

# Methodologische und empirische Studien zur Validierung von Achtsamkeitsskalen

Dem Fachbereich Psychologie und Sportwissenschaften der  
Goethe-Universität Frankfurt am Main  
zur Erlangung des akademischen Grades eines  
Dr. phil.

eingereichte Dissertation

von  
Herrn Dipl.-Psych. Volkmar Höfling  
aus  
Gemünden am Main

Datum der Einreichung: 09.05.2010

Dekan: Prof. Dr. Dr. W. Banzer

Erstgutachter: Prof. Dr. H. Moosbrugger

Zweitgutachter: Prof. Dr. U. Stangier

Datum der Disputation: 07.07.2010

# Zusammenfassung

Die vorliegende Arbeit leistet einen methodologischen und empirischen Beitrag zur Validierung zweier bedeutender Achtsamkeitsskalen, den Deutschen Adaptationen der Mindful Attention and Awareness Scale (MAAS; Brown & Ryan, 2003) und des Kentucky Inventory of Mindfulness Skills (KIMS; Baer, Smith & Allen, 2004). Zunächst stehen methodologische Fragen im Zentrum der Betrachtung. Es werden Methodeneffekte, wie sie in psychologischen Fragebogen vorkommen, vorgestellt. Weiterhin wird der spezifische Effekt der Itempolung beleuchtet, indem empirische Untersuchungen zur Analyse von Itempolungseffekten aus dem Bereich der Self-Esteem-Forschung vorgestellt werden. Anschließend werden statistische Verfahren zur Analyse der sogenannten Multitrait-Multimethod (MTMM)-Daten dargestellt. Hierbei wird die traditionelle MTMM-Analyse von Campbell und Fiske (1959) in Bezug zur konfirmatorischen Faktorenanalyse (CFA) gestellt und eine Übersicht über verschiedene Möglichkeiten zur Spezifikation sog. MTMM-CFA-Modelle gegeben.

Die methodologischen Überlegungen dienen als Grundlage für empirische Untersuchungen zur MAAS. Es sollten die Fragen geklärt werden, ob die MAAS auch dann als eindimensional bzw. homogen bezeichnet werden kann, wenn nicht nur die 15 negativ gepolten Items der MAAS, sondern auch positiv gepolte Items zusätzlich zu den negativ gepolten Items eingesetzt werden. Zur Klärung dieser Fragen wurden verschiedene CFA- bzw. MTMM-CFA-Modelle spezifiziert und auf ihre Modellanpassung hin überprüft. Es ergab sich, dass für eine modifizierte 10-Item-Version der MAAS (MAAS-Short) mit negativ und positiv gepolten Items Eindimensionalität und Homogenität nachgewiesen werden kann, wenn der Effekt der Itempolung in Form eines Methodenfaktors Berücksichtigung im Rahmen eines MTMM-CFA-Modells findet. Wichtige Kennwerte der MAAS-Short werden berichtet.

Mit dem KIMS liegt ein vierdimensionaler Achtsamkeitsfragebogen mit 39 Items vor, der sich für die Forschungspraxis als hilfreich, aber zu wenig ökonomisch erweist. Mögliche Homogenitätsmängel im KIMS werden mit verschiedenen CFA untersucht und führen zu einer 20-Item-Version des KIMS (KIMS-Short). Die vierfaktorielle Struktur des KIMS konnte für das KIMS-Short im Wesentlichen bestätigt werden, wobei für die Skala „Beobachten“ keine Eindimensionalität, sondern – theoretisch plausibel – Zweidimensionalität gefunden wurde. Zusammenhänge des KIMS-Short mit der MAAS, mit Persönlichkeitsvariablen und mit Maßen psychischer Belastung belegen die diskriminante und konvergente Validität des KIMS-Short.

Die Ergebnisse aus den methodologischen und empirischen Untersuchungen werden in einer zusammenfassenden Diskussion kritisch gewürdigt und eingeordnet. Mit vorliegender Arbeit werden vielfältige Analysemöglichkeiten von Methodeneffekten aufgezeigt sowie zwei revidierte Versionen von Achtsamkeitsskalen vorgelegt.

*Schlagwörter:* Achtsamkeit, Multitrait-Multimethod-Analysen, Validierung, Kentucky Inventory of Mindfulness Skills, Mindful Attention and Awareness Scale.

# Inhaltsverzeichnis

<b>1</b>	<b>Achtsamkeit (<i>mindfulness</i>) und ihre psychometrische Erfassung</b>	<b>1</b>
1.1	Eine Definition von Achtsamkeit . . . . .	1
1.2	Achtsamkeit im Kontext therapeutischer Interventionen . . . . .	2
1.3	Fragebogen zur Achtsamkeit . . . . .	4
1.4	Ansätze für Revisionsmöglichkeiten in Achtsamkeitsskalen . . . . .	7
1.4.1	Beschränkung auf negativ gepolte Items im Rahmen der Mindful Attention and Awareness Scale (MAAS) notwendig? . . . . .	7
1.4.2	Heterogenität in der Mindful Attention and Awareness Scale (MAAS) – Sind weniger Items ausreichend? . . . . .	8
1.4.3	Heterogenität im Kentucky Inventory of Mindfulness Skills (KIMS) – Chance für die Entwicklung einer Kurzform? . . . . .	9
1.5	Fragestellungen . . . . .	10
<b>2</b>	<b>Methodeneffekte im Kontext psychologischer Datenerhebung und deren Analyse</b>	<b>12</b>
2.1	Methodeneffekte im Rahmen von Selbstbeurteilungsinstrumenten . . . . .	12
2.2	Der spezifische Methodeneffekt der Itempolung im Kontext von Self-Esteem	13
2.3	Die Konzeptualisierung von Methodeneffekten im Kontext der Klassischen Testtheorie (KTT) . . . . .	16
2.4	Analyseverfahren zur Berücksichtigung von Methodeneffekten – Die Multitrait- Multimethod (MTMM)-Analyse . . . . .	16
2.4.1	„Übersetzung“ der vier Kriterien der klassischen MTMM-Analyse in den Bezugsrahmen der konfirmatorischen Faktorenanalyse (CFA) . . . . .	18
2.4.2	MTMM-Modelle im Kontext der CFA – Eine „theoretische“ Systematik zu MTMM-CFA-Modellen . . . . .	20
2.4.3	Zum CT-C( $M-1$ )-Modell und zur Konzeption von Methodenfaktoren	22
<b>3</b>	<b>Untersuchungen und Ergebnisse zur <i>Mindful Attention and Awareness Scale</i> (MAAS)</b>	<b>23</b>
3.1	Stichprobe . . . . .	23

3.2	Itempolung in der MAAS – MTMM-CFA-Modelle . . . . .	24
3.3	Homogenität der Items der MAAS . . . . .	26
3.4	Ergebnisse . . . . .	27
3.4.1	Evaluation der verschiedenen CFA- bzw. MTMM-CFA-Modelle . . .	27
3.4.2	Itemkennwerte der MAAS-Short . . . . .	28
<b>4</b>	<b>Untersuchungen und Ergebnisse zum <i>Kentucky Inventory of Mindfulness Skills</i> (KIMS)</b>	<b>31</b>
4.1	Stichproben . . . . .	31
4.2	Identifikation homogener Items – Die Kurzversion des KIMS (KIMS-Short)	32
4.3	Konvergente und diskriminante Validität des KIMS-Short . . . . .	36
<b>5</b>	<b>Diskussion</b>	<b>39</b>
5.1	Zusammenfassende Beantwortung der Fragestellungen . . . . .	39
5.1.1	Zur Methodologie von MTMM-Analysen . . . . .	39
5.1.2	Eindimensionalität und Homogenität in der MAAS . . . . .	40
5.1.3	Eine Kurzversion der Deutschen Adaptation des KIMS (KIMS-Short)	42
5.2	Grenzen der vorliegenden Arbeit und Resümee . . . . .	43
<b>A</b>	<b>Schrift (A)</b>	<b>52</b>
<b>B</b>	<b>Schrift (B)</b>	<b>66</b>
<b>C</b>	<b>Schrift (C)</b>	<b>87</b>

# Abbildungsverzeichnis

2.1	Schematische Pfaddiagramme sechs alternativer CFA- bzw. MTMM-CFA-Modelle zur Analyse von Itempolungseffekten im Rahmen der RSES (vgl. Horan et al., 2003, S. 442). Modell A und Modell B berücksichtigen keine Methodeneffekte, Modell C und Modell D spezifizieren Methodenfaktoren für die Itempolung, Modell E und Modell F setzen Kovarianzen zwischen den Residualvariablen der vier positiv bzw. drei negativ gepolten Items frei.	15
2.2	Pfaddiagramm eines beispielhaften Strukturmodells zu MTMM-Daten, in welchen die drei Traits Extraversion, Neurotizismus und Gewissenhaftigkeit mit den drei Methoden Selbstbeurteilung, Fremdbeurteilung A und Fremdbeurteilung B erfasst wurden. Jede latente Variable $TM$ repräsentiert eine Trait-Methodeneinheit, die „vertikalen“ Korrelationen stellen Heterotrait-Monomethod- und die „horizontalen“ Korrelationen Monotrait-Heteromethod-Koeffizienten dar. . . . .	19

# Tabellenverzeichnis

1.1	Auflistung von Gütekriterien zu Modellanpassungen aus CFA zur MAAS in verschiedenen Untersuchungen. . . . .	9
2.1	Taxonomie alternativer CFA- bzw. MTMM-CFA-Modelle nach Widaman (1985). In Abhängigkeit von apriorischen Annahmen zu Trait- bzw. Methodeneinflüssen in den vorliegenden MTMM-Daten können verschiedene Modelle spezifiziert werden. . . . .	21
3.1	Prozentuale Häufigkeitsverteilung in den Variablen „Geschlecht“, „Schulabschluss“ und „Meditationserfahrung“ in der Gesamt- und den Teilstichproben.	24
3.2	Übersicht über die spezifizierten CFA- und MTMM-CFA-Modelle auf Basis aller 28 positiv und negativ gepolter Items der MAAS. . . . .	26
3.3	Übersicht über die spezifizierten CFA- und MTMM-CFA-Modelle auf Basis der fünf positiv und negativ gepolten Items der MAAS-Short mit den höchsten Faktorladungen. . . . .	27
3.4	Gütekriterien aus den Analysen zu den auf Basis der MAAS spezifizierten CFA- bzw. MTMM-CFA-Modellen. . . . .	28
3.5	Gütekriterien aus den Analysen zu den auf Basis der MAAS-Short spezifizierten CFA- bzw. MTMM-CFA-Modellen. . . . .	29
3.6	Auflistung der Iteminhalte und Itemkennwerte (Mittelwerte, Trennschärfe-koeffizienten und Varianzen) der MAAS-Short. . . . .	30
4.1	Prozentuale Häufigkeitsverteilung in den Variablen „Geschlecht“, „Schulabschluss“ und „Praxis geistiger Übung“ in Stichprobe B und ihren Teilstichproben. . . . .	32
4.2	Die Faktoren, Iteminhalte und Faktorladungen der Items des KIMS-Short auf Basis der CFA mit allen Items (Stichprobe A). . . . .	34

4.3	Auflistung der Gütekriterien der Modellanpassungen der CFA für die vier Skalen des KIMS auf Basis der Items der jeweiligen Subskala des KIMS-Short in Stichprobe B (Zeilen BEO, BES, MAH und AOB). In der untersten Zeile sind die Gütekriterien einer CFA mit allen Items des KIMS-Short und korrelierten Faktoren aufgeführt (Stichprobe A). . . . .	35
4.4	Latente Faktorinterkorrelationen des KIMS-Short (Stichprobe A). . . . .	35
4.5	Korrelationen der fünf Faktoren des KIMS-Short mit der MAAS, mit Persönlichkeitsvariablen und Maßen psychischer Belastung. . . . .	37

# Kapitel 1

## Achtsamkeit (*mindfulness*) und ihre psychometrische Erfassung

Die Verbreitung achtsamkeitsbasierter Interventionen zur Behandlung somatischer und psychischer Syndrome nahm in den vergangenen Jahren stetig zu, bereits im Jahr 1998 wurde die Zahl der therapeutischen Behandlungsprogramme mit achtsamkeitsbasierten Interventionen auf 240 beziffert (vgl. Salmon, Santorelli & Kabat-Zinn, 1998). Achtsamkeit (*mindfulness*), ein facettenreiches spirituelles Konzept aus dem Buddhismus (vgl. Buchheld & Walach, 2004), war seither der wissenschaftlich-empirischen Betrachtung ausgesetzt. Hierzu wurde es notwendig, Achtsamkeit als psychologisches Konstrukt psychometrisch erfassbar bzw. untersuchbar zu machen.

Im diesem ersten Kapitel wird zunächst eine weithin akzeptierte Definition des Konstrukts Achtsamkeit vorgestellt (vgl. Kapitel 1.1), anschließend wird Achtsamkeit als ein Konzept beschrieben, welches im Kontext therapeutischer Interventionen verankert war und ist (vgl. Kapitel 1.2). Weiterhin werden wichtige Achtsamkeitsskalen (Selbstbeurteilungsinstrumente) zur Erfassung von Achtsamkeit präsentiert (vgl. Kapitel 1.3) und Hinweise auf mögliche Validitätsmängel zweier wichtiger Achtsamkeitsskalen erörtert (vgl. Kapitel 1.4). Am Ende des Kapitels werden die Fragestellungen für diese Dissertation herausgearbeitet (vgl. Kapitel 1.5) und den Schriften aus dem Anhang zugeordnet, welche die Grundlage dieser Dissertationsschrift darstellen.

### 1.1 Eine Definition von Achtsamkeit

Die vorliegende Arbeit versteht Achtsamkeit als psychologischen Prozess, als eine spezifische Form der Aufmerksamkeitslenkung (vgl. Bishop et al., 2004; Heidenreich & Michalak, 2003). Willentlich gelenkte Aufmerksamkeit wird auf im Hier und Jetzt ablaufende externe (z.B. Dinge oder Ereignisse) und internale Erlebnisinhalte (z.B. Gedanken, Gefühlen oder Körperempfindungen) gerichtet und automatische Reaktionen auf diese Erlebnisin-



halte werden unterlassen, ihre Impulse aber beobachtet bzw. verbalisiert. Die achtsame Aufmerksamkeitslenkung ist gekennzeichnet durch eine offene und annehmende (nicht bewertende) Haltung gegenüber äußeren und inneren Erlebnisinhalten unabhängig davon, ob diese als angenehm oder unangenehm erlebt werden. Im Rahmen achtsamer Aufmerksamkeitslenkung sollen Erlebnisinhalte ohne Bewertung und ohne Einordnung in Kategorien so angenommen werden wie sie sind. Die so verstandene Achtsamkeit wird in einer bedeutenden buddhistischen Lehrrede zur Achtsamkeitsmeditation auch als „Reines Beobachten“ (Nyanaponika, 2000) bezeichnet:

„Reines Beobachten ist das klare, unabgelenkte Beobachten dessen, was im Augenblick der jeweils gegenwärtigen Erfahrung (einer äußeren oder inneren) wirklich vor sich geht. Es ist die unmittelbare Anschauung der eigenen körperlichen und geistigen Daseinsvorgänge, soweit sie in den Spiegel unserer Aufmerksamkeit fallen. Dieses Beobachten gilt als ‚rein‘, weil sich der Beobachter dem Objekt gegenüber rein aufnehmend verhält, ohne mit dem Gefühl, dem Willen oder Denken bewertend Stellung zu nehmen und ohne durch Handeln auf das Objekt einzuwirken. Es sind die ‚reinen Tatsachen‘, die hier zu Wort kommen sollen. Wenn sich nun aber an ein anfänglich reines Registrieren dieser Tatsachen aus alter Gewohnheit doch wieder gleich Bewertungen und andere Reaktionen anschließen, so sollen dann eben diese Reaktionen selber sofort wieder zum Gegenstand Reinen Beobachtens gemacht werden. Eine so gewonnene innere Freiheit dem Objekt gegenüber wird durch Einübung allmählich zu einer vertrauten Geisteshaltung, die leicht verfügbar ist, wenn sie benötigt wird.“

Angefügt sei, dass die Wurzeln des psychologischen Verständnisses von Achtsamkeit dem Buddhismus entlehnt und an anderer Stelle ausführlich beschrieben worden sind (vgl. Buchheld & Walach, 2004).

## 1.2 Achtsamkeit im Kontext therapeutischer Interventionen

Zum besseren Verständnis von Achtsamkeit und ihrer Bedeutung ist es sinnvoll, näher auf die wichtigsten achtsamkeitsbasierten Interventionen, auf ihr Verständnis von Achtsamkeit und auf ihre Wirksamkeit einzugehen. Eine zentrale Stellung unter den therapeutischen Interventionen zur Achtsamkeit nimmt jene von Kabat-Zinn (2003) ein. In seiner Achtsamkeitsdefinition geht er von drei wesentlichen Aspekten aus, von der bereits erwähnten absichtsvollen Aufmerksamkeitslenkung (*paying attention on purpose*), dem vorurteilsfreien Wahrnehmen und Nicht-Bewerten (*nonjudgmentally to the unfolding of experience*) und der Präsenz im Augenblick (*awareness in the present moment*). In dem von ihm entwickelten Programm der Achtsamkeitsbasierten Stressbewältigung (*Mindfulness-based Stress Reduction* (MBSR); vgl. Kabat-Zinn, 1990) nehmen Patienten an einem achtwöchigen Programm

teil, welches tägliche Übungen zur Achtsamkeitspraxis (z.B. Body Scan, Yoga, Atemmeditation) vorsieht. Davidson et al. (2003) führte eine randomisierte Kontrollstudie mit einer MBSR-Trainingsgruppe und einer Kontrollgruppe (Warteliste) durch. Es ergab sich durch das MBSR-Training eine signifikant erhöhte linksseitige frontale Aktivierung im EEG und eine signifikant erhöhte Immunreaktion auf ein verabreichtes Influenzavakzin der Probanden der MBSR-Trainingsgruppe im Vergleich zu den Probanden der Kontrollgruppe.

Eine Forschungslinie sieht Achtsamkeit als bedeutsamen Einflussfaktor im Rahmen von metakognitiven Prozessen (vgl. Teasdale et al., 2002). So wird in der *differential activation hypothesis* (DAH; Teasdale, 1988) davon ausgegangen, dass depressive Stimmungen die Informationsverarbeitung negativ beeinflussen („kognitive Reaktivität“), insofern depressiogene Erklärungen von Erfahrungen (Empfindungen, Gefühle bzw. Gedanken) bei depressiver Stimmung leichter zugänglich werden als nicht depressiogene Erklärungen. Weiterhin führten die vermehrt zugänglichen negativen Interpretationen von Erfahrungen ihrerseits zu depressiver Stimmung („Spirale der Depression“). Während im Beckschen Depressionsmodell (vgl. Beck, 1967) depressive Verstimmungen u. a. als Folge negativer Kognitionen („Kognitive Triade“) bzw. depressive Rezidive als Folge kaum veränderbarer negativer Kognitionen gesehen werden, besteht die zentrale Annahme der DAH darin, depressive Rezidive durch eine gegenseitige Verstärkung depressiver Stimmung und negativer Interpretationen („Spirale der Depression“) zu erklären. Segal, Williams und Teasdale (2001) gehen deshalb davon aus, dass ein spezifisches Achtsamkeitstraining in Form der Achtsamkeitsbasierten Kognitiven Therapie (*Mindfulness-based Cognitive Therapy* (MBCT)), das sich stark an der MBSR orientiert, zu funktionaleren metakognitiven Reaktionen (*metacognitive awareness*) und damit zu einer Vermeidung der „Spirale der Depression“ und somit zu einer geringeren Wahrscheinlichkeit des Auftretens von depressiven Rezidiven führen kann. Es konnte mehrfach nachgewiesen werden, dass MBCT für Patienten mit zumindest drei depressiven Episoden in der Vorgeschichte eine signifikante Verringerung der Auftretenswahrscheinlichkeit eines erneuten Rezidivs bewirkt (vgl. Ma & Teasdale, 2004; Teasdale et al., 2000).

Im Kontext der Dialektischen Verhaltenstherapie (*Dialectical Behavioral Therapy* (DBT); vgl. Linehan, 1993) wird Achtsamkeit für Patienten anhand der drei „Was“-Aspekte bzw. drei „Wie“-Aspekte veranschaulicht. Die drei „Was“-Aspekte der Achtsamkeit umfassen „wahrnehmen“ (*observing*), „beschreiben“ (*describing*) und „teilnehmen“ (*participating*) als unabhängig voneinander zu trainierende psychologische Prozesse. Die drei „Wie“-Aspekte der Achtsamkeit verdeutlichen wesentliche Charakteristika von Achtsamkeit, die stets „nicht-bewertend“ (*nonjudgmentally*), „gegenwärtig“ (*in the moment*) und „zielgerichtet“ (*with effectiveness*) sein soll. In der DBT wird Achtsamkeit als ein Bündel von Fertigkeiten (sog. *skills*) aufgefasst, d.h. es wird versucht, Achtsamkeit in möglichst konkrete, alltagsnahe Verhaltensweisen zu übersetzen und diese zu trainieren bzw. anwenden zu lernen. Die DBT

als hochstrukturiertes Programm zur Behandlung der Borderline-Persönlichkeitsstörung (BPS) kombiniert traditionelle verhaltenstherapeutische Interventionen (z.B. Kontingenzmanagement, Stimuluskontrolle oder Exposition) mit individuell auf den jeweiligen Patienten ausgerichteten Achtsamkeitsübungen (*skills*). Die DBT gilt als Verfahren mit nachgewiesener Effektivität in der Behandlung der BPS, es konnten signifikante Verbesserungen hinsichtlich selbstverletzenden bzw. selbstschädigenden Verhaltens und anderer psychopathologischer Variablen dokumentiert werden (vgl. Bohus et al., 2004).

Als wesentlicher Aspekt von Achtsamkeit wird wiederholt „Akzeptanz“ genannt (vgl. Bishop et al., 2004; Kabat-Zinn, 1990; Linehan, 1993). Diese wird u. a. als „being ‚experientially open‘ to the reality of the present moment rather than being in a state of either belief or disbelief (or judging what is fair or unfair)“ (vgl. Roemer & Orsillo, 2002, S. 60) verstanden. Akzeptanz als Aspekt von Achtsamkeit wurde im Kontext der Akzeptanzbasierten Therapie (*Acceptance and Commitment Therapy* (ACT); Hayes, Strosahl & Wilson, 1999) negativ als Erlebnisvermeidung (*experiential avoidance*) bzw. positiv als Erlebnisbereitschaft (*willingness to experience*) konzeptualisiert. Achtsamkeit wird auch in diesem Kontext grundsätzlich als trainierbarer psychologischer Prozess verstanden. Der Nachweis therapeutischer Effekte von ACT wurde in zahlreichen Studien erbracht, z.B. ergaben sich vergleichbare Ergebnisse bei der Behandlung von Angst- bzw. depressiven Störungen zwischen einer Gruppe von Patienten, die mit kognitiver Therapie (CT), und einer Gruppe von Patienten, die mit ACT behandelt wurden (vgl. Forman, Herbert, Moitra, Yeomans & Geller, 2007).

Eine aktuelle Metaanalyse untersuchte die Wirksamkeit achtsamkeitsbasierter Interventionen bei Angst- bzw. depressiven Störungen (vgl. Hofmann, Sawyer, Witt & Oh, 2010). In diese Metaanalyse wurden insgesamt 39 Studien ( $N = 1140$ ) eingeschlossen, in denen sich die achtsamkeitsbasierten Interventionen im Wesentlichen eng an MBSR oder MBCT orientierten. Es ergaben sich moderate durchschnittliche Effektstärken für die Reduktion von Angst- (Hedges's  $g = 0.63$ ) bzw. depressiven Symptomen (Hedges's  $g = 0.59$ ). Baer (2003) führte eine Metaanalyse mit 15 Studien durch, die achtsamkeitsbasierte Interventionen dokumentiert hatten. Es wurden Effekte einer Reihe von Variablen untersucht (z.B. Angst, Depression, Schmerz etc.) und ein durchschnittlicher Effekt von  $d = 0.59$  ermittelt.

### 1.3 Fragebogen zur Achtsamkeit

Zur wissenschaftlich-empirischen Untersuchung der vielfältigen Effekte von Achtsamkeit und achtsamkeitsbasierten Interventionen war es notwendig geworden, sie psychometrisch zu erfassen. Hierfür wurden mehrere Achtsamkeitsskalen entwickelt, im Wesentlichen sollen fünf Selbstbeurteilungsinstrumente zur Erfassung von Achtsamkeit vorgestellt werden. Hierbei handelt es sich um die *Mindful Attention and Awareness Scale* (MAAS; Brown &

Ryan, 2003), das *Kentucky Inventory of Mindfulness Skills* (KIMS; Baer, Smith & Allan, 2004), den *Freiburger Fragebogen zur Achtsamkeit* (FFA; Walach, Buchheld, Buttmüller, Kleinknecht & Schmidt, 2006), die *Toronto Mindfulness Scale* (TMS; Lau et al., 2006) und das *Acceptance and Action Questionnaire* (AAQ; Hayes et al., 2004).

Die *Mindful Attention and Awareness Scale* (MAAS) besteht aus 15 Items, die negativ gepolt sind und auf einer 6-stufigen Likert-Skala *mindful attention and awareness* (MAA) „indirekt“ messen sollen. Die MAAS erfasst die offene und ungeteilte Beobachtung innerer und äußerer Stimuli, konzeptualisiert Achtsamkeit somit eindimensional als Grad an Unaufmerksamkeit in Bezug auf äußere bzw. innere Stimuli. Bei einer Retest-Reliabilität von  $r_{tt} = .81$  gehen die Autoren davon aus, dass die MAAS „individual differences in the frequency of mindful states over time“ (Brown & Ryan, 2003, S. 824) erfasst. In Korrelationsstudien (vgl. Brown & Ryan, 2003) zeigten sich positive Zusammenhänge zwischen der MAAS und „openness to experience“ des NEO-PI-R und negative Zusammenhänge zwischen der MAAS und beispielsweise Depressivität bzw. Ängstlichkeit.

Das *Kentucky Inventory of Mindfulness Skills* (KIMS) ist im Zusammenhang der Dialektischen Verhaltenstherapie (DBT) entstanden bzw. konzeptualisiert Achtsamkeit als mehrdimensionales Konstrukt anhand vier zentraler Achtsamkeitsfertigkeiten. Die Skala „Beobachten“ (*observing*) erfasst in Analogie zur MAAS den Aufmerksamkeitsgrad bei der Beobachtung äußerer und innerer Stimuli. Die Skala „Beschreiben“ (*describe*) erfasst die Tendenz, beobachtete Erlebnisinhalte in Worte zu fassen, dieses sog. „Labeln“ wird hierbei als achtsame Form der Distanzierung (Beobachtung) angesehen. Die Skala „Mit Aufmerksamkeit handeln“ (*acting with awareness*) erfasst die ungeteilte Aufmerksamkeit auf die augenblicklich ausgeführte Handlung und wird dem Konzept des „automatic pilot“, welches Handlungsausführung im stark automatisierten Modus meint, gegenübergestellt. Die Skala „Akzeptieren ohne Bewerten“ (*accepting without judgment*) erfasst den Grad an Nicht-Bewertung bzw. Nicht-Evaluierung wahrgenommener Stimuli. Die 39 Items des KIMS werden auf einer 5-stufigen Likert-Skala beurteilt, bis auf die Skala „observe“ ist von jeweils guten Retest-Reliabilitäten auszugehen (.65; .81; .86; .83). Baer et al. (2004) zeigten, dass ein konfirmatorisches Faktorenmodell mit vier Faktoren (vier Achtsamkeitsfertigkeiten) einen befriedigenden Modellfit erreicht, ohne allerdings einen Modellvergleich z.B. mit einem vierfaktoriellen Modell mit Superfaktor durchzuführen. Der von Baer et al. (2004) durchgeführte Vergleich eines vierfaktoriellen Modells mit einem einfaktoriellen Modell ergibt aufgrund der theoretischen Konzeption des KIMS absehbar in Bezug auf sein Ergebnis eine klare Überlegenheit des vierfaktoriellen Modells gegenüber dem einfaktoriellen Modell. Ein Modellvergleich zwischen einem vierfaktoriellen Modell und einem vierfaktoriellen Modell mit Superfaktor auf Basis der deutschen Adaptation des KIMS ergab, dass das vierfaktorielle Modell dem Modell mit Superfaktor überlegen ist (vgl. Ströhle, Nachtigall, Michalak & Heidenreich, 2010).

Im *Freiburger Fragebogen zur Achtsamkeit* (FFA) werden 30 Items auf einer 4-stufigen Likert-Skala beurteilt und verschiedene Aspekte von Achtsamkeit, darunter „Urteilslosigkeit“, „Akzeptanz“ oder „Nicht-Identifikation“ erfasst. Berechnet wird ein Summenscore, die Gesamtskala besitzt eine hohe interne Konsistenz ( $\alpha = .93$ ). Faktorenanalytische Untersuchungen legen einen Generalfaktor nahe, da sich die anfangs gefundenen vier Faktoren (gegenwärtige, nicht-identifizierte Aufmerksamkeit; akzeptierende, nicht-urteilende Haltung; ganzheitliches Annehmen; prozesshaftes, einsichtsvolles Verstehen) in weiteren Analysen nicht eindeutig bestätigen ließen (vgl. Walach et al., 2006). Die Langversion erweist sich als für den Meditationskontext geeignet, da sie Items mit hohem Schwierigkeitskoeffizienten enthält, die von Probanden ohne Meditationserfahrung nicht beurteilt werden können. Für letztgenannte Personengruppe steht eine Kurzversion zur Verfügung, die 14 Items enthält.

Die *Toronto Mindfulness Scale* (TMS; Lau et al., 2006) erfasst 13 Items auf einer 5-stufigen Likert-Skala. Die zu Grunde liegende Struktur mit den zwei Faktoren Dezentrierung (*decentering*) und Neugier (*curiosity*) konnte mittels konfirmatorischer Faktorenanalysen bestätigt werden. Als Besonderheit der TMS geben die Autoren an, dass diese Achtsamkeit als Zustand (*state*) erfasse, während die übrigen Skalen zur Achtsamkeit diese eher als überdauernde Eigenschaft (*trait-like*) verstünden. Die Erfassung als Zustand wird durch die Instruktion zur TMS gewährleistet, die für die Beantwortung der Items keinen vergangenen Zeitraum, sondern den jeweiligen Moment der Beantwortung zu Grunde legt. Studien zur TMS, die ähnlich wie der FFA starke Nähe zum Meditationskontext aufweist, zeigten, dass mehr Erfahrung in Meditation mit höheren Werten in beiden Skalen der TMS einherging. Auch ein MBSR-Training erhöhte die Werte in beiden Skalen der TMS signifikant (vgl. Lau et al., 2006).

Das *Acceptance and Action Questionnaire* (AAQ; Hayes et al., 2004) wurde im Kontext der Acceptance and Commitment Therapy (ACT) entwickelt und liegt in unterschiedlichen Versionen vor: Es existieren Versionen von 9, 16 und 19 Items, je nach Anwendungsfeld, wobei die 9-Item-Version mittels konfirmatorischer Faktorenanalysen empirisch überprüft wurde (vgl. Hayes et al., 2004). Auf einer 7-stufigen Likert-Skala können Probanden Items in Bezug auf das Konstrukt Erlebnisbereitschaft (*willingness to experience*) bzw. Erlebnisvermeidung (*experiential avoidance/control*) vornehmen. Hierbei gehen die Autoren offensichtlich davon aus, dass es sich beim positiv gepolten Konstrukt (Erlebnisbereitschaft) und beim negativ gepolten Konstrukt (Erlebnisvermeidung) um ein und dasselbe Konstrukt handelt. Eine Durchsicht aller Items des AAQ in inhaltlicher Hinsicht deutet weiterhin darauf hin, dass neben Items zur Erlebnisbereitschaft auch solche zur Handlungsorientierung, zur Wahrnehmung der Gegenwart bzw. zur kognitiven Rumination vorkommen.

## 1.4 Ansätze für Revisionsmöglichkeiten in Achtsamkeitsskalen

### 1.4.1 Beschränkung auf negativ gepolte Items im Rahmen der Mindful Attention and Awareness Scale (MAAS) notwendig?

Im Rahmen von Studien zur MAAS verglichen Brown und Ryan (2003) sog. direkte und indirekte Messungen von Achtsamkeit. Die Autoren verstehen unter einer „direkten Messung“ positiv gepolte, unter einer „indirekten Messung“ negativ gepolte Items. Zur Überprüfung der Frage, ob die negativ gepolten Items der MAAS dasselbe Konstrukt messen wie die positiv gepolten Items, formulierten die Autoren anhand der 15 negativ gepolten Items positiv gepolte Items und legten die insgesamt 28 Items einer Stichprobe von  $N = 145$  Probanden vor. Zum Nachweis der Homogenität der 28 Items führten die Autoren eine EFA durch, die eine einfaktorielle Lösung ergab. Die Korrelation des Summenwertes der positiv gepolten Items mit dem Summenwert der negativ gepolten Items ergab  $r = -.70$ . Diese Korrelation auf der manifesten Ebene zeigt einen deutlichen Zusammenhang zwischen „Achtsamkeit“ (positiv gepolte Items) und „Unachtsamkeit“ (negativ gepolte Items).

Schließlich führten die Autoren als weiteren, deskriptiven Beleg für die Übereinstimmung direkter und indirekter Maße von Achtsamkeit Korrelationen des Summenwertes der positiv gepolten Items und der negativ gepolten Items mit Summenwerten anderer Skalen im Vergleich an. Dieser Vergleich zeige, dass die MAAS-Skala negativer und die MAAS-Skala positiv gepolter Items „(...) had very similar relations to measures used to assess convergent and discriminant validity (...) the two scales showed consistently the same directions of effect (...)“ (vgl. Brown & Ryan, 2003, S. 832). Genauer betrachtet zeigen sich aber doch interessante Unterschiede: So weichen die Korrelationen beider Skalen mit bestimmten anderen Skalen erkennbar voneinander ab, so z.B. mit dem *Life Orientation Test* LOT (.34; .23)<sup>1</sup>, mit der Skala „Neurotizismus“ des *NEO Five Factor Inventory* NEO-FFI (-.56; -.45), mit dem *Beck's Depression Inventory* BDI (-.42; -.32), mit der Skala „Rumination“ des *Rumination-Reflection Questionnaire* RRQ (-.39; -.25), mit der Skala „unpleasant affect“ des *Profile of Mood States* POMS (-.42; -.26) und der *Mindfulness/Mindlessness Scale* MMS (.33; .44).

Bei der MMS zeigt sich eine erkennbar höhere Korrelation mit der MAAS-Skala positiv gepolter Items im Vergleich zur MAAS-Skala negativ gepolter Items, was sowohl auf inhaltliche (die MMS erfasst eher Achtsamkeit als Unachtsamkeit) als auch auf methodische Gründe (die MMS besteht aus positiv gepolten Items) zurückgeführt werden kann. Bei den Korrelationen mit dem LOT zeigt sich das umgekehrte Bild: „Optimismus“ erwies sich zur Erklärung der sog. Itempolungseffekte demzufolge als Persönlichkeitseigenschaft,

---

<sup>1</sup>Der zuerst genannte Koeffizient stellt die Produkt-Moment-Korrelation mit der MAAS-Skala der negativ gepolten, der an zweiter Stelle genannte Koeffizient die Produkt-Moment Korrelation mit der MAAS-Skala der positiv gepolten Items dar.

welche eher mit den negativ gepolten Items der MAAS korrelierte als mit den positiven. Die Neigung, ein „Unachtsamkeits-Item“ abzulehnen wäre bei einem „optimistischen“ Probanden größer als die Neigung, ein „Achtsamkeits-Item“ zu bejahen. In dieser Sichtweise provozierte die Persönlichkeitseigenschaft „Optimismus“ einen stärkeren Itempolungseffekt bei den negativ gepolten als bei den positiv gepolten Items der MAAS-Skala. Diese inhaltliche Interpretation ist im Falle des LOT deshalb wahrscheinlich, weil der LOT aus drei positiv und drei negativ gepolten Items besteht, somit bezüglich der Itempolung als balanciert gelten kann, was gegen eine methodische Interpretation der höheren Korrelation mit der MAAS-Skala negativer Items spricht. Bei den Korrelationen der MAAS-Skalen mit den übrigen Skalen, die alle eine im weitesten Sinne psychopathologische Verhaltensweise abbilden, ergibt sich ein ähnliches Resultat, insofern alle Korrelationen mit der MAAS-Skala negativ gepolter Items deutlich höher ausfallen als die Korrelationen mit der MAAS-Skala positiv gepolter Items. Dieser Unterschied in den Korrelationen könnte hierbei allerdings auch methodisch interpretiert werden, da die meisten der Skalen das jeweilige Konstrukt mithilfe von negativ gepolten Items abfragen. Die höheren Korrelationen wären somit auf die gemeinsame Methodenvarianz der negativen Itempolung zurückzuführen.

#### **1.4.2 Heterogenität in der Mindful Attention and Awareness Scale (MAAS) – Sind weniger Items ausreichend?**

Sowohl in Bezug auf die Originalfassung der MAAS (vgl. Brown & Ryan, 2003) als auch in Bezug auf die Deutsche Adaptation der MAAS (vgl. Michalak, Heidenreich, Ströhle & Nachtigall, 2008) liegen mehrere konfirmatorische Faktorenanalysen (CFA) vor (vgl. MacKillop & Anderson, 2007). Die Evaluation der Modellanpassungen für diese einfaktoriellen CFA liefern Hinweise auf verbesserungsfähige Homogenität der Skala (vgl. Tabelle 1.1).

Im Zusammenhang des Anlegens strenger Maßstäbe zur Beurteilung der Güte der Modellanpassung (vgl. Schweizer, 2010) wird eingewandt, dass letztlich nur akzeptable Gütekriterien bei der Beurteilung der Anpassung von Modellen gewährleisten, dass es sich bei den spezifizierten Modellen um wirklich passende handelte. Es wird festgestellt, dass ein Minimum an Gütekriterien aufzuführen ist, um eine adäquate Beurteilung überhaupt zu ermöglichen. Schweizer (2010) schlägt hierzu die Angabe folgender Kennwerte vor: Den Quotienten aus  $\chi^2$ -Wert und Freiheitsgraden (df), den Root Mean Square Error of Approximation (RMSEA), den Comparative Fit Index (CFI) und das Standardized Root Mean Square Residual (SRMR). Anstelle des CFI kann beispielsweise auch der Nonnormed Fit Index (NNFI) gewählt werden.

Aus den verschiedenen Untersuchungen zur MAAS geht hervor, dass die Modellanpassungen in den CFA als noch akzeptabel gelten können (vgl. Schermelleh-Engel, Moosbrugger & Müller, 2003), wobei anzumerken ist, dass keine der aufgeführten Untersuchungen

Tabelle 1.1: Auflistung von Gütekriterien zu Modellanpassungen aus CFA zur MAAS in verschiedenen Untersuchungen.

	Brown und Ryan, 2003		Michalak et al.,	MacKillop und
	Studentische Stichprobe $N = 327$	Allgemeine Stichprobe $N = 239$	2008 Studentische Stichprobe $N = 469$	Anderson, 2007 Studentische Stichprobe $N = 711$
$\chi^2/df$	2.1	2.0	2.8	–
CFI	.91	.92	–	.91
RMSEA	.06	.07	.06	.07
SRMR	–	–	–	.08

*Anmerkung.* CFI: Comparative Fit Index; RMSEA: Root Mean-Square Error of Approximation; SRMR: Standardized Root Mean-Square Residual.

alle von Schweizer (2010) für notwendig erachteten Gütekriterien aufführte. Es zeigt sich aber auch, dass die eindimensionale MAAS mit ausschließlich negativ gepolten Items keinen annähernd guten Modellfit erreicht, so dass sich deshalb die Frage stellt, ob die MAAS wirklich als ausreichend homogen bezeichnet bzw. ob das zugrundeliegende Konstrukt *mindful attention and awareness* (MAA) auch mit weniger, dafür aber homogeneren Items erfasst werden kann. Die mangelnde Modellanpassung ist auch vor dem Hintergrund der in den genannten Untersuchungen verwendeten Maximum Likelihood (ML)-Schätzmethode zu sehen, die von multivariater Normalverteilung intervallskalierter Indikatoren ausgeht, was im Falle von Fragebogenitems zumindest diskussionswürdig ist.

### 1.4.3 Heterogenität im Kentucky Inventory of Mindfulness Skills (KIMS) – Chance für die Entwicklung einer Kurzform?

Baer et al. (2004) führten Validierungsstudien für die Originalversion des KIMS durch, hierbei kamen auch CFA zur Bestätigung der vierfaktoriellen Struktur des Fragebogens zum Einsatz (vgl. Kapitel 1.3). In diesen CFA wurden nicht einzelne Items als Indikatoren herangezogen, sondern Item-Päckchen (*item parcels*). Auch in der deutschen Adaptation des KIMS (Ströhle, Nachtigall, Heidenreich & Michalak, 2010) wurden Item-Päckchen als Indikatoren in einer CFA verwendet. Dies ist aufgrund der Itemanzahl im KIMS mit 39 Items nachvollziehbar. Eine CFA mit allen 39 Items des KIMS als Indikatoren zöge eine erhebliche Anzahl an zu schätzenden Parametern nach sich und erfordere damit eine ausreichend große Stichprobe. Für eine solche CFA müsste ferner erwartet werden, dass Kovarianzen zwischen Residualvariablen von Items aufträten, die inhaltlich nur schwer interpretierbar wären. Festzuhalten bleibt, dass es bisher noch keine Studien gab, in denen



eine CFA des KIMS mit Items als Indikatoren durchgeführt wurde.

Aus der Studie von Ströhle et al. (2010) ergibt sich für das KIMS, dass die CFA eines vierfaktoriellen Modells (vier korrelierte Faktoren) mit Item-Päckchen einen noch akzeptablen Modellfit erreicht. Dies gilt auch für die Originalversion des KIMS, in der der CFA eines vierfaktoriellen Modells (vier korrelierte Faktoren) mit Item-Päckchen ein guter Modellfit bescheinigt wird ( $CFI = .95$ ;  $RMSEA = .07$ ;  $\chi^2$ -Wert wird nicht berichtet). Unter Zugrundelegung strenger Maßstäbe an die Gütekriterien von Strukturgleichungsmodellen (vgl. Schweizer, 2010) stellt sich die Frage, ob auf der Basis dieser mit Item-Päckchen erzielten Ergebnisse die einzelnen Skalen des KIMS als ausreichend homogen im Sinne der CFA gelten können.

In der Forschungspraxis wird das KIMS als einziges multidimensionales Instrument zur Erfassung von Achtsamkeit häufig zusammen mit anderen Fragebogen in sog. Fragebogen-Sets eingesetzt. Das KIMS ist aufgrund ihrer 39 Items in diesem Zusammenhang weniger ökonomisch als andere Achtsamkeitsskalen wie beispielsweise die MAAS mit 15 Items. Da für Fragebogen-Sets gilt, dass ihr Umfang die Compliance der Probanden nicht negativ beeinflussen sollte, wäre die Entwicklung einer kürzeren Version eines multidimensionalen Instruments zur Erfassung von Achtsamkeit demzufolge für die Forschungspraxis sehr wünschenswert.

## 1.5 Fragestellungen

Das übergeordnete Ziel dieser Arbeit besteht darin, einen Beitrag zur Validierung von Achtsamkeitsskalen zu leisten. Methodologische Überlegungen sollen in diesem Zusammenhang ein grundsätzliches Problem der psychometrischen Erfassung von psychologischen Variablen mit Selbstbeurteilungsinstrumenten beleuchten und statistische Analysen sollen bestehende Achtsamkeitsskalen weiterentwickeln. Folgende Fragestellungen stehen im Vordergrund:

1. Methodeneffekte spielen im Rahmen der Validierung eines psychologischen Konstrukts eine bedeutsame Rolle, insofern sie – nicht angemessen beachtet – diese nachteilig beeinflussen können. Im Kapitel 2 dieser Arbeit soll mittels methodologischer Untersuchungen die Frage geklärt werden, wie Methodeneffekte angemessen analysiert werden können. Grundlage für die Analyse von Methodeneffekten bilden sog. *multitrait multimethod* (MTMM)-Daten, also Daten aus der Erfassung von Persönlichkeitsvariablen mit zumindest zwei unterschiedlichen Methoden. Zur Analyse dieser Daten werden konfirmatorische Faktorenanalysen (CFA) herangezogen, die Methodeneffekte berücksichtigen können. Unterschiedliche MTMM-CFA-Modelle sollen daraufhin untersucht werden, wie sie Methodeneffekte spezifizieren und wie diese jeweils interpretiert werden können. Im Rahmen dieser Fragestellung wird auf Schrift

(A) zurückgegriffen (vgl. Anhang A):

**Höfling, V.**, Schermelleh-Engel, K., & Moosbrugger, H. (2009). Analyzing multitrait-multimethod data: A comparison of three approaches. *Methodology*, 5, 99–111.

2. Kapitel 3 beschäftigt sich mit der Validierung der MAAS und untersucht die Frage, inwieweit mögliche Validitätsmängel der MAAS mit Itempolungseffekten (vgl. Kapitel 1.4.1) bzw. nicht homogenen Items (vgl. Kapitel 1.4.2) in Verbindung gebracht werden können. Im Rahmen der MAAS brachten Brown und Ryan (2003) ausschließlich negativ gepolte Items zum Einsatz, da die Autoren von der konzeptuellen Äquivalenz zwischen Achtsamkeit und „Unachtsamkeit“ (*mindlessness*) ausgingen. Es gibt jedoch Hinweise darauf, dass diese konzeptuelle Äquivalenz Störeffekten ausgesetzt ist (vgl. Kapitel 1.4.1). Da positiv und negativ gepolte Items als unterschiedliche „Methoden“ betrachtet werden können, soll die Eindimensionalität der MAAS demzufolge einer Untersuchung in Analogie zu MTMM-CFA-Modellen unterzogen werden (vgl. Fragestellung 1). Als Ergebnis dieser Untersuchungen wird eine modifizierte Version der Deutschen Adaptation der MAAS vorgestellt, die MAAS-Short. Im Rahmen dieser Fragestellung wird auf Schrift (B) zurückgegriffen (vgl. Anhang B):

**Höfling, V.**, Moosbrugger, H., Schermelleh-Engel, K., & Heidenreich, T. (in press). Mindfulness or mindlessness? A modified version of the Mindful Attention and Awareness Scale (MAAS). *European Journal of Psychological Assessment*.

3. Hinweise auf mögliche Validitätsmängel aus verschiedenen CFA mit dem KIMS sprechen für weitere Untersuchungen zum KIMS bzw. für Modifikationen am KIMS (vgl. Kapitel 1.4.3). Als multidimensionales Instrument zur Erfassung von Achtsamkeit gehört das KIMS neben der MAAS zu den bedeutsamsten Achtsamkeitsskalen und wird in Studien zur Achtsamkeit deshalb häufig eingesetzt. Kapitel 4 beschreibt die Vorgehensweise in Untersuchungen zum KIMS, die erstmalig auch CFA auf Item-Ebene umfassen und schließlich in eine Kurzfassung der Deutschen Adaptation des KIMS (KIMS-Short) münden. Ziel dieser Untersuchung war es, die vier Faktoren des KIMS mit wenigen, aber maximal homogenen Items zu erfassen. Im Rahmen dieser Fragestellung wird auf Schrift (C) zurückgegriffen:

**Höfling, V.**, Ströhle, G., Michalak, J., & Heidenreich, T. (under review). A short version of the Kentucky Inventory of Mindfulness Skills (KIMS). *Journal of Clinical Psychology*.

## Kapitel 2

# Methodeneffekte im Kontext psychologischer Datenerhebung und deren Analyse

In diesem Kapitel soll das Problem von Methodeneffekten im Kontext psychologischer Datenerhebung dargestellt und mittels methodologischer Untersuchungen gezeigt werden, wie Methodeneffekte angemessen analysiert werden können. Zunächst wird ein Überblick über mögliche Methodeneffekte im Rahmen von Selbstbeurteilungsinstrumenten gegeben (vgl. Kapitel 2.1), anschließend wird auf den spezifischen Effekt der Itempolung eingegangen (vgl. Kapitel 2.2). Nach einer Einordnung der Methodeneffekte in den Kontext der klassischen Testtheorie (vgl. Kapitel 2.3) werden Verfahren zur Analyse von Methodeneffekten vorgestellt (vgl. Kapitel 2.4).

### 2.1 Methodeneffekte im Rahmen von Selbstbeurteilungsinstrumenten

Im Rahmen von Selbstbeurteilungsinstrumenten ist von verschiedenen spezifischen Methodeneffekten auszugehen, die im Wesentlichen auf Charakteristika der verwendeten Items bzw. Itemformate und Ratingskalen zurückzuführen sind (vgl. Podsakoff, MacKenzie, Lee & Podsakoff, 2003, S. 882), d.h. ein Proband wird in seinem Testverhalten neben seiner spezifischen Traitausprägung auch von den (vorwiegend formalen) Besonderheiten eines Items geleitet. Soziale Erwünschtheit wird in diesem Zusammenhang nicht nur als generelle Antworttendenz von Probanden diskutiert, sondern auch als itemspezifisches Charakteristikum (vgl. Thomas & Kilman, 1975; Nederhof, 1985), insofern z.B. Itemformulierungen neben einer spezifischen Traitausprägung auch soziale Erwünschtheit mit erfassen und so zu Verzerrungen führen können. Sind Itemformulierungen komplex oder mehrdeutig, werden

Probanden angeregt, zufällig zu antworten, oder sich bei ihrer Beantwortung auf implizite oder ad-hoc Theorien zu stützen (vgl. Goia & Sims, 1985). Werden unterschiedliche Traits mit gleichen Ratingskalen und Kategorienbezeichnungen (z.B. von „trifft nicht zu“ bis „trifft zu“) gemessen, so ist von einer erhöhten Kovariation aufgrund dieses methodenspezifischen Einflusses auszugehen (vgl. Tourangeau, Rips & Rasinski, 2000).

Neben diesen itemspezifischen Einflüssen bei einer Messung können auch Positionierungseffekte von Items eine Rolle spielen, also etwa Priming-Effekte (vgl. Salancik, 1984) oder kognitive Carryover-Effekte (Harrison & McLaughlin, 1993). Auch die Wirkung von Itemformulierungen auf die Stimmung von Probanden wird in diesem Zusammenhang diskutiert (vgl. Peterson, 2000).

Von Methodeneffekten wird auch ausgegangen, wenn innerhalb einer Skala sowohl negativ als auch positiv gepolte Items verwendet werden. Die gleichzeitige Verwendung positiv und negativ gepolter Items wird zur Reduktion von Antworttendenzen (z.B. Ja-sage-Tendenz oder Nein-sage-Tendenz) empfohlen (vgl. Winkler, Kanouse & Ware, 1982). Auf diese Weise sollen Verzerrungen abgemildert werden. Zusätzlich fungiert die gegenteilige Itempolung als „speed bump“, d.h. der Proband soll daran gehindert werden, bei seinem Antwortverhalten in einen automatisierten kognitiven Prozess zu verfallen (vgl. Podsakoff et al., 2003, S. 884). Aufgrund gegenteiliger Itempolung soll er jedes einzelne Item nicht-automatisiert, d.h. inhaltsbezogen bearbeiten. In exploratorischen Faktorenanalysen konnte wiederholt gezeigt werden, dass beispielsweise negativ gepolte Items (verschiedener Skalen) einen eigenen, weitgehend exklusiven Faktor bilden (Artefakt), welcher als „Methodenfaktor“ interpretiert werden könnte (vgl. Harvey, Billings & Nilan, 1985). Aus den verschiedenen Untersuchungen sollen nachfolgend jene zu *Rosenbergs Self-Esteem Skala* (RSES; Rosenberg, 1965) angeführt werden, da sie aufgrund ihrer methodischen Vorgehensweise von besonderer Bedeutung sind.

## 2.2 Der spezifische Methodeneffekt der Itempolung im Kontext von Self-Esteem

Bei verschiedenen Untersuchungen zu *Rosenbergs Self-Esteem Skala* (RSES; Rosenberg, 1965) konnte gezeigt werden, dass bei gleichzeitiger Verwendung von positiv und negativ gepolten Items von Methodeneffekten aufgrund Itempolung auszugehen ist (vgl. Horan, DiStefano & Motl, 2003). Die verschiedentlich berichteten Methodeneffekte aufgrund von Itempolung bei der Erfassung von Self-Esteem konnten auch mit bestimmten Traits in Verbindung gebracht werden, z.B. mit „Emotionaler Stabilität“ oder dem „Behavioral Inhibition System“ (vgl. DiStefano & Motl, 2006; Quilty, Oakman & Risko, 2006). Höhere Werte in diesen beiden Variablen gehen tendenziell mit stärkerer Antwortneigung bei negativ gepolten Items einher. Auch aus den Bereichen der Emotionsforschung und der Organi-

sationspsychologie wurde von Methodeneffekten bei gleichzeitiger Verwendung von positiv und negativ gepolten Items berichtet (vgl. Spector, Van Katwyk, Brannick & Chen, 1997; Vautier, Steyer, Jmel & Raufaste, 2005). Untersuchungen zu Methodeneffekten aufgrund von Itempolung wurden für relevante Konstrukte (wie z.B. Achtsamkeit) aus dem Bereich der Klinischen Psychologie bzw. Gesundheitspsychologie bisher nicht vorgenommen.

Von besonderem Interesse ist die methodische Vorgehensweise bei systematischen Untersuchungen von Itempolungseffekten bei der RSES. Horan et al. (2003) führten Modellvergleiche konfirmatorischer Faktorenanalysen zum RSES (vier positiv und drei negativ gepolte Items) durch. Hierbei spezifizierten sie insgesamt sechs alternative CFA- bzw. MTMM-CFA-Modelle (vgl. Abbildung 2.1): Zwei CFA-Modelle spezifizierten keinen Methodeneffekt aufgrund von Itempolung, wobei eines dieser CFA-Modelle einen Traitfaktor für Self-Esteem (Modell A in Abbildung 2.1), das andere CFA-Modell je einen Traitfaktor für die positiv bzw. negativ gepolten Items, also insgesamt zwei korrelierte Traitfaktoren spezifizierte (Modell B). Weitere zwei MTMM-CFA-Modelle spezifizierten einen latenten Methodenfaktor für die positiv gepolten bzw. für die negativ gepolten Items bei einem Traitfaktor für Self-Esteem (Modell C und Modell D). Zwei weitere MTMM-CFA-Modelle spezifizierten keine latenten Methodenfaktoren, ließen aber Korrelationen zwischen Residualvariablen zu, in einem Modell ausschließlich zwischen den positiv gepolten, im anderen Modell ausschließlich zwischen den negativ gepolten Items (Modell E und Modell F). Die Modelle mit Berücksichtigung von Itempolungseffekten erreichten die bessere Modellanpassung, am besten schnitten das MTMM-CFA-Modell mit einem Methodenfaktor für die negativ gepolten Items und das MTMM-CFA-Modell mit Residualvariablenkorrelationen für die negativ gepolten Items ab.

Aus den bisherigen Befunden ergibt sich die Schlussfolgerung, dass bei gleichzeitiger Verwendung positiv und negativ gepolter Items von Methodeneffekten aufgrund Itempolung auszugehen ist. Insofern sollten positiv und negativ gepolte Items im Rahmen von Dimensionalitätsuntersuchungen bzw. im Rahmen der Konstruktvalidierung bei der Konstruktion von psychometrischen Skalen, z.B. in Form von CFA berücksichtigt werden. Festzuhalten bleibt, dass zum einen die gleichzeitige Verwendung von positiv und negativ gepolten Items im Rahmen einer Skala zur Verhinderung von Antworttendenzen sinnvoll erscheint, zum anderen diese unterschiedliche Itempolung als Methodeneffekt unbedingt Berücksichtigung finden sollte. Nachfolgend werden die Konzeptualisierung von Methodeneffekten im Kontext der Klassischen Testtheorie (KTT) und Möglichkeiten zur Analyse solcher Methodeneffekte dargestellt.

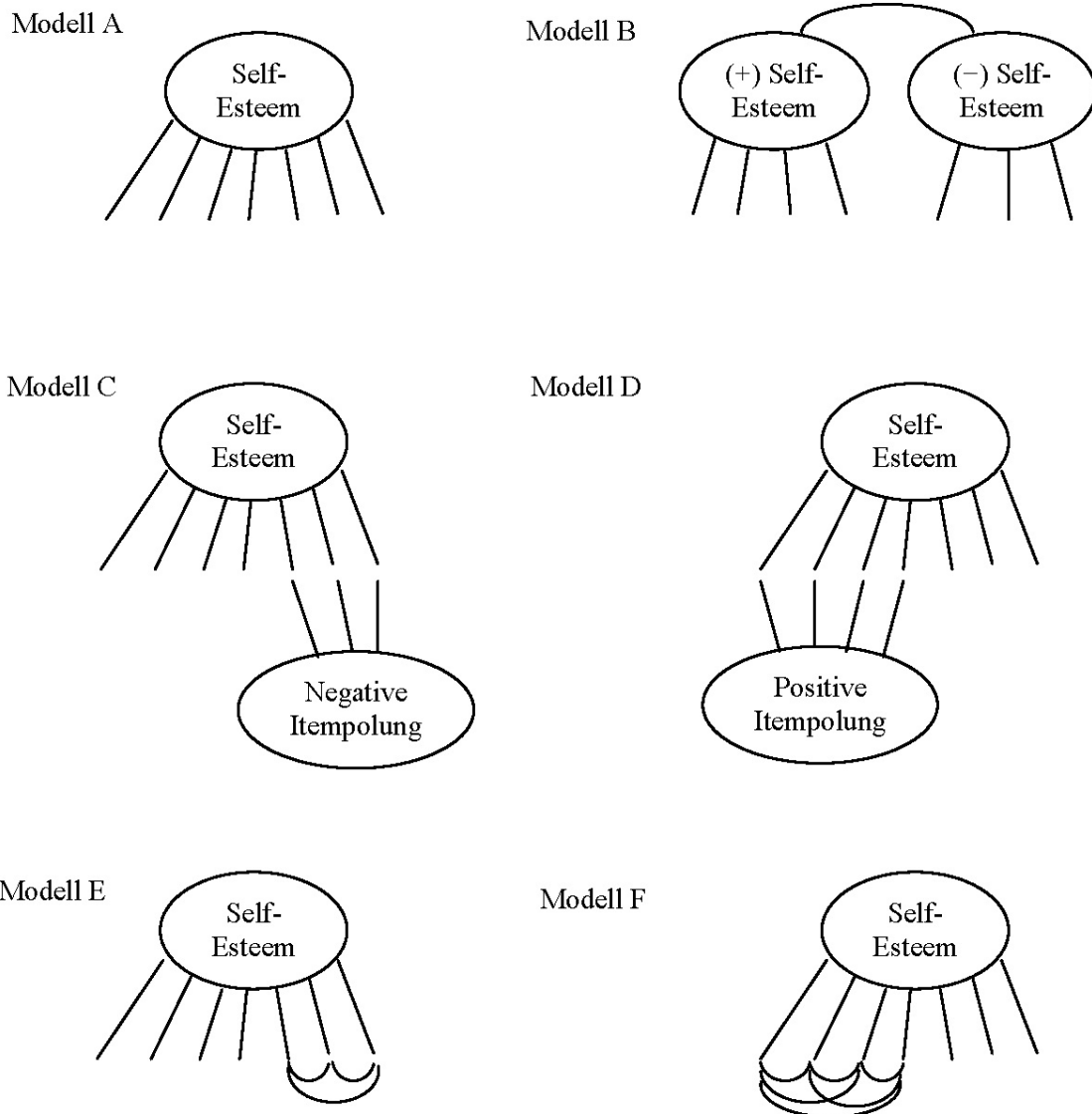


Abbildung 2.1: Schematische Pfaddiagramme sechs alternativer CFA- bzw. MTMM-CFA-Modelle zur Analyse von Itempolungseffekten im Rahmen der RSES (vgl. Horan et al., 2003, S. 442). Modell A und Modell B berücksichtigen keine Methodeneffekte, Modell C und Modell D spezifizieren Methodenfaktoren für die Itempolung, Modell E und Modell F setzen Kovarianzen zwischen den Residualvariablen der vier positiv bzw. drei negativ gepolten Items frei.

## 2.3 Die Konzeptualisierung von Methodeneffekten im Kontext der Klassischen Testtheorie (KTT)

Im Zusammenhang der Erfassung von Persönlichkeitsmerkmalen (Traits) wiesen Campbell und Fiske (1959) auf die Einflüsse von Erfassungsmethoden (z.B. Selbstbeurteilungsinstrumenten, Interviews etc.) auf die Messergebnisse hin. Diese Überlegungen erweitern das Verknüpfungaxiom der Klassischen Testtheorie (vgl. Moosbrugger, 2007, S. 101), welches besagt, dass sich jede Messung  $x_{vi}$  an Proband  $v$  im Item  $i$  aus einem wahren Wert  $\lambda_{vi}$  und einem zufälligen Fehlerwert  $\epsilon_{vi}$  zusammensetzt:

$$x_{vi} = \lambda_{vi} + \epsilon_{vi} \quad (2.1)$$

Die Überlegungen von Campbell und Fiske (1959) ergänzen somit das Verknüpfungaxiom um den Term  $\mu_{vi}$ , welcher als systematischer Methodenwert zu interpretieren ist:

$$x_{vi} = \tau_{vi} + \mu_{vi} + \epsilon_{vi} \quad (2.2)$$

Der für diagnostische Schlussfolgerungen relevante Term  $\tau_{vi}$  kann in Gleichung (2.2) unter Berücksichtigung eines konkreten Methodeneffekts (nämlich beispielsweise des spezifischen Einflusses eines Erhebungsinstruments) genauer bestimmt und somit die Validität von Schlussfolgerungen auf Basis der genaueren Traitausprägungen erhöht werden. Mit anderen Worten: Die unzureichende Berücksichtigung von Methodeneffekten kann zu einer verzerrten Schätzung von Traitausprägungen bei einer Person  $v$  und damit zu einer weniger validen Interpretation auf Basis dieser Traitausprägungen führen.

Mit „Methodeneffekt“ sind somit im Kontext der KTT verschiedene systematische Varianzquellen bezeichnet, die sich über den Trait hinaus auf das Messergebnis auswirken können (vgl. Schermelleh-Engel & Schweizer, 2007, S. 327). Systematiken von Methodeneffekten (vgl. Biesanz & West, 2004; Podsakoff et al., 2003) verweisen auf drei verschiedene Bereiche von Methodeneffekten, auf Einflüsse durch das Messinstrument (*method*), von Seiten des Beurteilers (*informant*) und des Kontextes (*occasion*).

Studien bzw. Analysen zu Methodeneffekten im Kontext eines interessierenden psychologischen Konstrukts dienen in statistischer Hinsicht dem Ziel der erweiterten Aufklärung systematischer Varianz des Messwerts  $x_{vi}$  und in diagnostischer Hinsicht einer verbesserten Validität aufgrund genauerer Bestimmungsmöglichkeiten von konkreten Traitausprägungen.

## 2.4 Analyseverfahren zur Berücksichtigung von Methodeneffekten – Die Multitrait-Multimethod (MTMM)-Analyse

Die klassische Vorgehensweise bei der deskriptiven Analyse der Multitrait-Multimethod-Matrix (MTMM-Matrix) nach Campbell und Fiske (1959) wurde verschiedentlich beschrie-

ben (vgl. z.B. Schermelleh-Engel & Schweizer, 2007, S. 332f.). Im Wesentlichen geht es darum, mittels MTMM-Matrix die diskriminante bzw. konvergente Validität korrelationsanalytisch anhand von vier Kriterien zu belegen. Die vier Kriterien, anhand derer das Vorliegen von konvergenter und diskriminanter Validität überprüft wird, sind:

- *Kriterium 1 für konvergente Validität.* Die Monotrait-Heteromethod-Korrelationen sind signifikant von null verschieden.
- *Kriterium 2 für diskriminante Validität.* Die Heterotrait-Monomethod-Korrelationen sollen signifikant kleiner sein als die Monotrait-Heteromethod-Korrelationen.
- *Kriterium 3 für diskriminante Validität.* Die Heterotrait-Heteromethod-Korrelationen sollen signifikant kleiner sein als die Monotrait-Heteromethod-Korrelationen und insgesamt eher die geringsten Korrelationen darstellen.
- *Kriterium 4 für Konstruktvalidität.* Identische Muster in den Monomethod- und Heteromethod-Teilmatrizen sollen das gemeinsame Vorliegen von konvergenter und diskriminanter Validität belegen.

Aus der Perspektive der MTMM-Analyse besteht der zentrale Aspekt der Konstruktvalidität im gleichzeitigen Nachweis der konvergenten und diskriminanten Validität. Von konvergenter Validität kann nur ausgegangen werden, wenn die Erfassung eines Traits durch verschiedene Methoden konvergiert. Diskriminante Validität hingegen liegt vor, wenn die Erfassung unterschiedlicher Traits trotz verschiedener bzw. trotz gleicher Methoden nicht konvergiert.

An dieser skizzierten Vorgehensweise bei der klassischen MTMM-Analyse sind einige Aspekte einzuwenden:

- Die MTMM-Analyse beruht auf der Auswertung von manifesten Korrelationen ohne Berücksichtigung der Tatsache, dass manifeste Variablen messfehlerbehaftet sind. Die Schlussfolgerungen hingegen betreffen latente Traits und Methoden.
- Die klassische MTMM-Analyse ist eine deskriptive und nicht im eigentlichen Sinne prüfende Verfahrensweise und bleibt somit in erheblichem Maße subjektiv.
- Jede Korrelation der MTMM-Matrix stellt eine Trait-Methoden-Einheit dar, d.h. jede einzelne Korrelation umfasst sowohl die Einflüsse eines spezifischen Traits als auch die Einflüsse der Methode, mit der dieser Trait erfasst wurde. Die klassische MTMM-Analyse zielt nicht auf eine Separierung von Trait- bzw. Methodeneinflüssen.

Eine Vielzahl von statistischen Analyseverfahren wurde seit dem einflussreichen Artikel von Campbell und Fiske (1959) zur Analyse von Methodeneffekten entwickelt. Es wurden log-lineare Modelle (vgl. Nussbeck, 2006) und IRT-Modelle (vgl. Rost & Carstensen,



2002; Rost & Walter, 2006) für die MTMM-Analyse kategorialer Daten und CFA-Modelle für die MTMM-Analyse kontinuierlicher Daten vorgestellt. Letztere Modelle (MTMM-CFA-Modelle) gehören zu den am häufigsten angewandten Verfahren zur Analyse von MTMM-Daten (vgl. Dumenci, 2000; Eid, Lischetzke & Nussbeck, 2006; Schermelleh-Engel & Schweizer, 2007).

Zunächst wird eine Vorgehensweise zur MTMM-Analyse vorgestellt, welche die Überprüfung der vier Kriterien in den Bezugsrahmen der CFA verlagert (vgl. Kapitel 2.4.1). Anschließend folgt eine theoretische Systematik zu möglichen MTMM-CFA-Modellen (vgl. Kapitel 2.4.2), bevor ein spezifisches MTMM-CFA-Modell charakterisiert wird (vgl. Kapitel 2.4.3).

#### 2.4.1 „Übersetzung“ der vier Kriterien der klassischen MTMM-Analyse in den Bezugsrahmen der konfirmatorischen Faktorenanalyse (CFA)

Die Korrelationen in der MTMM-Matrix repräsentieren Trait-Methoden-Einheiten (*trait method units*), d.h. jede einzelne Korrelation umfasst sowohl die Einflüsse eines spezifischen Traits als auch die Einflüsse der Methode, mit der dieser Trait erfasst wurde. Unter Beibehaltung dieser Trait-Methoden-Einheiten ist es unter bestimmten Umständen möglich, eine klassische MTMM-Matrix in den Bezugsrahmen der CFA zu überführen und die vier Kriterien der klassischen MTMM-Analyse im Kontext latenter Variablen zu überprüfen. Liegen mehrere Indikatoren einer Trait-Methoden-Einheit vor (z.B. in Form mehrerer Items bei der Erfassung eines Traits durch einen Beurteiler), so ist es möglich, für diese Trait-Methoden-Einheit ein Messmodell aufzustellen. In Abbildung 2.2 ist ein beispielhaftes Strukturmodell einer solchen „übersetzten“ MTMM-Matrix abgebildet. Es wurden die Traits „Neurotizismus“, „Extraversion“ und „Gewissenhaftigkeit“ mit den Methoden „Selbstbeurteilung“, „Fremdbeurteilung A“ und „Fremdbeurteilung B“ erfasst. Jede Trait-Methoden-Einheit (z.B.  $TM_{11}$ , die mittels Selbstbeurteilung erfasste Extraversion) wurde im zugrunde liegenden Beispiel durch vier Items operationalisiert, so dass es möglich ist, für jede Trait-Methoden-Einheit ein Messmodell aufzustellen (vgl. Höfling et al., 2009, S. 100; Anhang A). Die Interkorrelationen im daraus resultierenden Strukturmodell (vgl. Abbildung 2.2) können dann als Grundlage zur Überprüfung der vier Kriterien der klassischen MTMM-Analyse dienen:

- *Kriterium 1 für konvergente Validität.* Die Monotrait-Heteromethod-Korrelationen, in diesem Fall die Korrelationen  $\phi_M$  zwischen den latenten Variablen, sollen signifikant von null verschieden sein.
- *Kriterium 2 für diskriminante Validität.* Die Heterotrait-Monomethod-Korrelationen, hier die Korrelationen  $\phi_T$ , sollen signifikant kleiner sein als die Monotrait-Heteromethod-Korrelationen  $\phi_M$ .

- *Kriterium 3 für diskriminante Validität.* Die Heterotrait-Heteromethod-Korrelationen  $\phi_{TM}$  (aus Vereinfachungsgründen nicht in Abbildung 2.2 dargestellt, es handelt sich um „diagonale“ Interkorrelationen) sollen signifikant kleiner sein als die Monotrait-Heteromethod-Korrelationen  $\phi_M$  und insgesamt eher die geringsten Korrelationen darstellen.
- *Kriterium 4 für Konstruktvalidität.* Identische Muster in den Monomethod- und Heteromethod-Teilmatrizen sollen das gemeinsame Vorliegen von konvergenter und diskriminanter Validität belegen, z.B. sollte sich Rangreihe der Interkorrelationen von  $\phi_{T11,21}, \phi_{T21,31}$  und  $\phi_{T11,31}$  auch in der Rangreihe von  $\phi_{T12,22}, \phi_{T22,32}$  und  $\phi_{T12,32}$  wiederfinden.

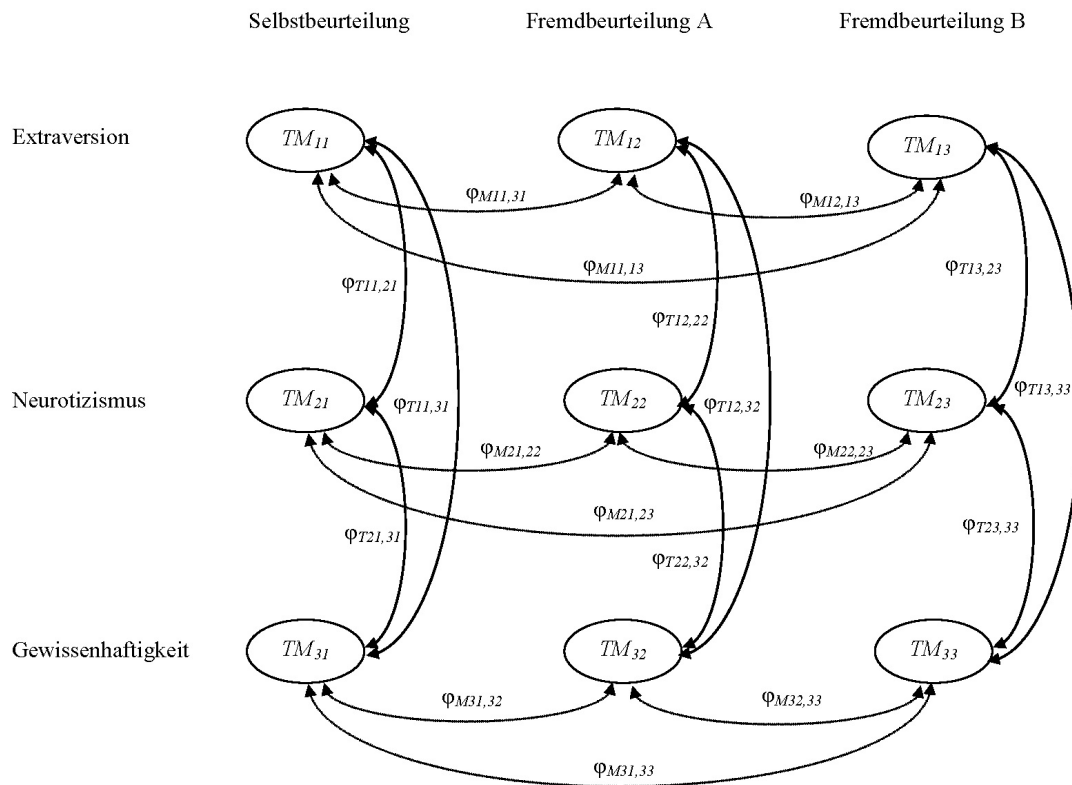


Abbildung 2.2: Pfaddiagramm eines beispielhaften Strukturmodells zu MTMM-Daten, in welchen die drei Traits Extraversion, Neurotizismus und Gewissenhaftigkeit mit den drei Methoden Selbstbeurteilung, Fremdbeurteilung A und Fremdbeurteilung B erfasst wurden. Jede latente Variable  $TM$  repräsentiert eine Trait-Methodeneinheit, die „vertikalen“ Korrelationen stellen Heterotrait-Monomethod- und die „horizontalen“ Korrelationen Monotrait-Heteromethod-Koeffizienten dar.

Die vorgestellte, in den Rahmen der CFA übersetzte MTMM-Analyse mittels der vier Kriterien der MTMM-Analyse nach Campbell und Fiske (1959) vermag den weiter oben genannten Kritikpunkt an den manifesten Korrelationen der klassischen MTMM-Analyse zu überwinden. Allerdings bleibt auch sie im Wesentlichen deskriptiv und ist damit nicht im eigentlichen Sinne eine prüfende Vorgehensweise. Schließlich hält sie an der ursprünglichen Idee der Trait-Methoden-Einheit fest und versucht nicht, Trait- und Methodeneinflüsse zu separieren.

#### 2.4.2 MTMM-Modelle im Kontext der CFA – Eine „theoretische“ Systematik zu MTMM-CFA-Modellen

Widaman (1985) entwickelte eine Taxonomie zur Analyse von MTMM-Daten ( $t$  Traits und  $m$  Methoden), die nicht mehr nur als deskriptiv, sondern als konfirmatorisch bezeichnet werden kann. Der Grundgedanke dieser Taxonomie besteht darin, dass eine Reihe von alternativen apriorischen MTMM-CFA-Modellen spezifiziert und hinsichtlich ihrer Modellgüte verglichen werden.

Zur Analyse der Traiteinflüsse wird wahlweise von keinem Traitfaktor, von einem Trait-Generalfaktor, von orthogonalen Traitfaktoren oder von obliquen Traitfaktoren ausgegangen. Zur Analyse der Methodeneinflüsse wird analog wahlweise von keinem Methodenfaktor, von einem Methoden-Generalfaktor, von orthogonalen Methodenfaktoren oder von obliquen Methodenfaktoren ausgegangen (vgl. Tabelle 2.1). Aus der Kombination dieser Möglichkeiten ergeben sich für die Taxonomie 16 theoretisch mögliche MTMM-CFA-Modelle. Widaman (1985) empfiehlt, drei Regeln bei der Spezifikation von MTMM-CFA-Modellen zu befolgen, nämlich 1) mit dem einfachsten Modell zu beginnen, 2) mittels Methodenfaktoren die Methodeneinflüsse abzubilden, um anschließend die verbleibende Varianz durch Traitfaktoren zu erklären, und 3) einen Vergleich der spezifizierten MTMM-CFA-Modelle aus der Taxonomie hinsichtlich ihrer Modellgüte vorzunehmen, um sich für das Modell mit der besten Modellgüte zu entscheiden.

Die Vorschläge von Widaman (1985) waren als Beitrag zur Objektivierung der MTMM-Analyse gedacht. Die Bestimmung konvergenter und diskriminanter Validität kann anhand eines aufgrund seiner Modellanpassung überlegenen MTMM-CFA-Modells erfolgen, welches eine Trennung von Trait- und Methodeneinflüssen vornimmt.

Faktisch findet nur eine geringe Auswahl der von Widaman (1985) vorgestellten MTMM-CFA-Modelle in der Praxis Anwendung: Das *correlated trait-correlated uniqueness* (CT-CU)-Modell (Kenny, 1976), das *correlated trait-uncorrelated method* (CT-UM)-Modell (Marsh, 1989; Marsh & Grayson, 1995) und das *correlated trait-correlated method* (CT-CM)-Modell (Marsh & Grayson, 1995). Dies hängt mit der Tatsache zusammen, dass bei vielen MTMM-CFA-Modellen Schätz- und Identifikationsprobleme auftreten (vgl. Eid, 2000; Eid & Diener, 2006; Grayson & Marsh, 1994; Marsh, 1989), z.B. negative Fehler-

Tabelle 2.1: Taxonomie alternativer CFA- bzw. MTMM-CFA-Modelle nach Widaman (1985). In Abhängigkeit von apriorischen Annahmen zu Trait- bzw. Methodeneinflüssen in den vorliegenden MTMM-Daten können verschiedene Modelle spezifiziert werden.

	<b>Kein Methoden-</b> <b>einfluss</b>	<b>Eine Methode</b>	<b><math>m</math> voneinander unabhän-</b> <b>gige Methoden</b>	<b><math>m</math> korrelierte Methoden</b>
<b>Kein Traiteinfluss</b>	Nullmodell	Ein Methodenfaktor	$m$ Methodenfaktoren (orthogonal)	$m$ Methodenfaktoren (oblique)
<b>Ein Trait</b>	Ein Traitfaktor	Ein Traitfaktor + ein Methodenfaktor	Ein Traitfaktor + $m$ Methodenfaktoren (orthogonal)	Ein Traitfaktor + $m$ Methodenfaktoren (oblique)
<b><math>t</math> voneinander un-</b> <b>abhängige Traits</b>	$t$ Traitfaktoren (orthogonal)	$t$ Traitfaktoren (orthogonal) + ein Methodenfaktor	$t$ Traitfaktoren (orthogonal) + $m$ Methodenfaktoren (orthogonal)	$t$ Traitfaktoren (orthogonal) + $m$ Methodenfaktoren (oblique)
<b><math>t</math> korrelierte</b> <b>Traits</b>	$t$ Traitfaktoren (oblique)	$t$ Traitfaktoren (oblique) + ein Methodenfaktor	$t$ Traitfaktoren (oblique) + $m$ Methodenfaktoren (orthogonal)	$t$ Traitfaktoren (oblique) + $m$ Methodenfaktoren (oblique)

varianzen oder standardisierte Faktorladungen größer als eins. Marsh und Bailey (1991) konnten zeigen, dass diese Schätzprobleme vor allem bei CT-CM- und CT-UM-Modellen auftreten. Aus diesem Grund kam es auch zur Weiterentwicklung des CT-C( $M-1$ )-Modells von Eid (2000).

### 2.4.3 Zum CT-C( $M-1$ )-Modell und zur Konzeption von Methodenfaktoren

Eine besondere Variante der vorgestellten Modelle stellt das von Eid (2000) vorgeschlagene Modell korrelierter Traits dar, in welchem ein Methodenfaktor weniger als Methoden insgesamt modelliert wird, das so genannte *correlated trait-correlated method minus one* [CT-C( $M-1$ )]-Modell. Bei diesem MTMM-CFA-Modell wird eine Methode nach theoretischen Überlegungen als Referenzmethode gewählt und nicht modelliert. Für diese Referenzmethode gilt, dass Trait- und Methodeneinflüsse nicht separiert, sondern als Trait-Methodeneinheit betrachtet werden. Alle anderen Methoden werden im Vergleich zur Referenzmethode interpretiert. Wird ein Persönlichkeitskonstrukt durch Selbst- und Fremdbeurteilungssitems gemessen, so kann beispielsweise die Selbstbeurteilung als Referenzmethode gewählt und die methodische Varianz der Fremdbeurteilungssitems (Nicht-Referenzmethoden) explizit modelliert werden. Ein Problem im Zusammenhang des CT-C( $M-1$ )-Modells ist allerdings die Frage der Wahl der geeigneten Referenzmethode, die idealerweise nach theoretischen Überlegungen erfolgen sollte. Eid (2000) konnte zeigen, dass die zuvor beschriebenen Identifikationsprobleme vieler MTMM-CFA-Modelle beim CT-C( $M-1$ )-Modell nicht auftreten, was auf die Vermeidung einer Überparametrisierung zurückgeführt wird.

Hinsichtlich der Konzeptualisierung von Methodenfaktoren ist darüber hinaus auf eine wichtige Unterscheidung hinzuweisen, nämlich auf die Frage, ob von generellen, Trait-übergreifenden Methodenfaktoren oder von Trait-spezifischen Methodeneffekten ausgegangen wird. Nussbeck, Eid, Geiser, Courvoisier und Lischetzke (2009) spezifizierten ein CT-C( $M-1$ )-Modell für drei Traits und drei Erfassungsmethoden dieser Traits und bestimmten den allgemeinen (Trait-übergreifenden) Einfluss der Nicht-Referenzmethoden (*common method-specificity*) und den Trait-spezifischen Einfluss der Nicht-Referenzmethoden (*unique method-specificity*). Dabei zeigte sich, dass der Trait-spezifische Methodeneinfluss stets erheblich höher war als der Trait-übergreifende Methodeneinfluss der Erfassungsmethoden (vgl. Höfling et al., 2009, S. 107). Eid et al. (2008) gehen davon aus, dass dieser Befund bei Fragebogendaten häufig zu finden sei.

## Kapitel 3

# Untersuchungen und Ergebnisse zur *Mindful Attention and Awareness* *Scale* (MAAS)

In diesem Kapitel werden Untersuchungen zur Eindimensionalität und Homogenität der MAAS und diesbezügliche Ergebnisse im Überblick dargestellt. Zunächst wird hierbei auf die dabei erhobenen und analysierten Daten eingegangen (vgl. Kapitel 3.1). Im Kapitel 3.2 wird die analytische Vorgehensweise zur Überprüfung von Itempolungseffekten und damit der Eindimensionalität in der MAAS erläutert, bevor die Frage der Homogenität der MAAS behandelt wird (vgl. Kapitel 3.3). Im Kapitel 3.4 werden schließlich die Ergebnisse aller diesbezüglichen Untersuchungen zusammenfasst.

### 3.1 Stichprobe

Die Stichprobe umfasst  $N = 606$  Teilnehmer und setzt sich aus drei Teilstichproben zusammen. Bei der größten Teilstichprobe „Frankfurt“ mit  $n_1 = 367$  handelt es sich um eine nicht-studentische Stichprobe, die in Zusammenarbeit mit den Teilnehmerinnen und Teilnehmern der Lehrveranstaltung „Testtheorie und Testkonstruktion“ im Sommersemester 2007 am Institut für Psychologie der Goethe-Universität, Frankfurt am Main, erhoben wurde. Die zweite Teilstichprobe „Esslingen“ mit  $n_2 = 188$  wurde an der Hochschule Esslingen in Lehrveranstaltungen zur „Psychologie für Soziale Arbeit und Pflege“ (Prof. Dr. Heidenreich) erhoben. Die dritte Teilstichprobe „Pädagogik“ mit  $n_3 = 51$  stammt aus einer Lehrveranstaltung für Lehramtsstudierende zu „Motivationspsychologie und ihre Anwendung“ (Prof. Dr. Vollmeyer).

In der Stichprobe ergibt sich für die Variable Alter ein Range von 16 bis 73 Jahren bei einem  $M = 31.1$  ( $SD = 12.0$ ). In Teilstichprobe „Frankfurt“ ergibt sich bei einem Range von 16 bis 73 Jahren der Mittelwert  $m_1 = 34.9$  ( $sd_1 = 13.3$ ), in Teilstichprobe

Tabelle 3.1: Prozentuale Häufigkeitsverteilung in den Variablen „Geschlecht“, „Schulabschluss“ und „Meditationserfahrung“ in der Gesamt- und den Teilstichproben.

		„Frankfurt“	„Esslingen“	„Pädagogik“	Gesamt
		$n_1 = 367$	$n_2 = 188$	$n_3 = 51$	
Geschlecht	weiblich	55.9%	79.1%	84.3%	65.6%
	männlich	44.1%	20.9%	15.7%	34.4%
Schulabschluss	Hauptschule	7.5%	0.0%	0.0%	4.5%
	Mittlere Reife	17.9%	0.0%	0.0%	11.8%
	(Fach-) Abitur	74.6%	100.0%	100.0%	83.7%
Meditationserfahrung	keine	79.9%	72.0%	72.5%	76.9%
	etwas	16.2%	20.9%	25.5%	18.4%
	viel	3.9%	7.1%	2.0%	4.7%

„Esslingen“ bei einem Range von 19 bis 51 Jahren ein Mittelwert  $m_2 = 25.6$  ( $sd_2 = 6.5$ ) und in Teilstichprobe „Pädagogik“ bei einem Range von 18 bis 43 Jahren ein Mittelwert  $m_3 = 24.4$  ( $sd_3 = 5.9$ ).

Aus Tabelle 3.1 geht hervor, dass die Gesamtsstichprobe und die beiden Teilstichproben „Esslingen“ und „Pädagogik“ dem Geschlecht nach nicht ausgewogen sind. Hinsichtlich der Qualifikation durch den Schulabschluss zeigt sich, dass die weitaus größte Zahl der Probanden, auch in Teilstichprobe „Frankfurt“, die diesbezüglich etwas heterogener erscheint, über die Fachgebundene bzw. Allgemeine Hochschulreife verfügt. In Bezug auf Meditationserfahrung ergibt sich ein einheitliches Bild im Vergleich der drei Teilstichproben, etwa dreiviertel der Probanden verfügen über „keine“, im Durchschnitt knapp 5% verfügen über „viel“ Meditationserfahrung. Aufgrund systematisch fehlender Werte gingen  $N = 602$  Datensätze in die Analyse ein.

### 3.2 Itempolung in der MAAS – MTMM-CFA-Modelle

Die *Mindful Attention and Awareness Scale* (MAAS; Brown & Ryan, 2003) wurde zur Erfassung einer wesentlichen Facette von dispositioanaler Achtsamkeit entwickelt, nämlich zur Erfassung des Konstrukts *mindful attention and awareness* (MAA). Die Autoren gingen von konzeptueller Äquivalenz zwischen zwei „Erfassungsmethoden“ aus, zwischen der Erfassung von Achtsamkeit mittels positiv gepolter Items und der Erfassung von „Unachtsamkeit“ (*mindlessness*) mittels negativ gepolter Items. Um die konzeptuelle Äquivalenz dieser beiden Erfassungsmethoden zu belegen, bestimmten sie zum einen die Korrelation zwischen dem jeweiligen Summenwert der negativ bzw. positiv gepolten Items mit  $r = -.70$ . Zum anderen berechneten sie Korrelationen zwischen den beiden Summenwer-

ten und weiteren Skalen, um zu dokumentieren, dass von einer vergleichbaren Validität (*comparative validity*) der beiden Erfassungsmethoden auszugehen sei (vgl. Kapitel 1.4.1). Aufgrund der für Brown und Ryan (2003) damit nachgewiesenen konzeptuellen Äquivalenz der Erfassungsmethoden und aufgrund der Tatsache, dass die positiv gepolten Items in hohem Maße schief verteilt und damit weniger geeignet zur Erfassung von MAA seien (vgl. Brown & Ryan, 2003, S. 826), hatten sich die Autoren dafür entschieden, ausschließlich negativ gepolte Items in der MAAS zu verwenden.

Die von Brown und Ryan (2003) für den Beleg der Eindimensionalität von MAA gewählte methodische Vorgehensweise kann allerdings nicht vollends überzeugen (vgl. Kapitel 1.4.1). Daher wurde im Rahmen der vorliegenden Arbeit ein alternatives methodisches Vorgehen zum Nachweis der Eindimensionalität von MAA gewählt. Mittels der Evaluation spezifischer MTMM-CFA-Modelle sollte dasjenige Modell identifiziert werden, welches die beste Modellanpassung zeigt, d.h. die Traiteinflüsse und die etwaig vorhandenen Methodeneinflüsse der Itempolung am besten modelliert. Bei der Konzeption der methodischen Vorgehensweise wurde auf Studien zu Itempolungseffekten aus dem Kontext von Self-Esteem (vgl. DiStefano & Motl, 2006; Horan et al., 2003, Quilty et al., 2006) bzw. Optimismus (vgl. Rauch, Schweizer & Moosbrugger, 2007; Schweizer & Rauch, 2008) zurückgegriffen. Diese Studien spezifizieren alternative MTMM-CFA-Modelle, um verschiedene Annahmen zu Trait- und Methodeneffekten zu überprüfen. Anschließend werden diese alternativen MTMM-CFA-Modelle hinsichtlich ihrer Modellgüte evaluiert, um die verschiedenen Annahmen zu Trait- und Methodeneffekten bestätigen bzw. verwerfen zu können (vgl. Kapitel 2.2). Bei der Spezifikation alternativer MTMM-CFA-Modelle wurde auch auf die Überlegungen von Widaman (1985) zurückgegriffen (vgl. Kapitel 2.4.2), der vorschlägt, alternative MTMM-CFA-Modelle zu spezifizieren, um Methoden- bzw. Traiteffekte zu überprüfen.

Zur Untersuchung der Eindimensionalität von MAA bei gleichzeitiger Verwendung positiv und negativ gepolter Items wurden analog zur Studie von Brown und Ryan (2003) 13 der 15 Items positiv umformuliert und der Stichprobe zur Bearbeitung vorgelegt (vgl. Höfling et al., in press, S. 6f.; Anhang B). Mit der Spezifikation alternativer MTMM-CFA-Modelle sollte überprüft werden, 1) ob ein Methodeneinfluss hinsichtlich Itempolung vorliegt oder nicht und 2) ob der Traiteinfluss besser durch einen Faktor MAA oder zwei Faktoren für MAA, jeweils einen Faktor für positiv gepolte Items und einen Faktor für negativ gepolte Items, erklärt wird. Um dies zu überprüfen, wurden vier Modelle spezifiziert, ein erstes mit einem Traitfaktor für MAA und ohne Methodenfaktor für Itempolungseffekte (Modell 1), ein zweites mit korrelierten Traitfaktoren, einem Traitfaktor MAA für positiv gepolte Items und einem Traitfaktor MAA für negativ gepolte Items (Modell 2), ein drittes mit einem Traitfaktor für MAA und einem Methodenfaktor für negativ gepolte Items (Modell 3) und ein viertes mit einem Traitfaktor für MAA und einem Methodenfaktor für positiv gepolte Items (Modell 4; vgl. Tabelle 3.2). Erwartet wurde eine bessere Modellanpassung



für die beiden Modelle mit Methodenfaktor (Modell 3 und Modell 4).

Tabelle 3.2: Übersicht über die spezifizierten CFA- und MTMM-CFA-Modelle auf Basis aller 28 positiv und negativ gepolter Items der MAAS.

Modell	Modellart	Traitfaktor(en)	Indikatoren	Methodenfaktor
1	CFA	MAA	28	–
2	CFA	MAA(+) und MAA(-)	28	–
3	MTMM-CFA	MAA	28	NIP
4	MTMM-CFA	MAA	28	PIP

*Anmerkung.* MAA: „mindful attention and awareness“; NIP: Negative Itempolung; PIP: Positive Itempolung.

### 3.3 Homogenität der Items der MAAS

Bisherige CFA zur MAAS werfen die Frage auf, ob diese Skala als hinreichend homogen bezeichnet werden kann (vgl. Kapitel 1.4.2). Weiteres Ziel der Untersuchungen zur MAAS im Rahmen dieser Arbeit war es daher, die fünf Items mit den höchsten Faktorladungen unter den negativ gepolten Items und die fünf Items mit den höchsten Faktorladungen unter den positiv gepolten Items zu ermitteln, also eine Reduktion der Itemanzahl der MAAS zur Erhöhung ihrer Homogenität zu erreichen. Hierzu wurden schrittweise CFA separat für die negativ und die positiv gepolten Items durchgeführt. „Schrittweise“ bedeutet, dass zunächst eine CFA mit allen positiv bzw. negativ gepolten Items (jeweils separat) durchgeführt und das Item mit der geringsten Faktorladung ermittelt wurde. Bei einer erneuten CFA (ohne eben dieses Item) wurde ein weiteres Item mit der geringsten Faktorladung identifiziert. Darauf folgte eine erneute CFA ohne dieses Item, wobei ein weiteres Item mit der geringsten Faktorladung ermittelt wurde. Dieses schrittweise Vorgehen wurde so lange durchgeführt, bis die fünf Items mit den höchsten Faktorladungen sowohl für die positiv als auch die negativ gepolten Items ermittelt waren.

Mit der auf diese Weise modifizierten MAAS, der MAAS-Short, wurden weitere MTMM-CFA-Modelle spezifiziert (vgl. Tabelle 3.3), um zu überprüfen, 1) ob auch in der MAAS-Short ein Methodeneinfluss hinsichtlich Itempolung vorliegt oder nicht, und 2) ob der Traiteinfluss besser durch einen Faktor MAA oder zwei Faktoren für MAA, jeweils einen Faktor für positiv gepolte Items und einen Faktor für negativ gepolte Items, erklärt wird. Hierzu wurde ein Modell mit einem Traitfaktor für MAA und einem Methodenfaktor für die negativ gepolten Items (Modell 5) und ein Modell mit einem Traitfaktor MAA und einem Methodenfaktor für die positiv gepolten Items (Modell 6) spezifiziert. Zwei weitere CFA-Modelle wurden analysiert, um dem möglichen Einwand zu begegnen, dass nach Eli-

Tabelle 3.3: Übersicht über die spezifizierten CFA- und MTMM-CFA-Modelle auf Basis der fünf positiv und negativ gepolten Items der MAAS-Short mit den höchsten Faktorladungen.

Modell	Modellart	Traitfaktor(en)	Indikatoren	Methodenfaktor
5	MTMM-CFA	MAA	10	NIP
6	MTMM-CFA	MAA	10	PIP
7	CFA	MAA(+) und MAA(-)	10	–
8	CFA	MAA	10	–

*Anmerkung.* MAA: „mindful attention and awareness“; NIP: Negative Itempolung; PIP: Positive Itempolung.

mination wenig homogener Items die Spezifikation eines Methodenfaktors obsolet geworden sein könnte. Modell 7 spezifizierte daher einen Traitfaktor für die positiv gepolten Items und einen Traitfaktor für die negativ gepolten Items, wobei beide Traitfaktoren korreliert waren, Modell 8 spezifizierte einen Traitfaktor für alle Items der MAAS-Short. Die Modelle 7 und 8 spezifizierten keinen Methodenfaktor.

Die Modelle wurden mit der Software *Mplus* Version 5 (Muthén & Muthén, 2006) unter Verwendung des Weighted Least Squares Means and Variance (WLSMV)-Schätzverfahrens (Muthén & Muthén, 2006, S. 484) analysiert. WLSMV kann bei ausreichend großer Stichprobe verwendet werden, wenn die Indikatoren substantiell von der Normalverteilung abweichen. Anzumerken ist freilich, dass Analysen mit diesem Schätzverfahren im Vergleich zum ML-Schätzverfahren in der Regel zu insgesamt etwas schlechteren Gütekriterien führen.

## 3.4 Ergebnisse

### 3.4.1 Evaluation der verschiedenen CFA- bzw. MTMM-CFA-Modelle

Die Ergebnisse der Untersuchungen zum Methodeneffekt der Itempolung in der MAAS bzw. der MAAS-Short ergeben sich aus der Evaluation der Gütekriterien der verschiedenen CFA- bzw. MTMM-CFA-Modelle. Die CFA für die MAAS (15 negativ gepolte Items) in der Untersuchungsstichprobe unter Verwendung des WLSMV-Schätzverfahrens ergab eine nur in Teilen akzeptable Modellanpassung ( $\chi^2/df = 5.03$ ; CFI = .90; NNFI = .94; RMSEA = .08; WRMR = 1.26) und bestätigt die nicht völlig überzeugenden Modellanpassungen der CFA zur MAAS in früheren Untersuchungen (vgl. Kapitel 1.4.2). Anhand der Modellanpassungen (Gütekriterien) der verschiedenen Modelle in Tabelle 3.4 bzw. Tabelle 3.5 sind vier Aspekte grundsätzlich hervorzuheben:

- Im Rahmen der CFA-Modelle kann die Traitvarianz offensichtlich nicht wesentlich besser durch zwei separate Traitfaktoren für MAA – erfasst mit positiv gepolten Items

Tabelle 3.4: Gütekriterien aus den Analysen zu den auf Basis der MAAS spezifizierten CFA- bzw. MTMM-CFA-Modellen.

Modell	$\chi^2$	<i>df</i>	WRMR	RMSEA	NNFI	AIC
1	3298.54	100	4.40	.23	.71	48063.30
2	3323.74	100	4.40	.23	.71	47895.48
3	718.21	107	1.83	.10	.95	46303.41
4	745.25	109	1.59	.10	.95	46460.45

*Anmerkung.* WRMR: Weighted Root Mean-Square Residual; RMSEA: Root Mean-Square Error of Approximation; NNFI: Nonnormed Fit Index; AIC: Akaike Information Criterion.

bzw. negativ gepolten Items – erklärt werden (Modell 2 im deskriptiven Vergleich zu Modell 1).

- Die Modellanpassung verbessert sich erkennbar in den MTMM-CFA-Modellen, wenn ein Methodenfaktor für die Itempolung spezifiziert wird (Modell 3 und 4 im deskriptiven Vergleich zu Modell 1 und 2), allerdings kann die Modellanpassung von Modell 3 und 4 noch nicht als akzeptabel gelten.
- Erst nach schrittweiser Reduktion der Items auf die jeweils fünf positiv bzw. negativ gepolten Items mit der höchsten Faktorladung zeigen die MTMM-CFA-Modelle eine nicht mehr akzeptable (Modell 6) bis gute Modellanpassung (Modell 5) für die modifizierte Version der MAAS (MAAS-Short) mit ihren nunmehr 10 Items (vgl. Tabelle 3.5).
- Für die CFA-Modelle der MAAS-Short (Modell 7 und 8, jeweils ohne Methodenfaktor) zeigt sich, dass deren Modellanpassung wiederum als nicht mehr akzeptabel gelten kann. Im Modell 7 wurde die Korrelation zwischen den beiden latenten Faktoren mit  $r = .86$  bestimmt.

Für die MAAS-Short kann somit davon ausgegangen werden, dass sie das zugrundeliegende Konstrukt *mindful attention and awareness* (MAA) eindimensional mit positiv und negativ gepolten Items erfasst. Aufgrund der Ergebnisse zu Modell 5 muss davon ausgegangen werden, dass die Eindimensionalität von MAA bei einem gleichzeitig vorhandenen Itempolungseffekt gegeben ist.

### 3.4.2 Itemkennwerte der MAAS-Short

Die Entscheidung von Brown und Ryan (2003), für die MAAS ausschließlich negativ gepolte Items zu verwenden, wurde von den Autoren mit „besseren“ Itemkennwerten für

Tabelle 3.5: Gütekriterien aus den Analysen zu den auf Basis der MAAS-Short spezifizierten CFA- bzw. MTMM-CFA-Modellen.

Modell	$\chi^2$	<i>df</i>	WRMR	RMSEA	NNFI	AIC
<b>5</b>	87.82	21	0.94	.07	.99	14733.70
<b>6</b>	189.03	21	1.45	.12	.98	14775.44
<b>7</b>	334.73	24	1.38	.15	.96	14858.67
<b>8</b>	377.78	23	1.49	.16	.96	14909.35

*Anmerkung.* WRMR: Weighted Root Mean-Square Residual; RMSEA: Root Mean-Square Error of Approximation; NNFI: Nonnormed Fit Index; AIC: Akaike Information Criterion.

die negativ gepolten Items im Vergleich zu den positiv gepolten begründet (vgl. Kapitel 3.1). Aus diesem Grund werden in Tabelle 3.6 die Mittelwerte und Varianzen sowie die Trennschärfekoeffizienten der Items der MAAS-Short vorgestellt. Der Summenwert der MAAS-Short in der untersuchten Stichprobe liegt bei  $M = 42.46$  ( $SD = 7.12$ ) und der Range zwischen 16 und 60.

Die Probanden nutzen die gesamte 6-stufige Likert-Skala bei allen Items bis auf Item 10 in der positiv gepolten Form aus. Für vier der fünf positiv gepolten Items ergibt sich aus Tabelle 3.6, dass ihre Mittelwerte höher sind und dass ihre Varianz geringer ist als die der negativ gepolten Items. Brown und Ryan (2003) nannten als einen wesentlichen Grund für die Nicht-Verwendung positiv gepolter Items, dass diese schief verteilt seien. Dies gilt auch für die Kennwerte der im Rahmen dieser Arbeit erhobenen Daten zur Deutschen Version der MAAS. Allerdings sind auch die negativ gepolten Items schief verteilt, wenn auch nicht ganz so stark. Auch der Befund geringerer Varianz ist aus Untersuchungen zur Originalversion der MAAS bekannt.

Tabelle 3.6: Auflistung der Iteminhalte und Itemkennwerte (Mittelwerte, Trennschärfekoeffizienten und Varianzen) der MAAS-Short.

Item	Iteminhalt	Trennschärfe- koeffizienten	Mittelwert	Varianz
maas03	Ich finde es schwierig, auf das konzentriert zu bleiben, was im gegenwärtigen Augenblick passiert.	.59	4.00	1.32
maas07	Es sieht so aus, als würde ich „automatisch funktionieren“, ohne viel Bewusstsein für das, was ich tue.	.51	4.27	1.28
maas10	Ich erledige Aufträge oder Aufgaben automatisch, ohne mir bewusst zu sein, was ich tue.	.61	4.26	1.05
maas11	Ich bemerke, wie ich jemandem nur mit einem Ohr zuhöre, während ich gleichzeitig etwas anderes tue.	.62	4.00	1.09
maas14	Ich merke, wie ich Dinge tue, ohne auf sie zu achten.	.58	3.88	1.13
maas03(+)	Ich finde es leicht, auf das konzentriert zu bleiben, was im gegenwärtigen Augenblick passiert.	.65	3.92	1.14
maas07(+)	Ich mache das, was ich tue, mit viel Bewusstsein.	.68	4.34	0.91
maas10(+)	Ich erledige Aufträge oder Aufgaben und bin mir dabei bewusst, was ich tue.	.64	4.81	0.76
maas11(+)	Wenn ich jemandem zuhöre, tue ich gleichzeitig nichts anderes.	.72	4.43	0.79
maas14(+)	Wenn ich Dinge tue, achte ich auch auf sie.	.63	4.54	0.88

## Kapitel 4

# Untersuchungen und Ergebnisse zum *Kentucky Inventory of Mindfulness Skills* (KIMS)

In diesem Kapitel werden Untersuchungen zur Homogenität, faktoriellen Validität und Konstruktvalidität des KIMS und diesbezügliche Ergebnisse im Überblick berichtet. Zunächst wird hierbei auf die dabei erhobenen und analysierten Daten eingegangen (vgl. Kapitel 4.1). Im Kapitel 4.2 werden die Vorgehensweise und Ergebnisse zur Identifikation der homogensten Items der Subskalen des KIMS und zur Faktorenstruktur des KIMS-Short dargestellt, bevor Zusammenhänge des KIMS-Short zu Persönlichkeitsvariablen und Maßen psychischer Belastung aufgezeigt werden (vgl. Kapitel 4.3).

### 4.1 Stichproben

Im Rahmen der Untersuchungen dieser Arbeit zum KIMS wurden zwei Stichproben zur Analyse herangezogen, im Folgenden als Stichprobe A (vgl. Kapitel 3.1) und Stichprobe B bezeichnet.

Stichprobe B aus der Untersuchung von Ströhle et al. (2010) umfasst  $N = 469$  Teilnehmer und setzt sich aus drei Teilstichproben zusammen, aus Teilstichprobe „Bochum“ mit  $n_1 = 127$ , aus Teilstichprobe „Frankfurt“ mit  $n_2 = 214$  und aus Teilstichprobe „Jena“ mit  $n_3 = 128$ . Für Stichprobe B ergibt sich für die Variable Alter ein Range von 18 bis 48 Jahren bei einem  $M = 24.0$  ( $SD = 6.0$ ). In Teilstichprobe „Bochum“ ergibt sich der Mittelwert  $m_1 = 24.6$  ( $sd_1 = 6.6$ ), in Teilstichprobe „Frankfurt“ der Mittelwert  $m_2 = 25.2$  ( $sd_2 = 6.5$ ) und in Teilstichprobe „Jena“ der Mittelwert  $m_3 = 21.7$  ( $sd_3 = 3.7$ ).

Aus Tabelle 4.1 geht hervor, dass Stichprobe B dem Geschlecht nach nicht ausgewogen ist. Hinsichtlich der Qualifikation durch den Schulabschluss zeigt sich, dass die weitaus größte Zahl der Probanden über die Allgemeine Hochschulreife verfügt. In Bezug auf die Praxis

in geistiger Übung ergibt sich ein einheitliches Bild im Vergleich der drei Teilstichproben, mindestens vier Fünftel der Probanden verfügen über „keine“, maximal 20% verfügen über regelmäßige Achtsamkeitspraxis. In Stichprobe B ist das Durchschnittsalter aufgrund des höheren Studentenanteils niedriger als in Stichprobe A.

Eine Substichprobe der Stichprobe B (Stichprobe B.1;  $N = 253$ ) von Ströhle et al. (2010) bearbeitete zusätzlich zu der Deutschen Adaptation des KIMS das *NEO-Fünf-Faktoren-Inventar* (NEO-FFI; Borkenau & Ostendorf, 1993), den *Fragebogen zur Erfassung dispositioneller Selbstaufmerksamkeit* (SAM; Philipp & Freudenberg, 1989) und die Skala „Symptombezogene Rumination“ aus dem *Response Styles Questionnaire* (RSQ; Kühner, Huffziger & Nolen-Hoeksema, 2007). Eine weitere Substichprobe (Stichprobe B.2;  $N = 216$ ) bearbeitete zusätzlich das *Beck Depressions-Inventar* (BDI; Hautzinger, Bailer, Worall & Keller, 1996) und das *Beck Angst-Inventar* (BAI; Margraf & Ehlers, 2002).

Tabelle 4.1: Prozentuale Häufigkeitsverteilung in den Variablen „Geschlecht“, „Schulabschluss“ und „Praxis geistiger Übung“ in Stichprobe B und ihren Teilstichproben.

		„Bochum“ $n_1 = 127$	„Frankfurt“ $n_2 = 214$	„Jena“ $n_3 = 128$	Gesamt
Geschlecht	weiblich	85.8%	85.6%	81.9%	84.6%
	männlich	14.2%	14.4%	18.1%	15.4%
Schulabschluss	Hauptschule	0.0%	0.0%	0.0%	0.0%
	Mittlere Reife	0.0%	0.0%	0.8 %	0.4%
	Abitur	95.3%	92.6%	96.9%	94.2%
	Universität	4.7%	7.4%	2.3%	5.4%
Praxis geistiger Übung	nein	88.9%	85.7%	79.5%	84.8%
	ja	11.1%	14.3%	20.5%	15.2%

## 4.2 Identifikation homogener Items – Die Kurzversion des KIMS (KIMS-Short)

Die Daten aus der Bearbeitung der Deutschen Adaptation des KIMS (Ströhle et al., 2010) durch Stichprobe B wurden zur Identifikation der homogensten Items in jeder Subskala des KIMS herangezogen. Hierzu wurden auf Ebene der Subskalen „schrittweise“ konfirmatorische Faktorenanalysen (CFA) durchgeführt (vgl. Kapitel 3.2), um die Faktorladungen der zu einer Subskala gehörenden Items zu bestimmen. Diese CFA auf Ebene der Subskalen spezifizierten einen latenten Faktor und die zugehörigen Items als Indikatoren. So konnte das Item mit der geringsten Faktorladung ermittelt werden, welches bei einer erneuten

CFA weggelassen wurde. Bei dieser erneuten CFA wurde ein weiteres Item mit der geringsten Faktorladung identifiziert, welches wiederum weggelassen wurde. Dieses schrittweise Vorgehen wurde so lange durchgeführt, bis die Items mit den Faktorladungen größer als .5 bzw. die fünf Items mit den höchsten Faktorladungen ermittelt waren.

Die Modelle wurden mit der Software *Mplus* Version 5 (Muthén & Muthén, 2006) unter Verwendung des Maximum-Likelihood Mean-adjusted (MLM)-Schätzverfahrens analysiert, da Stichprobe B für ein verteilungsfreies Schätzverfahren als zu klein erachtet wurde. Für die Subskalen „Beschreiben“ (BES) blieben fünf Items, für die Skala „Mit Aufmerksamkeit handeln“ (MAH) vier Item und für die Skala „Akzeptieren ohne Bewertung“ (AOB) fünf Items übrig (vgl. Tabelle 4.2). Diese Items können als die homogensten der jeweiligen Subskala bezeichnet werden und bilden insgesamt 14 Items der 20-Item-Version des KIMS-Short.

Im Rahmen der CFA für die Subskala „Beobachten“ (BEO) zeigte sich an sehr niedrigen Faktorladungen bzw. einer mangelhaften Modelanpassung, dass die zugehörigen Items dieser Skala nicht wirklich eindimensional sein können. Daher wurden in Bezug auf die Subskala BEO zwei latente (korrelierte) Faktoren spezifiziert. Auf den einen Faktor (BEO-IN) luden die Items, die sich auf das Beobachten innerer Stimuli bezogen, auf den anderen Faktor (BEO-AU) luden die Items, die sich auf das Beobachten äußerer Stimuli bezogen. Hierbei verblieben für jeden der beiden Faktoren drei Items, wobei die Korrelation zwischen BEO-IN und BEO-AU  $r = .86$  betrug (Stichprobe B; vgl. Tabelle 4.3). In Bezug auf die Modellanpassungen der durchgeführten CFA sei angefügt, dass die CFA auf Ebene aller Subskalen inakzeptable Modellanpassungen erreichten, solange alle Items des KIMS als Indikatoren spezifiziert wurden. Aus Tabelle 4.3 ergibt sich, dass die Modellanpassungen der CFA auf Ebene der Subskalen mit den homogensten Items jeder Subskala, den Items des KIMS-Short, hingegen durchweg gut waren.

Zur Bestätigung der Faktorenstruktur des KIMS-Short wurde eine CFA mit allen Items des KIMS-Short durchgeführt, wobei entsprechend jeder Subskala ein latenter Faktor spezifiziert wurde, in Bezug auf die Skala BEO wurden zwei latente Faktoren spezifiziert. Korrelationen zwischen den Faktoren wurden zugelassen. Für diese CFA wurde Stichprobe A herangezogen. Es ergab sich eine gute Modellanpassung (vgl. Tabelle 4.3) und alle latenten Faktoren korrelierten signifikant miteinander außer BEO-IN bzw. BEO-AU mit AOB (vgl. Tabelle 4.4). Das Muster der Interkorrelationen entspricht den bisherigen Befunden (vgl. Ströhle et al., 2010). Festgehalten werden soll, dass eine CFA mit einem Faktor zweiter Ordnung nicht konvergierte (vgl. Höfling et al., under review; Anhang C).



Tabelle 4.2: Die Faktoren, Iteminhalte und Faktorladungen der Items des KIMS-Short auf Basis der CFA mit allen Items (Stichprobe A).

<b>Faktor</b>	<b>Itemnummer (KIMS)</b>	<b>Iteminhalt</b>	<b>Faktor- ladung</b>
<b>BEO-IN</b>	9	Wenn ich gehe, dann nehme ich ganz bewusst wahr, wie sich die Bewegungen meines Körpers anfühlen.	.53
	13	Wenn ich dusche oder bade, bin ich mir des Gefühls des Wassers auf meinem Körper bewusst.	.73
	21	Ich achte auf Empfindungen wie zum Beispiel Wind in meinem Haar oder Sonnenschein auf meinem Gesicht.	.82
<b>BEO-AU</b>	25	Ich achte auf Geräusche wie beispielsweise das Ticken von Uhren, Vogelzwitschern oder das Geräusch vorüber fahrender Autos.	.57
	29	Ich nehme Gerüche und Düfte der Dinge wahr.	.60
<b>BES</b>	33	Ich bemerke visuelle Elemente sowohl in der Kunst als auch in der Natur, zum Beispiel Farben, Formen, Struktur oder Muster aus Licht und Schatten.	.62
	2	Ich kann meine Gefühle gut in Worte fassen.	.72
	14	Es fällt mir schwer, das, was ich denke, in Worte zu fassen.	.71
	18	Ich habe Schwierigkeiten, die richtigen Worte zu finden, um meine Gefühle auszudrücken.	.82
	22	Körperliche Empfindungen sind für mich schwer zu beschreiben, weil mir die richtigen Worte dazu fehlen.	.62
	26	Sogar wenn ich schrecklich verärgert bin, kann ich das in Worte fassen.	.63
	7	Ich konzentriere mich nur auf das, was ich gerade tue, und auf nichts anderes.	.68
	19	Wenn ich etwas tue, dann bin ich davon völlig eingenommen und denke an nichts anderes mehr.	.65
	31	Ich neige dazu, mehrere Dinge gleichzeitig zu tun, anstatt mich nur auf eine Sache zu konzentrieren.	.55
	38	Wenn ich etwas tue, werde ich so davon eingenommen, dass meine ganze Aufmerksamkeit darauf gerichtet ist.	.61
<b>AOB</b>	4	Ich kritisiere mich dafür, irrationale oder unangebrachte Gefühle zu haben.	.74
	16	Ich glaube, dass einige meiner Gedanken unnormal sind, und dass ich nicht so denken sollte.	.86
	20	Ich urteile darüber, ob meine Gedanken gut oder schlecht sind.	.76
	28	Ich sage mir, dass ich nicht so denken sollte, wie ich denke.	.85
	32	Ich denke, dass manche meiner Gefühle schlecht oder unangebracht sind, und dass ich sie nicht haben sollte.	.90

*Anmerkung.* BEO-IN: Skala Beobachten internaler Stimuli; BEO-AU: Skala Beobachten externaler Stimuli; BES: Skala Tendenz, Erlebnisse in Worte zu fassen; MAH: Skala Mit Aufmerksamkeit handeln; AOB: Skala Akzeptieren ohne Bewerten.

Tabelle 4.3: Auflistung der Gütekriterien der Modellanpassungen der CFA für die vier Skalen des KIMS auf Basis der Items der jeweiligen Subskala des KIMS-Short in Stichprobe B (Zeilen BEO, BES, MAH und AOB). In der untersten Zeile sind die Gütekriterien einer CFA mit allen Items des KIMS-Short und korrelierten Faktoren aufgeführt (Stichprobe A).

	<b>SB-<math>\chi^2</math></b>	<i>df</i>	<b>SRMR</b>	<b>RMSEA</b>	<b>CFI</b>
<b>BEO</b>	11.74	8	.02	.03	.99
<b>BES</b>	2.75	4	.01	.00	1.00
<b>MAH</b>	1.66	1	.01	.04	1.00
<b>AOB</b>	7.75	5	.02	.03	1.00
<b>KIMS-Short</b>	235.14	162	.04	.03	.98

*Anmerkung.* BEO: CFA zur Skala Beobachten mit einem latenten Traitfaktor für BEO-IN und BEO-AU; BES: CFA zur Skala Tendenz, Erlebnisse in Worte zu fassen; MAH: CFA zur Skala Mit Aufmerksamkeit handeln; AOB: CFA zur Skala Akzeptieren ohne Bewerten; KIMS-Short: CFA zum KIMS-Short; SB- $\chi^2$ : Satorra-Bentler  $\chi^2$ ; SRMR: Standardized Root Mean Square Residual; RMSEA: Root Mean Square Error of Approximation; CFI: Comparative Fit Index.

Tabelle 4.4: Latente Faktorinterkorrelationen des KIMS-Short (Stichprobe A).

	<b>BEO-AU</b>	<b>BES</b>	<b>MAH</b>	<b>AOB</b>
<b>BEO-IN</b>	.74	.21	.14	–
<b>BEO-AU</b>		.22	.16	–
<b>BES</b>			.18	.30
<b>MAH</b>				.37

*Anmerkung.* BEO-IN: Skala Beschreiben internaler Stimuli; BEO-AU: Skala Beschreiben externaler Stimuli; BES: Skala Tendenz, Erlebnisse in Worte zu fassen; MAH: Skala Mit Aufmerksamkeit handeln; AOB: Skala Akzeptieren ohne Bewerten.

### 4.3 Konvergente und diskriminante Validität des KIMS-Short

Das multidimensionale KIMS-Short erfasst vier verschiedene Aspekte von Achtsamkeit. Zur Validierung der Skalen des KIMS-Short ist neben dem Nachweis faktorieller Validität der Nachweis konvergenter und diskriminanter Validität erforderlich. Insbesondere der Nachweis diskriminanter Validität zur Differenzierung der KIMS-Short-Skalen von anderen Skalen zur Erfassung von Persönlichkeitsvariablen oder Symptomen psychischer Belastung sollte geführt werden (vgl. Clark & Watson, 1995). Baer et al. (2004) konnten in Untersuchungen zum KIMS ( $N = 130$ ) zeigen, dass das Erleben psychischer Belastung bzw. Neurotizismus mit den Skalen BES, MAH und AOB ( $.29 \leq r \leq .42$ ), nicht aber mit BEO, negativ korrelierte, dass diese Korrelationen von mittlerer Größe waren und daher auf miteinander verbundene, aber unterschiedliche Konstrukte hindeuteten. Zwischen dem Erleben negativer Affekte und den Achtsamkeitsaspekten BES, MAH und AOB bestand ein moderater inverser Zusammenhang (vgl. Baer et al., 2004, S. 200f.). Extraversion war moderat positiv korreliert mit BES, was auf einen Zusammenhang zwischen der Tendenz, eigene Erlebnisse in Worte zu fassen und Gesprächigkeit bzw. Geselligkeit hinweist. Offenheit für Erfahrungen war ausschließlich mit BEO verbunden, d.h. dass Probanden mit höheren Werten in Offenheit für Erfahrungen Erlebnisse intensiver wahrnehmen. In den Untersuchungen von Baer et al. (2004) konnte die diskriminante Validität der vier Achtsamkeitsaspekte des KIMS belegt werden, insofern moderate, geringe bzw. keine Zusammenhänge zu Variablen aus den Bereichen Persönlichkeit oder Psychopathologie gefunden wurden. Die Korrelationen zwischen den Skalen des KIMS und der MAAS ergaben einen hohen Wert für die Korrelation zwischen der MAAS und MAH ( $r = .57$ ), was auf die starke inhaltliche Überschneidung zurückgeführt werden kann. Zwischen der MAAS und BES bzw. AOB zeigten sich moderate Korrelationen ( $r = .24$  bzw.  $r = .30$ ), kein Zusammenhang wurde zwischen der MAAS und BEO gefunden (vgl. Baer et al., 2004, S. 202).

Mit dem KIMS-Short wurde eine Kurzversion eines multidimensionalen Fragebogens zur Erfassung von Achtsamkeit entwickelt, der sich vom KIMS in zweierlei Hinsicht unterscheidet. Zum einen weisen die Skalen des KIMS-Short eine höhere Homogenität auf als das ursprüngliche KIMS und zum anderen wird in Bezug auf die Skala BEO im KIMS-Short zwischen BEO-IN und BEO-AU unterschieden. Die Korrelationen – basierend auf den Daten von Ströhle et al. (2010) – zwischen den Skalen des KIMS-Short und Achtsamkeit (MAAS), des KIMS-Short und verschiedenen Persönlichkeitsvariablen (NEO-FFI-Skalen) und des KIMS-Short und Skalen psychischer Belastung (BDI, BAI und RSQ-Rumination) sind in Tabelle 4.5 aufgeführt. Auf Basis der Stichprobe A zeigen sich zwischen allen Skalen des KIMS-Short und der MAAS signifikante Korrelationen, wobei die zwischen der MAAS und MAH erwartungsgemäß am höchsten ausfällt. Die Korrelationen zu den Persönlichkeitsvariablen und den Skalen psychischer Belastung liegen in weitgehender Übereinstim-

Tabelle 4.5: Korrelationen der fünf Faktoren des KIMS-Short mit der MAAS, mit Persönlichkeitsvariablen und Maßen psychischer Belastung.

Bereich	Fragebogen	Stichprobe	Skala	Skalen des KIMS-Short					
				BEO-IN	BEO-OUT	BES	MAH	AOB	
<b>Achtsamkeit</b>	MAAS	A		.19**	.19**	.27**	.46**	.41**	
<b>Persönlichkeit</b>	NEO-FFI	B.1	Neurotizismus	-.21**	-.20**	-.38**	-.35**	-.52**	
		B.1	Extraversion	.17**	.13**	.39**	.07	.26**	
		B.1	Offenheit	.33**	.39**	.14	.05	.13	
		B.1	Verträglichkeit	-.02	.13	.02	.13	.26**	
		B.1	Gewissenhaftigkeit	-.01	.03	.29**	.40**	.16**	
	SAM	B.1	Private Selbstaufmerksamkeit	.33**	.28**	.09	.04	-.24**	
		B.1	Öffentliche Selbstaufmerksamkeit	.05	.01	-.04	-.17**	-.39**	
<b>Psychopathologie</b>	BDI	B.1	Selbstkenntnis	.30**	.19**	.46**	.27**	.19**	
	BAI	B.2		-.15	.08	-.22**	-.14**	-.58**	
	RSQ	B.2		-.12	-.05	-.19**	-.08	-.52**	
		B.1	Symptombezogene Rumination	-.04	-.01	-.19**	-.22**	-.50**	

*Anmerkung.* MAAS: Mindful Attention and Awareness Scale; NEO-FFI: NEO-Fünf-Faktoren-Inventar; SAM: Fragebogen zur dispositionellen Selbstaufmerksamkeit; BDI: Becks Depressions-Inventar; BAI: Becks Angst-Inventar; RSQ: Response Styles Questionnaire; Stichprobe A:  $N = 602$ ; Stichprobe B.1:  $N = 253$ ; Stichprobe B.2:  $N = 216$ ; BEO-IN: Skala Beobachten interner Stimuli; BEO-AU: Skala Beobachten externer Stimuli; BES: Skala Tendenz, Ergebnisse in Worte zu fassen; MAH: Skala Mit Aufmerksamkeit handeln; AOB: Skala Akzeptieren ohne Bewerten; \*\* =  $p < .01$ .

mung zu den Ergebnissen von Baer et al. (2004).

In Bezug auf die BEO-Skalen ergeben sich hohe Korrelationen zu Offenheit für Erfahrungen und Privater Selbstaufmerksamkeit. Keine signifikanten Korrelationen ergeben sich zu den Skalen psychischer Belastung. Hinsichtlich Selbstkenntnis ist die Korrelation von BEO-IN deutlich höher als BEO-AU. Für BES ergibt sich die höchste Korrelation mit Selbstkenntnis, weitere Korrelationen liegen mit Neurotizismus (negativ) und Extraversion vor. Niedrige negative Korrelationen ergeben sich mit den Skalen psychischer Belastung. Für MAH ergeben sich die höchsten Korrelationen mit Neurotizismus (negativ) und Gewissenhaftigkeit. Weiterhin liegen signifikante negative Korrelationen mit depressiven Symptomen und Rumination vor. Die höchsten Korrelationen ergeben sich zwischen der Subskala AOB und den Skalen psychischer Belastung bzw. Neurotizismus.

# Kapitel 5

## Diskussion

In diesem Kapitel werden die Ergebnisse der vorherigen Kapitel zusammengetragen und diskutiert. Hierbei sollen zunächst die Fragestellungen zur Methodologie von MTMM-Analysen, zur Eindimensionalität und Homogenität der MAAS und zur Homogenität, faktoriellen Validität und Konstruktvalidität des KIMS zusammenfassend beantwortet werden (vgl. Kapitel 5.1). Anschließend sollen Grenzen der vorgelegten Arbeit benannt und ein kurzes Resümee (vgl. Kapitel 5.2) gezogen werden.

### 5.1 Zusammenfassende Beantwortung der Fragestellungen

#### 5.1.1 Zur Methodologie von MTMM-Analysen

Psychologische Maße unterliegen einer Vielzahl von Einflüssen, die deren Validität negativ beeinträchtigen können (vgl. Podsakoff et al., 2003). Hierzu gehören Einflüsse durch Messinstrumente (vgl. Kapitel 2.1), Einflüsse durch Beurteiler oder durch spezifische Bedingungen der Untersuchungssituation (vgl. Biesanz & West, 2004; Pohl, Steyer & Kraus, 2008; Schermelleh-Engel, Keith, Moosbrugger & Hodapp, 2004). Zum Nachweis konvergenter und diskriminanter Validität unter Berücksichtigung solcher Einflüsse wurde die Erhebung und Analyse von Multitrait-Multimethod (MTMM)-Daten vorgeschlagen (vgl. Campbell & Fiske, 1959).

Es steht eine Vielzahl statistischer Verfahren zur Analyse von MTMM-Daten zur Verfügung (vgl. Eid & Diener, 2006). Am häufigsten werden Modelle aus dem Kontext der konfirmatorischen Faktorenanalyse (CFA) eingesetzt, sog. MTMM-CFA-Modelle. Ein erstes Ziel der vorliegenden Arbeit bestand im Aufzeigen der Möglichkeiten, Methodeneffekte mittels MTMM-CFA-Modellen zu untersuchen. Hierbei wurde dargelegt, wie der ursprünglich deskriptive Ansatz der MTMM-Analyse anhand von vier Kriterien zur Überprüfung der konvergenten und diskriminanten Validität von Campbell und Fiske (1959) in den Kontext der CFA übersetzt werden kann (vgl. Höfling et al., 2009). Weiterhin wurde deutlich, dass ein im eigentlichen Sinne konfirmatorisches Vorgehen den Vergleich verschiedener MTMM-

CFA-Modelle erfordert. Daher wurde anhand der Taxonomie von Widaman (1985) gezeigt, wie auf systematische Art und Weise MTMM-CFA-Modelle spezifiziert und auf ihre Modellanpassung hin verglichen werden können. Schließlich wurde ein weiteres MTMM-CFA-Modell, das CT-C( $M-1$ )-Modell, vorgestellt und auf ein spezifisches Charakteristikum bei der Analyse von Methodeneffekten im Rahmen von Fragebogendaten eingegangen, nämlich die Unterscheidung von Trait-übergreifenden und Trait-spezifischen Methodeneffekten (vgl. Nussbeck et al., 2009). Die Darstellung der Analysemöglichkeiten von Methodeneffekten wurde insbesondere in den Kontext der Methodeneffekte gestellt, die im Rahmen von Fragebogendaten auftreten können (vgl. Podsakoff et al., 2003). Hierbei wurde besonderes Augenmerk auf den spezifischen Methodeneffekt der Itempolung gelegt und eine Analysemöglichkeit zu Itempolungseffekten vorgestellt (vgl. Horan et al., 2003). Im Wesentlichen kann in Bezug auf Fragestellung 1 festgehalten werden, dass 1) Methodeneffekte bei der Erhebung psychologischer Daten unvermeidbar sind, dass 2) die Untersuchung von Itempolungseffekten im Kontext des Nachweises der faktoriellen Validität bzw. der Eindimensionalität einer Skala Anwendung finden sollte und dass 3) Methodeneffekte mittels mehrerer alternativer apriorischer MTMM-CFA-Modelle untersucht werden sollten.

### 5.1.2 Eindimensionalität und Homogenität in der MAAS

Die gleichzeitige Darbietung positiv und negativ gepolter Items kann Antworttendenzen reduzieren und zu einer besseren Itembearbeitung beitragen (vgl. Krosnick, 1999). Daher wurden im Rahmen der vorliegenden Untersuchung 1) analog zur Studie von Brown und Ryan (2003) positiv gepolte Items formuliert und 2) eine modifizierte Version der MAAS (MAAS-Short) entwickelt, die 3) mittels MTMM-CFA-Modellen auf ihre Eindimensionalität und Homogenität hin untersucht wurde. Brown und Ryan (2003) verzichteten bei der MAAS auf positiv gepolte Items, weil sie von der konzeptuellen Äquivalenz von Achtsamkeit und Unachtsamkeit ausgingen und weil sich für die negativ gepolten Items durchgängig bessere psychometrische Kennwerte ergeben hätten (vgl. Kapitel 3.4.2).

Die modifizierte Version der MAAS (MAAS-Short) umfasst fünf negativ und fünf positiv gepolte Items und kann als eindimensional gelten. Die Analyse mehrerer alternativer Modelle (vgl. Höfling et al., in press; Kapitel 3.4.1) ergab, dass die MTMM-CFA-Modelle, welche einen Methodenfaktor für die Itempolung spezifizierten (Modell 5 und Modell 6), deskriptiv eine tendenziell bessere Modellanpassung zeigten als die CFA-Modelle, die keinen Methodenfaktor spezifizierten (Modell 7 und Modell 8). Infolge dessen kann von einem Einfluss der Itempolung in der MAAS-Short aufgrund der gleichzeitigen Darbietung negativ und positiv gepolter Items ausgegangen werden. Dies bedeutet in Bezug auf Fragestellung 2 auch, dass unter der Berücksichtigung der Itempolung Eindimensionalität hinsichtlich des zentralen Achtsamkeitsaspekts *mindful attention and awareness* (MAA) vorliegt.

Die Analysen ergaben ferner, dass die MAAS-Short ausreichend homogen ist. Dabei

erwiesen sich sowohl bei den positiv als auch bei den negativ gepolten Items die inhaltlich identischen Items (sog. *mirror items*) als die homogensten. Hierbei ist hervorzuheben, dass der Prozess der Identifikation der homogensten Items für die positiv und negativ gepolten Items getrennt vorgenommen wurde. Hierdurch sollte ein mögliches methodisches Artefakt verhindert werden, welches entstehen kann, wenn positiv und negativ gepolte Items zusammen analysiert werden. In diesem Fall wäre mit hohen Interkorrelationen aufgrund der sehr ähnlichen Itemformulierungen zu rechnen, welche die Identifikation der homogensten Items verzerren könnten. Die „separate“ Identifikation der inhaltlich identischen Items im durchgeführten ergebnisoffenen Prozess deutet vielmehr darauf hin, dass es sich bei diesen Items um die in Bezug auf MAA homogensten handelt.

Ein weiterer Aspekt betrifft die psychometrischen Kennwerte der Items der MAAS-Short. Vier der fünf positiv gepolten Items erhalten die niedrigsten Varianzen und sind etwas schiefer verteilt als die übrigen Items. Dies entspricht bisherigen Befunden, denen zufolge Probanden positiv gepolten Achtsamkeits-Items stärker zustimmen als sie negativ gepolte Achtsamkeits-Items ablehnen. Für Probanden ohne Erfahrung in Achtsamkeitspraxis scheinen insbesondere die positiv gepolten Achtsamkeits-Items eine etwas andere Bedeutung zu haben als für Probanden mit umfangreicher Achtsamkeits-Erfahrung (vgl. Grossman, 2008). Ein empirischer Beleg hierfür ergibt sich aus der Optimismus-Forschung, wo im Rahmen von IRT-Modellen gefunden wurde, dass positiv gepolte Items für Probanden mit hohen Werten in Optimismus besser differenzieren als negativ gepolte Items, die im unteren und mittleren Bereich überlegen waren (vgl. Rauch, Schweizer & Moosbrugger, 2008). Demzufolge könnte es auch für eine Achtsamkeitsskala wie die MAAS-Short sinnvoll sein, positiv gepolte Items einzubeziehen, um auch bei Probanden mit längerer Achtsamkeitspraxis gut differenzieren zu können. Hierzu wären weiterführende Studien, beispielsweise unter Verwendung von IRT-Modellen denkbar.

Für die negativ und möglicherweise besonders die positiv gepolten Items ist auf eine weitere, nicht unerhebliche Besonderheit hinzuweisen. Im Zusammenhang der MAAS gibt es Belege dafür, dass sich die Selbstbeurteilung eigener Achtsamkeit mit beginnender Achtsamkeitspraxis derart verändert, dass die Werte für Achtsamkeit zunächst abnehmen (vgl. Brown & Ryan, 2003, S. 843). Dies wird damit erklärt bzw. deutet darauf hin, dass für Probanden eine zumindest geringe Erfahrung in Achtsamkeitspraxis nötig ist, um überhaupt valide Einschätzungen ihrer Achtsamkeit vornehmen zu können. Die beginnende Achtsamkeitspraxis führt bei diesen Probanden dazu, dass sie mit ihrer erst „anfänglichen“ Achtsamkeit konfrontiert werden, was sich entsprechend auf die Selbstbeurteilung ihrer Achtsamkeit auswirkt, die sie nunmehr zurückhaltender einstufen.

Mit der MAAS-Short liegt eine ökonomische 10-Item-Version vor, die den zentralen Achtsamkeitsaspekts MAA reliabel erfasst, in Bezug auf die interne Konsistenz der ursprünglichen MAAS sogar überlegen ist ( $\alpha = .88$ ) und daher für die Anwendung in Studi-



en zur Achtsamkeit besonders geeignet erscheint. Darauf hingewiesen werden muss freilich, dass die Reduktion der Items zum Zwecke höherer Homogenität mit dem Verlust an inhaltlicher Breite des Konstrukts MAA einhergeht.

### 5.1.3 Eine Kurzversion der Deutschen Adaptation des KIMS (KIMS-Short)

Bisherige konfirmatorische Faktorenanalysen (CFA) zum KIMS wurden stets auf Basis von Item-Päckchen (*item parcels*) und nicht auf Basis einzelner Items durchgeführt. Dies gilt in Bezug auf die Originalversion (vgl. Baer et al., 2004) und in Bezug auf die Deutsche Adaptation (vgl. Ströhle et al., 2010). Die Verwendung von Item-Päckchen anstelle von einzelnen Items als Indikatoren in CFA führt zu Variablen mit verkleinerten Intervallen, zur Annäherung an die Normalverteilung und damit zu einer Verbesserung des Modellfits (vgl. Bandalos, 2002; Little, Cunningham, Shahar & Widaman, 2002). Es können sich aber auch Nachteile aus der Anwendung von Item-Päckchen wie Fehleinschätzungen bezüglich der Eindimensionalität bzw. Homogenität einer Skala ergeben (vgl. Bandalos, 2008; Little et al., 2002). Daher scheint es empfehlenswert, bei der Validierung von Skalen CFA auf Basis einzelner Items und nicht allein auf Basis von Item-Päckchen durchzuführen.

Mit der Zielsetzung einer ökonomischen Kurzversion eines multidimensionalen Achtsamkeitsinventars und möglichst homogener Subskalen (vgl. Fragestellung 3) wurden somit erstmalig verschiedene CFA zum KIMS auf Item-Ebene durchgeführt. Daraus ergab sich die 20-Item-Version des KIMS-Short. Drei der vier Subskalen des KIMS konnten für das KIMS-Short mit den jeweils homogensten Items bestätigt werden, das waren „Beschreiben“ (BES), „Mit Aufmerksamkeit handeln“ (MAH) und „Akzeptieren ohne Bewerten“ (AOB). Für die Subskala „Beobachten“ (BEO) konnte keine befriedigende Modellanpassung erreicht werden, was die Eindimensionalität dieser Subskala in Frage stellte. Die theoretisch angezeigte Untersuchung eines CFA-Modells mit zwei korrelierten Faktoren für das Wahrnehmen internaler (BEO-IN) und externaler (BEO-AU) Reize erreichte eine gute Modellanpassung. Daher ist beim KIMS-Short von fünf Faktoren auszugehen, wobei die Faktoren BEO-IN und BEO-AU erwartungsgemäß am höchsten miteinander verbunden sind.

Aus den Ergebnissen der durchgeführten Untersuchungen (vgl. Höfling et al., under review; Kapitel 4.3) ergeben sich hinreichende Belege für die konvergente und diskriminante Validität des KIMS-Short. Zwischen den Subskalen des KIMS-Short und der MAAS ergaben sich signifikante Korrelationen, wobei die Korrelation zwischen MAH und der MAAS erwartungsgemäß am höchsten ausfiel. Dennoch muss in diesem Zusammenhang darauf hingewiesen werden, dass die Korrelationen zwischen BEO-IN, BEO-AU bzw. BES und der MAAS nicht sehr hoch waren. Dies deutet darauf hin, dass die verschiedenen Aspekte von Achtsamkeit relativ unterschiedlich sind und Achtsamkeit aus der Perspektive ihrer psychometrischen Erfassung ein eher heterogenes Konstrukt darstellt.

Die Skalen BEO-IN und BEO-AU sind erwartungsgemäß gleichermaßen mit Offenheit für Erfahrungen und mit Privater Selbstaufmerksamkeit (Reflexion über innere Erlebnis-inhalte) verbunden. Probanden, die internale Reize (BEO-IN) stärker wahrnehmen, neigen eher zu ausgeprägterem Selbstkenntnis in ihrer Selbsteinschätzung als Probanden, die externe Reize (BEO-AU) stärker wahrnehmen. Die Beschäftigung mit das eigene Selbst betreffenden internalen Reizen führt somit tendenziell zu einer ausgeprägteren Selbstkenntnis, einer „Expertise“ in Bezug auf das eigene Selbst, als das möglicherweise durch die Beschäftigung mit externalen Reizen geschieht. Höhere Selbstkenntnis resultiert auch bei stärkerer Tendenz, Erlebnisse in Worte zu fassen (BES). BES geht erwartungsgemäß auch mit höherer Ausprägung in Extraversion einher, MAH wiederum mit Gewissenhaftigkeit, d.h. für das regelmäßige und aufmerksame Ausführen von Handlungen ist Disziplin und Gewissenhaftigkeit nicht nur hilfreich, sondern auch erforderlich. Höhere Ausprägung in MAH geht mit geringer depressiver Symptomatik und weniger Ruminationen, aber nicht mit weniger (oder mehr) Angstsymptomatik einher. Körperliche Ängste, wie sie mit dem BAI erfasst werden, sind unabhängig davon, ob jemand geringe oder hohe Werte in MAH erhält.

Die Subskala AOB zeigt die höchsten Korrelationen mit den Skalen psychischer Belastung. Diese vergleichsweise hohen Korrelationen könnten einerseits aufgrund ihrer Beschaffenheit stärker durch positive oder negative Affektivität beeinflusst sein als andere Skalen (vgl. Podsakoff et al., 2003). Zudem könnte diese hohen Korrelationen andererseits aufgrund von Itempolungseffekten artifiziell erhöht sein, da alle beteiligten Skalen ausschließlich negativ gepolte Items verwenden.

Mit der 20-Item-Version des KIMS-Short liegt ein ökonomischer multidimensionaler Achtsamkeitsfragebogen vor, dessen Subskalen als befriedigend homogen gelten können und der wie die MAAS-Short als Fragebogen im Rahmen von Studien zur Achtsamkeit geeignet erscheint. Die Reduktion der Itemanzahl geht auch bei den Subskalen des KIMS-Short mit leicht eingeschränkter inhaltlicher Breite einher.

## 5.2 Grenzen der vorliegenden Arbeit und Resümees

Im Rahmen der vorliegenden Arbeit müssen gewisse Grenzen bzw. Einschränkungen benannt werden. Zunächst fällt hierunter die Tatsache, dass nur ein Ausschnitt der methodologisch relevanten Analyseverfahren im Kontext von MTMM-Daten erörtert werden konnte. MTMM-Analysen intervallskalierten Daten werden häufig mittels CFA durchgeführt und wurden deshalb bevorzugt in dieser Arbeit behandelt. MTMM-Modelle zu nominal- bzw. ordinalskalierten Daten wurden genannt (vgl. Kapitel 2.4), konnten aber im Rahmen dieser Arbeit keine weitere Berücksichtigung finden. Insbesondere log-lineare- (vgl. Nussbeck, 2006) oder IRT-Modelle (vgl. Rost & Carstensen, 2002; Rost & Walter, 2006) stellen anspruchsvolle und vielversprechende Analyseverfahren für MTMM-Daten dar. Die

Auseinandersetzung mit Methodeneffekten in dieser Arbeit beschränkt sich weiterhin vorwiegend auf Itempolungseffekte und geht nicht näher auf mögliche andere Methodeneffekte in Fragebogendaten ein. Allein aus der Arbeit von Podsakoff et al. (2003) wird deutlich, mit welchen Methodeneffekten bei Fragebogendaten zu rechnen ist bzw. welche Methodeneffekte in Validierungsstudien berücksichtigt werden sollten. Erwähnt werden sollte auch, dass MTMM-Analysen einer sorgfältigen Planung und aufwendigen Durchführung bedürfen. Wenn auch MTMM-Analysen zur Validierung psychometrischer Erfassungsinstrumente also sehr wünschenswert sind, erfordern sie doch einen hohen Arbeitsaufwand.

Eine weitere Einschränkung der vorliegenden Arbeit liegt in der Auswahl zweier bedeutsamer Achtsamkeitsskalen zur Modifikation bzw. weiteren Validierung. Dadurch wird auf die übrigen Fragebogen zur Achtsamkeit nicht eingegangen, die nicht nur wesentliche Aspekte von Achtsamkeit erfassen, sondern prinzipiell auch für Validierungsstudien geeignet gewesen wären. Hierbei muss insbesondere auf den bisher noch nicht erwähnten *Five Facet Mindfulness Questionnaire* (FFMQ; Baer, Smith, Hopkins, Krietemeyer & Toney, 2006) hingewiesen werden. Dieser wurde entwickelt, indem die Autoren einige der vorliegenden Achtsamkeitsskalen mit einer einheitlichen Ratingskala versahen und Probanden zur Bearbeitung vorlegten. Mittels exploratorischer Faktorenanalysen (EFA) und CFA wurden fünf Facetten von Achtsamkeit bestimmt, zu den vier Faktoren des KIMS kam ein weiterer Faktor „Nicht-Reagieren auf internale Erlebnisinhalte“ (*non-reactivity to inner experiences*) hinzu. Das FFMQ wurde in der vorliegenden Arbeit aus zwei Gründen nicht weiter berücksichtigt. Zum einen gibt es keine Deutsche Adaptation und damit keine Möglichkeit der Anwendung dieses Instruments im deutschen Sprachraum. Zum anderen besteht der Mehrgewinn des FFMQ gegenüber dem KIMS lediglich in der Erfassung einer weiteren Facette, zu der zwei Aspekte zu bemerken sind. Theoretisch ist die Facette des Nicht-Reagierens in Bezug auf das Verständnis von Achtsamkeit sehr plausibel (vgl. Teasdale, 1988; Kapitel 1.1; Kapitel 1.2), aus der Perspektive möglicher Methodeneffekte (aufgrund gleichlautender Itemformulierung) muss die Existenz dieses Faktors zumindest etwas in Frage gestellt werden. Vier der sieben dort zugeordneten Items formulieren nämlich jeweils zur Hälfte identisch aufgrund des gleichlautenden Beginns „Usually when I have distressing thoughts or images (...)“. Angeführt werden soll schließlich, dass die Ergebnisse zur MAAS-Short und zum KIMS-Short nicht direkt auf die englischsprachigen Originalversionen übertragen werden können. Hierzu wären Datenerhebungen und Analysen auf der Basis der Originalversionen notwendig.

Die vorliegende Arbeit umfasst einen methodologischen und einen empirischen bzw. anwendungsbezogenen Teil. Im methodologischen Teil werden wesentliche Aspekte zum Auftreten von Methodeneffekten in psychologischen Daten und zur Analyse von MTMM-Daten aufgezeigt. Im empirischen bzw. anwendungsbezogenen Teil werden die Ergebnisse von Validierungsstudien zu zwei bedeutsamen Achtsamkeitsskalen, den Deutschen Adap-

tationen der *Mindful Attention and Awareness Scale* (MAAS) und des *Kentucky Inventory of Mindfulness Skills* (KIMS), dargestellt. Diese Studien führten unter anderem zur Modifikation dieser beiden Skalen, so dass nun zwei ökonomische und in befriedigendem Maße homogene Achtsamkeitsfragebogen für die Anwendung im Rahmen von Forschungsfragenstellungen zur Verfügung stehen: Die MAAS-Short als modifizierte Version der MAAS und das KIMS-Short als modifizierte Kurzversion des multidimensionalen KIMS, wobei das KIMS-Short fünf anstelle von vier Faktoren im KIMS erfasst.

Vielfältige achtsamkeitsbasierte Interventionen sind bereits in Anwendung, deren Wirksamkeit nachgewiesen werden konnte (vgl. Kapitel 1.2). Die vorliegende Arbeit versteht sich als Beitrag zur weiteren Klärung der Frage, inwieweit Achtsamkeit im Rahmen solcher Interventionen bzw. eigener Achtsamkeitspraxis veränderbar ist und inwieweit dergleichen Veränderungen hinsichtlich Achtsamkeit den positiven Effekt achtsamkeitsbasierter Interventionen erklären können. Daher bleibt zu wünschen, dass die MAAS-Short und die KIMS-Short als ökonomische, reliable und valide Messinstrumente zur psychometrischen Erfassung von Achtsamkeit in diesem Sinne Anwendung finden und Hilfe sein können.

# Literaturverzeichnis

- Baer, R. A. (2003). Mindfulness training as a clinical intervention: A conceptual and empirical review. *Clinical Psychology: Science and Practice, 10*, 125–143.
- Baer, R. A., Smith, G. T. & Allen, K. B. (2004). Assessment of mindfulness by self-report: The Kentucky Inventory of Mindfulness Skills. *Assessment, 11*, 191–206.
- Baer, R. A., Smith, G. T., Hopkins, J., Krietemeyer, J. & Toney, L. (2006). Using self-report assessment methods to explore facets of mindfulness. *Assessment, 13*, 27–45.
- Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling, 9*, 78–102.
- Bandalos, D. L. (2008). Is parceling really necessary? A comparison of results from item parceling and categorical variable methodology. *Structural Equation Modeling, 15*, 211–240.
- Beck, A. T. (1967). *Depression: Causes and treatment*. Philadelphia: University of Pennsylvania Press.
- Biesanz, J. C. & West, S. G. (2004). Towards understanding assessments of the Big Five: Multitrait-multimethod analyses of convergent and discriminant validity across measurement occasion and type of observer. *Journal of Personality, 72*, 845–876.
- Bishop, S. R., Lau, M., Shapiro, S., Carlson, L., Anderson, N. D., Carmody, J., Segal, Z. V., Abbey, S., Speca, M., Velting, D. & Devins, G. (2004). Mindfulness: A proposed operational definition. *Clinical Psychology: Science and Practice, 11*, 230–241.
- Bohus, M., Haaf, B., Simms, T., Limberger, M. F., Schmahl, C., Unkel, C. & Linehan, M. M. (2004). Effectiveness of inpatient dialectical behavioral therapy for borderline personality disorder: a controlled trial. *Behavior Research and Therapy, 42*, 487–499.
- Borkenau, P. & Ostendorf, F. (1993). *NEO-Fünf-Faktoren-Inventar (NEO-FFI) nach Costa und McCrae*. Göttingen: Hogrefe.
- Brown, K. W. & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology, 84*, 822–848.
- Buchheld, N. & Walach, H. (2004). Die historischen Wurzeln der Achtsamkeitsmeditation – Ein Exkurs in Buddhismus und christliche Mystik. In T. Heidenreich & J. Michalak

- (Hrsg.), *Achtsamkeit und Akzeptanz in der Psychotherapie* (S. 25–46). Tübingen: Deutsche Gesellschaft für Verhaltenstherapie.
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105.
- Clark, L. A. & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*, 309–319.
- Davidson, R. J., Kabat-Zinn, J., Schumacher, J., Rosenkranz, M., Muller, D., Santorelli, S. F., Urbanowski, F., Harrington, A., Bonus, K. & Sheridan, J. F. (2003). Alterations in brain and immune function produced by mindfulness meditation. *Psychosomatic Medicine*, *65*, 564–570.
- DiStefano, C. & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling*, *13*, 440–464.
- Dumenci, L. (2000). Multitrait-multimethod analysis. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 583–611). San Diego, CA: Academic Press.
- Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika*, *65*, 241–261.
- Eid, M. & Diener, E. (2006). Introduction: The need for multimethod measurement in psychology. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 3–9). Washington, DC: APA.
- Eid, M., Lischetzke, T. & Nussbeck, F. W. (2006). Structural equation models for multitrait-multimethod data. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 283–299). Washington, DC: APA.
- Eid, M., Nussbeck, F. W., Geiser, C., Cole, D. A., Gollwitzer, M. & Lischetzke, T. (2008). Structural equation modelling of multitrait-multimethod data: Different models for different types of methods. *Psychological Methods*, *13*, 230–253.
- Filipp, S.-H. & Freudenberg, E. (1989). *Fragebogen zur Erfassung dispositionaler Selbstaufmerksamkeit (SAM)*. Handanweisung. Göttingen: Hogrefe.
- Forman, E. M., Herbert, J. D., Moitra, E., Yeomans, P. D. & Geller, P. A. (2007). A randomized controlled effectiveness trial of Acceptance and Commitment Therapy and Cognitive Therapy for anxiety and depression. *Behavior Modification*, *31*, 772–799.
- Gioia, D. A. & Sims, H. P., Jr. (1985). On avoiding the influence of implicit leadership theories in leader behavior descriptions. *Educational and Psychological Measurement*, *45*, 217–232.
- Grayson, D. & Marsh, H. W. (1994). Identification with deficient rankloading matrices in confirmatory factor analysis: Multitrait-multimethod models. *Psychometrika*, *59*,

121–134.

- Grossmann, P. (2008). On measuring mindfulness in psychosomatic and psychological research. *Journal of Psychosomatic Research*, *64*, 405–408.
- Harrison, D. A. & McLaughlin, M. E. (1993). Cognitive processes in self-report responses: Tests of item context effects in work attitude measures. *Journal of Applied Psychology*, *78*, 129–140.
- Harvey, R. J., Billings, R. S. & Nilan, K. J. (1985). Confirmatory factor analysis of the job diagnostic survey: Good news and bad news. *Journal of Applied Psychology*, *70*, 461–468.
- Hautzinger, M., Bailer, M., Worall, H. & Keller, F. (1994). *Beck-Depressions-Inventar*. Bern: Hans Huber Verlag.
- Hayes, S. C., Strosahl, K. & Wilson, K. G. (1999). *Acceptance and Commitment Therapy: An experiential approach to behavior change*. New York: Guilford Press.
- Hayes, S. C., Strosahl, K. D., Wilson, K. G., Bissett, R. T., Pistorello, J., Toarmino, D., Polusny, M. A., Dykstra, T. A., Batten, S. V., Bergan, J., Stewart, S. H., Zvolensky, M. J., Eifert, G. H., Bond, F. W., Forsyth, J. P., Karekla, M. & McCurry, S. M. (2004). Measuring experiential avoidance: A preliminary test of a working model. *The Psychological Record*, *54*, 553–578.
- Heidenreich, T. & Michalak, J. (2003). Achtsamkeit (mindfulness) als Therapieprinzip in Verhaltenstherapie und Verhaltensmedizin. *Verhaltenstherapie*, *13*, 264–274.
- Höfling, V., Schermelleh-Engel, K. & Moosbrugger, H. (2009). Analyzing multitrait-multimethod data: A comparison of three approaches. *Methodology*, *5*, 99–111.
- Hofmann, S. G., Sawyer, A. T., Witt, A. A. & Oh, D. (2010). The effect of mindfulness-based therapy on anxiety and depression: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, *78*, 169–183.
- Horan, P. M., DiStefano, C. & Motl, R. W. (2003). Wording effects in Self-Esteem Scales: Methodological artifact or response style? *Structural Equation Modeling*, *10*, 435–455.
- Kabat-Zinn, J. (1990). *Full catastrophe living: Using the wisdom of your body and mind to face stress, pain, and illness*. New York: Bantam Dell.
- Kabat-Zinn (2003). Mindfulness-based interventions in context: past, present, and future. *Clinical Psychology: Science and Practice*, *10*, 144–156.
- Kenny, D. A. (1976). An empirical application of confirmatory factor analysis to the multitrait-multimethod matrix. *Journal of Experimental Social Psychology*, *12*, 247–252.
- Kühner, C., Huffziger, S. & Nolen-Hoeksema, S. (2007). *Response Styles Questionnaire – Deutsche Version (RSQ-D)*. Manual. Göttingen: Hogrefe.
- Lau, M. A., Bishop, S. R., Segal, Z. V., Buis, T., Anderson, N. D., Carlson, L., Shapiro, S.

- & Carmody, J. (2006). The Toronto Mindfulness Scale: development and validation. *Journal of Clinical Psychology, 62*, 1445–1467.
- Linehan, M. M. (1993). *Cognitive behavioral treatment of Borderline Personality Disorder*. New York: Guilford Press.
- Little, T. D., Cunningham, W. A., Shahar, G. & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighting the merits. *Structural Equation Modeling, 9*, 151–173.
- Ma, S. H. & Teasdale, J. D. (2004). Mindfulness-based cognitive therapy for depression: Replication and exploration of differential relapse prevention effects. *Journal of Consulting and Clinical Psychology, 72*, 31–40.
- MacKillop, J. & Anderson, E. J. (2007). Further psychometric validation of the Mindful Attention Awareness Scale (MAAS). *Journal of Psychopathology and Behavioral Assessment, 29*, 289–293.
- Margraf, J. & Ehlers, A. (2002). *Beck Angstinventar – Deutsche Version (BAI)*. Hogrefe: Göttingen.
- Marsh, H. W. (1989). Confirmatory factor analysis of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement, 12*, 335–361.
- Marsh, H. W. & Bailey, M. (1991). Confirmatory factor analysis of multitrait-multimethod data: A comparison of alternative models. *Applied Psychological Measurement, 15*, 47–70.
- Marsh, H. W. & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 177–198). Thousand Oaks: Sage.
- Michalak, J., Heidenreich, T., Ströhle, G. & Nachtigall, C. (2008). Die deutsche Version der Mindful Attention and Awareness Scale (MAAS). *Zeitschrift für Klinische Psychologie und Psychotherapie, 37*, 200–208.
- Moosbrugger, H. (2007). Klassische Testtheorie (KTT). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 99–112). Heidelberg: Springer.
- Muthén, L. K. & Muthén, B. O. (2006). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.
- Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology, 15*, 263–280.
- Nussbeck, F. W. (2006). Assessing multimethod association with categorical variables. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 231–247). Washington, DC: APA.
- Nussbeck, F. W., Eid, M., Geiser, C., Courvoisier, D. S. & Lischetzke, T. (2009). A CTC(M-1) model for different types of raters. *Methodology, 5*, 88–98.
- Nyanaponika (2000). *Geistestraining durch Achtsamkeit*. Stammbach: Beyerlein & Stein-



schulte.

- Peterson, R. A. (2000). *Constructing effective questionnaires*. Thousand Oaks, CA: Sage.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y. & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879–903.
- Pohl, S., Steyer, R. & Kraus, K. (2008). Modelling method effects as individual causal effects. *Journal of the Royal Statistical Society, Series A, 171*, 41–63.
- Quilty, L. C., Oakman, J. M. & Risko, E. (2006). Correlates of the Rosenberg Self-Esteem Scale method effects. *Structural Equation Modeling, 13*, 99–117.
- Rauch, W. A., Schweizer, K. & Moosbrugger, H. (2007). Method effects due to social desirability as a parsimonious explanation of the deviation from unidimensionality in LOT-R scores. *Personality and Individual Differences, 42*, 1597–1607.
- Roemer, L. & Orsillo, S. M. (2002). Expanding our conceptualization of and treatment for generalized anxiety disorder: Integrating mindfulness/acceptance-based approaches with existing cognitive-behavioral models. *Clinical Psychology: Science and Practice, 9*, 54–68.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Rost, J. & Carstensen, C. (2002). Multidimensional Rasch measurement via item component models and faceted designs. *Applied Psychological Measurement, 26*, 42–56.
- Rost, J. & Walter, O. (2006). Multimethod item response theory. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 249–268). Washington, DC: APA.
- Salancik, G. R. (1984). On priming, consistency, and order effects in job attitude assessment: With a note on current research. *Journal of Management, 10*, 250–254.
- Salmon, P., Santorelli, S. F. & Kabat-Zinn, J. (1998). Intervention elements promoting high adherence to mindfulness-based stress reduction programs in the clinical behavioral medicine setting. In S. A. Shumaker, E. B. Schron, J. K. Ockene & W. L. McBee (Eds.), *Handbook of health behavior change* (2nd ed., pp. 239–266). New York: Springer.
- Schermelleh-Engel, K., Keith, N., Moosbrugger, H. & Hodapp, V. (2004). Decomposing person and occasion-specific effects: An extension of latent state-trait theory to hierarchical models. *Psychological Methods, 9*, 198–219.
- Schermelleh-Engel, K., Moosbrugger, H. & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*, 23–74.
- Schermelleh-Engel, K. & Schweizer, K. (2007). Multitrait-Multimethod-Analysen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 325–

- 342). Heidelberg: Springer.
- Schweizer, K. (2010). Some guidelines concerning the modeling of traits and abilities in test construction. *European Journal of Psychological Assessment, 26*, 1–2.
- Schweizer, K. & Rauch, W. A. (2008). An investigation of the structure of the Social Optimism Scale with respect to the dimensionality problem. *Journal of Individual Differences, 29*, 223–230.
- Segal, Z. V., Williams, J. M. G. & Teasdale, J. D. (2001). *Mindfulness-based cognitive therapy for depression: A new approach to preventing relapse*. New York: The Guilford Press.
- Spector, P. E., Van Katwyk, P. T., Brannick, M. T. & Chen, P. Y. (1997). When two factors don't reflect two constructs: How item characteristics can produce artifactual factors. *Journal of Management, 23*, 659–677.
- Ströhle, G., Nachtigall, C., Michalak, J. & Heidenreich, T. (2010). Die Erfassung von Achtsamkeit als mehrdimensionales Konstrukt: Die deutsche Version des Kentucky Inventory of Mindfulness Skills (KIMS). *Zeitschrift für Klinische Psychologie und Psychotherapie, 39*, 1–12.
- Teasdale, J. D. (1988). Cognitive vulnerability to persistent depression. *Cognition and Emotion, 2*, 247–274.
- Teasdale, J. D., Moore, R. G., Hayhurst, H., Pope, M., Williams, S. & Segal, Z. V. (2002). Metacognitive awareness and prevention of relapse in depression: Empirical evidence. *Journal of Consulting and Clinical Psychology, 70*, 278–287.
- Teasdale, J. D., Segal, Z. V., Williams, J. M. G., Ridgeway, V., Soulsby, J. & Lau, M. (2000). Prevention of relapse/recurrence in major depression by mindfulness-based cognitive therapy. *Journal of Consulting and Clinical Psychology, 68*, 615–623.
- Thomas, K. W. & Kilmann, R. H. (1975). The social desirability variable in organizational research: An alternative explanation for reported findings. *Academy of Management Journal, 18*, 741–752.
- Tourangeau, R., Rips, L. J. & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, England: Cambridge University Press.
- Vautier, S., Steyer, R., Jmel, S. & Raufaste, E. (2005). Imperfect or perfect dynamic bipolarity? The case of antonymous affective judgments. *Structural Equation Modeling, 12*, 391–410.
- Walach, H., Buchheld, N., Buttenmüller, V., Kleinknecht, N. & Schmidt, S. (2006). Measuring mindfulness – The Freiburg Mindfulness Inventory (FMI). *Personality and Individual Differences, 40*, 1543–1555.
- Widaman, K.-F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement, 9*, 1–26.
- Winkler, J. D., Kanouse, D. E. & Ware, J. E., Jr. (1982). Controlling for acquiescence response set in scale development. *Journal of Applied Psychology, 67*, 555–561.

## Anhang A

### Schrift (A)

**Höfling, V.**, Schermelleh-Engel, K., & Moosbrugger, H. (2009). Analyzing multitrait-multimethod data: A comparison of three approaches. *Methodology*, *5*, 99–111.

# Analyzing Multitrait-Multimethod Data

## A Comparison of Three Approaches

Volkmar Höfling, Karin Schermelleh-Engel, and Helfried Moosbrugger

Department of Psychological Research Methods and Evaluation, Goethe University, Frankfurt, Germany

**Abstract.** Assessing construct validity is a core task in psychology. Since Campbell and Fiske's (1959) seminal article on multitrait-multimethod (MTMM) analysis, several different methodological approaches for the analysis of convergent and discriminant validity of MTMM data have been developed. In this article, two MTMM approaches are transferred to the general framework of confirmatory factor analysis and compared with the extended version of the correlated trait-correlated method minus one model (Nussbeck, Eid, Geiser, Courvoisier, & Lischetzke, 2009): The multilevel MTMM model (Maas, Lensvelt-Mulders, & Hox, 2009) and the three-mode model (Oort, 2009). Assessing the construct validity of a German Big Five MTMM data set these three MTMM approaches are compared with regard to convergent and discriminant validity estimates and with regard to method effects. Advantages and limitations of each methodological approach will be discussed in detail.

**Keywords:** convergent validity, discriminant validity, multitrait-multimethod analysis, multilevel analysis, three-mode factor model, correlated trait-correlated method minus one model, confirmatory factor analysis

Campbell and Fiske's (1959) seminal article on the *Convergent and discriminant validation by the multitrait-multimethod (MTMM) matrix* had an immense impact on the process of exploring construct validity of psychological measures. They proposed to inspect the MTMM matrix, a matrix of correlations between several traits that were measured by several methods, to verify the construct validity of psychological measures. In Campbell and Fiske's perspective the core aspects of construct validity are convergent and discriminant validity. Convergent validity can only be assumed if different methods converge in the measurement of one trait. Discriminant validity, however, should be demonstrated by the nonconvergence of measurements of different traits. The application of at least two different methods for the measurement of one or more traits is necessary to separate trait and method influences for each score of a psychological variable (= trait-method unit, TMU). However, within the Campbell and Fiske approach trait- and method-specific influences cannot be separated from unsystematic measurement errors. The separation of measurement error from trait- and method-specific influences can be obtained by appropriate modern methodological approaches. Therefore, the focus of modern MTMM studies is not only on the proof of convergent and discriminant validity, but also on the separation of trait- and method-specific effects from measurement error (Eid & Diener, 2006).

A variety of statistical approaches to determine convergent and discriminant validity and to assess method-specific influences have been developed. Methodological approaches differ with regard to the scaling of the variables (e.g.,

categorical or metric variables). Log-linear models (Nussbeck, 2006) and MTMM models of item response theory (Rost & Carstensen, 2002; Rost & Walter, 2006) can be applied for the MTMM analysis of categorical data. For the MTMM analysis of metrical variables multilevel models (ML-CFA; Bryk & Raudenbusch, 1992; Hox, 2002; Hox & Maas, 2006), models of confirmatory factor analysis (CFA) or structural equation modeling (SEM), and three-mode factor (TMF) models (Bentler & Lee, 1979; Bloxom, 1968; Oort, 1999) are available. Especially MTMM models of CFA and SEM are the most widely applied statistical approaches for MTMM data analysis (Dumenci, 2000; Eid, Lischetzke, & Nussbeck, 2006; Schermelleh-Engel & Schweizer, 2007). CFA-MTMM models can be specified with single or multiple indicators, but only the use of multiple indicators guarantees the separation of measurement error and trait-specific method influences (Eid, 2000; Marsh & Hocevar, 1988).

Widaman (1985) developed a taxonomy of 16 CFA-MTMM models with four different types of trait structures (no trait factor, general trait factor, orthogonal trait factors, and oblique trait factors) and four different types of method structures (no method factor, general method factor, orthogonal method factors, and oblique method factors). Three of these 16 CFA-MTMM models are most often applied: In the correlated trait-correlated uniqueness (CT-CU) model (Kenny, 1976) correlations between traits are allowed and the residual variables of the indicators of the same method may covary. The correlated trait-uncorrelated method (CT-UM) model (Marsh, 1989; Marsh & Grayson, 1995)

establishes latent factors for all methods with no correlations between these method factors. A less restrictive version of the CT-UM model is the correlated trait-correlated method (CT-CM) model (Marsh & Grayson, 1995) that allows correlations between method factors. The correlated trait-correlated method minus one [CT-C( $M-1$ )] model (Eid, 2000; Eid, Lischetzke, Nussbeck, & Trierweiler, 2003) specifies one method factor less than methods included in the model. One method has to be chosen as the reference method; for this reference method no method factor is specified, whereas for the other methods (nonreference methods) method factors have to be modeled that may or may not be correlated. Depending on the choice of a CFA-MTMM model the interpretation of parameter estimates of traits and methods differs enormously. If researchers are interested in traits as common factors and methods as deviations from this “common view”, CT-CU and CT-UM models seem to be most appropriate. Compared to the CT-CU model the CT-UM model implies a much stronger assumption, that is, unidimensionality of the method influences belonging to the same method. If researchers are interested in contrasting methods because they assume different method-specific influences, the CT-C( $M-1$ ) model should be applied (Eid et al., 2006).

Research with CFA-MTMM models often seems to be somewhat “blind”, following the rule that the model with the best fit to the data should be preferred. However, researchers should consider which CFA-MTMM model is most appropriate for their research interests. Two key aspects have to be considered if MTMM models are estimated and evaluated: (1) Which types of methods can be found in the MTMM data set, and (2) Which MTMM model may provide answers to the specific research questions?

Regarding aspect 1, researchers have to decide whether the methods in their MTMM design are interchangeable or structurally different. Interchangeable methods could be – for example – different raters assessing a person X’s behavior; in this case the raters are randomly drawn from a specific population and no structural difference between the raters should exist as they all should have the same (or at least no individually specific) relationship to person X. Structurally different methods, on the other hand, could be raters with basically different perspectives to person X’s behavior. This may be true for different types of raters with individually specific relationships to person X, for example, relatives like parents, children, or spouses (Eid, 2006).

The distinction between interchangeable and structurally different methods is important for the choice of the appropriate methodological approach for MTMM data. In their basic approach, Campbell and Fiske (1959) recommended the use of structurally different methods rather than interchangeable methods in order to have a substantial test of convergent validity, since structurally different methods provide information about a trait from different perspectives (e.g., mother and teacher ratings of a child’s behavior). In more recent approaches (e.g., Eid et al., 2008), researchers were additionally interested in differences between methods and in possible explanations of these differences. It might also be of interest to examine common method influences for interchangeable (e.g., student ratings of a teacher) or closely related (e.g., mother and father ratings of their child’s behav-

ior) methods, in order to determine to which degree these methods yield identical results. The differences between interchangeable methods are presumably lower than in the case of structurally different methods albeit interchangeable methods will generally not lead to identical results. It might be of interest to determine covariates that warrant close ratings for interchangeable raters or that may explain why raters differ in their ratings.

Regarding aspect 2 on the appropriateness of MTMM models to provide answers for specific research questions, researchers should become familiar with the meaning of the results from different statistical MTMM approaches. Therefore, it can be helpful to analyze the same MTMM data with different methodological approaches and to compare the findings with regard to two research questions:

1. What kind of information do the different MTMM approaches offer regarding convergent and discriminant validity?
2. How do the different MTMM approaches specify method influences and how can they be interpreted?

This article is organized as follows: First, we will describe the specifics of the MTMM data set that will be analyzed by all three methodological MTMM approaches. Second, we will briefly summarize the statistical procedure of each of the approaches considering especially the research questions one and two. We will start with the multilevel analysis, we will continue with the TMF model, and we will conclude with the CT-C( $M-1$ ) model. Third, we will evaluate the three MTMM approaches with regard to convergent and discriminant validity, reliability, and method-specific effects.

## Data Set

The data for the three MTMM approaches are taken from the German Big Five MTMM data set (Eid, Lischetzke, Nussbeck, & Geiser, 2004). It contains three personality traits (extraversion, neuroticism, and conscientiousness) assessed with three different types of rating (self-report, peer report A, and peer report B) and each TMU is measured by four items with rating scales of five categories. Four hundred eighty-one participants were randomly selected from a population of students at the University of Trier and the University of Applied Sciences of Trier (Germany). Every participant was asked to bring in two peers; all answered the same questionnaire, either the self-report or the peer report version.

The MTMM data set has some specifics that make it very suitable for the examination and comparison of different MTMM approaches. First, every TMU is measured by four items. Therefore, the underlying design is a multiple and not a single indicator MTMM design with the above mentioned advantage, that trait effects as well as method-specific effects can be separated from measurement error. Second, regarding the type of methods in the data set, both interchangeable and structurally different methods were

available. Peer report A and peer report B were chosen randomly and there were no signs of systematic structural differences between peer A and peer B. Hence, both peer reports are interchangeable but not structurally different. Compared with the self-report, the peer report is certainly structurally different as peers show a substantially different perspective than the assessed person for the assessment of personality traits in relation to the target person. It can be presumed that particularly the information used for the assessment of personality traits differs structurally between self-reports and peer reports as a target person utilizes substantially different information than peers (e.g., cognitive or emotional experiences vs. behavioral signs).

## The Three Different MTMM Approaches

Three different MTMM approaches will be compared regarding convergent and discriminant validity, method-specific influences, and the suitability to the delineated MTMM data set. For an easier understanding and for an easier contrasting of the three different MTMM approaches, the multilevel MTMM model (Maas, Lensvelt-Mulders, & Hox, 2009) and the TMF model (Oort, 2009) will be transferred to the framework of CFA. Thus, these two MTMM approaches can be easily compared with the extended version of the CT-C( $M-1$ ) model (Nussbeck, Eid, Geiser, Courvosier, & Lischetzke, 2009).

### Multilevel MTMM Analysis

Maas et al. (2009) provided multilevel MTMM analyses for the above delineated MTMM data set. In their article, they picture the essentials of the multilevel approach in contrast to ordinary CFA, apply multilevel MTMM models, and discuss these models on the basis of the results. Though the multilevel approach is a univariate approach originally, it can be extended to multivariate ML-CFA. First, we introduce the CFA framework for multilevel MTMM analyses. Second, we present the ML-CFA-MTMM model Maas et al. (2009) applied to the MTMM data set. Third, the results of the ML-CFA-MTMM analyses will be interpreted.

### CFA Framework for the Multilevel MTMM Analysis

A subscript with five different indices for the evaluation of the ML-CFA-MTMM analysis according to the notation of Snijders and Bosker (1999) and Eid et al. (2008) is needed. The level-1 unit is denoted with  $r$  (the rater), and the level-2 unit is denoted with  $t$  (the target). Furthermore,  $k$  stands for the traits,  $m$  stands for the methods, and  $i$  denotes the indicator measuring the TMU. Thus, the ML-CFA-MTMM model can be defined by

$$Y_{rtikm} = \mu_{ikm} + \lambda_{Tikm} \cdot T_{ik} + \lambda_{Mikm} \cdot M_{rmm} + E_{rtikm}, \quad (1)$$

where  $Y_{rtikm}$  is the outcome variable, the score assessed by rater  $r$  for target  $t$  on indicator  $i$  of trait  $k$  measured with method  $m$ ,  $\mu_{ikm}$  is the indicator-specific mean (intercept) of the corresponding trait  $k$  measured with method  $m$ ,  $T_{ik}$  is the mean-centered trait factor for target  $t$  in trait  $k$ ,  $M_{rmm}$  is the method factor  $m$  specifying the stable method effect of rater  $r$  across traits, and  $E_{rtikm}$  is the measurement error of indicator  $i$ . For the factor loadings a further index is added to the subscript ( $T$  or  $M$ ) indicating the correspondence to the trait or method factors:  $\lambda_{Tikm}$  is the factor loading on the trait factor  $T_{ik}$  and  $\lambda_{Mikm}$  is the factor loading on the method factor  $M_{rmm}$ . In this model, it is assumed that there is only one trait factor per construct; traits are assumed to be homogeneous across indicators and methods. And it is assumed that there is only one method factor per method; method effects are assumed to be unidimensional within traits and to generalize across traits.

### Application of the ML-CFA-MTMM Model

The ML-CFA-MTMM design is characterized by a nested, two-level pattern and by the above stated model parameters (cf. Table 1). Three ratings  $r$  (three raters) are nested within one assessed subject  $t$  (the target). Regarding the ML-CFA-MTMM models of Maas et al. (2009) for the delineated MTMM data set (cf. Figure 1), there is one method factor  $M_{rmm}$  for each type of rating  $m$  (self-report, peer report A, and peer report B) on level-1. The method factor  $M_{rmm}$  reflects the assumption that the method-specific influences are unidimensional within one method  $m$  across traits. Every rater  $r$  has a value on the corresponding method factor  $M_{rmm}$ . This value indicates the (measurement error-free) deviation score of rater  $r$  from the mean rating of the corresponding method factor  $M_{rmm}$ . As the mean rating of  $M_{rmm}$  is set to zero, positive values stand for overestimations and negative values stand for underestimations in the corresponding type of rating. Only the correlation  $\phi_{mm'}$  between peer report A and peer report B will be set free reflecting the common method effect of these interchangeable types of rating.

On level-2 every target  $t$  has a value in the corresponding trait factor  $T_{ik}$  (extraversion, neuroticism, and conscientiousness) that contains no measurement error or method-specific influences. This value indicates the (measurement error-free) deviation score of subject  $t$  in trait  $k$  from the mean factor value (mean = 0). The correlations between different traits  $\phi_{kk'}$  are set free indicating discriminant validity.

For these analyses the authors used sum scores of each trait measured by self-report, peer report A, and peer report B resulting in nine indicators, three for each trait and three for each type of rating. Thus, the subscript  $i$  was not necessary anymore and had therefore been omitted (cf. Figure 1). In the unrestricted ML-CFA-MTMM model, the factor loadings  $\lambda_{Mikm}$  and  $\lambda_{Tikm}$  were estimated freely, whereas in the restricted ML-CFA-MTMM model these factor loadings were restricted to one assuming equivalent loadings for the indicators of the same trait and the indicators of the same method.

Table 1. Variables and parameters of the ML-CFA-MTMM model and their interpretation

Symbol	Interpretation
<b>Level 1</b>	
$Y_{rtikm}$	Observed score of rater $r$ for target $t$ on indicator $i$ assessing trait $k$ (extraversion, neuroticism, or conscientiousness) measured with method $m$ (self, peer A, or peer B)
$\mu_{ikm}$	Mean value of indicator $i$ assessing trait $k$ (extraversion, neuroticism, or conscientiousness) and measured with method $m$ (self, peer A, or peer B)
$M_{rtm}$	Deviation score of a rater $r$ from the mean rating (mean = 0) of the corresponding method $m$ (self, peer A, or peer B) assessing target $t$
$\varphi_{Mmm'}$	Correlation between the method factors $m$ and $m'$ (self, peer A, and peer B) as a measure of common variance between methods
$\lambda_{Mikm}$	Factor loading of the observed score $Y_{rtikm}$ on method $m$
$E_{rtikm}$	Measurement error of the observed score $Y_{rtikm}$
<b>Level 2</b>	
$T_{tk}$	Deviation score from the mean factor value (mean = 0) of a target $t$ (assessed subject) in trait $k$ (extraversion, neuroticism, or conscientiousness)
$\lambda_{Tikm}$	Factor loading of the observed score $Y_{rtikm}$ on trait $k$
$\varphi_{Tkk'}$	Correlation between the trait variables $k$ and $k'$ (extraversion, neuroticism, or conscientiousness) as a measure of common variance between traits

Note. The index  $i$  within the subscript can be omitted if there is only one indicator for a TMU. For the factor loadings a further index is added to the subscript ( $T$  or  $M$ ) indicating the correspondence to the trait ( $T$ ) or method factors ( $M$ ).

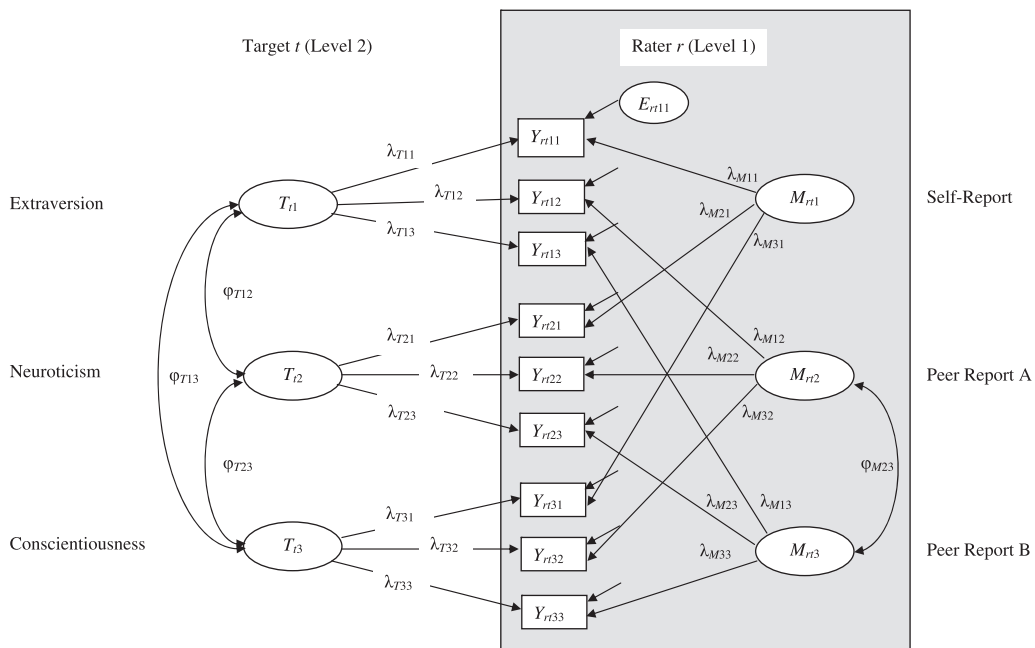


Figure 1. Path diagram of the ML-CFA-MTMM model for the MTMM data set transferred from the multilevel MTMM analyses of Maas et al. (2009) to the CFA framework. The nine indicators  $Y_{rtikm}$  have no subscript  $i$  because they represent the sum scores of each type of rating  $m$  (self-report, peer report A, or peer report B) assessing each trait  $k$  (extraversion, neuroticism, or conscientiousness). For ease of presentation only the error variance  $E_{r11}$  is depicted.

Results of the ML-CFA-MTMM Analyses

The authors conducted an unrestricted ML-CFA-MTMM analysis and a restricted ML-CFA-MTMM analysis

using AMOS (Arbuckle & Wotke, 2003). The model fit indices indicated that the unrestricted ML-CFA-MTMM model should be preferred rejecting the assumption of equivalent loadings. The superiority of the unrestricted

ML-CFA-MTMM model could be due to heterogeneous indicators measuring the same trait and, in principle, resolved by indicator-specific trait variables  $T_{ik}$ .

The trait factor variances  $\text{Var}(T_{ik})$  of the ML-CFA-MTMM model represent differences between the assessed subjects in each trait variable, the method factor variances  $\text{Var}(M_{rmm})$  represent the variation of each type of rating. Table 4 in the article of Maas et al. (2009) presents coefficients for the trait and method influences for the restricted ML-CFA-MTMM model. In the restricted ML-CFA-MTMM analysis the coefficients for convergent validity (the factor loadings) differed depending on the related trait but they were all quite substantial (for extraversion: .60; for neuroticism: .48; and for conscientiousness: .69). The coefficients for convergent validity in the superior unrestricted ML-CFA-MTMM model that were not reported by the authors differed probably slightly within the same trait. Furthermore, there is only a small general (trait-unspecific) method-specific influence of the same type of rating across traits (for self-report: .22; for peer report A: .15; and for peer report B: .11). All reliabilities of the single indicators are low ( $.24 \leq \text{Rel}(Y_{rikm}) \leq .53$ ). This is due to the fact that the measurement errors  $E_{rikm}$  are substantively large. According to the path diagram in Figure 1 the correlations  $\phi_{kk'}$  between the latent trait factors were estimated. These correlations account for the common variance of two trait variables; they demonstrate discriminant validity if they are moderate to low.

The correlations  $\phi_{mm'}$  between method factors were only estimated between peer report A and peer report B. The correlation  $\phi_{M23}$  quantifies the common variance of the corresponding method factors and shows the commonality of these types of ratings (peer A and peer B). With regard to the data set with interchangeable (peer report A and peer report B) and structurally different methods (self and peer reports), one would presume this correlation to be positive and substantial showing that both peer reports have more in common than self- and peer reports.

### Three-Mode Models for MTMM Data

Oort (2009) presents analyses with three-mode models for the delineated MTMM data set. The three modes refer to “traits”, “methods”, and “subjects”, whereas subjects are considered as random and traits and methods as fixed. In his article, Oort first applies direct product (DP) models, followed by composite direct product (CDP) models and finally TMF models. In this section, we will first illustrate the idea of DP models. Second, some of the findings of the DP model analyses of Oort will be presented and discussed. Third, the TMF models will be transferred to the framework of CFA. Fourth, some of the results of the TMF analyses of Oort will be redisplayed and interpreted.

#### DP Models

The underlying methodological approach of three-mode models derives from the analysis of MTMM data within

the MTMM matrix and is based on some strong assumptions. These assumptions will be illustrated drawing on the example of DP models. The DP model (Swain, 1975) decomposes the MTMM correlation matrix  $\Sigma$  (with its entries:  $\text{Cor}(Y_{km}, Y_{k'm'})$ , cf. Equation 2) into two correlation matrices. The first matrix contains the correlations of traits measured by the same method (heterotrait-monomethod correlations) and the second matrix contains correlations of the same trait measured with different methods (monotrait-heteromethod correlations). Thus, the correlations  $\text{Cor}(T_k, T_{k'})$  between different traits  $k$  and  $k'$  form the matrix  $\Sigma_T$  and the correlations  $\text{Cor}(M_m, M_{m'})$  between different methods  $m$  and  $m'$  form the matrix  $\Sigma_M$ . The total matrix is the product of these two matrices:  $\Sigma = \Sigma_T \times \Sigma_M$ . This decomposition implies that the correlations in  $\Sigma_T$  are invariant across methods and the correlations in  $\Sigma_M$  are invariant across traits. With this in mind the heterotrait-heteromethod correlations  $\text{Cor}(Y_{km}, Y_{k'm'})$  between an observed variable  $Y_{km}$  measuring trait  $k$  with method  $m$  and an observed variable  $Y_{k'm'}$  measuring trait  $k'$  with method  $m'$  can be decomposed as a product in the following way:

$$\text{Cor}(Y_{km}, Y_{k'm'}) = \text{Cor}(T_k, T_{k'}) \cdot \text{Cor}(M_m, M_{m'}). \quad (2)$$

The correlation  $\text{Cor}(T_k, T_{k'})$  quantifies the association between two different traits  $k$  and  $k'$  (discriminant validity) and the correlation  $\text{Cor}(M_m, M_{m'})$  quantifies the association between two different methods  $m$  and  $m'$  (convergent validity). It is possible to evaluate by means of model fit indices if these assumptions of the DP model fit a particular MTMM data set or not. The DP models are attractive because they are closely linked to the four criteria of Campbell and Fiske (1959) and allow a “quantitative” instead of just a “qualitative” evaluation of the MTMM matrix (cf. Eid, 2006, p. 226).

#### DP Model Analyses

For the DP model analysis (DP1 in Oort's article, 2009) the matrices  $\Sigma_T$  and  $\Sigma_M$  had to be estimated, whereas  $\Sigma_T$  contains the correlations between the 12 personality items (four items for each of the three personality traits) and  $\Sigma_M$  contains the correlations between the three methods (four items for each of the rating types: self, peer A, and peer B). DP1 showed acceptable model fit and the inspection of the matrices  $\Sigma_T$  and  $\Sigma_M$  indicated that the four criteria of Campbell and Fiske (1959) were fulfilled. The correlations in  $\Sigma_T$  (heterotrait-monomethod) hold for each of the three rating types and the correlations in  $\Sigma_M$  (monotrait-heteromethod) hold for each of the 12 items. This implies that the correlations between traits are the same irrespective if they are assessed by self-report or peer reports. Yet, this assumption of the DP model is generally rather restrictive. Another limitation of the DP model is that it does not account for measurement errors. In contrast to DP models CDP models (Browne, 1984) consider measurement error influences. As both DP and CDP models may be primarily appropriate for single indicator data analysis (using either one item or one scale as indicators of latent traits), TMF models seem to be better suited for multiple indicator data.



**CFA Framework for the TMF Model**

While DP models and CDP models examine convergent and discriminant validity by means of observed score correlations, TMF models examine convergent and discriminant validity by means of latent factor correlations within an MTMM matrix with correlations on the latent level. The methodological approach of TMF models can be transferred to the CFA framework. Applying the CFA framework's notation a TMF-CFA-MTMM model can be defined by

$$Y_{ikm} = \mu_{ikm} + \lambda_{ikm} \cdot TM_{km} + E_{ikm}, \tag{3}$$

where  $Y_{ikm}$  is the observed score of indicator  $i$  measuring trait  $k$  with method  $m$ ,  $\mu_{ikm}$  is the mean of indicator  $i$ ,  $TM_{km}$  is a latent variable specifying a trait  $k$  measured with method  $m$ ,  $\lambda_{ikm}$  is the factor loading of the corresponding indicator  $i$  on trait  $k$  measured with method  $m$ , and  $E_{ikm}$  is the measurement error (residual variable) of indicator  $i$ . Note that there are no separate method factors as  $TM_{km}$  can be regarded as a latent variable that associates both a particular trait and a specific method. Therefore,  $k$  traits and  $m$  methods result in  $k \times m$  latent variables  $TM_{km}$ . The correlation matrix of these latent variables ( $\Phi$ ) can be seen as an MTMM matrix. Additionally, this matrix can be subject to further restrictions as the DP model restrictions, for example (cf. Equation 8 in Oort's article).

**Application of the TMF-CFA-MTMM Model**

Since the three traits  $k$  (extraversion, neuroticism, and conscientiousness) were assessed with three rating types  $m$

(self, peer A, and peer B), nine latent variables  $TM_{km}$  had to be specified (cf. Figure 2). Additionally, some restrictions were introduced and tested. First, the covariances  $Cov(E_{ikm}, E_{ikm'})$  between the residual variables of an indicator  $i$  measured by different methods  $m$  were set free assuming indicator-specific influences on the observed scores of an indicator that generalize across methods. Second, the factor loadings  $\lambda_{ikm}$  were assumed to be invariant across methods ensuring that each trait factor retains its meaning irrespective of the measurement method. The assumptions one and two yielded a TMF model (TMF1 in Oort's article) that was evaluated by means of model fit indices. Third, the correlations between different traits within the same method (heterotrait-monomethod correlations) and the correlations between different methods within the same trait (monotrait-heteromethod correlations) were assumed to be invariant.

Thus, the assumptions one to three yielded an even more restricted TMF model (TMF2 in Oort's article) that was evaluated again by means of model fit indices and that was compared with TMF1. Fourth, TMF1 and TMF2 were compared with another TMF model (TMF3 in Oort's article) that took the type of method into account, that is, that peer A and peer B are interchangeable methods, whereas the self-report is structurally different. In TMF3 the correlations between the trait variables of peer A and peer B should be the same, the correlations of peer A or peer B variables with self-report variables should be the same, the correlations across peer A and peer B should be symmetric, and standard deviations of the latent variables should be the same for peer A and peer B.

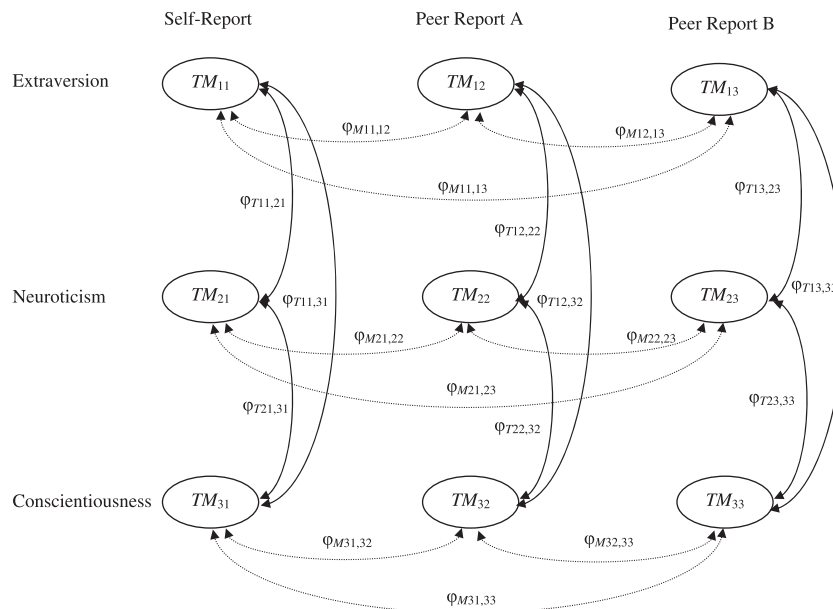


Figure 2. Path diagram of the TMF-CFA-MTMM model for the MTMM data set transferred from the TMF analyses of Oort (2009) to the CFA framework. Each latent variable  $TM_{km}$  represents a trait  $k$  (extraversion, neuroticism, or conscientiousness) assessed by a specific type of rating  $m$  (self-report, peer report A, or peer report B). The lines depict the correlations between the trait variables and the dotted lines depict the correlations between the types of rating. For ease of presentation the four indicators for each latent variable were omitted.

## Results of the TMF Model Analyses

Oort (2009) analyzed the TMF models for the MTMM data set using the Statistical Software Mx (Neale, Boker, Xie, & Maes, 1999) with the maximum likelihood estimation method. The estimations of TMF1, TMF2, and TMF3 resulted in good model fit indices, with TMF3 being slightly superior to TMF1 and TMF2 and with TMF1 being slightly superior to TMF2 according to the expected cross-validation index that can be used to evaluate differences in model fit (Browne & Cudeck, 1993).

In TMF2, being the most restrictive model, it is assumed that the correlations between different traits are the same for all monomethod blocks and that the correlations of the same trait measured with different methods are the same within one heteromethod block (cf. Table 4b in Oort's article). The idea that the correlations between trait variables measured by self-report are equivalent to the correlations between the trait variables measured by peer report A or peer report B is obviously too strong.

In TMF1 these restrictions are not imposed, hence, the results of TMF1 show that there are differences across methods regarding the correlations between traits within the heterotrait-monomethod blocks and that there are differences regarding the correlations across traits within the monotrait-heteromethod blocks (cf. Table 4a in Oort's article), whereas the patterns of heterotrait-monomethod correlations are very similar between peer report A and peer report B but not between self- and peer reports. This is also true for the monotrait-heteromethod correlations.

Therefore, it is not surprising that TMF3 outperforms the other models according to the model fit indices as it takes the differences of the types of methods into account (cf. Table 4c in Oort's article). The heterotrait-monomethod correlations indicating discriminant validity are different between self-report and peer reports but the same between peer report A and peer report B. For the self-report the correlation between extraversion and neuroticism was .13, between extraversion and conscientiousness it was .05, and between neuroticism and conscientiousness it was .06. For the peer reports the correlation between extraversion and neuroticism was  $-.05$ , between extraversion and conscientiousness it was  $-.06$ , and between neuroticism and conscientiousness it was .17. Furthermore the monotrait-heteromethod correlations indicating convergent validity were the same between self-report and peer report A and self-report and peer report B (.52 for extraversion, .31 for neuroticism, and .49 for conscientiousness), but somewhat different between peer report A and peer report B (.47 for extraversion, .34 for neuroticism, and .55 for conscientiousness). Further evidence for convergent validity is given by the factor loadings of the single items that are sufficiently large for extraversion ( $.45 \leq \lambda_{ikm} \leq .77$ ), neuroticism ( $.43 \leq \lambda_{ikm} \leq .84$ ), and conscientiousness ( $.57 \leq \lambda_{ikm} \leq .93$ ).

## CT-C(M-1) Model for Different Types of Raters

Nussbeck et al. (2009) present analyses with the CT-C(M-1) model (Eid, 2000; Eid et al., 2003). In their article, Nussbeck

et al. extend the standard CT-C(M-1) model for two structurally different methods to the CT-C(M-1) model with common method factors for structurally different and interchangeable methods. They depict this extended CT-C(M-1) model with its components and implications in detail and they specify the extended CT-C(M-1) model with common method factors for the MTMM data set. In this section, first the standard CT-C(M-1) model will be taken up again. Second, the extension of this standard CT-C(M-1) model will be pictured in detail and third, the results of the analyses with the extended CT-C(M-1) model will be interpreted.

## Standard CT-C(M-1) Model

The origin for the development of the standard CT-C(M-1) model is the "true score" of each measurement that contains all systematic influences; that is, trait, method, situation, and item influences. The idea of the standard CT-C(M-1) model is to uncover method influences by contrasting two structurally different methods that assess at least two traits. In this standard CT-C(M-1) model one method, the reference method, has to be chosen as comparison standard for which no method factor is specified. All other methods are compared to this reference method. Consequently, method factors are specified capturing the differences between reference and other methods. The underlying idea of this methodological approach is that a psychological measure depends on the measurement instrument applied in a study. As the nonreference method should be contrasted with the reference method, especially structurally different methods seem to be appropriate for an application of the standard CT-C(M-1) model. The choice of the reference method should be based on substantial reasons (Eid, 2000) because it has important implications for the interpretability of the factors of the nonreference method. Guidelines regarding the choice of the reference method are provided in Geiser, Eid, and Nussbeck (2008). The standard CT-C(M-1) model can be illustrated by two equations, one for the indicators of the reference method (cf. Equation 4) and the other for the indicators of the nonreference methods (cf. Equation 5). The indicators of the reference method (say  $m = 1$ ) are defined by:

$$Y_{ik1} = \mu_{ik1} + T_{ik1} + E_{ik1}, \quad (4)$$

where  $Y_{ik1}$  is the observed score of indicator  $i$  assessing trait  $k$  with the reference method  $m = 1$ ,  $\mu_{ik1}$  is the mean value of indicator  $i$ ,  $T_{ik1}$  is the indicator-specific trait  $zk$  measured with reference method  $m = 1$ , and  $E_{ik1}$  is the indicator-specific residual. Thus, each trait score  $T_{ik1}$  includes the trait influence as well as the effect of the reference method; the residual  $E_{ik1}$  of an indicator measured by the nonreference method covers both the nonreference specific method effect and the measurement error. Note that no method factor is specified for the reference method. The correlations between the indicator-specific trait variables  $T_{ik1}$  are estimated indicating convergent validity for correlations between method-specific trait variables of the same trait and indicating discriminant validity for

correlations between method-specific trait variables of different traits.

For the nonreference methods ( $m \neq 1$ ) of the standard CT-C( $M-1$ ) model a trait-specific method influence  $M_{ikm}$  of indicator  $i$  within trait  $k$  regarding method  $m$  is specified that models the method-specific effect of indicator  $i$  measured by method  $m$  within trait  $k$ . As researchers are interested in the specific method effects of the nonreference method, a method factor  $M_{km}$  can be defined due to the assumption that all method-specific residuals are unidimensional. Therefore, the indicators measured by the nonreference method ( $m \neq 1$ ) are defined by:

$$Y_{ikm} = \mu_{ikm} + \lambda_{Tikm} \cdot T_{ik1} + \lambda_{Mikm} \cdot M_{km} + E_{ikm}, \quad (5)$$

where  $Y_{ikm}$  is the observed score of indicator  $i$  assessing trait  $k$  with the nonreference method  $m$ ,  $\mu_{ikm}$  is the mean value,  $\lambda_{Tikm} \cdot T_{ik1}$  is the indicator-specific influence of trait  $k$  measured with the nonreference method  $m$ ,  $M_{km}$  is the trait-specific method factor of the nonreference method  $m$ ,  $\lambda_{Mikm}$  is the loading parameter of indicator  $i$  on the trait-specific method effect, and  $E_{ikm}$  is the indicator-specific residual. The correlations between the method factors  $M_{km}$  are estimated quantifying the amount of common method variance across traits.

### Extended CT-C( $M-1$ ) Model

As the delineated MTMM data set contains not only structurally different methods (self-report and peer report) but

also interchangeable methods (peer report A and peer report B), the standard CT-C( $M-1$ ) model had to be extended. Since the self-report appears as structurally different, it should be chosen as reference method. Thus, for the self-report no method factor will be specified. The nonreference methods (peer report A and peer report B) can easily be interpreted as deviations of the respective peer ratings from the self-rating (Eid et al., 2003). The nonreference methods are interchangeable and thus, they might share common method variance due to the latent correlations between  $M_{km}$ . Common method factors  $CM_k$  should be specified if researchers want to quantify this common method variance of the nonreference methods. Common method factors can be interpreted as general nonreference method effects with respect to a particular trait. Within such an extended CT-C( $M-1$ ) model the method factor  $M_{km}$  is decomposed into the common method factor and the unique method influence. Both method influences are trait-specific, whereas the unique method influence is specific to a particular nonreference method within a specific trait and thus, not correlated with any other latent variable except for unique method effects of the same method. The extended CT-C( $M-1$ ) model with common method factors (cf. Table 2) is defined by two equations, whereas only the second equation differs from the standard CT-C( $M-1$ ) model:

$$Y_{ikm} = \mu_{ikm} + \lambda_{Tikm} \cdot T_{ik1} + \lambda_{CMikm} \cdot CM_k + \lambda_{UMikm} \cdot UM_{km} + E_{ikm}, \quad (6)$$

Table 2. Variables and parameters of the CT-C( $M-1$ ) model with common method factors

Symbol	Interpretation
$Y_{ikm}$	Observed score on indicator $i$ assessing trait $k$ (extraversion, neuroticism, or conscientiousness) measured with method $m$ (self, peer A, or peer B)
$\mu_{ikm}$	Mean value of indicator $i$ assessing trait $k$ (extraversion, neuroticism, or conscientiousness) and measured with method $m$ (self, peer A, or peer B)
$T_{ik1}$	Deviation score from the mean factor value (mean = 0) assessing trait $k$ (extraversion, neuroticism, or conscientiousness) with indicator $i$ for the reference method $m = 1$
$CM_k$	Deviation score from the mean rating (mean = 0) of the common method factor of the nonreference methods (peer report A and peer report B) within trait $k$
$UM_{km}$	Deviation score from the mean rating (mean = 0) of the unique method factor of the nonreference method $m$ (peer A or peer B) within trait $k$
$E_{ikm}$	Measurement error of the observed score $Y_{ikm}$
$\varphi_{Tik,i'k'}$	Correlation between trait $k$ (extraversion, neuroticism, or conscientiousness) assessed by indicator $i$ and trait $k'$ assessed by indicator $i'$ as a measure of discriminant validity
$\varphi_{Tik,i'k}$	Correlation between trait variable $k$ (extraversion, neuroticism, or conscientiousness) assessed by indicator $i$ and the same trait $k$ assessed by indicator $i'$ as a measure of convergent validity
$\varphi_{CMkk'}$	Correlation between the common method factor of trait $k$ and the common method factor of trait $k'$ indicating the general nonreference method effect (the general peer-specific effect)
$\varphi_{UMkm,k'm}$	Correlation between the unique method factors of the nonreference method $m$ within trait $k$ and the same nonreference method $m$ within trait $k'$ indicating the general unique method effect
$\lambda_{Tikm}$	Factor loading of the observed score $Y_{ikm}$ on the trait $k$
$\lambda_{CMikm}$	Factor loading of the observed score $Y_{ikm}$ on the common method factor CM
$\lambda_{UMikm}$	Factor loading of the observed score $Y_{ikm}$ on the unique method factor UM

where  $Y_{ikm}$  is the observed score of indicator  $i$  assessing trait  $k$  with the reference method  $m = 1$ ,  $T_{ik1}$  is the indicator-specific influence of trait  $k$  measured with the reference method  $m = 1$ ,  $CM_k$  is the common method factor across the nonreference methods within trait  $k$ ,  $UM_{km}$  is the unique method factor of the nonreference method  $m$  within trait  $k$ , and  $E_{ikm}$  is the indicator-specific residual. Another assumption should be valid for the CT-C( $M-1$ ) model with common method factors for interchangeable nonreference methods. As the raters of peer report A and peer report B are taken out of one population, they should not differ regarding the distributions of their ratings. Therefore the intercepts, the factor loadings, and the variances of the latent variables have to be restricted to the same value.

**Application of the Extended CT-C( $M-1$ ) Model**

For the analysis of the CT-C( $M-1$ ) model for structurally different and interchangeable methods Nussbeck et al. (2009) formed two parcels by summing up two indicators of the four items for every TMU (two test halves of self-report, two test halves of peer report A, and two test halves of peer report B). Thus, six indicators (three parcels  $\times$  three methods) were available for each trait, two for each of the three rating types (self-report, peer report A, and peer report B). Three indicators (three test halves across methods) formed a trait-specific latent variable  $T_{ik}$ , whereas  $k$  indicates the trait the latent variable belongs to and  $i$  the indicators (test halves) that measure this trait with one of these

methods (cf. Figure 3). As a result of this two latent variables for each trait were obtained.

**Results of the Extended CT-C( $M-1$ ) Model Analysis**

The analysis of the CT-C( $M-1$ ) model was performed using *Mplus 4* (Muthén & Muthén, 2006) and showed a good fit to the data. Applying the methodological approach of the CT-C( $M-1$ ) model one can obtain a very useful variance decomposition (Nussbeck et al., 2009). First, the reliability coefficients are quite different across and within scales but they appear acceptable except for some indicators of neuroticism ( $.44 \leq \text{Rel}(Y_{ikm}) \leq .85$ ). Then, the consistency coefficients  $\text{CO}(Y_{ikm})$  quantifying the proportion of true variance of a nonreference indicator that is explained by the reference method show substantial values for all indicators ( $.17 \leq \text{CO}(Y_{ikm}) \leq .44$ ). The consistency coefficients of the indicators of neuroticism are noticeably smaller. This may be due to the fact that neuroticism is less visible in human behavior than extraversion and conscientiousness. The unique method-specificity coefficients  $\text{UMS}(Y_{ikm})$  quantify the proportion of true variance of a nonreference method indicator that is caused by the unique view of a rater. In this application, these coefficients are generally larger than the consistency coefficients and common method-specificity coefficients ( $.39 \leq \text{UMS}(Y_{ikm}) \leq .54$ ). This indicates that there is a remarkable influence of the individual view of each rater. The common method effects  $\text{CMS}(Y_{ikm})$  quantify the proportion of true variance of a nonreference method that is common with another nonreference method. These

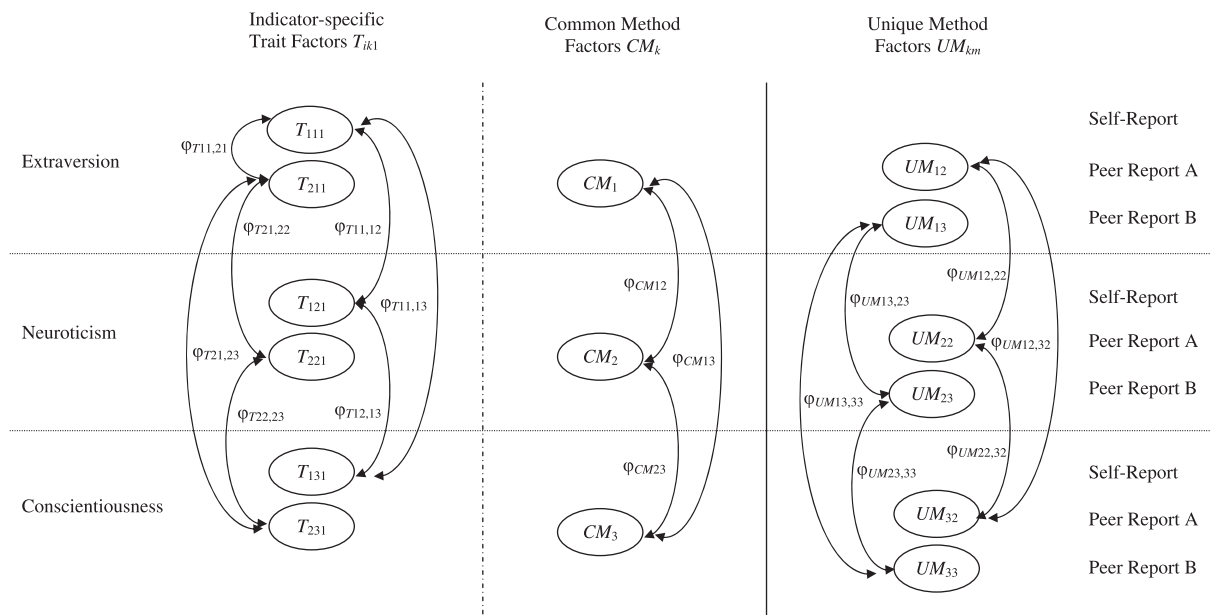


Figure 3. Path diagram of the extended CT-C( $M-1$ ) model with all latent variables arranged according to indicator-specific trait factors  $T_{ik1}$ , common method factors  $CM_k$ , and unique method factors  $UM_{km}$ . Within  $T_{ik1}$  and  $CM_k$  all possible correlations are allowed. For  $UM_{km}$  correlations are allowed only within the same method. The dotted vertical line indicates that there are correlations between  $T_{ik1}$  and  $CM_k$  if  $k \neq k'$ . The nondotted line indicates that there are no correlations between  $UM_{km}$  and  $T_{ik1}$  or  $UM_{km}$  and  $CM_k$ . Easing the illustration only  $\varphi_{T11,21}$  and the horizontal lines were added, and the correlations between  $CM_k$  and  $T_{ik1}$  were omitted.

coefficients are somewhat smaller for extraversion than for neuroticism and conscientiousness ( $.15 \leq \text{CMS}(Y_{ikm}) \leq .34$ ). The correlations  $\varphi_{Tiki'k}$  between the two latent variables of the same trait indicate sufficient convergent validity ( $.75 \leq \varphi_{Tiki'k} \leq .96$ ), while the correlations  $\varphi_{Tiki'k'}$  between two latent variables of different traits demonstrate high discriminant validity ( $.04 \leq |\varphi_{Tiki'k'}| \leq .16$ ). None of the correlations  $\varphi_{CMkk'}$  between the common method factors was significant and the correlations  $\varphi_{UMkmk'm}$  between the unique method factors of the same nonreference method were very small. As a result of this, one may conclude that the method effects in the MTMM data set are obviously both substantial and very trait-specific.

### Comparison of the Three MTMM Approaches

A variety of different methodological approaches for the analysis of MTMM data are available that calls for guidelines for the choice of the best suited MTMM approach for a specific empirical data set and specific research questions (cf. Eid et al., 2008). Therefore, the main objective of this article is the comparison of three different MTMM approaches regarding their conceptualization of trait and method effects, the way convergent and discriminant validity estimates can be obtained, the way method-specific influences are quantified, and their suitability for this specific MTMM data set. In this respect, the multilevel multitrait-multimethod model (ML-CFA-MTMM), the TMF model for multitrait-multimethod data (TMF-CFA-MTMM), and the extended CT-C( $M-1$ ) model with common method factors will be evaluated.

All MTMM approaches were presented within the framework of CFA, a very flexible tool that offers several advantages. In particular all CFA-MTMM approaches separate measurement error from latent trait and method effects. As a consequence, the convergent and discriminant validity can be assessed on the level of latent variables. Within all three MTMM approaches the construct validity of the delineated MTMM data set can be examined, though estimates of convergent and discriminant validity differed considerably between the approaches. Convergent validity could be demonstrated for all approaches by substantial factor loadings on the trait variables or by correlations of different latent variables measuring the same trait. The results also revealed that all discriminant validity coefficients were of negligible size irrespective of the MTMM approach used for the analysis. The three approaches differ with respect to their conceptualization of latent trait and method effects. Subsequently, the extent of method effects also differed considerably between the three MTMM approaches. In the following, we will first compare the MTMM approaches with regard to their ability to determine convergent and discriminant validity. Second, we will compare the three MTMM approaches with regard to the different specifications and interpretations of the method effects. Third, we will list the advantages and fourth, we will list the limitations of each MTMM approach.

### Convergent and Discriminant Validity

In the restricted ML-CFA-MTMM model, the estimated convergent validity coefficients (factor loadings) for the single indicators (sum scores) were large for conscientiousness (.69), somewhat smaller for extraversion (.60), and again smaller for neuroticism (.48). The analysis of the TMF-CFA-MTMM model (TMF3) resulted in coefficients for convergent validity that are higher for extraversion and conscientiousness than for neuroticism. Hence, the pattern is different between the self-report measures (extraversion: .52, conscientiousness: .49, and neuroticism: .31) and the peer report measures (extraversion: .47, conscientiousness: .55, neuroticism: .34). The factor loadings for the single items within the TMF-CFA-MTMM analysis were sufficiently large. In the CT-C( $M-1$ ) model, the consistency coefficients can be interpreted as indicators of convergent validity, and thus, the square root of this coefficient corresponds to the latent correlation between trait and indicator (the factor loading). Therefore, for the self-report (reference method) the latent correlations between latent traits and indicators were 1.0, while for the peer reports (nonreference methods) the correlations between latent traits (self-reported) and indicators were large for extraversion (first test half: .66 and second test half: .60), somewhat smaller for conscientiousness (first test half: .57 and second test half: .47), and again smaller for neuroticism (first test half: .42, second test half: .41). Within all three MTMM approaches neuroticism shows the smallest coefficients for convergent validity. This might be due to the fact that extraversion and conscientiousness are more "visible" traits than neuroticism (cf. Nussbeck et al., 2009).

All three approaches yielded comparable results with respect to the discriminant validity. In the ML-CFA-MTMM approach, the factor loadings on the traits were much higher than the factor loadings on the method factors (self-report: .22, peer report A: .15 and peer report B: .11). In the TMF-CFA-MTMM model, discriminant validity was also high with all correlation coefficients  $\leq |.17|$ . The same result was found in the CT-C( $M-1$ ) model (all correlation coefficients  $\leq |.16|$ ).

### Method Influences

The method effects in the ML-CFA-MTMM analysis accounted for general (nontrait-specific) method influences (cf. Figure 1). This kind of method influence can also be inspected by means of the TMF-CFA-MTMM analyses with regard to heterotrait-heteromethod correlations (cf. Figure 2). Both MTMM approaches showed rather small general method influences. The CT-C( $M-1$ ) model resulted in trait-specific method effects that were quite substantial for the MTMM data set. The analysis showed that the unique method-specificity coefficients were higher than the common method-specificity coefficients. High values of the unique method-specificity coefficients indicate that the self-report measures cannot be used to predict the peer ratings very well (cf. Figure 3) and, moreover, that both peers differ considerably in their views of the target's traits.

According to the conceptualization of the method effect, the interpretation of the latent traits differs. The latent trait factors of the CT-C( $M-1$ ) model represent the trait values measured with the reference method. Note that the nonreference method factors are residual factors in the CT-C( $M-1$ ) model indicating deviation scores from the reference method self-report. The latent trait factors of the ML-CFA-MTMM model can be seen as common factors and the method factors as deviations from this common view. Statements about trait influences with the TMF-CFA-MTMM approach can only be done by means of correlations between the corresponding latent TMUs.

### Advantages

The ML-CFA-MTMM approach enables researchers to deal with multilevel data, interchangeable methods, and to include covariates into their models. The multilevel analysis showed that gender differences exist: Females scored significantly higher on the three traits than males. The TMF-CFA-MTMM model also has some advantages. This model is best suited for the evaluation of the Campbell and Fiske (1959) criteria. Since a DP structure can be imposed on the factor correlations, multiplicative trait-method interactions can be tested. Several other hypotheses can be tested with regard to parameter restrictions, that is, equal factor loadings or factor correlations, and this method is also suited for interchangeable as well as structurally different methods. The CT-C( $M-1$ ) model offers some possibilities that are unique for this approach. First, it contains one method factor less than methods included in the model. This has consequences for the interpretation of the latent traits. The method without method factor is interpreted as the “comparison standard” (reference method). Its latent true scores are explicitly modeled as latent traits. Thus, the trait variable contains influences due to the trait under consideration but also influences due to the measurement method. The latent trait variables are used to predict the scores of the nonreference method indicators in a latent regression. All method factors (of nonreference methods) are residual factors that represent systematic (unidimensional) deviation of all indicators measuring the same trait of one nonreference method. Method factors represent thus deviations from the expected scores given the score on the reference method. The CT-C( $M-1$ ) model has one additional advantage regarding the exploration of method influences. The combination of interchangeable and structurally different methods can be analyzed with the extension of the CT-C( $M-1$ ) model. This extension allows for disentangling individual rater biases (unique method effects) from shared rater biases (common method effects).

### Limitations

The multilevel approach is univariate originally. Multivariate data can be analyzed in the multilevel regression framework by adding an additional level for the variables. If the focus is mainly on analyzing classical MTMM data, the

multilevel approach has no additional advantage over classical CFA models. The ML-CFA-MTMM model, furthermore, cannot account for unique and common method effects. Furthermore, it is not best suited for MTMM data with structurally different methods, but is well suited for interchangeable methods. The TMF-CFA-MTMM model does not separate the variance due to trait and method influences, it does not differentiate between unique and common method effects, and it cannot evaluate a multilevel data structure. A limitation of the CT-C( $M-1$ ) model is that the choice of an adequate “comparison standard” method is not always self-evident. At least one structurally different method is needed in order to select a well-defined comparison standard. This approach may not be appropriate if all methods are interchangeable and no “golden standard” exists. Additionally, in its current form, the model cannot be applied to multilevel data with differing numbers of peer raters per self-rating but has to be transformed to a multilevel structural equation model (Eid et al., 2008).

## Conclusions

Three methodological approaches for the analysis of MTMM data with certain advantages and limitations were presented and compared. First, the CFA framework appears very suitable for the analysis of MTMM data as it offers a very flexible analysis tool allowing for the examination of three rather different methodological approaches. Second, depending on the choice of the CFA-MTMM model the size and the interpretation of validity coefficients and method effects differ. In the ML-CFA-MTMM, trait variables represent common factors across all methods and, therefore, method factors are deviations from this common score. Loading parameters serve as validity coefficients and for estimating the size of method-specific effects. In the TMF-CFA-MTMM models, the inspection of the latent correlation matrix (with restrictions) reveals the convergent and discriminant validity as well as the size of method effects. In the CT-C( $M-1$ ) model, validity coefficients and method effects depend on the reference method taken as the “comparison standard”. Within this MTMM approach trait variables include the variance of the reference method, and all other methods are interpreted as residuals. Therefore all trait correlations are method-specific. Third, using multiple indicators instead of single indicators should be considered in order to separate method-specific influences more adequately from error-specific influences (Eid et al., 2008). Fourth, within the framework of CFA it is also possible to modify or extend existing methodological approaches. As an example of such a further extension, Eid et al. (2008) propose a multilevel CT-C( $M-1$ ) model.

It should be pointed out that psychological measures generally underlie a wide variety of potential sources of “method” effects. These comprise influences due to different measurement instruments, different raters or informants, or differences of occasions or contexts (cf. Biesanz & West, 2004; Courvoisier, Nussbeck, Eid, Geiser, & Cole, 2008; Pohl, Steyer, & Kraus, 2008; Schermelleh-Engel, Keith,

Moosbrugger, & Hodapp, 2004). Further potential sources of method biases are presented by Podsakoff, MacKenzie, Lee, and Podsakoff (2003). These method effects may be considered simultaneously within MTMM data analyses if the respective measures are available.

Researchers assessing the construct validity of measures by means of MTMM data should consider (a) the types of methods within this data as well as (b) the most appropriate MTMM design to answer their specific research questions. Thus, this article served the purpose to facilitate the choice of the most appropriate MTMM approach.

## References

- Arbuckle, J. L., & Wotke, W. (2003). *Amos 5.0 Update to the user's guide*. Chicago: Small Waters Corporation.
- Bentler, P. M., & Lee, S. Y. (1979). A statistical development of three-mode factor analysis. *British Journal of Mathematical and Statistical Psychology*, *32*, 87–104.
- Biesanz, J. C., & West, S. G. (2004). Towards understanding assessments of the Big Five: Multitrait-multimethod analyses of convergent and discriminant validity across measurement occasion and type of observer. *Journal of Personality*, *72*, 845–876.
- Bloxom, B. (1968). A note on invariance in three-mode factor analysis. *Psychometrika*, *33*, 347–350.
- Browne, M. W. (1984). The decomposition of multitrait-multimethod matrices. *British Journal of Mathematical and Statistical Psychology*, *37*, 1–21.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Beverly Hills, CA: Sage.
- Bryk, A. S., & Raudenbusch, S. W. (1992). *Hierarchical linear models for social and behavioral research: Application and data analysis methods*. Newbury Park, CA: Sage.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105.
- Courvoisier, D. S., Nussbeck, F. W., Eid, M., Geiser, C., & Cole, D. A. (2008). Analyzing the convergent validity of states and traits: Development and application of multimethod latent state-trait models. *Psychological Assessment*, *20*, 270–280.
- Dumenci, L. (2000). Multitrait-multimethod analysis. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 583–611). San Diego, CA: Academic Press.
- Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika*, *65*, 241–261.
- Eid, M. (2006). Methodological approaches for analyzing multimethod data. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 223–230). Washington, DC: APA.
- Eid, M., & Diener, E. (2006). Introduction: The need for multimethod measurement in psychology. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 3–9). Washington, DC: APA.
- Eid, M., Lischetzke, T., & Nussbeck, F. W. (2006). Structural equation models for multitrait-multimethod data. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 283–299). Washington, DC: APA.
- Eid, M., Lischetzke, T., Nussbeck, F. W., & Geiser, C. (2004). *Die Multitrait-Multimethod-Analyse: Entwicklung neuer Modelle und ihre Anwendung in der Differentiellen und Diagnostischen Psychologie* [The multitrait-multimethod analysis: development of new models and their application in psychology]. Landau in der Pfalz: Universität Koblenz-Landau DFG-Projekt EI 379/5-2.
- Eid, M., Lischetzke, T., Nussbeck, F. W., & Trierweiler, L. I. (2003). Separating trait effects from trait-specific method effects in multitrait-multimethod models: A multiple indicator CT-C(M-1) model. *Psychological Methods*, *8*, 38–60.
- Eid, M., Nussbeck, F. W., Geiser, C., Cole, D. A., Gollwitzer, M., & Lischetzke, T. (2008). Structural equation modelling of multitrait-multimethod data: Different models for different types of methods. *Psychological Methods*, *13*, 230–253.
- Geiser, C., Eid, M., & Nussbeck, F. W. (2008). On the meaning of latent variables in the CT-C(M-1) model: A comment on Maydeu-Olivares and Coffman (2006). *Psychological Methods*, *13*, 49–57.
- Hox, J. J. (2002). *Multilevel analysis. Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hox, J. J., & Maas, C. (2006). Multilevel models for multimethod measurements. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 269–281). Washington, DC: APA.
- Kenny, D. A. (1976). An empirical application of confirmatory factor analysis to the multitrait-multimethod matrix. *Journal of Experimental Social Psychology*, *12*, 247–252.
- Maas, C. J. M., Lensvelt-Mulders, G. J. L. M., & Hox, J. J. (2009). A multilevel multitrait-multimethod analysis. *Methodology*, *5*, 72–77.
- Marsh, H. W. (1989). Confirmatory factor analysis of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement*, *12*, 335–361.
- Marsh, H. W., & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 177–198). Thousand Oaks: Sage.
- Marsh, H. W., & Hocevar, D. (1988). A new, more powerful approach to multitrait-multimethod analyses: Application of second-order confirmatory factor analysis. *Journal of Applied Psychology*, *73*, 107–117.
- Muthén, L. K., & Muthén, B. O. (2006). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.
- Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (1999). *Mx: Statistical modeling* (5th ed.). Available at: <http://www.vcu.edu/mx>.
- Nussbeck, F. W. (2006). Assessing multimethod association with categorical variables. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 231–247). Washington, DC: APA.
- Nussbeck, F. W., Eid, M., Geiser, C., Courvoisier, D. S., & Lischetzke, T. (2009). A CTC(M-1) model for different types of raters. *Methodology*, *5*, 88–98.
- Oort, F. J. (1999). Stochastic three-mode models for mean and covariances structures. *British Journal of Mathematical and Statistical Psychology*, *52*, 243–272.
- Oort, F. J. (2009). Three-mode models for multitrait-multimethod data. *Methodology*, *5*, 78–87.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*, 879–903.
- Pohl, S., Steyer, R., & Kraus, K. (2008). Modelling method effects as individual causal effects. *Journal of the Royal Statistical Society, Series A*, *171*, 41–63.
- Rost, J., & Carstensen, C. (2002). Multidimensional Rasch measurement via item component models and faceted designs. *Applied Psychological Measurement*, *26*, 42–56.
- Rost, J., & Walter, O. (2006). Multimethod item response theory. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 249–268). Washington, DC: APA.

- Schermelleh-Engel, K., Keith, N., Moosbrugger, H., & Hodapp, V. (2004). Decomposing person and occasion-specific effects: An extension of latent state-trait theory to hierarchical models. *Psychological Methods, 9*, 198–219.
- Schermelleh-Engel, K., & Schweizer, K. (2007). Multitrait-Multimethod-Analysen. In H. Moosbrugger & A. Kelava [Multitrait-multimethod analyses] (Eds.), *Testtheorie und Fragebogenkonstruktion* [Test theory and questionnaire designing]. Heidelberg: Springer.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.
- Swain, A. J. (1975). *Analysis of parametric structures for variance matrices*. Australia: University of Adelaide Unpublished doctoral dissertation.
- Widaman, K.-F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement, 9*, 1–26.

---

Volkmar Höfling

---

Department of Psychological Research Methods and Evaluation  
Goethe University  
Mertonstrasse 17  
60054 Frankfurt am Main  
Germany  
Tel. +49 69 798 22894  
Fax +49 69 798 23847  
E-mail v.hoeffling@psych.uni-frankfurt.de

---



## Anhang B

### Schrift (B)

**Höfling, V.**, Moosbrugger, H., Schermelleh-Engel, K., & Heidenreich, T. (in press). Mindfulness or mindlessness? A modified version of the Mindful Attention and Awareness Scale (MAAS). *European Journal of Psychological Assessment*.

Running Head: Mindfulness or Mindlessness?

Mindfulness or Mindlessness?

A Modified Version of the Mindful Attention and Awareness Scale (MAAS)

Volkmar Höfling <sup>a</sup>

Helfried Moosbrugger <sup>a</sup>

Karin Schermelleh-Engel <sup>a</sup>

Thomas Heidenreich <sup>b</sup>

<sup>a</sup> Department of Psychological Research Methods and Evaluation, Goethe University Frankfurt,  
Frankfurt a. M., Germany

<sup>b</sup> Esslingen University of Applied Sciences, Esslingen, Germany

Corresponding author:

Volkmar Höfling

Goethe University Frankfurt

Department of Psychological Research Methods and Evaluation

Mertonstrasse 17

60054 Frankfurt am Main

Germany

E-Mail: [v.hoeffling@psych.uni-frankfurt.de](mailto:v.hoeffling@psych.uni-frankfurt.de)

Tel. +49 69 798 22894

Fax. +49 69 798 23847

**Abstract**

The 15 items of the Mindful Attention and Awareness Scale (MAAS; Brown & Ryan, 2003) are negatively worded and assumed to assess mindfulness. However, there were indications of non-equivalence between the original MAAS and a version with the positively rephrased MAAS items (“mirror items”). The present study examines if the mindfulness facet “mindful attention and awareness” (MAA) can be measured with positively and negatively worded items if method effects due to item wording are taken into account. Thus, the 15 negatively worded items of the MAAS and additionally 13 positively rephrased items were assessed ( $N = 602$ ). Confirmatory factor analyses (CFA) models with and without regard to method effects were carried out and evaluated by means of model fit. As a result, the positively and negatively worded items should be seen as different methods that influence the construct validity of mindfulness. Furthermore, a modified version of the MAAS (MAAS-Short) with five negatively worded items (taken from the MAAS) and five positively worded items (“mirror items”) was introduced as an alternative to assess MAA. The MAAS-Short appears superior to the original MAAS. The results and the limitations of the present study are discussed.

**Keywords:** mindfulness; multitrait-multimethod analysis; method effects; item wording effects; positively and negatively worded items.

## 1. Introduction

The way of directing one's attention on the experience occurring in the present moment is generally defined as a core aspect of mindfulness (Kabat-Zinn, 1994). An important self-report instrument on mindfulness, the Mindful Attention and Awareness Scale (MAAS), is assumed to measure the tendency to be attentive to and aware of present-moment experiences in daily life (Brown & Ryan, 2003). Within empirical analyses, mindfulness was proved to be distinct from constructs of reflexive consciousness like self-monitoring or private self-consciousness, and it was moderately related to measures of psychological well-being like self-esteem and optimism. Another important self-report questionnaire, the Kentucky Inventory of Mindfulness Skills (KIMS; Baer, Smith, & Allen, 2004), however, conceptualised mindfulness as a multifaceted construct including the mindfulness skills observing, describing, accepting without judgement, and acting with awareness.

Confirmatory factor analyses (CFA) with the MAAS revealed indications of lacking homogeneity (Brown & Ryan, 2003, p. 827). While there is a growing number of studies employing mindfulness measures (e.g. Michal, Beutel, Jordan, Zimmermann, Wolters, & Heidenreich, 2007), methodological concerns have been raised. Grossman (2008), for example, queries the exclusive focus on the assessment of lack of attentiveness as a measure of mindfulness (conceptualised as the converse of inattentiveness) as is the case within the MAAS.

### *1.1. Item wording effects within the Mindful Attention and Awareness Scale (MAAS)?*

Since mindless states are assumed to be much more common than mindful states, indirect items should be more "diagnostic" than direct claims to mindfulness (Brown & Ryan, 2003, p. 825f). Thus, the MAAS respondents rate how often they have experiences of mindlessness, e.g., acting on automatic pilot, or not paying attention to the present moment ("I find myself doing things without paying attention"). In their original paper, Brown and Ryan (2003) argue that the assessment of mindlessness (indirect-item approach via negatively worded items) is

empirically equivalent to mindfulness (direct-item approach via positively rephrased “mirror items”) by measuring the magnitude of correlation between the direct-item and indirect-item approaches ( $r = -.70$ ). To demonstrate the conceptual equivalence of the direct-item and the indirect-item approach, the authors obtained the comparative validity of both approaches by showing the same directions of effect to other measures. However, the correlations of the indirect-item approach (negatively worded items) with neuroticism, depression, rumination-reflection, and anxiety were all significantly larger (Brown & Ryan, 2003, p. 831f.) than with the direct-item approach (positively worded mirror items). Additionally, there were no differences between the correlations of the direct-item and the indirect-item approach with self-esteem and openness. Taking this into consideration, the stated equivalence of mindfulness and mindlessness seems open to question. The question is warranted whether the two different approaches using positively and negatively worded items should be considered as two different “methods” (Marsh, 1989) measuring the trait mindfulness. Item wording effects may bias the convergent and discriminant validity, i.e. the extent to which mindfulness correlates with other measures (Cote & Buckley, 1988).

### *1.2. Research on item wording effects*

Item wording effects have been reported in many areas such as educational research (Weems, Onwuegbuzie, & Lustig, 2003), mood research (Vautier, Steyer, Jmel, & Raufaste, 2005), or organisational research (Spector, Van Katwyck, Brannick, & Chen, 1997). Item wording effects within the Rosenberg Self-Esteem Scale (RSES; Rosenberg, 1965) were demonstrated with confirmatory factor analysis (CFA) models and further explored by linking them to other substantial constructs (DiStefano & Motl, 2006; Quilty, Oakman, & Risko, 2006), e.g. the effect of negatively worded items could be linked to the behavioural inhibition system or to fear of negative evaluation. For the Life Orientation Test (LOT-R; Scheier, Carver, & Bridges, 1994), the effect of positively worded items could be linked to social desirability, especially impression management (IM; Rauch, Schweizer, & Moosbrugger, 2007). Depending on the

extent of IM the functioning of positively and negatively worded items differed which might be a reason for observed deviations of the one-dimensionality of optimism. The position of one-dimensionality of optimism (Roth, Decker, Herzberg, & Brähler, 2008) can be maintained if method effects are taken into account. Recently, Schweizer and Rauch (2008) demonstrated item wording effects within the Social Optimism Scale of the Personal and Social Optimism questionnaire (POSO-E; Schweizer & Koch, 2001) by comparing CFA models with and without method effects. The application of an item response (IR) model for the Personal Optimism Scale of the POSO-E showed differences between positively and negatively worded items with respect to the separation of persons according to their degree of optimism (Rauch, Schweizer, & Moosbrugger, 2008).

## **2. Objectives**

The aim of the present study is to evaluate the construct validity of the MAAS. We assume that validity problems of the MAAS are based on item wording effects and non-homogeneous items. Thus, four CFA models were compared including a one-dimensional (OD) model with one trait factor for all negatively and positively worded mirror items, a two-dimensional (TD) model with two separate trait factors for mindfulness (positively worded items) and mindlessness (negatively worded items), and two trait-method (TM) models, both with a trait factor mindfulness, one with a method factor for negative item wording (NIW), and the other with a method factor for positive item wording (PIW). The two TM models refer to multitrait-multimethod (MTMM) models that uncover method influences by contrasting different methods (Höfling, Schermelleh-Engel, & Moosbrugger, 2009). Finally, the two TM models were modified including only the five most valid negatively worded and the five most valid positively worded items. The resulting MAAS-Short is assumed to be best suited for the assessment of MAA.

### 3. Method

#### 3.1. Participants

Participants were drawn from student populations at the Institute of Psychology at the Goethe University Frankfurt, and at the University of Applied Sciences Esslingen. Thus, for the entire sample ( $N = 602$ ) the mean age is 31.1 years ( $SD = 12.0$ ; range = 16–73). The sample is not balanced with regard to gender (female = 65.6 %; male = 34.4 %).

#### 3.2. Measures

##### 3.2.1. Mindful Attention and Awareness Scale (MAAS; German adaptation: Michalak, Heidenreich, Ströhle, & Nachtigall, 2008)

Dispositional mindful attention and awareness (MAA) was assessed with the German adaptation of the 15-item scale (Brown & Ryan, 2003) that exclusively consists of negatively worded items. As with to Brown and Ryan (2003, p. 831) 13 of these 15 items were rephrased into positively worded items (“mirror items”), while two items could not be rephrased (maas2 and maas12). The 28 items were rated on a 6-point Likert scale and the ratings of the negatively worded items were recoded inversely. The following alpha coefficients were obtained within the present study: MAAS(–) (15 negatively worded items):  $\alpha = .79$ , MAAS(+) (13 positively worded items):  $\alpha = .80$ , the entire MAAS (28 items):  $\alpha = .88$ .

##### 3.2.2. Kentucky Inventory of Mindfulness Skills (KIMS; German adaptation: Ströhle, Nachtigall, Michalak, & Heidenreich, 2009)

The German adaptation of the KIMS is a 39-item instrument designed to measure four basic mindfulness skills as behavioural descriptions of mindfulness: observing (OBS), describing (DES), acting with awareness (AWA), and accepting without judgement (AWJ). Items are rated on a 5-point Likert scale. Internal consistencies range from .76 to .91 for the four subscales. AWA is strongly related to MAA whereas the other three mindfulness skills (OBS, DES, and AWJ) extend the mindfulness conceptualisation of the MAAS.

The adaptations of the KIMS and the MAAS were done with respect to the standards of test adaptation to assure the equivalence of the assessment of mindfulness (cf. Van de Vijver & Watkins, 2006).

### 3.3. Analyses

Six different CFA models were analysed evaluating the construct validity of the MAAS. The OD model and the TD model were ordinary CFA models without method factors for item wording effects and the other four TM models were CFA models with method factors for item wording effects (cf. Table 1).

--- Insert Table 1 about here ---

*Model 1 and Model 2.* The OD model comprised a single trait factor for MAA with loadings of all 28 positively and negatively worded items on MAA and the TD model had separate substantive factors for mindlessness (MAA(-)) and mindfulness (MAA(+)) whereas the negatively worded items loaded on MAA(-) and the positively worded items loaded on MAA(+).

*Model 3 and Model 4.* The first TM model (Model 3) posited a single trait factor MAA and a method factor for negative item wording (NIW). The trait factor had loadings on all 28 items, the method factor was linked with the negatively worded items. The second TM model (Model 4) included a trait factor MAA and a method factor for positive item wording (PIW) that was linked to the positively worded items.

*Model 5 and Model 6.* Since the assumption of many homogeneous indicators is rather restrictive and often questionable, two modified TM models were analysed focusing on the five most valid negatively worded items and on the five most valid positively worded items according to the largest factor loadings within Model 3 and 4. We assumed that fewer items are better suited for the assessment of MAA evaluating the fit statistics of Model 5 and 6 and



inspecting the correlations between the original MAAS, the MAAS-Short and the subscales of the KIMS.

### 3.4. Model estimation and model fit

Models 1 to 6 were estimated with *Mplus* version 5 (Muthén & Muthén, 2006) applying the weighted least squares means and variance adjusted estimator (WLSMV; Muthén & Muthén, 2006, p. 484) since only the MAAS items were used as indicators and many of them deviated substantially from the normal distribution. For model fit evaluation, the  $\chi^2$ -value and the degrees of freedom (*df*) were reported as well as the root mean-square error of approximation (RMSEA), the nonnormed fit index (NNFI), and the Akaike information criterion (AIC) for non-nested models (Schermele-Engel, Moosbrugger, & Müller, 2003). For the analysis with the WLSMV estimator, the weighted root mean-square residual (WRMR) was obtained with values  $\leq 1$  indicating an acceptable fit (Yu, 2002, p. 41). The procedure of model fit evaluation will be in line with the guidelines of this journal (Schweizer, 2010).

## 4. Results

*Model 1 and Model 2.* The OD model (Model 1) without regard to item wording effects seems not appropriate according to the fit statistics (cf. Table 2). The factor loadings within the OD model range from .27 to .78 with four factor loadings lower than .40. Establishing separate factors for mindfulness and mindlessness is not an alternative either. The fit of the TD model (Model 2) is nearly as poor as for the OD model with a correlation between the two trait factors of  $r = .99$ . The factor loadings of the negatively worded items ranged from .28 to .70 with four loadings below .40, the factor loadings of the positively worded items range from .31 to .78 with one loading below .40.

--- Insert Table 2 about here ---

*Model 3 and Model 4.* MAA can be assessed with negatively and positively worded items if the method-specific effect of item wording is taken into account. Model 3 and 4 accounted for item wording effects specifying TM models with MAA as the trait factor and PIW (Model 3) and NIW (Model 4), respectively, as method factors. Compared to Model 1 and 2, the model fit increases for both TM models for all fit indices (cf. Table 2). Although the model fit was still not acceptable, taking item wording effects into account results in a more parsimonious solution according to the AIC. Within Model 3, the factor loadings for MAA ranged from .21 to .83 with ten loadings lower than .40, and the loadings for NIW ranged from .08 to .55. Within Model 4, the factor loadings for MAA ranged from .14 to .74 with eleven loadings lower than .40, and the loadings for PIW ranged from .09 to .64. The correlations between the residuals of the corresponding negatively and positively rephrased items (“mirror items”) within Model 3 and 4 were of very different magnitudes ( $.18 \leq r \leq .80$ ); between the negatively and positively worded items of maas7, maas10, maas11, and maas14, they were not significant.

*Model 5 and Model 6.* The fit indices of Model 3 and 4 cannot be assumed as really sufficient as they indicate considerable misfit that should not be ignored. Thus, the ten items with the largest factor loadings from .56 to .82, i.e. the most valid items, five of each method, were selected. These items were maas3, maas7, maas10, maas11, and maas14 for both the negatively and the positively worded version. The resulting 10-item version of the MAAS (MAAS-Short) appeared as a reliable measure to assess MAA with high internal consistency ( $\alpha = .88$ ). Model 5 and 6 showed the best fit of all analysed models with Model 5 being superior according to all fit indices (cf. Table 2). Only the correlations between the corresponding negatively and positively rephrased items maas3 and maas10 were significant indicating specific common variance between the corresponding items (“mirror items”) that cannot be explained by the influence of MAA or item wording. Figure 1 depicts the path diagram of Model 5. Therefore, for the assessment of MAA, five negatively and five positively worded items appear to be more precise and parsimonious than the full version of the original MAAS. Model 5 was re-

analyzed with age and gender as covariates to quantify their influence on MAA. Older participants showed higher values for MAA ( $r = .30$ ) and female participants tended to have higher MAA values as well ( $r = .16$ ).

---Insert Figure 1 about here---

*Convergent Validity of the MAAS-Short.* The inspection of the correlation matrix in Table 3 shows the same descending order of correlations between MAAS and MAAS-Short and the KIMS subscales AWA, AWJ, DES, and OBS. This should be interpreted in terms of equivalence between the MAAS and the MAAS-Short. The high relation of MAAS or MAAS-Short with the KIMS subscale “acting with awareness” (AWA) is not surprising as the conceptualisation of “mindful attention and awareness” (MAA) is very similar. However, the relation between MAAS-Short and AWA is somewhat stronger than between MAAS and AWA since the MAAS-Short measures MAA more precise than the original MAAS. The convergent validity of the MAAS-Short could be demonstrated. A complete matrix of correlations can be obtained from the authors on request.

---Insert Table 3 about here---

## **5. Discussion**

The first aim of the present paper was to demonstrate that item wording threatens the validity of the psychometric assessment of mindfulness (Podsakoff et al., 2003). Our study shows evidence for the method influence of item wording within the MAAS, an instrument to measure the mindfulness facet “mindful attention and awareness” (MAA). Thus, the one-dimensionality of MAA is not queried if the method effect of item wording is taken into account (cf. Schweizer & Rauch, 2008). MAA can be assessed by positively as well as

negatively worded items if trait-method (TM) models are applied. Furthermore, the simultaneous administering of negatively and positively worded items can reduce response sets and contribute to a better response reporting (cf. Krosnick, 1999). This is in line with Grossman (2008) who argued against the exclusive indirect approach of the MAAS as well.

The second aim of this study was to identify the most homogeneous items. For the assessment of MAA, fewer items are sufficient. The 10-item version MAAS-Short (cf. Appendix) reliably measures MAA and is superior to the MAAS with regard to internal consistency. However, it should be noted that the content validity might be somewhat restricted with fewer items. Furthermore, the strong relation between “mindful attention and awareness” (MAA) and “acting with awareness” (AWA) demonstrates the construct validity of this basic mindfulness facet.

There are three limitations of the present study, 1) further measures (e.g. the behavioural activation system (BAS), the behavioural inhibition system (BIS), the NEO-FFI openness scale, or social desirability) are needed to explore the stated item wording effects, 2) the use of “mirror items” might affect the participants’ response behaviour unfavourably because of the similar item content of negatively and positively rephrased items, and, 3) the results are restricted to the German adaptation of the MAAS and should only carefully be generalised to the original version.

Taken together, our study could demonstrate that 1) mindfulness can be assessed via positively as well as negatively worded items, 2) appropriate TM models should be applied to control for method effects due to item wording, and, 3) the MAAS-Short is a valid instrument to measure MAA. Advanced multitrait-multimethod models can help to establish evidence for the possible role of method effects in current research. Thus, from our point of view, since quantitative research on mindfulness is still in its beginning it should be intensified and it should complete theoretical considerations and qualitative research.

## References

- Baer, R. A., Smith, G. T., & Allen, K. B. (2004). Assessment of mindfulness by self-report: The Kentucky Inventory of Mindfulness Skills. *Assessment, 11*, 191–206.
- Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology, 84*, 822–848.
- Cote, J. A., & Buckley, R. (1988). Measurement error and theory testing in consumer research: An illustration of the importance of construct validation. *Journal of Consumer Research, 14*, 579–582.
- DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling, 13*, 440–464.
- Grossmann, P. (2008). On measuring mindfulness in psychosomatic and psychological research. *Journal of Psychosomatic Research, 64*, 405–408.
- Höfling, V., Schermelleh-Engel, K., & Moosbrugger, H. (2009). Analyzing multitrait-multimethod data: A comparison of three approaches. *Methodology, 5*, 99–111.
- Kabat-Zinn, J. (1990). Full catastrophe living: Using the wisdom of your body and mind to face stress, pain, and illness. New York: Bantam Dell.
- Krosnick, J. A. (1999). Survey research. *Annual review of Psychology, 50*, 537–567.
- Marsh, H. W. (1989). Confirmatory factor analysis of multitrait-multimethod data: Many problems and few solutions. *Applied Psychological Measurement, 13*, 335–361.
- Michal, M., Beutel, M. E., Jordan, J., Zimmermann, M., Wolters, S., & Heidenreich, T. (2007). Depersonalisation, mindfulness, and childhood trauma. *Journal of Nervous and Mental Disease, 195*, 693–696.
- Michalak, J., Heidenreich, T., Ströhle, G., & Nachtigall, C. (2008). Die deutsche Version der Mindful Attention and Awareness Scale (MAAS) [The German version of the Mindful

- Attention and Awareness Scale (MAAS)]. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 37, 200–208.
- Muthén, L. K., & Muthén, B. O. (2006). *Mplus user's guide*. Los Angeles; CA: Muthén & Muthén.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *Journal of Applied Psychology*, 88, 879–903.
- Quilty, L. C., Oakman, J. M., & Risko, E. (2006). Correlates of the Rosenberg Self-Esteem Scale method effects. *Structural Equation Modeling*, 13, 99–117.
- Rauch, W. A., Schweizer, K., & Moosbrugger, H. (2007). Method effects due to social desirability as a parsimonious explanation of the deviation from unidimensionality in LOT-R scores. *Personality and Individual Differences*, 42, 1597–1607.
- Rauch, W. A., Schweizer, K., & Moosbrugger, H. (2008). An IRT analysis of the Personal Optimism Scale. *European Journal of Psychological Assessment*, 24, 49-56.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Roth, M., Decker, O., Herzberg, P. Y., & Brähler, E. (2008). Dimensionality and norms of the Rosenberg Self-esteem Scale in a German general population sample. *European Journal of Psychological Assessment*, 24, 190-197.
- Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the life orientation test. *Journal of Personality and Social Psychology*, 67, 1063–1078.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8, 23-74.

- Schweizer, K. (2010). Some guidelines concerning the modeling of traits and abilities in test construction. *European Journal of Psychological Assessment, 26*, 1-2.
- Schweizer, K. & Koch, W. (2001). A revision of Cattell's investment theory: Cognitive properties influencing learning. *Learning and Individual Differences, 13*, 57-82.
- Schweizer, K. & Rauch, W. (2008). An investigation of the structure of the Social Optimism Scale with respect to the dimensionality problem. *Journal of Individual Differences, 29*, 223-230.
- Spector, P. E., van Katwyk, P. T., Brannick, M. T., & Chen, P. Y. (1997). When two factors don't reflect two constructs: How item characteristic can produce artifactual factors. *Journal of Management, 23*, 659-677.
- Ströhle, Nachtigall, Michalak, & Heidenreich (in press). Die Erfassung von Achtsamkeit als mehrdimensionales Konstrukt: Die deutsche Version des Kentucky Inventory of Mindfulness Skills (KIMS) [The assessment of mindfulness as a multifacet construct: The German version of the Kentucky Inventory of Mindfulness Skills (KIMS)]. *Zeitschrift für Klinische Psychologie und Psychotherapie*.
- Van de Vijver, F. J. R. & Watkins, D. (2006). Assessing similarity of meaning at the individual and country level. *European Journal of Psychological Assessment, 22*, 69-77.
- Vautier, S., Steyer, R., Jmel, S., & Raufaste, E. (2005). Imperfect or perfect dynamic bipolarity? The case of antonymous affective judgements. *Structural Equation Modeling, 12*, 391-410.
- Weems, G. H., Onwuegbuzie, A. J., & Lustig, D. (2003). Profiles of respondents who respond inconsistently to positively- and negatively-worded items on rating scales. *Evaluation and Research in Education, 17*, 45-60.
- Yu, C.-Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Doctoral dissertation, University of California, Los Angeles.

Table 1

*The six CFA models: Two CFA models without method factors (Model 1 and 2) and four trait-method (TM) models (Model 3, 4, 5, and 6). Model 5 and 6 were modified including only the five most valid negatively and the five most valid positively worded items as indicators*

<b>Model</b>	<b>Type of model</b>	<b>Trait factor(s)</b>	<b>Indicators</b>	<b>Method factor</b>
<b>1</b>	One-dimensional	MAA	28	–
<b>2</b>	Two-dimensional	MAA(+) and MAA(–)	28	–
<b>3</b>	Trait-method	MAA	28	NIW
<b>4</b>	Trait-method	MAA	28	PIW
<b>5</b>	Trait-method	MAA	10	NIW
<b>6</b>	Trait-method	MAA	10	PIW

*Note.* MAA = mindful attention and awareness; MAA(+) = mindfulness; MAA(–) = mindlessness; NIW = negative item wording; PIW = positive item wording.



Table 2

*Fit statistics for all models*

<b>Model</b>	$\chi^2$	<i>df</i>	<b>WRMR</b>	<b>RMSEA</b>	<b>NNFI</b>	<b>AIC</b>
<b>1</b>	3298.54	100	4.40	.23	.71	48063.30
<b>2</b>	3323.74	100	4.40	.23	.71	47895.48
<b>3</b>	718.21	107	1.83	.10	.95	46303.41
<b>4</b>	745.25	109	1.59	.10	.95	46460.45
<b>5</b>	87.82	21	0.94	.07	.99	14733.70
<b>6</b>	189.03	21	1.45	.12	.98	14775.44

*Note.* WRMR = weighted root mean-square residual; RMSEA = root mean-square error of approximation; NNFI = nonnormed fit index; AIC = Akaike information criterion.

Table 3

*Pearson's correlations between the original MAAS (15-item version), the short version of the MAAS (MAAS-Short; 10-item version), and the four subscales of the KIMS*

	<b>DES</b>	<b>AWA</b>	<b>AWJ</b>	<b>MAAS</b>	<b>MAAS- Short</b>
<b>OBS</b>	.30**	.04	-.17**	.18**	.18**
<b>DES</b>		.18**	.19**	.27**	.28**
<b>AWA</b>			.39**	.59**	.72**
<b>AWJ</b>				.41**	.37**

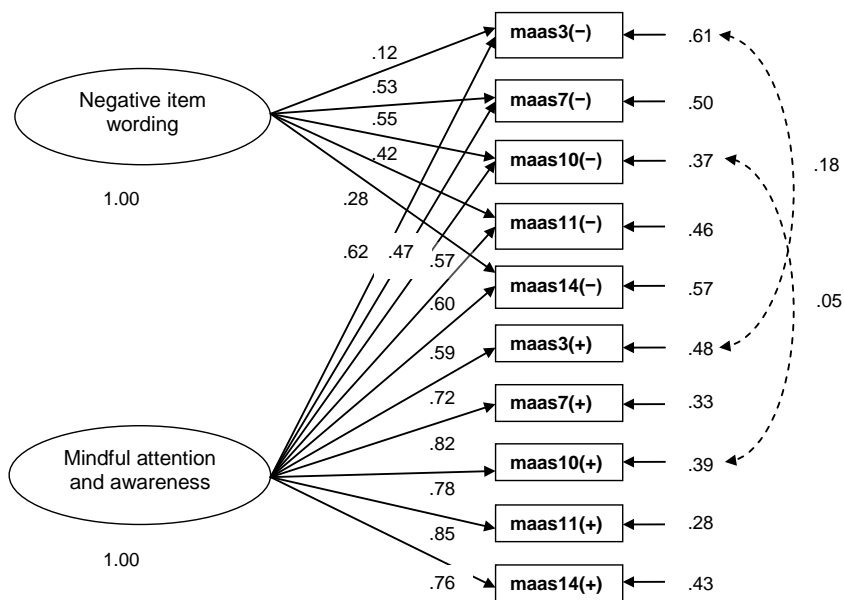
*Note.* OBS = KIMS subscale “observing”; DES = KIMS subscale “describing”; AWA = KIMS subscale “acting with awareness”; AWJ = KIMS subscale “accepting without judgement”; \*\* = correlation is significant ( $p = .01$ ).

Figure caption

Figure 1

Path diagram of the standardised solution of Model 5, a modified TM model with the trait factor mindful attention and awareness (MAA) and a method factor for negative item wording (NIW). The factor loadings of maas3(-) on the trait and on the method factor are fixed as scaling variables to achieve identification.

Figure 1



## Appendix – MAAS-Short

<b>Item</b>	<b>Characterisation of item content</b>
maas3(-)	difficult to stay focused in the present
maas7(-)	being without much awareness of what is done
maas10(-)	doing jobs or tasks automatically
maas11(-)	listening to someone doing something else at the same time
maas14(-)	doing things without paying attention
maas3(+)	easy to stay focused in the present
maas7(+)	being aware of what is done
maas10(+)	doing jobs or tasks with awareness
maas11(+)	listening to someone without doing something else at the same time
maas14(+)	doing things with paying attention

## Anhang C

### Schrift (C)

**Höfling, V.**, Ströhle, G., Michalak, J., & Heidenreich, T. (under review). A short version of the Kentucky Inventory of Mindfulness Skills (KIMS). *Journal of Clinical Psychology*.

Running Head: The KIMS-Short

A Short Version of the Kentucky Inventory of Mindfulness Skills (KIMS)

Volkmar Höfling <sup>a</sup>

Gunnar Ströhle <sup>b</sup>

Johannes Michalak <sup>c</sup>

Thomas Heidenreich <sup>d</sup>

<sup>a</sup> Goethe University Frankfurt, Germany

<sup>b</sup> Clinic of Psychotherapy, Psychosomatic, and Behavior Medicine Bad Gottleuba, Germany

<sup>c</sup> Ruhr University Bochum, Germany

<sup>d</sup> Esslingen University of Applied Sciences, Germany

**Keywords:** mindfulness; assessment; confirmatory factor analysis.

Correspondence concerning this article should be addressed to:

Volkmar Höfling, Goethe University Frankfurt, Institute of Psychology, Mertonstrasse 17, 60054

Frankfurt am Main, Germany, E-Mail: [v.hoefling@psych.uni-frankfurt.de](mailto:v.hoefling@psych.uni-frankfurt.de)

**Abstract**

The Kentucky Inventory of Mindfulness Skills (KIMS; Baer, Smith, & Allen, 2004) is a well validated self-report measure for the assessment of four different mindfulness factors. With its 39 items, the KIMS is less economic compared to other mindfulness scales. Therefore, this study aimed at developing a short version of the German adaptation of the KIMS (KIMS-Short) when applied in research studies on the effects of mindfulness. Confirmatory factor analyses were carried out with two samples ( $N = 469$  and  $N = 602$ ) to develop subscales with fewer items and to confirm the factor structure of the KIMS-Short. Furthermore, the relations between the KIMS-Short subscales and other scales were evaluated. The KIMS-Short contains 20 items enabling researchers to replicate the basic factor structure of four separate mindfulness skills. However, the analyses for the “observing” subscale revealed two different but strongly correlated factors depending on whether the observed stimuli were internal or external phenomena. The relations to other scales indicated that the KIMS-Short is a valid self-report measure for different factors of mindfulness.



## Introduction

The Kentucky Inventory of Mindfulness Skills (KIMS; Baer, Smith, & Allen, 2004) was designed to assess mindfulness in daily life via four dimensions, i.e. observing internal and external stimuli, describing and labelling phenomena nonjudgmentally, acting with awareness and undivided attention, and accepting events or experiences without judging them. Furthermore, the KIMS was developed as a self-report measure applicable to populations without meditation experience. It demonstrated good psychometric qualities and it was translated into other languages, e.g. into German (cf. Christopher, Charoensuk, Gilbert, Neary, & Pearce, 2009; Ströhle, Nachtigall, Michalak, & Heidenreich, 2009).

Empirical studies proved a salutogenic effect of mindfulness interventions such as the mindfulness-based stress reduction program (MBSR; Kabat-Zinn, 1990) on variables of psychological distress (Baer, 2003; Grossman, Niemann, Schmidt, & Walach, 2004). There is preliminary evidence that one or more mindfulness dimensions at least partly mediate such positive outcomes (cf. Carmody, Baer, Lykins, & Olendzki, 2009; Michalak, Meibert, Heidenreich & Schulte, 2008; Shapiro, Oman, Thoresen, Plante, & Flinders, 2008). While other mindfulness measures are one-dimensional, e.g. the Mindful Attention and Awareness Scale (MAAS; Brown & Ryan, 2003) which consists of 15 items, the KIMS assesses the four mindfulness factors containing 39 items. The recent development of the Five Facet Mindfulness Questionnaire (FFMQ; Baer et al., 2006) assesses five mindfulness factors with 39 items. Since this instrument has not been adapted to German yet it cannot be applied for research studies on mindfulness in Germany. Thus, the KIMS is the only multidimensional instrument for the assessment of mindfulness in Germany up to now.

The KIMS with 39 items makes this interesting measure less economic than one-dimensional mindfulness scales like the MAAS. Researchers that assess mindfulness with other measures want to provide questionnaire sets for research studies that guarantee the participants'

compliance, i.e. that do not exceed a certain item number. Hence, shortening the KIMS may be useful for future research studies on the various effects of mindfulness.

Therefore, the aim of this study was to introduce and evaluate a short version of the KIMS (KIMS-Short). Starting from the original KIMS, we identified items that were sufficient for the homogeneous assessment of the different factors of mindfulness measured with the KIMS resulting in a short version of the KIMS (KIMS-Short). We assumed that the factor structure of the KIMS-Short corresponds to the factor structure of the original KIMS (4 factors). Regarding the relations between the subscales of the KIMS-Short, we assumed that they reflect the same pattern as the subscales of the original KIMS. To demonstrate the validity of the different dimensions of mindfulness within the KIMS-Short we expect positive relations between the MAAS and the subscales of the KIMS-Short with the largest correlation between the MAAS and the “acting with awareness” subscale. With respect to the relation between the KIMS-Short and a measure of depressive symptoms, we expect negative relations with the largest correlation between the “accepting without judgement” subscale and a measure of depressive symptoms.

## Method

### *Participants*

*Sample 1.* 469 participants (female = 84.6 %; male = 15.4 %) were drawn from undergraduate psychology student populations at the Friedrich Schiller University of Jena, the Goethe University Frankfurt, and the Ruhr University Bochum. The average age was 24.0 years ( $SD = 6.0$ ; range = 18-48).

*Sample 2.* 602 participants (female = 65.6 %; male = 34.4 %) were drawn from different populations at the Goethe University Frankfurt (psychology and pedagogic students), and from

the University of Applied Sciences Esslingen (social work students). The average age of this sample was 31.1 years ( $SD = 12.0$ ; range = 16–73).<sup>1</sup>

### *Procedures*

Participants were asked if they would like to take part in this study. The participation was voluntarily (informed consent) and without any compensation. Participants were not drawn from clinical populations. They had to complete a questionnaire set individually with the measures described below. Instructions to the questionnaire set as a whole and to the single questionnaires were given in the set. Besides the questionnaires several demographic variables were assessed like age, gender, and experience with meditation.

### *Measures*

*Kentucky Inventory of Mindfulness Skills (KIMS)*. The KIMS (Baer et al., 2004) is a 39-item instrument designed to measure the four basic mindfulness skills “observing” (OBS), “describing” (DES), “acting with awareness” (AWA), and “accepting without judgement” (AWJ). The items of the German adaptation (Ströhle et al., 2009) were rated on a 5-point Likert scale. The KIMS subscales showed a good range of internal consistency (Cronbach’s alpha) within both samples (Sample 1:  $\alpha = .79-.92$ ; Sample 2:  $\alpha = .81-.89$ ).

*Mindful Attention and Awareness Scale (MAAS)*. The MAAS is a 15-item scale for the assessment of attention and awareness to present events and experiences (Brown & Ryan, 2003). The 15 items of the German adaptation (Michalak, Heidenreich, Ströhle, & Nachtigall, 2008) of the MAAS were rated on a 6-point Likert scale. The MAAS showed sufficient internal consistency within both samples (Sample 1:  $\alpha = .83$ ; Sample 2:  $\alpha = .79$ ).

*Beck Depression Inventory*. The Beck Depression Inventory (Beck & Steer, 1993) is one of the most frequently applied scales to assess depressive symptoms. The German adaptation

---

<sup>1</sup> In Germany the ethnic/racial background of participants is not routinely recorded. The proportion of students from Afrika, Amerika, and Asia is usually about 5 to 10 %.

(Hautzinger, Bailer, Worall, & Keller, 1994) including 21 items was only assessed in a subgroup of Sample 1 ( $n = 216$ ) and it showed good internal consistency ( $\alpha = .85$ ) within this subsample.

### *Analyses*

*Item selection.* Responses from Sample 1 were used to carry out four confirmatory factor analyses (CFA), one for each subscale of the KIMS to identify the most homogeneous items via factor loadings. Within each subscale CFA one latent factor was represented by the same number of indicators as items were belonging to the respective KIMS subscale. Those items with factor loadings below .50 were eliminated and the remaining items were defined as the KIMS-Short. Each subscale CFA (with the remaining items) was evaluated according to several fit indices (see below) and Cronbach's alpha, too, was calculated for each subscale of the KIMS-Short.

*Factor structure.* The factor structure of the KIMS-Short was replicated with Sample 2 to demonstrate that the item selection procedure leading to the KIMS-Short is valid not only for Sample 1 but for other samples as well. Thus, an overall CFA was conducted employing all latent factors represented by the remaining items of the respective subscales and correlations between these latent factors. The remaining items again were treated as indicators within this CFA.

*Relations with other scales.* Two scales were selected to demonstrate the validity of the KIMS-Short, i.e. the MAAS to demonstrate the convergent validity and the BDI to demonstrate the relation to a psychopathological measure. Pearson's correlations between the KIMS-Short subscales and the MAAS for Sample 2, as well as between the KIMS-Short subscales and the BDI for a part of Sample 1 ( $n = 216$ ) were calculated.

### *CFA estimation and evaluation and sample size*

All CFAs were estimated with *Mplus* version 5 (Muthén & Muthén, 2006) applying the maximum likelihood mean-adjusted (MLM) estimator since the KIMS items were used as indicators that revealed evidence of multivariate non-normality. The MLM estimator treats the indicators (KIMS

items) as continuously distributed and provides a chi-square test statistic (Satorra-Bentler  $\chi^2$ ) and standard errors that are robust to the non-normality of the indicators. For model fit evaluation, the Satorra-Bentler  $\chi^2$ -value (SB $\chi^2$ ) and the degrees of freedom (*df*) were reported as well as the root mean-square error of approximation (RMSEA), the comparative fit index (CFI) and the standardized root mean-square residual (SRMR).

Less than 30 parameters had to be estimated for the subscales' CFAs. As the ratio of 5 between observations and estimated parameters is traditionally considered as adequate for accurate parameter estimation, the size of Sample 1 can be regarded as large enough. For the overall CFA with the entire KIMS-Short the size of Sample 2 can be regarded as large enough as well since 68 parameters had to be estimated.

## Results

### *Item selection*

Finishing the item selection procedure the KIMS-Short contained 20 items with six items for OBS, five items for DES, four items for AWA, and five items for AWJ. Since the fit indices for the subscales of the KIMS-Short (cf. Table 1) indicate a good model fit (cf. Schermelleh-Engel, Moosbrugger, & Müller, 2003), these remaining items can be regarded as the most homogeneous for each subscale. However, the CFA for the subscale OBS showed the respective items not to be completely one-dimensional. Even the selection of the most homogeneous items did not eliminate this dimensionality problem. As two types of OBS items were put together within the original KIMS, i.e. the "observing inside" (OBS-IN) and the "observing outside" (OBS-OUT) items, two separate correlated latent factors were specified with the three most valid items as indicators for OBS-IN and the three most valid items as indicators for OBS-OUT. This CFA showed good fit indices (cf. Table 1) with a (latent) correlation between OBS-IN and OBS-OUT of  $r = .86$ . Consequently, the underlying factor structure of the KIMS-Short could be regarded as five- instead of four-dimensional.

The coefficients of internal consistency (Cronbach's alpha) ranged between adequate and good (Nunnally & Bernstein, 1994) for the KIMS-Short in Sample 1 and Sample 2, with .76 and .75 for OBS, .89 and .86 for DES, .70 and .78 for AWA, and .87 and .88 for AWJ. Since Cronbach's alpha decreases with the number of items within a scale, the internal consistency of the KIMS-Short subscales is remarkably good as there were only four (AWA) to six items (OBS) per scale.

### *Factor structure*

The five-factor structure of the KIMS-Short was confirmed with the overall CFA that showed good model fit indices (cf. Table 1). Table 2 indicates which items of the original KIMS belong to the factors of the KIMS-Short. All 20 items of the KIMS-Short showed factor loadings above .50, with 16 items above .60, and 10 items above .70. There were significant correlations between all five factors except for OBS-IN, OBS-OUT, and AWJ. The (latent) correlation between OBS-IN and OBS-OUT was  $r = .74$ , the correlations between OBS-IN resp. OBS-OUT with DES were .21 and .22, with respect to AWA they were .14 and .16. The correlation between DES and AWA was .18, between DES and AWJ it was .30, and between AWA and AWJ it was .37. It should be noted that a CFA with a second-order factor for mindfulness did not converge.

Another overall CFA was carried out with age and gender as covariates to quantify the influence of these demographic variables to the KIMS-Short. Older participants showed higher values for AWA ( $r = .20$ ) and AWJ ( $r = .10$ ) and female participants tended to have higher AWA values ( $r = .11$ ). Since there were only some participants with meditation experience this variable was not examined.

### *Relations to other scales*

The relation to the MAAS indicates the convergent validity of the KIMS-Short (cf. Table 3). The factors OBS-IN and OBS-OUT showed low correlations with the MAAS whereas the correlations

with DES, AWA, and AWJ were modest with the strongest correlation becoming evident between the MAAS and AWA.

The relations between the BDI and the KIMS-Short subscales, too, were in line with previous results regarding the KIMS. There were non-significant correlations between the two OBS scales and the BDI, and low or modest correlations between DES resp. AWA and the BDI. Only the correlation between the BDI and AWJ was quite substantial ( $r = -.58$ ).

### Discussion

The first aim of this study was to develop a short version of the KIMS (KIMS-Short) that measures the four mindfulness dimensions of the original KIMS with fewer items. The resulting KIMS-Short contains 20 items, i.e. half as much as the original KIMS. The correlations between the mindfulness factors were modest at most supporting the notion that the factors refer to distinct mindfulness skills. Only the high correlation between OBS-IN and OBS-OUT showed that there might be one underlying mindfulness skill “observing” with differences regarding the perspective of internal or external phenomena.

The second aim of this paper was to present indications of the validity of the KIMS-Short. The highest correlation was found, as anticipated before, between the MAAS and AWA, as these two dimensions have much in common. The pattern of correlations between the MAAS and the KIMS-Short subscales corresponds to previous findings with the original KIMS (cf. Baer et al., 2004). The correlations between the KIMS-Short and the BDI were negatively low to modest as it was the case with the correlations between symptoms of psychopathology and the original KIMS (cf. Baer et al., 2004). The high correlation of the BDI with AWJ needs further explanation. This correlation could be inflated by two different types of method biases (cf. Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). To begin with, the items of AWJ and the BDI could be influenced more strongly by mood states like positive or negative affectivity than the items of other scales. Thus, respondents would view these items in either generally negative or positive terms. Second, the

negative item wording of both all the AWJ and all the BDI items might cause an artifactual strong relationship as well.

There are at least three limitations to the present study. First, the results are limited to the German adaptation of the KIMS and should not simply be transferred to the English version of the KIMS. However, the introduced procedure for the development of a short version of the KIMS can be put into practice for other languages as well. Second, although a shorter version of the KIMS has the aforementioned advantages there remains the disadvantage of less diversity and heterogeneity within a mindfulness dimension with fewer items. Third, the results are based on student samples with less experience with meditation.

In a nutshell, our study provides encouraging evidence that the KIMS-Short is a reliable and valid instrument to assess different mindfulness factors with only a few more items than the MAAS includes.



## References

- Baer, R. A. (2003). Mindfulness training as a clinical intervention: A conceptual and empirical review. *Clinical Psychology: Science and Practice, 10*, 125-143.
- Baer, R. A., Smith, G. T., & Allen, K. B. (2004). Assessment of mindfulness by self-report: The Kentucky Inventory of Mindfulness Skills. *Assessment, 11*, 191–206.
- Baer, R. A., Smith, G. T., Hopkins, J., Krietemeyer, J., & Toney, L. (2006). Using self-report assessment methods to explore facets of mindfulness. *Assessment, 13*, 27-45.
- Beck, A. T. & Steer, R. A. (1993). *Manual for the Beck Depression Inventory*. San Antonio, TX: The Psychological Corporation Inc.
- Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology, 84*, 822–848.
- Carmody, J., Baer, R. A., Lykins, E. L. B., & Olendzki, N. (2009). An empirical study on the mechanisms of mindfulness in a mindfulness-based stress reduction program. *Journal of Clinical Psychology, 65*, 613-626.
- Christopher, M. S., Charoensuk, S., Gilbert, B. D., Neary, T. N., & Pearce, K. L. (2009). Mindfulness in Thailand and the United States: a case of apples versus oranges? *Journal of Clinical Psychology, 65*, 590-612.
- Grossman, P., Niemann, L., Schmidt, S., & Walach, H. (2004). Mindfulness-based stress reduction and health benefits: A meta-analysis. *Journal of Psychosomatic Research, 57*, 35-43.
- Hautzinger, M., Bailer, M., Worall, H. & Keller, F. (1994). Beck-Depressions-Inventar (BDI). Bern: Huber.
- Kabat-Zinn, J. (1990). Full catastrophe living: The program of the Stress Reduction Clinic at the University of Massachusetts Medical Center. New York: Delta.

- Michalak, J., Heidenreich, T., Meibert, P. & Schulte, D. (2008). Mindfulness predicts relapse/recurrence in major depressive disorder following Mindfulness-based Cognitive Therapy. *Journal of Nervous and Mental Disease, 196*, 630-633.
- Michalak, J., Heidenreich, T., Ströhle, G., & Nachtigall, C. (2008). Die deutsche Version der Mindful Attention and Awareness Scale (MAAS) [The German version of the Mindful Attention and Awareness Scale (MAAS)]. *Zeitschrift für Klinische Psychologie und Psychotherapie, 37*, 200–208.
- Muthén, L. K., & Muthén, B. O. (2006). *Mplus user's guide*. Los Angeles; CA: Muthén & Muthén.
- Nunally, J. & Bernstein, I. (1994). *Psychometric Theory* (3rd ed.). New York: McGraw Hill.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879–903.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*, 23-74.
- Shapiro, S. L., Oman, D., Thoresen, C. E., Plante, T. G., & Flinders, T. (2008). Cultivating mindfulness: effects on well-being. *Journal of Clinical Psychology, 64*, 840-862.
- Ströhle, Nachtigall, Michalak, & Heidenreich (in Press). Die Erfassung von Achtsamkeit als mehrdimensionales Konstrukt: Die deutsche Version des Kentucky Inventory of Mindfulness Skills (KIMS) [The assessment of mindfulness as a multifacet construct: The German version of the Kentucky Inventory of Mindfulness Skills (KIMS)]. *Zeitschrift für Klinische Psychologie und Psychotherapie*.

Table 1

*Model fit indices for the KIMS-Short subscales and overall CFA*

	<b>SB<math>\chi^2</math></b>	<b>df</b>	<b>SRMR</b>	<b>RMSEA</b>	<b>CFI</b>
<b>OBS*</b>	11.74	8	.02	.03	.99
<b>DES</b>	2.75	4	.01	.00	1.00
<b>AWA</b>	1.66	1	.01	.04	1.00
<b>AWJ</b>	7.75	5	.02	.03	1.00
<b>Overall<sup>o</sup></b>	235.14	162	.04	.03	.98

*Note.* SB $\chi^2$  = Satorra-Bentler scaled  $\chi^2$ ; SRMR = standardized root mean-square residual; RMSEA = root mean-square error of approximation; CFI = comparative fit index; OBS = observing; DES = describing; AWA = acting with awareness; AWJ = accepting without judgment; \* = CFA with separate factors for “observing inside” and “observing outside”; <sup>o</sup> = CFA with the correlated four mindfulness factors and the remaining 20 items as indicators.

Table 2

*Standardized factor loadings for the CFA with all items of the KIMS-Short (overall CFA)*

<b>Factor</b>	<b>Characterization of item content</b>	<b>Item Number (KIMS)</b>	<b>Factor loading</b>
<b>OBS-IN</b>	Noticing body sensations while walking	9	.53
	Staying alert to sensations while showering	13	.73
	Paying attention to sensations	21	.82
<b>OBS-OUT</b>	Paying attention to sounds	25	.57
	Noticing smells and aromas	29	.60
	Noticing visual elements	33	.62
<b>DES</b>	Finding words to describe feelings	2	.72
	Not finding words to describe thoughts <sup>a</sup>	14	.71
	Not expressing the feelings about things <sup>a</sup>	18	.82
	Not describing body sensations <sup>a</sup>	22	.62
	Finding words if upset	26	.63
<b>AWA</b>	Being focused only on what is done	7	.68
	Getting wrapped up in doing things	19	.65
	Doing several things at once <sup>a</sup>	31	.55
	Getting absorbed in what is done	38	.61
<b>AWJ</b>	Criticizing myself for having emotions <sup>a</sup>	4	.74
	Believing that my thoughts are abnormal <sup>a</sup>	16	.86
	Making judgments about my thoughts <sup>a</sup>	20	.76
	Telling myself that thoughts should not be <sup>a</sup>	28	.85
	Thinking emotions are not appropriate <sup>a</sup>	32	.90

*Note.* OBS-IN = observing inside; OBS-OUT = observing outside; DES = describing; AWA = acting with awareness; AWJ = accepting without judgment; <sup>a</sup> = negatively worded item.

Table 3

*Pearson's correlations between the factors of the KIMS-Short, the MAAS, and the BDI*

	<b>OBS- OUT</b>	<b>DES</b>	<b>AWA</b>	<b>AWJ</b>	<b>MAAS</b>	<b>BDI</b>
	<i>N</i> = 602	<i>N</i> = 602	<i>N</i> = 602	<i>N</i> = 602	<i>N</i> = 602	<i>n</i> = 216
<b>OBS-IN</b>	.48**	.16**	.12**	-.05	.19**	-.15
<b>OBS-OUT</b>		.17**	.08	-.03	.19**	-.08
<b>DES</b>			.15**	.25**	.27**	-.22**
<b>AWA</b>				.31**	.46**	-.14**
<b>AWJ</b>					.41**	-.58**

*Note.* OBS-IN = observing inside; OBS-OUT = observing outside; DES = describing; AWA = acting with awareness; AWJ = accepting without judgement; BDI = Beck Depression Inventory; \*\* = correlation is significant ( $p < .01$ ).

# Selbständigkeitserklärung

Hiermit erkläre ich, Volkmar Höfling, die vorliegende Dissertationsarbeit mit dem Titel „Methodologische und empirische Studien zur Validierung von Achtsamkeitsskalen“ eigenständig und ausschließlich unter Verwendung der angegebenen Hilfsmittel, angefertigt zu haben. Alle Quellen sind als solche kenntlich gemacht. Die vorliegende Arbeit ist in dieser oder anderer Form zuvor nicht als Dissertationsarbeit zur Begutachtung vorgelegt worden. Das beantragte Promotionsverfahren stellt das erste beantragte Promotionsverfahren dar.

Frankfurt am Main, 09.05.2010

Volkmar Höfling