation im Detail, um alle Ergebnisse der Regressionsanalysen korrekt interpretieren und einordnen zu können.

4.2 Methodik der Datensimulation und -Auswertung

Es folgt die technische Beschreibung der Datensimulation, der Datenaggregation sowie der Auswertung und der Analysen zwecks eindeutiger Interpretierbarkeit der nachfolgenden Analyseergebnisse. Die Beschreibungen beziehen sich hierbei auf die Analyse der Verzerrungen der Faktorladungen (vgl. Kapitel 4.4), die Effekt Effizienz Analyse (vgl. Kapitel 4.5) sowie die Standardfehler Präzisionsanalyse (vgl. Kapitel 4.6). Für die detaillierte Analyse der angepassten Benchmark Modelle erfolgt die Beschreibung der Methodik der Datensimulation und -Auswertung im entsprechenden Kapitel 4.7. Der gesamte Prozess der Simulation und Auswertung kann in die Schritte 1.) Datensimulation; 2.) Datenaggregation und 3.) Analyse der Schätzer unterteilt werden, wobei in Schritt 1 (Datensimulation) die benötigten Datensätze simuliert werden, auf deren Basis die entwickelten CFA-MTMM und NLC-CFA Modelle sowie das einfache CFA Modell in Schritt 2 (Datenaggregation) geschätzt werden. Die aggregierten Ergebnisse dieser einzelnen Schätzungen werden schließlich in Schritt 3 (Analyse der Schätzer) mittels Regressionsanalyse auf die variierenden Simulationsspezifikationen zurückgeführt.

Alle Simulationen und Berechnungen wurden mit dem Statistikprogramm „R“ v4.1.3 (R Core Team, 2021), der Library „lavaan“ v0.6-12 (Rosseel et al., 2022), der Library „moments“ v0.14.1 (Komsta & Novomestky, 2022), der Library „psych“ v2.2.5 (Revelle, 2022) und der Library „lm.beta“ v1.6-2 (Behrendt, 2022) durchgeführt und werden nachfolgend detailliert beschrieben. Der verwendete R-Syntax ist in Anhang 7.1 zu finden.

1. Datensimulation

Im 1. Schritt (Datensimulation) erfolgt die Simulation aller für die weiteren Analysen benötigten Daten. Die Datensimulation erfolgt dabei unter der Annahme, dass die in Kapitel 2 abgeleiteten theoretischen Befunde bezüglich des Auftretens von FRE korrekt sind. Konkret werden mehrere Indikatoren für drei Gruppen innerhalb von 10000 Iterationen simuliert, welche eine beliebige Fragebatterie abbilden, die zur Schätzung eines (beliebigen) latenten Konstruktes η

verwendet werden und innerhalb welcher FRE in zwei von drei Gruppen zwischen allen Kontext-Zielfragen-Dyaden auftreten. Die Datensimulation erfolgt dabei in den folgenden Schritten:

1. Per Pseudo-Zufallsgenerator werden Werte für die Fallzahl *n* (zwischen 200 und 5000), die Anzahl an Indikatoren *q* (mit *j* = 1, …, *q*) des latenten Faktors η, für dessen Indikatoren FRE erwartet werden (zwischen 3 und 5), die unstandardisierte Faktorladung λ_{sim} für Indikatoren des Faktors η (zwischen 0,3 und 0,9) sowie den relativen Anteil des systematischen Messfehlers innerhalb des gesamten Messfehlers $\frac{\sigma_{sys}^2}{\sigma_{sys}^2+\sigma_{rand}^2}$ für Indikatoren von η (zwischen 0,1 und 0,9) als gleichförmige Zufallszahl mit Zurücklegen innerhalb der jeweiligen Intervalle generiert.
2. Für alle drei Gruppen *z* wird je eine Variable η_{sim_z} (unabhängig voneinander) sowie über alle Gruppen hinweg zwei Variablen ξ_{sim_1} und ξ_{sim_2} mit $Cov(\xi_{sim_2}, \xi_{sim_1}) = 0$, $Cov(\xi_{sim_2}, \eta_{sim_z}) = 0$ und $Cov(\eta_{sim_z}, \xi_{sim_1}) = 0$ mit einer Fallzahl *n* und $\sim N(0,1)$ simuliert.
3. Je Gruppe werden insgesamt *q* Residuen $\varepsilon_{sim_{z,j}}$ sowie über alle Gruppen hinweg sechs Residuen δ_{sim_1} bis δ_{sim_6} mit einer Fallzahl von *n* und $\sim N(0,1)$ simuliert.
4. Über alle drei Gruppen hinweg wird $y_{z,j=1}$ wie folgt simuliert:

$$y_{z,j=1} = \lambda_{sim}\eta_{sim_z} + (1 - \lambda_{sim})\varepsilon_{sim_{z,j=1}}$$

5. Für Gruppe 1 wird $y_{z,j}$ mit $j > 1$ wie folgt simuliert:

$$y_{z,j>1} = \lambda_{sim}\eta_{sim_z} + (1 - \lambda_{sim})\left(\frac{\sigma_{sys}^2}{\sigma_{sys}^2 + \sigma_{rand}^2}\left(-0{,}5y_{z,j-1}\xi_{sim_1} - 0{,}5y_{z,j-1}\xi_{sim_2}\right)\right.$$
$$\left. + \left(1 - \frac{\sigma_{sys}^2}{\sigma_{sys}^2 + \sigma_{rand}^2}\right)\varepsilon_{sim_{z,j}}\right)$$

6. Für Gruppe 2 wird $y_{z,j}$ mit $j > 1$ wie folgt simuliert:

$$y_{z,j>1} = \lambda_{sim}\eta_{sim_z} + (1 - \lambda_{sim})\left(\frac{\sigma_{sys}^2}{\sigma_{sys}^2 + \sigma_{rand}^2}y_{zj-1}\xi_{sim_1}\xi_{sim_2}\right.$$
$$\left. + \left(1 - \frac{\sigma_{sys}^2}{\sigma_{sys}^2 + \sigma_{rand}^2}\right)\varepsilon_{sim_{z,j}}\right)$$

7. Für Gruppe 3 wird $y_{z,j}$ mit $j > 1$ wie folgt simuliert:

$$y_{z,j>1} = \lambda_{sim}\eta_{sim_z} + (1 - \lambda_{sim})\varepsilon_{sim_{z,j}}$$

8. Über alle Gruppen hinweg wird x_j mit $j = 1,..,3$ wie folgt simuliert:

$$x_j = 0{,}8\xi_{sim_1} + 0{,}2\delta_{sim_j}$$

9. Über alle Gruppen hinweg wird x_j mit $j = 4,..,6$ wie folgt simuliert:

$$x_j = 0{,}8\xi_{sim_2} + 0{,}2\delta_{sim_j}$$

10. Für alle Indikatoren *x* und *y* werden unter Berücksichtigung der jeweiligen Gruppe *z* alle möglichen Produkt-Indikatoren berechnet, welche für die Spezifikation der latenten Produkte gemäß Unconstrained Approach benötigt werden. Produkt-Indikatoren, welche somit nicht kalkuliert werden müssen umfassen quadratische Produktterme sowie alle Produkte, welche mit y_q gebildet werden, da es sich hierbei um die letzte Zielfrage innerhalb der Fragebatterie handelt, welche wiederum für keine weitere Variable als Kontextfrage dient.

2. Datenaggregation

Nach erfolgreicher Datensimulation erfolgt im nächsten Schritt die Datenaggregation. Dies bedeutet, dass die beiden in Kapitel 3.3 eingeführten Modelle (CFA-MTMM Modell und NLC-CFA Modell) sowie ein einfaches CFA Modell ohne weitere Spezifikation (vgl. Kapitel 4.1) geschätzt werden. Alle drei Modelle werden auf Basis der gleichen, simulierten Daten als Multigruppenmodell über alle drei Gruppen *z* hinweg geschätzt, die Ergebnisse der einzelnen Schätzungen werden für weitere Analysen gespeichert. Im Detail erfolgt die Datenaggregation in den folgenden Schritten:

11. Es wird ein MG-CFA Modell zur Schätzung von η_z auf Basis von y_{zj} mit konfiguraler Invarianz, ML-Schätzer, einer auf 1 fixierten Varianz von η_z und *q* Indikatoren y_{zj} ohne weitere Restriktionen frei geschätzt (einfaches CFA Modell).
12. Resultierende Faktorladungen und Standardfehler werden in einem Datensatz ***D*** innerhalb der Zeile *i* (entspricht der Iteration) und mit den geschätzten Parametern in den Spalten gespeichert.
13. Es wird ein CFA-MTMM Multigruppenmodell (gemäß Kapitel 3.3.1 und 4.1) zur Schätzung von η_z auf Basis von y_{zj} mit metrischer Invarianz, ML-Schätzer, einer auf 1 fixierten Varianz von η_z und *q* Indikatoren y_{zj} mit spezifizierter Messfehler-Kovarianz

zwischen aufeinanderfolgenden Kontext-Zielfragen-Dyaden $\varepsilon_{j+1,j}$ mit j = 1, …, q-1 über alle Gruppen z geschätzt. Die Messfehler-Kovarianz wird zwischen den Gruppen frei geschätzt, innerhalb der Gruppen zwischen den aufeinanderfolgenden Kontext-Zielfragen-Dyaden jedoch als gleich geschätzt.

14. Resultierende Faktorladungen, die Messfehler-Kovarianzen und deren Standardfehler werden in einem Datensatz ***D*** innerhalb der Zeile *i* (entspricht der Iteration) und mit den geschätzten Parametern in den Spalten gespeichert.
15. Es wird ein NLC-CFA Multigruppenmodell (gemäß Kapitel 3.3.2 und 4.1) zur Schätzung von η_z auf Basis von y_{zj} mit metrischer Invarianz, ML-Schätzer, einer auf 1 fixierten Varianz von η_z und q Indikatoren y_{zj} geschätzt. ξ_1 wird diesbezüglich mittels x_1 bis x_3, ξ_2 mittels x_4 bis x_6, ebenfalls mit einer auf 1 fixierten Varianz geschätzt. Latente Produkte werden auf Basis der jeweiligen Produkt-Indikatoren gemäß Unconstrained Approach, erneut mit einer auf 1 fixierten Varianz geschätzt. Um die Anzahl zu schätzender Parameter und damit die Dauer der gesamten Simulation zu verringern[8], wurden alle Faktorladungen von *x* sowie von allen Produkt-Indikatoren als gleich geschätzt. Für alle $y_{j>1}$ wurden zudem die folgenden Strukturgleichungen definiert:

$$y_j = \lambda_j\eta + \beta_1 y_{j-1} + \beta_2\xi_1 + \beta_3\xi_2 + \beta_4 y_{j-1}\xi_1 + \beta_5 y_{j-1}\xi_2 + \beta_6\xi_1\xi_2 + \beta_7 y_{j-1}\xi_1\xi_2 + \varepsilon_j$$

Diesbezüglich wurden die Effekte β_b (mit *b* als Laufindex der zu schätzenden Effekte, b = 1, …, 7) innerhalb einer Gruppe zwischen Kontext-Zielfragen-Dyaden als gleich, aber zwischen den Gruppen als verschieden geschätzt.

16. Resultierende Faktorladungen, Effekte und Standardfehler werden in Datenmatrix ***D*** innerhalb der Zeile *i* (entspricht der Iteration) und mit den geschätzten Parametern in den Spalten gespeichert.
17. Zusätzlich werden *n, q,* λ_{sim}, und $\frac{\sigma_{sys}^2}{\sigma_{sys}^2+\sigma_{rand}^2}$ in Datensatz ***D*** innerhalb der Zeile *i* (entspricht der Iteration) in den entsprechenden Zellen gespeichert.
18. Die Schritte 1 bis 17 werden für jede Iteration *i* (mit insgesamt 10000 Iterationen) wiederholt.

[8] Bei aktueller Spezifikation benötigt die gesamte Datensimulation auf einem 64-Bit-System mit 16 GB DDR4 RAM und Intel i7-1185G7 CPU (3 GHz) acht Tage. Weitere, zu schätzende Parameter erhöhen die Datensimulation aufgrund dessen, dass die zu schätzende Parametermatrix Theta bei Berücksichtigung weiterer Indikatoren nicht linear, sondern überproportional anwächst, ebenfalls überproportional statt linear.

3. Analyse der Schätzer

Im letzten Schritt erfolgt die Analyse der zuvor in den einzelnen Iterationen produzierten Schätzer, wobei es sich hierbei im Einzelnen um die Faktorladungen des CFA Modells, die Effekte, welche die FRE quantifizieren sowie deren Standardfehler handelt. Konkret wird analysiert, 1.) wie Faktorladungen durch unspezifizierte FRE verzerrt werden und durch welches Modell und unter welchen Simulationsbedingungen diese Verzerrung am geringsten ausfällt; 2.) welches der Modelle unter welchen Simulationsbedingungen besser dazu geeignet ist, den FRE zu quantifizieren; und 3.) wie Präzise die Quantifizierung des FRE für die verschiedenen Modelle und Simulationsbedingungen ausfällt. Im Detail bedeutet dies:

19. Auf Basis von Datensatz ***D*** werden Koeffizienten für die durchschnittliche Verzerrung der Faktorladungen innerhalb der drei Gruppen *z* relativ zu den in der Simulation spezifizierten wahren Werten, sowohl für Ergebnisse der einfachen CFA Analyse (ohne Kontrolle der FRE), als auch für die Ergebnisse der CFA-MTMM Analyse und der NLC-CFA Analyse, gemäß Formel 54 (vgl. Kapitel 4.1) berechnet.
20. Auf Basis von Datensatz ***D*** werden Koeffizienten für die durchschnittlichen Verzerrungen der geschätzten Effekte über alle drei Gruppen *z* hinweg relativ zu den in der Simulation spezifizierten wahren Werten gemäß Formel 56 und Formel 57 (vgl. Kapitel 4.1) berechnet.
21. Auf Basis von Datensatz ***D*** werden die durchschnittlichen Standardfehler der geschätzten Effekte über alle drei Gruppen *z* hinweg gemäß Formel 58 berechnet.
22. Zwecks vereinfachter Interpretation der Ergebnisse wird *n* um den Wert 1500 zentriert und durch 1000 dividiert. *q* wird als Dummy-Variable derart rekodiert, dass $q = 5$ der Referenzkategorie entspricht. λ_{sim} wird um den Wert 0,6 zentriert und mit 100 multipliziert. $\frac{\sigma_{sys}^2}{\sigma_{sys}^2+\sigma_{rand}^2}$ wird um den Wert 0,5 zentriert und ebenfalls mit 100 multipliziert.
23. Auf Basis von Datensatz ***D*** wird ein Interaktionsterm für λ_{sim} und $\frac{\sigma_{sys}^2}{\sigma_{sys}^2+\sigma_{rand}^2}$ berechnet.
24. Aus Datensatz ***D*** werden alle Iterationen *i* der Simulation gelöscht, bei welchen die geschätzten Modelle nicht konvergiert sind oder welche über nicht plausible Parameterschätzer (negative Varianzen und Faktorladungen außerhalb eines theoretisch möglichen Bereichs, vgl. Kapitel 4.3) verfügen.
25. Die Varianzen der verschiedenen Koeffizienten von $\mathrm{SV}_{\lambda_{iz}}$, $\mathrm{SV}_{\beta_{ib}}$, $\mathrm{SV}_{\varepsilon_{iz}}$, $\overline{\mathrm{SE}}_{ib}$ und SE_{iz} werden mittels mehrerer multipler Regression auf die Ausprägungen von n_i, p_i, λ_{sim_i}, $(\frac{\sigma_{sys}^2}{\sigma_{sys}^2+\sigma_{rand}^2})_i$ sowie das Produkt von λ_{sim_i} und $(\frac{\sigma_{sys}^2}{\sigma_{sys}^2+\sigma_{rand}^2})_i$ zurück geführt.

26. Alle Ergebnisse der Regressionsanalysen werden in zusammenfassenden Tabellen aggregiert.

Nach erfolgter Datensimulation und -Aggregation, aber noch vor Analyse der Schätzer, erfolgt als Zwischenschritt eine Analyse der Verteilungen aller abhängiger Variablen $SV_{\lambda_{iz}}$, $SV_{\beta_{ib}}$, $SV_{\varepsilon_{iz}}$, $\overline{SE}_{ib}$ und SE_{iz}. Ziel dieser Verteilungsanalyse ist es, das optimale Regressionsverfahren für die Analyse der Schätzer bestimmen zu können. Die entsprechenden Verteilungsanalysen sowie die Entscheidung für ein geeignetes Regressionsverfahren sind im nachfolgenden Kapitel zu finden.

4.3 Verteilungsanalysen der relevanten Variablen

Um die aus der Simulation resultierenden Faktorladungen, Effekte, Standardfehler und Verzerrungen per Regression auf die variierten Simulations- und Modellspezifikationen zurückführen zu können, muss zunächst eine Analyse der Verteilungen der relevanten Variablen durchgeführt werden, um das optimale Regressionsverfahren zu ermitteln. Die Verteilungsanalysen konzentrieren sich dabei auf die abhängigen Variablen, da die unabhängigen Variablen gleichförmig zufällig innerhalb der in Tabelle 4 aufgezeigten Intervalle und unabhängig voneinander simuliert wurden und die theoretischen Verteilungen demnach bekannt sind.

Verteilungen der *SV* der Faktorladungen

Beginnend mit den *SV* der Faktorladungen sind deren Verteilungen für die einzelnen Modelle in Tabelle 6 abgebildet. Da die einfachen CFA Modelle ohne Spezifikation des FRE nicht mit metrischer, sondern konfiguraler Invarianz geschätzt wurden, werden die *SV* dieser Modelle je Gruppe und nicht über alle Gruppen hinweg angegeben (vgl. „CFA Gruppe 1“ bis „CFA Gruppe 3“ in Tabelle 6). Die *SV* der NLC-CFA Modelle (vgl. „NLC-CFA“ in Tabelle 6) sowie der CFA-MTMM Modelle (vgl. „CFA-MTMM“ in Tabelle 6) können hingegen entsprechend ihrer Berechnung (vgl. Kapitel 4.1) über alle drei Gruppen hinweg angegeben werden, da die zugrundeliegenden Faktorladungen mit metrischer Invarianz geschätzt wurden. Zu beachten ist zudem, dass die geschätzten Faktorladungen zwecks vereinfachter Interpretation mit 100 multipliziert wurden. Eine geschätzt Faktorladung von 0,6 beispielsweise entspricht in dieser Auswertung also einem Wert von 60.

Tabelle 6: Verteilung der *SV* der Faktorladungen

$SV_{\lambda_{iz}}$	Min.	Max.	Mittelwert	Standardabweichung	Schiefe	Kurtosis	% Missing
CFA Gruppe 1	0,07	17,17	2,44	2,05	1,93	8,20	4,44
CFA Gruppe 2	0,06	98,58	6,67	11,00	3,92	21,95	4,44
CFA Gruppe 3	0,03	22,39	1,57	1,27	3,52	30,22	4,44
NLC-CFA	0,01	97,52	2,46	5,64	6,22	56,00	4,44
CFA-MTMM	0,06	21,74	3,52	2,86	1,22	4,72	4,44

$SV_{\lambda_{iz}}$: Standard-Verzerrung der Faktorladungen je Gruppe; **CFA Gruppe 1**: *SV* der Faktorladungen des einfachen CFA Modells ohne Spezifikation eines FRE innerhalb von Gruppe 1, welche mit Assimilationseffekten zwischen allen aufeinanderfolgenden Kontext-Zielfragen-Dyaden simuliert wurde; **CFA Gruppe 2**: *SV* der Faktorladungen des einfachen CFA Modells ohne Spezifikation eines FRE innerhalb von Gruppe 2, welche mit Kontrasteffekten zwischen allen aufeinanderfolgenden Kontext-Zielfragen-Dyaden simuliert wurde; **CFA Gruppe 3**: *SV* der Faktorladungen des einfachen CFA Modells ohne Spezifikation eines FRE innerhalb von Gruppe 3, welche ohne FRE simuliert wurde; **NLC-CFA**: *SV* der Faktorladungen des NLC-CFA Modells über alle drei Gruppen hinweg, geschätzt mit metrischer Invarianz; **CFA-MTMM**: *SV* der Faktorladungen des CFA-MTMM Modells über alle drei Gruppen hinweg, geschätzt mit metrischer Invarianz; **Min**: geringster simulierter Wert der Variable über alle Iterationen; **Max**: höchster simulierter Wert der Variable über alle Iterationen; **% Missing**: Anteil an von der weiteren Analyse ausgeschlossenen Iterationen in Prozent

Bezüglich der möglichen Reichweite der *SV* ist festzuhalten, dass die unstandardisierten Faktorladungen λ inhaltlich als eine Approximation des relativen Anteils der Common Variance innerhalb des jeweiligen Indikators (mal 100) zu interpretieren sind, wodurch sowohl $\lambda > 100$, als auch $\lambda < 0$ theoriegeleitet keine plausiblen Schätzungen sind. Zum einen kann ein Anteil nicht mehr als 100% betragen, zum anderen kann ein Anteil nicht negativ sein. Die maximal mögliche, plausible *SV* kann demnach ebenfalls maximal 100 betragen, was in den Analysen durch einen Ausschluss von Iterationen mit *SV* > 100 gewährleistet wurde. Insgesamt wurden so 4,44% von 10000 Iterationen aus den Analysen ausgeschlossen, weil einzelne *SV* zu hoch waren, aber auch weil negative Varianzen geschätzt wurden oder die Modelle nicht konvergiert sind.

Werden entsprechend die maximal geschätzten *SV* betrachtet, so fallen vor allem die Faktorladungen der einfachen CFA in Gruppe 2 (max = 98,58) sowie der NLC-CFA (max = 97,52) auf, so dass hier von einer erhöhten Streuung ausgegangen werden kann. Daneben ist für die CFA Gruppe 2 (vgl. CFA Gruppe 2 mit $\overline{SV} = 6{,}67$) ebenfalls der höchste Mittelwert sowie die höchste Standardabweichung (11,00) zu beobachten, was sich derart interpretieren lässt, dass

für Schätzungen von Faktorladungen auf Basis von Daten, in welchen Kontrasteffekte vorhanden sind, welche jedoch nicht innerhalb des Modells spezifiziert werden, im Durchschnitt die höchsten Verzerrungen und ebenfalls die heterogensten Schätzungen zu erwarten sind.

Die geringste durchschnittliche Verzerrung ist zu erwarten, wenn keine FRE simuliert sind und diese auch nicht im Modell spezifiziert werden (vgl. CFA Gruppe 3 mit $\overline{SV} = 1{,}57$). Eine ebenfalls geringe durchschnittliche Verzerrung ist sowohl zu erwarten, wenn Assimilationseffekte zu erwarten sind, diese jedoch nicht im Modell spezifiziert werden (vgl. CFA Gruppe 1 mit $\overline{SV} = 2{,}44$), oder wenn Faktorladungen auf Basis des NLC-CFA Modells geschätzt werden (vgl. NLC-CFA mit $\overline{SV} = 2{,}46$), wobei die Differenz der beiden Werte statistisch nicht signifikant ist ($p = 0{,}7446$). Die durchschnittliche Verzerrung der Faktorladungen bei Schätzung mittels CFA-MTMM beträgt zuletzt 3,52. Alles in allem führen, so wie die Daten simuliert wurden, Assimilationseffekte im Durchschnitt zu relativ geringen Verzerrungen im Vergleich zu Kontrasteffekten, während das NLC-CFA Modell im Durchschnitt zu einer geringeren Verzerrung der Faktorladungen führt als das CFA-MTMM Modell. Dies ist bereits an dieser Stelle ein Indiz für die Überlegenheit der NLC-CFA gegenüber der CFA-MTMM bei der Bereinigung von verzerrten Faktorladungen.

Die Schiefe-Werte der *SV* der Faktorladungen zuletzt sind alle > 1, wobei die höchsten Werte für die *SV* in der 2. und 3. Gruppe der einfachen CFA ohne Spezifikation des FRE (Simulation mit Kontrasteffekt und Simulation ohne FRE) sowie der NLC-CFA anfallen. Die Daten sind entsprechend eindeutig rechtsschief verteilt und weichen folglich von einer Normalverteilung ab, worauf ebenfalls Kurtosiswerte zwischen 4,72 (CFA-MTMM) und 56,00 (NLC-CFA) hindeuten. Für die Regressionsanalysen der *SV* der Faktorladungen wird demnach ein Verfahren benötigt, welches auch für rechtsschiefe Verteilungen zu unverzerrten Schätzern führt.

Verteilungen der *SV* der geschätzten FRE Effekte

Über die Verzerrungen der Faktorladungen hinaus werden ebenfalls die *SV* der in der NLC-CFA geschätzten Regressionseffekte analysiert, welche den FRE und dessen Abhängigkeit von weiteren Moderatorvariablen (Zugänglichkeit von Informationen gegenüber dem Objekt der Befragung und Framing der Befragtenrolle) quantifizieren. Die Verteilungen der besagten Verzerrungen sind in Tabelle 7 aufgezeigt.

Tabelle 7: Verteilung der *SV* der NLC-CFA Effekte

$SV_{\beta_{ib}}$	Min.	Max.	Mittelwert	Standardab-weichung	Schiefe	Kurtosis	% Missing
$\boldsymbol{\beta_1}$	0,01	90,58	3,08	6,85	5,50	41,26	4,44
$\boldsymbol{\beta_2}$	0,03	6,54	0,77	0,58	2,28	11,75	4,44
$\boldsymbol{\beta_3}$	0,01	7,28	0,77	0,59	2,47	13,68	4,44
$\boldsymbol{\beta_4}$	0,05	21,81	2,00	1,72	3,67	28,19	4,44
$\boldsymbol{\beta_5}$	0,04	23,71	1,99	1,72	3,76	29,67	4,44
$\boldsymbol{\beta_6}$	0,02	10,97	0,97	0,83	2,63	14,71	4,44
$\boldsymbol{\beta_7}$	0,06	94,05	4,26	8,01	5,92	45,66	4,44

$\mathbf{SV_{\beta_{ib}}}$: Standard-Verzerrung der Regressionsschätzer der NLC-CFA; $\boldsymbol{\beta}$**:** Regressionsschätzer gemäß Tabelle 2; **Min**: geringster simulierter Wert der Variable über alle Iterationen; **Max**: höchster simulierter Wert der Variable über alle Iterationen; **% Missing**: Anteil an von der weiteren Analyse ausgeschlossenen Iterationen in Prozent

Werden zunächst Schiefe und Kurtosis betrachtet um zu überprüfen, ob die *SV* der Regressionseffekte ebenfalls die Normalverteilungsannahme verletzten, so lässt sich ein deutliches Urteil bilden. Die Schiefe liegt in einem Bereich von 2,28 (*SV* von β_2) bis 5,92 (*SV* von β_7), die Kurtosis liegt in einem Bereich von 11,75 (*SV* von β_2) bis 45,66 (*SV* von β_7), beides lässt auf eine Verletzung der Normalverteilungsannahme mit Rechtsschiefe schließen. Für die Regressionsanalysen der *SV* der Effektschätzer der NLC-CFA wird demnach ebenfalls ein Verfahren benötigt, welches für rechtsschiefe Verteilungen zu unverzerrten Schätzern führt.

Daneben sind die höchste durchschnittliche *SV*, die höchsten maximalen Werte der *SV*, die höchsten Standardabweichungen der *SV* und die höchsten Schiefe- und Kurtosiswerte der *SV* für die beiden Effektschätzer β_1 und β_7 abzulesen, welche den Haupteffekt der Interaktion (β_1) sowie die 3-Wege-Interaktion (β_7) darstellen (vgl. Tabelle 2). Die stabilsten Schätzungen qua dem geringsten maximalen Wert, dem geringsten durchschnittlichen Wert, der geringsten Standardabweichung und der geringsten Schiefe und Kurtosis der *SV* sind hingegen für β_2, β_3 und β_6 zu beobachten, für welche keine Effekte simuliert wurden und welche inhaltlich nicht den FRE selbst oder eine der relevanten Moderatoreinflüsse quantifizieren, sondern zwecks Kontrolle der relevanten Interaktionsschätzer spezifiziert werden (vgl. Kapitel 3.3.2). Die *SV* der Effekte von β_4 und β_5, welche die interaktiven Einflüsse der beiden Moderatorvariablen quantifizieren, verfügen zuletzt über moderate Mittelwerte, Standardabweichungen, Maximalwerte

sowie Schiefe- und Kurtosiswerte. Diese Zwischenbefunde werden in der Effekt Effizienz Analyse in Kapitel 4.5 vertiefend diskutiert.

Tabelle 8: Verteilung der *SV* Messfehlerkovarianzen

$SV_{\varepsilon_{iz}}$	Min.	Max.	Mittelwert	Standardabweichung	Schiefe	Kurtosis	% Missing
Gruppe 1	0,00	13,04	1,67	2,61	1,91	5,73	4,44
Gruppe 2	0,00	21,03	2,10	2,89	1,68	5,27	4,44
Gruppe 3	0,00	12,91	1,71	2,62	1,89	5,67	4,44

$\mathbf{SV}_{\varepsilon_{iz}}$: Standard-Verzerrung der Messfehlerkovarianz der CFA-MTMM je Gruppe; **Gruppe 1**: Messfehlerkovarianz innerhalb von Gruppe 1, welche mit Assimilationseffekten zwischen allen aufeinanderfolgenden Kontext-Zielfragen-Dyaden simuliert wurde; **Gruppe 2**: Messfehlerkovarianz innerhalb von Gruppe 2, welche mit Kontrasteffekten zwischen allen aufeinanderfolgenden Kontext-Zielfragen-Dyaden simuliert wurde; **Gruppe 3**: Messfehlerkovarianz innerhalb von Gruppe 3, welche ohne FRE simuliert wurde; **Min**: geringster simulierter Wert der Variable über alle Iterationen; **Max**: höchster simulierter Wert der Variable über alle Iterationen; **% Missing**: Anteil an von der weiteren Analyse ausgeschlossenen Iterationen in Prozent

In Tabelle 8 sind die Verteilungen der einzelnen, gruppenspezifischen *SV* der Messfehlerkovarianzen über alle Iterationen hinweg abgebildet.

Bezüglich der Überprüfung der Normalverteilungsannahme liegen die Kurtosiswerte der drei Gruppen zwischen 5,27 und 5,73 und die Schiefe zwischen 1,68 und 1,91, was erneut eine Verletzung der Normalverteilungsannahme indiziert. Es ist darüber hinaus interessant zu beobachten, dass die höchste durchschnittliche *SV* für Gruppe 2 beobachtbar ist, also in jener Gruppe, in welcher der FRE mit der höchsten Komplexität als 3-Wege-Interaktion simuliert wurde. Dieser vorläufige Befund wird ebenfalls vertiefend in Kapitel 4.5 diskutiert.

Verteilungen der Standardfehler der geschätzten FRE Effekte

Die inhaltliche Interpretation der Verteilung der Standardfehler der Effektschätzer zuletzt erfolgt vertiefend und im Kontext der tatsächlichen Effektstärken in Kapitel 4.6. Relevant ist an dieser Stelle entsprechend vor allem eine Überprüfung der Verletzung der Normalverteilungsannahme der Standardfehler auf Basis von Schiefe und Kurtosis, welche Tabelle 9 zu entnehmen sind. Diesbezüglich liegen die Werte der Schiefe zwischen 1,86 (β_2 und β_3) und 3,70 (β_1) und die Werte der Kurtosis zwischen 7,85 ($\varepsilon_{z=3}$) und 21,73 (β_1), so dass auch für die Analysen der Standardfehler ein Verfahren angebracht ist, welches Verletzungen der Normalverteilungsannahme berücksichtigt.

Tabelle 9: Verteilung der Standardfehler der Effektschätzer

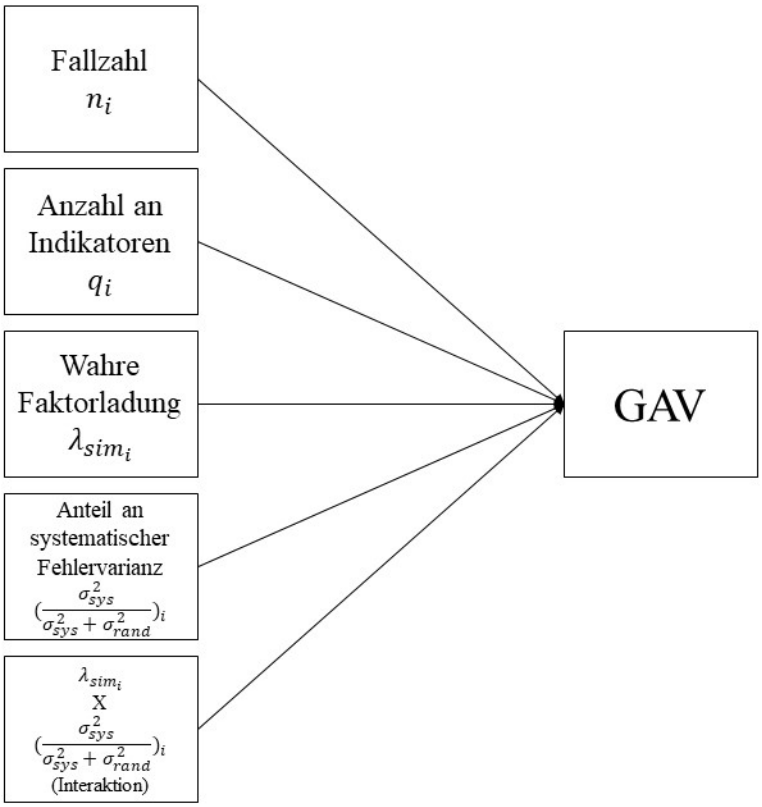
SE	Min.	Max.	Mittelwert	Standardabweichung	Schiefe	Kurtosis	% Missing
β_1	0,26	33,38	2,04	2,67	3,70	21,73	4,44
β_2	0,16	4,37	0,81	0,46	1,86	8,02	4,44
β_3	0,16	4,40	0,81	0,46	1,86	8,02	4,44
β_4	0,11	3,96	0,69	0,40	2,06	9,51	4,44
β_5	0,11	4,30	0,69	0,40	2,07	9,74	4,44
β_6	0,16	6,57	1,04	0,69	1,98	9,20	4,44
β_7	0,14	4,21	0,71	0,42	2,09	9,89	4,44
$\varepsilon_{z=1}$	0,00	2,30	0,25	0,24	1,98	9,69	4,44
$\varepsilon_{z=2}$	0,00	2,24	0,29	0,26	1,76	8,20	4,44
$\varepsilon_{z=3}$	0,01	2,90	0,45	0,37	1,74	7,85	4,44

SE: Standardfehler der Effektschätzer; ***β*:** Standardfehler der Regressionsschätzer gemäß Tabelle 2; ***ε*:** Standardfehler der Messfehlerkovarianz für Gruppe 1 (z = 1), Gruppe 2 (z = 2) und Gruppe 3 (z = 3); **Min**: geringster simulierter Wert der Variable über alle Iterationen; **Max**: höchster simulierter Wert der Variable über alle Iterationen; **% Missing**: Anteil an von der weiteren Analyse ausgeschlossenen Iterationen in Prozent

Entscheidung hinsichtlich eines geeigneten Regressionsverfahrens

Aufgrund dieser Analysen der Schiefe- und Kurtosiswerte der jeweils zu analysierenden, abhängigen Variablen ist die Notwendigkeit der Verwendung eines Verfahrens, welches Verletzungen der Normalverteilungsannahme berücksichtigt, ersichtlich. Ein solches Verfahren ist in SEM bei Verwendung des robusten ML Schätzalgorithmus gegeben, wie in Kapitel 3.1.2 diskutiert. Um entsprechend weiterhin innerhalb des bereits eingeführten und ausführlich diskutierten SEM Frameworks zu arbeiten und um die einzelnen Analysen hinsichtlich ihrer Methodik miteinander vergleichbar zu machen, wird für alle folgenden Regressionsanalysen der Verzerrungen der Faktorladungen, der Verzerrungen der Effektkoeffizienten und der Standardfehler SEM und der MLR Schätzer verwendet. Für alle Analysen wird dabei das gleiche Modell mit den gleichen unabhängigen Variablen und den gleichen Effekten spezifiziert, wobei sich zwischen den Analysen jeweils lediglich die abhängige Variable unterscheidet. Das entsprechende Modell wird in Abbildung 21 mit einer **g**enerischen **a**bhängigen **V**ariablen „GAV“ dargestellt.

Abbildung 21: Generisches Regressionsmodell



GAV: Generische abhängige Variable, entspricht den abhängigen Variablen aus Tabelle 5, Kapitel 4.1

4.4 Analyse der Verzerrungen von Faktorladungen

Beginnend mit der Analyse der Verzerrungen der Faktorladungen ist zunächst festzuhalten, dass sowohl die Verzerrungen für Modelle ohne Kontrolle von FRE (einfaches CFA Multigruppenmodell), als auch die Verzerrungen für Modelle mit Kontrolle von FRE (CFA-MTMM Multigruppenmodell und NLC-CFA Multigruppenmodell) analysiert werden, um die Höhe der Standard-Verzerrungen (*SV*, vgl. Kapitel 4.1 und 4.2) zwischen den Modellen zu vergleichen. Dies soll Aufschluss darüber geben, ob Faktorladungen bei Anwesenheit von FRE eine geringere Verzerrung erfahren, wenn innerhalb des CFA Modells eine Kontrolle besagter FRE stattfindet sowie, welches der beiden Kontrollverfahren (CFA-MTMM oder NLC-CFA) zu einer höheren Bereinigung der Verzerrung führt. Die Abhängige Variable der Modellevaluation stellt in allen Analysen die *SV* der Faktorladung dar, wobei das einfache CFA Multigruppenmodell ohne metrische Invarianz und das CFA-MTMM Multigruppenmodell sowie das NLC-CFA Multigruppenmodell mit metrischer Invarianz geschätzt wurde. Für das einfache CFA Modell

liegt entsprechend eine *SV* je Gruppe vor, da unterschiedliche Faktorladungen je Gruppe geschätzt wurden, während sowohl für das CFA-MTMM Modell, als auch für das NLC-CFA Modell je eine *SV* über alle Gruppen hinweg vorliegt, da die Faktorladungen hier für alle Gruppen identisch sind. Die erwähnten Gruppen *z* unterscheiden sich dabei gemäß der Beschreibung in Kapitel 4.1 und 4.2 in eine Gruppe mit simuliertem Assimilationseffekt (Gruppe 1), eine Gruppe mit simuliertem Kontrasteffekt (Gruppe 2) und in eine Gruppe ohne simulierten FRE (Gruppe 3). Für die definierten Modelle sind unstandardisierte Faktorladungen als Approximation des relativen Anteils der Common Variance zu interpretieren. Die *SV* hat einen positiven Wertebereich und ist als Wurzel der durchschnittlichen Verzerrungsquadrate über alle Faktorladungen innerhalb einer Gruppe Mal 100 zu interpretieren. Wäre die Wurzel der durchschnittlichen quadrierten Verzerrung der Faktorladungen beispielsweise 0,05, so wäre in der Auswertung eine *SV* von 5,00 angezeigt. Dabei gilt, dass je größer die *SV*, desto stärker die durchschnittliche Verzerrung, während eine *SV* gleich 0 einer unverzerrten Schätzung entspricht. Die abhängige *SV* wird für die Verzerrungsanalyse zwecks Evaluation der verschiedenen Modelle auf verschiedene unabhängige Variablen mittels SEM-basierter Regression und MLR Schätzer zurückgeführt. Als unabhängige Variablen innerhalb der Verzerrungsanalyse sind, wie ebenfalls zuvor beschrieben, n (Fallzahl), q (Anzahl an Indikatoren des zu schätzenden latenten Konstruktes), λ_{sim} (wahre Faktorladung, interpretierbar als approximierter Anteil an Common Variance bzw. approximierte Faktorladung), $\frac{\sigma_{sys}^2}{\sigma_{sys}^2+\sigma_{rand}^2}$ (Anteil des systematischen Messfehlers innerhalb des gesamten Messfehlers) sowie das Produkt von λ_{sim} und $\frac{\sigma_{sys}^2}{\sigma_{sys}^2+\sigma_{rand}^2}$ definiert. Die unabhängigen Variablen sind diesbezüglich durch Zentrierung derart rekodiert, dass die Konstante *a* innerhalb aller Regressionen als durchschnittlich geschätzte *SV* für ein Modell mit einer Fallzahl von 1500 und fünf Indikatorvariablen mit einer wahren Faktorladung von 0,6 und einem wahren Anteil des systematischen Messfehlers am gesamten Messfehler von 50% innerhalb der Indikatoren interpretiert werden kann, was im Folgenden als „Benchmark“ Spezifikation bzw. Modell bezeichnet wird. Darüber hinaus impliziert der Regressionskoeffizient für den Effekt von n die Veränderung der *SV* für zusätzliche 1000 Befragte. λ_{sim} und $\frac{\sigma_{sys}^2}{\sigma_{sys}^2+\sigma_{rand}^2}$ wurden derart rekodiert, dass der Regressionskoeffizient als Veränderung der *SV* für jeden zusätzlichen Prozentpunkt des (approximierten) Anteils der Common Variance an der gesamten Varianz bzw. des Anteils des systematischen Messfehlers am gesamten Messfehler interpretiert werden kann. Positive Effekte implizieren demnach eine Erhöhung der *SV* und eine geringere Effizienz,

während negative Effekte eine Verringerung der *SV* und damit eine erhöhte Effizienz der Schätzung implizieren. Die zugrundeliegenden geschätzten CFA Modelle (aus welchen sich die zur Analyse stehenden Faktorladungen ergeben) mit fünf Indikatorvariablen sind (vereinfacht) in Abbildung 18, Abbildung 19 und Abbildung 20 in Kapitel 4.1aufgezeigt.

Verzerrungen der Faktorladungen ohne Kontrolle der FRE

Aus Tabelle 10, welche die *SV* der Faktorladungen für die einfachen CFA Modelle ohne Spezifikation des FRE aufzeigen, ist die *SV* des Benchmark Modells je Gruppe aus Konstante *a* abzulesen. Für Gruppe 1 (FRE als Assimilationseffekt auf Basis von zwei separaten 2-Wege-Interaktionen simuliert) beträgt diese 2,721 ($p < 0{,}001$), was einer geringen, jedoch vorhandenen Verzerrung entspricht. So würde für das Benchmark Modell beispielsweise statt einer Faktorladung von 0,6 im Durchschnitt und vereinfacht (da es sich bei der *SV* nicht um die tatsächliche Verzerrung, sondern um die Wurzel der durchschnittlichen, quadrierten Verzerrung handelt) eine Faktorladung von 0,627 geschätzt werden. Unter Berücksichtigung der standardisierten Regressionskoeffizienten *b** können weiterhin die verschiedenen Faktoren gemäß ihrer Einflussstärke auf die Verzerrungen der Faktorladungen verglichen werden.

Den stärksten Einfluss auf die *SV* hat der (approximierte Anteil) an Common Variance mit einem standardisierten Regressionskoeffizienten von $b^* = -0{,}512$. Mit jedem zusätzlichen Prozentpunkt des Anteils an Common Variance reduziert sich die *SV* um 0,061. Dies bedeutet, dass je geringer der Anteil an Common Variance bei einem ansonsten konstanten Anteil des systematischen Messfehlers am gesamten Messfehler (was ebenfalls einem stärkeren FRE entspricht), desto größer die *SV* der Faktorladungen. Dabei ist jedoch zu beachten, dass es sich hierbei absolut betrachtet um keinen sonderlich starken Effekt handelt. So bräuchte es rechnerisch eine Erhöhung des (approximierten) Anteils der Common Variance um 45 Prozentpunkte (also beispielsweise eine Erhöhung von 50% auf 95%), um die absolute *SV* der Faktorladung des Benchmark Modells von 0,027 zu bereinigen.

Der Anteil systematischer Fehlervarianz an der gesamten Fehlervarianz wiederum hat trotz eines erhöhten *b** ebenfalls einen absolut betrachtet geringen, jedoch positiven Effekt auf die Verzerrung der Faktorladungen mit $b^* = 0{,}497$. So nimmt die *SV* der Faktorladungen mit jedem zusätzlichen Prozentpunkt-Anteil der systematischen Fehlervarianz an der gesamten Fehlervarianz bei einer ansonsten konstanten wahren Faktorladung (was einem stärkeren FRE entspricht, vgl. Kapitel 4.1 und 4.2) um einen Wert 0,045 zu. Beide Effekte weißen entsprechend darauf hin, dass Faktorladungen in Gruppe 1 umso stärker Verzerrt sind, desto stärker der FRE simuliert wurde. Verringert sich der Anteil an Common Variance in gleichem Maße wie sich

der Anteil der systematischen Fehlervarianz erhöht (was zu einem negativen Produkt von λ_{sis} und $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ führt), so erhöht sich der *SV* ebenfalls, was durch den negativen Interaktionseffekt impliziert wird, da ein negatives Produkt mal einen negativen Koeffizienten zu einer positiven Schätzung der *SV* führt. Die Analyse zeigt somit insgesamt, dass je stärker der FRE simuliert wurde, desto höher die *SV* und damit die zu erwartende Verzerrung ausfällt. Die konkreten Werte der unstandardisierten Regressionskoeffizienten weißen daneben aber ebenfalls auf einen absolut betrachtet geringen Einfluss hin.

Tabelle 10: Verzerrungen der Faktorladungen ohne Kontrolle von Fragereihenfolgeneffekten

	Gruppe 1			**Gruppe 2**			**Gruppe 3**		
	b	SE	b*	b	SE	b*	b	SE	b*
$\boldsymbol{a}$	2,721***	0,024	NA	9,729***	0,186	NA	1,967***	0,021	NA
$\boldsymbol{n}$	-0,239***	0,01	-0,163	-0,011	0,055	-0,001	-0,44***	0,01	-0,482
$\boldsymbol{q = 3}$	0,105***	0,03	0,024	-5,568***	0,208	-0,239	0,179***	0,027	0,066
$\boldsymbol{q = 4}$	-0,150***	0,027	-0,035	-3,542***	0,209	-0,152	0,056*	0,023	0,021
$\boldsymbol{\lambda_{sis}}$	-0,061***	0,001	-0,512	-0,287***	0,005	-0,446	-0,024***	0,001	-0,32
$\frac{\boldsymbol{\sigma^2_{sys}}}{\boldsymbol{\sigma^2_{sys}+\sigma^2_{rand}}}$	0,045***	0,001	0,497	0,22***	0,004	0,448	<0,001	<0,001	-0,006
$\boldsymbol{int}$	-0,002***	<0,001	-0,376	-0,011***	<0,001	-0,383	<0,001	<0,001	-0,019

Gruppe 1: Gruppe mit simuliertem, aber im Modell nicht spezifizierten Assimilationseffekt; **Gruppe 2**: Gruppe mit simuliertem, aber im Modell nicht spezifiziertem Kontrasteffekt; **Gruppe 3**: Gruppe ohne simulierten und ohne spezifizierten FRE; **a**: Konstante der Regressionsschätzung; **n**: Effekt der Fallzahl auf die *SV* der Faktorladungen; **q**: Einfluss der Anzahl an Indikatoren innerhalb des geschätzten Modells auf die *SV* der Faktorladungen; $\boldsymbol{\lambda_{sis}}$: Einfluss der wahren Faktorladung bzw. des Anteils an Common Variance auf die *SV* der geschätzten Faktorladungen; $\frac{\boldsymbol{\sigma^2_{sys}}}{\boldsymbol{\sigma^2_{sys}+\sigma^2_{rand}}}$**:** Einfluss des Anteils des systematischen Messfehlers am gesamten Messfehler auf die *SV* der Faktorladungen; **int**: Einfluss der Interaktion von λ_{sis} und $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ auf die *SV* der Faktorladungen;

b***: p ≤ 0,001; b**: p ≤ 0,01; b*: p ≤ 0,05

a für n = 1500; p = 5; λ_{sis} = 0,6; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ = 0,5

Neben den Einflüssen des Anteils der Common Variance und der systematischen Fehlervarianz weißen (für Gruppe 1) die Effekte der Anzahl an Indikatorvariablen *q* darauf hin, dass die Faktorladungen eines ansonsten mit dem Benchmark Modell identischen Modelles mit lediglich

drei Indikatorvariablen stärker verzerrt sind (b = 0,105), während ein Modell mit vier Indikatoren eine geringere *SV* aufweisen (b = -0,150). Wird jedoch berücksichtigt, dass die Faktorladung zwecks Vereinfachter Interpretation mit 100 multipliziert wurde, so beträgt die absolute *SV* lediglich 0,001 bzw. -0,002 für eine wahre Faktorladung von 0,6. Absolut betrachtet ist somit auch der Einfluss der Modellkomplexität äußerst gering. Die Fallzahl zuletzt hat einen negativen Effekt auf die *SV* mit b^* = -0,163 und b = -0,239. Es lässt sich schließen, dass je höher die Fallzahl, desto geringer die *SV*, auch wenn dieser Effekt absolut betrachtet ebenfalls als eher gering betrachtet werden kann.

Für Gruppe 1 kann somit geschlossen werden, dass auch wenn die Verzerrung der Faktorladungen absolut betrachtet lediglich gering ausfallen, diese dennoch vorhanden sind und systematisch durch den Anteil an Common Variance und den Anteil an systematischer Fehlervarianz bzw. die Stärke des simulierten FRE sowie die Modellkomplexität und die Fallzahl beeinflusst werden. Ein höherer, simulierter FRE führt zu einer erhöhten Verzerrung der Faktorladungen, während eine höhere Fallzahl zu einer geringeren Verzerrung führt. Darüber hinaus sind Faktorladungen mit q = 3 Indikatoren stärker und Modelle mit q = 4 Indikatoren geringer verzerrt als Modelle mit q = 5 Indikatoren. Der relativ geringe, absolute Einfluss der einzelnen Prädikatoren auf die absolute Höhe der *SV* impliziert daneben jedoch eine relativ konsistente Schätzung der Faktorladungen innerhalb der simulierten Daten.

Für Gruppe 2, innerhalb welcher ein Kontrasteffekt als 3-Wege-Interaktion simuliert wurde, impliziert Konstante a für Benchmark Modelle eine *SV* der Faktorladungen von 9,729, was einer erhöhten Verzerrung entspricht. Der Effekt des (approximierten) Anteils an Common Variance (b^* = -0,446) ist erneut negativ und lässt sich derart interpretieren, dass die *SV* um 0,287 zunimmt, wenn der (approximierte) Anteil an Common Variance um 1% abnimmt, was absolut betrachtet einen erhöhten Einfluss repräsentiert. Der Effekt des Anteils des systematischen Messfehlers (b^* = 0,448) lässt sich derart interpretieren, dass die *SV* um 0,220 zunimmt, wenn der Anteil des systematischen Messfehlers um 1% zunimmt, was absolut betrachtet ebenfalls einem erhöhten Einfluss entspricht. Der negative Interaktionseffekt (b^* = -0,383) lässt sich ebenfalls erneut analog zu Gruppe 1 interpretieren. Faktorladungen sind demnach umso stärker verzerrt, desto stärker der simulierte FRE ausfällt.

Daneben hat die Modellkomplexität ebenfalls einen systematischen Einfluss auf Verzerrungen der Faktorladungen mit b = -3,542 für q = 4 Modelle und b = -5,568 für q = 3 Modelle. Faktorladungen in Gruppe 2 sind somit am geringsten Verzerrt, wenn das Modell mit drei statt fünf

Indikatorvariablen geschätzt wird, während die Verwendung von vier statt fünf Indikatorvariablen die Verzerrung ebenfalls reduziert. Absolut betrachtet entspricht dies ebenfalls einem relevanten Einfluss. Die Fallzahl zuletzt hat keinen statistisch signifikanten Einfluss auf die *SV*.

Für Gruppe 2 bedeutet dies entsprechend, dass ein systematischer Einfluss der Stärke des FRE und der Modellkomplexität auf die *SV* der Faktorladungen vorliegt, welcher hinsichtlich der absoluten Höhe der Verzerrungen keinesfalls als trivial zu bewerten ist. Desto stärker der FRE simuliert ist, desto stärker die Verzerrung und desto weniger Indikatoren an der Schätzung beteiligt sind, desto geringer die Verzerrung.

Für Gruppe 3 der CFA, welche ohne FRE simuliert wurde, zeigt sich eine *SV* der Faktorladungen für das Benchmark Modell von 1,967, was insgesamt als geringe Verzerrung angesehen werden kann. Der Anteil der systematischen Fehlervarianz und die Interaktion des Anteils an Fehlervarianz mit dem Anteil an Common Variance haben keinen signifikanten Effekt auf die *SV* der Faktorladungen was insofern plausibel ist, als dass in Gruppe 3 kein FRE als systematische Fehlervarianz simuliert wurde. Daneben zeigt sich ein negativer und statistisch signifikanter Effekt des Anteils an Common Variance von b = -0,024 (b^* = -0,320) was entsprechend derart zu interpretieren ist, dass je höher die wahre Faktorladung, desto geringer die Verzerrung der Faktorladung, auch wenn es sich hierbei absolut betrachtet in der Tendenz um einen schwachen Effekt handelt. Auch die Fallzahl hat einen negativen Effekt von b = -0,440 (b^* = -0,482). Je höher die Fallzahl, desto geringer ist die Verzerrung, wobei es sich hierbei absolut betrachtet um einen bedeutsamen Effekt handelt. Modelle mit drei statt fünf Indikatorvariablen haben eine erhöhte *SV* mit b = 0,179 (b^* = 0,066), während die *SV* von Modelle mit vier Indikatorvariablen ebenfalls, jedoch in geringerem Maße mit b = 0,056 (b^* = 0,0.21) erhöht ist, wobei beide Effekte absolut betrachtet als vernachlässigbar gering interpretiert werden können.

Faktorladungen können somit, so keine FRE simuliert wurden, insgesamt effizient geschätzt werden. Sind hingegen FRE vorhanden und werden diese nicht in der Modellspezifikation berücksichtigt, so können moderate bis erhebliche Verzerrungen resultieren.

Verzerrungen der Faktorladungen bei gegebener Kontrolle der FRE

Dem gegenüber sind in Tabelle 11 die zu erwartenden *SV* entsprechend der jeweiligen Simulationsparameter für das CFA-MTMM und das NLC-CFA Modell aufgezeigt, innerhalb welchen gemäß der Beschreibung in Kapitel 4.1 und 3.3 FRE kontrolliert werden. Konstante *a* zeigt diesbezüglich erneut die durchschnittlich zu erwartende *SV* der Faktorladungen für Benchmark

Modelle, welche 3,210 für das CFA-MTMM Modell und 1,495 für das NLC-CFA Modell beträgt. Bei gegebener Kontrolle der FRE werden die Faktorladungen demnach für beide Modelle lediglich mit geringer Verzerrung geschätzt, wobei die Verzerrung durch die NLC-CFA für Benchmark Modelle statistisch signifikant ($p \leq 0{,}001$) geringer ist als für die CFA-MTMM. Es lässt sich somit auf Basis der geschätzten Parameter eindeutig schließen, dass die NLC-CFA für Benchmark Modelle zu einer höheren Bereinigung der Verzerrungen durch FRE führt und Faktorladungen somit effizienter geschätzt werden können, als dies bei Spezifikation eines CFA-MTMM Modells der Fall ist.

Tabelle 11: Verzerrungen der Faktorladungen bei gegebener Kontrolle von Fragereihenfolgeneffekten

	CFA-MTMM			**NLC-CFA**		
	b	SE	b*	b	SE	b*
a	3,21***	0,03	NA	1,495***	0,055	NA
n	-0,103***	0,013	-0,05	-0,481***	0,04	-0,119
$q = 3$	1,458***	0,045	0,241	4,217***	0,155	0,353
$q = 4$	-0,181***	0,034	-0,03	0,024	0,047	0,002
λ_{sis}	-0,062***	0,001	-0,373	-0,031***	0,003	-0,094
$\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$	0,066***	0,001	0,517	-0,064***	0,003	-0,254
int	-0,003***	<0,001	-0,424	0,001***	<0,001	0,048

CFA-MTMM: Schätzung auf Basis des CFA-MTMM Modells gemäß Kapitel 3.3.1 und 4.1; **NLC-CFA**: Schätzung auf Basis des NLC-CFA Modells gemäß Kapitel 3.3.2 und 4.1; **n**: Effekt der Fallzahl auf die *SV* der Faktorladungen; **q**: Einfluss der Anzahl an Indikatoren innerhalb des geschätzten Modells auf die *SV* der Faktorladungen; λ_{sis}: Einfluss der wahren Faktorladung bzw. des Anteils an Common Variance auf die *SV* der geschätzten Faktorladungen; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$**:** Einfluss des Anteils des systematischen Messfehlers am gesamten Messfehler auf die *SV* der Faktorladungen; **int**: Einfluss der Interaktion von λ_{sis} und $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ auf die *SV* der Faktorladungen;

b***: $p \leq 0{,}001$; b**: $p \leq 0{,}01$; b*: $p \leq 0{,}05$

a für $n = 1500$; $p = 5$; $\lambda_{sis} = 0{,}6$; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}} = 0{,}5$

Bei Kontrolle von FRE mittels NLC-CFA hat der Anteil der systematischen Fehlervarianz an der gesamten Fehlervarianz bezogen auf die absolute Verzerrung einen geringen und negativen Effekt auf die *SV* der Faktorladungen mit $b = -0{,}064$ ($b^* = -0{,}254$). Gleiches gilt für den (approximierten) Anteil an Common Variance, wobei der Effekt mit $b = -0{,}031$ ($b^* = -0{,}094$) noch

geringer und nahezu vernachlässigbar ist. Der Effekt der Interaktion zwischen den beiden Variablen ist noch geringer, jedoch positiv mit $b = 0{,}001$ ($b^* = 0{,}048$). Bei gegebener Kontrolle von vorhandenen FRE mittels NLC-CFA verringert somit sowohl ein erhöhter Anteil an Common Variance, als auch ein erhöhter Anteil der systematischen Fehlervarianz die *SV*. Nimmt hingegen der Anteil an Common Variance ab und der Anteil an systematischer Fehlervarianz zu, was in einem negativen Produkt aus den beiden Variablen resultiert und einem stärkeren, simulierten FRE entsprich, so verringert sich der *SV* ebenfalls minimal, was durch den positiven Interaktionseffekt indiziert wird (ein negatives Produkt mal einem positiven Koeffizienten resultiert in einer negativen Schätzung der *SV*). Gleichzeitig sind die Effekte absolut betrachtet sehr schwach, so dass selbst wenn ein systematischer Einfluss besteht, die eigentliche Verzerrung der Faktorladungen lediglich gering ausfällt. Beispielsweise müsste der (approximierte) Anteil an Common Variance um 100 Prozentpunkte sinken (also von Maximalwert auf den Minimalwert der Skala), damit die durchschnittliche, quadrierte Verzerrung der Faktorladung um 0,03 zunimmt.

Die Fallzahl hat bei gegebener Kontrolle des FRE mittels NLC-CFA einen statistisch signifikanten Einfluss auf die *SV* von Faktorladungen, wie ein b von -0,481 ($b^* = -0{,}119$) aufzeigt. Erhöht sich die Fallzahl, so verringert sich die *SV*. Absolut betrachtet handelt es sich dabei um einen nennenswerten Einfluss, da rechnerisch lediglich etwa 3000 zusätzliche Befragte benötigt werden, um die gesamte, erwartete *SV* für NLC-CFA Benchmark Modelle zu bereinigen. Die Modellkomplexität zuletzt, indiziert durch die Anzahl an Indikatorvariablen q, hat bei drei statt fünf Indikatorvariablen ebenfalls einen Einfluss auf die Verzerrung von Faktorladungen mit $b = 4{,}217$ ($b^* = 0{,}353$), während der Einfluss von $q = 4$ Modellen statistisch nicht signifikant ist. Modelle, welche mit drei statt fünf Indikatoren geschätzt wurden verfügen somit über eine erhöhte und absolut betrachtet erhebliche Verzerrung, während die Verzerrung bei Modellen mit vier Indikatoren mit den Verzerrungen von Modellen mit fünf Indikatoren vergleichbar ist. Es lässt sich somit festhalten, dass Faktorladungen in NLC-CFA Modellen lediglich sehr gering verzerrt sind und dass die Höhe der Verzerrung in nennenswertem Umfang vor allem durch die Modellkomplexität und die Fallzahl systematisch beeinflusst wird.

Bei gegebener Kontrolle von FRE durch CFA-MTMM Modelle hat der Anteil der systematischen Fehlervarianz an der gesamten Fehlervarianz den stärksten Einfluss auf die *SV* der Faktorladungen mit $b = 0{,}066$ ($b^* = 0{,}517$), auch wenn der Einfluss absolut betrachtet erneut eher gering ausfällt (nimmt der Anteil der systematischen Fehlervarianz an der gesamten Fehlervarianz um 100 Prozentpunkte zu, erhöht sich die durchschnittliche quadrierte Verzerrung der Faktorladung um 0,07). Gleiches gilt für den (approximierten) Anteil an Common Variance mit

b = -0,062 (b^* = -0,373) sowie für die Interaktion zwischen den beiden Variablen mit einem Effekt von b = -0,003 (b^* = -0,424), wobei die beiden zuletzt genannten Effekte negativ sind. In CFA-MTMM Modellen verringert sich somit die *SV* bei Erhöhung des Anteils an Common Variance oder bei Verringerung des Anteils des systematischen Messfehlers, was beides ebenfalls Verringerung des absoluten FRE impliziert. Nehmen der Anteil an Common Variance ab und der Anteil an systematischer Fehlervarianz zu, was in einem negativen Produkt aus den beiden Variablen resultiert und einem stärkeren, simulierten FRE entsprich, so erhöht sich die *SV*, was durch den negativen Interaktionseffekt indiziert wird (ein negatives Produkt mal einem negativen Koeffizienten resultiert in einer positiven Schätzung der *SV*). Ein stärkerer FRE resultiert somit insgesamt in einer stärkeren Verzerrung, auch wenn die Effekte für die CFA-MTMM insgesamt gering sind, so dass selbst wenn ein systematischer Einfluss besteht, die eigentliche Verzerrung der Faktorladungen erneut lediglich gering ausfällt.

Die Fallzahl hat für CFA-MTMM Modelle einen statistisch signifikanten, jedoch absolut betrachtet sehr geringen Einfluss auf die *SV* von Faktorladungen von -0,103 (b^* = -0,050). Erhöht sich also die Fallzahl, dann verringert sich die *SV*. Die Modellkomplexität hat bei drei statt fünf Indikatorvariablen einen Einfluss auf die Verzerrung von Faktorladungen mit b = 1,458 (b^* = 0,241) und bei vier statt fünf Indikatorvariablen mit b = -0,181 (b^* = -0,030). Modelle welche mit drei statt fünf Indikatoren geschätzt wurden verfügen somit über eine erhöhte Verzerrung, während die Verzerrung bei Modellen mit vier Indikatoren minimal geringer ist als bei Modellen mit fünf Indikatoren. Es lässt sich somit auch für CFA-MTMM Modelle festhalten, dass Faktorladungen lediglich gering verzerrt sind und die Höhe der Verzerrung im Wesentlichen durch die Modellkomplexität beeinflusst wird.

Vergleich der Modelle

Die Ergebnisse der Analysen der CFA Modelle ohne Kontrolle des FRE mit den Modellen mit Kontrolle des FRE vergleichend (vgl. Tabelle 12) zeigt sich, dass die zu erwartende *SV* der Faktorladungen innerhalb des Benchmark Modells am geringsten ausfällt, wenn FRE mittels NLC-CFA kontrolliert werden. Die durchschnittliche Verzerrung ist hier sogar geringer als bei Schätzung eines einfachen CFA Modells, wenn keine FRE simuliert wurden (vgl. Gruppe 3 in Tabelle 12). Für Benchmark Modelle resultiert das NLC-CFA Modell somit in den effizientesten Schätzungen. Demgegenüber führt die Schätzung mittels CFA-MTMM zu geringeren Verzerrungen, als wenn Kontrasteffekte vorhanden sind, diese jedoch nicht kontrolliert werden (vgl. Gruppe 2 in Tabelle 12). Hingegen ist das CFA-MTMM Modell weniger Effizient und führt zu mehr Verzerrungen, als wenn Assimilationseffekte vorhanden sind und diese nicht

kontrolliert werden (vgl. Gruppe 1 in Tabelle 12). Die Effizienz des CFA-MTMM Modells ist somit Abhängig von der Art der zugrundeliegenden Verzerrung, während die NLC-CFA über alle Verzerrungen hinweg die effizientesten Schätzer produziert.

Tabelle 12: Gegenüberstellung der Faktorladungsverzerrung (unstand. Regressionskoeffizienten *b*)

	CFA-MTMM	**NLC-CFA**	**Gruppe 1**	**Gruppe 2**	**Gruppe 3**
a	3,21***	1,495***	2,721***	9,729***	1,967***
n	-0,103***	-0,481***	-0,239***	-0,011	-0,44***
$q = 3$	1,458***	4,217***	0,105***	-5,568***	0,179***
$q = 4$	-0,181***	0,024***	-0,15***	-3,542***	0,056*
λ_η	-0,062***	-0,031***	-0,061***	-0,287***	-0,024***
$\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$	0,066***	-0,064***	0,045***	0,22***	<0,001
int	-0,003***	0,001***	-0,002***	-0,011***	<0,001

CFA-MTMM: Schätzung auf Basis des CFA-MTMM Modells gemäß Kapitel 3.3.1 und 4.1; **NLC-CFA**: Schätzung auf Basis des NLC-CFA Modells gemäß Kapitel 3.3.2 und 4.1; **Gruppe 1**: Gruppe mit simuliertem, aber im Modell nicht spezifizierten Assimilationseffekt; **Gruppe 2**: Gruppe mit simuliertem, aber im Modell nicht spezifiziertem Kontrasteffekt; **Gruppe 3**: Gruppe ohne simulierten und ohne spezifizierten FRE; **a**: Konstante der Regressionsschätzung; **n**: Effekt der Fallzahl auf die *SV* der Faktorladungen; **q**: Einfluss der Anzahl an Indikatoren innerhalb des geschätzten Modells auf die *SV* der Faktorladungen; λ_{sis}: Einfluss der wahren Faktorladung bzw. des Anteils an Common Variance auf die *SV* der geschätzten Faktorladungen; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$**:** Einfluss des Anteils des systematischen Messfehlers am gesamten Messfehler auf die *SV* der Faktorladungen; **int**: Einfluss der Interaktion von λ_{sis} und $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ auf die *SV* der Faktorladungen; b***: $p \leq 0{,}001$; b**: $p \leq 0{,}01$; b*: $p \leq 0{,}05$

a für $n = 1500$; $p = 5$; $\lambda_{sis} = 0{,}6$; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}} = 0{,}5$

Darüber hinaus weißen sowohl der (approximierte) Anteil der Common Variance, der Anteil des systematischen Messfehlers am gesamten Messfehler, als auch die Interaktion der beiden Variablen (welche in Kombination den Einfluss der Stärke des FRE auf die durchschnittlich zu erwartende *SV* der Faktorladungen quantifizieren) innerhalb der Gruppen, in welchen FRE simuliert, jedoch nicht kontrolliert wurden (vgl. Gruppe 1 und Gruppe 2 in Tabelle 12) einen signifikanten Einfluss auf die Verzerrungen der Faktorladungen in dem Sinne auf, dass je stärker der FRE simuliert ist, desto höher die *SV* und damit die Verzerrung der Faktorladungen. Gleiches ist bei gegebener Kontrolle der FRE durch das CFA-MTMM zu beobachten. Werden

FRE hingegen durch das NLC-CFA Modell kontrolliert, so reduziert sich die Verzerrung insgesamt bei höheren, wahren Faktorladungen, bei höheren Anteilen des systematischen Messfehlers und bei stärkeren FRE. Absolut betrachtet ist der Einfluss der Höhe der simulierten Varianzanteile bzw. der Höhe des simulierten FRE außer für die einfache CFA in Gruppe 2 jedoch insgesamt relativ gering. Es scheint entsprechend so, dass stärkere FRE bzw. höhere Varianzanteile innerhalb des NLC-CFA Modells durch eindeutige Spezifikation der einzelnen Effektpfade klarer voneinander differenziert werden können, was insgesamt zu effizienteren Schätzungen führt.

Aus dieser Differenz der durchschnittlichen Verzerrung der Faktorladungen der Benchmark Modelle, abzulesen in Konstante *a*, sowie dem Befund, dass stärkere FRE in Modellen ohne Kontrolle und für das CFA-MTMM Modell zu einer höheren Verzerrung führen, während höhere Varianzanteile und stärkere FRE in Modellen mit Kontrolle der FRE mittels NLC-CFA zu einer verringerten Verzerrungen der Faktorladungen führen, lässt sich insgesamt ableiten, dass FRE dort Faktorladungen verzerren, wo diese nicht kontrolliert werden, während Verzerrungen durch Kontrolle von FRE mittels NLC-CFA am effizientesten bereinigt werden können und die Effizienz des CFA-MTMM Modells abhängig ist von der Art der zugrundeliegenden Verzerrung. Dabei bleibt zu beachten, dass die Einflüsse der unterschiedlichen Varianzanteile auf die Höhe der Verzerrung (außer für die einfache CFA in Gruppe 2) sehr gering sind, so dass die Höhe der Verzerrungen auch bei unterschiedlichen Simulationsspezifikationen relativ stabil ist.

Weitere Befunde des Vergleichs der jeweiligen Modelle beziehen sich zum einen auf die Fallzahl und zum anderen auf die Modellkomplexität. So ist der Einfluss der Fallzahl, mit Ausnahme für Gruppe 2, innerhalb welcher der Kontrasteffekt als 3-Wege-Interaktion simuliert wurde, in allen anderen Gruppen und Modellen statistisch signifikant und negativ. Die Schätzung wird somit umso effizienter, desto höher die Fallzahl, wobei dieser Einfluss absolut betrachtet vor allem für die Schätzung mittels NLC-CFA und CFA-MTMM, jedoch auch für die Schätzung des einfachen CFA Modells in Gruppe 3 nennenswert und relevant ist. Die Modellkomplexität zuletzt hat einen uneinheitlichen Effekt auf die *SV*. So ist die *SV* für das NLC-CFA Modell sowie für Modelle ohne Kontrolle und ohne simulierten FRE (Gruppe 3 des einfachen CFA Modells) umso höher, desto weniger Indikatoren an der Schätzung beteiligt sind. Für Modelle ohne Kontrolle und als 2-Wege-Interaktionen simulierten Assimilationseffekt (Gruppe 1 des einfachen CFA Modells) sowie für das CFA-MTMM Modell ist die *SV* relativ zum Benchmark Modell höher bei $q = 3$ Indikatoren und geringer bei $q = 4$ Indikatoren. Für Modelle ohne Kontrolle und als 3-Wege-Interaktion simulierter Kontrasteffekt (Gruppe 2 des einfachen CFA Modells) ist die *SV* geringer bei $q = 4$ Indikatoren und am geringsten bei $q = 3$ Indikatoren. Am

stärksten sind diese Einflüsse der Modellkomplexität dabei für das NLC-CFA Modell sowie für Gruppe 2 (als 3-Wege-Interaktion simulierter Kontrasteffekt, der nicht im Modell spezifiziert wurde).

Zusammenfassung

Zusammenfassend und hinsichtlich der einleitend formulierten Frage für die Analyse der Verzerrungen von Faktorladungen lässt sich somit festhalten, dass die Kontrolle von FRE mittels NLC-CFA zu einer Verringerung der Verzerrung von Faktorladungen führt und somit zu effizienteren Messungen der Indikatoren von Konstruktvalidität und Kompositreliabilität beiträgt. Hinsichtlich der Verringerung der Verzerrung ist das NLC-CFA Modell zudem der CFA-MTMM überlegen. Diesbezüglich hat die Fallzahl insgesamt einen negativen Einfluss auf die Höhe der Verzerrungen von Faktorladungen, während der Einfluss der Modellkomplexität insgesamt eher unsystematisch ist, auch wenn bei Verwendung der NLC-CFA die Empfehlung gegeben werden kann, Modelle mit mindestens vier Indikatorvariablen zu schätzen, da die Verwendung von lediglich drei Indikatorvariablen in solchen Modellen zu einer stark erhöhten *SV* führt.

4.5 Effekt Effizienz Analyse

Nachdem die Verzerrungen der Faktorladungen hinsichtlich ihrer Abhängigkeit von den verschiedenen Simulationsspezifikationen per Regression analysiert wurden, erfolgt nun die Analyse der Verzerrungen der tatsächlich geschätzten FRE in Form der Haupt- und Interaktionseffekte innerhalb der NLC-CFA sowie der Messfehlerkovarianzen innerhalb der CFA-MTMM, um die Effizienz der beiden vorgeschlagenen Modelle hinsichtlich der Möglichkeit zur Identifikation von FRE zu evaluieren und miteinander zu vergleichen. Diesbezüglich werden die Daten der 10000 geschätzten NLC-CFA Modelle sowie der CFA-MTMM Modelle verwendet, in welchen FRE spezifiziert und welche mit metrischer Invarianz geschätzt wurden, wobei innerhalb der vorliegenden Analyse für NLC-CFA Modelle sowie für CFA-MTMM Modelle die jeweiligen die Standard-Verzerrung (*SV*) der geschätzten Effekte gemäß Kapitel 4.1 im Fokus der Betrachtung stehen. Die *SV* (welche zwecks vereinfachter Interpretation mit 100 multipliziert sind) der NLC-CFA der Effekte über alle Gruppen *z*, differenziert nach Effekt β_1 bis β_7 (vgl. Kapitel 4.1 und Tabelle 3), sowie der CFA-MTMM der einzelnen Gruppen *z* werden dabei erneut per Regression auf die Fallzahl, die Modellkomplexität (indiziert durch die Anzahl an

Indikatoren *q*), die wahre simulierte Faktorladung (interpretierbar als approximierter Anteil an Common Variance bzw. Annäherung an die standardisierte Faktorladung), den Anteil der systematischen Fehlervarianz an der gesamten Fehlervarianz, sowie die Interaktion der beiden zuletzt genannten Variablen zurückgeführt.

Die Höhe der geschätzten NLC-CFA Effekte können als Anteil der abhängigen Variablen interpretiert werden, welche durch den entsprechenden Faktor oder Moderator bedingt ist. Die Höhe der geschätzten Messfehlerkovarianzen wiederum können als absolute Messfehlerkovarianzen der geschätzten Modelle interpretiert werden, wobei als „wahrer Wert" 0 angenommen werden kann. Innerhalb der durchgeführten Regressionsanalysen ist Konstante *a* für NLC-CFA Modelle als durchschnittliche *SV* der geschätzten Effekte des Benchmark Modells und für CFA-MTMM Modelle als durchschnittliche *SV* der Messfehlerkovarianz des Benchmark Modells zu interpretieren, was einem Modell mit einer Fallzahl von 1500, fünf Indikatoren, einer Faktorladung von 0,6 und einem Anteil der systematischen Fehlervarianz an der gesamten Fehlervarianz von 50% entspricht. Die entsprechend zu schätzenden Modelle je Iteration sind in Abbildung 19 und Abbildung 20 dargestellt (vgl. Kapitel 4.1). Für die Evaluation der CFA Modelle impliziert der Regressionskoeffizient für den Effekt von n erneut die Veränderung der *SV* für zusätzliche 1000 Befragte. λ_{sim} und $\frac{\sigma_{sys}^2}{\sigma_{sys}^2+\sigma_{rand}^2}$ wurden derart rekodiert, dass der Regressionskoeffizient als Veränderung der *SV* für jeden zusätzlichen Prozentpunkt des (approximierten) Anteils der Common Variance an der gesamten Varianz bzw. des Anteils des systematischen Messfehlers am gesamten Messfehler interpretiert werden kann.

Verzerrung der Haupteffekte der NLC-CFA

Tabelle 13 zeigt die *SV* des Haupteffektes (über alle Gruppen *z*) der Kontextfrage y_{j-1} auf die Zielfrage y_j in Abhängigkeit der variablen Simulationsparameter für das NLC-CFA Modell, welcher für alle Simulationsmodelle und Gruppen mit einem wahren Wert von „0" simuliert wurde. Für das Benchmark Modell ist entsprechend aus Konstante *a* abzulesen, dass die *SV* des Schätzers bei 1,585 liegt. Dies bedeutet, dass für das Benchmark Modell im Schnitt für einzelne Effektschätzer eine *SV* von 0,0159 erwartet werden kann, wobei der Wert als Wurzel der durchschnittlichen quadrierten Abweichung des Schätzers vom wahren Wert über alle drei Modellgruppen zu verstehen ist. Dies lässt den Schluss zu, dass besagter Haupteffekt, zumindest für das Benchmark Modell, relativ effizient geschätzt werden kann.

Tabelle 13: Durchschnittliche Verzerrung der Effekte von β_1 (y_{j-1})

	b	SE	b*
a	1,585***	0,065	NA
n	-0,653***	0,051	-0,133
***q* = 3**	6,023***	0,182	0,415
***q* = 4**	0,262***	0,052	0,018
λ_{sim}	-0,020***	0,003	-0,049
$\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$	-0,087***	0,004	-0,284
int	< 0,001	< 0,001	0,008

a: Konstante der Regressionsschätzung; **n**: Effekt der Fallzahl auf die *SV* der Effekte; **q**: Einfluss der Anzahl an Indikatoren innerhalb des geschätzten Modells auf die *SV* der Effekte; λ_{sis}: Einfluss der wahren Faktorladung bzw. des Anteils an Common Variance auf die *SV* der Effekte; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$**:** Einfluss des Anteils des systematischen Messfehlers am gesamten Messfehler auf die *SV* der Effekte; **int**: Einfluss der Interaktion von λ_{sis} und $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ auf die *SV* der Effekte;

b***: p ≤ 0,001; b**: p ≤ 0,01; b*: p ≤ 0,05

a für *n* = 1500; *p* = 5; λ = 0,6; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ = 0,5

Auf Basis der standardisierten Regressionskoeffizienten *b** lassen sich darüber hinaus die Stärke der Effekte der Modell- und Simulationsspezifikationen auf die *SV* der Schätzer systematisieren. So hat beispielsweise der (approximierte) Anteil der Common Variance lediglich einen geringen Einfluss auf die Höhe der *SV* des Schätzers (*b** = -0,049, *b* = -0,020). Dabei ist zu beachten, dass es sich dabei um einen negativen Effekt handelt: Je höher der Anteil an Common Variance bei einem konstanten Anteil systematischer Fehlervarianz (und desto geringer somit der FRE), desto geringer die *SV* bzw. die Verzerrung des Schätzers. Daneben weißt der Anteil der systematischen Fehlervarianz an der gesamten Fehlervarianz ebenfalls einen negativen, jedoch weitaus höheren Effekt auf (*b** = -0,284, *b* = -0,087), so dass ein höherer Anteil des systematischen Messfehlers bei einem konstanten Anteil an Common Variance (was in einem höheren, absoluten FRE resultiert) zu einer verringerten Verzerrung führt. Von der Interaktion der beiden Variablen geht darüber hinaus kein statistisch signifikanter Effekt aus. Dies lässt sich entsprechend derart interpretieren, dass je höher die jeweiligen systematischen Varianzanteile (der Anteil an Common Variance und der Anteil an systematischer Fehlervarianz) und desto geringer entsprechend der Varianzanteil, welcher auf zufällige Messfehler zurückgeht,

desto geringer die *SV* und desto effizienter die Schätzung. In anderen Worten bedeutet dies, dass desto geringer die zufällige Fehlervarianz der abhängigen Variablen, desto unverzerrter die Schätzung des Effektes. Werden die Effekte hingegen für die absolut zu erwartende *SV* betrachtet, so müsste sich der (approximierte) Anteil an Common Variance etwa um 50 Prozentpunkte bzw. der Anteil der systematischen Fehlervarianz etwa um 10 Prozentpunkte reduzieren, damit die *SV* des Schätzers um 0,01 zunimmt. Absolut betrachtet handelt es sich hierbei also um geringe Veränderungen der erwarteten Verzerrungen.

Daneben hat die Fallzahl einen statistisch signifikanten, negativen und absolut betrachtet erhöhten Einfluss auf die *SV* des Schätzers mit b^* = -0,133 und b = -0,653. Mit zunehmender Fallzahl reduziert sich somit die *SV*, die Effizienz der Schätzung erhöht sich (etwa 2400 zusätzliche Befragte reduzieren die durchschnittlich erwartete *SV* von Benchmark Modellen auf 0). Die Modellkomplexität zuletzt hat ebenfalls einen Einfluss auf die *SV*. So wird die *SV* für Effekte von Modellen mit drei Indikatorvariablen relativ zum Benchmark Modell mit b = 6,023 (b^* = 0,415) geschätzt, während Modelle mit vier Indikatorvariablen mit b = 0,262 (b^* = 0,018) geschätzt werden. Desto weniger Indikatoren an der Schätzung beteiligt sind, desto höher ist somit die *SV* und damit die Verzerrung, auch wenn die zusätzliche Verzerrung bei Verwendung von vier Indikatorvariablen sehr gering ist.

Der Haupteffekt der Kontextfrage y_{j-1} auf die Zielfrage y_j kann somit insgesamt relativ effizient geschätzt werden. Die absolut zu erwartende *SV* berücksichtigend führt für die simulierten Daten vor allem die Verwendung von drei statt fünf Indikatorvariablen sowie eine erhebliche Reduktion der Fallzahl zu einer relevanten Verzerrung des Schätzers.

Zusätzlich zum Haupteffekt der Kontextfrage y_{j-1} auf die Zielfrage y_j wurden ebenfalls zwecks vollständiger statistischer Kontrolle aller möglichen Interaktionen die Haupteffekte von ξ_1 und ξ_2 geschätzt, um eine vollständige Verzerrungsanalyse aller beteiligten Effekte zu realisieren, auch wenn die benannten Effekte gemäß der Diskussion in Kapitel 3.3.2 inhaltlich für eine Analyse der FRE nicht bedeutsam sind. Die geschätzten, durchschnittlichen Verzerrungen der beiden Haupteffekte von ξ_1 und ξ_2 unterscheiden sich dabei nicht wesentlich, so dass die Interpretation der Verzerrungen beider Effektschätzer zusammengefasst wird. Verzerrungen des Haupteffektes von ξ_1 sind in Tabelle 14 einzusehen, Tabelle 15 zeigt die Verzerrungen der Haupteffekte von ξ_2. Der wahre, simulierte Effekt für beide Schätzer entspricht über alle Simulationen und Gruppen hinweg 0.

Tabelle 14: Durchschnittliche Verzerrung der Effekte von β_2 (ξ_1)

	b	**SE**	**b***
a	0,997***	0,010	NA
n	-0,215***	0,004	-0,514
***q* = 3**	0,029*	0,011	0,023
***q* = 4**	-0,014	0,011	-0,011
$\boldsymbol{\lambda_{sim}}$	-0,013***	< 0,001	-0,380
$\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$	-0,001*	< 0,001	-0,020
int	< 0,001	< 0,001	0,001

a: Konstante der Regressionsschätzung; **n**: Effekt der Fallzahl auf die *SV* der Effekte; **q**: Einfluss der Anzahl an Indikatoren innerhalb des geschätzten Modells auf die *SV* der Effekte; $\boldsymbol{\lambda_{sis}}$: Einfluss der wahren Faktorladung bzw. des Anteils an Common Variance auf die *SV* der Effekte; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$**:** Einfluss des Anteils des systematischen Messfehlers am gesamten Messfehler auf die *SV* der Effekte; **int**: Einfluss der Interaktion von λ_{sis} und $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ auf die *SV* der Effekte;

b***: p ≤ 0,001; b**: p ≤ 0,01; b*: p ≤ 0,05

a für n = 1500; p = 5; λ = 0,6; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ = 0,5

Auf Basis der Konstante *a* ist entsprechend für das Benchmark Modell eine geringe *SV* der Effekte von etwa 1,0 indiziert. Für das Benchmark Modell ist somit eine effiziente Schätzung der beiden Haupteffekte von ξ_1 und ξ_2 gegeben. Analog zur *SV* des Haupteffektes der Kontextfrage y_{j-1} ist auch die *SV* der beiden Haupteffekte der latenten Konstrukte ξ_1 und ξ_2 systematisch abhängig vom (approximierten) Anteil an Common Variance mit b = -0,013 ($b^* \approx$ -0,37). Je höher also der Anteil an Common Variance bei konstantem Anteil an systematischer Fehlervarianz (was in einem absolut betrachtet geringeren FRE resultiert), desto geringer die Verzerrung des Schätzers. Dabei ist jedoch zu berücksichtigen, dass die Verzerrung trotz erhöhtem b^* insgesamt als gering zu interpretieren ist, da eine Verringerung des Anteils an Common Variance um 100 Prozentpunkte nötig wäre, damit die absolut zu erwartende *SV* des Schätzers um 0,01 zunimmt. Der Effekt des Anteils an systematischer Fehlervarianz ist zwar für beide Haupteffekte ebenfalls statistisch signifikant, jedoch derart gering, dass er vernachlässigt werden kann (b = -0,001 und b^* = -0,020 für den Haupteffekt von ξ_1 und b = 0,000 und b^* = -0,018 für den Haupteffekt von ξ_2). Die Interaktion der beiden Variablen ist daneben statistisch nicht signifikant.

Tabelle 15: Durchschnittliche Verzerrung der Effekte von β_3 (ξ_2)

	b	**SE**	**b***
a	1,010***	0,011	NA
n	-0,219***	0,004	-0,516
***q* = 3**	0,012	0,012	0,010
***q* = 4**	-0,018	0,011	-0,014
$\boldsymbol{\lambda_{sim}}$	-0,013***	< 0,001	-0,370
$\frac{\sigma_{sys}^2}{\sigma_{sys}^2 + \sigma_{rand}^2}$	< 0,001*	< 0,001	-0,018
int	< 0,001	< 0,001	-0,005

a: Konstante der Regressionsschätzung; **n**: Effekt der Fallzahl auf die *SV* der Effekte; **q**: Einfluss der Anzahl an Indikatoren innerhalb des geschätzten Modells auf die *SV* der Effekte; $\boldsymbol{\lambda_{sis}}$: Einfluss der wahren Faktorladung bzw. des Anteils an Common Variance auf die *SV* der Effekte; $\frac{\sigma_{sys}^2}{\sigma_{sys}^2+\sigma_{rand}^2}$**:** Einfluss des Anteils des systematischen Messfehlers am gesamten Messfehler auf die *SV* der Effekte; **int**: Einfluss der Interaktion von λ_{sis} und $\frac{\sigma_{sys}^2}{\sigma_{sys}^2+\sigma_{rand}^2}$ auf die *SV* der Effekte;

b***: p ≤ 0,001; b**: p ≤ 0,01; b*: p ≤ 0,05

a für n = 1500; p = 5; λ = 0,6; $\frac{\sigma_{sys}^2}{\sigma_{sys}^2+\sigma_{rand}^2}$ = 0,5

Die Fallzahl hat für beide Haupteffekte einen statistisch signifikanten Einfluss auf die *SV* mit $b^* \approx$ -0,51 und $b \approx$ -0,22, so dass insgesamt geschlossen werden kann, dass die beiden Haupteffekte umso effizienter geschätzt werden können, desto höher die Fallzahl ist. Das hohe b^* soll jedoch auch hier nicht zu der Annahme eines bedeutsamen Effekts verleiten, da es im Schnitt 5000 Befragte weniger benötigt, damit die absolut zu erwartende *SV* des Schätzers um 0,01 zunimmt. Bezüglich der Modellkomplexität führt die Verwendung von vier statt fünf Indikatorvariablen für beide Haupteffekte zu keiner statistisch signifikanten Veränderung der *SV*. Bei Verwendung von drei statt fünf Indikatoren ist die *SV* für den Haupteffekt von ξ_1 statistisch signifikant um 0,029 höher, während der gleiche Effekt für den Haupteffekt von ξ_2 nicht statistisch signifikant ist. Die Erhöhung der Verzerrung ist also erneut derart gering, dass sie für die entsprechenden Effekte vernachlässigt werden kann.

Die beiden Haupteffekte können somit über alle Simulationen und relativ unabhängig von den variierenden Simulationsparametern, stabil und effizient geschätzt werden so der relativ schwache Einfluss der Prädiktoren auf die tatsächliche Verzerrung berücksichtigt wird.

Verzerrung der 2-Wege-Interaktionen der NLC-CFA

Nach Analyse der Verzerrungen der Haupteffekte werden als nächstes die durchschnittlichen Verzerrungen der 2-Wege Interaktionseffekte der Kontextfrage mit den exogenen latenten Variablen $y_{j-1}\xi_1$ und $y_{j-1}\xi_2$ sowie der exogenen latenten Variablen miteinander ($\xi_1\xi_2$) betrachtet. Da sich die Verzerrungen zwischen den beiden Interaktionseffekten von $y_{j-1}\xi_1$ und $y_{j-1}\xi_2$ kaum unterscheiden (vgl. Tabelle 16 für Verzerrung der Effekte von $y_{j-1}\xi_1$ und Tabelle 17 für Verzerrungen der Effekte von $y_{j-1}\xi_2$), wird die Analyse der Verzerrungen beider Interaktionseffekte im Folgenden erneut zusammengefasst.

Tabelle 16: Durchschnittliche Verzerrung der Effekte von β_4 ($y_{j-1}\xi_1$)

	b	**SE**	**b***
a	2,205***	0,032	NA
n	-0,085***	0,009	-0,069
$q = 3$	-0,064	0,036	-0,018
$q = 4$	-0,268***	0,037	-0,073
λ_{sim}	-0,058***	0,001	-0,578
$\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$	0,027***	0,001	0,350
int	-0,001***	< 0,001	-0,193

a: Konstante der Regressionsschätzung; **n**: Effekt der Fallzahl auf die *SV* der Effekte; **q**: Einfluss der Anzahl an Indikatoren innerhalb des geschätzten Modells auf die *SV* der Effekte; λ_{sis}: Einfluss der wahren Faktorladung bzw. des Anteils an Common Variance auf die *SV* der Effekte; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$**:** Einfluss des Anteils des systematischen Messfehlers am gesamten Messfehler auf die *SV* der Effekte; **int**: Einfluss der Interaktion von λ_{sis} und $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ auf die *SV* der Effekte;

b***: p ≤ 0,001; b**: p ≤ 0,01; b*: p ≤ 0,05

a für n = 1500; p = 5; λ = 0,6; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ = 0,5

Bezüglich der Interpretation der *SV* beider Interaktionseffekte innerhalb des Benchmark Modells, abzulesen in Konstante *a*, ist zunächst festzuhalten, dass für dieses Modell in Gruppe 1 (innerhalb welcher ein Assimilationseffekt simuliert wurde) ein wahrer Effekt von -0,1 simuliert wurde (der reverse approximierter Anteil an Common Variance von 0,4 Mal ein Anteil von

0,5 des systematischen Messfehlers am gesamten Fehler Mal -0,5 als technischer Simulationskoeffizient, siehe Kapitel 4.1), während für Gruppe 2 (innerhalb welcher ein Kontrasteffekt simuliert wurde) und Gruppe 3 (innerhalb welcher kein FRE simuliert wurde) ein wahrer Effekt von 0 simuliert wurde. Gemäß Konstante *a* beträgt die *SV* beider Interaktionseffekte ungefähr 2,2 (was als Wurzel der durchschnittlichen quadrierten Verzerrungen über alle Gruppen *z* aufgrund der Multiplikation mit 100 einem Wert von 0,022 entspricht) und ist statistisch hoch signifikant, so dass die Effekte für Benchmark Modelle insgesamt noch relativ effizient geschätzt werden. Die Abhängigkeit der FRE von weiteren, exogenen Variablen oder latenten Konstrukten ist somit grundsätzlich identifizierbar.

Darüber hinaus haben sowohl der Anteil des systematischen Messfehlers ($b^* \approx 0{,}35$ und $b = 0{,}027$), der (approximierte) Anteil an Common Variance ($b^* = -0{,}57$ und $b = -0{,}058$), als auch die Interaktion zwischen beiden ($b^* = 0{,}19$ und $b = 0{,}001$) einen Einfluss auf die Verzerrung der geschätzten Interaktionseffekte, auch wenn die Effekte absolut betrachtet erneut lediglich einen äußerst geringen Einfluss auf die erwartete Höhe der *SV* haben. Bezüglich des Anteils des systematischen Messfehlers lässt sich beobachten, dass wenn dieser bei einer ansonsten konstanten Faktorladung von 0,6 zunimmt (was in einem erhöhten, absoluten Anteil des FRE resultiert), die *SV* und damit die Verzerrung zunimmt. Der Anteil an Common Variance hat hingegen einen negativen Einfluss auf die *SV* in dem Sinne, dass je geringer der Anteil bei einem ansonsten konstanten Anteil des systematischen Messfehlers (je höher entsprechend der FRE), desto höher die *SV* und damit die Verzerrung. Der Einfluss des Interaktionseffektes der beiden Anteile lässt sich derart interpretieren, dass wenn der Anteil an Common Variance abnimmt und der Anteil des systematischen Messfehlers zunimmt (was insgesamt in einem stärkeren FRE und als Produkt in einem negativen Ergebnis resultiert), die *SV* und damit die Verzerrung ebenfalls zunimmt (da ein negatives Produkt mal ein negativer Koeffizient zu einer positiven Schätzung der *SV* führt). Insgesamt lässt sich somit schließen, dass die beiden Interaktionseffekte von $y_{j-1}\xi_1$ und $y_{j-1}\xi_2$ umso mehr verzerrt sind, desto stärker der FRE simuliert wurde, auch wenn der Einfluss der Höhe des FRE auf die Verzerrung, wie bereits beschrieben, absolut betrachtet relativ gering ist (der approximierte Anteil an Common Variance, also des Simulationsparameters mit dem stärksten Effekt gemäß standardisiertem Regressionskoeffizient b^*, müsste sich beispielsweise um etwa 17 Prozentpunkte verringern, damit sich die absolute *SV* der beiden Interaktionseffekte um 0,01 erhöht).

Tabelle 17: Durchschnittliche Verzerrung der Effekte von β_5 ($y_{j-1}\xi_2$)

	b	**SE**	**b***
a	2,188***	0,032	NA
n	-0,083***	0,009	-0,067
***q* = 3**	-0,059	0,036	-0,016
***q* = 4**	-0,247***	0,037	-0,068
$\boldsymbol{\lambda_{sim}}$	-0,058***	0,001	-0,574
$\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$	0,027***	0,001	0,356
int	-0,001***	< 0,001	-0,197

a: Konstante der Regressionsschätzung; **n**: Effekt der Fallzahl auf die *SV* der Effekte; **q**: Einfluss der Anzahl an Indikatoren innerhalb des geschätzten Modells auf die *SV* der Effekte; $\boldsymbol{\lambda_{sis}}$: Einfluss der wahren Faktorladung bzw. des Anteils an Common Variance auf die *SV* der Effekte; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$**:** Einfluss des Anteils des systematischen Messfehlers am gesamten Messfehler auf die *SV* der Effekte; **int**: Einfluss der Interaktion von λ_{sis} und $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ auf die *SV* der Effekte;

b***: p ≤ 0,001; b**: p ≤ 0,01; b*: p ≤ 0,05

a für *n* = 1500; *p* = 5; λ = 0,6; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ = 0,5

Zusätzlich hat die Komplexität des Modells ebenfalls einen Einfluss auf die geschätzten Verzerrungen bei Verwendung von vier statt fünf Indikatorvariablen mit $b^* \approx -0{,}07$ und $b \approx 0{,}25$, während die Verwendung von drei statt fünf Indikatorvariablen keinen statistisch signifikanten Einfluss auf die *SV* nimmt. Der negative Effekt impliziert eine Verringerung der *SV* für Modelle mit vier statt fünf Indikatorvariablen, auch wenn der Effekt absolut betrachtet erneut sehr gering ist. Die Fallzahl zuletzt hat ebenfalls einen sehr geringen negativen Effekt auf die *SV* mit standarisiertem $b^* \approx -0{,}07$ und unstandardisiertem $b \approx -0{,}08$. Entsprechend kann interpretiert werden, dass die *SV* der Effekte mit zunehmender Fallzahl umso geringer wird, was einer erhöhten Effizienz bei zunehmender Fallzahl entspricht.

Für die beiden Interaktionseffekte ist somit festzuhalten, dass auch wenn systematische Einflüsse der Simulationsspezifikationen auf die *SV* der beiden Interaktionseffekte identifiziert werden konnten, diese alle derart gering sind, so dass insgesamt von einer stabilen und effizienten Schätzung der Interaktionseffekte für die simulierten Daten ausgegangen werden kann.

Tabelle 18: Durchschnittliche Verzerrung der Effekte von β_6 ($\xi_1\xi_2$)

	b	**SE**	**b***
a	1,260***	0,014	NA
n	-0,274***	0,006	-0,464
***q* = 3**	-0,008	0,015	-0,004
***q* = 4**	0,015	0,015	0,009
$\boldsymbol{\lambda_{sim}}$	-0,024***	< 0,001	-0,492
$\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$	< 0,001	< 0,001	-0,010
int	< 0,001	< 0,001	-0,015

a: Konstante der Regressionsschätzung; **n**: Effekt der Fallzahl auf die *SV* der Effekte; **q**: Einfluss der Anzahl an Indikatoren innerhalb des geschätzten Modells auf die *SV* der Effekte; $\boldsymbol{\lambda_{sis}}$: Einfluss der wahren Faktorladung bzw. des Anteils an Common Variance auf die *SV* der Effekte; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$**:** Einfluss des Anteils des systematischen Messfehlers am gesamten Messfehler auf die *SV* der Effekte; **int**: Einfluss der Interaktion von λ_{sis} und $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ auf die *SV* der Effekte;

b***: p ≤ 0,001; b**: p ≤ 0,01; b*: p ≤ 0,05

a für n = 1500; p = 5; λ = 0,6; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ = 0,5

Zudem werden die *SV* des Interaktionseffektes von $\xi_1\xi_2$ analysiert, welcher Gemäß der Diskussion in Kapitel 3.2.2 keine inhaltliche Bedeutung hinsichtlich der Interpretation der Höhe der FRE hat und vor allem zwecks Kontrolle der relevanten Interaktionseffekte spezifiziert wird. Dabei gilt für den Effekt grundsätzlich, dass dieser in allen Iterationen und Gruppen mit einem wahren Wert von 0 simuliert wurde. Die diesbezügliche *SV* der Effekte ist in Tabelle 18 abzulesen. Gemäß Konstante *a* (*b* = 1,260) ist für das Benchmark Modell entsprechend ersichtlich, dass hier im Schnitt lediglich eine sehr geringe *SV* zu erwarten ist. Die Schätzung des Interaktionseffektes ist somit innerhalb des Benchmark Modells effizient. Daneben wird die *SV* des Effektes lediglich durch die Fallzahl (*b** = -0,464, *b* = 0,274) und den (approximierten) Anteil an Common Variance (*b** = 0,492, *b* = 0,024) statistisch signifikant beeinflusst. Dies bedeutet für die Schätzung des Effektes, dass dieser umso geringer verzerrt ist, desto höher die Fallzahl. Zudem verringert sich die Verzerrung bei einem zunehmenden Anteil an Common Variance. Erneut ist jedoch darauf hinzuweißen, dass es sich bei den Effekten trotz eines hohen *b** absolut betrachtet um Effekte mit schwachem Einfluss auf die zu erwartende *SV* handelt. Der Anteil an

Common Variance beispielsweise müsste etwa um 50 Prozentpunkte abnehmen, damit die absolut zu erwartende *SV* um 0,01 zunimmt. Über alle simulierten Daten hinweg ist somit auch für diesen Interaktionseffekt eine effiziente Schätzung zu erwarten.

Verzerrung der 3-Wege-Interaktion der NLC-CFA

Nach Betrachtung der Verzerrungen der Haupteffekte sowie der 2-Wege-Interaktionseffekte, welche den isolierten Einfluss von ξ_1 und ξ_2 auf den Effekt zwischen Kontextfrage y_{j-1} und Zielfrage y_j quantifizieren, werden nachfolgend die Verzerrungen der 3-Wege-Interaktion $y_{j-1}\xi_1\xi_2$, welche den Einfluss des Produktes von ξ_1 und ξ_2 auf den Effekt zwischen Kontextfrage y_{j-1} und Zielfrage y_j quantifizieren, analysiert. Die *SV* ist in Tabelle 19 einzusehen. Diesbezüglich ist zunächst festzuhalten, dass bei Benchmark-Spezifikation für Gruppe 1 (mit simuliertem Assimilationseffekt) für den entsprechenden Effekt ein wahrer Wert von 0, für Gruppe 2 (mit simuliertem Kontrasteffekt) ein wahrer Wert von 0,2 (ein reverser approximierter Anteil an Common Variance von 0,4 Mal ein Anteil von 0,5 des systematischen Messfehlers am gesamten Fehler Mal 1 als technischer Simulationskoeffizient, siehe Kapitel 4.1) und für Gruppe 3 (kein FRE simuliert) erneut ein wahrer Wert von 0 simuliert wurde.

Diesbezüglich ist aus Konstante *a* abzulesen, dass die *SV* des Effektes über alle Gruppen im Schnitt 6,237 (0,062 so berücksichtigt wird, dass die *SV* mit 100 multipliziert wurde) beträgt, was einer substanziellen Verzerrung entspricht. Da der wahre Effekt 0,2 beträgt, ist der Effekt somit grundsätzlich identifizierbar, auch wenn eine verzerrte Schätzung zu erwarten ist. Zusätzlich zur Identifizierbarkeit der Abhängigkeit der FRE von exogenen Variablen oder latenten Konstrukten ist somit (unter der Bedingung der Benchmark Spezifikation) ebenfalls die Abhängigkeit der FRE von 3-Wege-Interaktionen identifizierbar, auch wenn es sich hierbei nicht um eine effiziente Schätzung handeln muss.

Der (approximierte) Anteil an Common Variance (b^* = 0,379, b = -0,178), der Anteil des systematischen Messfehlers am gesamten Messfehler (b^* = 0,282, b = 0,101) sowie die Interaktion der beiden Variablen (b^* = -0,258, b = 0,005) haben absolut betrachtet ebenfalls einen substanziellen Einfluss auf die *SV* des 3-Wege-Interaktionseffektes. In diesem Sinne erhöht sich die *SV* relativ zum Benchmark Modell bei einem abnehmenden Anteil an Common Variance und einem konstanten Anteil des systematischen Messfehlers, also bei stärkeren FRE. Ebenfalls erhöht sich die *SV* relativ zum Benchmark Modell mit zunehmendem Anteil des systematischen Messfehlers am gesamten Messfehler bei einem ansonsten konstanten Anteil an Common Variance, also ebenfalls bei stärkerem FRE. Verringert sich der Anteil an Common Variance und

erhöht sich der Anteil des systematischen Messfehlers, wird der FRE also stärker, so erhöht sich die *SV* des Effektes aufgrund des negativen Interaktionseffektes ebenfalls. Erneut wird der Effekt entsprechend umso stärker verzerrt, desto stärker der FRE simuliert wurde.

Tabelle 19: Durchschnittliche Verzerrung der Effekte von β_7 ($y_{j-1}\xi_1\xi_2$)

	b	**SE**	**b***
a	6,237***	0,177	NA
n	0,134**	0,049	0,023
***q* = 3**	-3,205***	0,200	-0,189
***q* = 4**	-3,199***	0,204	-0,188
$\boldsymbol{\lambda_{sim}}$	-0,178***	0,005	-0,379
$\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$	0,101***	0,004	0,282
int	-0,005***	< 0,001	-0,258

a: Konstante der Regressionsschätzung; **n**: Effekt der Fallzahl auf die *SV* der Effekte; **q**: Einfluss der Anzahl an Indikatoren innerhalb des geschätzten Modells auf die *SV* der Effekte; $\boldsymbol{\lambda_{sis}}$: Einfluss der wahren Faktorladung bzw. des Anteils an Common Variance auf die *SV* der Effekte; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$**:** Einfluss des Anteils des systematischen Messfehlers am gesamten Messfehler auf die *SV* der Effekte; **int**: Einfluss der Interaktion von λ_{sis} und $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ auf die *SV* der Effekte;

b***: p ≤ 0,001; b**: p ≤ 0,01; b*: p ≤ 0,05

a für $n = 1500$; $p = 5$; $\lambda = 0{,}6$; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}} = 0{,}5$

Die Fallzahl hat ebenfalls einen signifikanten Einfluss auf die Höhe der *SV* mit $b = 0{,}134$ und $b^* = 0{,}023$, was derart zu interpretieren ist, dass je höher die Fallzahl, desto höher die Verzerrung. Absolut betrachtet ist dieser Einfluss jedoch nicht substanziell. Der Einfluss der Modellkomplexität ist hingegen erneut nennenswert mit $b^* = -0{,}189$ und $b = -3{,}205$ für Modelle mit drei Indikatorvariablen und $b^* = -0{,}188$ und $b = -3{,}199$ für Modelle mit vier Indikatorvariablen. Dies bedeutet, dass Modelle mit drei oder vier statt fünf Indikatorvariablen eine geringere *SV* und damit eine geringere Verzerrung aufweisen, wobei die Differenz der Effektstärken zwischen drei und vier Indikatoren lediglich marginal ist.

Für die geschätzte 3-Wege-Interaktion kann also relativ stabil und für alle simulierten Daten eine substanzielle Verzerrung des Schätzers festgestellt werden, welche sich durch Verwendung von weniger als fünf Indikatorvariablen zumindest teilweise verringert und bei stärkeren FRE erhöht ist. Aufgrund der absoluten Höhe des simulierten Interaktionseffektes ist dieser innerhalb des Benchmark Modells trotz der substanziellen Verzerrung prinzipiell identifizierbar.

Effizienz der Effektschätzungen innerhalb der NLC-CFA

Bezogen auf das Benchmark Modell implizieren die analysierten Ergebnisse grundsätzlich relativ geringe *SV* für alle Schätzer bis auf jenen der 3-Wege-Interaktion. Die effizientesten Schätzungen mit den geringsten *SV* für Benchmark Modelle sind dabei für die Effekte von ξ_1, ξ_2 und $\xi_1\xi_2$ gegeben, also für jene exogenen Variablen, welche nicht als Interaktionseffekt mit der Kontextfrage gebildet sind oder den Haupteffekt der Kontextfrage darstellen und somit auch keine Varianzanteile beinhalten, welche mit dem FRE assoziiert sind. In dieser Reihenfolge zunehmend verzerrt sind der Haupteffekt von y_{j-1}, die 2-Wege-Interaktionseffekte von $y_{j-1}\xi_1$ und $y_{j-1}\xi_2$ und schließlich der 3-Wege-Interaktionseffekt von $y_{j-1}\xi_1\xi_2$. Die *SV* der 2-Wege-Interaktionseffekt von $y_{j-1}\xi_1$ und $y_{j-1}\xi_2$ sowie des Haupteffektes von y_{j-1} sind diesbezüglich nach wie vor gering, so dass auch für diese eine effiziente Schätzung innerhalb des Benchmark Modells zu erwarten ist. Lediglich die *SV* des 3-Wege-Interaktionseffektes von $y_{j-1}\xi_1\xi_2$ ist erhöht, wobei die Werte auch hier noch für eine allgemeine Identifizierbarkeit des entsprechenden Effektes innerhalb des Benchmark Modells sprechen.

Bezüglich des Einflusses des (approximierten) Anteils an Common Variance, des Anteils an systematischer Fehlervarianz sowie der Interaktion der beiden Variablen lässt sich für die Interaktionseffekte von $y_{j-1}\xi_1$, $y_{j-1}\xi_2$ und $y_{j-1}\xi_1\xi_2$ feststellen, dass je stärker der simulierte FRE, desto höher die Verzerrung der Schätzer. Dies ist insofern relevant, als dass eben jene Effekte zur Quantifizierung des FRE dienen. Je stärker der zu schätzende Effekt ausfällt, desto stärker also die Verzerrung der Schätzung. Für den Haupteffekt von y_{j-1} zeigt sich hingegen, dass dieser umso effizienter (mit geringerer *SV*) geschätzt wird, desto höher insgesamt die systematischen Varianzanteile (also der approximierte Anteil an Common Variance und der Anteil an systematischer Fehlervarianz) bzw. desto geringer der relative Anteil der zufälligen Fehlervarianz. Die *SV* der Effekte von ξ_1, ξ_2 und $\xi_1\xi_2$ sind hingegen lediglich durch den Anteil an Common Variance in dem Sinne beeinflusst, dass ein höherer Anteil an Common Variance zu einer geringeren *SV* und damit einer effizienteren Schätzung führt.

Unter Kontrolle des (approximierten) Anteils an Common Variance, des Anteils an systematischer Fehlervarianz sowie der Interaktion der beiden Variablen innerhalb der durchgeführten Regressionsanalysen lässt sich bezüglich der Fallzahl feststellen, dass diese für alle Effekte außer für die 3-Wege-Interaktion von $y_{j-1}\xi_1\xi_2$ einen negativen Effekt auf die *SV* hat. Für diese Effekte lässt sich somit sagen, dass je höher die Fallzahl, desto geringer die *SV* und desto effizienter die Schätzung, während dieser Zusammenhang für die 3-Wege-Interaktion umgekehrt ist.

Für alle bisher genannten Einflussfaktoren ist jedoch zu berücksichtigen, dass der tatsächliche Einfluss auf die absolut zu erwartende *SV* der einzelnen Schätzer (außer für den Einfluss der Fallzahl auf die *SV* des Haupteffektes von y_{j-1}) äußerst gering ist. Die geschätzte *SV* ist somit im Schnitt über alle simulierten Daten relativ stabil.

Lediglich die Modellkomplexität hat einen substanziellen Einfluss auf die *SV* der Effekte der Kontextfrage y_{j-1} sowie der 3-Wege Interaktion. Für den Effekt der Kontextfrage führt die Verwendung von drei statt fünf Indikatoren zu einer substanziellen Erhöhung der erwarteten *SV*, für die 3-Wege-Interaktion führt die Verwendung von drei oder vier statt fünf Indikatorvariablen hingegen zu einer substanziellen Verringerung der erwarteten *SV*. Es kann somit insgesamt geschlossen werden, dass eine Schätzung eines NLC-CFA Modells bei Verwendung von vier Indikatorvariablen zu einer effizienten Schätzung der Effektparameter führt, wobei für die 3-Wege-Interaktion nach wie vor eine Verzerrung erwartet werden kann welche jedoch geringer ausfällt als bei Verwendung von fünf Indikatorvariablen.

Verzerrungen der Messfehlerkovarianz der CFA-MTMM

Zusätzlich zu den FRE der NLC-CFA, welche als Regressionskoeffizienten in Form von Haupt- und Interaktionseffekten spezifiziert und geschätzt wurden, soll die Effizienz der geschätzten FRE der CFA-MTMM analysiert werden, welche als Messfehlerkovarianzen spezifiziert wurden. Wie bereits in Kapitel 4.1 und 4.2 beschrieben erfolgt die Analyse ebenfalls auf Basis der *SV* der Messfehlerkovarianzen, wobei die *SV* je analysierter Gruppe dargestellt werden, da die Messfehlerkovarianzen innerhalb der Gruppen als gleich und zwischen den Gruppen als verschieden geschätzt wurden. Die *SV* wurden zudem erneut mit 100 multipliziert, um die Interpretation zu vereinfachen. Die *SV* sind, inklusive der Rückführung auf die variierenden Simulationsspezifikationen, in Tabelle 20 für Gruppe 1 (simuliert mit Assimilationseffekt), in Tabelle 21 für Gruppe 2 (simuliert mit Kontrasteffekt) sowie in Tabelle 22 für Gruppe 3 (ohne FRE simuliert) dargestellt. Der Wert in Konstante *a* entspricht dabei erneut der durchschnittlich

geschätzten *SV* für Benchmark Modelle während die übrigen Regressionskoeffizienten die durchschnittlich zu erwartenden Veränderungen der *SV* angeben.

Tabelle 20: Durchschnittliche Verzerrung der Messfehlerkovarianz in Gruppe 1

	b	**SE**	**b***
a	0,252***	0,023	NA
n	-0,018	0,013	-0,010
$q = 3$	3,886***	0,049	0,703
$q = 4$	0,438***	0,030	0,079
λ_{sim}	0,041***	0,001	0,266
$\frac{\sigma^2_{sys}}{\sigma^2_{sys} + \sigma^2_{rand}}$	0,005***	0,001	0,040
int	< 0,001***	< 0,001	-0,063

Gruppe 1: Gruppe mit simuliertem Assimilationseffekt; **a**: Konstante der Regressionsschätzung; **n**: Effekt der Fallzahl auf die *SV* der Effekte; **q**: Einfluss der Anzahl an Indikatoren innerhalb des geschätzten Modells auf die *SV* der Effekte; λ_{sis}: Einfluss der wahren Faktorladung bzw. des Anteils an Common Variance auf die *SV* der Effekte; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$**:** Einfluss des Anteils des systematischen Messfehlers am gesamten Messfehler auf die *SV* der Effekte; **int**: Einfluss der Interaktion von λ_{sis} und $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ auf die *SV* der Effekte;
b***: p ≤ 0,001; b**: p ≤ 0,01; b*: p ≤ 0,05
a für n = 1500; p = 5; λ = 0,6; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ = 0,5

Konstante *a* aus Tabelle 20 zeigt für Benchmark Modelle in Gruppe 1 eine erwartete *SV* der Messfehlerkovarianz von 0,252. Wird dabei berücksichtigt, dass die Varianz der Indikatorvariablen um den Wert 1 schwankt und der angezeigt Wert in Konstante *a* mit 100 multipliziert wurde, so handelt es sich bei 0,252 um eine sehr geringe Verzerrung für Benchmark Modelle. Der Einfluss des (approximierten) Anteils an Common Variance ist mit b = 0,041 (b* = 0,266) positiv in dem Sinne, des je höher der Anteil (und desto geringer entsprechend der absolute FRE), desto höher die *SV* der Messfehlerkovarianz. Gleiches gilt für den positiven Effekt des Anteils des systematischen Messfehlers (b = 0,005 und b* = 0,040): Je höher der Anteil (und desto höher entsprechend der FRE), desto höher die *SV*. Die Interaktion der beiden Anteile zuletzt verfügt über einen negativen Effekt (b > 0,000 und b* = -0,063) was derart zu interpretieren ist, dass wenn der Anteil an Common Variance abnimmt und der Anteil des systematischen Messfehlers zunimmt (was insgesamt in einem stärkeren FRE und als Produkt in einem

negativen Ergebnis resultiert), die *SV* und damit die Verzerrung ebenfalls zunimmt (da ein negatives Produkt mal ein negativer Koeffizient zu einer positiven Schätzung der *SV* führt). Für die Schätzung der Messfehlerkovarianz lässt sich also beobachten, dass diese für höhere FRE, aber auch für höhere Anteile der Common Variance, also für absolut betrachtet geringere Anteile des zufälligen Messfehlers zunehmend verzerrt ist. Dabei ist jedoch analog zum Einfluss der Höhe des FRE in NLC-CFA Modellen zu berücksichtigen, dass die *SV* absolut betrachtet nur sehr gering durch die Varianzanteile beeinflusst wird. So muss der (approximierte) Anteil an Common Variance beispielsweise um etwa 25 Prozentpunkte erhöht werden (also z.B. von 50% auf 75%), um die *SV* um 0,01 zu erhöhen.

Daneben hat die Modellkomplexität einen substanziellen Einfluss auf die *SV* in Gruppe 1 für Modelle mit drei statt fünf indikatorvariablen mit $b = 3{,}886$ und $b^* = 0{,}703$ sowie mit $b = 0{,}438$ und $b^* = 0{,}079$ für Modelle mit vier statt fünf Indikatorvariablen. Gerade bei Verwendung von drei Indikatoren erhöht sich die Verzerrung somit substanziell. Die Fallzahl zuletzt hat keinen statistisch signifikanten Einfluss auf die *SV* der Messfehlerkovarianz in Gruppe 1.

Tabelle 21 zeigt für Benchmark Modelle in Gruppe 2 eine erwartete *SV* Messfehlerkovarianz von 1,149 in Konstante *a*, was einer stärkeren *SV* als in Gruppe 1, aber nach wie vor einer geringen Verzerrung entspricht. Relevant hierbei ist, dass der systematische Messfehler in Gruppe 2 auf Basis einer 3-Wege-Interaktion und in Gruppe 1 auf Basis einer 2-Wege-Interaktion simuliert wurde, die zugrundeliegende Systematik des Messfehlers in Gruppe 2 also komplexer ist. Eine komplexere Systematik des zugrundeliegenden Messfehlers führt in den vorliegenden Modellen und auf Basis der simulierten Daten somit zu einer höheren Verzerrung. Der Einfluss des (approximierten) Anteils an Common Variance ($b = 0{,}004$ und $b^* = 0{,}026$) ist daneben erneut positiv, jedoch vernachlässigbar gering. Gleiches gilt für den positiven Effekt des Anteils des systematischen Messfehlers ($b = 0{,}020$ und $b^* = 0{,}158$) sowie für den negativen Interaktionseffekt ($b = -0{,}002$ und $b^* = -0{,}238$). Für die Schätzung der Messfehlerkovarianz lässt sich somit erneut, analog zu Gruppe 1 beobachten, dass diese für höhere FRE, aber auch für höhere Anteile der Common Variance, also für absolut betrachtet geringere Anteile des zufälligen Messfehlers zunehmend verzerrt ist aber auch, dass die *SV* absolut betrachtet nur sehr gering durch die Varianzanteile beeinflusst wird (der Anteil an systematischer Fehlervarianz beispielsweise muss um 50 Prozentpunkte erhöht werden, um die *SV* um 0,01 zu erhöhen).

Tabelle 21: Durchschnittliche Verzerrung der Messfehlerkovarianz in Gruppe 2

	b	**SE**	**b***
a	1,149***	0,044	NA
n	-0,034	0,018	-0,016
***q* = 3**	2,788***	0,066	0,455
***q* = 4**	0,148**	0,051	0,024
$\boldsymbol{\lambda_{sim}}$	0,004*	0,002	0,026
$\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$	0,020***	0,001	0,158
int	-0,002***	< 0,001	-0,238

Gruppe 2: Gruppe mit simuliertem Kontrasteffekt; **a**: Konstante der Regressionsschätzung; **n**: Effekt der Fallzahl auf die *SV* der Effekte; **q**: Einfluss der Anzahl an Indikatoren innerhalb des geschätzten Modells auf die *SV* der Effekte; $\boldsymbol{\lambda_{sis}}$: Einfluss der wahren Faktorladung bzw. des Anteils an Common Variance auf die *SV* der Effekte; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$**:** Einfluss des Anteils des systematischen Messfehlers am gesamten Messfehler auf die *SV* der Effekte; **int**: Einfluss der Interaktion von λ_{sis} und $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ auf die *SV* der Effekte;

b***: p ≤ 0,001; b**: p ≤ 0,01; b*: p ≤ 0,05

a für n = 1500; p = 5; λ = 0,6; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ = 0,5

Daneben ist es die Modellkomplexität, welche in Gruppe 2 einen substanziellen Einfluss auf die *SV* mit b = 2,788 (b* = 0,455) für Modelle mit drei statt fünf indikatorvariablen und b = 0,148 (b* = 0,024) für Modelle mit vier statt fünf Indikatorvariablen aufweist. Gerade bei Verwendung von drei Indikatoren erhöht sich die Verzerrung somit substanziell. Die Fallzahl zuletzt hat keinen statistisch signifikanten Einfluss auf die *SV* der Messfehlerkovarianz.

Tabelle 22 zuletzt zeigt eine erwartete *SV* Messfehlerkovarianz von 0,382 in Konstante *a* für Benchmark Modelle in Gruppe 3, was einer stärkeren Verzerrung als in Gruppe 1 und einer geringeren Verzerrung als in Gruppe 2 entspricht, wobei die Verzerrung absolut betrachtet erneut als gering eingeschätzt werden kann. Dieser Befund ist insofern relevant, als dass in Gruppe 3 kein FRE in Form eines systematischen Messfehlers simuliert wurde, die Messfehlerkovarianz aber dennoch in höherem Maße verzerrt ist als für Gruppe 1, in welcher ein Assimilationseffekt simuliert wurde. Die Spezifikation einer Messfehlerkovarianz in Gruppe 3 entspricht dabei einer Fehlspezifikation des Modells, was scheinbar entsprechend zu einer erhöhten Verzerrung führt. Hinsichtlich des Einflusses des (approximierten) Anteils an Common Variance, des Einflusses des Anteils der systematischen Fehlervarianz sowie der Interaktion der

beiden sind erneut die gleichen Vorzeichen wie in Gruppe 1 und Gruppe 2 sowie erneut ähnlich geringe Effekte zu beobachten. Für die Schätzung der Messfehlerkovarianz lässt sich somit beobachten, dass diese für höhere FRE, aber auch für höhere Anteile der Common Variance, also für absolut betrachtet geringere Anteile des zufälligen Messfehlers zunehmend verzerrt ist sowie, dass die *SV* absolut betrachtet nur sehr gering durch die jeweiligen Varianzanteile beeinflusst wird.

Tabelle 22: Durchschnittliche Verzerrung der Messfehlerkovarianz in Gruppe 3

	b	**SE**	**b***
a	0,382***	0,025	NA
n	-0,044**	0,013	-0,024
$q = 3$	3,837***	0,049	0,692
$q = 4$	0,304***	0,031	0,055
λ_{sim}	0,041***	0,001	0,267
$\frac{\sigma^2_{sys}}{\sigma^2_{sys} + \sigma^2_{rand}}$	0,004***	0,001	0,031
int	< 0,001***	< 0,001	-0,051

Gruppe 3: Gruppe ohne simulierten FRE; **a**: Konstante der Regressionsschätzung; **n**: Effekt der Fallzahl auf die *SV* der Effekte; **q**: Einfluss der Anzahl an Indikatoren innerhalb des geschätzten Modells auf die *SV* der Effekte; λ_{sis}: Einfluss der wahren Faktorladung bzw. des Anteils an Common Variance auf die *SV* der Effekte; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$**:** Einfluss des Anteils des systematischen Messfehlers am gesamten Messfehler auf die *SV* der Effekte; **int**: Einfluss der Interaktion von λ_{sis} und $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ auf die *SV* der Effekte;
b***: p ≤ 0,001; b**: p ≤ 0,01; b*: p ≤ 0,05
a für n = 1500; p = 5; λ = 0,6; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ = 0,5

Ebenfalls mit den Befunden aus Gruppe 1 und Gruppe 2 vergleichbar ist der substanzielle Einfluss der Modellkomplexität mit b = 3,837 (b* = 0,692) für Modelle mit drei statt fünf indikatorvariablen und b = 0,304 (b* = 0,055) für Modelle mit vier statt fünf Indikatorvariablen. Bei Verwendung von drei Indikatoren erhöht sich die Verzerrung somit auch in Gruppe 3 substanziell. Die Fallzahl zuletzt hat für Gruppe 3 einen negativen, aber vernachlässigbar geringen Einfluss auf die *SV* mit b = -0,044 und b* = -0,024.

Zusammenfassend lässt sich für die Messfehlerkovarianzen somit festhalten, dass diese in den Benchmark Modellen für Gruppe 1 (mit Assimilationseffekt simuliert) am geringsten verzerrt sind und für Gruppe 2 (mit Kontrasteffekt simuliert) im höchsten Maße verzerrt sind. Die höchste Verzerrung in Gruppe 2 lässt sich dabei auf eine komplexe Systematik des zugrundeliegenden systematischen Messfehlers (als 3-Wege-Interaktion simuliert) zurückführen, während die erhöhte Verzerrung in Gruppe 3, innerhalb welcher keine FRE simuliert wurden, auf eine Fehlspezifikation des Modells zurückgeführt werden kann. Daneben haben die verschiedenen Varianzanteile in allen drei Gruppen einen Einfluss auf die Höhe der *SV* in dem Sinne, dass je geringer der Anteil des zufälligen Messfehlers, desto stärker sind die Messfehlerkovarianzen verzerrt. Dieser Einfluss ist jedoch, ebenso wie der Einfluss der Fallzahl auf die Verzerrung, vernachlässigbar gering, so dass insgesamt von einer stabilen Verzerrung der Schätzung über die verschiedenen Konfigurationen der Fallzahl und der Varianzanteile in den simulierten Daten hinweg ausgegangen werden kann. Lediglich die Modellkomplexität hat einen substanziellen Einfluss auf die Verzerrung, wenn lediglich drei statt fünf Indikatorvariablen an der Schätzung beteiligt sind.

Gegenüberstellung und Fazit

Für die Effizienz der Schätzungen der konkreten Effekte mittels NLC-CFA (auf Basis von regressionsbasierten Haupt- und Interaktionseffekten) sowie mittels CFA-MTMM (mittels Messfehlerkovarianzen) lässt sich insgesamt feststellen, dass die Effekte im Wesentlichen effizient, das heißt relativ unverzerrt, geschätzt werden können. Für Benchmark Modelle ist innerhalb der NLC-CFA lediglich der Effekt der 3-Wege-Interaktion substanziell verzerrt, auch wenn die Höhe der Verzerrung nach wie vor eine korrekte Identifikation des Effektes innerhalb des Benchmark Modells zulässt. Innerhalb der CFA-MTMM ist die Messfehlerkovarianz in Gruppe 2 am stärksten verzerrt, innerhalb welcher der FRE mittels 3-Wege-Interaktion simuliert wurde. Eine zu komplexe Systematik des zugrundeliegenden systematischen Messfehlers führt somit in beiden Modellen zu einer Verringerung der Effizienz. Jenseits des Benchmark Modells ist es vor allem die Modellkomplexität, welche einen substanziellen Einfluss auf die *SV*, sowohl innerhalb der NLC-CFA, als auch der CFA-MTMM, hat. In beiden Modellen führt die Verwendung von drei statt fünf Indikatorvariablen zu einer erhöhten Verzerrung. Für NLC-CFA Modelle wird die Verwendung von vier statt fünf Indikatorvariablen für eine möglichst effiziente Schätzung aller Effekte empfohlen, während dies in CFA-MTMM Modellen zumindest nicht zu einer substanziellen Erhöhung der Verzerrung führt. Für alle sonstigen simulierten Daten

führt sowohl die NLC-CFA, als auch die CFA-MTMM zu relativ konstant und gering verzerrten Schätzungen.

Die konkrete Vergleichbarkeit der Effizienz der geschätzten Effekte zwischen NLC-CFA und CFA-MTMM ist ansonsten insofern limitiert, als dass die *SV* der NLC-CFA auf Regressionskoeffizienten und der CFA-MTMM auf Messfehlerkovarianzen basiert. Von einem direkten Vergleich der Koeffizienten a wird daher abgesehen. Aber auch wenn insgesamt evaluiert wurde, dass beide Modelle zu relativ unverzerrten Schätzungen führen, wird an dieser Stelle trotzdem ein deutlicher Vorteil in der Verwendung der NLC-CFA gesehen, da dieses Modell mit einem deutlichen Informationsgewinn einhergeht. Eine konkrete Spezifikation der einzelnen Haupt- und Interaktionseffekte und die effiziente Schätzung dieser lassen eindeutige Rückschlüsse auf die zugrundeliegenden Wirkmechanismen systematischer Messfehler zu, welche durch das CFA-MTMM Modell in dieser Form schlicht nicht möglich sind.

4.6 Standardfehler Präzisionsanalyse

Nachfolgend werden die Standardfehler (SE) der geschätzten Effektparameter von β_1 bis β_7 der NLC-CFA sowie der Messfehlerkovarianzen aus Gruppe 1 bis Gruppe 3 der CFA-MTMM ebenfalls hinsichtlich ihrer Abhängigkeit von den variierenden Simulationsspezifikationen je Iteration analysiert, da diese die Präzision der Schätzer indizieren. Wie in der vorangegangenen Analyse werden hierfür die Daten der je 10000 geschätzten NLC-CFA und CFA-MTMM Modelle, in welchen FRE spezifiziert und die mit metrischer Invarianz geschätzt wurden, verwendet. Zur Analyse stehen entsprechend die geschätzten SE der Effekte, jedoch nicht der Faktorladungen. Die durchschnittlichen SE der NLC-CFA Effekte über alle Gruppen z, differenziert nach Effekt β_1 bis β_7 sowie die SE der CFA-MTMM Effekte innerhalb der Gruppen z (vgl. Kapitel 4.1) werden dabei analog zur vorangegangenen Effekt Effizienz Analyse per Regression auf die Fallzahl n, die Modellkomplexität (indiziert durch die Anzahl an Indikatoren q), die wahre unstandardisierte Faktorladung λ_{sim} (interpretierbar als approximierter Anteil an Common Variance an der gesamten Varianz bzw. Annäherung an die standardisierte Faktorladung), den Anteil der systematischen Fehlervarianz an der gesamten Fehlervarianz $\frac{\sigma_{sys}^2}{\sigma_{sys}^2+\sigma_{rand}^2}$, sowie die Interaktion der beiden zuletzt genannten Variablen zurückgeführt. Hinsichtlich der Interpretation der konkreten Höhe der Standardfehler ist erneut darauf hinzuweisen, dass die Höhe der zugehörigen NLC-CFA Effekte als Anteil der abhängigen Zielfrage zu interpretieren

sind, welcher durch den entsprechenden Faktor oder die entsprechende Interaktion bedingt ist, während die Höhe der zugehörigen Messfehlerkovarianzen der CFA-MTMM durch die Varianz der Indikatoren limitiert ist, welche um den Wert 1 schwankt. Zudem wurden alle SE mit 100 multipliziert, um die Interpretation der Ergebnisse zu vereinfachen. Abermals entspricht der Effekt der Konstante *a* dem zu erwartenden SE für Effekte des Benchmark Modells (n = 1500, q = 5, λ_{sim} = 0,6, $\frac{\sigma_{sys}^2}{\sigma_{sys}^2+\sigma_{rand}^2}$ = 0,5). Die entsprechend zu schätzenden Modelle je Iteration sind in Abbildung 19 und Abbildung 20 einzusehen (vgl. Kapitel 4.1). Ferner wurden die unabhängigen Variablen der Regressionsanalyse, wie bereits für die Analyse der Verzerrungen der Faktorladungen sowie die Effekt Effizienz Analyse, derart recodiert, dass der Effekt von n die Veränderung der SE für zusätzliche 1000 Befragte und die Effekte von λ_{sim} und $\frac{\sigma_{sys}^2}{\sigma_{sys}^2+\sigma_{rand}^2}$ die Veränderung der SE für jeden zusätzlichen Prozentpunkt des (approximierten) Anteils der Common Variance an der gesamten Varianz bzw. des Anteils des systematischen Messfehlers am gesamten Messfehler indiziert. Die Modellkomplexität q wurde als Dummy mit der Referenzkategorie q = 5 codiert. Generell gilt für die Analyse der SE, dass je kleiner diese sind, desto präziser die Schätzung ausfällt.

SE der Haupteffekte der NLC-CFA

Beginnend mit dem durchschnittlichen SE des Haupteffektes von y_{j-1} über alle Gruppen *z* der NLC-CFA ist zunächst festzuhalten, dass der durchschnittliche SE für entsprechende Effektschätzer des Benchmark Modells gemäß Konstante *a* (vgl. Tabelle 23) 1,336, was (da der tatsächlich geschätzte SE mit 100 multipliziert wurde) 0,0134 entspricht, wobei *a* statistisch signifikant ist. Für den Effekt des Benchmark Modells wird der SE somit als sehr klein geschätzt, was insgesamt einer äußerst präzisen Schätzung entspricht.

Der Anteil der systematischen Fehlervarianz an der gesamten Fehlervarianz hat einen negativen Effekt auf die Höhe des durchschnittlichen SE mit einem standardisierten Regressionskoeffizienten b^* = 0,413 (b = -0,049), was bedeutet, dass je höher der absolute Anteil des FRE ausfällt, der Effekt umso präziser geschätzt wird. Demgegenüber hat der (approximierte) Anteil an Common Variance einen positiven, aber schwachen Einfluss auf die Höhe des SE (b^* = 0,073, b = 0,011). Die Schätzung des Effektes wird somit bei einem abnehmenden Anteil an Common Variance (und einem höheren, absoluten FRE) ebenfalls präziser. Die Interaktion der beiden Anteile zuletzt hat erneut einen negativen und geringen Effekt (b^* = -0,127, b = -0,001) auf die Höhe des SE. Nimmt der Anteil an Common Variance ab und der Anteil des systematischen

Messfehlers zu, erhöht sich also der absolute FRE, so erhöht sich der SE, was in einer verringerten Präzision des Schätzers resultiert. Dabei ist der Interaktionseffekt jedoch um ein Vielfaches geringer als die beiden jeweiligen Haupteffekte, so dass deren Wirkrichtung nicht durch die Interaktion aufgehoben wird. Es gilt somit insgesamt, dass je stärker der FRE, desto präziser die Schätzung des Effektes von y_{j-1}, wobei der Anteil der systematischen Fehlervarianz den stärksten Einfluss nimmt, während der Einfluss des Anteils an Common Variance tatsächlich absolut betrachtet schwach und damit wenig substanziell ist.

Tabelle 23: Durchschnittliche SE der Effekte von β_1 ($\overline{SE}_{y_{j-1}}$)

	b	**SE**	**b***
a	1,336***	0,024	NA
n	-0,468***	0,015	-0,244
$q = 3$	3,181***	0,055	0,562
$q = 4$	0,269***	0,026	0,047
λ_{sim}	0,011***	0,001	0,073
$\frac{\sigma^2_{sys}}{\sigma^2_{sys} + \sigma^2_{rand}}$	-0,049***	0,001	-0,413
int	-0,001***	< 0,001	-0,127

a: Konstante der Regressionsschätzung; **n**: Effekt der Fallzahl auf SE der Effekte; **q**: Einfluss der Anzahl an Indikatoren innerhalb des geschätzten Modells auf die SE der Effekte; λ_{sis}: Einfluss der wahren Faktorladung bzw. des Anteils an Common Variance auf die SE der Effekte; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$: Einfluss des Anteils des systematischen Messfehlers am gesamten Messfehler auf die SE der Effekte; **int**: Einfluss der Interaktion von λ_{sis} und $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ auf die SE der Effekte;

b***: p ≤ 0,001; b**: p ≤ 0,01; b*: p ≤ 0,05

a für $n = 1500$; $p = 5$; $\lambda = 0{,}6$; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}} = 0{,}5$

Die Modellkomplexität wiederum hat ebenfalls einen relevanten Einfluss auf die Höhe des SE mit $b^* = 0{,}562$ und $b = 3{,}181$ für $q = 3$ Modelle und $b^* = 0{,}047$ und $b = 0{,}269$ für $q = 4$ Modelle. Modelle mit drei Indikatorvariablen verfügen somit, relativ zum Benchmark Modell, über einen stark erhöhten SE, Modelle mit vier Indikatorvariablen über einen leicht erhöhten SE. Schätzungen der Parameter sind somit umso weniger präzise, desto weniger Indikatoren an der Schätzung beteiligt sind, wobei vor allem die Verwendung von drei statt fünf Indikatoren zu einer substanziellen Erhöhung des SE führt. Die Fallzahl wiederum hat einen negativen Einfluss auf

die Höhe des SE innerhalb der getesteten Modelle mit b^* = -0,244 und b = -0,468. Desto höher die Fallzahl, desto geringer somit der durchschnittliche SE und desto höher die Präzision.

Der Haupteffekt von y_{j-1} kann demnach relativ stabil und präzise geschätzt werden. Die Präzision der Schätzung wird jedoch auch substanziell durch den Anteil der systematischen Fehlervarianz an der gesamten Fehlervarianz, die Fallzahl und die Modellkomplexität (bei Verwendung von drei statt fünf Indikatorvariablen) beeinflusst.

Tabelle 24: Durchschnittliche SE der Effekte von β_2 ($\overline{SE}_{\xi_1}$)

	b	**SE**	**b***
a	1,023***	0,006	NA
n	-0,226***	0,003	-0,690
$q = 3$	0,058***	0,006	0,060
$q = 4$	0,026***	0,006	0,027
λ_{sim}	-0,012***	< 0,001	-0,460
$\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$	-0,001***	< 0,001	-0,062
int	< 0,001***	< 0,001	0,046

a: Konstante der Regressionsschätzung; **n**: Effekt der Fallzahl auf SE der Effekte; **q**: Einfluss der Anzahl an Indikatoren innerhalb des geschätzten Modells auf die SE der Effekte; λ_{sis}: Einfluss der wahren Faktorladung bzw. des Anteils an Common Variance auf die SE der Effekte; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$**:** Einfluss des Anteils des systematischen Messfehlers am gesamten Messfehler auf die SE der Effekte; **int**: Einfluss der Interaktion von λ_{sis} und $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ auf die SE der Effekte;

b***: p ≤ 0,001; b**: p ≤ 0,01; b*: p ≤ 0,05

a für n = 1500; p = 5; λ = 0,6; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ = 0,5

Nach Analyse des SE desjenigen Haupteffektes, welcher gemäß Beschreibung in Kapitel 3 unmittelbar relevant für die Identifikation und Interpretation der FRE ist (Haupteffekt von y_{j-1}), werden die SE derjenigen Haupteffekte analysiert, welche zwecks Berücksichtigung aller möglichen Interaktionsterme als Kontrollvariablen innerhalb der Modelle spezifiziert wurden (SE der Effekte von ξ_1 und ξ_2). Die durchschnittlich geschätzten SE und Verzerrungsquellen der Effekte von ξ_1 (vgl. Tabelle 24) und ξ_2 (vgl. Tabelle 25) sind diesbezüglich derart ähnlich, dass deren Analyse im Folgenden zusammengefasst wird.

Aus Konstante *a* geht ein durchschnittlicher SE (mal 100) für beide Effekte innerhalb der Benchmark Analyse von 1,023 hervor, was einem SE von 0,0102 entspricht. Der SE ist somit ebenfalls relativ gering und insbesondere geringer als der SE des Haupteffektes von y_{j-1}. Die Haupteffekte von ξ_1 und ξ_2 können somit für Benchmark Modelle ebenfalls präzise geschätzt werden.

Tabelle 25: Durchschnittliche SE der Effekte von β_3 ($\overline{SE}_{\xi_2}$)

	b	**SE**	**b***
a	1,024***	0,006	NA
n	-0,226***	0,003	-0,690
***q* = 3**	0,057***	0,006	0,059
***q* = 4**	0,026***	0,006	0,026
$\boldsymbol{\lambda_{sim}}$	-0,012***	< 0,001	-0,460
$\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$	-0,001***	< 0,001	-0,062
int	< 0,001***	< 0,001	0,046

a: Konstante der Regressionsschätzung; **n**: Effekt der Fallzahl auf SE der Effekte; **q**: Einfluss der Anzahl an Indikatoren innerhalb des geschätzten Modells auf die SE der Effekte; $\boldsymbol{\lambda_{sis}}$: Einfluss der wahren Faktorladung bzw. des Anteils an Common Variance auf die SE der Effekte; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$**:** Einfluss des Anteils des systematischen Messfehlers am gesamten Messfehler auf die SE der Effekte; **int**: Einfluss der Interaktion von λ_{sis} und $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ auf die SE der Effekte;

b***: p ≤ 0,001; b**: p ≤ 0,01; b*: p ≤ 0,05

a für $n = 1500$; $p = 5$; $\lambda = 0{,}6$; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}} = 0{,}5$

Die Einflüsse des (approximierten) Anteils an Common Variance ($b^* = -0{,}460$, $b = -0{,}012$), des Anteils des systematischen Messfehlers ($b^* = -0{,}062$, $b = -0{,}001$) und des Interaktionseffektes der beiden Variablen ($b^* = 0{,}046$, $b = 0{,}000$) sind zwar statistisch signifikant und lassen sich derart interpretieren, dass je höher die systematischen Varianzanteile ausfallen und desto geringer somit der Varianzanteil des zufälligen Messfehlers ist, desto geringer der SE und desto höher die Präzision ausfällt. Jedoch sind die Effekte absolut betrachtet sehr gering und damit wenig substanziell, so dass von einer konstant stabilen Schätzung des SE unabhängig von der simulierten Höhe der einzelnen Varianzanteile ausgegangen werden kann.

Der Einfluss der Fallzahl ist hingegen substanziell und negativ mit b^* = -0,690 und b = -0,226. Bei einer Erhöhung der Fallzahl verringert sich die durchschnittlich geschätzte SE, die Präzision des Schätzers erhöht sich mit zusätzlichen Befragten. Der Einfluss der Modellkomplexität hingegen ist insgesamt gering mit b^* = 0,03 und b = 0,03 für q = 4 Modelle und für q = 3 Modelle mit b^* = 0,06 und b = 0,06. Entsprechend lässt sich interpretieren, dass je geringer die Anzahl an Indikatoren q, desto höher der SE und desto weniger präzise die Schätzung.

Die beiden Haupteffekte können somit über alle Simulationen relativ stabil und präzise geschätzt werden, wobei hauptsächlich die Fallzahl einen substanziellen Einfluss auf die Präzision der Schätzung hat.

SE der 2-Wege-Interaktionen der NLC-CFA

Nach Analyse der Präzision der Haupteffekte wird die Präzision der 2-Wege-Interaktionseffekte der Kontextfrage mit den exogenen latenten Variablen $y_{j-1}\xi_1$ und $y_{j-1}\xi_2$ sowie der exogenen latenten Variablen miteinander ($\xi_1\xi_2$) analysiert. Da sich die Regressionseffekte zwischen den beiden Interaktionseffekten von $y_{j-1}\xi_1$ und $y_{j-1}\xi_2$ kaum unterscheiden (vgl. Tabelle 26 für den SE der Effekte von $y_{j-1}\xi_1$ und Tabelle 27 für den SE der Effekte von $y_{j-1}\xi_2$), wird die Analyse der SE beider Interaktionseffekte im Folgenden erneut zusammengefasst. Bezüglich des durchschnittlichen SE der Effekte innerhalb des Benchmark Modells über alle Gruppen hinweg indizieren die Konstanten a einen durchschnittlichen Schätzwert von etwa 0,84, was einem zu erwartenden durchschnittlichen SE von 0,0084 entspricht, so dass für Benchmark Modelle und die entsprechenden Effekte geschlossen werden kann, dass der SE als äußerst gering geschätzt wird, was erneut einer sehr präzisen Schätzung entspricht. Wird darüber hinaus berücksichtigt, dass die beiden zugehörigen Interaktionseffekte innerhalb des Benchmark Modells in der Gruppe mit simuliertem Assimilationseffekt mit -0,1 simuliert wurden (vgl. Kapitel 4.1 und 4.2), so ist der Effekt bei einem zu erwartenden durchschnittlichen SE von 0,0084 theoretisch statistisch signifikant identifizierbar.

Daneben hat sowohl der (approximierte) Anteil an Common Variance (b^* = -0,435, b = -0,010), der Anteil an systematischer Fehlervarianz (b^* = 0,21, b = -0,004) sowie die Interaktion der beiden Variablen (b^* = 0,06, b < 0,001) einen statistisch signifikanten Einfluss auf die Höhe des SE, auch wenn die Effekte insgesamt derart gering sind, dass sie im Grunde vernachlässigt werden können. Inhaltlich formal kann dennoch interpretiert werden, dass je geringer der Anteil an Common Variance (bei konstantem Anteil an systematischer Fehlervarianz) ausfällt, was in einem insgesamt stärkeren FRE resultiert, desto höher der SE und damit umso geringer die

Präzision ist. Ein höherer Anteil an Common Variance verringert somit den SE und erhöht die Präzision. Ebenfalls gilt, dass umso höher der Anteil an systematischer Fehlervarianz bei konstantem Anteil an Common Variance, was ebenfalls in einem stärkeren FRE resultiert, desto geringer der SE und desto höher somit die Präzision ist. Die Präzision des Schätzers ist somit unabhängig von der Stärke des FRE und ist stattdessen umso höher, desto höher die systematischen Varianzanteile (approximierter Anteil an Common Variance und Anteil an systematischer Fehlervarianz) bzw. desto geringer der Varianzanteil des zufälligen Messfehlers ausfällt. Dieser Zusammenhang ist jedoch, wie bereits besprochen, kaum substanziell.

Tabelle 26: Durchschnittliche SE der Effekte von β_4 ($\overline{SE}_{y_{j-1}\xi_1}$)

	b	**SE**	**b***
a	0,846***	0,005	NA
n	-0,193***	0,002	-0,668
***q* = 3**	0,128***	0,006	0,150
***q* = 4**	0,036***	0,005	0,042
$\boldsymbol{\lambda_{sim}}$	-0,010***	< 0,001	-0,435
$\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$	-0,004***	< 0,001	-0,208
int	< 0,001***	< 0,001	0,061

a: Konstante der Regressionsschätzung; **n**: Effekt der Fallzahl auf SE der Effekte; **q**: Einfluss der Anzahl an Indikatoren innerhalb des geschätzten Modells auf die SE der Effekte; $\boldsymbol{\lambda_{sis}}$: Einfluss der wahren Faktorladung bzw. des Anteils an Common Variance auf die SE der Effekte; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$**:** Einfluss des Anteils des systematischen Messfehlers am gesamten Messfehler auf die SE der Effekte; **int**: Einfluss der Interaktion von λ_{sis} und $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ auf die SE der Effekte;

b***: p ≤ 0,001; b**: p ≤ 0,01; b*: p ≤ 0,05

a für n = 1500; p = 5; λ = 0,6; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ = 0,5

Daneben hat die Modellkomplexität ebenfalls einen Einfluss auf die Höhe des SE mit b^* = 0,13 und b = 0,15 für q = 3 Modelle und b^* = 0,04 und $b \approx 0{,}04$ für q = 4 Modelle. Relativ zum Benchmark Modell ist der SE entsprechend umso höher, desto weniger Indikatoren an der Schätzung beteiligt sind. Die Verwendung von weniger als fünf Indikatorvariablen führt somit zu einer geringeren Präzision der Schätzung. Dabei ist jedoch ebenfalls zu bemerken, dass der Einfluss erneut als sehr gering und damit als wenig substanziell interpretierbar ist.

Tabelle 27: Durchschnittliche SE der Effekte von β_5 ($\overline{SE}_{y_{j-1}\xi_2}$)

	b	SE	b*
a	0,844***	0,005	NA
n	-0,193***	0,002	-0,668
$q = 3$	0,130***	0,006	0,152
$q = 4$	0,038***	0,005	0,045
λ_{sim}	-0,010***	< 0,001	-0,435
$\frac{\sigma^2_{sys}}{\sigma^2_{sys} + \sigma^2_{rand}}$	-0,004***	< 0,001	-0,209
int	< 0,001***	< 0,001	0,062

a: Konstante der Regressionsschätzung; **n**: Effekt der Fallzahl auf SE der Effekte; **q**: Einfluss der Anzahl an Indikatoren innerhalb des geschätzten Modells auf die SE der Effekte; λ_{sis}: Einfluss der wahren Faktorladung bzw. des Anteils an Common Variance auf die SE der Effekte; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$: Einfluss des Anteils des systematischen Messfehlers am gesamten Messfehler auf die SE der Effekte; **int**: Einfluss der Interaktion von λ_{sis} und $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ auf die SE der Effekte;

b***: p ≤ 0,001; b**: p ≤ 0,01; b*: p ≤ 0,05

a für n = 1500; p = 5; λ = 0,6; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ = 0,5

Für die Höhe des SE der beiden Effekte liegt für die Fallzahl ebenfalls ein statistisch signifikanter Effekt vor mit b^* = 0,668 und b = -0,193. Eine Erhöhung der Fallzahl verringert die SE der Schätzer und erhöht somit die Präzision der Schätzung der Effekte.

Die Präzision der beiden 2-Wege-Interaktionseffekte wird somit lediglich durch die Fallzahl substanziell beeinflusst. Die Effekte können ansonsten stabil, konsistent und vor allem präzise innerhalb der simulierten Daten geschätzt werden.

Zudem wird der durchschnittliche SE des Interaktionseffektes von $\xi_1\xi_2$ sowie dessen Einflussfaktoren analysiert (vgl. Tabelle 28), welcher zwecks Kontrolle der übrigen Haupt- und Interaktionseffekte innerhalb des NLC-CFA Modells spezifiziert und geschätzt wurde. Gemäß Konstante *a* beträgt der durchschnittliche SE des Interaktionseffektes von $\xi_1\xi_2$ für das Benchmark Modell 1,340 bzw. umgerechnet (da der SE mit 100 multipliziert wurde) 0,0134. Entsprechend wird der SE für Benchmark Modelle und den entsprechenden Effekt als gering geschätzt, was einer präzisen Schätzung entspricht, auch wenn der SE mit dem des Schätzers des Effektes von y_{j-1} vergleichbar ist und somit zu den höchsten, hier geschätzten SE gehört.

Tabelle 28: Durchschnittliche SE der Effekte von β_6 ($\overline{SE}_{\xi_1\xi_2}$)

	b	**SE**	**b***
a	1,340***	0,008	NA
n	-0,292***	0,004	-0,591
***q* = 3**	0,019*	0,009	0,013
***q* = 4**	0,024**	0,008	0,017
$\boldsymbol{\lambda_{sim}}$	-0,025***	< 0,001	-0,622
$\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$	-0,001***	< 0,001	-0,030
int	< 0,001	< 0,001	0,003

a: Konstante der Regressionsschätzung; **n**: Effekt der Fallzahl auf SE der Effekte; **q**: Einfluss der Anzahl an Indikatoren innerhalb des geschätzten Modells auf die SE der Effekte; $\boldsymbol{\lambda_{sis}}$: Einfluss der wahren Faktorladung bzw. des Anteils an Common Variance auf die SE der Effekte; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$**:** Einfluss des Anteils des systematischen Messfehlers am gesamten Messfehler auf die SE der Effekte; **int**: Einfluss der Interaktion von λ_{sis} und $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ auf die SE der Effekte;

b***: p ≤ 0,001; b**: p ≤ 0,01; b*: p ≤ 0,05

a für n = 1500; p = 5; λ = 0,6; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ = 0,5

Die Einflüsse des (approximierten) Anteils an Common Variance (b^* = -0,622, b = -0,025) und des Anteils an systematischer Fehlervarianz (b^* = -0,030, b = -0,001) sind signifikant und negativ, jedoch auch absolut betrachtet sehr gering und wenig substanziell, die Interaktion der beiden Variablen verfügt über keinen statistisch signifikanten Effekt. Es lässt sich somit zwar erneut theoretisch-inhaltlich interpretieren, dass je höher die systematischen Varianzanteile und desto geringer der relative Varianzanteil des zufälligen Messfehlers, desto geringer der SE und desto höher die Präzision. Die Effekte sind jedoch ebenfalls derart gering, dass die Einflüsse kaum substanziell sind.

Daneben hat die Fallzahl einen negativen und substanziellen Einfluss auf die Höhe des SE mit b^* = -0,591 und b = -0,292. Erneut zeigt sich somit, dass je höher die Fallzahl, desto geringer der SE und desto präziser die Schätzung. Hinsichtlich der Modellkomplexität hat die Verwendung von vier statt fünf Indikatorvariablen einen signifikanten, aber sehr geringen Einfluss mit b^* = 0,017 und b = 0,024, während die Verwendung von drei statt fünf Indikatorvariablen einen geringeren Einfluss mit b^* = 0,013 und b = 0,019 nimmt. Die Verwendung von drei oder vier

statt fünf Indikatorvariablen erhöht somit den SE geringfügig, auch wenn der Einfluss kaum als substanziell interpretiert werden kann.

Die Präzision der 2-Wege-Interaktion wird somit im Wesentlichen durch die Fallzahl beeinflusst. Der Effekt kann ansonsten stabil und präzise innerhalb der simulierten Daten geschätzt werden.

SE der 3-Wege-Interaktion der NLC-CFA

Nach Analyse der Präzision der Haupt- und 2-Wege-Interaktionseffekte wird nachfolgend der SE der 3-Wege-Interaktion von $y_{j-1}\xi_1\xi_2$ (vgl. Tabelle 29) betrachtet. Gemäß Konstante *a* ist ein durchschnittlicher SE von 0,903 (entspricht umgerechnet einem SE von 0,0090, da der Wert für die Analyse mit 100 multipliziert wurde) zu erwarten, welcher relativ vergleichbar ist mit den SE der vorangegangenen Effekte, was in einer vergleichbaren Präzision des Effektes für Benchmark Modelle resultiert. Auch die Schätzung der 3-Wege-Interaktion kann somit für Benchmark Modelle als Präzise interpretiert werden. Wird zudem berücksichtigt, dass die zugehörige 3-Wege-Interaktion innerhalb des Benchmark Modells in der Gruppe mit simuliertem Kontrasteffekt mit 0,2 simuliert wurden (vgl. Kapitel 4.1 und 4.2), so ist dieser Effekt bei einem zu erwartenden SE von 0,0090 theoretisch statistisch signifikant identifizierbar.

Wie für den SE der Interaktionsterme von $y_{j-1}\xi_1$ und $y_{j-1}\xi_2$ ist auch die Höhe des SE der 3-Wege-Interaktion substanziell durch die Fallzahl bedingt (b^* = 0,647, b = -0,197). Bei einer Erhöhung der Fallzahl verringert sich die durchschnittlich geschätzte SE, die Präzision des Schätzers erhöht sich somit mit zusätzlichen Befragten.

Darüber hinaus hat der (approximierte) Anteil an Common Variance einen negativen Einfluss auf die Höhe des SE (b^* = -0,485, b = -0,012), der Anteil des systematischen Messfehlers hat ebenfalls einen negativen Einfluss (b^* = -0,130, b = -0,002), während die Interaktion der beiden Variablen über keinen statistisch signifikanten Einfluss verfügt. Analog zur Diskussion der 2-Wege-Interaktionen lassen sich die Befunde entsprechend erneut derart interpretieren, dass je höher die systematischen Varianzanteile ausfallen und desto geringer somit der Varianzanteil des zufälligen Messfehlers, desto geringer der SE und desto höher die Präzision. Absolut betrachtet sind die Effekte jedoch erneut derart gering, dass an dieser Stelle von keinem substanziellen Effekt ausgegangen werden kann

Tabelle 29: Durchschnittliche SE der Effekte von β_7 ($\overline{SE}_{y_{j-1}\xi_1\xi_2}$)

	b	SE	b*
a	0,903***	0,006	NA
n	-0,197***	0,002	-0,647
$q = 3$	0,067***	0,006	0,074
$q = 4$	-0,002	0,006	-0,002
λ_{sim}	-0,012***	< 0,001	-0,485
$\frac{\sigma^2_{sys}}{\sigma^2_{sys} + \sigma^2_{rand}}$	-0,002***	< 0,001	-0,130
int	< 0,001	< 0,001	-0,008

a: Konstante der Regressionsschätzung; **n**: Effekt der Fallzahl auf SE der Effekte; **q**: Einfluss der Anzahl an Indikatoren innerhalb des geschätzten Modells auf die SE der Effekte; λ_{sis}: Einfluss der wahren Faktorladung bzw. des Anteils an Common Variance auf die SE der Effekte; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$: Einfluss des Anteils des systematischen Messfehlers am gesamten Messfehler auf die SE der Effekte; **int**: Einfluss der Interaktion von λ_{sis} und $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ auf die SE der Effekte;

b***: p ≤ 0,001; b**: p ≤ 0,01; b*: p ≤ 0,05

a für $n = 1500$; $p = 5$; $\lambda = 0{,}6$; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}} = 0{,}5$

Die Modellkomplexität zuletzt hat erneut einen Einfluss auf die Höhe des SE mit $b^* = 0{,}074$ und $b = 0{,}067$ für $q = 3$ Modelle. Der durchschnittliche SE bei Verwendung von drei Indikatorvariablen ist somit relativ zum Benchmark Modell erhöht, während der SE bei Verwendung von vier Indikatoren vergleichbar mit der Verwendung von fünf Indikatorvariablen ist. Die Schätzung wird somit relativ zum Benchmark Modell weniger präzise bei Verwendung von drei Indikatorvariablen, auch wenn der Effekt insgesamt sehr gering und damit wenig substanziell ist.

Die Präzision der geschätzten 3-Wege-Interaktion ist somit insgesamt ebenfalls im Wesentlichen durch die Fallzahl beeinflusst. Der Effekt kann ansonsten innerhalb der simulierten Daten konstant präzise geschätzt werden.

Präzision der Effektschätzungen innerhalb der NLC-CFA

Bezogen auf die SE-Schätzer des Benchmark Modells lässt sich grundsätzlich zusammenfassen, dass diese für alle Effekte eine präzise Schätzung implizieren. Vorhandene, simulierte Effekte sind entsprechend der Höhe des zu erwartenden SE und des daraus resultierenden t-Wertes statistisch signifikant identifizierbar.

Bezüglich der Stärke des simulierten FRE lässt sich für den Haupteffekt von y_{j-1} festhalten, dass dieser mit einem umso geringeren SE geschätzt wird, desto höher der Anteil der systematischen Fehlervarianz, sprich, desto stärker der FRE simuliert wurde. Für alle anderen Effekte sind, bezogen auf die absolute Effektstärke, keine substanziellen Einflüsse des (approximierten) Anteils an Common Variance bzw. der systematischen Fehlervarianz auf die Höhe des SE zu beobachten. In der Tendenz gilt lediglich, dass der SE umso geringer geschätzt wird, desto höher der Anteil an Common Variance bzw. desto höher der Anteil an systematischer Fehlervarianz, was in beiden Fällen gleichbedeutend ist mit einer Verringerung des relativen Anteils der Varianz des zufälligen Messfehlers. Eine höhere Residualvarianz verringert also tendenziell die Präzision der Schätzung.

Hinsichtlich der Modellkomplexität zeigt sich hingegen ein uneinheitliches Bild. So werden die Haupteffekte ξ_1 und ξ_2 mit umso höherem SE und entsprechend umso geringerer Präzision geschätzt, desto weniger Indikatoren an der Schätzung beteiligt sind, während der Haupteffekt von y_{j-1} lediglich bei Verwendung von drei statt fünf Indikatoren mit substanziell erhöhtem SE geschätzt wird. Für die sonstigen Effekte ist, bezogen auf die absolute Effektstärke, kein substanzieller Einfluss der Modellkomplexität auf die Präzision der Schätzung festzustellen. Die Fallzahl zuletzt hat über alle Effekte hinweg einen gleichgerichteten und substanziellen Einfluss auf den SE und damit die Präzision: Erhöht sich die Fallzahl, erhöht sich die Präzision der Schätzung. Dabei ist die Fallzahl für alle Schätzer der relevanteste Einflussfaktor für die Präzision.

SE der Messfehlerkovarianzen der CFA-MTMM

Über die Präzisionsanalyse der Schätzer der NLC-CFA hinaus wird ebenfalls die Präzision der geschätzten Messfehlerkovarianzen der CFA-MTMM analysiert. Wie bereits in Kapitel 4.1 und 4.2 beschrieben erfolgt die Analyse auf Basis der konkreten SE der Messfehlerkovarianzen je Gruppe, welche direkt aus den geschätzten CFA-MTMM Modellen extrahiert wurden. Die jeweiligen SE wurden erneut mit 100 multipliziert, um die Interpretation zu vereinfachen. Alle SE sind, inklusive der Rückführung auf die variierenden Simulationsspezifikationen, in Tabelle

30 für Gruppe 1 (simuliert mit Assimilationseffekt), in Tabelle 31 für Gruppe 2 (simuliert mit Kontrasteffekt) sowie in Tabelle 32 für Gruppe 3 (ohne FRE simuliert) dargestellt. Der Wert in Konstante *a* entspricht wie in allen vorangegangenen Analysen dem durchschnittlich geschätzten SE der Messfehlerkovarianz für Benchmark Modelle, während die übrigen Regressionskoeffizienten die durchschnittlich zu erwartende Veränderung des SE angeben.

Tabelle 30: Durchschnittlicher SE der Messfehlerkovarianz in Gruppe 1

	b	**SE**	**b***
a	0,188***	0,002	NA
n	-0,069***	0,001	-0,398
***q* = 3**	0,317***	0,003	0,616
***q* = 4**	0,085***	0,002	0,165
$\boldsymbol{\lambda_{sim}}$	-0,007***	< 0,001	-0,457
$\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$	-0,003***	< 0,001	-0,302
int	< 0,001***	< 0,001	0,143

Gruppe 1: Gruppe mit simuliertem Assimilationseffekt; **a**: Konstante der Regressionsschätzung; **n**: Effekt der Fallzahl auf SE der Effekte; **q**: Einfluss der Anzahl an Indikatoren innerhalb des geschätzten Modells auf die SE der Effekte; $\boldsymbol{\lambda_{sis}}$: Einfluss der wahren Faktorladung bzw. des Anteils an Common Variance auf die SE der Effekte; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$**:** Einfluss des Anteils des systematischen Messfehlers am gesamten Messfehler auf die SE der Effekte; **int**: Einfluss der Interaktion von λ_{sis} und $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ auf die SE der Effekte;

b***: p ≤ 0,001; b**: p ≤ 0,01; b*: p ≤ 0,05

a für *n* = 1500; *p* = 5; λ = 0,6; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ = 0,5

Für Gruppe 1 des Benchmark Modells (innerhalb welcher ein Assimilationseffekt in Form von zwei 2-Wege-Interaktionen simuliert wurde) ist gemäß Konstante *a* im Schnitt ein SE von 0,0019 zu erwarten (Wert in Tabelle wurde vor der Analyse mit 100 multipliziert), was in Anbetracht eines Varianzbereichs der zugehörigen Indikatoren um den Wert 1 einem geringen SE der Messfehlerkovarianz entspricht. Der (approximierte) Anteil der Common Variance hat einen negativen Einfluss auf den SE mit *b** = 0,457 und *b* = -0,007, wobei zu berücksichtigen ist, dass dieser Wert mit 100 multipliziert wurde. Bezogen auf den tatsächlichen SE müsste sich der (approximierte) Anteil an Common Variance somit um 10 Prozentpunkte verringern, damit sich der SE auf 0,0026 erhöht. Wird bedacht, welcher große Unterschied ein Anteil an Common

Variance von 10 Prozentpunkten ist und um welche kleinen, tatsächlichen Veränderungen eines ansonsten generell kleinen SE es sich hierbei handelt, kann der Einfluss des Anteils an Common Variance absolut betrachtet als gering interpretiert werden. Gleiches gilt entsprechend für den Ebenfalls negativen und noch geringeren Effekt des Anteils an systematischer Fehlervarianz von $b^* = -0{,}302$ und $b = -0{,}003$ sowie den positiven Effekt der Interaktion der beiden Variablen von $b^* = 0{,}143$ ($b < 0{,}001$). Auch wenn die Effekte insgesamt sehr gering sind und der Effekt der Interaktion sogar vernachlässigbar gering ist, können sich die Effektrichtungen derart interpretieren lassen, dass je höher die systematischen Varianzanteile ausfallen und desto geringer somit der Varianzanteil des zufälligen Messfehlers ist, desto geringer der SE und desto höher die Präzision ausfällt.

Die Fallzahl hingegen hat einen negativen und substanziellen ($b = -0{,}069$ und $b^* = -0{,}398$) Einfluss auf die Höhe des SE. Je höher also die Fallzahl, desto geringer der SE. Den stärksten Einfluss hat hingegen die Modellkomplexität mit $b = 0{,}317$ und $b^* = 0{,}616$ für Modelle mit drei statt fünf Indikatorvariablen und $b = 0{,}085$ und $b^* = 0{,}165$ für Modelle mit vier statt fünf Indikatorvariablen. Je weniger Indikatorvariablen an der Schätzung beteiligt sind, desto höher ist also der SE und desto geringer entsprechend die Präzision der Schätzung.

Der durchschnittliche SE der Messfehlerkovarianz für Gruppe 2 (innerhalb welcher ein Kontrasteffekt als 3-Wege-Interaktion simuliert wurde) des Benchmark Modells entspricht gemäß Tabelle 31 $a = 0{,}0025$ (Wert in Tabelle wurde vor der Analyse mit 100 multipliziert), wodurch dieser größer ist als der SE in Gruppe 1. Trotz dessen ist auch dieser SE in Anbetracht eines Varianzbereichs der zugehörigen Indikatorvariablen, welcher um den Wert 1 schwankt, als gering zu betrachten, so dass auch für Gruppe 2 von einer Präzisen Schätzung ausgegangen werden kann.

Die Effekte des (approximierten) Anteils an Common Variance ($b = -0{,}009$ und $b^* = 0{,}584$), des Anteils der systematischen Fehlervarianz ($b = -0{,}002$ und $b^* = 0{,}134$) sowie der Interaktion zwischen beiden Variablen (nicht signifikant) haben für den SE der Messfehlerkovarianz in Gruppe 2 die gleiche Richtung bzw. eine vergleichbare Stärke mit den Effekten in Gruppe 1, so dass auch hier eine vergleichbare Interpretation angewandt werden kann: Zwar verringert sich der SE mit einem verringerten Anteil an zufälliger Fehlervarianz innerhalb der simulierten Indikatoren, jedoch ist dieser Einfluss sehr gering und damit wenig substanziell.

Tabelle 31: Durchschnittlicher SE der Messfehlerkovarianz in Gruppe 2

	b	**SE**	**b***
a	0,254***	0,003	NA
n	-0,078***	0,001	-0,418
$q = 3$	0,285***	0,003	0,519
$q = 4$	0,073***	0,003	0,133
λ_{sim}	-0,009***	< 0,001	-0,584
$\frac{\sigma^2_{sys}}{\sigma^2_{sys} + \sigma^2_{rand}}$	-0,002***	< 0,001	-0,134
int	< 0,001	< 0,001	-0,006

Gruppe 2: Gruppe mit simuliertem Kontrasteffekt; **a**: Konstante der Regressionsschätzung; **n**: Effekt der Fallzahl auf SE der Effekte; **q**: Einfluss der Anzahl an Indikatoren innerhalb des geschätzten Modells auf die SE der Effekte; λ_{sis}: Einfluss der wahren Faktorladung bzw. des Anteils an Common Variance auf die SE der Effekte; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$**:** Einfluss des Anteils des systematischen Messfehlers am gesamten Messfehler auf die SE der Effekte; **int**: Einfluss der Interaktion von λ_{sis} und $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ auf die SE der Effekte;
b***: p ≤ 0,001; b**: p ≤ 0,01; b*: p ≤ 0,05
a für n = 1500; p = 5; λ = 0,6; $\frac{\sigma^2_{sys}}{\sigma^2_{sys}+\sigma^2_{rand}}$ = 0,5

Sowohl für die Fallzahl, als auch für die Modellkomplexität sind in Gruppe 2 ebenfalls mit Gruppe 1 vergleichbare Effekte identifizierbar, so dass auch hier die Interpretation analog verläuft: Je höher die Fallzahl, desto geringer der SE und je weniger Indikatorvariablen an der Schätzung beteiligt sind, desto höher ist der SE und desto geringer entsprechend die Präzision der Schätzung. Beide Effekte sind stark genug, dass sie als substanziell bezeichnet werden können.

In Gruppe 3 zuletzt wurde kein FRE simuliert, die spezifizierte Messfehlerkovarianz entspricht hier also einer Fehlspezifikation, wodurch sich ein durchschnittlicher SE von 0,0044 (Wert in Tabelle wurde vor der Analyse mit 100 multipliziert) für das Benchmark Modell erklärt, was der Höchste SE aller drei Gruppen ist. Der SE verringert sich insgesamt, wenn sich der Anteil der zufälligen Fehlervarianz verringert, wobei vor allem ein zunehmender (approximierter) Anteil an Common Variance einen Einfluss auf die Reduktion des SE nimmt, auch wenn der absolute Einfluss der Varianzanteile insgesamt als gering betrachtet werden kann. Daneben gilt erneut, dass je höher die Fallzahl, desto geringer der SE sowie, dass je weniger Indikatoren an der Schätzung beteiligt sind, desto höher ist der SE.

Tabelle 32: Durchschnittlicher SE der Messfehlerkovarianz in Gruppe 3

	b	**SE**	**b***
a	0,437***	0,003	NA
n	-0,125***	0,002	-0,476
***q* = 3**	0,346***	0,004	0,445
***q* = 4**	0,103***	0,004	0,132
$\boldsymbol{\lambda_{sim}}$	-0,014***	< 0,001	-0,631
$\frac{\sigma_{sys}^2}{\sigma_{sys}^2 + \sigma_{rand}^2}$	< 0,001*	< 0,001	-0,010
int	< 0,001**	< 0,001	-0,015

Gruppe 3: Gruppe ohne simulierten FRE; **a**: Konstante der Regressionsschätzung; **n**: Effekt der Fallzahl auf SE der Effekte; **q**: Einfluss der Anzahl an Indikatoren innerhalb des geschätzten Modells auf die SE der Effekte; $\boldsymbol{\lambda_{sis}}$: Einfluss der wahren Faktorladung bzw. des Anteils an Common Variance auf die SE der Effekte; $\frac{\sigma_{sys}^2}{\sigma_{sys}^2+\sigma_{rand}^2}$**:** Einfluss des Anteils des systematischen Messfehlers am gesamten Messfehler auf die SE der Effekte; **int**: Einfluss der Interaktion von λ_{sis} und $\frac{\sigma_{sys}^2}{\sigma_{sys}^2+\sigma_{rand}^2}$ auf die SE der Effekte;

b***: p ≤ 0,001; b**: p ≤ 0,01; b*: p ≤ 0,05

a für n = 1500; p = 5; λ = 0,6; $\frac{\sigma_{sys}^2}{\sigma_{sys}^2+\sigma_{rand}^2}$ = 0,5

Zusammenfassend lässt sich für die Präzision der geschätzten Messfehlerkovarianzen festhalten, dass der SE in den Benchmark Modellen für Gruppe 1 (mit Assimilationseffekt simuliert) am geringsten und für Gruppe 3 (ohne FRE simuliert) am höchsten ist, auch wenn der SE nach wie vor in allen drei Gruppen als gering betrachtet werden kann. Scheinbar hat die Fehlspezifikation des Modells in Gruppe 3 einen stärkeren, negativen Einfluss auf die Präzision der Schätzung als eine komplexere Systematik des zugrundeliegenden systematischen Messfehlers in Gruppe 2 gegenüber Gruppe 1. Daneben haben die verschiedenen Varianzanteile in allen drei Gruppen einen geringen, aber vorhandenen Einfluss auf die Höhe des SE in dem Sinne, dass je geringer der Anteil des zufälligen Messfehlers, desto geringer ist auch der SE. Die Fallzahl hat einen substanziellen, negativen Einfluss auf den SE was bedeutet, dass desto höher die Fallzahl, desto höher die Präzision der Schätzung. Daneben hat auch die Modellkomplexität einen substanziellen Einfluss auf die Verzerrung in dem Sinne, dass je weniger Indikatoren an der Schätzung beteiligt sind, desto weniger Präzise wird die Schätzung.

Gegenüberstellung und Fazit

Für die Präzision der Schätzungen der konkreten Effekte mittels NLC-CFA (auf Basis von regressionsbasierten Haupt- und Interaktionseffekten) sowie mittels CFA-MTMM (mittels Messfehlerkovarianzen) lässt sich insgesamt feststellen, dass die Effekte im Wesentlichen präzise, das heißt mit relativ geringem SE, geschätzt werden können. Die Höhe der geschätzten SE impliziert zudem für Benchmark Modelle, dass die konkreten Haupt- und Interaktionseffekte der NLC-CFA, welche den FRE quantifizieren, theoretisch statistisch signifikant identifiziert werden können.

Auffällig ist vor allem für CFA-MTMM Modelle, dass der SE in den Benchmark Modellen für Gruppe 1 (mit Assimilationseffekt simuliert) am geringsten und für Gruppe 3 (ohne FRE simuliert) am höchsten ist, was eventuell auf die Fehlspezifikation des Modells in Gruppe 3 und eine komplexere Systematik des zugrundeliegenden systematischen Messfehlers in Gruppe 2 gegenüber Gruppe 1 zurückgeführt werden kann. Über Benchmark Modelle hinaus ist es vor allem die Fallzahl, welche einen substanziellen Einfluss auf die Höhe des SE, sowohl in NLC-CFA Modellen als auch in CFA-MTMM Modellen in dem Sinne hat, dass je höher die Fallzahl, desto höher die Präzision der Schätzung. Auch die Modellkomplexität weißt sowohl in NLC-CFA Modellen, als auch in CFA-MTMM Modellen über einen substanziellen Einfluss auf den SE auf. In beiden Modellen führt vor allem die Verwendung von drei statt fünf Indikatorvariablen zu einer substanziellen Erhöhung des SE, während die Verwendung von vier statt fünf Indikatorvariablen auch in CFA-MTMM Modellen zu einer geringfügigen Erhöhung des SE führt. Bezüglich des Einflusses des (approximierten) Anteils an Common Variance und des Anteils des systematischen Messfehlers ist zuletzt sowohl für NLC-CFA Modelle, als auch für CFA-MTMM Modelle tendenziell zu beobachten, dass ein höherer, absoluter Anteil des zufälligen Messfehlers zu einem höheren SE und damit einer geringeren Präzision führt. In beiden Modellen sind die entsprechenden Effekte (mit Ausnahme des Effektes des Anteils der systematischen Fehlervarianz für den Haupteffekt von y_{j-1} in NLC-CFA Modellen) jedoch relativ gering, so dass diese kaum als substanziell betrachtet werden können.

Es lässt sich somit zusammenfassend festhalten, dass alle analysierten Effekte für Benchmark Modelle präzise geschätzt werden können. In der Tendenz verringert sich die Präzision bei einer geringeren Anzahl an Indikatorvariablen, während sich die Präzision aller Effekte bei Erhöhung der Fallzahl erhöht. Ein geringer Anteil der Varianz des zufälligen Messfehlers wirkt sich tendenziell ebenfalls positiv auf die Präzision der Schätzer aus.

4.7 Detaillierte Analyse der angepassten Benchmark Modelle

Abschließend wird auf Basis der Analysen der Verzerrungen und der Standardfehler zwecks Illustration der konkreten Schätzwerte innerhalb der Gruppen und der Richtung der Abweichungen das bisherige Benchmark Modell angepasst und innerhalb einer weiteren Simulationsstudie analysiert (siehe Anhang 7.2). Die simulierten Daten aus i = 1000 Iterationen werden dabei genutzt, um einfache CFA Modelle ohne weitere Spezifikationen, NLC-CFA Modelle sowie CFA-MTMM Modelle je Iteration zu schätzen, wobei alle drei Modelle als Multigruppenmodelle auf Basis der gleichen, simulierten Daten geschätzt werden. Die einfachen CFA Modelle wurden mit konfiguraler-, die CFA-MTMM und die NLC-CFA Modelle mit metrischer Invarianz geschätzt. Die Modellspezifikation entspricht dabei für alle Iterationen dem Benchmark Modell gemäß den vorangegangenen Kapiteln mit Ausnahme einer Anpassung der Modellkomplexität. Dies bedeutet, dass alle Daten und alle Modelle in allen Iterationen und Gruppen mit einer Fallzahl n = 1500, einer Anzahl an Indikatoren von q = 4 (statt q = 5), einer unstandardisierten Faktorladung von 0,6 und einem relativen Anteil des systematischen Messfehlers am gesamten Messfehler von 50% simuliert und geschätzt wurden. Die Anpassung der Modellkomplexität durch Reduktion der verwendeten Indikatoren von fünf auf vier ergibt sich, da hierfür innerhalb der Effekt Effizienz Analyse eine geringere Verzerrung der Schätzer identifiziert wurde und es neben den Faktorladungen vor allem die Effektschätzer (Regressionskoeffizienten und Messfehlerkovarianzen) selbst sind, welche nachfolgend analysiert werden. Innerhalb der Verzerrungsanalyse der Faktorladungen wurde vor allem identifiziert, dass Modelle mit mindestens vier Indikatorvariablen geschätzt werden sollten, um zu unverzerrten Schätzungen der Faktorladungen zu gelangen, so dass diesbezüglich keine Probleme aufgrund der Anpassung des Modells zu erwarten sind. Daneben wurde innerhalb der Präzisionsanalyse für NLC-CFA Modelle eine Erhöhung der geschätzten SE vor allem bei Verwendung von drei Indikatoren identifiziert, so dass eine Verwendung von vier Indikatoren hinsichtlich dieser Modelle keine Verringerung der Präzision impliziert, wodurch eine Anpassung des Benchmark Modells für NLC-CFA Modelle hinsichtlich der Präzision ebenfalls unproblematisch ist. Anders ist dies hingegen für CFA-MTMM Modelle, für welche auch bei Verwendung von vier statt fünf Indikatoren eine geringfügige Erhöhung der SE der geschätzten Messfehlerkovarianzen erwartet werden kann. Dies ist bei der nachfolgenden Analyse zu berücksichtigen.

Die zugrundeliegende Simulation impliziert für NLC-CFA Modelle in Gruppe 1 (simuliert mit Assimilationseffekt) einen wahren Effekt von -0,1 für β_4 und β_5, in Gruppe 2 (simuliert mit Kontrasteffekt) einen wahren Effekt von 0,2 für β_7 und einen wahren Effekt von 0 für alle

sonstigen Effekte in allen sonstigen Gruppen. Für CFA-MTMM Modelle wird hingegen eine Messfehlerkovarianz von 0 erwartet, da sowohl Assimilations- als auch Kontrasteffekte konditional in Abhängigkeit zweier exogener Variablen ξ_1 und ξ_2 simuliert wurden, welche normalverteilt um den Mittelwert 0 streuen, wobei für die mittlere Merkmalsausprägung kein FRE simuliert wurde. Bei Konstanthaltung von ξ_1 und ξ_2 durch eine fehlende Berücksichtigung der jeweiligen Effekte innerhalb der CFA-MTMM kann somit nur der Effekt für $\xi_1 = \xi_2 = 0$ geschätzt werden, welcher, da konditional, ebenfalls 0 beträgt, so dass auch die Messfehlerkovarianzen trotz vorhandener FRE 0 betragen müssen (vgl. Kapitel 4.1). Für die Interpretation der Ergebnisse ist zudem hervorzuheben, dass für Gruppe 3 in keiner der Iterationen ein FRE simuliert wurde. Für Schätzer der Faktorladungen und der Effekte werden, da es sich um eine Monte Carlo Analyse ohne variierende Simulationsspezifikationen handelt, die Mittelwerte der jeweiligen Schätzer aus allen Iterationen präsentiert, der 95% Konfidenzintervall (KI95) der jeweiligen Schätzer, berechnet auf Basis der einzelnen Schätzer der 1000 Iterationen, wird ebenfalls angegeben. Die Bestimmung des Signifikanzniveaus der jeweiligen Effektschätzer wird diesbezüglich aus dem Mittelwert der t-Werte der jeweiligen Effektschätzer aus allen Iterationen abgeleitet. Alle unstandardisierten Faktorladungen sind analog zu den vorangegangenen Analysen als approximierter Anteil an Common Variance bzw. als Annäherung an die standardisierte Faktorladung des entsprechenden Indikators zu interpretieren. Die unstandardisierten Regressionseffekte der NLC-CFA können als Anteil der abhängigen Variablen, welche durch den entsprechenden Faktor oder die entsprechende Interaktion bedingt ist, interpretiert werden. Die Höhe der geschätzten Messfehlerkovarianzen wiederum sind dahingehend einzuordnen, dass die Varianz der jeweiligen Indikatorvariablen je nach Iteration um den Wert 1 schwankt.

Analyse der Faktorladungen der angepassten Benchmark Modelle

Beginnend mit der Analyse der Faktorladungen sind diese als Mittelwerte inklusive 95% Konfidenzintervall je Iteration für die drei einzelnen Gruppen der einfachen CFA Modelle (getestet mit konfiguraler Invarianz und ohne gesonderte Spezifikation des FRE) sowie für die NLC-CFA Modelle und die CFA-MTMM Modelle (getestet mit metrischer Invarianz, insofern keine Differenzierung zwischen den Gruppen) in Tabelle 33 abgebildet. Zu beachten ist dabei, dass für Gruppe 3 kein FRE bzw. keine systematische Abhängigkeit der Indikatoren von weiteren, über das latente Konstrukt hinausgehenden Faktoren simuliert wurde. Bei den Faktorladungen für Gruppe 3 handelt es sich somit um Extraktionen, die keinerlei systematischen Verzerrungen unterliegen, während die Faktorladungen in Gruppe 1 und 2 sowohl auf Basis der Common

Variance, als auch auf Basis der unkontrollierten, systematischen Fehlervarianz geschätzt sind, wodurch Verzerrungen gegenüber Gruppe 3 erwartet werden können. Bei den Faktorladungen der NLC-CFA sowie der CFA-MTMM handelt es sich schließlich um Schätzungen (über alle drei Gruppen hinweg) unter Kontrolle der simulierten systematischen Fehlervarianz, namentlich des FRE.

Tabelle 33: Geschätzte Faktorladungen der angepassten Benchmark Analyse

λ_j	**CFA Gruppe 1 (KI95)**	**CFA Gruppe 2 (KI95)**	**CFA Gruppe 3 (KI95)**	**NLC-CFA (KI95)**	**CFA-MTMM (KI95)**
λ_1	0,595 (0,594 - 0,596)	0,587 (0,586 - 0,588)	0,600 (0,599 - 0,601)	0,601 (0,601 - 0,602)	0,594 (0,593 - 0,594)
λ_2	0,606 (0,606 - 0,607)	0,612 (0,611 - 0,612)	0,600 (0,599 - 0,601)	0,604 (0,604 - 0,605)	0,606 (0,605 - 0,606)
λ_3	0,618 (0,617 - 0,619)	0,638 (0,637 - 0,639)	0,600 (0,599 - 0,601)	0,605 (0,604 - 0,605)	0,625 (0,624 - 0,625)
λ_4	0,619 (0,618 - 0,620)	0,648 (0,647 - 0,65)	0,600 (0,599 - 0,601)	0,604 (0,604 - 0,605)	0,626 (0,625 - 0,626)
$\frac{\sum_{j=1}^{4}\lambda_j}{4}$	0,610	0,621	0,600	0,604	0,613

CFA Gruppe 1: Schätzung innerhalb der Gruppe mit simuliertem, aber im Modell nicht spezifizierten Assimilationseffekt (mit konfiguraler Invarianz); **CFA Gruppe 2**: Schätzung innerhalb der Gruppe mit simuliertem, aber im Modell nicht spezifiziertem Kontrasteffekt (mit konfiguraler Invarianz); **CFA Gruppe 3**: Schätzung innerhalb der Gruppe ohne simulierten und ohne spezifizierten FRE (mit konfiguraler Invarianz); **NLC-CFA**: Schätzung auf Basis des NLC-CFA Modells gemäß Kapitel 3.3.2 und 4.1 (mit metrischer Invarianz); **CFA-MTMM**: Schätzung auf Basis des CFA-MTMM Modells gemäß Kapitel 3.3.1 und 4.1 (mit metrischer Invarianz); **KI 95**: 95% Konfidenzintervall der Schätzer auf Basis der Iterationen der Simulation; λ_j: Mittelwert der unstandardisierten Faktorladung über alle Iterationen der Simulation; alle unstandardisierten Faktorladungen λ sind statistisch hoch signifikant mit $p < 0{,}001$

Die durchschnittliche Faktorladung für die vier Items aus Gruppe 3 (keine FRE simuliert oder kontrolliert) beträgt 0,6, wobei dieser Wert tatsächlich im Durchschnitt aller Iterationen für alle Items geschätzt wurde. Der 95% Konfidenzintervall kann für alle Items insgesamt als sehr schmal eingeschätzt werden. Hieraus ergibt sich zum einen eine konsistente Schätzung der Faktorladungen über alle Iterationen hinweg, zum anderen kann die Schätzung als äußerst effizient und damit unverzerrt betrachtet werden.

Hiermit vergleichend beträgt die durchschnittliche Faktorladung für Gruppe 1, innerhalb welcher Assimilationseffekte simuliert, im Modell aber nicht kontrolliert wurden, 0,610 mit einer

Spanne von 0,595 (λ_1) bis 0,619 (λ_4). Die schmalen Konfidenzintervalle indizieren erneut für alle Items eine konsistente Schätzung über alle Iterationen. Auffällig ist, dass die geschätzten Faktorladungen von Item 1 (minimale Unterschätzung) bis Item 4 (geringfügige Überschätzung) tendenziell ansteigen und somit (wenn auch lediglich geringfügig) verzerrt sind. Die durch FRE bedingten Messfehler scheinen sich entsprechend in aufeinanderfolgenden Items zu kumulieren, so dass die Verzerrung der Faktorladung von Item zu Item zunimmt, was in Kapitel 5.3 vertiefend diskutiert wird.

Die durchschnittliche Faktorladung für Gruppe 2 (Kontrasteffekte wurden simuliert, jedoch nicht kontrolliert) beträgt über alle Items hinweg 0,621 mit einer Spanne von 0,587 (λ_1) bis 0,648 (λ_4). Es ist somit eine ähnliche Verzerrungssystematik wie für Gruppe 1 vorhanden, obwohl die Verzerrungen in Gruppe 2 insgesamt stärker sind als in Gruppe 1. Der enge Konfidenzintervall der Faktorladungen impliziert darüber hinaus erneut eine konsistente Schätzung über alle Iterationen. Es wiederholt sich somit das Muster, dass die Faktorladung von Item 1 bis Item 4 zunehmen. Das Vorhandensein von FRE führt somit in Modellen, in welchen dieser nicht kontrolliert wird und auf Basis der simulierten Daten tatsächlich zu einer Verzerrung der Faktorladungen. In der durchgeführten Simulation kumulieren sich Verzerrungen darüber hinaus von Item zu Item, was mit zunehmenden Verzerrungen von Faktorladung zu Faktorladung einhergeht. Somit ist Evidenz dafür gegeben, dass der Anteil des systematischen Messfehlers, gegeben durch FRE, bei fehlender statistischer Kontrolle inkorrekt als Anteil der Common Variance geschätzt wird, was in verzerrten Faktorladungen resultiert.

Die durchschnittliche Faktorladung der NLC-CFA, innerhalb welcher FRE mittels regressiver Haupt- und Interaktionseffekte kontrolliert sind, beträgt 0,604 mit einer Spanne von 0,601 (λ_1) bis 0,605 (λ_3) mit einem erneut sehr schmalen Konfidenzintervall für alle Items. Faktorladungen sind somit im Schnitt lediglich um den Wert 0,004 verzerrt, wenn FRE mittels NLC-CFA kontrolliert werden. Darüber hinaus ist die Verzerrung der Faktorladungen bei gegebener statistischer Kontrolle deutlich geringer, als wenn vorhandene FRE nicht kontrolliert werden, wie in CFA Gruppe 1 und Gruppe 2 der Fall. Tatsächlich ist die Verzerrung der Faktorladungen bei gegebener statistischer Kontrolle derart gering, dass die geschätzten Faktorladungen annährend mit denjenigen Faktorladungen vergleichbar sind, die auf Basis von Daten geschätzt wurden, welche keine FRE enthalten (CFA Gruppe 3). Insofern kann geschlossen werden, dass die statistische Kontrolle von FRE tatsächlich zu einer Bereinigung verzerrter Faktorladungen führt, wenn diese eigentlich durch FRE verzerrt sind. Interessant ist daneben, dass die Faktorladungen ähnlich wie in den vorangegangenen Analysen von Gruppe 1 und Gruppe 2 zunächst ansteigen, dann aber in λ_3 mit 0,605 den höchsten Wert erreichen und in λ_4 wieder geringfügig auf 0,604

abnehmen. Eine Messfehlerkumulation wie für die vorangegangenen Gruppen diskutiert scheint sich hier also für weitere Indikatoren zu bereinigen.

Die durchschnittliche Faktorladung der CFA-MTMM, innerhalb welcher FRE als Messfehlerkovarianzen spezifiziert sind, beträgt insgesamt 0,613 mit einer Spanne von 0,594 (λ_1) bis 0,626 (λ_4). Der Konfidenzintervall der Faktorladungen impliziert zudem erneut eine konsistente Schätzung über alle Iterationen. Es lässt sich somit beobachten, dass Faktorladungen stärker verzerrt sind als bei Kontrolle durch die NLC-CFA. Zudem sind die Faktorladungen, mit metrischer Invarianz über alle drei Gruppen hinweg geschätzt, ebenfalls stärker verzerrt, als wenn die Faktorladungen lediglich innerhalb von Gruppe 1 (mit Assimilationseffekt) simuliert, dort aber nicht kontrolliert werden. Eine Bereinigung findet entsprechend nur gegenüber Gruppe 2 (Kontrasteffekt simuliert und nicht kontrolliert) statt, innerhalb welcher ein stärkerer FRE mit höherer Komplexität simuliert wurde als in Gruppe 1. Es wiederholt sich zudem das Muster, dass die Faktorladung von Item 1 bis Item 4 zunehmen. Es ist somit für Benchmark Modelle Evidenz gegeben, dass die CFA-MTMM vor allem dazu geeignet ist, Verzerrungen auf Basis schwerwiegender systematischer Messfehler mit erhöhter Effektstärke zu bereinigen, insgesamt jedoch zu weniger effizient geschätzten Faktorladungen führt als bei Verwendung der NLC-CFA.

Analyse der Effekte der angepassten Benchmark Modelle

Als nächstes zur Analyse stehen die spezifizierten Effekte innerhalb des NLC-CFA Modells und des CFA-MTMM Modells. Die Analyse erfolgt differenziert nach den drei Gruppen *z*. In Gruppe 1 (simuliert mit Assimilationseffekt) ist diesbezüglich für die NLC-CFA ein wahrer Effekt für β_4 und β_5 von -0,1 simuliert, welche den Einfluss der Interaktionsterme von $y_{j-1}\xi_1$ und $y_{j-1}\xi_2$ auf die Zielfrage y_j repräsentieren. In Gruppe 2 (simulierter Kontrasteffekt) ist für die NLC-CFA ein wahrer Effekt für β_7 von 0,2 simuliert, welcher den Einfluss der 3-Wege-Interaktion von $y_{j-1}\xi_1\xi_2$ auf die Zielfrage y_j repräsentiert. Für Gruppe 3 (ohne FRE simuliert) ist für die NLC-CFA hingegen insgesamt, ebenso wie für allen sonstigen, hier nicht genannten Variablen, kein Einfluss simuliert. Für die Messfehlerkovarianz der CFA-MTMM ist theoriebedingt für keine der Gruppen ein Effekt zu erwarten. In Tabelle 34 sind die Mittelwerte der geschätzten Effekte über alle Iterationen inklusive des 95% Konfidenzintervalls abgebildet.

Tabelle 34: Geschätzte Effekte der angepassten Benchmark Analyse

Effekt	Gruppe 1		Gruppe 2		Gruppe 3	
	Wahrer Effekt	Geschätzter Effekt (KI95)	Wahrer Effekt	Geschätzter Effekt (KI95)	Wahrer Effekt	Geschätzter Effekt (KI95)
β_1	0	-0,004 (-0,004 - -0,003)	0	-0,004 (-0,005 - -0,003)	0	-0,003 (-0,004 - -0,003)
β_2	0	0 (-0,001 - 0,001)	0	0 (-0,001 - 0)	0	0 (0 - 0,001)
β_3	0	0 (0 - 0,001)	0	0 (-0,001 - 0,001)	0	0 (-0,001 - 0)
β_4	-0,1	-0,067*** (-0,068 - -0,067)	0	0 (-0,001 - 0)	0	0 (-0,001 - 0,001)
β_5	-0,1	-0,068*** (-0,068 - -0,067)	0	0 (0 - 0,001)	0	0 (-0,001 - 0)
β_6	0	0 (-0,001 - 0)	0	0 (-0,001 - 0)	0	0,001 (0 - 0,002)
β_7	0	-0,001 (-0,001 - 0)	0,2	0,144*** (0,143 - 0,145)	0	0 (0 - 0,001)
$\varepsilon_{j+1,j}$	0	-0,002 (-0,002 - -0,002)	0	-0,003 (-0,003 - -0,002)	0	-0,001 (-0,001 - -0,001)

Gruppe 1: Schätzung innerhalb der Gruppe mit simuliertem Assimilationseffekt; **Gruppe 2**: Schätzung innerhalb der Gruppe mit simuliertem Kontrasteffekt; **Gruppe 3**: Schätzung innerhalb der Gruppe ohne simulierten FRE; **Wahrer Effekt**: Tatsächlicher, simulierter Effekt in den einzelnen Iterationen der Simulation; **Geschätzter Effekt**: Mittelwert der geschätzten Effekte über alle Iterationen der Simulation; **KI 95**: 95% Konfidenzintervall der geschätzten Effekte auf Basis der Iterationen der Simulation; β_b: Unstandardisierter Regressionseffekt der NLC-CFA; $\varepsilon_{j+1,j}$: Unstandardisierte Messfehlerkovarianz der CFA-MTMM; *Effekt****: $p \leq 0{,}001$; *Effekt* **: $p \leq 0{,}01$; *Effekt* *: $p \leq 0{,}05$

Beginnend mit Gruppe 1 zeigt sich, dass keiner der Effekte mit Ausnahme der Effekte von β_4 und β_5 statistisch signifikant ist. Der Effekt von β_4 und von β_5 wird für Gruppe 1 mit -,067 bzw. -0,068 ($p < 0{,}001$) geschätzt, beide Effekte verfügen über einem sehr schmalen 95% Konfidenzintervall und sind somit konsistent über alle Iterationen geschätzt. Die Effekte sind dabei etwa um 0,033 bzw. 0,032 unterschätzt, was annähernd mit der geschätzten Abweichung der Effekt Effizienz Analyse (vgl. Kapitel 4.5) vergleichbar ist. Dennoch werden die Effekte, auch wenn geringfügig verzerrt, mit korrektem Vorzeichen und als statistisch signifikant geschätzt, während alle anderen Effekte in Gruppe 1 ebenfalls korrekt als nicht statistisch signifikant geschätzt werden, was auch für die Messfehlerkovarianz der CFA-MTMM gilt. Für Gruppe 1

liegt somit für keinen der Effekte ein Typ 1 oder Typ 2 Fehler vor, die Schätzung erfolgt für beide Modelle präzise und im Wesentlichen effizient.

Für Gruppe 2 zeigt sich ebenfalls, wie es den durch die Simulation bedingten wahren Werten entspricht, lediglich ein signifikanter Effekt für β_7 ($p < 0{,}001$). Dieser beträgt 0,144 mit einem erneut sehr schmalen Konfidenzintervall, was eine konsistente Schätzung des Effektes über alle Iterationen hinweg impliziert. Der Effekt wird somit ebenfalls unterschätzt. Darüber hinaus entspricht auch diese Verzerrung annähernd der Verzerrung, welche in der Effekt Effizienz Analyse geschätzt wurde, auch wenn die Verzerrung gemäß der dort identifizierten Effektstärken durch eine Anpassung der Modellkomplexität geringer ausfallen müsste, als es hier tatsächlich der Fall ist. Dennoch zeigt sich für keinen der Regressionseffekte der NLC-CFA sowie ebenfalls nicht für die Messfehlerkovarianz der CFA-MTMM in Gruppe 2 ein Typ 1 oder Typ 2 Fehler, wodurch es sich alles in allem um eine präzise Schätzung handelt, welche jedoch in ihrer Effizienz beeinträchtigt ist.

Für Gruppe 3 zuletzt wird korrekterweise keiner der Effekte als statistisch signifikant geschätzt. Die Schätzung für Gruppe 3 kann somit insgesamt als Präzise und Effizient angesehen werden.

Gegenüberstellung und Fazit

Bezüglich der beobachteten Faktorladungen lässt sich zunächst feststellen, dass diese, wo unkontrollierte FRE vorhanden sind, verzerrt sind. Bei fehlender Kontrolle von FRE steigen die Faktorladungen vom ersten bis zum letzten Item sukzessiv an, so dass von einer Messfehlerkumulation durch den FRE in nachfolgenden Items ausgegangen werden kann. Die Anwesenheit von FRE bei fehlender Kontrolle führt somit tatsächlich zu verzerrten Schätzungen der Faktorladungen. Findet hingegen eine statistische Kontrolle von FRE mittels NLC-CFA statt, so wird die Verzerrung der Faktorladungen im Wesentlichen bereinigt. Eine Kontrolle von FRE auf Basis der NLC-CFA führt somit zu effizienteren Schätzungen von Faktorladungen. Eine Kontrolle von FRE auf Basis des CFA-MTMM Modell ist hingegen vor allem dazu geeignet, Verzerrungen auf Basis schwerwiegender systematischer Messfehler mit erhöhter Effektstärke zu bereinigen. Die CFA-MTMM führt dabei jedoch insgesamt zu weniger effizient geschätzten Faktorladungen als die NLC-CFA

Bezüglich der beobachteten Effekte der NLC-CFA Modelle selbst lässt sich insgesamt feststellen, dass auch wenn diese in der Tendenz unterschätzt werden, keine Typ 1 oder Typ 2 Fehler auftreten. Die Effektschätzungen sind somit teilweise effizient, hingegen jedoch präzise hinsichtlich der allgemeinen Identifizierbarkeit der Effekte auf Basis der geschätzten t-Werte der

jeweiligen Effekte. Die Messfehlerkovarianzen der CFA-MTMM werden daneben ebenfalls korrekt als statistisch nicht signifikant geschätzt. Daneben geht die Verwendung der CFA-MTMM für die hier simulierten Daten jedoch mit einem Informationsverlust gegenüber der NLC-CFA einher, da auf Basis der so geschätzten Messfehlerkovarianzen keine relevanten Informationen bezüglich bestehender FRE gegeben sind. Hierbei handelt es sich jedoch nicht um eine generelle Schwäche der CFA-MTMM, sondern um eine Besonderheit der vorliegenden Datensimulation und -Aggregation. Wären das zugrundeliegende Gruppendesign beispielsweise mit weitere Gruppen entsprechend eines dichotomisierten ξ_1 bzw. ξ_2 multipliziert worden, so wären die konditional simulierten FRE durchaus theoretisch identifizierbar gewesen. Da dies die Komplexität der vorliegenden Analyse jedoch um ein Vielfaches erhöht hätte, wurde hierauf bewusst verzichtet.

Insgesamt lässt sich somit für das angepasste Benchmark Modell und auf Basis der simulierten Daten festhalten, dass Faktorladungen bei Anwesenheit und fehlender Kontrolle von FRE verzerrt sind, dass das NLC-CFA Modell dazu in der Lage ist, diese Verzerrungen fast vollständig zu bereinigen und dass das NLC-CFA Modell für diese Aufgabe dem CFA-MTMM Modell deutlich überlegen ist. Konkret quantifizierte FRE können innerhalb des NLC-CFA Modells darüber hinaus präzise und hinreichend effizient geschätzt werden.

5 Diskussion und Zusammenfassung

Ziel dieser Arbeit ist es, das Auftreten von FRE theoretisch fundiert modellhaft darzustellen sowie auf Basis dieser theoretischen Fundierung ein innovatives CFA Modell (NLC-CFA Modell) vorzuschlagen, mit welchem FRE innerhalb von Fragebatterien sowohl identifiziert, als auch kontrolliert werden können, um so zu messfehlerbereinigten Schätzungen zu gelangen. Das vorgeschlagene CFA Modell soll dabei einem etablierten CFA Modell (CFA-MTMM Modell) gegenübergestellt werden, um den Mehrwert des Modells aufzuzeigen.

Diesbezüglich erfolgte in Kapitel 2 zunächst eine allgemeine theoretische Diskussion des Befragtenverhaltens aus kognitionspsychologischer, soziologischer und methodologischer Perspektive. Nachfolgend wurden FRE innerhalb der jeweils diskutierten Theorien und Modelle eingeordnet, um deren verallgemeinerbare kognitive Einflussfaktoren, welche sich zwischen den verschiedenen, diskutierten Modellen identifizieren lassen, innerhalb von Thesen zu aggregieren. Dabei wurde identifiziert, dass aus kognitiver Perspektive vor allem der Modus der Informationsverarbeitung, der Zugang von Informationen bezüglich des Objekts der Befragung sowie das Framing der Befragtenrolle einen konkreten Einfluss auf das Befragtenverhalten und somit das Auftreten von FRE nehmen. Schließlich wurden eben jene Thesen verwendet, um ein einfaches Erklärungsmodell für das Auftreten von FRE zu synthetisieren, aus welchem empirisch überprüfbare, theoretische Hypothesen abgeleitet wurden.

Nachfolgend wurden in Kapitel 3 sowohl etablierte, als auch innovative statistische Verfahren und Modelle diskutiert, welche zur Analyse und Bereinigung von FRE genutzt werden und genutzt werden können. Konkret wurde diesbezüglich die Möglichkeit aufgezeigt, wie FRE mithilfe der etablierten MTMM Analyse (und eines adaptierten CFA-MTMM Modells), aber auch mittels eines neu eingeführten NLC-CFA Modells (welches multiple latente Interaktionseffekte auf Indikatorebene eines Multigruppen-CFA Modells utilisiert) identifiziert und kontrolliert werden können, was theoretisch fundiert zu einer Bereinigung der Messfehler und somit zu einer verbesserten Messung der Datenqualität führen sollte. Von dem NLC-CFA Modell wird dabei insofern eine höhere Messfehlerbereinigung als von dem CFA-MTMM Modell erwartet, als dass innerhalb der NLC-CFA die drei in Kapitel 2 identifizierten Einflussfaktoren (Modus der Informationsverarbeitung, Zugänglichkeit von Informationen, Framing der Befragtenrolle) gemäß ihrer konkreten, auf theoretischer Ebene identifizierten Einflusssystematik durch Pfadmodellierung berücksichtigt werden können, was innerhalb des CFA-MTMM Mo-

dells in dieser Form nicht möglich ist. Innerhalb des NLC-CFA Modells können somit potenziell mehr gegebene Informationen bei der Schätzung von Parametern berücksichtigt werden, was du effizienteren und präziseren Schätzungen führen sollte.

In Kapitel 4 wurde entsprechend eine erweiterte Monte-Carlo Analyse mit variierenden Simulations- und Modellspezifikationen je Iteration durchgeführt, um die Tauglichkeit des eingeführten NLC-CFA Modells sowie des vorgeschlagenen CFA-MTMM Modells zur Identifikation von FRE und zur Bereinigung von Parameterverzerrungen zu überprüfen. Die je Iteration auftretenden Verzerrungen der Faktorladungen und der Effektkoeffizienten sowie die Höhe der geschätzten Standardfehler der einzelnen Modelle wurden diesbezüglich per Regression auf die variierenden Simulations- und Modellspezifikationen zurückgeführt, um Einflussfaktoren der Effizienz und Präzision der jeweiligen Schätzer zu identifizieren und die Tauglichkeit der Modelle zu evaluieren. Eine Gegenüberstellung der beiden Modelle fand darüber hinaus ebenfalls statt. Auf Basis der Analyseergebnisse wurden die beiden Modelle abschließend angepasst und erneut innerhalb einer Monte-Carlo Simulation, jedoch ohne variierende Simulations- und Modellspezifikationen, getestet, um zu abschließenden Erkenntnissen hinsichtlich der Präzision und Effizienz der vorgeschlagenen Modelle zu gelangen.

Die Ergebnisse dieser Analysen werden nachfolgend innerhalb ihres methodischen (vgl. Kapitel 5.1) sowie ihres theoretischen (vgl. Kapitel 5.2) Kontextes eingeordnet und diskutiert, bevor Randbedingungen, Limitationen und Möglichkeiten bei der konkreten Anwendung von NLC-CFA Modellen in empirischen Untersuchungen diskutiert werden (vgl. Kapitel 5.3). Eine Gesamtzusammenfassung bildet den Abschluss dieser Arbeit (vgl. Kapitel 5.4).

5.1 Diskussion hinsichtlich der methodischen Implikationen

Es ist eine der einleitend formulierten Ziele dieser Arbeit, ein statistisches Verfahren zur Identifikation und Kontrolle von FRE in beliebigen Fragebatterien zu entwickeln (das NLC-CFA Modell), welches zu einer höheren Bereinigung von durch FRE bedingte Verzerrungen in der Lage ist als etablierte Verfahren (das CFA-MTMM Modell). Diesbezüglich wurde diskutiert, dass vor allem Faktorladungen der Items solcher Fragebatterien als Indikatoren der Konstruktvalidität und der Komposit-Reliabilität durch FRE verzerrt werden. Es besteht somit die implizite Annahme, dass unkontrollierte FRE Faktorladungen verzerren sowie die impliziten Forderungen, dass besagte Verzerrung durch Schätzung eines NLC-CFA Modells (zumindest teilweiße) bereinigt werden kann und dass die Bereinigung durch NLC-CFA Modelle höher ist als

die Bereinigung durch etablierte CFA-MTMM Modelle. NLC-CFA Modelle sollen darüber hinaus den Mehrwert bieten, die konkrete Höhe der FRE zu quantifizieren und die zugrundeliegenden Wirkmechanismen dieses systematischen Messfehlers somit sichtbar zu machen. All diese Punkte wurden in Kapitel 4 innerhalb mehrerer Simulationsstudien untersucht, die Ergebnisse der Analysen werden nachfolgend zusammengefasst und diskutiert.

Bereinigung der Faktorladungen

Bezüglich der Bereinigung von Faktorladungen geht aus der Analyse der Verzerrungen von Faktorladungen (vgl. Kapitel 4.4) für alle innerhalb der Monte-Carlo Analyse berücksichtigten Modell- und Simulationsspezifikationen hervor, 1.) dass Faktorladungen dort verzerrt geschätzt werden, wo FRE erwartet und nicht kontrolliert werden; 2.) dass die Schätzung eines NLC-CFA Modells für die simulierten Daten zu einer fast vollständigen Bereinigung der Verzerrungen führt; und 3.) dass NLC-CFA Modelle den etablierten CFA-MTMM Modellen hinsichtlich der Bereinigung von Faktorladungen überlegen sind. Verzerrungen der Faktorladungen wurden diesbezüglich als Standard-Verzerrungen (*SV*) operationalisiert, wobei diese als 100 Mal die Wurzel der durchschnittlichen quadrierten Abweichung der geschätzten Faktorladung von der wahren Faktorladung innerhalb einer Gruppe *z* interpretiert werden können. Die Multiplikation mit 100 wird für die Präsentierten Schätzwerte in der nachfolgenden Diskussion jedoch aufgehoben, um die Interpretation der Höhen der Verzerrungen relativ zur wahren simulierten Faktorladung korrekt einschätzen zu können.

Für das Benchmark Modell (entspricht einem Modell mit fünf Indikatorvariablen, einer Fallzahl von 1500, einer simulierten wahren Faktorladung von 0,6 und einem simulierten Anteil der systematischen Messfehlervarianz an der gesamten Messfehlervarianz von 50%) wurde entsprechend eine *SV* von 0,020 geschätzt, wenn keine FRE simuliert und auch kein statistisches Verfahren zu deren Kontrolle angewandt wurde (entspricht „Gruppe 3“ des einfachen CFA Modells in Kapitel 4.4). Entsprechend kann dieser Wert für Benchmark Modelle als *SV* für korrekt spezifizierte Modelle ohne systematische Messfehler in den zugrundeliegenden Daten angenommen werden, was insgesamt einer geringen Verzerrung entspricht, wie ebenfalls aus der Betrachtung der konkreten Faktorladungen für Benchmark Modelle aus Kapitel 4.7 (detaillierte Analyse der angepassten Benchmark Modelle) hervorgeht.

Werden hingegen FRE als Assimilationseffekte in Form von zwei voneinander unabhängigen 2-Wege-Interaktionen simuliert und werden diese nicht mittels NLC-CFA oder CFA-MTMM kontrolliert (entspricht „Gruppe 1“ der einfachen CFA), so werden Faktorladungen mit einer *SV* von 0,027 geschätzt. Die Verzerrung der Faktorladungen ist also tatsächlich höher, wenn

Assimilationseffekte vorhanden sind, diese jedoch nicht innerhalb des Modells spezifiziert werden bzw. keine statistische Kontrolle dieser stattfindet. Daneben ist jedoch ebenfalls zu beachten, dass auch wenn eine erhöhte Verzerrung gegenüber „Gruppe 1“ (kein FRE simuliert) vorhanden ist, diese nicht sonderlich gravierend ist. Für die Modellschätzung mit unkontrolliertem Assimilationseffekt ist entsprechend lediglich eine geringe Verzerrung relativ zu einem korrekt spezifizierten Modell ohne simulierten FRE zu erwarten.

Demgegenüber werden für Modelle ohne spezifizierten FRE Faktorladungen mit $SV = 0{,}097$ geschätzt, wenn FRE als Kontrasteffekte mittels 3-Wege-Interaktion simuliert wurden (entspricht „Gruppe 2“ der einfachen CFA). Die Verzerrung der Faktorladungen ist also ebenfalls relativ zu „Gruppe 3“ (kein FRE simuliert) höher, wenn Kontrasteffekte vorhanden sind, diese jedoch nicht innerhalb des Modells spezifiziert werden. Wird die konkrete *SV* mit derjenigen aus „Gruppe 3“ verglichen so zeigt sich darüber hinaus, dass für unkontrollierte Kontrasteffekte nicht nur eine geringe, sondern eine gravierende Verzerrung zu erwarten ist. Die Verzerrung der Faktorladungen bei simuliertem FRE auf Basis einer 3-Wege-Interaktion ist entsprechend ebenfalls um ein Vielfaches höher als die Verzerrung der Faktorladungen bei simuliertem FRE auf Basis von zwei voneinander unabhängigen 2-Wege-Interaktionen. Es kann somit geschlossen werden, dass die Verzerrungen der Faktorladungen für Benchmark Modelle umso höher ausfallen, desto komplexer die Systematik der Abhängigkeiten des zugrundeliegenden Messfehlers ist. Eine unberücksichtigte und unspezifizierte 3-Wege-Interaktion erzeugt eine um ein Vielfaches höhere Verzerrung als zwei unspezifizierte und voneinander unabhängige 2-Wege-Interaktionen.

Die Verzerrung bzw. die *SV* für „Gruppe 1“ und für „Gruppe 2“ ist dabei umso höher, je geringer die Fallzahl und für „Gruppe 1“ höher, wenn insgesamt lediglich drei statt fünf Indikatorvariablen an der Schätzung beteiligt sind, demgegenüber jedoch auch geringer, wenn vier statt fünf Indikatorvariablen an der Schätzung beteiligt sind. Die Effekte sind dabei aber absolut betrachtet äußerst gering und damit vernachlässigbar. Für „Gruppe 2“ erhöht sich hingegen die Effizienz der Schätzung erheblich, desto weniger Indikatorvariablen (relativ zum Benchmark Modell mit fünf Indikatorvariablen) an der Schätzung beteiligt sind, so dass hinsichtlich der Modellkomplexität insgesamt geschlossen werden kann, dass die Verwendung von vier statt fünf Indikatorvariablen die *SV* der Faktorladungen für beide Gruppen verringert. Die Es zeigt sich zudem, dass je stärker der simulierte FRE ist, desto höher auch die *SV* relativ zum Benchmark Modell ausfällt, wen keine statistische Kontrolle besagter FRE stattfindet, wobei dieser Zusammenhang für „Gruppe 1“ erneut vernachlässigbar gering ist und lediglich für „Gruppe 2“ relevante Effektstärken aufweist.

Werden vorhandene FRE hingegen auf Basis des vorgeschlagenen CFA-MTMM Modells kontrolliert, so werden die Faktorladungen für Benchmark Modelle mit $SV = 0{,}032$ geschätzt. Eine Bereinigung der Faktorladungen findet somit lediglich relativ zu einer Modellschätzung mit unspezifizierten, aber vorhandenen Kontrasteffekten („Gruppe 2“ der einfachen CFA) statt, während die Schätzung weniger effizient ist als eine Schätzung bei unspezifizierten, aber vorhandenen Assimilationseffekten („Gruppe 1“ der einfachen CFA) sowie bei unspezifizierten und nicht vorhandenen FRE („Gruppe 3“ der einfachen CFA). Bezüglich dieses Befundes wurde diskutiert, dass das CFA-MTMM Modell insbesondere zu einer Bereinigung von Verzerrungen beiträgt, wenn die entsprechenden Verzerrungen auf einen Messfehler mit komplexen Wirkmechanismus (wie einer 3-Wege-Interaktion gegenüber zwei voneinander unabhängigen 2-Wege-Interaktion) zurückgeführt werden kann.

Die Verzerrungen der Faktorladungen innerhalb des CFA-MTMM Modells sind darüber hinaus umso höher, desto stärker der zugrundeliegende FRE simuliert wurde, wobei die entsprechenden Effekte derart gering sind, dass selbst wenn systematische Einflüsse vorhanden sind, diese derart gering sind, dass sie vernachlässigt werden können. Daneben ist es insbesondere die Modellkomplexität, welche einen substanziellen Einfluss auf die Verzerrungen hat: Modelle, welche mit drei statt fünf Indikatoren geschätzt wurden verfügen über eine erhöhte Verzerrung, während die Verzerrung bei Modellen mit vier Indikatoren minimal geringer ist als bei Modellen mit fünf Indikatoren.

Demgegenüber werden die Verzerrungen der geschätzten Faktorladungen fast vollständig bereinigt, wenn eine statistische Kontrolle der FRE mittels NLC-CFA stattfindet. Dieser Schluss ist dadurch begründet, dass die *SV* der Faktorladungen für NLC-CFA Benchmark Modelle als 0,015 geschätzt wird, während die *SV* für Benchmark Modelle bei korrekter Modellspezifikation und ohne simulierten FRE („Gruppe 3“ der einfachen CFA) 0,020 beträgt. Die Schätzung der Faktorladungen ist somit bei gegebener statistischer Kontrolle von FRE mittels NLC-CFA sogar effizienter als die Schätzung von Faktorladungen mittels einfacher CFA (ohne simulierte FRE) für die hier verwendeten Daten und daneben ebenfalls effizienter als die Schätzung von Faktorladungen in Benchmark Modellen, in welchen vorhandene FRE nicht statistisch kontrolliert werden sowie die Schätzung auf Basis des CFA-MTMM Modells.

Daneben erhöht sich die Effizienz der NLC-CFA Schätzung relativ zum Benchmark Modell weiter bei erhöhter Fallzahl, während sich die Effizienz der Schätzung der Faktorladungen bei Verwendung von lediglich drei statt fünf Indikatorvariablen substanziell verringert. Zudem erhöht sich die Effizienz der Schätzung ebenfalls bei einem geringeren Anteil derjenigen Varianz

innerhalb der Indikatorvariablen, welche auf den zufälligen Messfehler zurückgeht, auch wenn dieser Zusammenhang als vernachlässigbar gering angesehen werden kann.

Insgesamt zeigt sich somit, dass Faktorladungen tatsächlich dort verzerrt sind, wo systematische Messfehler (hier: der FRE) vorhanden, jedoch nicht spezifiziert sind, wobei die Höhe der zu erwartenden Verzerrung von der Komplexität der Zusammenhänge und Abhängigkeiten des systematischen Messfehlers abhängig zu sein scheint. Auf Basis des Gruppenvergleiches lässt sich vorsichtig schließen, dass je komplexer die Systematik des zugrundeliegenden, unberücksichtigten Messfehlers, desto höher die Verzerrung. Die Kontrolle von FRE durch das NLC-CFA Modell führt demgegenüber zu einer fast vollständigen Bereinigung der Verzerrung von Faktorladungen und ist darüber hinaus der Kontrolle durch das vorgeschlagene CFA-MTMM Modell überlegen, was bei Verwendung der NLC-CFA insgesamt zu effizienteren Messungen der Indikatoren von Konstruktvalidität und Kompositreliabilität führt. Dabei kann insgesamt die Empfehlung gegeben werden, dass für eine effiziente Bereinigung der Verzerrung eine erhöhte Fallzahl vorhanden sein sollte (n = 1500 für Benchmark Modelle, wobei auch geringere Fallzahlen möglich sind, da sich die *SV* für 1000 Befragte weniger lediglich geringfügig erhöht), sowie dass keine Modelle geschätzt werden sollten, welche lediglich über drei Indikatorvariablen verfügen (die *SV* erhöht sich hierdurch relativ zum Benchmark Modell um 0,042). Muss aus praktischen Gründen hingegen ein CFA-MTMM Modell zur Bereinigung von Faktorladungen verwendet werden so gilt zu Bedenken, dass hierfür lediglich für Messfehler mit komplexer zugrundeliegender Systematik eine effiziente Bereinigung zu erwarten ist.

Präzision und Effizienz der Effektschätzer

Bezüglich der Präzision und der Effizienz der Effektschätzer innerhalb von NLC-CFA Benchmark Modellen geht aus den entsprechenden Analysen (vgl. Kapitel 4.5, Effekt Effizienz Analyse und Kapitel 4.6, Standardfehler Präzisionsanalyse) hervor, dass Effekte im Allgemeinen mit hoher Präzision und mit überwiegend hoher Effizienz geschätzt werden können. Die Messfehlerkovarianzen der CFA-MTMM Benchmarkmodelle werden hingegen alle mit hoher Präzision und Effizienz geschätzt. Verzerrungen wurden diesbezüglich erneut als Standard-Verzerrungen (*SV*) operationalisiert, wobei diese als 100 Mal die Wurzel der durchschnittlichen quadrierten Abweichung der geschätzten Effekte (Regressionseffekte für NLC-CFA Modelle und Messfehlerkovarianzen für CFA-MTMM Modelle) von den wahren Effekten je Effektschätzer über alle drei Modellgruppen z hinweg für NLC-CFA Modelle und innerhalb der jeweiligen Modellgruppen für CFA-MTMM Modelle interpretiert werden können.

Für das NLC-CFA Benchmark Modell implizieren die *SV* der analysierten Effekte grundsätzlich eine relativ geringe Verzerrung und damit eine hohe Effizienz aller Schätzer mit Ausnahme der 3-Wege-Interaktion, welche mit erhöhter Verzerrung geschätzt wird. Analog hierzu ist auch innerhalb der CFA-MTMM die Messfehlerkovarianz in derjenigen Gruppe am stärksten verzerrt (auch wenn die Verzerrungen innerhalb der CFA-MTMM absolut betrachtet äußerst gering ausfallen), innerhalb welcher der FRE als Kontrasteffekt mittels 3-Wege-Interaktion simuliert wurde (Gruppe 2). Eine zu komplexe Systematik des zugrundeliegenden systematischen Messfehlers führt somit in beiden Modellen zu einer Verringerung der Effizienz. Daneben sind die Verzerrungen der Effekte in NLC-CFA Benchmark Modellen am geringsten für Schätzer, welche unabhängig von der Kontextfrage y_{j-1} sind, also auch nicht an Interaktionen mit dieser beteiligt sind und somit nicht durch FRE beeinflusst werden (betrifft ξ_1, ξ_2 und $\xi_1\xi_2$). Effekte, welche unabhängig von systematischen Messfehlern simuliert sind, können somit mit der höchsten Effizienz geschätzt werden. Für die konkreten Effekte, welche die entsprechenden konditionalen Einflüsse der FRE innerhalb der NLC-CFA quantifizieren (die beiden 2-Wege-Interaktionen $y_{j-1}\xi_1$ und $y_{j-1}\xi_2$ sowie die 3-Wege-Interaktion $y_{j-1}\xi_1\xi_2$) zeigt die detaillierte Analyse der angepassten Benchmark Modelle darüber hinaus eine generelle Unterschätzung der Effekte, wobei die entsprechenden Effekte gemäß der analysierten SE dennoch Identifizierbar bleiben. Die Präzisionsanalyse offenbart zudem eine hohe Präzision aller Schätzer innerhalb beider Benchmark Modelle, so dass für beide Benchmark Modelle insgesamt eine präzise und tendenziell effiziente Schätzung der Effekte mit einer beobachtbaren Unterschätzung derjenigen Effekte innerhalb der NLC-CFA, welche den FRE quantifizieren, angenommen werden kann.

Für Variationen beider Benchmark Modelle zeigt sich hinsichtlich der Stärke des zugrundeliegenden FRE bzw. der unterschiedlichen Varianzanteile innerhalb der Indikatorvariablen, dass diese in beiden Modellen weder einen substanziellen Einfluss auf die Effizienz, noch auf die Präzision der Effektschätzungen hat. Daneben ist die Effizienz der Schätzer vor allem substanzielle durch die Modellkomplexität beeinflusst. Die Verwendung von drei statt fünf Indikatorvariablen führt generell zu einer erhöhten Verzerrung, während für NLC-CFA Modelle die Verwendung von vier statt fünf Indikatorvariablen für eine möglichst effiziente Schätzung empfohlen wird, was in CFA-MTMM Modellen die Effizienz zumindest nicht substanziell verringert. Auch die Präzision der Schätzung ist bei Verwendung von drei statt fünf Indikatorvariablen substanziell erhöht, während die Verwendung von vier statt fünf Indikatorvariablen in CFA-MTMM Modellen zu einer geringfügigen Verringerung der Präzision führt. Dies gilt es insgesamt bei der Auswahl der Modelle im Kontext der Anzahl verfügbarer Indikatorvariablen zu

beachten. Die Fallzahl zuletzt hat einen substanziellen Einfluss vor allem auf die Präzision der Schätzung in dem Sinne, dass je höher die Fallzahl, desto präziser die Schätzung. Bezüglich der Präzision der Schätzer fällt zuletzt innerhalb von CFA-MTMM Modellen auf, dass diese für die Schätzung der Messfehlerkovarianz in Gruppe 3 (ohne FRE simuliert) am geringsten in allen CFA-MTMM Modellgruppen ist. Eine Fehlspezifikation des Modells in Gruppe 3, in welcher trotz fehlendem FRE eine Messfehlerkovarianz spezifiziert wurde verringert die Präzision der Schätzung somit in höherem Maße als ein vorhandener und durch eine Messfehlerkovarianz spezifizierter systematischer Messfehler.

Es zeigt sich somit insgesamt, dass die Effekte sowohl für NLC-CFA Modelle, als auch für CFA-MTMM Modelle präzise und ausreichend effizient hinsichtlich der Identifizierbarkeit der tatsächlichen Effekte geschätzt werden können. Einen substanziellen Einfluss auf die Präzision und Effizienz der Schätzungen ist insbesondere durch die Modellkomplexität gegeben, wobei vor allem die Verwendung von drei statt fünf Indikatorvariablen die Effizienz und Präzision der Schätzung in beiden Modellen verringert. Die Verwendung von vier statt fünf Indikatorvariablen erhöht die Effizienz in NLC-CFA Modellen, während sich die Effizienz in CFA-MTMM Modellen hierdurch zumindest nicht substanziell verschlechtert. Die Präzision nimmt bei Verwendung von vier Indikatorvariablen hingegen für beide Modelle geringfügig ab. Die Fallzahl wiederum hat vor allem auf die Präzision der Schätzungen einen substanziellen Einfluss in beiden Modellen in dem Sinne, dass je höher die Fallzahl, desto höher die Präzision. Die konkrete Höhe des FRE bzw. der jeweiligen Varianzanteile innerhalb der Indikatorvariablen haben hingegen im Großen und Ganzen einen vernachlässigbar geringen Einfluss auf Präzision und Effizienz beider Modelle. Dies alles berücksichtigend wird insgesamt ein deutlicher Informationsgewinn bei der Schätzung der Effekte mittels NLC-CFA gegenüber der Schätzung mittels CFA-MTMM gesehen, da innerhalb der NLC-CFA eine konkrete Spezifikation der einzelnen Haupt- und Interaktionseffekte möglich ist, was eindeutige Rückschlüsse auf die zugrundeliegenden Wirkmechanismen systematischer Messfehler zulässt, welche durch das CFA-MTMM Modell in dieser Form nicht möglich sind.

Methodische Gesamtevaluation der durchgeführten Analysen

Insgesamt können auf Basis der durchgeführten Analysen in Kapitel 4 die folgenden methodischen Erkenntnisse hinsichtlich der Anwesenheit von FRE sowie deren statistischer Kontrolle auf Basis des in Kapitel 3 vorgeschlagenen NLC-CFA Modells sowie des CFA-MTMM Modells zusammengefasst werden:

Erstens zeigt sich grundsätzlich, dass vorhandene systematische Messfehler auf Indikatorebene dort Faktorladungen verzerren, wo diese nicht innerhalb der zu schätzenden Modelle spezifiziert werden, da die unberücksichtigte Varianz der Messfehler innerhalb der Indikatoren ansonsten als Common Variance extrahiert wird, wenn mehrere Indikatoren des zu schätzenden latenten Konstruktes durch den systematischen Messfehler beeinflusst werden. Dabei zeigen die Simulationsstudien, dass ein systematischer Messfehler, welcher auf Basis einer 3-Wege-Interaktion simuliert wurde, zu einer höheren Verzerrung der Faktorladungen führt, als wenn der systematische Messfehler auf Basis von zwei 2-Wege-Interaktionen simuliert wurde. Hieraus kann mit Vorsicht abgeleitet werden, dass die Verzerrung der Faktorladungen umso größer zu sein scheint, desto komplexer die zugrundeliegenden Abhängigkeiten eines unspezifizierten, systematischen Messfehlers, auch wenn eine solche Aussage weiterer Forschung bedarf.

Zweitens zeigt sich, dass die Verzerrungen der Faktorladungen innerhalb der in der Simulation analysierten Benchmark Modelle fast vollständig bereinigt werden, wenn auf Basis des in Kapitel 3 vorgeschlagenen NLC-CFA Modell systematische Messfehler (hier: der FRE) auf Indikatorebene kontrolliert werden, was insgesamt zu effizienteren Schätzungen der Faktorladungen führt, wodurch Messungen der Konstruktvalidität und Kompositreliabilität ebenfalls effizienter werden. Die Schätzung von Modellen mit lediglich drei Indikatorvariablen sollte dabei vermieden werden, da dies einen erheblichen Einfluss auf die effiziente Schätzung der Faktorladungen hat.

Drittens zeigt sich, dass die Schätzung von CFA-MTMM Modellen lediglich zu einer teilweisen Bereinigung von stark verzerrten Faktorladungen beitragen. Die Effizienz der Schätzung ist dabei insgesamt der Effizienz der Schätzung durch NLC-CFA Modelle unterlegen. Auch für CFA-MTMM Modelle sollte daneben eine Schätzung mittels drei statt fünf Indikatorvariablen vermieden werden, da dies insgesamt in einer höheren Verzerrung der Faktorladungen resultiert.

Viertens zeigt sich, dass die Spezifikation der systematischen Messfehler mittels NLC-CFA nicht nur Verzerrungen von Faktorladungen bereinigt, sondern dass die geschätzten Effekte der systematischen Messfehler ebenfalls präzise geschätzt werden. Eine spezifische Analyse des Vorhandenseins und der Abhängigkeiten von systematischen Messfehlern ist somit in Abgrenzung zu Schätzungen mittels CFA-MTMM Modellen möglich. Zu beachten ist dabei jedoch, dass die konkreten, die systematischen Messfehler quantifizierenden Effekte innerhalb der Benchmark Modelle unterschätzt werden, so dass an dieser Stelle lediglich von einer annähernd

effizienten Schätzung der entsprechenden Effekte ausgegangen werden kann. Eine höhere Fallzahl wirkt sich dabei positiv auf die Präzision der Schätzer der entsprechenden Effekte aus. Bezüglich der Modellkomplexität kann Empfohlen werden, NLC-CFA Modelle mit vier statt fünf Indikatorvariablen zu schätzen, da dies mit einer Erhöhung der Effizienz und zumindest mit keiner substanziellen Verschlechterung der Präzision einhergeht.

Fünftens kann für das neu eingeführte NLC-CFA Modell insgesamt geschlossen werden, dass für eine Bereinigung der Faktorladungen und eine präzise und effiziente Schätzung der Effekte, welche den Einfluss systematischer Messfehler quantifizieren, ein Modell mit vier oder fünf Indikatorvariablen geschätzt werden und eine für CFA Multigruppenmodelle insgesamt ausreichenden Fallzahl vorhanden sein sollte. Durch systematische Messfehler bedingte Verzerrungen von Faktorladungen können somit durch das NLC-CFA Modell bereinigt werden, die annähernde Stärke, die Richtung und die Signifikanz der Effekte der systematischen Messfehler werden korrekt interpretierbar geschätzt. Das NLC-CFA Modell ist dem CFA-MTMM Modell dabei sowohl hinsichtlich der Bereinigung von Faktorladungen, als auch hinsichtlich des Informationsgewinnes durch die geschätzten Parameter überlegen.

5.2 Diskussion hinsichtlich der theoretischen Implikationen

Theoretisch fundiert wurden in Kapitel 2 eine Reihe von Hypothesen und notwendiger Bedingungen hinsichtlich des Auftretens von FRE in Fragebatterien von Umfragen aufgestellt. Diesbezüglich war es das Ziel, hierauf aufbauend ein CFA Modell zu entwickeln, mit welchem besagte Hypothesen und Bedingungen überprüft werden können. Um das so entwickelte NLC-CFA Modell zu testen, wurden verschiedene Datensätze mit und ohne FRE entsprechend der theoretischen Diskussion in Kapitel 2 simuliert. Die Schätzung des NLC-CFA Modells auf Basis dieser simulierten Daten, für welche wahre Werte bekannt sind, gibt schließlich Aufschluss darüber, unter welchen Bedingungen die jeweiligen Hypothesen und Bedingungen erfolgreich (d.h. unter Ausschluss des Auftretens von Typ 1 oder Typ 2 Fehlern) getestet werden können. Die Diskussion der einzelnen Hypothesen und Bedingungen orientiert sich dabei vor allem an den Ergebnissen der Analyse des angepassten Benchmark Modells (vgl. Kapitel 4.7), greift dabei jedoch auch auf die Ergebnisse der Verzerrungsanalysen (vgl. Kapitel 4.5) und der Präzisionsanalysen (vgl. Kapitel 4.6) zurück. Die Überprüfung der Hypothesen und Bedingungen mittels des CFA-MTMM Modells ist hingegen aufgrund der fehlenden Rückführung der FRE auf zugrundeliegende Faktoren (betrifft die formulierten Hypothesen) sowie der spezifischen

Systematik der Datensimulation (betrifft die formulierten notwendigen Bedingungen) nicht möglich, so dass dieses Modell an dieser Stelle nicht weiter berücksichtigt wird.

Notwendige Bedingung 1 & 2

Beginnend mit den Notwendigen Bedingungen 1 & 2 wurden FRE, differenziert nach Assimilations- und Kontrasteffekten, hinsichtlich des das Auftreten bedingenden Grades an Antwortelaboration evaluiert. Der Grad an Antwortelaboration stellt in den in Kapitel 2 diskutierten Modellen ein Indikator für den Modus der Informationsverarbeitung dar, welcher, je nach Modell, unter anderem in einen automatisch-spontanen sowie einen geplant-kontrollierten Modus (Fazio, 1990), einen impulsiven sowie einen reflexiven Modus (Strack & Deutsch, 2004) bzw. einen automatisch-spontanen und reflexiv-kalkulierenden Modus (Esser, 2001) differenziert werden kann. Diesbezüglich wurde diskutiert, dass Assimilationseffekte einen automatisch-spontanen bzw. impulsiven Antwortmodus mit geringer Elaboration bedingen, während Kontrasteffekte einen geplant-kontrollierten, reflexiven bzw. reflexiv-kalkulierenden Modus mit hoher Elaboration bedingen. Hieraus schließend wurden die folgenden beiden notwendigen Bedingungen formuliert:

Notwendige Bedingung 1: Geringe Elaboration ist eine notwendige Bedingung für einen Assimilationseffekt.

Notwendige Bedingung 2: Hohe Elaboration ist eine notwendige Bedingung für einen Kontrasteffekt.

Entsprechend kann Bedingung 1 bei Schätzung mittels NLC-CFA dann widerlegt werden, wenn Assimilationseffekte auch außerhalb der Befragtengruppe mit geringer Elaboration auftreten, während Bedingung 2 widerlegt werden kann, wenn Kontrasteffekte auch außerhalb der Befragtengruppe mit hoher Elaboration auftreten.

Innerhalb des NLC-CFA Modells sind FRE gemäß Kapitel 3 als Interaktionseffekte zwischen dem Effekt von Kontext- auf Zielfrage und weiteren Bedingungsfaktoren, in diesem Fall die chronische Zugänglichkeit von Informationen mit Relevanz für das Objekt der Befragung (ξ_1 innerhalb der Analysen in Kapitel 4) sowie dem Framing der Befragtenrolle (ξ_2 innerhalb der Analysen in Kapitel 4) spezifiziert und in den zugrundeliegenden Daten simuliert. Entsprechend werden Assimilationseffekte bei gegebener Simulations- und Modellspezifikation durch negative Effekte der Interaktionsterme $y_{j-1}\xi_1$ und $y_{j-1}\xi_2$ identifiziert. Die Effekte sind theoretisch begründbar insofern negativ, als dass für Assimilationseffekte erwartet wird, dass diese 1.) umso geringer sind, desto höher der Grad der chronischen Zugänglichkeit (impliziert einen

negativen Effekt für $y_{j-1}\xi_1$); und 2.) umso geringer sind, desto positiver das Framing der Befragtenrolle ausfällt (impliziert einen negativen Effekt für $y_{j-1}\xi_2$). Kontrasteffekte werden hingegen bei gegebener Simulations- und Modellspezifikation durch einen positiven Effekt des Interaktionsterms $y_{j-1}\xi_1\xi_2$ identifiziert. Der Kontrasteffekt ist insofern durch eine 3-Wege-Interaktion bedingt und positiv als das erwartet wird, dass der Kontrasteffekt lediglich bei geringer chronischer Zugänglichkeit von Informationen gegenüber dem Objekt der Befragung sowie bei einem positiven Framing der Befragtenrolle auftritt, was in einem negativen Produkt der beiden Faktoren resultiert. Da der Kontrasteffekt zudem als Entfernung des Urteils der Zielfrage vom Urteil der Kontextfrage definiert ist, was einem negativen Effekt der Kontextfrage auf die Zielfrage entspricht, so wird wiederum ein positiver Effekt erwartet, da ein negatives Produkt zu einem negativen Effekt führt (vgl. Kapitel 2.3.3).

Tabelle 35: Geschätzte Effekte der angepassten Benchmark Analysen je Gruppe (Zusammenfassung)

Koeffizient (für unabhängige Variable)	Gruppe 1 (geringe Elaboration)	Gruppe 2 (hohe Elaboration)	Gruppe 3 (Kein FRE simuliert)
$\boldsymbol{\beta_1}$ (y_{j-1})	-0,004	-0,004	-0,003
$\boldsymbol{\beta_2}$ (ξ_1)	0,000	0,000	0,000
$\boldsymbol{\beta_3}$ (ξ_2)	0,000	0,000	0,000
$\boldsymbol{\beta_4}$ $(y_{j-1}\xi_1)$	-0,067***	0,000	0,000
$\boldsymbol{\beta_5}$ $(y_{j-1}\xi_2)$	-0,068***	0,000	0,000
$\boldsymbol{\beta_6}$ $(\xi_1\xi_2)$	0,000	0,000	0,001
$\boldsymbol{\beta_7}$ $(y_{j-1}\xi_1\xi_2)$	-0,001	0,144***	0,000

Gruppe 1: Schätzung innerhalb der Gruppe mit simuliertem Assimilationseffekt; **Gruppe 2**: Schätzung innerhalb der Gruppe mit simuliertem Kontrasteffekt; **Gruppe 3**: Schätzung innerhalb der Gruppe ohne simulierten FRE; $\boldsymbol{\beta_b}$: Durchschnittliche unstandardisierte Regressionseffekte der NLC-CFA über die Iterationen der angepassten Benchmark Analyse; $\boldsymbol{\beta_b}$***: $p \leq 0{,}001$; $\boldsymbol{\beta_b}$**: $p \leq 0{,}01$; $\boldsymbol{\beta_b}$*: $p \leq 0{,}05$

Zu beachten ist dabei, dass sich zu erwartende Effekte in Interaktionsanalysen je nach Kodierung der zugrundeliegenden Variablen verändern und die hier besprochene Systematik somit

lediglich bei Anwendung der in Kapitel 4.1 vorgeschlagenen Modellspezifikation und der in Kapitel 4.2 aufgezeigten Datensimulation gilt. Dies betrifft zum einen die Rotation der Skalen, wenn hohe Werte nicht, wie hier impliziert, für eine hohe Zugänglichkeit oder ein positives Framing, sondern im Gegenteil für eine geringe Zugänglichkeit oder ein negatives Framing stehen. Dies führt entsprechend zu einer Umkehrung der Vorzeichen, welche bei der Interpretation der Ergebnisse zu beachten ist. Weiterhin ist die Schätzung des Haupteffektes von y_{j-1} auf y_j abhängig von den Nullpunkten derjenigen Variablen, welche zur Bildung eines Interaktionsterms verwendet werden (in diesem Fall ξ_1 und ξ_2) (Hayes, 2013). Bei Anwendung des hier vorgeschlagenen Verfahrens zur Analyse von realen, empirisch erhobenen Daten ist die Interpretation der Ergebnisse entsprechend an der real vorliegenden Datenstruktur, insbesondere der Rotation der Variablen sowie deren Nullpunkte, anzupassen.

Um den Prozess der Überprüfung der beiden Bedingungen zu illustrieren, sind in Tabelle 35 die Ergebnisse der angepassten Benchmark Analyse aus Kapitel 4.7 zusammengefasst. Wird nun zugrunde gelegt, dass Assimilationseffekte durch negative Effekte von $y_{j-1}\xi_1$ und $y_{j-1}\xi_2$ identifiziert sind und Bedingung 1 postuliert, dass diese lediglich bei geringer Antwortelaboration auftreten, so ist die Bedingung ungültig, wenn die zu erwartenden Effekte 1.) statt lediglich in Gruppe 1 ebenfalls in Gruppe 2 und/oder Gruppe 3 auftreten oder 2.) wenn der zu erwartende Effekt innerhalb von Gruppe 1 nicht negativ oder nicht statistisch signifikant ist, wobei es für die Widerlegung der Bedingung genügt, wenn 1.) oder 2.) gegeben ist. Gemäß Tabelle 35 sind die entsprechenden Effekte innerhalb der Benchmark Modelle in Gruppe 1 negativ und statistisch signifikant, während in Gruppe 2 und Gruppe 3 keine signifikanten Effekte vorliegen, so wie es der zugrundeliegenden Simulation und Theorie entspricht. Zu beachten ist dabei jedoch, dass die entsprechenden Effekte gemäß Kapitel 4.7 unterschätzt werden. Sind Assimilationseffekte entsprechend innerhalb der Daten vorhanden, so kann Bedingung 1 auf Basis des NLC-CFA Modells und den gegebenen Daten erfolgreich getestet und korrekterweise akzeptiert werden.

Über diese Befunde hinaus zeigt die Analyse der Standardfehler in Kapitel 4.6, dass die Schätzung innerhalb des Benchmark Modells präzise ist und die Präzision tendenziell durch eine Verringerung der Anzahl an Indikatoren innerhalb der für die Analyse verwendeten Fragebatterie, in welcher FRE erwartet werden oder durch eine Verringerung der Fallzahl verringert wird. Hinsichtlich der Effizienz der Schätzung (vgl. Kapitel 4.5) ist diesbezüglich zu beachten, dass die genannten Effekte innerhalb des Benchmark Modells über eine statistisch signifikante,

jedoch geringe Standard-Verzerrung ungleich 0 verfügen, während die Verzerrung bei Verwendung von weniger als fünf, aber insbesondere bei Verwendung von vier Indikatoren innerhalb der Fragebatterie geringer wird.

Wird daneben zugrunde gelegt, dass Kontrasteffekte durch einen positiven Effekt von $y_{j-1}\xi_1\xi_2$ identifiziert sind und Bedingung 2 postuliert, dass diese lediglich bei erhöhter Antwortelaboration auftreten, so ist die Bedingung ungültig, wenn der zu erwartende Effekte 1.) statt lediglich in Gruppe 2 ebenfalls in Gruppe 1 und/oder Gruppe 3 auftritt oder 2.) wenn der zu erwartende Effekt innerhalb von Gruppe 2 nicht positiv oder nicht statistisch signifikant ist, wobei es für die Widerlegung der Bedingung erneut genügt, wenn 1.) oder 2.) erfüllt ist. In Tabelle 35 ist diesbezüglich ein statistisch signifikanter, positiver Effekt von $y_{j-1}\xi_1\xi_2$ in Gruppe 2 aufgezeigt, während der entsprechende Effekt weder in Gruppe 1, noch in Gruppe 3 statistisch signifikant ist. Zu beachten ist, dass der Effekt gemäß Kapitel 4.7 in Benchmark Modellen unterschätzt wird. Sind Kontrasteffekte also entsprechend innerhalb der Daten vorhanden, so kann auch Bedingung 2 auf Basis des NLC-CFA Modells und der gegebenen Daten erfolgreich getestet und korrekterweise vorläufig akzeptiert werden.

Hinsichtlich der Präzision der Schätzung zeigt die Analyse der Standardfehler (vgl. Kapitel 4.6), dass diese für Benchmark Modelle eine präzise Schätzung indizieren, während die Schätzung vor allem bei Verringerung der Fallzahl weniger präzise wird. Daneben indizieren die Analysen (vgl. Kapitel 4.5 und Kapitel 4.7), dass der Effekt in Benchmark Modellen substanziell verzerrt ist, wobei vor allem die Verwendung von weniger als fünf Indikatorvariablen die Effizienz der Schätzung erhöht

Hypothesen H_{1a}

FRE bzw. Assimilations- und Kontrasteffekte werden nicht nur durch den Grad der Antwortelaboration bzw. den Modus des Informationsverarbeitungsprozesses bedingt. In Kapitel 2 wurde darüber hinaus ebenfalls der Einfluss der kognitiven, chronischen Zugänglichkeit von Informationen mit Relevanz für das Objekt der Befragung als Einflussfaktor diskutiert. Aus den aggregierten Theorien und Modellen des Befragtenverhaltens wurde diesbezüglich geschlussfolgert, dass Assimilationseffekte dann auftreten, wenn situativ zugängliche Informationen, welche z.B. durch die vorangegangene Kontextfrage aktiviert wurden, für die Urteilsbildung verwendet werden. Sind die für die Urteilsbildung verwendeten Informationen hingegen chronisch Zugänglich, basieren sie also auf bereits vorhandenen Einstellungen, Urteilen etc., so werden eher die chronisch zugänglichen statt der situativ zugänglichen Informationen für die

Urteilsbildung verwendet. Speziell für Assimilationseffekte wurde zudem definiert, dass diese sich durch eine Annäherung des Urteils der Zielfrage an das Urteil der Kontextfrage auszeichnen. Führen vorhandene, chronisch zugängliche Informationen also bei der Urteilsbildung zu einer geringeren Verwendung von situativ zugänglichen Informationen, so führt dies zu einem geringeren Auftreten von Assimilationseffekten, also zu einer geringeren Annäherung des Urteils der Zielfrage an das Urteil der Kontextfrage. Entsprechend gilt

H_{1a}: Desto geringer die kognitive chronische Zugänglichkeit von Information mit Relevanz für das Objekt der Befragung bei geringer Antwortelaboration, desto mehr nähert sich die Antwort auf eine Zielfrage der Antwort auf die Kontextfrage an.

Dies bedeutet, dass H_{1a} durch Inspektion des Interaktionseffektes ($y_{j-1}\xi_1$) zwischen dem Konstrukt, welches die chronische Zugänglichkeit von Informationen repräsentiert (ξ_1), und der Kontextfrage (y_{j-1}) innerhalb derjenigen Gruppe, welche Befragte mit geringer Antwortelaboration abbilden (Gruppe 1), überprüft werden kann. Erwartet wird entsprechend, wie für Bedingung 1 und Bedingung 2 bereits beschrieben, ein negativer Interaktionseffekt von $y_{j-1}\xi_1$ in Gruppe 1, da sich die Antwort der Zielfrage umso mehr der Antwort auf die Kontextfrage annähert, desto geringer die chronische Verfügbarkeit von Informationen ausfällt (da Antworten somit umso eher auf Basis situativ zugänglicher Informationen prozessiert werden). H_{1a} wird entsprechend falsifiziert, wenn der Effekt in Gruppe 1 nicht negativ oder nicht statistisch signifikant ist, wobei in H_{1a} keine Aussage bezüglich der Effektausprägung innerhalb der anderen Gruppen getroffen wird. Zu beachten ist dabei erneut, wie bereits für Bedingung 1 und Bedingung 2 beschrieben, dass die genannte Falsifikationssystematik für die Ergebnisse der Simulationsstudie mit gegebener Spezifikation gilt und dass sich die zu erwartenden Effekte bei Anwendung der NLC-CFA in empirischen Studien je nach Rotation und Lage der Mittelwerte der beteiligten Variablen unterscheiden.

Bezüglich der Effektschätzung von $y_{j-1}\xi_1$ innerhalb der durchgeführten Benchmark Analyse ist in Tabelle 35 für Gruppe 1 ein statistisch signifikanter und negativer Effekt abzulesen. Auch wenn dies für die Falsifikation nicht relevant ist, wird der gleiche Effekt darüber hinaus in Gruppe 2 und Gruppe 3 als nicht signifikant geschätzt. Innerhalb der Benchmark Analyse ist der Effekt somit präzise und ohne Typ 1 oder Typ 2 Fehler geschätzt, was auf Basis der Interpretation des Konfidenzintervalls mit hoher Wahrscheinlichkeit für mindestens 95% durchgeführten Iterationen der angepassten Benchmark Analyse gilt (vgl. Kapitel 4.7). Die Überprüfung von H_{1a} führt somit bei gegebener Modellspezifikation und auf Basis der simulierten Daten zu einem korrekten Ergebnis, die Abhängigkeit eines Assimilationseffektes von weiteren

Faktoren ist auf Basis des vorgeschlagenen NLC-CFA Modells identifizierbar. Weitere Implikationen, welche sich aus den durchgeführten Präzisions- und Effizienzanalysen ergeben, wurden bereits im Kontext von Bedingung 1 und Bedingung 2 diskutiert.

Hypothesen H_{1b}

Neben dem Modus des Informationsverarbeitungsprozesses und der chronischen Zugänglichkeit von Informationen mit Relevanz für das Objekt der Befragung sind Assimilationseffekte entsprechend der Diskussion in Kapitel 2 ebenfalls durch das Framing der Befragtenrolle bedingt. Zusammenfassend wurde diskutiert, dass das Framing über die Differenzierung in chronisch und situativ zugängliche Informationen hinaus generell bedingt, welche Informationen im Kontext der Antwortgenerierung konkret als relevant assoziiert werden. Die Annahme lautet entsprechend, dass Befragte mit indifferentem oder ablehnendem Framing (als negatives Framing verstanden) andere, weniger relevante Informationen für die Urteilsbildung prozessieren, als dies von kooperativen Befragten (als positives Framing verstanden) zu erwarten ist, wozu ebenfalls die situativ zugänglichen Informationen, welche durch die Kontextfrage aktiviert wurden, gehören können. Führt ein negatives Framing also bei der Urteilsbildung zu einer erhöhten Verwendung von situativ zugänglichen Informationen, so führt dies zu einem erhöhten Auftreten von Assimilationseffekten, also zu einer erhöhten Annäherung des Urteils der Zielfrage an das Urteil der Kontextfrage. Entsprechend gilt:

H_{1b}: Bei indifferenter Befragtenrolle und geringer Antwortelaboration nähert sich die Antwort auf eine Zielfrage der Antwort auf die Kontextfrage an.

Entsprechend kann H_{1b} durch Untersuchung des Interaktionseffektes ($y_{j-1}\xi_2$) zwischen dem Konstrukt, welches das Framing der Befragtenrolle repräsentiert (ξ_2), und der Kontextfrage (y_{j-1}) innerhalb derjenigen Gruppe, welche Befragte mit geringer Antwortelaboration abbilden (Gruppe 1), überprüft werden. Erwartet wird, wie ebenfalls für Bedingung 1 und Bedingung 2 beschrieben, ein negativer Interaktionseffekt von $y_{j-1}\xi_2$ in Gruppe 1, da sich die Antwort der Zielfrage umso mehr der Antwort auf die Kontextfrage annähert, desto negativer das Framing der Befragtenrolle ausfällt (da Antworten somit umso eher auf Basis situativ zugänglicher Informationen prozessiert werden). H_{1b} wird demnach falsifiziert, wenn der Effekt in Gruppe 1 nicht negativ oder nicht statistisch signifikant ist, wobei in H_{1b} erneut keine Aussage bezüglich der Effektausprägung innerhalb der anderen Gruppen getroffen wird. Erneut ist die genannte Falsifikationssystematik lediglich für die Ergebnisse der Simulationsstudie mit gegebener Spezifikation gültig, die zu erwartenden Effekte können sich bei Anwendung des Verfahrens in

empirischen Studien je nach Rotation und Lage der Mittelwerte der beteiligten Variablen unterscheiden.

Hinsichtlich der konkreten Falsifikation der Hypothese ist innerhalb der Benchmark Analyse in Tabelle 35 für Gruppe 1 ein statistisch signifikanter und negativer Effekt für $y_{j-1}\xi_2$ abzulesen. Die Effekte in Gruppe 2 und Gruppe 3 sind für die Falsifikation erneut nicht relevant, dennoch ist es interessant, dass diese als nicht signifikant geschätzt werden. Der Effekt ist somit innerhalb der Benchmark Analyse präzise und ohne Typ 1 oder Typ 2 Fehler geschätzt, was auf Basis der Konfidenzintervalle mit hoher Wahrscheinlichkeit für mindestens 95% der durchgeführten Iterationen der Benchmark Analyse gilt (vgl. Kapitel 4.7). Bei gegebener Modellspezifikation und auf Basis der simulierten Daten führt die Überprüfung von H_{1b} entsprechend zu einem korrekten Ergebnis. Weitere Implikationen für den Effekt von $y_{j-1}\xi_2$ bezüglich der Präzision und Effizienz über das Benchmark Modell hinaus wurden ebenfalls bereits im Kontext von Bedingung 1 und Bedingung 2 diskutiert.

Hypothesen H_2

Für Kontrasteffekte wurde in Kapitel 2 diskutiert, dass diese unter der Bedingung einer erhöhten Antwortelaboration, einer geringen chronischen Zugänglichkeit von Informationen mit Relevanz für das Objekt der Befragung sowie eines positiven Framings der Befragtenrolle auftreten, wobei die Bedingungsfaktoren als Interaktiv zu verstehen sind. Dies bedeutet, dass Kontrasteffekte lediglich für Befragte mit einem positiven Framing und einer geringen chronischen Zugänglichkeit von Informationen und einer erhöhten Antwortelaboration auftreten. Zusammenfassend wurde dies damit begründet, dass Kontrasteffekte in dem Sinne eine erhöhte Elaboration bedingen, als dass Informationen bewusst subtrahiert werden, um Redundanz zu vermeiden. Daneben muss ebenfalls eine kooperative Kommunikationsnorm für die generelle Intention gegeben sein, Redundanz zu vermeiden. Zudem müssen die durch die Kontextfrage situativ aktivierten Informationen, welche subtrahiert werden, bewusst kognitiv zugänglich sein, was im Umkehrschluss eine geringe chronische Zugänglichkeit von Informationen mit Relevanz für das Objekt der Befragung impliziert. Innerhalb eines hoch elaborierten Informationsverarbeitungsprozesses führt ein positives Framing der Befragtenrolle sowie eine geringe chronische Zugänglichkeit von Informationen mit Relevanz für das Objekt der Befragung somit zu einem Kontrasteffekt, welcher durch eine Entfernung des Urteils der Zielfrage vom Urteil der Kontextfrage definiert ist. Es gilt:

H_2: Bei kooperativer Befragtenrolle und hoher Antwortelaboration entfernt sich die Antwort auf eine Zielfrage von der Antwort auf die Kontextfrage, desto geringer die kognitive chronische Zugänglichkeit von Information mit Relevanz für das Objekt der Befragung.

H_2 kann folglich durch Analyse der 3-Wege-Interaktion ($y_{j-1}\xi_1\xi_2$) zwischen dem Konstrukt, welches die chronische Zugänglichkeit repräsentiert (ξ_1), dem Konstrukt, welches das Framing der Befragtenrolle repräsentiert (ξ_2), und der Kontextfrage (y_{j-1}) innerhalb derjenigen Gruppe, welche Befragte mit hoher Antwortelaboration abbilden (Gruppe 2), getestet werden. Erwartet wird ein positiver Interaktionseffekt von $y_{j-1}\xi_1\xi_2$ in Gruppe 2. Dies ist dadurch begründbar, da für das Framing und die Zugänglichkeit jeweils ein positiver, sowie ein negativer Effekt erwartet werden, was bei Bildung des Produktes zu einem negativen Vorzeichen führt. Da sich das Urteil der Zielfrage von Kontrasteffekten jedoch ebenfalls negativ von dem Urteil der Kontextfrage entfernt, resultiert eine negative Verzerrung bei einem gegebenen negativen Produkt von $\xi_1\xi_2$ in einem positiven Effekt. H_2 kann entsprechend falsifiziert werden, wenn der Effekt in Gruppe 2 nicht positiv oder nicht statistisch signifikant ist, wobei in H_2 erneut keine Aussage bezüglich der Effektausprägung innerhalb der anderen Gruppen getroffen wird. Und auch für H_2 gilt, dass die genannte Falsifikationssystematik lediglich für die Ergebnisse der Simulationsstudie mit der in Kapitel 4 diskutierten Spezifikation gültig ist. Entsprechend können sich auch für H_2 die zu erwartenden Effekte bei Anwendung des Verfahrens in empirischen Studien je nach Rotation und Lage der Mittelwerte der beteiligten Variablen unterscheiden.

Soll H_2 nun konkret getestet werden, muss der Effekt von $y_{j-1}\xi_1\xi_2$ aus Gruppe 2 in Tabelle 35 analysiert werden. Dieser wird gemäß Konfidenzintervall (vgl. Kapitel 4.7) mit hoher Wahrscheinlichkeit für mindestens 95% der Iterationen der angepassten Benchmark Analyse als positiv und statistisch signifikant geschätzt. Auch wenn dies für die Falsifikation der Hypothese keine Rolle spielt ist darüber hinaus für Gruppe 1 sowie für Gruppe 3 abzulesen, dass der entsprechende Effekt jeweils als statistisch nicht signifikant geschätzt wird. Der Effekt ist somit innerhalb der Benchmark Analyse präzise und ohne Typ 1 oder Typ 2 Fehler geschätzt. Die Überprüfung von H_2 bei gegebener Modellspezifikation und Datenstruktur führt demnach zu einem korrekten Ergebnis. Bezüglich der Präzision und Effizienz bei Schätzung des Effektes von $y_{j-1}\xi_1\xi_2$ über das Benchmark Modell hinaus wurden bereits im Kontext von Bedingung 1 und Bedingung 2 mögliche Einflussfaktoren diskutiert.

Zusammenfassung der theoretischen Implikationen der NLC-CFA Schätzung

Bezüglich der Überprüfbarkeit der theoretischen Hypothesen und notwendigen Bedingungen auf Basis des NLC-CFA Modells zeigt sich somit insgesamt, dass bei gegebener Modell- und Simulationsspezifikation für das angepasste Benchmark Modell alle Hypothesen mit korrektem Resultat, das heißt ohne das Auftreten von Typ 1 oder Typ 2 Fehlern, getestet werden konnten. Zu berücksichtigen bleibt dabei eine generelle Unterschätzung der Effekte, was bei der Interpretation dieser zu berücksichtigen ist sowie, je nach Höhe der tatsächlichen Effekte und der Standardfehler, in bestimmten Fällen zu einem Typ 1 Fehler führen kann. Die Präzision der Schätzungen verringert sich dabei über das Benchmark Modell hinaus für alle innerhalb der Monte-Carlo Analyse berücksichtigten Simulations- und Modellspezifikationen für die relevanten Effekte ($y_{j-1}\xi_1$, $y_{j-1}\xi_2$ und $y_{j-1}\xi_1\xi_2$) substanziell bei Verringerung der Fallzahl (vgl. Kapitel 4.6). Hinsichtlich der Effizienz der Schätzung lässt sich für die genannten Effekte zusammenfassen, dass eine Verringerung der Anzahl an Indikatoren von fünf auf vier (und für $y_{j-1}\xi_1\xi_2$ ebenfalls von fünf auf drei) die Verzerrung verringert (vgl. Kapitel 4.5). Auf Basis der Ergebnisse der Simulationsstudie kann somit geschlossen werden, dass das NLC-CFA Modell mit hoher Präzision dazu genutzt werden kann, die hier formulierten Hypothesen zu testen, auch wenn berücksichtigt werden sollte, dass die Effekte tendenziell unterschätzt sein können.

5.3 Anwendung der NLC-CFA in der Praxis

Die Entwicklung eines innovativen Modells zur Identifikation und Kontrolle von FRE ist kein reiner Selbstzweck. Stattdessen soll das entwickelte und evaluierte NLC-CFA Modell als konkretes Werkzeug für empirische Untersuchungen in der wissenschaftlichen Praxis dienen und ebenso für spezielle Anwendungsfälle individuell adaptiert werden können. Entsprechend erfolgt zum Abschluss dieser Arbeit eine Diskussion der speziellen Rahmenbedingungen, Limitationen und Implikationen der durchgeführten Evaluation des NLC-CFA Modells, da alle durchgeführten Analysen grundsätzlich auf spezifisch simulierten Daten basieren, welche in dieser Form nicht für jeden empirischen Anwendungsfall gegeben sein müssen. Dies betrifft insbesondere 1.) die Güte der Schätzungen derjenigen exogenen, latenten Variablen, welche zur Kontrolle der systematischen Messfehler verwendet werden (ξ_1 und ξ_2 in Kapitel 3 und 4); 2.) das Problem der Multikollinearität zwischen den exogenen Variablen; 3.) die Simulation innerhalb der Logik serieller Mediation auf Indikatorebene; und 4.) die Verwendung geeigneter

Model Fit Maße. Eine Diskussion der entsprechenden Rahmenbedingungen, Limitationen und Implikationen soll insgesamt dazu beitragen, auf Basis des eingeführten NLC-CFA Modells geschätzte Effekte besser interpretieren und in den Kontext der gesamten Datensimulation einordnen zu können. Hierauf aufbauend erfolgt abschließend 5.) eine Diskussion der weiteren Anwendungsmöglichkeiten des hier vorgeschlagenen Modells hinsichtlich dessen situativer Anpassung für weitere Anwendungsfälle.

Güte der Messung exogener, latenter Variablen

Die erste zu diskutierende Rahmenbedingung bzw. Limitation der durchgeführten Simulationsanalysen liegt in der Simulationsspezifikation der exogenen, latenten Variablen ξ_1 und ξ_2 begründet. Beide latente Variablen wurden normalverteilt mit einem Mittelwert von 0 und einer Varianz von 1 sowie unabhängig voneinander simuliert. Die drei jeweils simulierten Indikatorvariablen x_j, welche später zur Schätzung der entsprechenden latenten Konstrukte verwendet wurden, setzen sich zu 80% aus der Varianz des entsprechenden latenten Konstruktes (Common Variance) und zu 20% aus der Varianz des jeweiligen zufälligen Messfehlers zusammen. Die potentielle Varianzausschöpfung der entsprechenden Indikatorvariablen x_j liegt demnach bei $R_j^2 = 0{,}8$. Dies entspricht innerhalb dieses Messmodells einer standardisierten, wahren Faktorladung von $\lambda_j = 0{,}89$, also einer sehr hohen Faktorladung als Indikator einer sehr hohen Konstruktvalidität und einer sehr hohen Kompositreliabilität (Cho, 2016; Schermelleh-Engel & Schweizer, 2012; Urban & Mayerl, 2014), was einer annähernd „perfekten“ Messung der entsprechenden latenten Konstrukte entspricht, wie sie in der empirischen Realität kaum zu finden ist.

Problematisch ist dies insofern, als dass die Effizienz aller Schätzungen in SEM, also auch der Schätzung der Effekte exogener latenter Variablen auf endogene, abhängige Variablen von der Güte der Messung der entsprechenden latenten Variablen, sprich, der Höhe der standardisierten Faktorladungen abhängig ist (Brown, 2006). Dies wird ebenfalls durch Simulationsstudien unterstützt welche z.B. aufzeigen, dass die Höhe der Reliabilität eines latenten Konstruktes einen Einfluss auf die Güte der Schätzung von Parametern nimmt (Yang & Green, 2010) oder dass ein Zusammenhang zwischen effizienten Parameterschätzungen und der Höhe von Faktorladungen in Interaktion mit der Fallzahl besteht (Wolf, Harrington, Clark, & Miller, 2013). Allgemein lässt sich formulieren, dass je höher die standardisierten Faktorladungen eines latenten Konstruktes, umso effizienter kann die Varianz des latenten Konstruktes und damit auch die

Kovarianzen mit anderen Faktoren geschätzt werden, was zu effizienteren Schätzungen von Effekten auf Ebene des Strukturmodells führt.

Im Umkehrschluss bedeutet dies, dass Messungen der exogenen latenten Variablen ξ_1 und ξ_2 mit geringeren, standardisierten Faktorladungen bzw. einem geringeren Anteil an Common Variance, als in der vorliegenden Studie simuliert wurde, zu Verzerrungen der Schätzungen der FRE führen können. Dies sollte dabei insbesondere die beiden Haupteffekte von ξ_1 und ξ_2, aber auch diejenigen Interaktionseffekte, welche als Produkt aus ξ_1 oder ξ_2 gebildet sind, betreffen ($y_{j-1}\xi_1$, $y_{j-1}\xi_2$, $\xi_1\xi_2$ und $y_{j-1}\xi_1\xi_2$), wobei der Haupteffekt von y_{j-1} ebenfalls durch eine verzerrte Schätzung von ξ_1 und ξ_2 beeinflusst werden kann, da die unabhängigen Variablen des Modells sich ebenfalls gegenseitig kontrollieren. Eine geringe Güte der Schätzung der latenten Konstrukte ξ_1 und ξ_2, impliziert durch geringe, standardisierte Faktorladungen, kann somit die Schätzung aller Effekte verzerren.

Daneben ist zu beachten, dass in der vorliegenden Simulation und Schätzung angenommen wurde, dass ξ_1 und ξ_2 durch keine systematischen Messfehler bedingt sind und in diesen Beispielsweise kein FRE vorhanden ist. Wird jedoch angenommen, dass FRE ein Phänomen sind, welches grundsätzlich in jeder Batterie aufeinanderfolgender Fragen auftreten können, so ist die prinzipielle Unabhängigkeit von ξ_1 und ξ_2 von systematischen Messfehlern nicht plausibel. Entsprechend kann empfohlen werden, dass wenn latente Konstrukte zur Kontrolle und damit zur Bereinigung systematischer Messfehler verwendet werden, zuvor getestet werden sollte, ob diese selbst durch systematische Messfehler beeinflusst werden und ob diese über ausreichend hohe Faktorladungen verfügen. Sind die latenten Konstrukte entsprechend nicht unabhängig von systematischen Messfehlern oder verfügen diese über eine zu geringe Faktorladung, so sollten eher manifeste exogene Variablen zur Kontrolle systematischer Messfehler verwendet werden.

Multikollinearität zwischen exogenen Variablen

Eine weitere Rahmenbedingung bzw. Limitation der vorliegenden Simulation und der Modellschätzungen ergibt sich erneut aus der Spezifikation von ξ_1 und ξ_2, da die beiden latenten Variablen zusätzlich unabhängig voneinander, sprich mit $Cov(\xi_2, \xi_1) = 0$ simuliert wurden. Die vorliegende Simulationsspezifikation impliziert entsprechend grundsätzlich, dass keine Multikollinearität zwischen ξ_1 und ξ_2 vorliegt, was für empirische Anwendungen kaum realistisch ist, insbesondere wenn die inhaltliche Bedeutung der latenten Konstrukte gemäß Kapitel 2 und Kapitel 3 berücksichtigt wird. ξ_1 entspricht der Zugänglichkeit von Informationen gegenüber

dem Objekt der Befragung, ξ_2 entspricht dem Framing der Befragtenrolle, wobei die Diskussion in Kapitel 2 aufzeigt, dass diese beiden Faktoren nicht zwangsläufig unabhängig voneinander sind. Situationsbedingt kann für die empirische Anwendung des vorgeschlagenen Modells entsprechend eine Vorhandene, wenn nicht gar eine erhöhte Kovarianz zwischen den exogenen Variablen erwartet werden, wodurch das Thema der Multikollinearität in das Zentrum der Betrachtung rückt.

Grundsätzlich gilt bei vorhandener Multikollinearität, also einer vorhandenen und erhöhten Kovarianz bzw. Korrelation zwischen unabhängigen, exogenen Variablen (z.B. ξ_1 und ξ_2), insbesondere für nichtlineare Modelle (wie das hier vorliegende), dass das zugrundeliegende Verfahren nicht trennscharf zwischen derjenigen Varianz der abhängigen Variable unterscheiden kann, welche durch ξ_1 und welche durch ξ_2 erklärt wird. Desto höher also die Korrelation zwischen ξ_1 und ξ_2, desto eher sind Effekte, welche auf ξ_1 und ξ_2 zurückgehen, verzerrt, wobei insbesondere die Präzision von Schätzungen abnimmt, da Standardfehler zunehmen. Dabei sollte bei der Betrachtung jedoch zwischen nicht-essenzieller und essenzieller Multikollinearität unterschieden werden. Die erste Form der Multikollinearität ist lediglich durch die Skalierung der beteiligten Variablen bedingt und kann durch simple Zentrierung der Prädiktoren ausgeglichen werden. Essenzielle Multikollinearität hingegen geht auf fundamentale Zusammenhänge zwischen den Messungen von ξ_1 und ξ_2 zurück und kann entsprechend nicht durch Zentrierung oder andere, einfache Datentransformationen ausgeglichen werden. Entsprechend führt letztere zur Verzerrung von Effekten (Moosbrugger et al., 2009).

Demnach ist es bei Schätzung des vorgeschlagenen NLC-CFA Modells ratsam die Korrelation zwischen den exogenen Variablen ξ_1 und ξ_2 zu überprüfen. Im optimalen Fall ist keine bzw. lediglich eine sehr geringe Korrelation vorhanden. Ist hingegen eine relevante Korrelation vorhanden, so kann diese durch Fixierung einer Kovarianz zwischen den exogenen Variablen quantifiziert werden, so dass diese zumindest nicht unberücksichtigt bleibt.

Serielle Mediation auf Indikatorebene

Weiterhin sind die Ergebnisse der Simulationsstudien dadurch limitiert, dass die Zusammenhänge zwischen aufeinanderfolgenden Kontext-Zielfragen-Dyaden innerhalb der Logik serieller Mediation simuliert sind. Dies soll am Beispiel der Effekte von y_j auf y_{j+1}, welcher in den Simulationsstudien als β_1 bezeichnet wird, illustriert werden. Dabei ist zu beachten, dass in den Simulationsstudien die jeweiligen Effekte zwischen y_1 und y_2, zwischen y_2 und y_3 etc. innerhalb einer Gruppe grundsätzlich als gleich simuliert wurde. Um nun also das Problem serieller

Mediation auf Indikatorebene darzustellen, wird als einfaches Beispiel ein Gleichungssystem (Formel 59.1a, Formel 59.2a und Formel 59.3a) definiert, innerhalb welchem die Messungen von drei Items y_j (mit j als Laufindex der Variablen y) abhängig sind von der entsprechenden Common Variance $\lambda_j\eta$ sowie einem zufälligen Messfehler ε_j. Zusätzlich sind y_2 und y_3 abhängig von der vorangegangenen Frage, dargestellt durch $\beta_1 y_1$ für y_2 und $\beta_1 y_2$ für y_3. Der FRE β_1 ist entsprechend für beide Variablen y_2 und y_3 analog zu den Simulationsstudien durch den gleichen Wert dargestellt. Um das Beispiel einfach zu halten, wurde zudem auf Interaktionsterme verzichtet.

$$(1a)\quad y_1 = \lambda_1\eta + \varepsilon_1$$

$$(2a)\quad y_2 = \lambda_2\eta + \beta_1 y_1 + \varepsilon_2$$

$$(3a)\quad y_3 = \lambda_3\eta + \beta_1 y_2 + \varepsilon_3$$

$$(2b)\quad y_2 = \lambda_2\eta + \beta_1(\lambda_1\eta + \varepsilon_1) + \varepsilon_2 = \lambda_2\eta + \beta_1\lambda_1\eta + \beta_1\varepsilon_1 + \varepsilon_2$$

$$(2c)\quad y_2 = \eta(\lambda_2 + \beta_1\lambda_1) + \beta_1\varepsilon_1 + \varepsilon_2$$

Formel 59

$$(3b)\quad y_3 = \lambda_3\eta + \beta_1(\eta(\lambda_2 + \beta_1\lambda_1) + \beta_1\varepsilon_1 + \varepsilon_2) + \varepsilon_3$$

$$(3c)\quad y_3 = \lambda_3\eta + \beta_1\eta(\lambda_2 + \beta_1\lambda_1) + \beta_1^2\varepsilon_1 + \beta_1\varepsilon_2 + \varepsilon_3$$

$$(3d)\quad y_3 = \eta(\lambda_3 + \beta_1\lambda_2 + \beta_1^2\lambda_1) + \beta_1^2\varepsilon_1 + \beta_1\varepsilon_2 + \varepsilon_3$$

Wird Formel 59.1a nun in Formel 59.2a eingesetzt, so ergibt sich Formel 59.2b und durch Umstellung Formel 59.2c. Formel 59.2c kann wiederum in Formel 59.3a eingesetzt werden (führt zu Formel 59.3b), was wiederum zu Formel 59.3d führt. Aus den Formeln Formel 59.2c und Formel 59.3d lassen sich nun zwei Erkenntnisse ziehen:

1.) entspricht die zufällige Messfehlervarianz in y_1 lediglich der wahren Varianz von ε_1, als geschätzte zufällige Messfehlervarianz von y_2 wird hingegen $\beta_1\varepsilon_1 + \varepsilon_2$ extrahiert, die zufällige Messfehlervarianz von y_3 entspricht zuletzt $\beta_1^2\varepsilon_1 + \beta_1\varepsilon_2 + \varepsilon_3$, wenn FRE nicht spezifiziert werden. Die führt, bei gegebener positiver Ausprägung von β_1, ε_1, ε_2 und ε_3 zu einer erhöhten,

geschätzten Varianz der zufälligen Messfehler in Items, welche später in der Fragebatterie präsentiert wurden und für welche FRE vorliegen. Bei negativer Ausprägung von β_1 führt dies hingegen zu einer verringerten, geschätzten Varianz der zufälligen Messfehler. Je nachdem, ob der zugrundeliegende FRE nun ein Assimilationseffekt (entspricht einem positiven Effekt, da sich das Urteil der Zielfrage an das Urteil der Kontextfrage annähert) oder ein Kontrasteffekt (entspricht einem negativen Effekt, da sich das Urteil der Zielfrage vom Urteil der Kontextfrage entfernt) darstellt, ist der Varianzanteil, welcher auf zufällige Messfehler zurückgeht, in Items, welche später innerhalb der Fragebatterie präsentiert werden, höher (bei positivem β_1) oder tiefer (bei negativem β_1), wenn die einzelnen Effekte nicht korrekt spezifiziert werden, wie es in den einfachen CFA Modellen ohne Spezifikation der FRE in Kapitel 4 geschehen ist. Dies resultiert letztlich in verzerrten Schätzungen der Residualvarianzen und damit zu verzerrten Messungen des Model Fit.

2.) entspricht die geschätzte Common Variance in y_1 lediglich der Varianz von $\lambda_1\eta$, der geschätzte Anteil der Common Variance von y_2 entspricht hingegen $(\lambda_2 + \beta_1\lambda_1)\eta$, die geschätzte Common Variance von y_3 entspricht zuletzt $(\lambda_3 + \beta_1\lambda_2 + \beta_1^2\lambda_1)\eta$, wenn FRE nicht spezifiziert werden. Dies führt ebenfalls zu einer verzerrten Schätzung des Anteils an Common Variance und entsprechend zu verzerrten Schätzungen von Faktorladungen und/oder der latenten Varianz, je nachdem, ob die latente Varianz oder eine Faktorladung auf 1 fixiert wurde. Die Verzerrung ist dabei erneut umso höher, desto später das entsprechende Item innerhalb der Fragebatterie präsentiert wurde, da sich der Einfluss des FRE in aufeinanderfolgenden Items akkumuliert.

Somit ist aus mathematischer Sicht dargestellt, warum in den Simulationsstudien Faktorladungen verzerrt geschätzt werden, wenn FRE nicht spezifiziert und die einzelnen Varianzanteile nicht zerlegt werden. In der detaillierten Analyse der angepassten Benchmark Modelle (vgl. Kapitel 4.7) zeigt sich darüber hinaus ebenfalls, dass Faktorladungen umso verzerrter sind, desto später diese innerhalb der Fragebatterie präsentiert werden. Zudem sollte ähnliches für die entsprechenden Residualvarianzen beobachtbar sein, auch wenn diese nicht gezielt in den Simulationsstudien analysiert wurden.

Relevant ist dieser grundsätzliche Zusammenhang darüber hinaus entsprechend insofern, als dass die Stärke der FRE β_1, welche wiederum gemäß Formel 59 einen erheblichen Einfluss auf die Dekomposition der verschiedenen Varianzanteile haben, zwischen allen Kontext-Zielfragen-Dyaden als gleich simuliert und geschätzt wurden. Unberücksichtigt bleibt dabei, dass diesbezüglich die theoretische Annahme getroffen wurde, dass FRE über Zeit die gleiche Stärke

behalten und nicht etwa mit zunehmender Exposition abnehmen oder gar zunehmen. Ein abnehmender Anteil der Stärke der FRE würde entsprechend zu einer geringeren Verzerrung der extrahierten Varianzanteile bei fehlender Spezifikation der FRE in späteren Items einer Fragebatterie führen, als wenn die Stärke der FRE zwischen allen Kontext-Zielfragen-Dyaden konstant ist. Es kann somit empfohlen werden, dass ebenfalls Modelle geschätzt werden, in welchen die Stärke der FRE nicht zwischen den Dyaden fixiert, sondern frei geschätzt werden, so hierfür genügend Freiheitsgrade vorliegen. Ein Vergleich der Residualvarianzen der entsprechenden Indikatoren unter Berücksichtigung der Vorzeichen der Regressionskoeffizienten zwischen den Modellen mit und ohne Fixierung kann dann Aufschluss darüber geben, welche Modellspezifikation eher durch die Daten unterstützt wird.

Model Fit

In den Analysen komplett unberücksichtigt blieb bisher der Model Fit, welchen es für gewöhnlich bei Anwendung von CFA und SEM zu interpretieren gilt, da auf Basis der entsprechenden Maße abgeleitet werden kann, ob (und in welchem Ausmaß) das spezifizierte Modell durch die zugrundeliegenden Daten unterstützt wird.

Grundsätzlich gilt dabei jedoch zu berücksichtigen, dass keine theoretische Rechtfertigung traditioneller Model Fit Maße bei der Verwendung nichtlinearer Modelle existiert, da diese auf der Annahme linearer Abhängigkeit basieren und somit theoretisch verzerrt und faktisch ineffizient sind (Klein & Moosbrugger, 2000; Moosbrugger et al., 2009). Die Evaluation des Model Fit bei nichtlinearen Modellen ist somit ein ungelöstes Problem der Methodenforschung (Schermelleh-Engel et al., 2014), wobei vorgeschlagene Verfahren weiterer Analysen und Evaluationen bedürfen (Pfleger, 2019).
Eine solche mögliche Lösung zur Evaluation des Modell Fit wird von Klein und Moosbrugger (2000) im Kontext des LMS-Verfahrens vorgeschlagen. Empfohlen wird, LMS-Modelle in zwei Schritten zu schätzen (Klein & Moosbrugger, 2000; Muthén, 2012), wobei die nachfolgend beschriebene Technik ebenfalls für Modelle, die wie das vorliegende NLC-CFA Modell nach dem Unconstrained Approach spezifiziert sind, angewandt werden kann: Im ersten Schritt wird das Strukturmodell als lineares Modell ohne Interaktionsterme geschätzt, wobei Modell Fit Maße wie CFI, SRMR, RMSEA und Chi² diesem Modell entnommen werden. In einem zweiten Schritt wird besagtes Strukturmodell um die Interaktionsterme ergänzt, wobei diesem Modell die finalen Regressionskoeffizienten und Signifikanzniveaus entnommen werden. Zudem wird ein log-likelihood-ratio Test durchgeführt, mit dessen Hilfe der relative Fit der beiden Modelle verglichen und somit getestet wird, ob das Modell ohne Interaktionsterme mit einem

signifikanten Verlust des Fit relative zum komplexeren Modell mit Interaktionstermen einhergeht (Satorra, 2000; Satorra & Bentler, 2010). Sollte das weniger komplexe Modell über einen guten Fit verfügen und per log-likelihood Test einen signifikanten Verlust des Fit relativ zum komplexeren Modell aufweisen, kann angenommen werden, dass das komplexere Modell ebenfalls über einen guten Fit verfügt (Maslowsky et al. 2015). Somit ist zumindest eine Annäherung an den totalen Model Fit durch Ableitung eines relativen Fit gegeben, auch wenn die Höhe des tatsächlichen Fit nach wie vor nicht abgeleitet werden kann.

Eine weitere Möglichkeit zur Annäherung an den Model Fit in nichtlinearen Modellen wird von Klein und Schermelleh-Engel (2010) diskutiert. Diesbezüglich wird ein neues Maß vorgeschlagen, welches eine mögliche heteroskedastische Verteilung der Residuen analysiert, wobei die Likelihood der Residuen unter der Annahme gegebener Heteroskedastizität mit der Likelihood der Residuen unter der Annahme gegebener Homoskedastizität verglichen wird. Die theoretische Annahme dahinter lautet, dass unberücksichtigte nichtlineare Beziehungen innerhalb des Strukturmodells zu heteroskedastisch verteilten Residuen führen. Wird also eine eigentlich nichtlineare Beziehung fälschlicherweise als lineare Beziehung geschätzt, so drückt sich dies in einer heteroskedastischen Residualverteilung aus, was durch die vorgeschlagene Maßzahl quantifiziert wird und somit identifizierbar ist. Die entsprechenden Annahmen werden durch eine Monte Carlo Simulationsanalyse bestätigt. Der eigentliche Model Fit ist somit nach wie vor nicht quantifiziert, hingegen ist auf Basis dieses Verfahrens zumindest eine Möglichkeit gegeben, fehlerhaft spezifizierte Modelle zu identifizieren.

Soll also entsprechend das vorgeschlagene NLC-CFA Modell zur Identifikation und Kontrolle von FRE verwendet werden, so kann grundsätzlich der von Klein und Moosbrugger (2000) vorgeschlagene Log-Likelihood Test durchgeführt werden, um sich dem Model Fit zumindest anzunähern.

Verfahren als Framework

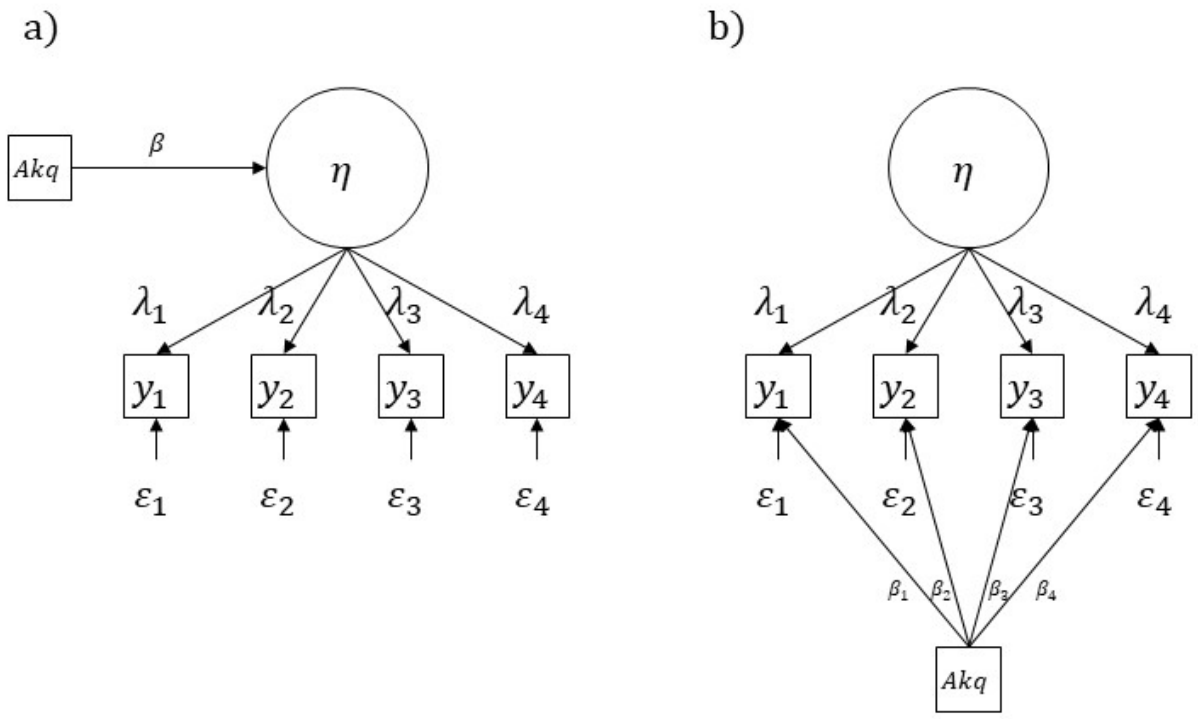
Abschließend ist grundsätzlich festzustellen, dass bei Schätzung des vorgeschlagenen NLC-CFA Modells ein äußerst komplexes Modell zu spezifizieren ist, da insgesamt mehrere latente Interaktionen auf Indikatorebene, innerhalb serieller Mediationen und innerhalb eines Multigruppen Modells mit metrischer Invarianz geschätzt werden, wobei jede einzelne dieser Spezifikationen für sich schon zu relativ komplexen Modellen führt. Entsprechend bedingt das vorliegende Verfahren auch eine Vielzahl an Daten, so dass genügend Freiheitsgrade vorhanden sind, um das Modell schätzen zu können.

Diesbezüglich gilt jedoch zu berücksichtigen, dass die vorliegende Arbeit nicht nur gezeigt hat, dass mittels des in Kapitel 3 eingeführten NLC-CFA Modells FRE quantifiziert und hierdurch verzerrte Faktorladungen bereinigt werden können. Vielmehr zeigt sich, dass Faktorladungen grundsätzlich durch unspezifizierte systematische Messfehler verzerrt sind und dass die Spezifikation regressiver Beziehungen zwischen methodenrelevanten, exogenen Variablen und Indikatorvariablen eines CFA Modells zu einer Bereinigung der Faktorladungen führt. Entsprechend kann angenommen werden, dass diese grundsätzliche Technik der Kontrolle durch Regression auf Indikatorebene für viele verschiedene Modelle möglich ist, entsprechend auch für weniger komplexe Modelle, so dass das Verfahren insgesamt als Framework für die Bereinigung von Faktorladungen angesehen werden kann. Das hier vorgeschlagene NLC-CFA Modell kann demnach in vielerlei Hinsicht adaptiert werden, darunter zum Beispiel als

1. **C**ontrolled CFA (**C-CFA**): Kontrolle von Indikatorvariablen durch lineare Regressionen (statt nichtlinearer Regressionen) mit manifesten exogenen Variablen (statt mit latenten exogenen Variablen)
2. **L**atent **c**ontrolled CFA (**LC-CFA**): Kontrolle von Indikatorvariablen durch lineare Regressionen (statt nichtlinearer Regressionen) mit latenten exogenen Variablen
3. **N**onlinear **c**ontrolled CFA (**NC-CFA**): Kontrolle von Indikatorvariablen durch nichtlineare Regressionen mit manifesten exogenen Variablen (statt mit latenten exogenen Variablen)

Ein Beispiel für einen möglichen Anwendungsfall eines C-CFA Modells ist bei der Kontrolle von Akquieszenzeffekten (vgl. Kapitel 2.2.1) innerhalb beliebiger latenter Konstrukte auf Indikatorebene durch lineare Rückführung auf eine manifeste Variable gegeben (vgl. Abbildung 22b), ähnlich wie Mayerl und Giehl (2018) dies bereits auf der Ebene des latenten Konstruktes kontrollieren (vgl. Abbildung 22a). Diesbezüglich wird zunächst über alle Befragungsitems hinweg die Häufigkeit absoluter Zustimmung gezählt und als Anteil aller absoluten Zustimmungen an allen möglichen Befragungsitems der gesamten Befragung in Prozent dargestellt („Akq“ in Abbildung 22). Der so resultierende manifeste Wert kann dann als Approximation der Tendenz zur Zustimmung angesehen werden. Durch Spezifikation der entsprechenden Regressionen auf Indikatorebene sollte die Varianz der einzelnen Items dann korrekt in die Common Variance, die durch die Tendenz zur Zustimmung bedingte, systematische Fehlervarianz, sowie die zufällige Fehlervarianz zerlegt werden.

Abbildung 22: C-CFA zur Bereinigung von Akquieszenz-Effekten

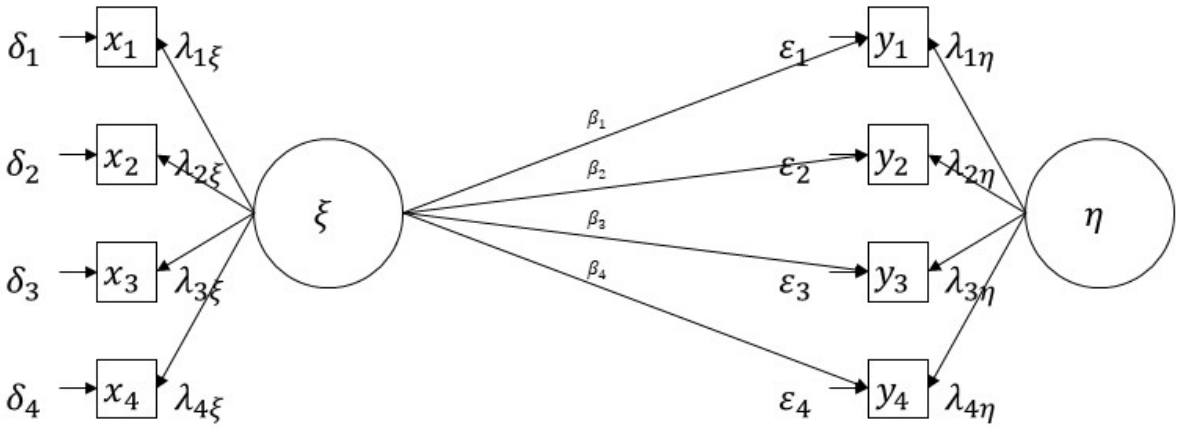


Der Vorteil der Kontrolle auf Indikatorebene gegenüber der Kontrolle auf der Ebene der latenten Konstrukte liegt hierbei in der Würdigung dessen, dass nicht jede Indikatorvariable gleich stark durch eine Tendenz zur Zustimmung beeinflusst werden muss, was durch die verschiedenen Regressionspfade spezifiziert wird. Entsprechend ist lediglich darauf zu achten, dass genügend manifeste Variablen an der Schätzung beteiligt sind, so dass genügend Freiheitsgrade für die Schätzung weiterer Effekte zur Verfügung stehen. Für das Beispiel aus Abbildung 22b mit vier Indikatorvariablen und einer exogenen, manifesten Variable sind zum Beispiel 12 Parameter zu schätzen (vier Faktorladungen, vier Messfehler und vier Regressionskoeffizienten, wenn die Varianz des latenten Konstruktes auf 1 fixiert wird), während die Kovarianz-Matrix insgesamt 15 Informationen enthält (zehn Kovarianzen und fünf Varianzen bei insgesamt fünf manifesten Messungen, welche sich aus den vier Indikatoren *y* und der Tendenz zur Zustimmung „*Akq*“ ergeben). Das Modell könnte also mit drei Freiheitsgraden geschätzt werden.

Die Kontrolle auf Indikatorebene durch Regression ist dabei, wie weiter oben bereits aufgeführt, nicht auf die Verwendung manifester, exogener Variablen und lineare Beziehungen beschränkt. Die lineare Kontrolle auf Basis latenter exogener Variablen (LC-CFA) ist entsprechend dieser Logik genauso möglich, was in vielfältigen Möglichkeiten der statistischen Kontrolle systematischer Messfehler auf Indikatorebene resultiert. Die Kontrolle von Indikatoren,

welche als sensibel gelten, mittels eines latenten Konstruktes zur Schätzung des Bedürfnisses nach sozialer Anerkennung im Kontext der Forschung rund um sozial erwünschtes Antwortverhalten (vgl. Kapitel 2.2.1) ist hier nur ein weiteres Beispiel und in der folgenden Abbildung 23 dargestellt:

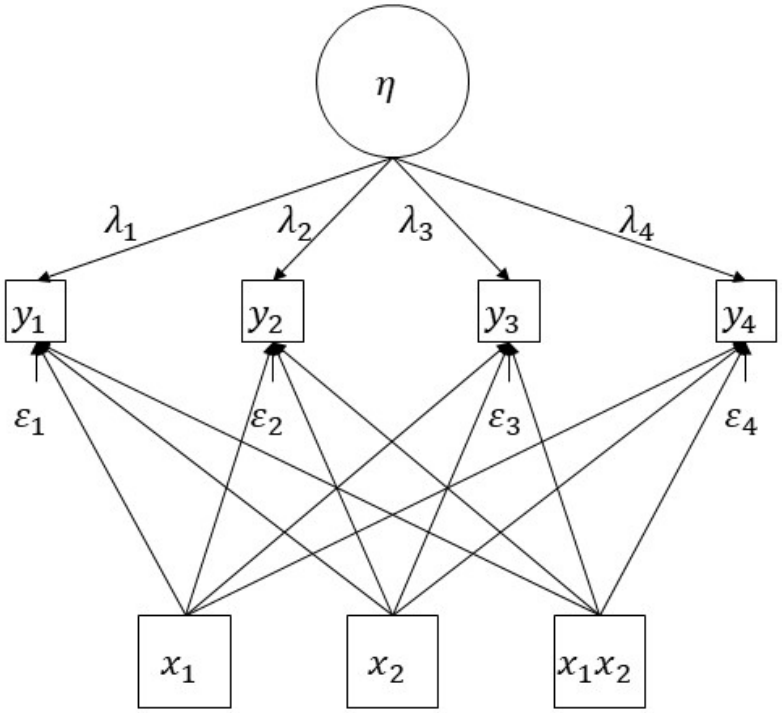
Abbildung 23: LC-CFA zur Bereinigung des Bedürfnisses nach sozialer Anerkennung



Das Bedürfnis nach sozialer Anerkennung wird hier latent durch ξ dargestellt und beispielhaft mit vier Indikatorvariablen geschätzt. Das zu schätzende Merkmal von Interesse ist generisch, durch η dargestellt und ebenfalls beispielhaft durch vier Indikatorvariablen geschätzt. Die Regressionskoeffizienten β geben die diversen methodischen Einflüsse an. Bei Schätzung eines solchen Modells ist dabei jedoch erneut auf eine ausreichende Anzahl an Freiheitsgraden zu achten.

Daneben wurden die Zugänglichkeit von Informationen bezüglich des Objektes von Interesse und das Framing der Befragtenrolle innerhalb der theoretischen Diskussion des Antwortverhaltens und des Auftretens von Antwortverzerrungen (vgl. Kapitel 2.1 und Kapitel 2.2) nicht nur als mögliche Prädiktoren für das Auftreten von FRE, sondern generell als relevant hinsichtlich des Auftretens von Antwortverzerrungen und systematischen Messfehlern diskutiert. Über die Kontrolle von FRE hinaus können die beiden Faktoren somit generell zur Kontrolle von Indikatoren in Fragebatterien hinsichtlich aller möglicher Antwortverzerrungen verwendet werden. Ist es dabei zudem möglich, die beiden Faktoren als manifeste Variablen zu messen statt als latente Variable zu schätzen und wird dabei eine mögliche interaktive Beziehung zwischen beiden Variablen berücksichtigt, so ist ein generisches Beispiel für eine NC-CFA gemäß Abbildung 24 gegeben.

Abbildung 24: Generisches NC-CFA Modell



Die Zugänglichkeit von Informationen gegenüber dem Objekt der Befragung wird für dieses Beispiel manifest durch x_1, das Framing der Befragtenrolle manifest durch x_2 gemessen, x_1x_2 repräsentiert das Produkt der beiden Variablen. Das generische Objekt der Befragung wird beispielhaft durch vier Indikatorvariablen *y* geschätzt und durch die latente Variable η repräsentiert. Die nichtlinearen Effekte zwischen *x* und *y* werden aus Gründen der Übersichtlichkeit der Abbildung nicht dargestellt. Haben die Zugänglichkeit von Informationen gegenüber dem Objekt der Befragung sowie das Framing der Befragtenrolle einen Einfluss auf die Messung von *y*, so wird dieser Einfluss entsprechend durch ein solches Modell kontrolliert, ansonsten verzerrte Faktorladungen der Schätzung werden bereinigt. Entsprechend der beiden vorangegangenen beispielhaften Modelle ist bei der Schätzung eines solchen Modells ebenfalls auf ausreichend verfügbare Freiheitsgrade zu achten.

Neben allen anderen, hier vorgestellten Ergebnissen ist durch die situative Adaption des NLC-CFA Modells zur Bereinigung von Verzerrungen durch FRE also vor allem gezeigt, dass durch unspezifizierte systematische Messfehler verzerrte Faktorladungen durch spezifizierte Regressionen auf Indikatorebene bereinigt werden können. Welche exogenen Variablen dabei zur Kontrolle der Indikatoren verwendet werden, ist situativ abhängig von dem innerhalb der Fra-

gebatterie erwarteten systematischen Messfehler und der theoretischen Diskussion der jeweiligen Einflussfaktoren dieses Messfehlers. Die verwendeten exogenen Variablen sollten dabei lediglich trennscharf vom eigentlich zu schätzenden Merkmal von Interesse abgrenzbar sein, um diskriminante Validität zu etablieren und um Verzerrungen durch Multikollinearität zu vermeiden. Generisch betrachtet wurden jedoch in Kapitel 2 die Zugänglichkeit von Informationen gegenüber dem Objekt der Befragung, das Framing der Befragtenrolle und darüber hinaus der Modus der Informationsverarbeitung als allgemeine Einflussfaktoren auf das Befragtenverhalten diskutiert, so dass sich diese bei ansonsten fehlender Theorie bezüglich des Auftretens beliebiger systematischer Messfehler allgemein zur Bereinigung von Indikatoren eignen. Neben den inhaltlichen Anwendungsmöglichkeiten kann das vorgeschlagene NLC-CFA Modell somit insgesamt als komplexer Anwendungsfall eines Frameworks zur Bereinigung von Faktorladungen und der Quantifizierung systematischer Messfehler durch Regressionen auf Indikatorebene dienen.

Zusammenfassung: Anwendung der NLC-CFA in der Praxis

Zusammenfassend können auf Basis der hier geführten Diskussion die folgenden, weiterführenden Erkenntnisse hinsichtlich der Anwendung der NLC-CFA in der wissenschaftlichen Praxis geschlossen werden:

Erstens wird empfohlen, dass bei Verwendung latenter Konstrukte zur Kontrolle auf Indikatorebene zuvor getestet wird, ob diese selbst von systematischen Messfehlern abhängig sind und ob die entsprechenden standardisierten Faktorladungen ausreichend hoch sind. Sind die latenten Konstrukte nicht unabhängig von systematischen Messfehlern oder verfügen sie über eine zu geringe Faktorladung, so sollten stattdessen manifeste exogene Variablen verwendet werden.

Zweitens wird bei Schätzung eines NLC-CFA Modells empfohlen, auf Multikollinearität zwischen den exogenen Variablen zu testen. Im Optimalfall ist keine Multikollinearität vorhanden. Ist Multikollinearität hingegen gegeben, so sollte diese zumindest durch Fixierung einer Kovarianz zwischen den exogenen Variablen berücksichtigt werden.

Drittens führt die durch den FRE bedingte serielle Mediation zwischen den Indikatorvariablen zu Problemen bei der Schätzung der Parameter, wenn die entsprechenden, systematischen Effekte nicht korrekt (mittels des vorgeschlagenen NLC-CFA Modells) spezifiziert werden. In diesem Kontext ist ebenfalls relevant, dass es einer theoretischen Fundierung bedarf, ob der FRE zwischen verschiedenen Kontext-Zielfragen-Dyaden als gleich oder frei geschätzt wird.

Werden die Effekte als gleich geschätzt, so verringert sich hierdurch die Anzahl an zu schätzenden Parametern sowie der Interpretationsaufwand. Entspricht eine solche Gleichsetzung hingegen nicht der empirischen Realität, so handelt es sich hierbei um eine fehlerhafte Spezifikation des Modells. Es wird entsprechend empfohlen, dass sowohl Modelle geschätzt werden, in welchen die Stärke der FRE nicht zwischen den Dyaden fixiert, sondern frei geschätzt werden, so hierfür genügend Freiheitsgrade vorliegen, als auch, dass Modelle geschätzt werden, in welchen die entsprechenden Effekte als gleich geschätzt werden. Werden dann die Residualvarianzen der entsprechenden Indikatoren unter Berücksichtigung der Vorzeichen der Regressionskoeffizienten zwischen den Modellen mit und ohne Fixierung verglichen kann hieraus geschlossen werden, welche Modellspezifikation eher durch die Daten unterstützt wird.

Viertens ist zu beachten, dass gängige Model Fit Maße nicht für nichtlineare Modelle interpretiert werden sollten. Zur Evaluation des Model Fit für solche Modelle kann grundsätzlich der von Klein und Moosbrugger (2000) vorgeschlagene Log-Likelihood Test durchgeführt werden, auch wenn es sich hierbei letztlich lediglich um eine Einschätzung des relativen Fit im Vergleich zu einem Basismodell ohne Interaktionsterme handelt.

Fünftens und letztens soll erneut betont werden, dass verzerrte Faktorladungen generell durch spezifizierte Regressionen auf Indikatorebene bereinigt werden können. Das vorgeschlagene NLC-CFA Modell stellt sich somit als komplexer Anwendungsfall eines Frameworks zur Bereinigung von Faktorladungen und der Quantifizierung systematischer Messfehler durch Regression der Indikatoren auf weitere, methodenrelevante Variablen dar, welches situativ und je nach erwartetem systematischen Messfehler innerhalb der Fragebatterie adaptiert werden kann.

5.4 Gesamtzusammenfassung

Die einleitend formulierte Forschungsfrage dieser Arbeit lautet: „Wie können FRE möglichst präzise und effizient quantifiziert und kontrolliert werden und wie wirkt sich diese Kontrolle auf die Schätzungen der Faktorladungen und damit die Indikatoren der Datenqualität aus?“

Diesbezüglich wurden in den vorangegangenen Kapiteln zum einen Determinanten für das Auftreten von FRE identifiziert und hierauf Aufbauend ein CFA Modell (NLC-CFA Modell) zur Identifikation und Kontrolle von FRE vorgeschlagen und evaluiert sowie einem etablierten CFA Modell (CFA-MTMM Modell) gegenübergestellt. Die Ergebnisse dieser Evaluation wurden umfassend diskutiert. Entsprechend werden nachfolgend zuletzt alle Ergebnisse dieser Arbeit zusammengefasst:

Erstens wurden auf inhaltlicher Ebene Thesen des Befragtenverhaltens und des Auftretens von Antwortverzerrungen als kleinster gemeinsamer Nenner verschiedener etablierter Verhaltensmodelle aggregiert. FRE wurden dabei konkret in Assimilations- und Kontrasteffekte differenziert. Für Assimilationseffekte wurde diskutiert, dass diese notwendig durch einen geringen Grad an Antwortelaboration bedingt sind und ansonsten auftreten, wenn für Befragte diejenigen Informationen mit Relevanz für das Objekt der Befragung chronisch nicht oder in geringem Maße zugänglich sind und ein indifferentes Framing der Befragtenrolle vorliegt. Für Kontrasteffekte wurde hingegen diskutiert, dass diese notwendig durch einen hohen Grad an Antwortelaboration bedingt sind und ansonsten auftreten, wenn für Befragte diejenigen Informationen mit Relevanz für das Objekt der Befragung chronisch nicht oder in geringem Maße zugänglich sind und ein kooperatives Framing der Befragtenrolle vorliegt (vgl. Kapitel 2.2.3).

Zweitens wurde ein konkretes CFA Modell zur Identifikation und Kontrolle von Fragereihenfolgeneffekten in Fragebatterien auf Basis des Grades an Antwortelaboration, der Informationszugänglichkeit und des Framings der Befragtenrolle eingeführt und diskutiert, das sogenannte NLC-CFA Modell. FRE werden dabei als systematische Messfehler auf Indikatorebene angenommen, welche konditional von den oben genannten Prädiktoren abhängig sind. Eine Bereinigung besagter Antwortverzerrungen findet durch nichtlineare Regression der Indikatoren auf die latent geschätzten Prädiktoren statt. Dies soll zu einer Bereinigung verzerrter Parameterschätzungen, allen voran der Faktorladungen beitragen, da die Varianz der Indikatoren durch Spezifikation der entsprechenden Methodeneffekte korrekt in den Anteil an Common Variance, den Anteil an systematischer Fehlervarianz sowie den Anteil an zufälliger Fehlervarianz zerlegt werden kann. Zwecks Evaluation der gesamten Effektivität des NLC-CFA Modells zur Identifikation von FRE und zur Bereinigung der Faktorladungen wurde auf Basis der etablierten MTMM Analyse zum Vergleich ein CFA-MTMM Modell adaptiert. Der Anspruch an das NLC-CFA Modell ist entsprechend, Faktorladungen effizienter zu schätzen, als dies durch das etablierte MTMM Verfahren möglich ist (vgl. Kapitel 3.3).

Drittens zeigt sich auf Basis umfangreicher Simulationsanalysen, dass vorhandene systematische Messfehler tatsächlich dort Faktorladungen verzerren, wo diese unberücksichtigt bleiben. Hingegen werden die Verzerrungen fast vollständig bereinigt, wenn systematische Messfehler (hier: der FRE) auf Basis des NLC-CFA Modell kontrolliert werden, was insgesamt zu effizienteren Schätzungen der Faktorladungen führt, wodurch Messungen der Konstruktvalidität und Kompositreliabilität ebenfalls effizienter werden. Die Schätzung von Modellen mit lediglich drei Indikatorvariablen sollte bei Anwendung des Modells jedoch vermieden werden, da dies die Effizienz der Schätzung der Faktorladungen verringert. Daneben zeigt sich, dass Schätzung

des adaptierten CFA-MTMM Modells lediglich zu einer teilweisen Bereinigung bei stark verzerrten Faktorladungen beitragen. Die Effizienz der Schätzung von Faktorladungen durch NLC-CFA Modelle ist dabei insgesamt der Effizienz der Schätzung durch CFA-MTMM Modelle überlegen (vgl. Kapitel 4.4, Kapitel 4.7 und Kapitel 5.1).

Viertens zeigt sich, dass die Spezifikation der systematischen Messfehler mittels NLC-CFA nicht nur Verzerrungen von Faktorladungen bereinigt, sondern dass die geschätzten Effekte der systematischen Messfehler ebenfalls präzise geschätzt werden. Eine spezifische Analyse des Vorhandenseins und der Abhängigkeiten von systematischen Messfehlern ist somit mittels NLC-CFA möglich. Zu beachten ist dabei jedoch, dass die konkreten Effekte tendenziell unterschätzt werden, so dass an dieser Stelle lediglich von einer annährend effizienten Schätzung der entsprechenden Effekte durch die NLC-CFA ausgegangen werden kann. Bezüglich der Modellkomplexität kann Empfohlen werden, NLC-CFA Modelle mit vier statt fünf Indikatorvariablen zu schätzen, da dies mit einer Erhöhung der Effizienz und zumindest mit keiner substanziellen Verschlechterung der Präzision der geschätzten Effekte einhergeht. Es zeigt sich dabei insgesamt bezüglich der Überprüfbarkeit der in Kapitel 2 abgeleiteten notwendigen Bedingungen und Hypothesen auf Basis des NLC-CFA Modells, dass bei gegebener Modell- und Simulationsspezifikation alle Hypothesen mit korrektem Resultat, das heißt ohne das Auftreten von Typ 1 oder Typ 2 Fehlern, getestet werden konnten (vgl. Kapitel 4.5, Kapitel 4.6, Kapitel 4.7, Kapitel 5.1 und Kapitel 5.2).

Fünftens wird bei konkreter Schätzung des NLC-CFA Modells auf Basis realer empirischer Daten empfohlen, dass bei Verwendung latenter Konstrukte zur Kontrolle auf Indikatorebene zuvor getestet werden sollte, ob besagte Konstrukte (bzw. deren Indikatoren) ihrerseits selbst von systematischen Messfehlern beeinflusst sind und ob die entsprechenden standardisierten Faktorladungen ausreichend hoch sind. Sind die Schätzungen der latenten Konstrukte nicht unabhängig von systematischen Messfehlern oder verfügen sie über eine zu geringe Faktorladung, so sollten stattdessen manifeste exogene Variablen verwendet werden. Die verwendeten exogenen latenten Konstrukte sollten darüber hinaus auf Multikollinearität überprüft werden. Ist Multikollinearität gegeben, so sollte diese zumindest durch Fixierung einer Kovarianz zwischen den exogenen Variablen berücksichtigt werden. Zusätzlich wird empfohlen, dass sowohl Modelle geschätzt werden, in welchen die Stärke der FRE nicht zwischen den Dyaden fixiert, sondern frei geschätzt werden, so hierfür genügend Freiheitsgrade vorliegen, als auch, dass Modelle geschätzt werden, in welchen die entsprechenden Effekte als gleich geschätzt werden. Die Interpretation der Residualvarianzen der entsprechenden Indikatoren zwischen den Modellen

mit und ohne Fixierung gibt dann Aufschluss darüber, welches der Modelle eher durch die Daten unterstützt wird. Hinsichtlich des Model Fit des geschätzten NLC-CFA Modells wird empfohlen, einen Log-Likelihood Test durchzuführen, auch wenn es sich hierbei letztlich lediglich um eine Einschätzung des relativen Fit im Vergleich zu einem Basismodell ohne Interaktionsterme handelt (vgl. Kapitel 5.3).

Sechstens und letztens soll erneut betont werden, dass das eingeführte und evaluierte NLC-CFA Modell zur Identifikation und Kontrolle von FRE vor allem als spezieller Anwendungsfall eines allgemeinen Frameworks zur Messfehlerbereinigung auf Indikatorebene durch regressive Beziehungen angesehen werden kann. Das eingeführte Modell kann dabei situativ beliebig adaptiert werden, etwa als „**c**ontrolled CFA“ (**C-CFA**, Kontrolle von Indikatorvariablen durch lineare Regressionen statt nichtlinearer Regressionen mit manifesten exogenen Variablen statt mit latenten exogenen Variablen), als „**l**atent **c**ontrolled CFA“ (**LC-CFA**, Kontrolle von Indikatorvariablen durch lineare Regressionen statt nichtlinearer Regressionen mit latenten exogenen Variablen) oder als „**n**onlinear **c**ontrolled CFA „(**NC-CFA**, Kontrolle von Indikatorvariablen durch nichtlineare Regressionen mit manifesten exogenen Variablen statt mit latenten exogenen Variablen), da verzerrte Faktorladungen generell (und nicht nur auf Basis latenter Interaktionen) durch spezifizierte Regressionen auf Indikatorebene bereinigt werden können. Die zur Kontrolle verwendeten exogenen Variablen werden bei einer solchen Adaption situativ und je nach erwartetem systematischen Messfehler innerhalb der Fragebatterie ausgewählt, wobei bei fehlender Theorie bezüglich des zu kontrollierenden systematischen Messfehlers generell auf den Modus der Informationsverarbeitung, die Zugänglichkeit von Informationen bezüglich dem Objekt der Befragung und das Framing der Befragtenrolle als allgemeine Determinanten des Antwortverhaltens zurückgegriffen werden kann.

6 Literaturverzeichnis

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50(2)*, S. 179-211.

Ajzen, I., & Fishbein, M. (1980). *Understanding Attitudes and Predicting Social Behavior.* London: Prentice-Hall.

Allen, N., & Meyer, J. (1996). Affective, Continuance, and Normative Commitment to the Organization: An Examination of Construct Validity. *Journal of Vocational Behavior, 49(3)*, S. 252-276.

Andersen, H., & Mayerl, J. (2017). Social Desirability and Undesirability Effects on Survey Response latencies. *Bulletin of Sociological Methodology, 135(1)*.

Andersen, H., & Mayerl, J. (2019). Responding to Socially Desirable and Undesirable Topics: Different Types of Response Behaviour? *Methods, data, analyses: a journal for quantitative methods and survey methodology (mda), 13(1)*, S. 7-35.

Aquilina, W. (1994). Interview Mode Effects in Surveys of Drug and Alcohol use: a Field Experiment. *Public Opinion Quarterly, 58(2)*, S. 210-240.

Atteslander, P. (2010). *Methoden der empirischen Sozialforschung (13. Aufl.).* Berlin: Erich Schmidt Verlag.

Ayidiya, S., & McClendon, M. (1990). Response Effects in Mail Surveys. *Public Opinion Quarterly, 54(2)*, S. 229–247.

Bagozzi, R., Yi, Y., & Phillips, L. (1991). Assessing Construct Validity in Organizational Research. *Administrative Science Quarterly, 36(3)*, S. 421-458.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review, 84(2)*, S. 191-215.

Bandura, A. (1994). Self-efficacy. In V. Ramachaudran, *Encyclopedia of human behavior (Vol. 4)* (S. 71-81). New York: Academic Press.

Bard, G., & Weinstein, Y. (2017). The effect of question order on evaluations of test performance: Can the bias dissolve? *The Quarterly Journal of Experimental Psychology, 70(10)*, S. 2130-2140.

Barendse, M., Oort, F., & Garst, G. (2010). Using restricted factor analysis with latent moderated structures to detect uniform and nonuniform measurement bias; a simulation study. *AStA Advances in Statistical Analysis, 94(2)*, S. 117-127.

Bargh, J. (1989). Conditional Automaticity: Varieties of Automatic Influence in Social Perception and Cognition. In J. Uleman, & J. Bargh, *Unintended Thought* (S. 3-51). New York: Guilford Press.

Bar-Hillel, M., Peer, E., & Acquisti, A. (2014). "Heads or tails?"—A reachability bias in binary choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40(6)*, S. 1656–1663.

Barnett, L., Vazou, S., Abbott, G., Bowe, S., Robinson, L., Ridgers, N., & Salmon, J. (2016). Construct validity of the pictorial scale of Perceived Movement Skill Competence. *Psychology of Sport and Exercise, 22*, S. 294-302.

Bassili, J., & Krosnick, J. (2000). Do Strength-Related Attitude Properties Determine Susceptibility to Response Effects? New Evidence From Response Latency, Attitude Extremity, and Aggregate Indices. *Political Psychology, 21(1)*, S. 107-132.

Behrendt, S. (2022). *Package 'lm.beta'. Add Standardized Regression Coefficients to lm-Objects.* Von https://cran.r-project.org/web/packages/lm.beta/lm.beta.pdf abgerufen

Best, H., & Mayerl, J. (2013). Values, beliefs, attitudes: An empirical study on the structure of environmental concern and recycling participation. *Social Science Quarterly, 94(3)*, S. 691–714.

Biesanz, J. (2012). Autoregressive longitudinal models. In R. Hoyle, *Handbook of structural equation modeling* (S. 459–471). New York: The Guilford Press.

Billiet, J., & Davidov, E. (2008). Testing the stability of an acquiescence style factor behind two interrelated substantive variables in a panel design. *Sociological Methods & Research, 36(4)*, S. 542-562.

Bishop, G., & Smith, A. (2001). Response-order effects and the early Gallup split-ballots. *Public Opinion Quarterly, 65(4)*, 479-505.

Bogner, K., & Landrock, U. (2014). *Antworttendenzen in standardisierten Umfragen.* Hg. v. GESIS - Leibniz Institute for the Social Sciences.

Bollen, K. (1989). *Structural Equations with Latent Variables.* New York: Wiley.

Borger, N., Sikkel, D., & Hox, J. (2004). Response Effects in Surveys on Children and Adolescents: The Effect of Number of Response Options, Negative Wording, and Neutral Mid-Point. *Quality and Quantity, 38(1)*, S. 17-33.

Bradburn, N. (1979). Respondent Burden. In L. Reeder, *Health Survey Research Methods: Second Biennial Conference.* Williamsburg, Va: U.S. Government Printing Office.

Bradburn, N. (1983). Response Effects. In P. Rossi, J. Wright, & A. Anderson, *Handbook of Survey Research* (S. 289-328). Orlando, Florida: Academic Press.

Brookes, S., Chalmers, K., Avery, K., Coulman, K., & Blazeby, J. (2018). Impact of question order on prioritisation of outcomes in the development of a core outcome set: a randomised controlled trial. *Trials, 66(19)*.

Brown, T. (2003). Confirmatory factor analysis of the Penn State Worry Questionnaire: Multiple factors or method effects? *Behaviour Research and Therapy, 41(12)*, S. 1411-1426.

Brown, T. (2006). *Confirmatory Factor Analysis for Applied Research.* New York, London: Guilford Press.

Browne, M., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research, 21(2)*, S. 230–258.

Buchner, A., Erdfelder, E., & Faul, F. (1996). Teststärkeanalysen. In E. Erdfelder, R. Mausfeld, T. Meiser, & G. Rudinger, *Handbuch Quantitative Methoden* (S. 123–136). Weinheim: Psychologie Verlags Union.

Bühner, M. (2004). *Einführung in die Test- und Fragebogenkonstruktion.* München: Pearson Studium.

Byrne, B., Shavelson, R., & Muthén, B. (1989). Testing for Equivalence of Factor Covariance and Mean Structures: The Issue of Partial Measurement Invariance. *Psychological Bulletin, 195(3)*, S. 456-466.

Cacioppo, J., Petty, R., Feinstein, J., & Jarvis, W. (1996). Dispositional differences in cognitive motivation: the life and times of individuals varying in need for cognition. *Psychological Bulletin, 119(2)*, S. 197-253.

Campanelli, P., & O'Muircheartaigh, C. (1999). Interviewers, Interviewer Continuity, and Panel Survey Nonresponse. *Quality and Quantity, 33(1)*, S. 59-76.

Campbell, D., & Fiske, D. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56(2)*, S. 81–105.

Cannell, C., Miller, P., & Oksenberg, L. (1981). Research on interviewing techniques. In S. Leinhardt, *Social Methodology.* San Francisco: Jossey-Bass Publishers.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39(5)*, S. 752-766.

Chaiken, S., & Trope, Y. (1999). *Dual process theories in social psychology.* New York & London: Guilford Press.

Cham, H., West, S., Ma, Y., & Aiken, L. (2012). Estimating latent variable interactions with non-normal observed data: A comparison of four approaches. *Multivariate Behavioral Research, 47(6)*, S. 840-876.

Chen, P. (2010). Item Order Effects on Attitude Measures. Electronic Theses and Dissertations, 778. Abgerufen am 22. August 2022 von https://digitalcommons.du.edu/etd/778

Chen, S., & Chaiken, S. (1999). The Heuristic-Systematic Model in its broader context. In S. Chaiken, & Y. Trope, *Dual-process theories in social psychology.* New York & London: The Guilford Press.

Cho, E. (2016). Making Reliability Reliable: A Systematic Approach to Reliability Coefficients. *Organizational Research Methods, 19(4)*, S. 651-682.

Cieciuch, J., Davidov, E., Algesheimer, R., & Schmidt, P. (2018). Testing for Approximate Measurement Invariance of Human Values in the European Social Survey. *Sociological Methods & Research, 47(4)*, S. 665–686.

Clark, H. (1985). Language use and language users. In G. Lindzey, & E. Aronson, *Handbook of social psychology (Vol. 2)* (S. 179-232). New York: Random House.

Cooke, R., & Sheeran, P. (2004). Moderation of cognition-intention and cognition-behaviour relations: A meta-analysis of properties of variables from the theory of planned behaviour. *British Journal of Social Psychology, 43(2)*, S. 159-186.

Cronbach, L., & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52(4)*, S. 281-302.

Cunningham, J., Kushnir, V., & McCambridge, J. (2016). The Impact of Asking About Interest in Free Nicotine Patches on Smoker's Stated Intent to Change: Real Effect or Artefact of Question Ordering? *Nicotine & Tobacco Research, 18(5)*, S. 1215-1217.

Diekmann, A. (2007). *Empirische Sozialforschung. Grundlagen, Methoden, Anwendung (15. Aufl.).* Reinbek bei Hamburg: Rowohlt.

Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften.* Berlin: Springer.

Eagly, A., & Chaiken, S. (1993). *The Psychology of Attitudes.* Fort Worth: Harcourt Brace Jovanovich.

Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika, 65(2)*, 241–261.

Eirmbter, W., & Jacob, R. (1996). Fragen zu Fragen: Instrumentbedingte Antwortverzerrungen? *ZUMA-Nachrichten, 38(20)*, S. 90-111.

Erbslöh, B., & Koch, A. (1988). Die Non-Response-Studie zum Allbus 1986: Problemstellung, Design, erste Ergebnisse. *ZUMA-Nachrichten, 22(12)*, S. 29-44.

Esser, H. (1991). Die Erklärung systematischer Fehler in Interviews. Befragtenverhalten als "rational choice". In R. Wittenberg, *Person - Situation - Institution - Kultur : Günter Büschges zum 65. Geburtstag* (S. 59–78). Berlin: Duncker & Humblot.

Esser, H. (1996). Die Definition der Situation. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 48(1)*, S. 1-34.

Esser, H. (2001). *Soziologie. Spezielle Grundlagen. Band 6: Sinn und Kultur.* . Frankfurt a.M.: Campus.

Esser, H. (2006a). Eines für alle(s)? Das Weber-Paradigma, das Konzept des moderaten methodologischen Holismus und das Modell der soziologischen Erklärung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 58(2)*, S. 352–364.

Esser, H. (2006b). Affektuelles Handeln: Emotionen und das Modell der Frame-Selektion. In R. Schützeichel, *Emotionen und Sozialtheorie. Disziplinäre Ansätze* (S. 143–174). Frankfurt a.M.: Campus.

Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior, 26(2)*, S. 132-139.

Fazio, R. (1986). How do attitudes guide behavior? In R. Sorrentino, & E. Higgins, *The handbook of motivation and cognition: Foundation of social behavior* (S. 204-243). New York: Guilford Press.

Fazio, R. (1990). Multiple Processes by which Attitudes guide Behavior: the MODE Model as an integrative framework. *Advances in Experimental Social Psychology, 23*, S. 75-109.

Fazio, R., & Olson, M. (2014). The MODE Model. Attitude-Behavior Processes as a Function of Motivation and Opportunity. In J. Sherman, B. Gawronski, & Y. Trope, *Dual-Process Theories of the Social Mind.* New York & London: The Guilford Press.

Finney, S., & DiStefano, C. (2008). Non-normal and categorical data in structural equation modeling. In G. Hancock, & R. Mueller, *Structural Equation Modeling: A Second Course* (S. 269-314). Information Age Publishing.

Fischer, L., & Lück, H. (2014). Allgemeine Arbeitszufriedenheit. In *Zusammenstellung sozialwissenschaftlicher Items und Skalen.*

Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behavior: The reasoned action approach.* New York: Psychology Press.

Friedrichs, J. (1990). *Methoden empirischer Sozialforschung (14. Aufl.).* Opladen: Westdeutscher Verlag.

Garbarski, D., Schaeffer, N., & Dykema, J. (2015). The effects of response option order and question order on self-rated health. *Quality of Life Research, 24(6)*, S. 1443-1453.

Gawronski, B., & Strack, F. (2004). On the propositional nature of cognitive consistency: Dissonance changes explicit, but not implicit attitudes. *Journal of Experimental Social Psychology, 40*, S. 535-542.

Geiser, C., Eid, M., West, S., Lischetzke, T., & Nussbeck, F. (2012). A comparison of method effects in two confirmatory factor models for structurally different methods. *Structural Equation Modeling, 19(3)*, S. 409–436.

Glantz, A., & Michael, T. (2014). Interviewereffekte. In N. Baur, & J. Blasius, *Handbuch Methoden der empirischen Sozialforschung* (S. 313–322). Wiesbaden: Springer Fachmedien.

Goyder, J. (1986). Surveys on Surveys: Limitations and Potentialities. *Public Opinion Quarterly, 50(1)*, S. 27-41.

Greenberger, E., Chen, C., Dmitrieva, J., & Farruggia, S. (2003). Item-wording and the dimensionality of the Rosenberg Self-Esteem Scale: Do they matter? *Personality and Individual Dierences, 35(6)*, S. 1241–1254.

Greenleaf, E. (1992). Measuring extreme response style. *Public Opinion Quarterly, 56(3)*, S. 328-351.

Groves, R. (1989). *Survey errors and survey costs.* New York: Wiley.

Groves, R., Cialdini, R., & Couper, M. (1992). Understanding the Decision to Participate in a Survey. *Public Opinion Quarterly, 56(4)*, S. 475-595.

Gulliksen, H. (1950). *Theory of Mental Tests.* New York: Wiley.

Hartig, J., & Jude, N. (2007). Empirische Erfassung von Kompetenzen und psychometrische Kompetenzmodelle. In J. Hartig, & E. Klieme, *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik. Eine Expertise im Auftrag des Bundesministeriums für Bildung und Forschung (Bildungsforschung, 20).* Bonn, Berlin.

Hartmann, P. (1994). Interviewing when the spouse is present. *International Journal of Public Opinion Research, 6(3)*, S. 298-306.

Haunberger, S. (2006). Das standardisierte Interview als soziale Interaktion: Interviewereffekte in der Umfrageforschung. *ZA-Information / Zentralarchiv für Empirische Sozialforschung, 58*, S. 23–46.

Hayes, A. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach.* New York: The Guilford Press.

Heerwegh, D., & Loosveldt, G. (2009). Explaining the intention to participate in a web survey: a test of the theory of planned behaviour. *International Journal of Social Research Methodology 12(3)*, S. 181-195.

Higgins, E., & Bargh, J. (1987). Social cognition and social perception. *Annual Review of Psychology, 38*, S. 369-345.

Hilton, D., & Slugoski, B. (1986). Knowledge-based causal attribution: The abnormal conditions focus model. *Psychological Review, 93(1)*, S. 75-88.

Hlawatsch, A., & Krickl, T. (2019). Einstellungen zu Befragungen. In N. Baur, & J. Blasius, *Handbuch Methoden der empirischen Sozialforschung* (S. 357-364). Wiesbaden: Springer.

Hox, J., & Bechger, T. (1999). An Introduction to Structural Equation Modelling. *Family Science Review, 11*, S. 354-373.

Hsiao, Y., & Lai, M. (2018). The Impact of Partial Measurement Invariance on Testing Moderation for Single and Multi-Level Data. *Frontiers in Psychology, 740(9)*.

Huang, F., & Cornell, D. (2016). Question Order Affects the Measurement of Bullying Victimization Among Middle School Students. *Educational and Psychological Measurement, 76(5)*, S. 724-740.

Huddy, L., Billig, J., Bracciodieta, J., Hoeffler, L., Moynihan, P., & Pugliani, P. (1997). The Effect of Interviewer Gender on the Survey Response. *Political Behavior, 19(3)*, S. 197-220.

Jäckle, A., Lynn, P., Sinibaldi, J., & Tipping, S. (2011). The effect of interviewer personality, skills and attitudes on respondent co-operation with face-to-face surveys. *ISER Working Paper Series No. 2011-14*.

Jackman, M., Leite, W., & Cochrane, D. (2011). Estimating Latent Variable Interactions with the Unconstrained Approach: A Comparison of Methods to Form Product Indicators for Large, Unequal Numbers of Items. *Structural Equation Modeling: A Multidisciplinary Journal, 18(2)*, S. 274-288.

Johnson, A., Villanti, A., Glasser, A., Pearson, J., & Delnevo, C. (2019). Impact of Question Type and Question Order on Tobacco Prevalence Estimates in US Young Adults: A Randomized Experiment. *Nicotine and Tobacco Research, 21(8)*, S. 1144-1146.

Jöreskog, K. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika, 36(2)*, S. 109-133.

Jöreskog, K., & Yang, F. (1996). Nonlinear structual equation models: The Kenny-Judd model with interaction effects. In G. Markoulides, & R. Schumacker, *Advanced structural equation modelling* (S. 57-87). Mawah, NJ: Lawrence Erlbaum Association.

Kahlor, L., Dunwoody, S., Griffin, R., Neuwirth, K., & Giese, J. (2003). Studying heuristic-systematic processing of risk communication. *Risk Analysis, 23(2)*, S. 355-368.

Kahn, R., & F., C. C. (1968). Interviewing. *International Encyclopaedia of the Social Sciences 8*, S. 149-161.

Kahneman, D. (2011). *Thinking, fast and slow.* New York: Farrar, Straus and Giroux.

Kam, C., & Meyer, J. (2015). How Careless Responding and Acquiescence Response Bias Can Influence Construct Dimensionality: The Case of Job Satisfaction. *Organizational Research Methods, 18(3)*, S. 512-541.

Kelava, A., Werner, C., Schermelleh-Engel, K., Moosbrugger, H., Zapf, D., Ma, Y., West, S. (2011). Advanced Nonlinear Latent Variable Modeling: Distribution Analytic LMS and QML Estimators of Interaction and Quadratic Effects. *Structural Equation Modeling: A Multidisciplinary Journal, 18(3)*, S. 465-491.

Keller, J., Bohner, G., & Erb, H. (2000). Intuitive und heuristische Urteilsbildung - verschiedene Prozesse? Präsentation einer deutschen Fassung des "Rational-Experiential Inventory" sowie neuer Selbstberichtskalen zur Heuristiknutzung. *Zeitschrift für Sozialpsychologie, 31(2)*, S. 87-101.

Kenny, D., & Judd, C. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin, 96(1)*, S. 201-210.

Keusch, F. (2012). How to Increase Response Rates in List-Based Web Survey Samples. *Social Science Computer Review, 30(2)*.

Kiesler, S., & Sproull, L. (1986). Response Effects in the Electronic Survey. *Public Opinion Quarterly, 50(3)*, S. 402-413.

Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika, 65(4)*, S. 457-474.

Klein, A., & Muthén, B. (2008). Quasi-Maximum Likelihood Estimation of Structural Equation Models with Multiple Interaction and Quadratic Effects. *Multivariate Behavioral Research, 42(4)*, S. 647-673.

Klein, A., & Schermelleh-Engel, K. (2010). Introduction of a new measure for detecting poor fit due to omitted nonlinear terms in SEM. *AStA – Advances in Statistical Analysis, 94(2)*, S. 157-166.

Kline, R. (2016). *Principles and practice of structural equation modeling (4th ed.).* New York: Guilford Press.

Koch, T., Holtmann, J., Bohn, J., & Eid, M. (2017). Explaining general and specific factors in longitudinal, multimethod, and bifactor Models: Some Caveats and Recommendations. *Psychological Methods. Advance online publication.*

Koch, T., Kelava, A., & Eid, M. (2018). Analyzing Different Types of Moderated Method Effects in Confirmatory Factor Models for Structurally Different Methods. *Structural Equation Modeling: A Multidisciplinary Journal, 25(2)*, S. 179-200.

Komsta, L., & Novomestky, F. (2022). *Package 'moments'. Moments, Cumulants, Skewness, Kurtosis and Related Tests.* Von https://cran.r-project.org/web/packages/moments/moments.pdf abgerufen

Krebs, D., & Menold, N. (2014). Gütekriterien quantitativer Sozialforschung. In N. Baur, & J. Blasius, *Handbuch Methoden der empirischen Sozialforschung* (S. 425–438). Wiesbaden: Springer Fachmedien.

Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and web surveys. *Public Opinion Quarterly, 72(5)*, S. 847-865.

Krishna, A., & Strack, F. (2017). Reflection and Impulse as Determinants of Human Behavior. In P. Meusburger, B. Werlen, & L. Suarsana, *Knowledge and Action. Knowledge and Space, Vol 9.* Springer.

Kroneberg, C. (2005). Die Definition der Situation und die variable Rationalität der Akteure. Ein allgemeines Modell des Handelns. *Zeitschrift für Soziologie, 34(5)*, S. 344–363.

Kroneberg, C. (2007). Wertrationalität und das Modell der Frame-Selektion. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 59*, S. 215–239.

Kroneberg, C. (2011). Das Modell der Frame-Selektion (MFS). In C. Kroneberg, *Die Erklärung sozialen Handelns* (S. 119-164). VS Verlag für Sozialwissenschaften.

Krosnick, J. (1991). Response Strategies for Coping with the Cognitive Demands Of Attitude Measures in Surveys. *Applied Cognitive Psychology, 5(3)*, S. 213-236.

Krosnick, J. (1999a). Survey Research. *Annual Review of Psychology, 50*, S. 537-567.

Krosnick, J. (1999b). Maximizing questionnaire quality. In J. Robinson, P. Shaver, & L. Wrightsman, *Measures of Political Attitudes* (S. 35-57). San Diego: Academic Press.

Krosnick, J., & Fabrigar, L. (1997). Designing rating scales for effective measurement in surveys. In L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, & D. Trewin, *Survey Measurement and Process Quality.* New York: Wiley.

Krosnick, J., & Fabrigar, L. (1998). *Designing Good Questionnaires: Insights from Psychologie.* New York: Oxford University Press.

Krosnick, J., & Petty, R. (1995). Attitude Strength: An Overview. In R. Petty, & J. Krosnick, *Attitude strength: Antecedents and consequences* (S. 1-24). Hillsdale: Lawrence Erlbaum.

Krosnick, J., & Presser, S. (2010). Question and questionnaire design. In P. Marsden, & J. Wright, *Handbook of Survey Research* (S. 263-313). Bingley: Emerald Group Publishing.

Krosnick, J., & Schuman, H. (1988). Attitude intensity, importance, and certainty and susceptibility to response effects. *Journal of Personality and Social Psychology, 54(6)*, S. 940-952.

Krosnick, J., Narayan, S., & Smith, W. (1996). Satisficing in surveys: Initial evidence. In M. Braverman, & J. Slater, *Advances in survey research* (S. 29-44). San Francisco: Jossey-Bass.

Kruglanski, A., & Freund, T. (1983). The freezing and unfreezing of lay-inferences: Effects of impressional primacy, ethnic stereotyping, and numerical anchoring. *Jourrnal of Experimental Social Psvchology, 19(5)*, S. 448-468.

Kumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality and Quantity, 47(4)*, S. 2025-2047.

Lai, M., Richardson, G., & Mak, H. (2019). Quantifying the impact of partial measurement invariance in diagnostic research: An application to addiction research. *Addictive Behaviors, 94*, S. 50-56.

Lander, B. (2000). Anwesenheitseffekte im Wandel. Eine Sekundäranalyse zur Anwesenheit des Partners im Interview anhand des ALLBUS 1980 bis 1998. *Zeitschrift für Soziologie, 29(3)*, S. 227-239.

LeBel, E. (2010). Attitude Accessibility as a Moderator of Implicit and Explicit Self-esteem Correspondence. *Self and Identity, 9(2)*, S. 195-208.

Lee, S., Schwarz, N., & Goldstein, L. (2014). Culture-Sensitive Question Order Effects of Self-Rated Health Between Older Hispanic and Non-Hispanic Adults in the United States. *Journal of Aging and Health, 26(5)*, S. 860-883.

Li, C. (2016a). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavioral Research Methods, 48(3)*, S. 936–949.

Li, C. (2016b). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychological Methods, 21(3)*, S. 369-387.

Lienert, G. (1969). *Testaufbau und Testanalyse (3. durch einen Anhang über Faktorenanalyse ergänzte Aufl).* Weinheim, Berlin und Basel: Julius Beltz.

Lienert, G., & Raatz, U. (1994). *Testaufbau und Testanalyse.* Weinheim: Beltz.

Little, T. (2013). *Longitudinal structural equation modeling.* New York: Guilford Press.

Little, T., Bovaird, J., & Widaman, K. (2006). On the Merits of Orthogonalizing Powered and Product Terms: Implications for Modeling Interactions Among Latent Variables. *Structural Equation Modeling: A Multidisciplinary Journal, 13(4)*, S. 497-519.

Liu, Y., & Thompson, M. (2022). The Impact of Partial Measurement Invariance on Between-group Comparisons of Second-Order Factor Means. *Structural Equation Modeling: A Multidisciplinary Journal, 29(1)*, S. 86-100.

Lord, F., & Novick, M. (1968). *Statistical theories of mental test scores.* Reading, Mass.: Addison-Wesley.

Lubke, G., & Muthén, B. (2004). Applying multigroup confirmatory factor models for continuous outcomes to Likert scale data complicates meaningful group comparisons. *Structural Equation Modeling, 11(4)*, S. 514–534.

Maccoby, E., & Maccoby, N. (1976). Das Interview: ein Werkzeug der Sozialforschung. In R. König, *Das Interview. Formen, Technik, Auswertung (6. Aufl.)* (S. 37-85). Köln: Kiepenheuer & Witsch GmbH.

Magelssen, M., Supphellen, M., Nortvedt, P., & Materstvedt, L. (2016). Attitudes towards assisted dying are influenced by question wording and order: a survey experiment. *BMC Medical Ethics, 24(17)*.

Mangione, T., Fowler, F., & Louis, T. (1992). Question characteristics and interviewer effects. *Journal of Official Statistics, 8(3)*, S. 293-307.

Marsh, H. (1986). Negative item bias in ratings scales for preadolescent children: A cognitive-developmental phenomenon. *Developmental Psychology, 22(1)*.

Marsh, H. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology, 70(4)*, S. 810-819.

Marsh, H., & Bailey, M. (1991). Confirmatory factor analysis of multitrait-multimethod data: A comparison of alternative models. *Applied Psychological Measurement, 15(1)*, S. 47–70.

Marsh, H., & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. Hoyle, *Structural equation modeling: Concepts, issues, and applications* (S. 177–187). Thousand Oaks und London: Sage.

Marsh, H., Wen, Z., & Hau, K. (2004). Structural equation models of latent interactions: evaluation of alternative estimation strategies and indicator construction. *Psychological Methods, 9(3)*, S. 275-300.

Marsh, H., Wen, Z., & Hau, K. (2006). Structural Equation Models of Latent Interaction and Quadratic Effects. In G. Hancock, & R. Mueller, *Structural Equation Modeling: A Second Course* (S. 225-265). Greenwich, CT: Information Age.

Marsh, H., Wen, Z., Hau, K., Little, T., Bovaird, J., & Widaman, K. (2007). Unconstrained Structural Equation Models of Latent Interactions: Contrasting Residual- and Mean-Centered Approaches. *Structural Equation Modeling, 14(4)*, S. 570-580.

Martin, L. (1986). Set/Reset: The use and disuse of concepts in impression formation. *Journal of Personality and Social Psychology, 51(3)*, S. 493-504.

Maslowsky, J., Jager, J., & Hemken, D. (2015). Estimating and interpreting latent variable interactions: A tutorial for applying the latent moderated structural equations method. *International Journal of Behavioral Development, 39(1)*, S. 87-96.

Mayerl, J. (2009). *Kognitive Grundlagen sozialen Verhaltens. Framing, Einstellungen und Rationalität.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Mayerl, J. (2013). Response Latency Measurement in Surveys. Detecting Strong Attitudes and Response Effects. *Survey Methods: Insights from the Field.* .

Mayerl, J., & Giehl, C. (2018). A Closer Look at Attitude Scales with Positive and Negative Items. Response Latency Perspectives on Measurement Quality. *Survey Research Methods 12(3)*, S. 9999 - 10016.

Mayerl, J., & Urban, D. (2008). *Antwortreaktionszeiten in Survey-Analysen. Messung, Auswertung und Anwendungen* . Wiesbaden: VS Verlag für Sozialwissenschaften.

Mayerl, J., & Urban, D. (2019). Der Mythos von der „gefährlichen“ Multikollinearität bei der Schätzung von Interaktionseffekten. In *Beitrag für die Festschrift „Grundlagen –*

Methoden – Anwendungen in den Sozialwissenschaften. Festschrift zu Ehren von Steffen M. Kühnel. Wiesbaden: VS Springer Verlag.

Mayerl, J., Andersen, H., & Giehl, C. (2019). Identification of Measurement Problems of Survey Items and Scales Using Paradata. In N. Menold, & T. Wolbring, *Qualitätssicherung sozialwissenschaftlicher Erhebungsinstrumente* (S. 9-36). Wiesbaden: VS Verlag für Sozialwissenschaften.

Maynard, D., Houtkoop-Steenstra, H., & Schaeffer, N. (2002). *Standardization and tacit knowledge. Interaction and practise in the survey interview. .* New York: Wiley.

Mazurek, M., Carlson, C., Baker-Ericzén, M., Butter, E., Norris, M., & Kanne, S. (2020). Construct Validity of the Autism Impact Measure (AIM). *Journal of Autism and Developmental Disorders, 50(7)*, S. 1-13.

McBride, L., & Moran, G. (1967). Double agreement as a function of item ambiguity and susceptibility to demand implications of the psychological situation. *Journal of Personality and Social Psychology, 6(1)*, S. 115-118.

Messick, S. (1989). Validity. In R. Linn, *Educational measurement (3. Aufl.)* (S. 13–103). New York: Macmillan.

Mohr, G., & Müller, A. (2014). Psychosomatische Beschwerden im nichtklinischen Kontext. In *Zusammenstellung sozialwissenschaftlicher Items und Skalen.*

Moosbrugger, H. (2012). Klassische Testtheorie (KTT). In H. Moosbrugger, & A. Kelava, *Testtheorie und Fragebogenkonstruktion (2. Aufl.)* (S. 103-117). Heidelberg: Springer.

Moosbrugger, H., & Brandt, H. (2020). Itemkonstruktion und Antwortverhalten. In H. Moosbrugger, & A. Kelava, *Testtheorie und Fragebogenkonstruktion* (S. 67-89). Berlin & Heidelberg: Springer.

Moosbrugger, H., Schermelleh-Engel, K., & Klein, A. (1997). Methodological problems of estimating latent interaction effects. *Methods of Psychological Research Online, 2*, S. 95-111.

Moosbrugger, H., Schermelleh-Engel, K., Kelava, A., & Klein, A. (2009). Testing Multiple Nonlinear Effects in Structural Equation Modeling: A Comparison of Alternative Estimation Approaches. In T. Teo, & M. Khine, *Structural Equation Modelling in Educational Research: Concepts and Applications* (S. 103-136). Rotterdam: Sense Publishers.

Muthén, B. (2012). *Latent variable interactions.* Von http://www.statmodel.com/download/LV%20Inter action.pdf abgerufen

Muthén, B. O., & Asparouhov, T. (2013). *BSEM Measurement Invariance Analysis. Mplus Web Notes, 17.* Von https://www.statmodel.com/examples/webnotes/webnote17.pdf abgerufen

Muthén, L., & Muthén, B. (2007). *Mplus user's guide.* Los Angeles, CA: Muthén & Muthén.

Paulhus, D. (2002). Socially desirable responding: The evolution of a construct. In H. Braun, D. Jackson, & D. Wiley, *The role of constructs in psychological and educational measurement* (S. 49-69). Mahwah, NY: Erlbaum.

Petty, R., & Cacioppo, J. (1986). The Elaboration Likelihood Model of Persuasion. *Advances in Experimental Social Psychology, 19(1)*, S. 124-205.

Petty, R., Cacioppo, J., & Kasmer, J. (2015). The role of affect in the elaboration likelihood model of persuasion. In I. Donohew, H. Sypher, & E. Higgins, *Communication, Social Cognition, and Affect.* New York: Psychology Press.

Pfleger, P. (2019). Exploring Fit for Nonlinear Structural Equation Models. In *Theses and Dissertations, 7370.* Brigham Young University.

Phillips, P., Phillips, J., & Aaron, B. (2013). *Survey Basics.* Alecandria, VA: ASTD Press.

Ping, R. (1996). Latent Variable Interaction and Quadratic Effect Estimation: A Two-Step Technique Using Structural Equation Analysis. *The Psychological Bulletin, 119(1)*, S. 166-175.

Pokropek, A., Davidov, E., & Schmidt, P. (kein Datum). A Monte Carlo Simulation Study to Assess The Appropriateness of Traditional and Newer Approaches to Test for Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 26(5)*, S. 724-74.

Porst, R. (1998). Erfahrung mit und Bewertung von Umfragen: was unsere Befragten über Umfragen denken. *ZUMA-Arbeitsbericht, 3*.

R Core Team. (2021). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* Von https://www.R-project.org/ abgerufen

Rammstedt, B., Goldberg, L., & Borg, I. (2010). The measurement equivalence of Big-Five factor markers for persons with different levels of education. *Journal of Research in Personality, 44(4)*, S. 53-61.

Reimer, T. (2003). Direkte und indirekte Effekte der Argumentqualität – Der Einfluss der Argumentstärke auf die wahrgenommene Expertise eines Kommunikators. *Zeitschrift für Sozialpsychologie, 34(4)*, S. 243-255.

Reinecke, J. (1991). *Interviewer- und Befragtenverhalten: theoretische Ansätze und methodische Konzepte.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Reinecke, J. (2005). *Strukturgleichungsmodelle in den Sozialwissenschaften.* München: Oldenbourg Verlag.

Reinecke, J., & Pöge, A. (2010). Strukturgleichungsmodelle. In C. Wolf, & H. Best, *Handbuch der sozialwissenschaftlichen Datenanalyse* (S. 775-804). Wiesbaden: VS Verlag für Sozialwissenschaften.

Revelle, W. (2022). *Package 'psych'. Procedures for Psychological, Psychometric, and Personality.* Von https://cran.r-project.org/web/packages/psych/psych.pdf abgerufen

Rogelberg, S., Fisher, G., Maynard, D., Hakel, M., & Horvath, M. (2001). Attitudes Toward Surveys: Development of a Measure and its Relationship to Respondent Behavior. *Organizational Research Methods, 4(1)*, S. 3-25.

Roper, B. (1986). Evaluating Polls with Poll Data. *Public Opinion Quarterly, 50*, S. 10-16.

Rosseel, Y., Jorgensen, T. D., Rockwood, N., Oberski, D., Byrnes, J., L., V., Du, H. (2022). *Package 'lavaan'. Latent Variable Analysis.* Von https://cran.r-project.org/web/packages/lavaan/lavaan.pdf abgerufen

Samples, T., Woods, A., Davis, T., Rhodes, M., Shahane, A., & Kaslow, N. (2012). Race of Interviewer Effect on Disclosures of Suicidal Low-Income African American Women. *Journal of Black Psychology, 40(1)*.

Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In R. Heijmans, D. Pollock, & A. Satorra, *Innovations in multivariate statistical analysis. Festschrift for Heinz Neudecker* (S. 233-247). London: Kluwer Academic Publishers.

Satorra, A., & Bentler, P. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika, 75(2)*, S. 243–248.

Schaeffer, N., Dykema, J., & Maynard, D. (2010). Interviewers and Interviewing. In P. Marsden, & J. Wright, *Handbook of Survey Research* (S. 437-470). Bingley: Emerald Group.

Schermelleh-Engel, K., & Schweizer, K. (2012). Multitraid-Multimethod-Analysen. In H. Moosbrugger, & A. Kelava, *Testtheorie und Fragebogenkonstruktion* (S. 345-362). Berlin und Heidelberg: Springer-Verlag.

Schermelleh-Engel, K., Kerwer, M., & Klein, A. (2014). Evaluation of model fit in nonlinear multilevel structural equation modeling. *Frontiers in Psychology, 181(5)*, S. 1-11.

Schnell, R., Hill, P., & Esser, E. (2018). *Methoden der empirischen Sozialforschung (11. Aufl.).* München & Wien: Oldenburg.

Schreier, J., Biethahn, N., & Drewes, F. (2018). Question order effects in partial least squares path modelling: an empirical investigation. *Quality and Quantity, 52(1)*, S. 71-84.

Schuette, R., & Fazio, R. (1995). Attitude Accessibility and Motivation as Determinants of Biased Processing: A Test of the MODE Model. *Personality and Social Psychology Bulletin, 21(7)*, S. 704-710.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys.* Orlando: Academic Press.

Schuman, H., Presser, S., & Ludwig, J. (1981). Context effects on survey responses to questions about abortions. *Public Opinion Quarterly, 45(2)*, S. 216-223.

Schupp, J., & Gerlitz, J. (2014). Big Five Inventory-SOEP (BFI-S). In *Zusammenstellung sozialwissenschaftlicher Items und Skalen.*

Schwarz, N. (1991). In welcher Reihenfolge fragen? Kontexteffekte in standardisierten Befragungen. *ZUMA-Arbeitsbericht, 16*.

Schwarz, N., & Bless, H. (1992). Assimilation and contrast effcts in attitude measurement: An inclusion/exclusion model. *Advances of Consumer Research, 19*, S. 72-77.

Schwarz, N., Strack, F., & Mai, H. (1991). Assimilation and contrast effects in part-whole question sequences: A conversational logic analysis. *Public Opinion Quarterly, 55(1)*, S. 3-23.

Schwarz, N., Strack, F., Hippler, H., & Bishop, G. (1991). The impact of administration mode on response effects in survey measurement. *Applied Cognitive Psychology, 5(3)*, S. 193-212.

Sharp, L., & Frankel, J. (1983). Respondent Burden: A Test of Some Common Assumptions. *Public Opinion Quarterly, 47(1)*, S. 36-53.

Shrum, L., & O'Guinn, T. (1993). Processes and Effects in the Construction of Social Reality. Construct Accessibility as an Explanatory Variable. *Communication Research, 20(3)*, S. 436-471.

Silber, H., Danner, D., & Rammstedt, B. (2018). The impact of respondent attentiveness on reliability and validity. *International Journal of Social Research Methodology, 22(2)*, S. 153-164.

Sivo, S. A. (2001). Multiple Indicator Stationary Time Series Models. *Structural Equation Modeling: A Multidisciplinary Journal, 8(4)*, S. 599-612.

Skarbek-Kozietulska, A., Preisendörfer, P., & Wolter, F. (2012). Leugnen oder gestehen? Bestimmungsfaktoren wahrer Antworten in Befragungen. *Zeitschrift für Soziologie, 41(1)*, S. 5-23.

Smith, E., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4(2)*, S. 108-131.

Stark, T., Silber, H., Krosnick, J., Blom, A., Aoyagi, M., Belchior, A., . . . Yu, R. (2018). Generalization of Classic Question Order Effects Across Cultures. *Sociological Methods & Research, 49(3)*, S. 1-36.

Steinmetz, H., Davidov, E., & Schmidt, P. (2011). Three Approaches to Estimate Latent Interaction Effects: Intention and Perceived Behavioral Control in the Theory of Planned Behavior. *Methodological Innovations Online, 6(1)*, S. 95-110.

Stinchcombe, A., Jones, C., & Sheatsley, P. (1981). Nonresponse Bias for Attitude Questions. *Public Opinion Quarterly, 45(1)*, S. 359-375.

Stocké, V. (2002). Die Vorhersage von Fragenreihenfolgeeffekten durch Antwortlatenzen: eine Validierungsstudie. *ZUMA Nachrichten, 50(26)*, S. 26–53.

Stocké, V. (2004). Entstehungsbedingungen von Antwortverzerrungen durch soziale Erwünschtheit. Ein Vergleich der Prognosen der Rational-Choice Theorie und des Modells der Frame-Selektion. *Zeitschrift für Soziologie, 33(4)*, S. 303–320.

Stocké, V. (2006). Attitudes towards Surveys, Attitude Accessibility and the Effect on Respondents' Susceptibility to Nonresponse. *Quality and Quantity, 40*, S. 259-288.

Stocké, V., & Becker, B. (2004). Determinanten und Konsequenzen der Umfrageeinstellung. Bewertungsdimensionen unterschiedlicher Umfragesponsoren und die Antwortbereitschaft der Befragten. *Sonderforschungsbereich 504: Rationalitätskonzepte, Entscheidungsverhalten und ökonomische Modellierung*, S. 04-17.

Stocké, V., & Langfeldt, B. (2003). Umfrageeinstellung und Umfrageerfahrung. Die relative Bedeutung unterschiedlicher Aspekte der Interviewerfahrung für die generalisierte Umfrageeinstellung. *ZUMA-Nachrichten, 27(52)*, S. 55–88.

Strack, F. (2013). *Zur Psychologie der standardisierten Befragung: Kognitive und kommunikative Prozesse. Band 48 von Lehr- und Forschungstexte Psychologie.* Berlin: Springer.

Strack, F., & Deutsch, R. (2004). Reflective and Impulsive Determinants of Social Behavior. *Personality and Social Psychology Review, 8(3)*, S. 220-247.

Strack, F., & Deutsch, R. (2012). A theory of impulse and reflection. In P. Van Lange, A. Kruglanski, & E. Higgins, *Handbook of theories of social psychology* (S. 97-117). Sage Publications Ltd.

Strack, F., & Martin, L. (1986). Verfügbarkeit und Verwendung selbstbezogener Informationen in der Befragungssituation. *Paper präsentiert auf dem 35. Kongress der Deutschen Gesellschaft für Psychologie.* Heidelberg.

Strack, F., & Martin, L. (1987). Thinking, Judging, and Communicating: A Process Account of Context Effects in Attitude Surveys. In H. Hippler, N. Schwarz, & S. Sudman, *Social Information Processing and Survey Methodology.* New York: Springer.

Strack, F., Martin, L., & Schwarz, N. (1987). *The Context Paradox in Attitude Surveys: Assimilation or Contrast? ZUMA-Arbeitsbericht Nr. 87/07* . Mannheim: Zentrum für Umfragen, Methoden und Analysen e.V. (ZUMA).

Suchman, L., & Jordan, B. (1990). Interactional troubles in face-to-face survey interviews. *Journal of the American Statistical Association, 85(409)*, S. 232-241.

Sudman, S., Bradburn, N., & Schwarz, N. (1996). *Thinking about answers. The application of cognitive processes to survey methodology.* San Francisco: Jossey-Bass.

Thurstone, L. (1947). *Multiple factor analysis. A development and expansion of vectors of the mind.* Chicago: University of Chicago Press.

Tomás, J., Oliver, A., Galiana, L., Sancho, P., & Lila, M. (2013). Explaining method effects associated with negatively worded items in trait and state global and domain-specific self-esteem scales. *Structural Equation Modeling: A Multidisciplinary Journal, 20(2)*, S. 299–313.

Tormala, Z., Petty, R., & Brinol, P. (2002). Ease of retrieval effects in persuasion: A self-validation analysis. *Personality and Social Psychology Bulletin, 28(12)*, S. 1700-1712.

Tourangeau, R., & Rasinski, K. (1988). Cognitive Processes Underlying Context Effects in Attitude Measurement. *Psychological Bulletin, 103(3)*, S. 299-314.

Tourangeau, R., Rips, L., & Rasinski, K. (2000). *The Psychology of Survey Response.* Cambridge: Cambridge University Press.

Trott, D., & Jackson, D. (1967). An experimental analysis of acquiescence. *Journal of Experimental Research in Personality, 2*, S. 278-288.

Urban, D., & Mayerl, J. (2014). *Strukturgleichungsmodellierung. Ein Ratgeber für die Praxis.* Wiesbaden: Springer VS.

Urban, D., & Mayerl, J. (2018). *Angewandte Regressionsanalyse: Theorie, Technik und Praxis.* Wiesbaden: Springer VS.

Vandenberg, R. (2006). Software Review: MPlus 3.0. *Organizational Research Methods, 9(3)*, S. 408-412.

Verdam, M., Oort, F., & Sprangers, M. (2017). Item bias detection in the Hospital Anxiety and Depression Scale using structural equation modeling: comparison with other item bias detection methods. *Quality of Life Research, 26(6)*, S. 1439-1450.

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing, 27(3)*, S. 236-247.

Weisberg, H. (2005). *The Total Survey Error Approach.* Chicago: University of Chicago Press.

Wijnen, K., Vermeir, I., & Van Kenhovea, P. (2007). The relationships between traits, personal values, topic involvement, and topic sensitivity in a mail survey context. *Personality and Individual Differences, 42(1)*, S. 61-73.

Wilding, S., Conner, M., Sandberg, T., Prestwich, A., Lawton, R., Wood, C., Sheeran, P. (2016). The question-behaviour effect: A theoretical and methodological review and meta-analysis. *European Review of Social Psychology, 27(1)*, S. 196-230.

Williams, T., Scalco, M., & Simms, L. (2018). The construct validity of general and specific dimensions of personality pathology. *Psychological Medicine, 48(5)*, S. 834-848.

Wingerter, C. (2014). Allgemeines Umweltbewusstsein. In *Zusammenstellung sozialwissenschaftlicher Items und Skalen.*

Wolf, E., Harrington, K., Clark, S., & Miller, M. (2013). Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. *Educational and Psychological Measurement, 73(6)*, S. 913-934.

Wyer, R., & Srull, T. (1986). Human cognition in its social context. *Psychological Review, 93(3)*, S. 322-359.

Yan, T., & Tourangeau, R. (2008). Fast times and easy questions: the effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology, 22(1)*, S. 51-68.

Yang, Y., & Green, S. (2010). A Note on Structural Equation Modeling Estimates of Reliability. *Structural Equation Modeling: A Multidisciplinary Journal, 17(1)*, S. 66-81.

Yerkes, R., & Dodson, J. (1908). The relation of strength of stimulus to rapidity of habitformation. *Journal of Comparative Neurology and Psychology, 18*, S. 459-482.

Young, A., & Fazio, R. (2013). Attitude accessibility as a determinant of object construal and evaluation. *Journal of Experimental Social Psychology, 49(3)*, S. 404-418.

7 Anhang

7.1 R Syntax der Monte-Carlo Analyse mit 10000 Iterationen

```
###-----------------------------------------------------------------###
###--- Schritt 1 -----------------------------------------------------###
###--- Lade benötigte Packages, setze Seed ---------------------------###
###-----------------------------------------------------------------###
library(lavaan)
library(moments)
library(psych)
library(lm.beta)

set.seed(1234)

###-----------------------------------------------------------------###
###--- Schritt 2 -----------------------------------------------------###
###--- Definiere Ergebnistabellen und Parameter ----------------------###
###-----------------------------------------------------------------###
j <- 10000 #Anzahl an Iterationen


CFAEM <- matrix(ncol = 57, nrow = j) #Monte Carlo CFA Ergebnisse Matrix
colnames(CFAEM) <-
c("g1f1","g1f2","g1f3","g1f4","g1f5","g1se1","g1se2","g1se3","g1se4","g1se5","g2f1","g2f2","g2
f3","g2f4","g2f5","g2se1","g2se2","g2se3","g2se4","g2se5","g3f1","g3f2","g3f3","g3f4","g3f5","
g3se1","g3se2","g3se3","g3se4","g3se5","Chisq","df","p","cfi","rmsea","srmr","k","g4f1","g4f2"
,"g4f3","g4f4","g4f5","g4se1","g4se2","g4se3","g4se4","g4se5","g5f1","g5f2","g5f3","g5f4","g5f
5","g5se1","g5se2","g5se3","g5se4","g5se5")


MONTE_CARLO <- matrix(ncol = 78, nrow = j) #Relevante Ergebnisse der Monte-Carlo Loops nach
Iteration
colnames(MONTE_CARLO) <- c("Wahre f","f1","f2","f3","f4","f5","SE f1","SE f2","SE f3","SE
f4","SE f5",
                           "Wahres b1 g1","Wahres b4 g1","Wahres b5 g1","Wahres b7 g1","b1
g1","b4 g1","b5 g1","b7 g1","SE b1 g1","SE b4 g1","SE b5 g1","SE b7 g1",
                           "Wahres b1 g2","Wahres b4 g2","Wahres b5 g2","Wahres b7 g2","b1
g2","b4 g2","b5 g2","b7 g2","SE b1 g2","SE b4 g2","SE b5 g2","SE b7 g2",
                           "Wahres b1 g3","Wahres b4 g3","Wahres b5 g3","Wahres b7 g3","b1
g3","b4 g3","b5 g3","b7 g3","SE b1 g3","SE b4 g3","SE b5 g3","SE b7 g3",
                           "n","k","cv","sm",
                           "Wahres b2 g1","Wahres b3 g1","Wahres b6 g1","b2 g1","b3 g1","b6
g1","SE b2 g1","SE b3 g1","SE b6 g1",
                           "Wahres b2 g2","Wahres b3 g2","Wahres b6 g2","b2 g2","b3 g2","b6
g2","SE b2 g2","SE b3 g2","SE b6 g2",
                           "Wahres b2 g3","Wahres b3 g3","Wahres b6 g3","b2 g3","b3 g3","b6
g3","SE b2 g3","SE b3 g3","SE b6 g3")


MCMTMM <- matrix(ncol = 20, nrow = j) #Monte Carlo MTMM Ergebnisse Matrix
colnames(MCMTMM) <-
c("f1","f2","f3","f4","f5","se1","se2","se3","se4","se5","n","k","cv","sm","cov1","cov2","cov3
","sec1","sec2","sec3")


FAKTOR <- matrix(ncol = 28, nrow = 7) #Regressionsergebnisse Faktorladungen
colnames(FAKTOR) <- c("b CFA1","SE CFA1","b* CFA1","p CFA1",
```

```
                      "b CFA2","SE CFA2","b* CFA2","p CFA2",
                      "b CFA3","SE CFA3","b* CFA3","p CFA3",
                      "b CFA4","SE CFA4","b* CFA4","p CFA4",
                      "b CFA5","SE CFA5","b* CFA5","p CFA5",
                      "b SEM","SE SEM","b* SEM","p SEM",
                      "b MTMM","SE MTMM","b* MTMM","p MTMM")
rownames(FAKTOR) <- c("n","k3","k4","cv","sm","cv*sm","a")

EFFEKTS <- matrix(ncol = 40, nrow = 7) #Regressionsergebnisse Fragereihenfolgeneffekte
colnames(EFFEKTS) <- c("b (b1)","SE (b1)","b* (b1)","p (b1)",
                       "b (b4)","SE (b4)","b* (b4)","p (b4)",
                       "b (b5)","SE (b5)","b* (b5)","p (b5)",
                       "b (b7)","SE (b7)","b* (b7)","p (b7)",
                       "b (b2)","SE (b2)","b* (b2)","p (b2)",
                       "b (b3)","SE (b3)","b* (b3)","p (b3)",
                       "b (b6)","SE (b6)","b* (b6)","p (b6)",
                       "b (cov1)","SE (cov1)","b* (cov1)","p (cov1)",
                       "b (cov2)","SE (cov2)","b* (cov2)","p (cov2)",
                       "b (cov3)","SE (cov3)","b* (cov3)","p (cov3)")
rownames(EFFEKTS) <- c("n","k3","k4","cv","sm","cv*sm","a")

SE <- matrix(ncol = 40, nrow = 7) #Regressionsergebnisse SE
colnames(SE) <- c("b (b1)","SE (b1)","b* (b1)","p (b1)",
                  "b (b4)","SE (b4)","b* (b4)","p (b4)",
                  "b (b5)","SE (b5)","b* (b5)","p (b5)",
                  "b (b7)","SE (b7)","b* (b7)","p (b7)",
                  "b (b2)","SE (b2)","b* (b2)","p (b2)",
                  "b (b3)","SE (b3)","b* (b3)","p (b3)",
                  "b (b6)","SE (b6)","b* (b6)","p (b6)",
                  "b (cov1)","SE (cov1)","b* (cov1)","p (cov1)",
                  "b (cov2)","SE (cov2)","b* (cov2)","p (cov2)",
                  "b (cov3)","SE (cov3)","b* (cov3)","p (cov3)")
rownames(SE) <- c("n","k3","k4","cv","sm","cv*sm","a")

###-----------------------------------------------------------------###
###--- Schritt 3 -----------------------------------------------------###
###--- Öffne Monte-Carlo Loop und Error Handler ----------------------###
###-----------------------------------------------------------------###
for(i1 in 1:j){
  tryCatch({

###-----------------------------------------------------------------###
###--- Schritt 4 -----------------------------------------------------###
###--- Definiere Zufallsparameter  -----------------------------------###
###-----------------------------------------------------------------###
n <- round(runif(1,200,5000))
k <- ceiling(runif(1,2,5))
cv <- round(runif(1,0.3,0.9),digits=3)
sm <- round(runif(1,0.1,0.9),digits=3)

###-----------------------------------------------------------------###
```

```
###--- Schritt 5 ----------------------------------------------------###
###--- Berechne Indikatorvariablen je Gruppe und erstelle Datensatz --###
###-------------------------------------------------------------------###
eta1 <- rnorm(n,0,1)
eta2 <- rnorm(n,0,1)
eta3 <- rnorm(n,0,1)

xi1 <- rnorm(n,0,1)
xi2 <- rnorm(n,0,1)

for(i2 in 1:21){
  nam_e <- paste("e", i2, sep = "")
  assign(nam_e, rnorm(n,0,1))
}

y1 <- cv*eta1 + (1-cv)*e1
y2 <- cv*eta1 + (1-cv)*(sm*(-0.5*y1*xi1 - 0.5*y1*xi2) + (1-sm)*e2)
y3 <- cv*eta1 + (1-cv)*(sm*(-0.5*y2*xi1 - 0.5*y2*xi2) + (1-sm)*e3)
y4 <- cv*eta1 + (1-cv)*(sm*(-0.5*y3*xi1 - 0.5*y3*xi2) + (1-sm)*e4)
y5 <- cv*eta1 + (1-cv)*(sm*(-0.5*y4*xi1 - 0.5*y4*xi2) + (1-sm)*e5)

y6 <- cv*eta2 + (1-cv)*e6
y7 <- cv*eta2 + (1-cv)*(sm*(y6*xi1*xi2) + (1-sm)*e7)
y8 <- cv*eta2 + (1-cv)*(sm*(y7*xi1*xi2) + (1-sm)*e8)
y9 <- cv*eta2 + (1-cv)*(sm*(y8*xi1*xi2) + (1-sm)*e9)
y10 <- cv*eta2 + (1-cv)*(sm*(y9*xi1*xi2) + (1-sm)*e10)

y11 <- cv*eta3 + (1-cv)*e11
y12 <- cv*eta3 + (1-cv)*e12
y13 <- cv*eta3 + (1-cv)*e13
y14 <- cv*eta3 + (1-cv)*e14
y15 <- cv*eta3 + (1-cv)*e15

x1 <- 0.8*xi1 + 0.2*e16
x2 <- 0.8*xi1 + 0.2*e17
x3 <- 0.8*xi1 + 0.2*e18

x4 <- 0.8*xi2 + 0.2*e19
x5 <- 0.8*xi2 + 0.2*e20
x6 <- 0.8*xi2 + 0.2*e21

gr1 <- 1
gr2 <- 2
gr3 <- 3

XY1 <- data.frame(x1,x2,x3,x4,x5,x6,y1,y2,y3,y4,y5,gr1)
colnames(XY1) <- c("x1","x2","x3","x4","x5","x6","y1","y2","y3","y4","y5","Gruppe")
XY2 <- data.frame(x1,x2,x3,x4,x5,x6,y6,y7,y8,y9,y10,gr2)
colnames(XY2) <- c("x1","x2","x3","x4","x5","x6","y1","y2","y3","y4","y5","Gruppe")
XY3 <- data.frame(x1,x2,x3,x4,x5,x6,y11,y12,y13,y14,y15,gr3)
colnames(XY3) <- c("x1","x2","x3","x4","x5","x6","y1","y2","y3","y4","y5","Gruppe")
```

```
XY <- rbind(XY1,XY2,XY3)

###------------------------------------------------------------------###
###--- Schritt 6 ----------------------------------------------------###
###--- Berechne Interaktions-Indikatoren (unconstr. approach) --------###
###------------------------------------------------------------------###
#xi1*xi2
XY$x1x4 <- XY$x1*XY$x4
XY$x1x5 <- XY$x1*XY$x5
XY$x1x6 <- XY$x1*XY$x6
XY$x2x4 <- XY$x2*XY$x4
XY$x2x5 <- XY$x2*XY$x5
XY$x2x6 <- XY$x2*XY$x6
XY$x3x4 <- XY$x3*XY$x4
XY$x3x5 <- XY$x3*XY$x5
XY$x3x6 <- XY$x3*XY$x6

#y*xi1
XY$y1x1 <- XY$y1*XY$x1
XY$y1x2 <- XY$y1*XY$x2
XY$y1x3 <- XY$y1*XY$x3
XY$y2x1 <- XY$y2*XY$x1
XY$y2x2 <- XY$y2*XY$x2
XY$y2x3 <- XY$y2*XY$x3
XY$y3x1 <- XY$y3*XY$x1
XY$y3x2 <- XY$y3*XY$x2
XY$y3x3 <- XY$y3*XY$x3
XY$y4x1 <- XY$y4*XY$x1
XY$y4x2 <- XY$y4*XY$x2
XY$y4x3 <- XY$y4*XY$x3

#y*xi2
XY$y1x4 <- XY$y1*XY$x4
XY$y1x5 <- XY$y1*XY$x5
XY$y1x6 <- XY$y1*XY$x6
XY$y2x4 <- XY$y2*XY$x4
XY$y2x5 <- XY$y2*XY$x5
XY$y2x6 <- XY$y2*XY$x6
XY$y3x4 <- XY$y3*XY$x4
XY$y3x5 <- XY$y3*XY$x5
XY$y3x6 <- XY$y3*XY$x6
XY$y4x4 <- XY$y4*XY$x4
XY$y4x5 <- XY$y4*XY$x5
XY$y4x6 <- XY$y4*XY$x6

#y*xi1*xi2
XY$y1x1x4 <- XY$y1*XY$x1*XY$x4
XY$y1x1x5 <- XY$y1*XY$x1*XY$x5
XY$y1x1x6 <- XY$y1*XY$x1*XY$x6
XY$y1x2x4 <- XY$y1*XY$x2*XY$x4
XY$y1x2x5 <- XY$y1*XY$x2*XY$x5
```

```
XY$y1x2x6 <- XY$y1*XY$x2*XY$x6
XY$y1x3x4 <- XY$y1*XY$x3*XY$x4
XY$y1x3x5 <- XY$y1*XY$x3*XY$x5
XY$y1x3x6 <- XY$y1*XY$x3*XY$x6
XY$y2x1x4 <- XY$y2*XY$x1*XY$x4
XY$y2x1x5 <- XY$y2*XY$x1*XY$x5
XY$y2x1x6 <- XY$y2*XY$x1*XY$x6
XY$y2x2x4 <- XY$y2*XY$x2*XY$x4
XY$y2x2x5 <- XY$y2*XY$x2*XY$x5
XY$y2x2x6 <- XY$y2*XY$x2*XY$x6
XY$y2x3x4 <- XY$y2*XY$x3*XY$x4
XY$y2x3x5 <- XY$y2*XY$x3*XY$x5
XY$y2x3x6 <- XY$y2*XY$x3*XY$x6
XY$y3x1x4 <- XY$y3*XY$x1*XY$x4
XY$y3x1x5 <- XY$y3*XY$x1*XY$x5
XY$y3x1x6 <- XY$y3*XY$x1*XY$x6
XY$y3x2x4 <- XY$y3*XY$x2*XY$x4
XY$y3x2x5 <- XY$y3*XY$x2*XY$x5
XY$y3x2x6 <- XY$y3*XY$x2*XY$x6
XY$y3x3x4 <- XY$y3*XY$x3*XY$x4
XY$y3x3x5 <- XY$y3*XY$x3*XY$x5
XY$y3x3x6 <- XY$y3*XY$x3*XY$x6
XY$y4x1x4 <- XY$y4*XY$x1*XY$x4
XY$y4x1x5 <- XY$y4*XY$x1*XY$x5
XY$y4x1x6 <- XY$y4*XY$x1*XY$x6
XY$y4x2x4 <- XY$y4*XY$x2*XY$x4
XY$y4x2x5 <- XY$y4*XY$x2*XY$x5
XY$y4x2x6 <- XY$y4*XY$x2*XY$x6
XY$y4x3x4 <- XY$y4*XY$x3*XY$x4
XY$y4x3x5 <- XY$y4*XY$x3*XY$x5
XY$y4x3x6 <- XY$y4*XY$x3*XY$x6

###-------------------------------------------------------------------###
###--- Schritt 7 -----------------------------------------------------###
###--- CFA Basismodell -----------------------------------------------###
###-------------------------------------------------------------------###
if(k == 3){
  CFA <- 'eta_est =~ NA*y1 + y2 + y3
          eta_est ~~ 1*eta_est'
  fit_cfa <- cfa(CFA,XY,group="Gruppe")
  results_cfa_ust <- parameterEstimates(fit_cfa)

  MTMM <- 'eta_est =~ NA*y1 + y2 +y3
           eta_est ~~ 1*eta_est
           y1 ~~ c(cov1,cov2,cov3)*y2
           y2 ~~ c(cov1,cov2,cov3)*y3'
  fit_MTMM <- cfa(MTMM,XY,group="Gruppe")
  results_MTMM_ust <- parameterEstimates(fit_MTMM)

  CFAEM[i1,1] <- results_cfa_ust[1,6]
  CFAEM[i1,2] <- results_cfa_ust[2,6]
```

```
CFAEM[i1,3] <- results_cfa_ust[3,6]
CFAEM[i1,6] <- results_cfa_ust[1,7]
CFAEM[i1,7] <- results_cfa_ust[2,7]
CFAEM[i1,8] <- results_cfa_ust[3,7]
CFAEM[i1,11] <- results_cfa_ust[12,6]
CFAEM[i1,12] <- results_cfa_ust[13,6]
CFAEM[i1,13] <- results_cfa_ust[14,6]
CFAEM[i1,16] <- results_cfa_ust[12,7]
CFAEM[i1,17] <- results_cfa_ust[13,7]
CFAEM[i1,18] <- results_cfa_ust[14,7]
CFAEM[i1,21] <- results_cfa_ust[23,6]
CFAEM[i1,22] <- results_cfa_ust[24,6]
CFAEM[i1,23] <- results_cfa_ust[25,6]
CFAEM[i1,26] <- results_cfa_ust[23,7]
CFAEM[i1,27] <- results_cfa_ust[24,7]
CFAEM[i1,28] <- results_cfa_ust[25,7]
CFAEM[i1,31] <- fitMeasures(fit_cfa)[3]
CFAEM[i1,32] <- fitMeasures(fit_cfa)[4]
CFAEM[i1,33] <- fitMeasures(fit_cfa)[5]
CFAEM[i1,34] <- fitMeasures(fit_cfa)[9]
CFAEM[i1,35] <- fitMeasures(fit_cfa)[23]
CFAEM[i1,36] <- fitMeasures(fit_cfa)[29]
CFAEM[i1,37] <- 3

CFAEM[i1,38] <- results_MTMM_ust[1,7]
CFAEM[i1,39] <- results_MTMM_ust[2,7]
CFAEM[i1,40] <- results_MTMM_ust[3,7]
CFAEM[i1,43] <- results_MTMM_ust[1,8]
CFAEM[i1,44] <- results_MTMM_ust[2,8]
CFAEM[i1,45] <- results_MTMM_ust[3,8]
CFAEM[i1,48] <- results_MTMM_ust[14,7]
CFAEM[i1,49] <- results_MTMM_ust[15,7]
CFAEM[i1,50] <- results_MTMM_ust[16,7]
CFAEM[i1,53] <- results_MTMM_ust[14,8]
CFAEM[i1,54] <- results_MTMM_ust[15,8]
CFAEM[i1,55] <- results_MTMM_ust[16,8]

MTMM <- 'eta_est =~ NA*y1 + y2 +y3
         eta_est ~~ 1*eta_est
         y1 ~~ c(cov1,cov2,cov3)*y2
         y2 ~~ c(cov1,cov2,cov3)*y3'
fit_MTMM <- cfa(MTMM,XY,group="Gruppe",group.equal = c("loadings"))
results_MTMM_ust <- parameterEstimates(fit_MTMM)

MCMTMM[i1,1] <- results_MTMM_ust[1,7]
MCMTMM[i1,2] <- results_MTMM_ust[2,7]
MCMTMM[i1,3] <- results_MTMM_ust[3,7]
MCMTMM[i1,6] <- results_MTMM_ust[1,8]
MCMTMM[i1,7] <- results_MTMM_ust[2,8]
MCMTMM[i1,8] <- results_MTMM_ust[3,8]
MCMTMM[i1,11] <- n
```

```
  MCMTMM[i1,12] <- k
  MCMTMM[i1,13] <- cv
  MCMTMM[i1,14] <- sm
  MCMTMM[i1,15] <- results_MTMM_ust[5,7]
  MCMTMM[i1,16] <- results_MTMM_ust[18,7]
  MCMTMM[i1,17] <- results_MTMM_ust[31,7]
  MCMTMM[i1,18] <- results_MTMM_ust[5,8]
  MCMTMM[i1,19] <- results_MTMM_ust[18,8]
  MCMTMM[i1,20] <- results_MTMM_ust[31,8]
}

if(k == 4){
  CFA <- 'eta_est =~ NA*y1 + y2 + y3 + y4
          eta_est ~~ 1*eta_est'
  fit_cfa <- cfa(CFA,XY,group="Gruppe")
  results_cfa_ust <- parameterEstimates(fit_cfa)

  MTMM <- 'eta_est =~ NA*y1 + y2 +y3 + y4
           eta_est ~~ 1*eta_est
           y1 ~~ c(cov1,cov2,cov3)*y2
           y2 ~~ c(cov1,cov2,cov3)*y3
           y3 ~~ c(cov1,cov2,cov3)*y4'
  fit_MTMM <- cfa(MTMM,XY,group="Gruppe")
  results_MTMM_ust <- parameterEstimates(fit_MTMM)

  CFAEM[i1,1] <- results_cfa_ust[1,6]
  CFAEM[i1,2] <- results_cfa_ust[2,6]
  CFAEM[i1,3] <- results_cfa_ust[3,6]
  CFAEM[i1,4] <- results_cfa_ust[4,6]
  CFAEM[i1,6] <- results_cfa_ust[1,7]
  CFAEM[i1,7] <- results_cfa_ust[2,7]
  CFAEM[i1,8] <- results_cfa_ust[3,7]
  CFAEM[i1,9] <- results_cfa_ust[4,7]
  CFAEM[i1,11] <- results_cfa_ust[15,6]
  CFAEM[i1,12] <- results_cfa_ust[16,6]
  CFAEM[i1,13] <- results_cfa_ust[17,6]
  CFAEM[i1,14] <- results_cfa_ust[18,6]
  CFAEM[i1,16] <- results_cfa_ust[15,7]
  CFAEM[i1,17] <- results_cfa_ust[16,7]
  CFAEM[i1,18] <- results_cfa_ust[17,7]
  CFAEM[i1,19] <- results_cfa_ust[18,7]
  CFAEM[i1,21] <- results_cfa_ust[29,6]
  CFAEM[i1,22] <- results_cfa_ust[30,6]
  CFAEM[i1,23] <- results_cfa_ust[31,6]
  CFAEM[i1,24] <- results_cfa_ust[32,6]
  CFAEM[i1,26] <- results_cfa_ust[29,7]
  CFAEM[i1,27] <- results_cfa_ust[30,7]
  CFAEM[i1,28] <- results_cfa_ust[31,7]
  CFAEM[i1,29] <- results_cfa_ust[32,7]
  CFAEM[i1,31] <- fitMeasures(fit_cfa)[3]
  CFAEM[i1,32] <- fitMeasures(fit_cfa)[4]
```

```
  CFAEM[i1,33] <- fitMeasures(fit_cfa)[5]
  CFAEM[i1,34] <- fitMeasures(fit_cfa)[9]
  CFAEM[i1,35] <- fitMeasures(fit_cfa)[23]
  CFAEM[i1,36] <- fitMeasures(fit_cfa)[29]
  CFAEM[i1,37] <- 4

  CFAEM[i1,38] <- results_MTMM_ust[1,7]
  CFAEM[i1,39] <- results_MTMM_ust[2,7]
  CFAEM[i1,40] <- results_MTMM_ust[3,7]
  CFAEM[i1,41] <- results_MTMM_ust[4,7]
  CFAEM[i1,43] <- results_MTMM_ust[1,8]
  CFAEM[i1,44] <- results_MTMM_ust[2,8]
  CFAEM[i1,45] <- results_MTMM_ust[3,8]
  CFAEM[i1,46] <- results_MTMM_ust[4,8]
  CFAEM[i1,48] <- results_MTMM_ust[18,7]
  CFAEM[i1,49] <- results_MTMM_ust[19,7]
  CFAEM[i1,50] <- results_MTMM_ust[20,7]
  CFAEM[i1,51] <- results_MTMM_ust[21,7]
  CFAEM[i1,53] <- results_MTMM_ust[18,8]
  CFAEM[i1,54] <- results_MTMM_ust[19,8]
  CFAEM[i1,55] <- results_MTMM_ust[20,8]
  CFAEM[i1,56] <- results_MTMM_ust[21,8]

  MTMM <- 'eta_est =~ NA*y1 + y2 +y3 + y4
           eta_est ~~ 1*eta_est
           y1 ~~ c(cov1,cov2,cov3)*y2
           y2 ~~ c(cov1,cov2,cov3)*y3
           y3 ~~ c(cov1,cov2,cov3)*y4'
  fit_MTMM <- cfa(MTMM,XY,group="Gruppe",group.equal = c("loadings"))
  results_MTMM_ust <- parameterEstimates(fit_MTMM)

  MCMTMM[i1,1] <- results_MTMM_ust[1,7]
  MCMTMM[i1,2] <- results_MTMM_ust[2,7]
  MCMTMM[i1,3] <- results_MTMM_ust[3,7]
  MCMTMM[i1,4] <- results_MTMM_ust[4,7]
  MCMTMM[i1,6] <- results_MTMM_ust[1,8]
  MCMTMM[i1,7] <- results_MTMM_ust[2,8]
  MCMTMM[i1,8] <- results_MTMM_ust[3,8]
  MCMTMM[i1,9] <- results_MTMM_ust[4,8]
  MCMTMM[i1,11] <- n
  MCMTMM[i1,12] <- k
  MCMTMM[i1,13] <- cv
  MCMTMM[i1,14] <- sm
  MCMTMM[i1,15] <- results_MTMM_ust[6,7]
  MCMTMM[i1,16] <- results_MTMM_ust[23,7]
  MCMTMM[i1,17] <- results_MTMM_ust[40,7]
  MCMTMM[i1,18] <- results_MTMM_ust[6,8]
  MCMTMM[i1,19] <- results_MTMM_ust[23,8]
  MCMTMM[i1,20] <- results_MTMM_ust[40,8]
}
```

```
if(k == 5){
  CFA <- 'eta_est =~ NA*y1 + y2 + y3 + y4 + y5
          eta_est ~~ 1*eta_est'
  fit_cfa <- cfa(CFA,XY,group="Gruppe")
  results_cfa_ust <- parameterEstimates(fit_cfa)

  MTMM <- 'eta_est =~ NA*y1 + y2 +y3 + y4 + y5
           eta_est ~~ 1*eta_est
           y1 ~~ c(cov1,cov2,cov3)*y2
           y2 ~~ c(cov1,cov2,cov3)*y3
           y3 ~~ c(cov1,cov2,cov3)*y4
           y4 ~~ c(cov1,cov2,cov3)*y5'
  fit_MTMM <- cfa(MTMM,XY,group="Gruppe")
  results_MTMM_ust <- parameterEstimates(fit_MTMM)

  CFAEM[i1,1] <- results_cfa_ust[1,6]
  CFAEM[i1,2] <- results_cfa_ust[2,6]
  CFAEM[i1,3] <- results_cfa_ust[3,6]
  CFAEM[i1,4] <- results_cfa_ust[4,6]
  CFAEM[i1,5] <- results_cfa_ust[5,6]
  CFAEM[i1,6] <- results_cfa_ust[1,7]
  CFAEM[i1,7] <- results_cfa_ust[2,7]
  CFAEM[i1,8] <- results_cfa_ust[3,7]
  CFAEM[i1,9] <- results_cfa_ust[4,7]
  CFAEM[i1,10] <- results_cfa_ust[5,7]
  CFAEM[i1,11] <- results_cfa_ust[18,6]
  CFAEM[i1,12] <- results_cfa_ust[19,6]
  CFAEM[i1,13] <- results_cfa_ust[20,6]
  CFAEM[i1,14] <- results_cfa_ust[21,6]
  CFAEM[i1,15] <- results_cfa_ust[22,6]
  CFAEM[i1,16] <- results_cfa_ust[18,7]
  CFAEM[i1,17] <- results_cfa_ust[19,7]
  CFAEM[i1,18] <- results_cfa_ust[20,7]
  CFAEM[i1,19] <- results_cfa_ust[21,7]
  CFAEM[i1,20] <- results_cfa_ust[22,7]
  CFAEM[i1,21] <- results_cfa_ust[35,6]
  CFAEM[i1,22] <- results_cfa_ust[36,6]
  CFAEM[i1,23] <- results_cfa_ust[37,6]
  CFAEM[i1,24] <- results_cfa_ust[38,6]
  CFAEM[i1,25] <- results_cfa_ust[39,6]
  CFAEM[i1,26] <- results_cfa_ust[35,7]
  CFAEM[i1,27] <- results_cfa_ust[36,7]
  CFAEM[i1,28] <- results_cfa_ust[37,7]
  CFAEM[i1,29] <- results_cfa_ust[38,7]
  CFAEM[i1,30] <- results_cfa_ust[39,7]
  CFAEM[i1,31] <- fitMeasures(fit_cfa)[3]
  CFAEM[i1,32] <- fitMeasures(fit_cfa)[4]
  CFAEM[i1,33] <- fitMeasures(fit_cfa)[5]
  CFAEM[i1,34] <- fitMeasures(fit_cfa)[9]
  CFAEM[i1,35] <- fitMeasures(fit_cfa)[23]
  CFAEM[i1,36] <- fitMeasures(fit_cfa)[29]
```

```
CFAEM[i1,37] <- 5

CFAEM[i1,38] <- results_MTMM_ust[1,7]
CFAEM[i1,39] <- results_MTMM_ust[2,7]
CFAEM[i1,40] <- results_MTMM_ust[3,7]
CFAEM[i1,41] <- results_MTMM_ust[4,7]
CFAEM[i1,42] <- results_MTMM_ust[5,7]
CFAEM[i1,43] <- results_MTMM_ust[1,8]
CFAEM[i1,44] <- results_MTMM_ust[2,8]
CFAEM[i1,45] <- results_MTMM_ust[3,8]
CFAEM[i1,46] <- results_MTMM_ust[4,8]
CFAEM[i1,47] <- results_MTMM_ust[5,8]
CFAEM[i1,48] <- results_MTMM_ust[22,7]
CFAEM[i1,49] <- results_MTMM_ust[23,7]
CFAEM[i1,50] <- results_MTMM_ust[24,7]
CFAEM[i1,51] <- results_MTMM_ust[25,7]
CFAEM[i1,52] <- results_MTMM_ust[26,7]
CFAEM[i1,53] <- results_MTMM_ust[22,8]
CFAEM[i1,54] <- results_MTMM_ust[23,8]
CFAEM[i1,55] <- results_MTMM_ust[24,8]
CFAEM[i1,56] <- results_MTMM_ust[25,8]
CFAEM[i1,57] <- results_MTMM_ust[26,8]

MTMM <- 'eta_est =~ NA*y1 + y2 +y3 + y4 + y5
         eta_est ~~ 1*eta_est
         y1 ~~ c(cov1,cov2,cov3)*y2
         y2 ~~ c(cov1,cov2,cov3)*y3
         y3 ~~ c(cov1,cov2,cov3)*y4
         y4 ~~ c(cov1,cov2,cov3)*y5'
fit_MTMM <- cfa(MTMM,XY,group="Gruppe",group.equal = c("loadings"))
results_MTMM_ust <- parameterEstimates(fit_MTMM)

MCMTMM[i1,1] <- results_MTMM_ust[1,7]
MCMTMM[i1,2] <- results_MTMM_ust[2,7]
MCMTMM[i1,3] <- results_MTMM_ust[3,7]
MCMTMM[i1,4] <- results_MTMM_ust[4,7]
MCMTMM[i1,5] <- results_MTMM_ust[5,7]
MCMTMM[i1,6] <- results_MTMM_ust[1,8]
MCMTMM[i1,7] <- results_MTMM_ust[2,8]
MCMTMM[i1,8] <- results_MTMM_ust[3,8]
MCMTMM[i1,9] <- results_MTMM_ust[4,8]
MCMTMM[i1,10] <- results_MTMM_ust[5,8]
MCMTMM[i1,11] <- n
MCMTMM[i1,12] <- k
MCMTMM[i1,13] <- cv
MCMTMM[i1,14] <- sm
MCMTMM[i1,15] <- results_MTMM_ust[7,7]
MCMTMM[i1,16] <- results_MTMM_ust[28,7]
MCMTMM[i1,17] <- results_MTMM_ust[49,7]
MCMTMM[i1,18] <- results_MTMM_ust[7,8]
MCMTMM[i1,19] <- results_MTMM_ust[28,8]
```

```
  MCMTMM[i1,20] <- results_MTMM_ust[49,8]
}

###-----------------------------------------------------------------###
###--- Schritt 8 -------------------------------------------------###
###--- Monte-Carlo Iterationen ------------------------------------###
###-----------------------------------------------------------------###
if(k==3){
  SEM <- 'eta_est =~ NA*y1 + y2 + y3
          xi1_est =~ NA*x1 + f2*x1 + f2*x2 + f2*x3
          xi2_est =~ NA*x4 + f2*x4 + f2*x5 + f2*x6

          eta_est ~~ 1*eta_est
          xi1_est ~~ 1*xi1_est
          xi2_est ~~ 1*xi2_est

          eta_est ~~ 0*xi1_est
          eta_est ~~ 0*xi2_est
          xi1_est ~~ 0*xi2_est

          y1xi1_est =~ NA*y1x1 + f3*y1x1 + f3*y1x2 + f3*y1x3
          y2xi1_est =~ NA*y2x1 + f3*y2x1 + f3*y2x2 + f3*y2x3

          y1xi1_est ~~ 1*y1xi1_est
          y2xi1_est ~~ 1*y2xi1_est

          y1xi2_est =~ NA*y1x4 + f3*y1x4 + f3*y1x5 + f3*y1x6
          y2xi2_est =~ NA*y2x4 + f3*y2x4 + f3*y2x5 + f3*y2x6

          y1xi2_est ~~ 1*y1xi2_est
          y2xi2_est ~~ 1*y2xi2_est

          xi1xi2_est =~ NA*x1x4 + f4*x1x4 + f4*x1x5 + f4*x1x6 + f4*x2x4 + f4*x2x5 + f4*x2x6 +
f4*x3x4 + f4*x3x5 + f4*x3x6

          xi1xi2_est ~~ 1*xi1xi2_est

          y1xi1xi2_est =~ NA*y1x1x4 + f5*y1x1x4 + f5*y1x1x5 + f5*y1x1x6 + f5*y1x2x4 +
f5*y1x2x5 + f5*y1x2x6 + f5*y1x3x4 + f5*y1x3x5 + f5*y1x3x6
          y2xi1xi2_est =~ NA*y2x1x4 + f5*y2x1x4 + f5*y2x1x5 + f5*y2x1x6 + f5*y2x2x4 +
f5*y2x2x5 + f5*y2x2x6 + f5*y2x3x4 + f5*y2x3x5 + f5*y2x3x6

          y1xi1xi2_est ~~ 1*y1xi1xi2_est
          y2xi1xi2_est ~~ 1*y2xi1xi2_est

          y2 ~ c(b1,b2,b3)*y1 + c(b4,b5,b6)*xi1_est + c(b7,b8,b9)*xi2_est +
c(b10,b11,b12)*y1xi1_est + c(b13,b14,b15)*y1xi2_est + c(b16,b17,b18)*xi1xi2_est +
c(b19,b20,b21)*y1xi1xi2_est
          y3 ~ c(b1,b2,b3)*y2 + c(b4,b5,b6)*xi1_est + c(b7,b8,b9)*xi2_est +
c(b10,b11,b12)*y2xi1_est + c(b13,b14,b15)*y2xi2_est + c(b16,b17,b18)*xi1xi2_est +
c(b19,b20,b21)*y2xi1xi2_est'
  fit_sem <- sem(SEM,XY,group="Gruppe",group.equal = c("loadings"))
  results_sem_k3 <- parameterEstimates(fit_sem)
```

```
MONTE_CARLO[i1,1] <- cv
MONTE_CARLO[i1,2] <- results_sem_k3[1,7]
MONTE_CARLO[i1,3] <- results_sem_k3[2,7]
MONTE_CARLO[i1,4] <- results_sem_k3[3,7]
MONTE_CARLO[i1,7] <- results_sem_k3[1,8]
MONTE_CARLO[i1,8] <- results_sem_k3[2,8]
MONTE_CARLO[i1,9] <- results_sem_k3[3,8]
MONTE_CARLO[i1,12] <- 0
MONTE_CARLO[i1,13] <- (1-cv)*(sm*-0.5)
MONTE_CARLO[i1,14] <- (1-cv)*(sm*-0.5)
MONTE_CARLO[i1,15] <- 0
MONTE_CARLO[i1,16] <- results_sem_k3[62,7]
MONTE_CARLO[i1,17] <- results_sem_k3[65,7]
MONTE_CARLO[i1,18] <- results_sem_k3[66,7]
MONTE_CARLO[i1,19] <- results_sem_k3[68,7]
MONTE_CARLO[i1,20] <- results_sem_k3[62,8]
MONTE_CARLO[i1,21] <- results_sem_k3[65,8]
MONTE_CARLO[i1,22] <- results_sem_k3[66,8]
MONTE_CARLO[i1,23] <- results_sem_k3[68,8]
MONTE_CARLO[i1,24] <- 0
MONTE_CARLO[i1,25] <- 0
MONTE_CARLO[i1,26] <- 0
MONTE_CARLO[i1,27] <- (1-cv)*sm
MONTE_CARLO[i1,28] <- results_sem_k3[285,7]
MONTE_CARLO[i1,29] <- results_sem_k3[288,7]
MONTE_CARLO[i1,30] <- results_sem_k3[289,7]
MONTE_CARLO[i1,31] <- results_sem_k3[291,7]
MONTE_CARLO[i1,32] <- results_sem_k3[285,8]
MONTE_CARLO[i1,33] <- results_sem_k3[288,8]
MONTE_CARLO[i1,34] <- results_sem_k3[289,8]
MONTE_CARLO[i1,35] <- results_sem_k3[291,8]
MONTE_CARLO[i1,36] <- 0
MONTE_CARLO[i1,37] <- 0
MONTE_CARLO[i1,38] <- 0
MONTE_CARLO[i1,39] <- 0
MONTE_CARLO[i1,40] <- results_sem_k3[508,7]
MONTE_CARLO[i1,41] <- results_sem_k3[511,7]
MONTE_CARLO[i1,42] <- results_sem_k3[512,7]
MONTE_CARLO[i1,43] <- results_sem_k3[514,7]
MONTE_CARLO[i1,44] <- results_sem_k3[508,8]
MONTE_CARLO[i1,45] <- results_sem_k3[511,8]
MONTE_CARLO[i1,46] <- results_sem_k3[512,8]
MONTE_CARLO[i1,47] <- results_sem_k3[514,8]
MONTE_CARLO[i1,48] <- n
MONTE_CARLO[i1,49] <- k
MONTE_CARLO[i1,50] <- cv
MONTE_CARLO[i1,51] <- sm
MONTE_CARLO[i1,52] <- 0
MONTE_CARLO[i1,53] <- 0
MONTE_CARLO[i1,54] <- 0
MONTE_CARLO[i1,55] <- results_sem_k3[63,7]
```

```
  MONTE_CARLO[i1,56] <- results_sem_k3[64,7]
  MONTE_CARLO[i1,57] <- results_sem_k3[67,7]
  MONTE_CARLO[i1,58] <- results_sem_k3[63,8]
  MONTE_CARLO[i1,59] <- results_sem_k3[64,8]
  MONTE_CARLO[i1,60] <- results_sem_k3[67,8]
  MONTE_CARLO[i1,61] <- 0
  MONTE_CARLO[i1,62] <- 0
  MONTE_CARLO[i1,63] <- 0
  MONTE_CARLO[i1,64] <- results_sem_k3[286,7]
  MONTE_CARLO[i1,65] <- results_sem_k3[287,7]
  MONTE_CARLO[i1,66] <- results_sem_k3[290,7]
  MONTE_CARLO[i1,67] <- results_sem_k3[286,8]
  MONTE_CARLO[i1,68] <- results_sem_k3[287,8]
  MONTE_CARLO[i1,69] <- results_sem_k3[290,8]
  MONTE_CARLO[i1,70] <- 0
  MONTE_CARLO[i1,71] <- 0
  MONTE_CARLO[i1,72] <- 0
  MONTE_CARLO[i1,73] <- results_sem_k3[509,7]
  MONTE_CARLO[i1,74] <- results_sem_k3[510,7]
  MONTE_CARLO[i1,75] <- results_sem_k3[513,7]
  MONTE_CARLO[i1,76] <- results_sem_k3[509,8]
  MONTE_CARLO[i1,77] <- results_sem_k3[510,8]
  MONTE_CARLO[i1,78] <- results_sem_k3[513,8]
}

if(k==4){
  SEM <- 'eta_est =~ NA*y1 + y2 + y3 + y4
          xi1_est =~ NA*x1 + f2*x1 + f2*x2 + f2*x3
          xi2_est =~ NA*x4 + f2*x4 + f2*x5 + f2*x6

          eta_est ~~ 1*eta_est
          xi1_est ~~ 1*xi1_est
          xi2_est ~~ 1*xi2_est

          eta_est ~~ 0*xi1_est
          eta_est ~~ 0*xi2_est
          xi1_est ~~ 0*xi2_est

          y1xi1_est =~ NA*y1x1 + f3*y1x1 + f3*y1x2 + f3*y1x3
          y2xi1_est =~ NA*y2x1 + f3*y2x1 + f3*y2x2 + f3*y2x3
          y3xi1_est =~ NA*y3x1 + f3*y3x1 + f3*y3x2 + f3*y3x3

          y1xi1_est ~~ 1*y1xi1_est
          y2xi1_est ~~ 1*y2xi1_est
          y3xi1_est ~~ 1*y3xi1_est

          y1xi2_est =~ NA*y1x4 + f3*y1x4 + f3*y1x5 + f3*y1x6
          y2xi2_est =~ NA*y2x4 + f3*y2x4 + f3*y2x5 + f3*y2x6
          y3xi2_est =~ NA*y3x4 + f3*y3x4 + f3*y3x5 + f3*y3x6

          y1xi2_est ~~ 1*y1xi2_est
```

```
          y2xi2_est ~~ 1*y2xi2_est
          y3xi2_est ~~ 1*y3xi2_est


          xi1xi2_est =~ NA*x1x4 + f4*x1x4 + f4*x1x5 + f4*x1x6 + f4*x2x4 + f4*x2x5 + f4*x2x6 +
f4*x3x4 + f4*x3x5 + f4*x3x6


          xi1xi2_est ~~ 1*xi1xi2_est


          y1xi1xi2_est =~ NA*y1x1x4 + f5*y1x1x4 + f5*y1x1x5 + f5*y1x1x6 + f5*y1x2x4 +
f5*y1x2x5 + f5*y1x2x6 + f5*y1x3x4 + f5*y1x3x5 + f5*y1x3x6
          y2xi1xi2_est =~ NA*y2x1x4 + f5*y2x1x4 + f5*y2x1x5 + f5*y2x1x6 + f5*y2x2x4 +
f5*y2x2x5 + f5*y2x2x6 + f5*y2x3x4 + f5*y2x3x5 + f5*y2x3x6
          y3xi1xi2_est =~ NA*y3x1x4 + f5*y3x1x4 + f5*y3x1x5 + f5*y3x1x6 + f5*y3x2x4 +
f5*y3x2x5 + f5*y3x2x6 + f5*y3x3x4 + f5*y3x3x5 + f5*y3x3x6


          y1xi1xi2_est ~~ 1*y1xi1xi2_est
          y2xi1xi2_est ~~ 1*y2xi1xi2_est
          y3xi1xi2_est ~~ 1*y3xi1xi2_est


          y2 ~ c(b1,b2,b3)*y1 + c(b4,b5,b6)*xi1_est + c(b7,b8,b9)*xi2_est +
c(b10,b11,b12)*y1xi1_est + c(b13,b14,b15)*y1xi2_est + c(b16,b17,b18)*xi1xi2_est +
c(b19,b20,b21)*y1xi1xi2_est
          y3 ~ c(b1,b2,b3)*y2 + c(b4,b5,b6)*xi1_est + c(b7,b8,b9)*xi2_est +
c(b10,b11,b12)*y2xi1_est + c(b13,b14,b15)*y2xi2_est + c(b16,b17,b18)*xi1xi2_est +
c(b19,b20,b21)*y2xi1xi2_est
          y4 ~ c(b1,b2,b3)*y3 + c(b4,b5,b6)*xi1_est + c(b7,b8,b9)*xi2_est +
c(b10,b11,b12)*y3xi1_est + c(b13,b14,b15)*y3xi2_est + c(b16,b17,b18)*xi1xi2_est +
c(b19,b20,b21)*y3xi1xi2_est'
  fit_sem <- sem(SEM,XY,group="Gruppe",group.equal = c("loadings"))
  results_sem_k4 <- parameterEstimates(fit_sem)


  MONTE_CARLO[i1,1] <- cv
  MONTE_CARLO[i1,2] <- results_sem_k4[1,7]
  MONTE_CARLO[i1,3] <- results_sem_k4[2,7]
  MONTE_CARLO[i1,4] <- results_sem_k4[3,7]
  MONTE_CARLO[i1,5] <- results_sem_k4[4,7]
  MONTE_CARLO[i1,7] <- results_sem_k4[1,8]
  MONTE_CARLO[i1,8] <- results_sem_k4[2,8]
  MONTE_CARLO[i1,9] <- results_sem_k4[3,8]
  MONTE_CARLO[i1,10] <- results_sem_k4[4,8]
  MONTE_CARLO[i1,12] <- 0
  MONTE_CARLO[i1,13] <- (1-cv)*(sm*-0.5)
  MONTE_CARLO[i1,14] <- (1-cv)*(sm*-0.5)
  MONTE_CARLO[i1,15] <- 0
  MONTE_CARLO[i1,16] <- results_sem_k4[81,7]
  MONTE_CARLO[i1,17] <- results_sem_k4[84,7]
  MONTE_CARLO[i1,18] <- results_sem_k4[85,7]
  MONTE_CARLO[i1,19] <- results_sem_k4[87,7]
  MONTE_CARLO[i1,20] <- results_sem_k4[81,8]
  MONTE_CARLO[i1,21] <- results_sem_k4[84,8]
  MONTE_CARLO[i1,22] <- results_sem_k4[85,8]
  MONTE_CARLO[i1,23] <- results_sem_k4[87,8]
  MONTE_CARLO[i1,24] <- 0
  MONTE_CARLO[i1,25] <- 0
  MONTE_CARLO[i1,26] <- 0
```

```
MONTE_CARLO[i1,27] <- (1-cv)*sm
MONTE_CARLO[i1,28] <- results_sem_k4[398,7]
MONTE_CARLO[i1,29] <- results_sem_k4[401,7]
MONTE_CARLO[i1,30] <- results_sem_k4[402,7]
MONTE_CARLO[i1,31] <- results_sem_k4[404,7]
MONTE_CARLO[i1,32] <- results_sem_k4[398,8]
MONTE_CARLO[i1,33] <- results_sem_k4[401,8]
MONTE_CARLO[i1,34] <- results_sem_k4[402,8]
MONTE_CARLO[i1,35] <- results_sem_k4[404,8]
MONTE_CARLO[i1,36] <- 0
MONTE_CARLO[i1,37] <- 0
MONTE_CARLO[i1,38] <- 0
MONTE_CARLO[i1,39] <- 0
MONTE_CARLO[i1,40] <- results_sem_k4[715,7]
MONTE_CARLO[i1,41] <- results_sem_k4[718,7]
MONTE_CARLO[i1,42] <- results_sem_k4[719,7]
MONTE_CARLO[i1,43] <- results_sem_k4[721,7]
MONTE_CARLO[i1,44] <- results_sem_k4[715,8]
MONTE_CARLO[i1,45] <- results_sem_k4[718,8]
MONTE_CARLO[i1,46] <- results_sem_k4[719,8]
MONTE_CARLO[i1,47] <- results_sem_k4[721,8]
MONTE_CARLO[i1,48] <- n
MONTE_CARLO[i1,49] <- k
MONTE_CARLO[i1,50] <- cv
MONTE_CARLO[i1,51] <- sm
MONTE_CARLO[i1,52] <- 0
MONTE_CARLO[i1,53] <- 0
MONTE_CARLO[i1,54] <- 0
MONTE_CARLO[i1,55] <- results_sem_k4[82,7]
MONTE_CARLO[i1,56] <- results_sem_k4[83,7]
MONTE_CARLO[i1,57] <- results_sem_k4[86,7]
MONTE_CARLO[i1,58] <- results_sem_k4[82,8]
MONTE_CARLO[i1,59] <- results_sem_k4[83,8]
MONTE_CARLO[i1,60] <- results_sem_k4[86,8]
MONTE_CARLO[i1,61] <- 0
MONTE_CARLO[i1,62] <- 0
MONTE_CARLO[i1,63] <- 0
MONTE_CARLO[i1,64] <- results_sem_k4[399,7]
MONTE_CARLO[i1,65] <- results_sem_k4[400,7]
MONTE_CARLO[i1,66] <- results_sem_k4[403,7]
MONTE_CARLO[i1,67] <- results_sem_k4[399,8]
MONTE_CARLO[i1,68] <- results_sem_k4[400,8]
MONTE_CARLO[i1,69] <- results_sem_k4[403,8]
MONTE_CARLO[i1,70] <- 0
MONTE_CARLO[i1,71] <- 0
MONTE_CARLO[i1,72] <- 0
MONTE_CARLO[i1,73] <- results_sem_k4[716,7]
MONTE_CARLO[i1,74] <- results_sem_k4[717,7]
MONTE_CARLO[i1,75] <- results_sem_k4[720,7]
MONTE_CARLO[i1,76] <- results_sem_k4[716,8]
MONTE_CARLO[i1,77] <- results_sem_k4[717,8]
```

```
  MONTE_CARLO[i1,78] <- results_sem_k4[720,8]
}

if(k==5){
  SEM <- 'eta_est =~ NA*y1 + y2 + y3 + y4 + y5
          xi1_est =~ NA*x1 + f2*x1 + f2*x2 + f2*x3
          xi2_est =~ NA*x4 + f2*x4 + f2*x5 + f2*x6

          eta_est ~~ 1*eta_est
          xi1_est ~~ 1*xi1_est
          xi2_est ~~ 1*xi2_est

          eta_est ~~ 0*xi1_est
          eta_est ~~ 0*xi2_est
          xi1_est ~~ 0*xi2_est

          y1xi1_est =~ NA*y1x1 + f3*y1x1 + f3*y1x2 + f3*y1x3
          y2xi1_est =~ NA*y2x1 + f3*y2x1 + f3*y2x2 + f3*y2x3
          y3xi1_est =~ NA*y3x1 + f3*y3x1 + f3*y3x2 + f3*y3x3
          y4xi1_est =~ NA*y4x1 + f3*y4x1 + f3*y4x2 + f3*y4x3

          y1xi1_est ~~ 1*y1xi1_est
          y2xi1_est ~~ 1*y2xi1_est
          y3xi1_est ~~ 1*y3xi1_est
          y4xi1_est ~~ 1*y4xi1_est

          y1xi2_est =~ NA*y1x4 + f3*y1x4 + f3*y1x5 + f3*y1x6
          y2xi2_est =~ NA*y2x4 + f3*y2x4 + f3*y2x5 + f3*y2x6
          y3xi2_est =~ NA*y3x4 + f3*y3x4 + f3*y3x5 + f3*y3x6
          y4xi2_est =~ NA*y4x4 + f3*y4x4 + f3*y4x5 + f3*y4x6

          y1xi2_est ~~ 1*y1xi2_est
          y2xi2_est ~~ 1*y2xi2_est
          y3xi2_est ~~ 1*y3xi2_est
          y4xi2_est ~~ 1*y4xi2_est

          xi1xi2_est =~ NA*x1x4 + f4*x1x4 + f4*x1x5 + f4*x1x6 + f4*x2x4 + f4*x2x5 + f4*x2x6 +
f4*x3x4 + f4*x3x5 + f4*x3x6

          xi1xi2_est ~~ 1*xi1xi2_est

          y1xi1xi2_est =~ NA*y1x1x4 + f5*y1x1x4 + f5*y1x1x5 + f5*y1x1x6 + f5*y1x2x4 +
f5*y1x2x5 + f5*y1x2x6 + f5*y1x3x4 + f5*y1x3x5 + f5*y1x3x6
          y2xi1xi2_est =~ NA*y2x1x4 + f5*y2x1x4 + f5*y2x1x5 + f5*y2x1x6 + f5*y2x2x4 +
f5*y2x2x5 + f5*y2x2x6 + f5*y2x3x4 + f5*y2x3x5 + f5*y2x3x6
          y3xi1xi2_est =~ NA*y3x1x4 + f5*y3x1x4 + f5*y3x1x5 + f5*y3x1x6 + f5*y3x2x4 +
f5*y3x2x5 + f5*y3x2x6 + f5*y3x3x4 + f5*y3x3x5 + f5*y3x3x6
          y4xi1xi2_est =~ NA*y4x1x4 + f5*y4x1x4 + f5*y4x1x5 + f5*y4x1x6 + f5*y4x2x4 +
f5*y4x2x5 + f5*y4x2x6 + f5*y4x3x4 + f5*y4x3x5 + f5*y4x3x6

          y1xi1xi2_est ~~ 1*y1xi1xi2_est
          y2xi1xi2_est ~~ 1*y2xi1xi2_est
          y3xi1xi2_est ~~ 1*y3xi1xi2_est
```

```
          y4xi1xi2_est ~~ 1*y4xi1xi2_est


          y2 ~ c(b1,b2,b3)*y1 + c(b4,b5,b6)*xi1_est + c(b7,b8,b9)*xi2_est +
c(b10,b11,b12)*y1xi1_est + c(b13,b14,b15)*y1xi2_est + c(b16,b17,b18)*xi1xi2_est +
c(b19,b20,b21)*y1xi1xi2_est

          y3 ~ c(b1,b2,b3)*y2 + c(b4,b5,b6)*xi1_est + c(b7,b8,b9)*xi2_est +
c(b10,b11,b12)*y2xi1_est + c(b13,b14,b15)*y2xi2_est + c(b16,b17,b18)*xi1xi2_est +
c(b19,b20,b21)*y2xi1xi2_est

          y4 ~ c(b1,b2,b3)*y3 + c(b4,b5,b6)*xi1_est + c(b7,b8,b9)*xi2_est +
c(b10,b11,b12)*y3xi1_est + c(b13,b14,b15)*y3xi2_est + c(b16,b17,b18)*xi1xi2_est +
c(b19,b20,b21)*y3xi1xi2_est

          y5 ~ c(b1,b2,b3)*y4 + c(b4,b5,b6)*xi1_est + c(b7,b8,b9)*xi2_est +
c(b10,b11,b12)*y4xi1_est + c(b13,b14,b15)*y4xi2_est + c(b16,b17,b18)*xi1xi2_est +
c(b19,b20,b21)*y4xi1xi2_est'
  fit_sem <- sem(SEM,XY,group="Gruppe",group.equal = c("loadings"))
  results_sem_k5 <- parameterEstimates(fit_sem)


  MONTE_CARLO[i1,1] <- cv
  MONTE_CARLO[i1,2] <- results_sem_k5[1,7]
  MONTE_CARLO[i1,3] <- results_sem_k5[2,7]
  MONTE_CARLO[i1,4] <- results_sem_k5[3,7]
  MONTE_CARLO[i1,5] <- results_sem_k5[4,7]
  MONTE_CARLO[i1,6] <- results_sem_k5[5,7]
  MONTE_CARLO[i1,7] <- results_sem_k5[1,8]
  MONTE_CARLO[i1,8] <- results_sem_k5[2,8]
  MONTE_CARLO[i1,9] <- results_sem_k5[3,8]
  MONTE_CARLO[i1,10] <- results_sem_k5[4,8]
  MONTE_CARLO[i1,11] <- results_sem_k5[5,8]
  MONTE_CARLO[i1,12] <- 0
  MONTE_CARLO[i1,13] <- (1-cv)*(sm*-0.5)
  MONTE_CARLO[i1,14] <- (1-cv)*(sm*-0.5)
  MONTE_CARLO[i1,15] <- 0
  MONTE_CARLO[i1,16] <- results_sem_k5[100,7]
  MONTE_CARLO[i1,17] <- results_sem_k5[103,7]
  MONTE_CARLO[i1,18] <- results_sem_k5[104,7]
  MONTE_CARLO[i1,19] <- results_sem_k5[106,7]
  MONTE_CARLO[i1,20] <- results_sem_k5[100,8]
  MONTE_CARLO[i1,21] <- results_sem_k5[103,8]
  MONTE_CARLO[i1,22] <- results_sem_k5[104,8]
  MONTE_CARLO[i1,23] <- results_sem_k5[106,8]
  MONTE_CARLO[i1,24] <- 0
  MONTE_CARLO[i1,25] <- 0
  MONTE_CARLO[i1,26] <- 0
  MONTE_CARLO[i1,27] <- (1-cv)*sm
  MONTE_CARLO[i1,28] <- results_sem_k5[520,7]
  MONTE_CARLO[i1,29] <- results_sem_k5[523,7]
  MONTE_CARLO[i1,30] <- results_sem_k5[524,7]
  MONTE_CARLO[i1,31] <- results_sem_k5[526,7]
  MONTE_CARLO[i1,32] <- results_sem_k5[520,8]
  MONTE_CARLO[i1,33] <- results_sem_k5[523,8]
  MONTE_CARLO[i1,34] <- results_sem_k5[524,8]
  MONTE_CARLO[i1,35] <- results_sem_k5[526,8]
  MONTE_CARLO[i1,36] <- 0
  MONTE_CARLO[i1,37] <- 0
```

```
  MONTE_CARLO[i1,38] <- 0
  MONTE_CARLO[i1,39] <- 0
  MONTE_CARLO[i1,40] <- results_sem_k5[940,7]
  MONTE_CARLO[i1,41] <- results_sem_k5[943,7]
  MONTE_CARLO[i1,42] <- results_sem_k5[944,7]
  MONTE_CARLO[i1,43] <- results_sem_k5[946,7]
  MONTE_CARLO[i1,44] <- results_sem_k5[940,8]
  MONTE_CARLO[i1,45] <- results_sem_k5[943,8]
  MONTE_CARLO[i1,46] <- results_sem_k5[944,8]
  MONTE_CARLO[i1,47] <- results_sem_k5[946,8]
  MONTE_CARLO[i1,48] <- n
  MONTE_CARLO[i1,49] <- k
  MONTE_CARLO[i1,50] <- cv
  MONTE_CARLO[i1,51] <- sm
  MONTE_CARLO[i1,52] <- 0
  MONTE_CARLO[i1,53] <- 0
  MONTE_CARLO[i1,54] <- 0
  MONTE_CARLO[i1,55] <- results_sem_k5[101,7]
  MONTE_CARLO[i1,56] <- results_sem_k5[102,7]
  MONTE_CARLO[i1,57] <- results_sem_k5[105,7]
  MONTE_CARLO[i1,58] <- results_sem_k5[101,8]
  MONTE_CARLO[i1,59] <- results_sem_k5[102,8]
  MONTE_CARLO[i1,60] <- results_sem_k5[105,8]
  MONTE_CARLO[i1,61] <- 0
  MONTE_CARLO[i1,62] <- 0
  MONTE_CARLO[i1,63] <- 0
  MONTE_CARLO[i1,64] <- results_sem_k5[521,7]
  MONTE_CARLO[i1,65] <- results_sem_k5[522,7]
  MONTE_CARLO[i1,66] <- results_sem_k5[525,7]
  MONTE_CARLO[i1,67] <- results_sem_k5[521,8]
  MONTE_CARLO[i1,68] <- results_sem_k5[522,8]
  MONTE_CARLO[i1,69] <- results_sem_k5[525,8]
  MONTE_CARLO[i1,70] <- 0
  MONTE_CARLO[i1,71] <- 0
  MONTE_CARLO[i1,72] <- 0
  MONTE_CARLO[i1,73] <- results_sem_k5[941,7]
  MONTE_CARLO[i1,74] <- results_sem_k5[942,7]
  MONTE_CARLO[i1,75] <- results_sem_k5[945,7]
  MONTE_CARLO[i1,76] <- results_sem_k5[941,8]
  MONTE_CARLO[i1,77] <- results_sem_k5[942,8]
  MONTE_CARLO[i1,78] <- results_sem_k5[945,8]
}

###-----------------------------------------------------------------###
###--- Schritt 9 -------------------------------------------------###
###--- Schließe Monte-Carlo Loop und Error Handler ----------------###
###-----------------------------------------------------------------###
  print(paste0("Abgeschlossene Iterationen: ", i1))
  }, error=function(e){})
}
```

```
###-----------------------------------------------------------------###
###--- Schritt 10 ------------------------------------------------###
###--- Bias Kalkulation ------------------------------------------###
###-----------------------------------------------------------------###
n_emp <- (MONTE_CARLO[,48]-1500)/1000
k_emp <-  MONTE_CARLO[,49]
k3_emp <- ifelse(k_emp == 3,1,0)
k4_emp <- ifelse(k_emp == 4,1,0)
cv_emp <- (MONTE_CARLO[,50]-0.6)*100
sm_emp <- (MONTE_CARLO[,51]-0.5)*100
cvXsm_emp <- cv_emp*sm_emp


g1_f_cfa <- data.frame((CFAEM[,1]-MONTE_CARLO[,1])^2,(CFAEM[,2]-MONTE_CARLO[,1])^2,(CFAEM[,3]-
MONTE_CARLO[,1])^2,(CFAEM[,4]-MONTE_CARLO[,1])^2,(CFAEM[,5]-MONTE_CARLO[,1])^2)
g2_f_cfa <- data.frame((CFAEM[,11]-MONTE_CARLO[,1])^2,(CFAEM[,12]-
MONTE_CARLO[,1])^2,(CFAEM[,13]-MONTE_CARLO[,1])^2,(CFAEM[,14]-MONTE_CARLO[,1])^2,(CFAEM[,15]-
MONTE_CARLO[,1])^2)
g3_f_cfa <- data.frame((CFAEM[,21]-MONTE_CARLO[,1])^2,(CFAEM[,22]-
MONTE_CARLO[,1])^2,(CFAEM[,23]-MONTE_CARLO[,1])^2,(CFAEM[,24]-MONTE_CARLO[,1])^2,(CFAEM[,25]-
MONTE_CARLO[,1])^2)


g4_f_cfa <- data.frame((CFAEM[,38]-MONTE_CARLO[,1])^2,(CFAEM[,39]-
MONTE_CARLO[,1])^2,(CFAEM[,40]-MONTE_CARLO[,1])^2,(CFAEM[,41]-MONTE_CARLO[,1])^2,(CFAEM[,42]-
MONTE_CARLO[,1])^2)
g5_f_cfa <- data.frame((CFAEM[,48]-MONTE_CARLO[,1])^2,(CFAEM[,49]-
MONTE_CARLO[,1])^2,(CFAEM[,50]-MONTE_CARLO[,1])^2,(CFAEM[,51]-MONTE_CARLO[,1])^2,(CFAEM[,52]-
MONTE_CARLO[,1])^2)


bias_g1_cfa <- round(sqrt(rowMeans(g1_f_cfa,na.rm=TRUE))*100,digits=3)
bias_g2_cfa <- round(sqrt(rowMeans(g2_f_cfa,na.rm=TRUE))*100,digits=3)
bias_g3_cfa <- round(sqrt(rowMeans(g3_f_cfa,na.rm=TRUE))*100,digits=3)


bias_g4_cfa <- round(sqrt(rowMeans(g4_f_cfa,na.rm=TRUE))*100,digits=3)
bias_g5_cfa <- round(sqrt(rowMeans(g5_f_cfa,na.rm=TRUE))*100,digits=3)


gx_f_MTMM <- data.frame((MCMTMM[,1]- MCMTMM[,13])^2,(MCMTMM[,2]- MCMTMM[,13])^2,(MCMTMM[,3]-
MCMTMM[,13])^2,(MCMTMM[,4]- MCMTMM[,13])^2,(MCMTMM[,5]- MCMTMM[,13])^2)
bias_f_MTMM <- round(sqrt(rowMeans(gx_f_MTMM,na.rm = TRUE))*100,digits=3)


gx_f_sem <- data.frame((MONTE_CARLO[,2]- MONTE_CARLO[,1])^2,(MONTE_CARLO[,3]-
MONTE_CARLO[,1])^2,(MONTE_CARLO[,4]- MONTE_CARLO[,1])^2,(MONTE_CARLO[,5]-
MONTE_CARLO[,1])^2,(MONTE_CARLO[,6]- MONTE_CARLO[,1])^2)
bias_f_sem <- round(sqrt(rowMeans(gx_f_sem,na.rm = TRUE))*100,digits=3)


bias_b1_true <- round(sqrt(((MONTE_CARLO[,16] - MONTE_CARLO[,12])^2 + (MONTE_CARLO[,28] -
MONTE_CARLO[,24])^2 + (MONTE_CARLO[,40] - MONTE_CARLO[,36])^2)/3)*100,digits=3)
bias_b4_true <- round(sqrt(((MONTE_CARLO[,17] - MONTE_CARLO[,13])^2 + (MONTE_CARLO[,29] -
MONTE_CARLO[,25])^2 + (MONTE_CARLO[,41] - MONTE_CARLO[,37])^2)/3)*100,digits=3)
bias_b5_true <- round(sqrt(((MONTE_CARLO[,18] - MONTE_CARLO[,14])^2 + (MONTE_CARLO[,30] -
MONTE_CARLO[,26])^2 + (MONTE_CARLO[,42] - MONTE_CARLO[,38])^2)/3)*100,digits=3)
bias_b7_true <- round(sqrt(((MONTE_CARLO[,19] - MONTE_CARLO[,15])^2 + (MONTE_CARLO[,31] -
MONTE_CARLO[,27])^2 + (MONTE_CARLO[,43] - MONTE_CARLO[,39])^2)/3)*100,digits=3)


bias_b2_true <- round(sqrt(((MONTE_CARLO[,55] - MONTE_CARLO[,52])^2 + (MONTE_CARLO[,64] -
MONTE_CARLO[,61])^2 + (MONTE_CARLO[,73] - MONTE_CARLO[,70])^2)/3)*100,digits=3)
bias_b3_true <- round(sqrt(((MONTE_CARLO[,56] - MONTE_CARLO[,53])^2 + (MONTE_CARLO[,65] -
MONTE_CARLO[,62])^2 + (MONTE_CARLO[,74] - MONTE_CARLO[,71])^2)/3)*100,digits=3)
bias_b6_true <- round(sqrt(((MONTE_CARLO[,57] - MONTE_CARLO[,54])^2 + (MONTE_CARLO[,66] -
MONTE_CARLO[,63])^2 + (MONTE_CARLO[,75] - MONTE_CARLO[,72])^2)/3)*100,digits=3)
```

```
se_b1_sem <- round(((MONTE_CARLO[,20] + MONTE_CARLO[,32] + MONTE_CARLO[,44])/3)*100,digits=3)
se_b4_sem <- round(((MONTE_CARLO[,21] + MONTE_CARLO[,33] + MONTE_CARLO[,45])/3)*100,digits=3)
se_b5_sem <- round(((MONTE_CARLO[,22] + MONTE_CARLO[,34] + MONTE_CARLO[,46])/3)*100,digits=3)
se_b7_sem <- round(((MONTE_CARLO[,23] + MONTE_CARLO[,35] + MONTE_CARLO[,47])/3)*100,digits=3)

se_b2_sem <- round(((MONTE_CARLO[,58] + MONTE_CARLO[,67] + MONTE_CARLO[,76])/3)*100,digits=3)
se_b3_sem <- round(((MONTE_CARLO[,59] + MONTE_CARLO[,68] + MONTE_CARLO[,77])/3)*100,digits=3)
se_b6_sem <- round(((MONTE_CARLO[,60] + MONTE_CARLO[,69] + MONTE_CARLO[,78])/3)*100,digits=3)

cov1_MTMM <- round(sqrt((MCMTMM[,15]*100)^2),digits=3)
cov2_MTMM <- round(sqrt((MCMTMM[,16]*100)^2),digits=3)
cov3_MTMM <- round(sqrt((MCMTMM[,17]*100)^2),digits=3)

secov1_MTMM <- round(MCMTMM[,18]*100,digits=3)
secov2_MTMM <- round(MCMTMM[,19]*100,digits=3)
secov3_MTMM <- round(MCMTMM[,20]*100,digits=3)

###-----------------------------------------------------------------###
###--- Schritt 11 ------------------------------------------------###
###--- Analysiere Verteilung von Y und X -----------------------###
###-----------------------------------------------------------------###
DF_Y <-
data.frame(bias_g1_cfa,bias_g2_cfa,bias_g3_cfa,bias_g4_cfa,bias_g5_cfa,bias_f_sem,bias_f_MTMM,
                      bias_b1_true,bias_b4_true,bias_b5_true,bias_b7_true,
                      bias_b2_true,bias_b3_true,bias_b6_true,
                      cov1_MTMM,cov2_MTMM,cov3_MTMM,
                      se_b1_sem,se_b4_sem,se_b5_sem,se_b7_sem,
                      se_b2_sem,se_b3_sem,se_b6_sem,
                      secov1_MTMM,secov2_MTMM,secov3_MTMM)

outl <- 100
DF_Y[DF_Y == "NaN"] <- NA
DF_Y[DF_Y > outl] <- NA
DF_Y[!!rowSums(is.na(DF_Y)),] <- NA

NORM_Y <- matrix(ncol=7,nrow=27)
colnames(NORM_Y) <- c("min","max","mean","sd","skew","kurt","% Missing")
rownames(NORM_Y) <- colnames(DF_Y)

for(i in 1:27){
  NORM_Y[i,1] <- round(min(DF_Y[,i],na.rm=TRUE),digits=4)
  NORM_Y[i,2] <- round(max(DF_Y[,i],na.rm=TRUE),digits=4)
  NORM_Y[i,3] <- round(mean(DF_Y[,i],na.rm=TRUE),digits=4)
  NORM_Y[i,4] <- round(sd(DF_Y[,i],na.rm=TRUE),digits=4)
  NORM_Y[i,5] <- round(skewness(DF_Y[,i],na.rm=TRUE),digits=4)
  NORM_Y[i,6] <- round(kurtosis(DF_Y[,i],na.rm=TRUE),digits=4)
  NORM_Y[i,7] <- round((sum(is.na(DF_Y[[i]]))/10000)*100,digits=2)
}

write.table(NORM_Y,"NORM_Y.txt",sep = "\t",row.names = TRUE, col.names = TRUE)
```

```
DF_X <- data.frame(MONTE_CARLO[,48],MONTE_CARLO[,49],MONTE_CARLO[,50],MONTE_CARLO[,51])

NORM_X <- matrix(ncol=6,nrow=4)
colnames(NORM_X) <- c("min","max","mean","sd","skew","kurt")
rownames(NORM_X) <- c("n","k","cv","sm")

for(i in 1:4){
  NORM_X[i,1] <- round(min(DF_X[,i],na.rm=TRUE),digits=4)
  NORM_X[i,2] <- round(max(DF_X[,i],na.rm=TRUE),digits=4)
  NORM_X[i,3] <- round(mean(DF_X[,i],na.rm=TRUE),digits=4)
  NORM_X[i,4] <- round(sd(DF_X[,i],na.rm=TRUE),digits=4)
  NORM_X[i,5] <- round(skewness(DF_X[,i],na.rm=TRUE),digits=4)
  NORM_X[i,6] <- round(kurtosis(DF_X[,i],na.rm=TRUE),digits=4)
}

write.table(NORM_X,"NORM_X.txt",sep = "\t",row.names = TRUE, col.names = TRUE)

for(i in 1:27){
  x_breaks <- 100
  x_lim <- 25
  if(NORM_Y[i,2] > 100){x_breaks <- NORM_Y[i,2]}
  if(NORM_Y[i,2] < 25){x_lim <- NORM_Y[i,2]}
  png(file=paste0(colnames(DF_Y)[i],".png"))
  hist(DF_Y[,i],main=paste0("Histogram für
",colnames(DF_Y)[i]),xlab=colnames(DF_Y)[i],breaks=x_breaks,xlim=range(c(0,x_lim)))
  dev.off()
}

png(file="n.png")
hist(DF_X[,1],main=paste0("Histogram für
",rownames(NORM_X)[1]),xlab=rownames(NORM_X)[1],breaks=24)
dev.off()
factor_k <- as.factor(DF_X[,2])
png(file="k.png")
plot(factor_k,main="Balkendiagramm für k")
dev.off()
png(file="cv.png")
hist(DF_X[,3],main=paste0("Histogram für ",rownames(NORM_X)[3]),xlab=rownames(NORM_X)[3])
dev.off()
png(file="sm.png")
hist(DF_X[,4],main=paste0("Histogram für ",rownames(NORM_X)[4]),xlab=rownames(NORM_X)[4])
dev.off()

###-----------------------------------------------------------------###
###--- Schritt 12 ------------------------------------------------###
###--- MLR Regressionen ------------------------------------------###
###-----------------------------------------------------------------###
DF_ALL <- data.frame(DF_Y,n_emp,k_emp,k3_emp,k4_emp,cv_emp,sm_emp,cvXsm_emp)

sem_g1_cfa <- sem("bias_g1_cfa ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)
```

```
sem_g2_cfa <- sem("bias_g2_cfa ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)

sem_g3_cfa <- sem("bias_g3_cfa ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)

sem_g4_cfa <- sem("bias_g4_cfa ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)

sem_g5_cfa <- sem("bias_g5_cfa ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)

sem_f_sem <- sem("bias_f_sem ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)


sem_f_MTMM <- sem("bias_f_MTMM ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)


sem_b1_sem <- sem("bias_b1_true ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)

sem_b4_sem <- sem("bias_b4_true ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)

sem_b5_sem <- sem("bias_b5_true ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)

sem_b7_sem <- sem("bias_b7_true ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)


sem_b2_sem <- sem("bias_b2_true ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)

sem_b3_sem <- sem("bias_b3_true ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)

sem_b6_sem <- sem("bias_b6_true ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)


sem_cov1_MTMM <- sem("cov1_MTMM ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)

sem_cov2_MTMM <- sem("cov2_MTMM ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)

sem_cov3_MTMM <- sem("cov3_MTMM ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)


sem_se1_sem <- sem("se_b1_sem ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)

sem_se4_sem <- sem("se_b4_sem ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)

sem_se5_sem <- sem("se_b5_sem ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)

sem_se7_sem <- sem("se_b7_sem ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)


sem_se2_sem <- sem("se_b2_sem ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)

sem_se3_sem <- sem("se_b3_sem ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)

sem_se6_sem <- sem("se_b6_sem ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)


sem_secov1_MTMM <- sem("secov1_MTMM ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)

sem_secov2_MTMM <- sem("secov2_MTMM ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)

sem_secov3_MTMM <- sem("secov3_MTMM ~ n_emp + k3_emp + k4_emp + cv_emp + sm_emp +
cvXsm_emp",DF_ALL,estimator="MLR",meanstructure=TRUE)


for(i1 in 1:7){
  for(i2 in 1:4){
```

```
    FAKTOR[i1,i2] <- round(parameterestimates(sem_g1_cfa)[i1,i2+3],digits=3)
    FAKTOR[i1,i2+4] <- round(parameterestimates(sem_g2_cfa)[i1,i2+3],digits=3)
    FAKTOR[i1,i2+8] <- round(parameterestimates(sem_g3_cfa)[i1,i2+3],digits=3)
    FAKTOR[i1,i2+12] <- round(parameterestimates(sem_g4_cfa)[i1,i2+3],digits=3)
    FAKTOR[i1,i2+16] <- round(parameterestimates(sem_g5_cfa)[i1,i2+3],digits=3)
    FAKTOR[i1,i2+20] <- round(parameterestimates(sem_f_sem)[i1,i2+3],digits=3)
    FAKTOR[i1,i2+24] <- round(parameterestimates(sem_f_MTMM)[i1,i2+3],digits=3)
    if(i1 == 7){
      FAKTOR[i1,i2] <- round(parameterestimates(sem_g1_cfa)[29,i2+3],digits=3)
      FAKTOR[i1,i2+4] <- round(parameterestimates(sem_g2_cfa)[29,i2+3],digits=3)
      FAKTOR[i1,i2+8] <- round(parameterestimates(sem_g3_cfa)[29,i2+3],digits=3)
      FAKTOR[i1,i2+12] <- round(parameterestimates(sem_g4_cfa)[29,i2+3],digits=3)
      FAKTOR[i1,i2+16] <- round(parameterestimates(sem_g5_cfa)[29,i2+3],digits=3)
      FAKTOR[i1,i2+20] <- round(parameterestimates(sem_f_sem)[29,i2+3],digits=3)
      FAKTOR[i1,i2+24] <- round(parameterestimates(sem_f_MTMM)[29,i2+3],digits=3)
    }
    if(i2 == 3){
      FAKTOR[i1,i2] <- round(standardizedSolution(sem_g1_cfa)[i1,4],digits=3)
      FAKTOR[i1,i2+4] <- round(standardizedSolution(sem_g2_cfa)[i1,4],digits=3)
      FAKTOR[i1,i2+8] <- round(standardizedSolution(sem_g3_cfa)[i1,4],digits=3)
      FAKTOR[i1,i2+12] <- round(standardizedSolution(sem_g4_cfa)[i1,4],digits=3)
      FAKTOR[i1,i2+16] <- round(standardizedSolution(sem_g5_cfa)[i1,4],digits=3)
      FAKTOR[i1,i2+20] <- round(standardizedSolution(sem_f_sem)[i1,4],digits=3)
      FAKTOR[i1,i2+24] <- round(standardizedSolution(sem_f_MTMM)[i1,4],digits=3)
    }
    if(i2 == 3 && i1 == 7){
      FAKTOR[i1,i2] <- NA
      FAKTOR[i1,i2+4] <- NA
      FAKTOR[i1,i2+8] <- NA
      FAKTOR[i1,i2+12] <- NA
      FAKTOR[i1,i2+16] <- NA
      FAKTOR[i1,i2+20] <- NA
      FAKTOR[i1,i2+24] <- NA
    }
  }
}

for(i1 in 1:7){
  for(i2 in 1:4){
    EFFEKTS[i1,i2] <- round(parameterestimates(sem_b1_sem)[i1,i2+3],digits=3)
    EFFEKTS[i1,i2+4] <- round(parameterestimates(sem_b4_sem)[i1,i2+3],digits=3)
    EFFEKTS[i1,i2+8] <- round(parameterestimates(sem_b5_sem)[i1,i2+3],digits=3)
    EFFEKTS[i1,i2+12] <- round(parameterestimates(sem_b7_sem)[i1,i2+3],digits=3)
    EFFEKTS[i1,i2+16] <- round(parameterestimates(sem_b2_sem)[i1,i2+3],digits=3)
    EFFEKTS[i1,i2+20] <- round(parameterestimates(sem_b3_sem)[i1,i2+3],digits=3)
    EFFEKTS[i1,i2+24] <- round(parameterestimates(sem_b6_sem)[i1,i2+3],digits=3)
    EFFEKTS[i1,i2+28] <- round(parameterestimates(sem_cov1_MTMM)[i1,i2+3],digits=3)
    EFFEKTS[i1,i2+32] <- round(parameterestimates(sem_cov2_MTMM)[i1,i2+3],digits=3)
    EFFEKTS[i1,i2+36] <- round(parameterestimates(sem_cov3_MTMM)[i1,i2+3],digits=3)
    if(i1 == 7){
      EFFEKTS[i1,i2] <- round(parameterestimates(sem_b1_sem)[29,i2+3],digits=3)
```

```
      EFFEKTS[i1,i2+4] <- round(parameterestimates(sem_b4_sem)[29,i2+3],digits=3)
      EFFEKTS[i1,i2+8] <- round(parameterestimates(sem_b5_sem)[29,i2+3],digits=3)
      EFFEKTS[i1,i2+12] <- round(parameterestimates(sem_b7_sem)[29,i2+3],digits=3)
      EFFEKTS[i1,i2+16] <- round(parameterestimates(sem_b2_sem)[29,i2+3],digits=3)
      EFFEKTS[i1,i2+20] <- round(parameterestimates(sem_b3_sem)[29,i2+3],digits=3)
      EFFEKTS[i1,i2+24] <- round(parameterestimates(sem_b6_sem)[29,i2+3],digits=3)
      EFFEKTS[i1,i2+28] <- round(parameterestimates(sem_cov1_MTMM)[29,i2+3],digits=3)
      EFFEKTS[i1,i2+32] <- round(parameterestimates(sem_cov2_MTMM)[29,i2+3],digits=3)
      EFFEKTS[i1,i2+36] <- round(parameterestimates(sem_cov3_MTMM)[29,i2+3],digits=3)
    }
    if(i2 == 3){
      EFFEKTS[i1,i2] <- round(standardizedSolution(sem_b1_sem)[i1,4],digits=3)
      EFFEKTS[i1,i2+4] <- round(standardizedSolution(sem_b4_sem)[i1,4],digits=3)
      EFFEKTS[i1,i2+8] <- round(standardizedSolution(sem_b5_sem)[i1,4],digits=3)
      EFFEKTS[i1,i2+12] <- round(standardizedSolution(sem_b7_sem)[i1,4],digits=3)
      EFFEKTS[i1,i2+16] <- round(standardizedSolution(sem_b2_sem)[i1,4],digits=3)
      EFFEKTS[i1,i2+20] <- round(standardizedSolution(sem_b3_sem)[i1,4],digits=3)
      EFFEKTS[i1,i2+24] <- round(standardizedSolution(sem_b6_sem)[i1,4],digits=3)
      EFFEKTS[i1,i2+28] <- round(standardizedSolution(sem_cov1_MTMM)[i1,4],digits=3)
      EFFEKTS[i1,i2+32] <- round(standardizedSolution(sem_cov2_MTMM)[i1,4],digits=3)
      EFFEKTS[i1,i2+36] <- round(standardizedSolution(sem_cov3_MTMM)[i1,4],digits=3)
    }
    if(i2 == 3 && i1 == 7){
      EFFEKTS[i1,i2] <- NA
      EFFEKTS[i1,i2+4] <- NA
      EFFEKTS[i1,i2+8] <- NA
      EFFEKTS[i1,i2+12] <- NA
      EFFEKTS[i1,i2+16] <- NA
      EFFEKTS[i1,i2+20] <- NA
      EFFEKTS[i1,i2+24] <- NA
      EFFEKTS[i1,i2+28] <- NA
      EFFEKTS[i1,i2+32] <- NA
      EFFEKTS[i1,i2+36] <- NA
    }
  }
}

for(i1 in 1:7){
  for(i2 in 1:4){
    SE[i1,i2] <- round(parameterestimates(sem_se1_sem)[i1,i2+3],digits=3)
    SE[i1,i2+4] <- round(parameterestimates(sem_se4_sem)[i1,i2+3],digits=3)
    SE[i1,i2+8] <- round(parameterestimates(sem_se5_sem)[i1,i2+3],digits=3)
    SE[i1,i2+12] <- round(parameterestimates(sem_se7_sem)[i1,i2+3],digits=3)
    SE[i1,i2+16] <- round(parameterestimates(sem_se2_sem)[i1,i2+3],digits=3)
    SE[i1,i2+20] <- round(parameterestimates(sem_se3_sem)[i1,i2+3],digits=3)
    SE[i1,i2+24] <- round(parameterestimates(sem_se6_sem)[i1,i2+3],digits=3)
    SE[i1,i2+28] <- round(parameterestimates(sem_secov1_MTMM)[i1,i2+3],digits=3)
    SE[i1,i2+32] <- round(parameterestimates(sem_secov2_MTMM)[i1,i2+3],digits=3)
    SE[i1,i2+36] <- round(parameterestimates(sem_secov3_MTMM)[i1,i2+3],digits=3)
    if(i1 == 7){
      SE[i1,i2] <- round(parameterestimates(sem_se1_sem)[29,i2+3],digits=3)
```

```
      SE[i1,i2+4] <- round(parameterestimates(sem_se4_sem)[29,i2+3],digits=3)
      SE[i1,i2+8] <- round(parameterestimates(sem_se5_sem)[29,i2+3],digits=3)
      SE[i1,i2+12] <- round(parameterestimates(sem_se7_sem)[29,i2+3],digits=3)
      SE[i1,i2+16] <- round(parameterestimates(sem_se2_sem)[29,i2+3],digits=3)
      SE[i1,i2+20] <- round(parameterestimates(sem_se3_sem)[29,i2+3],digits=3)
      SE[i1,i2+24] <- round(parameterestimates(sem_se6_sem)[29,i2+3],digits=3)
      SE[i1,i2+28] <- round(parameterestimates(sem_secov1_MTMM)[29,i2+3],digits=3)
      SE[i1,i2+32] <- round(parameterestimates(sem_secov2_MTMM)[29,i2+3],digits=3)
      SE[i1,i2+36] <- round(parameterestimates(sem_secov3_MTMM)[29,i2+3],digits=3)
    }
    if(i2 == 3){
      SE[i1,i2] <- round(standardizedSolution(sem_se1_sem)[i1,4],digits=3)
      SE[i1,i2+4] <- round(standardizedSolution(sem_se4_sem)[i1,4],digits=3)
      SE[i1,i2+8] <- round(standardizedSolution(sem_se5_sem)[i1,4],digits=3)
      SE[i1,i2+12] <- round(standardizedSolution(sem_se7_sem)[i1,4],digits=3)
      SE[i1,i2+16] <- round(standardizedSolution(sem_se2_sem)[i1,4],digits=3)
      SE[i1,i2+20] <- round(standardizedSolution(sem_se3_sem)[i1,4],digits=3)
      SE[i1,i2+24] <- round(standardizedSolution(sem_se6_sem)[i1,4],digits=3)
      SE[i1,i2+28] <- round(standardizedSolution(sem_secov1_MTMM)[i1,4],digits=3)
      SE[i1,i2+32] <- round(standardizedSolution(sem_secov2_MTMM)[i1,4],digits=3)
      SE[i1,i2+36] <- round(standardizedSolution(sem_secov3_MTMM)[i1,4],digits=3)
    }
    if(i2 == 3 && i1 == 7){
      SE[i1,i2] <- NA
      SE[i1,i2+4] <- NA
      SE[i1,i2+8] <- NA
      SE[i1,i2+12] <- NA
      SE[i1,i2+16] <- NA
      SE[i1,i2+20] <- NA
      SE[i1,i2+24] <- NA
      SE[i1,i2+28] <- NA
      SE[i1,i2+32] <- NA
      SE[i1,i2+36] <- NA
    }
  }
}

write.table(FAKTOR,"faktor_out.txt",sep = "\t",row.names = TRUE, col.names = TRUE)
write.table(EFFEKTS,"effekt_out.txt",sep = "\t",row.names = TRUE, col.names = TRUE)
write.table(SE,"se_out.txt",sep = "\t",row.names = TRUE, col.names = TRUE)
```

7.2 R Syntax der angepassten Benchmark Analyse

```
###-----------------------------------------------------------------###
###--- Schritt 1 -------------------------------------------------###
###--- Lade benötigte Packages, setze Seed ---------------------------###
###-----------------------------------------------------------------###
library(lavaan)
```

```
library(Rmisc)

set.seed(1234)

###-----------------------------------------------------------------###
###--- Schritt 2 -----------------------------------------------------###
###--- Definiere Ergebnistabellen und Parameter ----------------------###
###-----------------------------------------------------------------###
j <- 1000 #Anzahl an Iterationen
n <- 1500 #Fallzahl
cv <- 0.6 #Anteil common variance per Item (= unstand. Lambda)
sm <- 0.5 #Anteil der systematischen Fehlervarianz an der gesamten Fehlervarianz

FAKTOR <- matrix(ncol = 20, nrow = j) #Monte Carlo Faktor Ladungen
colnames(FAKTOR) <- c("f1g1CFA","f2g1CFA","f3g1CFA","f4g1CFA",
                      "f1g2CFA","f2g2CFA","f3g2CFA","f4g2CFA",
                      "f1g3CFA","f2g3CFA","f3g3CFA","f4g3CFA",
                      "f1SEM","f2SEM","f3SEM","f4SEM",
                      "f1MTMM","f2MTMM","f3MTMM","f4MTMM")

EFFEKTE <- matrix(ncol = 24, nrow = j) #Monte Carlo Effekte
colnames(EFFEKTE) <- c("b1g1","b2g1","b3g1","b4g1","b5g1","b6g1","b7g1",
                       "b1g2","b2g2","b3g2","b4g2","b5g2","b6g2","b7g2",
                       "b1g3","b2g3","b3g3","b4g3","b5g3","b6g3","b7g3",
                       "covg1","covg2","covg3")

z_EFFEKTE <- matrix(ncol = 24, nrow = j) #Monte Carlo z-Wert der Effekte
colnames(z_EFFEKTE) <- c("b1g1","b2g1","b3g1","b4g1","b5g1","b6g1","b7g1",
                         "b1g2","b2g2","b3g2","b4g2","b5g2","b6g2","b7g2",
                         "b1g3","b2g3","b3g3","b4g3","b5g3","b6g3","b7g3",
                         "covg1","covg2","covg3")

###-----------------------------------------------------------------###
###--- Schritt 3 -----------------------------------------------------###
###--- Öffne Monte-Carlo Loop und Error Handler ----------------------###
###-----------------------------------------------------------------###
for(i1 in 1:j){
  tryCatch({

    ###-----------------------------------------------------------------###
    ###--- Schritt 4 -----------------------------------------------------###
    ###--- Berechne Indikatorvariablen je Gruppe und erstelle Datensatz --###
    ###-----------------------------------------------------------------###
    eta1 <- rnorm(n,0,1)
    eta2 <- rnorm(n,0,1)
    eta3 <- rnorm(n,0,1)

    xi1 <- rnorm(n,0,1)
    xi2 <- rnorm(n,0,1)

    for(i2 in 1:21){
```

```
  nam_e <- paste("e", i2, sep = "")
  assign(nam_e, rnorm(n,0,1))
}

y1 <- cv*eta1 + (1-cv)*e1
y2 <- cv*eta1 + (1-cv)*(sm*(-0.5*y1*xi1 - 0.5*y1*xi2) + (1-sm)*e2)
y3 <- cv*eta1 + (1-cv)*(sm*(-0.5*y2*xi1 - 0.5*y2*xi2) + (1-sm)*e3)
y4 <- cv*eta1 + (1-cv)*(sm*(-0.5*y3*xi1 - 0.5*y3*xi2) + (1-sm)*e4)

y6 <- cv*eta2 + (1-cv)*e6
y7 <- cv*eta2 + (1-cv)*(sm*(y6*xi1*xi2) + (1-sm)*e7)
y8 <- cv*eta2 + (1-cv)*(sm*(y7*xi1*xi2) + (1-sm)*e8)
y9 <- cv*eta2 + (1-cv)*(sm*(y8*xi1*xi2) + (1-sm)*e9)

y11 <- cv*eta3 + (1-cv)*e11
y12 <- cv*eta3 + (1-cv)*e12
y13 <- cv*eta3 + (1-cv)*e13
y14 <- cv*eta3 + (1-cv)*e14

x1 <- 0.8*xi1 + 0.2*e16
x2 <- 0.8*xi1 + 0.2*e17
x3 <- 0.8*xi1 + 0.2*e18

x4 <- 0.8*xi2 + 0.2*e19
x5 <- 0.8*xi2 + 0.2*e20
x6 <- 0.8*xi2 + 0.2*e21

gr1 <- 1
gr2 <- 2
gr3 <- 3

XY1 <- data.frame(x1,x2,x3,x4,x5,x6,y1,y2,y3,y4,gr1)
colnames(XY1) <- c("x1","x2","x3","x4","x5","x6","y1","y2","y3","y4","Gruppe")
XY2 <- data.frame(x1,x2,x3,x4,x5,x6,y6,y7,y8,y9,gr2)
colnames(XY2) <- c("x1","x2","x3","x4","x5","x6","y1","y2","y3","y4","Gruppe")
XY3 <- data.frame(x1,x2,x3,x4,x5,x6,y11,y12,y13,y14,gr3)
colnames(XY3) <- c("x1","x2","x3","x4","x5","x6","y1","y2","y3","y4","Gruppe")
XY <- rbind(XY1,XY2,XY3)

###-------------------------------------------------------------------###
###--- Schritt 5 -----------------------------------------------------###
###--- Berechne Interaktions-Indikatoren (unconstr. approach) --------###
###-------------------------------------------------------------------###
#xi1*xi2
XY$x1x4 <- XY$x1*XY$x4
XY$x1x5 <- XY$x1*XY$x5
XY$x1x6 <- XY$x1*XY$x6
XY$x2x4 <- XY$x2*XY$x4
XY$x2x5 <- XY$x2*XY$x5
XY$x2x6 <- XY$x2*XY$x6
XY$x3x4 <- XY$x3*XY$x4
```

```
XY$x3x5 <- XY$x3*XY$x5
XY$x3x6 <- XY$x3*XY$x6

#y*xi1
XY$y1x1 <- XY$y1*XY$x1
XY$y1x2 <- XY$y1*XY$x2
XY$y1x3 <- XY$y1*XY$x3
XY$y2x1 <- XY$y2*XY$x1
XY$y2x2 <- XY$y2*XY$x2
XY$y2x3 <- XY$y2*XY$x3
XY$y3x1 <- XY$y3*XY$x1
XY$y3x2 <- XY$y3*XY$x2
XY$y3x3 <- XY$y3*XY$x3

#y*xi2
XY$y1x4 <- XY$y1*XY$x4
XY$y1x5 <- XY$y1*XY$x5
XY$y1x6 <- XY$y1*XY$x6
XY$y2x4 <- XY$y2*XY$x4
XY$y2x5 <- XY$y2*XY$x5
XY$y2x6 <- XY$y2*XY$x6
XY$y3x4 <- XY$y3*XY$x4
XY$y3x5 <- XY$y3*XY$x5
XY$y3x6 <- XY$y3*XY$x6

#y*xi1*xi2
XY$y1x1x4 <- XY$y1*XY$x1*XY$x4
XY$y1x1x5 <- XY$y1*XY$x1*XY$x5
XY$y1x1x6 <- XY$y1*XY$x1*XY$x6
XY$y1x2x4 <- XY$y1*XY$x2*XY$x4
XY$y1x2x5 <- XY$y1*XY$x2*XY$x5
XY$y1x2x6 <- XY$y1*XY$x2*XY$x6
XY$y1x3x4 <- XY$y1*XY$x3*XY$x4
XY$y1x3x5 <- XY$y1*XY$x3*XY$x5
XY$y1x3x6 <- XY$y1*XY$x3*XY$x6
XY$y2x1x4 <- XY$y2*XY$x1*XY$x4
XY$y2x1x5 <- XY$y2*XY$x1*XY$x5
XY$y2x1x6 <- XY$y2*XY$x1*XY$x6
XY$y2x2x4 <- XY$y2*XY$x2*XY$x4
XY$y2x2x5 <- XY$y2*XY$x2*XY$x5
XY$y2x2x6 <- XY$y2*XY$x2*XY$x6
XY$y2x3x4 <- XY$y2*XY$x3*XY$x4
XY$y2x3x5 <- XY$y2*XY$x3*XY$x5
XY$y2x3x6 <- XY$y2*XY$x3*XY$x6
XY$y3x1x4 <- XY$y3*XY$x1*XY$x4
XY$y3x1x5 <- XY$y3*XY$x1*XY$x5
XY$y3x1x6 <- XY$y3*XY$x1*XY$x6
XY$y3x2x4 <- XY$y3*XY$x2*XY$x4
XY$y3x2x5 <- XY$y3*XY$x2*XY$x5
XY$y3x2x6 <- XY$y3*XY$x2*XY$x6
XY$y3x3x4 <- XY$y3*XY$x3*XY$x4
```

```
XY$y3x3x5 <- XY$y3*XY$x3*XY$x5
XY$y3x3x6 <- XY$y3*XY$x3*XY$x6

###-------------------------------------------------------------------###
###--- Schritt 6 -----------------------------------------------------###
###--- CFA Basismodell -----------------------------------------------###
###-------------------------------------------------------------------###
CFA <- 'eta_est =~ NA*y1 + y2 + y3 + y4
      eta_est ~~ 1*eta_est'
fit_cfa <- cfa(CFA,XY,group="Gruppe")
results_cfa_ust <- parameterEstimates(fit_cfa)

for(i in 1:4){
  FAKTOR[i1,i] <- results_cfa_ust[i,6]
  FAKTOR[i1,i+4] <- results_cfa_ust[14+i,6]
  FAKTOR[i1,i+8] <- results_cfa_ust[28+i,6]
}

###-------------------------------------------------------------------###
###--- Schritt 7 -----------------------------------------------------###
###--- NLC-CFA Modell ------------------------------------------------###
###-------------------------------------------------------------------###
SEM <- 'eta_est =~ NA*y1 + y2 + y3 + y4
      xi1_est =~ NA*x1 + f2*x1 + f2*x2 + f2*x3
      xi2_est =~ NA*x4 + f2*x4 + f2*x5 + f2*x6

      eta_est ~~ 1*eta_est
      xi1_est ~~ 1*xi1_est
      xi2_est ~~ 1*xi2_est

      eta_est ~~ 0*xi1_est
      eta_est ~~ 0*xi2_est
      xi1_est ~~ 0*xi2_est

      y1xi1_est =~ NA*y1x1 + f3*y1x1 + f3*y1x2 + f3*y1x3
      y2xi1_est =~ NA*y2x1 + f3*y2x1 + f3*y2x2 + f3*y2x3
      y3xi1_est =~ NA*y3x1 + f3*y3x1 + f3*y3x2 + f3*y3x3

      y1xi1_est ~~ 1*y1xi1_est
      y2xi1_est ~~ 1*y2xi1_est
      y3xi1_est ~~ 1*y3xi1_est

      y1xi2_est =~ NA*y1x4 + f3*y1x4 + f3*y1x5 + f3*y1x6
      y2xi2_est =~ NA*y2x4 + f3*y2x4 + f3*y2x5 + f3*y2x6
      y3xi2_est =~ NA*y3x4 + f3*y3x4 + f3*y3x5 + f3*y3x6

      y1xi2_est ~~ 1*y1xi2_est
      y2xi2_est ~~ 1*y2xi2_est
      y3xi2_est ~~ 1*y3xi2_est
```

```
          xi1xi2_est =~ NA*x1x4 + f4*x1x4 + f4*x1x5 + f4*x1x6 + f4*x2x4 + f4*x2x5 + f4*x2x6 +
f4*x3x4 + f4*x3x5 + f4*x3x6


          xi1xi2_est ~~ 1*xi1xi2_est


          y1xi1xi2_est =~ NA*y1x1x4 + f5*y1x1x4 + f5*y1x1x5 + f5*y1x1x6 + f5*y1x2x4 +
f5*y1x2x5 + f5*y1x2x6 + f5*y1x3x4 + f5*y1x3x5 + f5*y1x3x6
          y2xi1xi2_est =~ NA*y2x1x4 + f5*y2x1x4 + f5*y2x1x5 + f5*y2x1x6 + f5*y2x2x4 +
f5*y2x2x5 + f5*y2x2x6 + f5*y2x3x4 + f5*y2x3x5 + f5*y2x3x6
          y3xi1xi2_est =~ NA*y3x1x4 + f5*y3x1x4 + f5*y3x1x5 + f5*y3x1x6 + f5*y3x2x4 +
f5*y3x2x5 + f5*y3x2x6 + f5*y3x3x4 + f5*y3x3x5 + f5*y3x3x6


          y1xi1xi2_est ~~ 1*y1xi1xi2_est
          y2xi1xi2_est ~~ 1*y2xi1xi2_est
          y3xi1xi2_est ~~ 1*y3xi1xi2_est


          y2 ~ c(b1,b2,b3)*y1 + c(b4,b5,b6)*xi1_est + c(b7,b8,b9)*xi2_est +
c(b10,b11,b12)*y1xi1_est + c(b13,b14,b15)*y1xi2_est + c(b16,b17,b18)*xi1xi2_est +
c(b19,b20,b21)*y1xi1xi2_est
          y3 ~ c(b1,b2,b3)*y2 + c(b4,b5,b6)*xi1_est + c(b7,b8,b9)*xi2_est +
c(b10,b11,b12)*y2xi1_est + c(b13,b14,b15)*y2xi2_est + c(b16,b17,b18)*xi1xi2_est +
c(b19,b20,b21)*y2xi1xi2_est
          y4 ~ c(b1,b2,b3)*y3 + c(b4,b5,b6)*xi1_est + c(b7,b8,b9)*xi2_est +
c(b10,b11,b12)*y3xi1_est + c(b13,b14,b15)*y3xi2_est + c(b16,b17,b18)*xi1xi2_est +
c(b19,b20,b21)*y3xi1xi2_est'
    fit_sem <- sem(SEM,XY,group="Gruppe",group.equal = c("loadings"))
    results_sem_k4 <- parameterEstimates(fit_sem)


    for(i in 1:4){
      FAKTOR[i1,i+12] <- results_sem_k4[i,7]
    }


    for(i in 1:7){
      EFFEKTE[i1,i] <- results_sem_k4[i+80,7]
      EFFEKTE[i1,i+7] <- results_sem_k4[i+397,7]
      EFFEKTE[i1,i+14] <- results_sem_k4[i+714,7]
    }


    for(i in 1:7){
      z_EFFEKTE[i1,i] <- results_sem_k4[i+80,9]
      z_EFFEKTE[i1,i+7] <- results_sem_k4[i+397,9]
      z_EFFEKTE[i1,i+14] <- results_sem_k4[i+714,9]
    }


    ###-------------------------------------------------------------------###
    ###--- Schritt 8 -----------------------------------------------------###
    ###--- CFA-MTMM Modell -----------------------------------------------###
    ###-------------------------------------------------------------------###
    MTMM <- 'eta_est =~ NA*y1 + y2 + y3 + y4
            eta_est ~~ 1*eta_est


            y1 ~~ c(cov1,cov2,cov3)*y2
            y2 ~~ c(cov1,cov2,cov3)*y3
            y3 ~~ c(cov1,cov2,cov3)*y4'
    fit_MTMM <- sem(MTMM,XY,group="Gruppe",group.equal = c("loadings"))
```

```
    results_MTMM_ust <- parameterEstimates(fit_MTMM)

    for(i in 1:4){
      FAKTOR[i1,i+16] <- results_MTMM_ust[i,7]
    }

    EFFEKTE[i1,c(22,23,24)] <- results_MTMM_ust[c(6,23,40),7]
    z_EFFEKTE[i1,c(22,23,24)] <- results_MTMM_ust[c(6,23,40),9]

###-----------------------------------------------------------------###
###--- Schritt 8 -------------------------------------------------###
###--- Schließe Monte-Carlo Loop und Error Handler ------------------###
###-----------------------------------------------------------------###
    print(paste0("Abgeschlossene Iterationen: ", i1))
  }, error=function(e){})
}

###-----------------------------------------------------------------###
###--- Schritt 9 -------------------------------------------------###
###--- Aggregation -----------------------------------------------###
###-----------------------------------------------------------------###
F_ERGEBNISSE <- matrix(ncol = 5, nrow = 8) #Faktorladungen aggregierte Ergebnisse
colnames(F_ERGEBNISSE) <- c("CFA g1", "CFA g2","CFA g3","NLC","MTMM")
rownames(F_ERGEBNISSE) <- c("f1","CI95 f1","f2","CI95 f2","f3","CI95 f3","f4","CI95 f4")

E_ERGEBNISSE <- matrix(ncol = 3, nrow = 16) #Effekte aggregierte Ergebnisse
colnames(E_ERGEBNISSE) <- c("g1","g2","g3")
rownames(E_ERGEBNISSE) <- c("b1","CI95 b1","b2","CI95 b2","b3","CI95 b3","b4","CI95
b4","b5","CI95 b5","b6","CI95 b6","b7","CI95 b7",
                            "Cov","CI95 Cov")

Z_ERGEBNISSE <- matrix(ncol = 3, nrow = 16) #z-Werte der Effekte aggregierte Ergebnisse
colnames(Z_ERGEBNISSE) <- c("g1","g2","g3")
rownames(Z_ERGEBNISSE) <- c("b1","CI95 b1","b2","CI95 b2","b3","CI95 b3","b4","CI95
b4","b5","CI95 b5","b6","CI95 b6","b7","CI95 b7",
                            "Cov","CI95 Cov")

for(i in 1:4){
  F_ERGEBNISSE[(2*i)-1,1] <- round(mean(FAKTOR[,i]),digits=3)
  F_ERGEBNISSE[(2*i)-1,2] <- round(mean(FAKTOR[,i+4]),digits=3)
  F_ERGEBNISSE[(2*i)-1,3] <- round(mean(FAKTOR[,i+8]),digits=3)
  F_ERGEBNISSE[(2*i)-1,4] <- round(mean(FAKTOR[,i+12]),digits=3)
  F_ERGEBNISSE[(2*i)-1,5] <- round(mean(FAKTOR[,i+16]),digits=3)

  F_ERGEBNISSE[(2*i),1] <- paste0(round(CI(FAKTOR[,i],ci=0.95)[3],digits=3)," -
",round(CI(FAKTOR[,i],ci=0.95)[1],digits=3))
  F_ERGEBNISSE[(2*i),2] <- paste0(round(CI(FAKTOR[,i+4],ci=0.95)[3],digits=3)," -
",round(CI(FAKTOR[,i+4],ci=0.95)[1],digits=3))
  F_ERGEBNISSE[(2*i),3] <- paste0(round(CI(FAKTOR[,i+8],ci=0.95)[3],digits=3)," -
",round(CI(FAKTOR[,i+8],ci=0.95)[1],digits=3))
  F_ERGEBNISSE[(2*i),4] <- paste0(round(CI(FAKTOR[,i+12],ci=0.95)[3],digits=3)," -
",round(CI(FAKTOR[,i+12],ci=0.95)[1],digits=3))
  F_ERGEBNISSE[(2*i),5] <- paste0(round(CI(FAKTOR[,i+16],ci=0.95)[3],digits=3)," -
",round(CI(FAKTOR[,i+16],ci=0.95)[1],digits=3))
```

```
}


for(i in 1:7){
  E_ERGEBNISSE[(2*i)-1,1] <- round(mean(EFFEKTE[,i]),digits=3)
  E_ERGEBNISSE[(2*i)-1,2] <- round(mean(EFFEKTE[,i+7]),digits=3)
  E_ERGEBNISSE[(2*i)-1,3] <- round(mean(EFFEKTE[,i+14]),digits=3)


  E_ERGEBNISSE[(2*i),1] <- paste0(round(CI(EFFEKTE[,i],ci=0.95)[3],digits=3)," -
",round(CI(EFFEKTE[,i],ci=0.95)[1],digits=3))
  E_ERGEBNISSE[(2*i),2] <- paste0(round(CI(EFFEKTE[,i+7],ci=0.95)[3],digits=3)," -
",round(CI(EFFEKTE[,i+7],ci=0.95)[1],digits=3))
  E_ERGEBNISSE[(2*i),3] <- paste0(round(CI(EFFEKTE[,i+14],ci=0.95)[3],digits=3)," -
",round(CI(EFFEKTE[,i+14],ci=0.95)[1],digits=3))


  Z_ERGEBNISSE[(2*i)-1,1] <- round(mean(z_EFFEKTE[,i]),digits=3)
  Z_ERGEBNISSE[(2*i)-1,2] <- round(mean(z_EFFEKTE[,i+7]),digits=3)
  Z_ERGEBNISSE[(2*i)-1,3] <- round(mean(z_EFFEKTE[,i+14]),digits=3)


  Z_ERGEBNISSE[(2*i),1] <- paste0(round(CI(z_EFFEKTE[,i],ci=0.95)[3],digits=3)," -
",round(CI(z_EFFEKTE[,i],ci=0.95)[1],digits=3))
  Z_ERGEBNISSE[(2*i),2] <- paste0(round(CI(z_EFFEKTE[,i+7],ci=0.95)[3],digits=3)," -
",round(CI(z_EFFEKTE[,i+7],ci=0.95)[1],digits=3))
  Z_ERGEBNISSE[(2*i),3] <- paste0(round(CI(z_EFFEKTE[,i+14],ci=0.95)[3],digits=3)," -
",round(CI(z_EFFEKTE[,i+14],ci=0.95)[1],digits=3))
}


for(i in 1:3){
  E_ERGEBNISSE[15,i] <- round(mean(EFFEKTE[,21+i]),digits=3)
  E_ERGEBNISSE[16,i] <- paste0(round(CI(EFFEKTE[,21+i],ci=0.95)[3],digits=3)," -
",round(CI(EFFEKTE[,21+i],ci=0.95)[1],digits=3))
  Z_ERGEBNISSE[15,i] <- round(mean(z_EFFEKTE[,21+i]),digits=3)
  Z_ERGEBNISSE[16,i] <- paste0(round(CI(z_EFFEKTE[,21+i],ci=0.95)[3],digits=3)," -
",round(CI(z_EFFEKTE[,21+i],ci=0.95)[1],digits=3))
}


write.table(F_ERGEBNISSE,"faktorBA.txt",sep = "\t",row.names = TRUE, col.names = TRUE)
write.table(E_ERGEBNISSE,"effektBA.txt",sep = "\t",row.names = TRUE, col.names = TRUE)
write.table(Z_ERGEBNISSE,"zBA.txt",sep = "\t",row.names = TRUE, col.names = TRUE
```

Eidesstattliche Erklärungen

Hiermit erkläre Ich, Christoph Giehl, geboren am 20.12.1988 in Rockenhausen, dass Ich die vorliegende Dissertation selbstständig und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt habe. Mit einer elektronischen Überprüfung dieser Dissertation auf etwaige Plagiate bin ich einverstanden.

Weiterhin erkläre Ich, dass diese Doktorarbeit als Monographie verfasst wurde. Die Arbeit wurde von Prof. Dr. Jochen Mayerl, Institut für Soziologie, betreut.

Zuletzt erkläre Ich, dass Ich mit dieser Doktorarbeit in einer früheren Form bereits einen Antrag auf Zulassung zur Promotion an der Technischen Universität Kaiserslautern gestellt und bewilligt bekommen habe. Aufgrund des Wechsels meines Betreuers Prof. Dr. Jochen Mayerl von der TU Kaiserslautern zur TU Chemnitz innerhalb des Anfertigungszeitraums dieser Doktorarbeit habe Ich das Promotionsvorhaben an der TU Kaiserslautern formal abgemeldet und unter Einhaltung der Promotionsordnung in ihrer Fassung vom 20. Mai 2022 an der TU Chemnitz angemeldet.

Ort, Datum	Unterschrift

Wissenschaftlicher Lebenslauf und Veröffentlichungen

Allgemeine Angaben

Name: Christoph Giehl

Staatsangehörigkeit: Deutsch

Geburtsdatum: 20.12.1988

Geburtsort: Rockenhausen

Ausbildung und Weiterbildungen

2020	Teilnahme an der GESIS Summer School in Survey Methodology, Kurs: „Mathematical Tools for Social Scientists: A Refresher Course with R"
2018	Teilnahme am GESIS Frühjahrsworkshop, Kurs: "Einführung in die Logik der Bayesschen Statistik"
2017	Teilnahme an der GESIS Summer School in Survey Methodology, Kurs: "Structural Equation Modeling with Panel Data"
2016	Teilnahme am GESIS Frühjahrsworkshop, Kurs: "Mehrebenenanalyse mit HLM"
2015	Teilnahme an der GESIS Summer School in Survey Methodology, Kurs: "Introduction to Structural Equation Modeling Framework: Confirmatory Factor Analysis with Mplus"
2014 -	Promotion an der Technischen Universität Kaiserslautern, Fachbereich Sozialwissenschaften. Thema: „Identifikation und Kontrolle von Fragereihenfolgeneffekten. Eine simultane Analyse multipler, latenter Interaktionen auf Messfehlerebene mit Strukturgleichungsmodellen."
2012 - 2014	Abschluss in integrative Sozialwissenschaften MA an der Technischen Universität Kaiserslautern
2009 - 2012	Abschluss in integrative Sozialwissenschaften BA an der Technischen Universität Kaiserslautern
1999 - 2008	Abitur an der Integrierten Gesamtschule Rockenhausen

Projektmitarbeit und Berufserfahrung

2020 – 2023	Projektmitarbeiter im Projekt „A-DigiKomp“ am Fachgebiet für Psychologie II und dem Center for Cognitive Science der TU Kaiserslautern
2020 – 2023	Projektmitarbeiter im Projekt „Equal-Life“ am Fachgebiet für Psychologie II und dem Center for Cognitive Science der TU Kaiserslautern
2019 - 2020	Projektmitarbeiter im Projekt „ConWearDi“ am Fachgebiet für Psychologie II und dem Center for Cognitive Science der TU Kaiserslautern
2017 - 2019	Projektmitarbeiter im Projekt „ConWearDi“ am Fachgebiet für Empirische Sozialforschung der TU Kaiserslautern in Assoziation mit dem DFKI und dem Center for Cognitive Science der TU Kaiserslautern
2016 - 2020	Mitarbeiter und Dozent am Fachgebiet für Stadtsoziologie der TU Kaiserslautern, Schwerpunkt: Statistische Datenanalyse
2016 - 2017	Projektmitarbeiter im Projekt „Lessons Learned? Grundlegende Erkenntnisse aus der Kommunikation der Rückverlegung aus einem Einsatzgebiet am Beispiel der Rückverlegung des DEU EinsKtgt ISAF aus Afghanistan 2014“ am Fachgebiet für Empirische Sozialforschung der TU Kaiserslautern in Assoziation mit der Universität der Bundeswehr München
2014 - 2017	Projektmitarbeiter im Projekt „Lehrevaluation Plus“ am Fachgebiet für Empirische Sozialforschung der TU Kaiserslautern in Assoziation mit dem Referat für Qualität in Studium und Lehre der TU Kaiserslautern
2014	Projektmitarbeit im Distance and Independent Studies Center der TU Kaiserslautern, insbesondere empirische Aufbereitung und Analyse quantitativer Daten
2013 - 2014	Wissenschaftliche Hilfskraft im Fachgebiet Pädagogik der TU Kaiserslautern, Verantwortungsbereich IT und Onlinebefragungsinstrument
2013 - 2014	Wissenschaftliche Hilfskraft im Fachgebiet für Empirische Sozialforschung der TU Kaiserslautern
2013	Projektmitarbeiter am Lehrstuhl für Erwachsenenbildung der TU Kaiserslautern
2012 - 2014	Leitung über das Projekt "keiner darf verloren gehen" bei der Förderinitiative Donnersberg
2011 -	Vorstandsmitglied der Förderinitiative Donnersberg

2011 Praktikum bei der Förderinitiative Donnersberg im Bereich Erwachsenenbildung und Qualitätsmanagement (8 Wochen)

2008 - 2009 Zivildienst in der Pfalzklinik für Psychotherapie Rockenhausen

Lehrtätigkeit

2018 - 2020 Dozent für Einführung in die deduktive und induktive Statistik am Fachgebiet für Stadtsoziologie der TU Kaiserslautern

2017 - 2020 Dozent für Umweltsoziologie am Fachgebiet für Stadtsoziologie der TU Kaiserslautern

2016 - 2018 Dozententätigkeit an der Universität Mannheim, Kurs: „Datenanalyse mit SPSS“

2014 - 2018 Dozent für Methoden der empirischen Sozialforschung am Fachgebiet für Empirische Sozialforschung der TU Kaiserslautern

2014 Tutor für Fortgeschrittene Techniken der Datenanalyse im Fachgebiet für Empirische Sozialforschung an der TU Kaiserslautern

2014 Tutor für Methoden der empirischen Sozialforschung im Fachgebiet für Empirische Sozialforschung an der TU Kaiserslautern

2011 - 2016 Dozent im Bereich Erwachsenenbildung, insbesondere Kommunikationstraining bei der Förderinitiative Donnersberg

Beiträge auf Konferenzen und Präsentationen

2019 Teilnahme an der ESRA Conference in Zagreb, Kroatien. Thema „ Identifying and controlling for systematic measurement errors using LMS: Findings from a simulation study“, Session „Quality of Survey Data“

2018 Teilnahme an der ISA Conference in Toronto, Kanada. Thema: „Explaining response styles and response bias using the attitude towards surveys and the attitude accessibility towards the research topic“, Session „Cultural Response Styles“

2018 Teilnahme an der ISA Conference in Toronto, Kanada. Thema: „Replicating Findings Regarding Attitudes Towards the International Security Assistance

Force (ISAF) in Afghanistan in Five European Countries“, Session „Replicability in the Social Sciences: Extent, Reasons and Consequences“

2017 Teilnahme an der ESRA Conference in Lissabon, Portugal. Thema „Explaining systematic measurement errors using cognitive-process models of response behavior and the attitude towards surveys “, Session „Cognition in Surveys“

2016 Teilnahme an der “RC33 Conference 2016, the 9th International Conference on Social Science Methodology” der ISA (International Sociology Association” in Leicester, UK. Präsentation zum Thema “The attitude towards surveys as a predictor of response behavior”, Session “Assessing the Quality of Survey Data”

2015 Teilnahme an der ESRA Conference in Reykjavic, Island. Thema „Attitude towards surveys as a general predictor of data quality in dual-process models of response behavior“, Session „Cognition in Surveys“

Wissenschaftlicher Beirat

2018 Wissenschaftlicher Beirat (in Vertretung) für die Studie „Umweltbewusstsein in Deutschland 2018“ (Bundesministerium für Umwelt, Naturschutz und nukleare Sicherheit; Umweltbundesamt; Institut für ökologische Wirtschaftsforschung)

Publikationen

Spilski, J., Exner, J.P., Schmidt, M., Schlittmeier, S., Giehl, C., Lachmann, T., Pietschmann, M., & Werth, D. (2019). Potential of VR in the vocational education and training of craftsmen. In: Proceedings of the 19th International Conference on Construction Applications of Virtual Reality. 13-15 November 2019, Bangkok, Thailand.

Berg, M., Giehl, C., Koch, M., Memmel, M., Spellerberg, A., & Walter, R. (2019). Gutachterliche Stellungnahme zu den Auswirkungen künstlicher Systeme im Speziellen und der Digitalisierung im Allgemeinen auf das kommunale Leben in Rheinland-Pfalz 2050. Technische Universität Kaiserslautern, Kaiserslautern.

Mayerl, J., Andersen, H., & Giehl, C. (2019). Identification of MeasurementProblems of Survey Items and Scales using Paradata. In N. Menold und T.Wolbring, *Qualitätssicherung sozialwissenschaftlicher Erhebungsinstrumente* (S. 9-36). Wiesbaden: Springer VS.

Mayerl, J., & Giehl, C. (2018): A Closer Look at Attitude Scales with Positive and Negative Items. Response Latency Perspectives on Measurement Quality. *Survey Research Methods, 12 (3)*, S. 9999 - 10016.

Giehl, C., Bogner, C., Mayerl, J., & Prescher, T. (2018). Entwicklung und Test eines Messinstruments zur überfachlichen Kompetenz. Eine konfirmatorische Faktoranalyse. In T. Prescher, C. Marquardt, & J. Fluck, *Zeitschrift für Empirische Pädagogik, Themenheft: Überfachliche Kompetenzen als kompetenzgenerierende Lösungsmuster, 32(1).*

Spellerberg, A., & Giehl, C. (2018). Armut und Wohnen. In P. Böhnke, J. Dittmann, & J. Goebel, *Handbuch Armut. Ursachen, Trends, Maßnahmen*. Opladen/Toronto: Verlag Barbara Budrich.

Prescher, T., & Giehl, C. (2017). Förderung personaler Kompetenz: körperorientierte Selbsterfahrung als Möglichkeit einer kompetenzorientierten Hochschule. In R. Arnold, M. Lermen, & M. Haberer, *Selbstlernangebote und Studienunterstützung. Band III zur Fachtagung „Selbstgesteuert, kompetenzorientiert und offen?!“.* Baltmannsweiler: Schneider Verlag Hohengehren.

Giehl, C., & Mayerl, J. (2016). Gesundheits- und Umweltbewusstsein als Ausdruck postmaterialistischer Werteorientierung? Empirische Analyse eines Strukturgleichungsmodells mit Daten des SOEP 2006. *Schriftenreihe der Empirischen Sozialforschung Kaiserslautern (1).*

Danksagung

Auch wenn eine Doktorarbeit von einer einzigen Person zu verfassen ist, würden wohl sehr wenige Doktorarbeiten ohne die Unterstützung weiterer Menschen verfasst werden.

So wäre diese Doktorarbeit nicht möglich gewesen ohne die umfassende Unterstützung meines Doktorvaters Prof. Dr. Jochen Mayerl, der mich bereits während meines Studiums als wissenschaftliche Hilfskraft unterstützt und gefördert hat. Ohne die vielen persönlichen Gespräche, in welchen er mir verschiedene statistische Verfahren erklärt und meine Faszination für diese geweckt hat, wäre ich sicher einen anderen Weg gegangen. Hierfür möchte Ich meine tiefe Dankbarkeit ausdrücken.

Mein Dank gebührt ebenfalls meinen Kollegen an der Technischen Universität Kaiserslautern für die vielen angeregten Diskussionen über statistische Verfahren und die wissenschaftliche Methodik, aus denen Ich sehr viel lernen konnte. Besonders hervorheben möchte Ich an dieser Stelle zum einen Jan Spilski für sein unermüdliches Feedback zu meiner Arbeit, welches für mich immer wieder ein Quell für meine weitere Entwicklung darstellt. Zum anderen möchte Ich aber auch gegenüber Prof. Dr. Thomas Lachmann meinen besonderen Dank ausdrücken der mir ermöglicht hat, meine Arbeit mit meinem Forschungsschwerpunkt in seinem Fachgebiet weiterzuführen, nachdem Prof. Mayerl dem Ruf aus Chemnitz gefolgt ist und Kaiserslautern verlassen hat.

Abschließend gilt mein Dank meiner Familie, die schon seit jeher meine Wissbegierde unterstützt und mich in anstrengenden Phasen in Liebe erträgt. Mama, Papa, Sabrina, Andreas, Peter, Heike und Thomas, danke, dass ihr da wart und da seid. Sarah, danke für einfach ALLES.