

Konvergente und diskriminante Validität über die Zeit: Integration von Multitrait-Multimethod-Modellen (MTMM-Modellen) und der Latent-State-Trait-Theorie (LST-Theorie)

Fridtjof W. Nussbeck, Michael Eid, Christian Geiser, Delphine S. Courvoisier und David A. Cole

Inhaltsverzeichnis

- 27.1 Einleitung – 715**
 - 27.1.1 MTMM-Modelle – 715
 - 27.1.2 Methodeneffekte – 716
 - 27.1.3 LST-Modelle – 718
 - 27.1.4 Beschränkungen der LST- und MTMM-Modelle – 719
 - 27.1.5 Verbindung beider Ansätze – 720

- 27.2 Längsschnittliche MTMM-Modelle – 721**
 - 27.2.1 Multioccasion-MTMM-Modell – 721
 - 27.2.2 Multiconstruct-LST-Modell – 722
 - 27.2.3 Multimethod-LST-Modell – 724
 - 27.2.4 Vergleich der drei längsschnittlichen MTMM-Modelle – 727

- 27.3 Multiconstruct-LST- und Multimethod-LST-Modell in der empirischen Anwendung – 730**
 - 27.3.1 Ergebnisse mit dem Multiconstruct-LST-Modell – 730
 - 27.3.2 Ergebnisse mit dem Multimethod-LST-Modell – 733
 - 27.3.3 Fazit der Anwendungen der beiden multimethodalen LST-Modelle – 735

- 27.4 Praktische Hinweise zur Analyse longitudinaler multimodaler Modelle – 735**

Konstanzer Online-Publikations-System (KOPS)

URL: <http://nbn-resolving.de/urn:nbn:de:bsz:352-2-1sb0whsai7u689>

27.5 Zusammenfassung – 736

27.6 EDV-Hinweise – 736

27.7 Kontrollfragen – 737

Literatur – 737

i Im diagnostisch-therapeutischen Kontext könnten folgende Fragen von hoher Relevanz sein: Wie stark ähneln sich Schüler- und Lehrereinschätzungen in Bezug auf z. B. Ängstlichkeit und Depressivität? Generalisieren Lehrer im Sinne eines Halo-Effekts (Thorndike 1920), kommen sie also zu nahezu identischen Einschätzungen von Ängstlichkeit und Depressivität? Wie sehr hängen die Einschätzungen der Lehrer von stabilen oder situativen Faktoren ab? Können Lehrer die Schwankungen der Ängstlichkeit ihrer Schüler nachvollziehen? Diese Fragen können weder mit LST- noch mit MTMM-Modellen umfassend beantwortet werden, sondern müssen in Kombinationen der beiden Ansätze untersucht werden. Im folgenden Abschnitt werden drei Modelle vorgestellt, die eine Beantwortung dieser Fragen ermöglichen.

27.1 Einleitung

Psychologische Merkmale und somit auch psychologische Messwerte unterliegen einer Vielzahl von Einflüssen. Beispielsweise hängt die Ausprägung der Ängstlichkeit von Schulkindern nicht nur von ihrer dispositionellen Ängstlichkeit als über Situationen hinweg stabile Eigenschaft ab, sondern auch von (momentan wirkenden) situativen Einflüssen, z. B. einem gerade wütenden Sturm oder dem Alptraum der letzten Nacht. Außerdem fallen die Messungen je nach Messmethode, z. B. ob die Schüler sich selbst einschätzen oder ob sie von ihren Lehrern oder Eltern eingeschätzt werden, unterschiedlich aus. In vorangegangenen Kapiteln wurden bereits statistische Modelle beschrieben, die verschiedene Einflussfaktoren auf Messungen trennbar und in ihrer Größe messbar machen können. Schermelleh-Engel, Geiser und Burns beschreiben in ► Kap. 25 Modelle, die den Einfluss verschiedener Messmethoden auf die Messergebnisse untersuchen. Die Messmethoden beinhalten ganz unterschiedliche inhaltliche Aspekte. Bezüglich der Messergebnisse verschiedener Fragebogen können mit den vorgestellten Multitrait-Multimethod-Modellen (MTMM-Modellen) z. B. die Einflüsse unterschiedlicher Beurteiler (*Rater*), unterschiedlicher Facetten eines Konstrukts oder verschiedener Messgelegenheiten analysiert werden. Liegen wiederholte Messungen vor, bieten sich Modelle der Latent-State-Trait-Theorie (LST-Theorie, vgl. ► Kap. 26; Steyer 1987, 1989; Steyer et al. 1992; Steyer et al. 2015) an. LST-Modelle teilen viele strukturelle Bestandteile mit den MTMM-Modellen mit latenten Variablen, entstammen jedoch einer eigenständigen Forschungstradition.

Psychologische Messungen unterliegen Einflüssen der Messmethode und der Situation

27.1.1 MTMM-Modelle

MTMM-Modelle können herangezogen werden, um die konvergente und die diskriminante Validität von psychologischen Messungen, beispielsweise von Fremdeinschätzungen (Lehrerratings) und Selbsteinschätzungen (Schülerratings) der Ängstlichkeit und Depressivität von Kindern zu bestimmen. Mithilfe von MTMM-Modellen lassen sich zunächst die Messfehler von den wahren Werten trennen. Die wahren Werte können dann in die Bestandteile, die auf den Einfluss des Konstrukts (z. B. Ängstlichkeit) und der Messmethode (Lehrer- oder Schülerrating) zurückzuführen sind, zerlegt werden. Je nach ausgewähltem Modell (s. Eid et al. 2006; Eid et al. 2008) erhält man einen gemeinsamen Faktor als Trait und zusätzlich Abweichungsvariablen für die Selbstratings der Schüler sowie für die Fremdratings der Lehrer als Methodenfaktoren. Beispiele hierfür sind Correlated-Trait-Unrelated-Method- (CTUM-) oder Correlated-Trait-Correlated-Method-Modelle (CTCM-Modelle, ► Abschn. 25.6). In diesen Modellen werden die Trait-Variablen wie ein Faktor in der Faktorenanalyse bestimmt. Alle Urteile mit allen Messmethoden tragen zur Schätzung des Faktorwertes (Ausprägung des Traits) bei. Die gemeinsamen systematischen Abweichungen der Urteile

Konvergente und diskriminante Validität

innerhalb der Methoden (also der gemeinsame Varianzanteil der Urteile, der nicht durch den Faktor erklärt werden kann) bilden die Methodenvariable.

Im Sinne des *Correlated-Trait-Correlated-(Method-minus-1)-Modells*, kurz CTC($M - 1$)-Modell (► Abschn. 25.7; Eid 2000), kann eine Methode als Standard gewählt werden, die allein zur Schätzung der Trait-Variablen herangezogen wird; die anderen Methoden werden gegen diese *Standardmethode* kontrastiert. Dienen beispielsweise die Ratings der Schüler als Standardmethode, so entspricht die Trait-Ausprägung dem True-Score-Wert der Schülerratings (also dem wahren Wert im Sinne der Klassischen Testtheorie, KTT, ► Kap. 13). In der True-Score-Variable der Schüler sind somit sowohl Einflüsse des tatsächlich zu messenden Merkmals als auch der Methode (Selbsteinschätzung) enthalten. Die Trait-Variable, die mit der Standardmethode gemessen wird, wird explizit als Variable aufgefasst, die diese beiden Komponenten enthält. Aus diesem Grund muss die Wahl der Standardmethode aus theoretischen Überlegungen erfolgen, wobei z. B. die Methode als Standard gewählt werden kann, von der die beste Einschätzung des zu messenden Merkmals zu erwarten ist. Die Trait-Variable dient als Prädiktor in einer latenten Regression zur Vorhersage der True-Score-Variablen der Lehrer (Nichtstandardmethoden). Die Abweichungen der True-Score-Werte der Lehrer werden dann in den Methodenfaktoren abgebildet (s. Eid et al. 2003; Eid et al. 2006, 2008; ► Kap. 25). Die Methodenfaktoren sind in diesem Modell Residuen einer latenten Regression; sie bilden Abweichung der Nichtstandardmethoden von der Vorhersage durch die Standardmethode ab.

27.1.2 Methodeneffekte

Im Rahmen von MTMM-Modellen gibt es verschiedene Möglichkeiten, Methodeneffekte zu konzeptualisieren. In den vorangegangenen Kapiteln (► Kap. 24, 25 und 26) wurden bereits mehrere auf Strukturgleichungsmodellen aufbauende Ansätze vorgestellt. Zur Modellierung ist es notwendig zu unterscheiden, welche inhaltlichen Fragen mit den Modellen beantwortet werden sollen und welche Datenstruktur vorliegt. Es geht darum, die Frage zu erörtern, wie der manifeste Messwert eines Konstrukts Y_{ikmt} ¹ (i = Indikator, in der Regel Items bzw. Itempäckchen², k = Konstrukt, m = Messmethode/Beurteiler/Rater/Informant und t = Messgelegenheit/Zeitpunkt/Occasion) mit bedeutungsvollen latenten Variablen erklärt werden kann.

Dabei ist es wichtig, die Art der eingesetzten Methoden zu beachten (Eid 2006; Eid et al. 2006, 2008). Messmethoden können untereinander austauschbar, strukturell unterschiedlich oder gleichwertig sein:

Austauschbare Methoden

a. *Die Methoden sind untereinander austauschbar:*

- Dieser Fall tritt z. B. ein, wenn mehrere Messmethoden (Beurteiler) eingesetzt werden, sich keine der Methoden von den anderen abhebt und die Messmethoden somit statistisch einer Zufallsauswahl entsprechen. Dies wäre der Fall, wenn Schüler einer Klasse per Zufall ausgewählt werden, um die Qualität des Unterrichts einzuschätzen. Alle Schüler entstammen der gleichen Population (Schüler einer Klasse) und liefern somit gleichwertige Einschätzungen der Lehrqualität.
- Ein weiteres Beispiel sind Messzeitpunkte, z. B. wenn Schüler in regelmäßigen Abständen in Bezug auf ihre Ängstlichkeit befragt würden. Dabei darf es keine Kriterien für die Auswahl der Messgelegenheiten geben, die mit

1 Der Index m beschreibt im Unterschied zu den vorangegangenen Kapiteln nicht die Zahl der Items (I, \dots, m), sondern eine ausgewählte Methode m .

2 Um metrische Indikatoren zu erhalten, können mehrere Items, die ein Konstrukt messen, gemittelt werden. Die neu entstandene Variable nennt man Itempäckchen (Item-Parcel). Dieses Vorgehen ist nicht gänzlich unumstritten (s. dazu Little et al. 2002).

der Ausprägung des Konstrukts zusammenhängen, sodass sich die Schüler zu jeder Messgelegenheit in einer zufälligen Situation befinden, deren Einfluss auf das Konstrukt unbekannt ist. Dies wäre nicht mehr der Fall, wenn sie vor und nach einem Selbstbewusstseinstraining befragt würden, da davon auszugehen ist, dass die Ängstlichkeitseinschätzungen nach dem Training geringer ausfallen sollten. Für den Fall austauschbarer Methoden (Occasions) ist es sinnvoll, den Trait als einen gemeinsamen Faktor zu definieren (z. B. im CTUM- oder LST-Modell; Eid et al. 2008; ■ Abb. 27.1a). Die Abweichungen der austauschbaren Methoden (einer Messgelegenheit: O_{+kmt} von Occasion) vom gemeinsamen Faktor sind oft von substantiellem Interesse. Erklärende Konstrukte können evtl. herangezogen werden, um die Abweichung der Methoden vom gemeinsamen Faktor zu erklären.

- b. *Die Methoden sind nicht austauschbar, sie unterscheiden sich strukturell:* Werden z. B. Schüler und Lehrer zur Ängstlichkeit der Schüler befragt, stehen den Beurteilern unterschiedliche Informationen zur Verfügung, da die Schüler eine Selbsteinschätzung abgeben, während die Lehrer aus der Fremdperspektive beurteilen müssen (vgl. ■ Abb. 27.1b). Im Schülerrating (Selbsteinschätzung) können die Schüler sowohl ihr eigenes Verhalten als auch ihr Erleben als Grundlage ihrer Einschätzungen nutzen. Im Lehrerrating (Fremdeinschätzung) müssen die Lehrer hingegen auf Verhaltensbeobachtungen oder Äußerungen der Schüler und Mutmaßungen über das Erleben der Schüler zurückgreifen. So gibt es Schüler, denen man ihre Angst „ansieht“, und wieder andere, die sie recht gut zu verbergen wissen. Manche Schüler sind vorsichtiger und werden alleine deshalb als ängstlicher eingeschätzt. Deshalb ist es sinnvoll, die Schülerratings als Standardmethode im CTC($M - 1$)-Modell einzusetzen und die Lehrerratings gegen diese zu kontrastieren. Ungünstig wäre es hingegen, einen Trait als das den verschiedenen Ratings gemeinsame Merkmal zu definieren: Denkt beispielsweise ein Kind, dass es das traurigste Kind der Welt ist, der Lehrer hingegen, dass es ein sehr glückliches Kind sei, so ist es nicht sinnvoll, den Mittelwert der beiden Ratings als „wahre Stimmung“ zu definieren. Viel aussagekräftiger ist es, beide Methoden zu kontrastieren, um untersuchen zu können, warum sich die Schüler- von den Lehrerratings unterscheiden.
- c. *Die Methoden sind gleichwertige Repräsentationen eines Traits,* z. B. Testhälften oder parallele Tests (■ Abb. 27.1c). In MTMM-Datensätzen kann es vorkommen, dass Indikatoren trotz sehr hoher Korrelationen mit den Indikatoren desselben Traits einen sehr spezifischen systematischen Varianzanteil binden. Dies liegt daran, dass sich die Inhalte der Items in den verschiedenen Indikatoren nicht vollständig replizieren, sondern neben den gemeinsamen sprachlichen Inhalten auch stets eigene Aspekte beinhalten.

Strukturell unterschiedliche Methoden

Gleichwertige Methoden

Vor allem in longitudinalen Modellen kommt es oft vor, dass Indikatoren über Messgelegenheiten hinweg stärker miteinander korrelieren als mit den anderen Indikatoren des gleichen Konstrukts zur gleichen Messgelegenheit (Eid 1996; Marsh und Grayson 1994; Steyer et al. 1992). Dieser *Autokorrelationseffekt* entspricht der spezifischen, zeitlich stabilen und reliablen Varianz eines einzelnen Indikators. Die Autokorrelation kann man auch als Methodeneffekt des Indikators (oder indikatorspezifischen Effekt) interpretieren.

Oft ist dieser Effekt (der Unterschied zwischen den Indikatoren) aber nicht von substantiellem Interesse, sondern eher nebensächlich, wenn es darum geht, die Varianzquellen Methode, Messgelegenheit und Messfehler vom Trait zu trennen. Aus diesem Grund spezifiziert man in vielen Anwendungen separate Trait-Faktoren für beide Testhälften, obwohl beide zum gleichen Konstrukt gehören (vgl. ► Kap. 26). In den meisten Fällen korrelieren diese Faktoren sehr hoch positiv miteinander, was darauf hinweist, dass die indicatorspezifischen Effekte eher gering ausfallen. Bei den hier vorgestellten Modellen ist jedoch davon auszugehen, dass ein Autokorrelationseffekt eintritt.

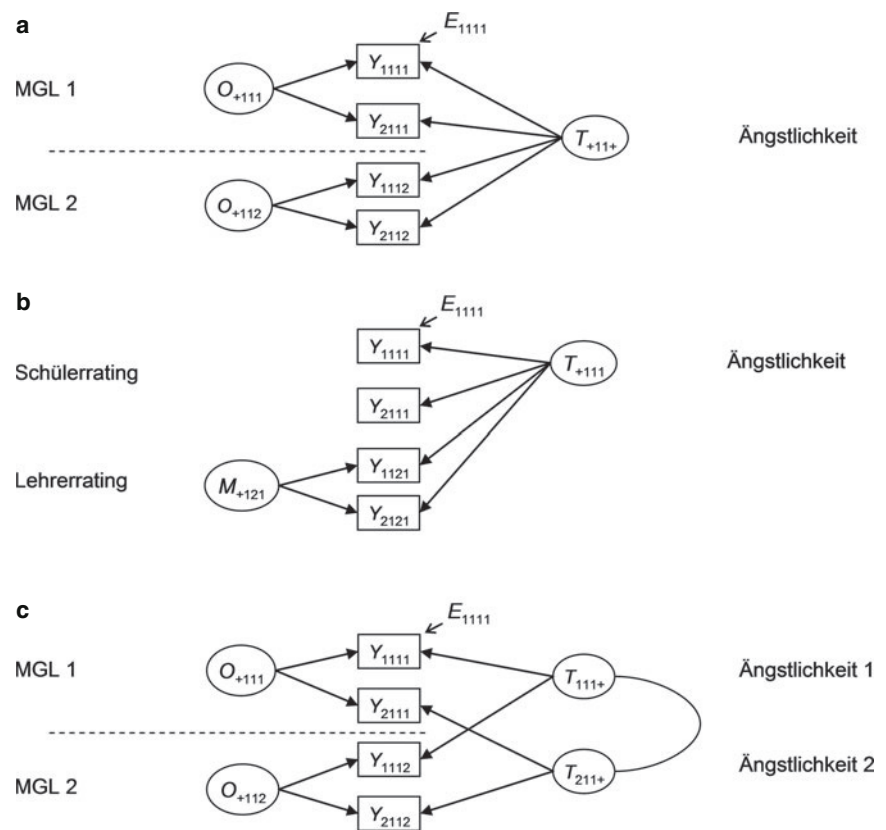


Abb. 27.1 Verschiedene Arten von Methodenfaktoren: **a** austauschbare Messgelegenheiten/Occasions (O_{+kmi}), **b** Methodenfaktor/Beurteiler als Residualfaktor [M_{+kmi} ; CTC($M - 1$)-Modell] und **c** Methodeneffekte, die durch separate Trait-Variablen abgebildet sind (T_{ikm+} und $T_{i'km+}$). MGL: Messgelegenheit; „+“ zeigt an, dass diese Variable sich auf mehrere Ziffern des betreffenden Indexes bezieht (Platzhalter). Fehlervariablen sind nur für den jeweils ersten Indikator (Item bzw. Itempäckchen) abgebildet. Die Trait-Variablen in den Modellen **a** und **c** sind als gemeinsame Faktoren definiert, sie sind zwischen den Methoden abgebildet. Die Trait-Variable im Modell **b** entspricht dem Faktor für das Schülerrating, sie ist deswegen auf der Höhe der Indikatoren des Schülerratings abgebildet

Methodeneffekte sind trait-spezifisch

Allen Konzeptualisierungen der verschiedenen Arten von Methoden ist gemeinsam, dass Methodeneffekte *trait-spezifisch* sind. Wir gehen also davon aus, dass der Effekt einer Methode bei verschiedenen Merkmalen unterschiedlich ausfällt. Daraus folgt, dass für jede Kombination von Trait und Methode ein Methodenfaktor geschätzt werden muss³. Fremdratings können beispielsweise bei Einschätzungen der Depressivität starke Verzerrungen, bei der Einschätzung der Ängstlichkeit hingegen nur geringe Verzerrungen aufweisen; ebenso können die Messwerte der Ängstlichkeit zu bestimmten Messgelegenheiten erhöht sein, die Messwerte der Depressivität hingegen nicht.

27.1.3 LST-Modelle

LST-Modelle können genutzt werden, um zeitstabile Einflüsse von messgelegenheitsspezifischen Einflüssen auf Merkmalsausprägungen zu trennen (vgl. ► Kap. 26). Hier werden ebenfalls zunächst wahre Werte und Messfehler getrennt; sodann werden die wahren Werte in zeitstabile Anteile (die Trait-Einflüsse)

³ Für die Standardmethode im CTC($M - 1$)-Modell wird jedoch kein separater Methodenfaktor spezifiziert.

und messgelegenheitsspezifische Anteile (Messgelegenheitseinflüsse) zerlegt. Die situativen Einflüsse schwanken dabei um den stabilen Trait-Wert.

Situative Schwankungen können viele Ursachen haben. Einerseits könnten die Schüler (aus obigem Beispiel) bei der Einschätzung ihrer Ängstlichkeit stark durch momentane Einflüsse gesteuert werden: Dies könnte ein vorherrschendes schweres Gewitter sein oder der gerade vergangene Aufenthalt im Schullandheim. Sie können aber auch durch nicht oder nur schwer messbare Faktoren beeinflusst sein: Dies könnte ein Albtraum in der vergangenen Nacht, eine Magenverstimmung oder ein Streit mit einem der Geschwister sein. Die messgelegenheitsspezifischen Faktoren umfassen alle möglichen momentanen Einflüsse und integrieren sie in einen sog. „inneren Zustand“.

Der wesentliche Unterschied zwischen MTMM- und LST-Modellen ist, dass in LST-Modellen typischerweise ein bestimmter Trait unter expliziter Berücksichtigung von zeitlichen Schwankungen über einen längeren Zeitraum wiederholt gemessen wird, während in MTMM-Modellen mehrere Traits ohne Berücksichtigung von zeitlichen Schwankungen mit verschiedenen Methoden gemessen werden. LST-Modelle werden vornehmlich in Längsschnittanalysen zur Bestimmung der Reliabilität, Konsistenz und Messgelegenheitsspezifität eingesetzt (vgl. ► Kap. 26; Yousfi und Steyer 2006), können aber auch zur Überprüfung inhaltlicher Hypothesen eingesetzt werden (s. dazu Courvoisier et al. 2007). MTMM-Modelle werden vornehmlich in Querschnittanalysen eingesetzt, um Konstrukte hinsichtlich ihrer konvergenten und diskriminanten Validität zu untersuchen (u. a. Eid et al. 2006, 2008).

Unterschied zwischen MTMM- und LST-Modellen

27.1.4 Beschränkungen der LST- und MTMM-Modelle

Die bislang vorgestellten Modelle erlauben eine Zerlegung der Varianz der beobachteten multimethodal erhobenen Daten in verschiedene Bestandteile. Im LST-Modell werden die Bestandteile geschätzt, die von stabilen Dispositionen (Traits), messgelegenheitsspezifischen Einflüssen und Messfehlern abhängen. In MTMM-Modellen werden die Bestandteile geschätzt, die von stabilen Dispositionen (Traits), unterschiedlichen Messmethoden (Methoden) und Messfehlern abhängen. Im ersten Fall können Hypothesen getestet werden, wie sehr Messungen über die Zeit um einen stabilen Trait-Wert schwanken und ob ein Trait tatsächlich zeitlich stabil ist; im zweiten Fall kann die konvergente und die diskriminante Validität psychologischer Messungen überprüft werden. In konventionellen LST-Modellen kann jedoch die Hypothese, dass Lehrer- und Schülerratings übereinstimmen, nicht überprüft werden. In MTMM-Modellen kann nicht überprüft werden, ob die konvergente und diskriminante Validität der Messungen zeitlich stabil sind.

Gerade in den empirischen Sozialwissenschaften können wir jedoch nicht uneingeschränkt davon ausgehen, dass die Einflüsse unterschiedlicher Methoden auf ein Messergebnis über die Zeit stabil bleiben oder sich homogen verändern. Dies bedeutet, dass das Ergebnis einer querschnittlichen MTMM-Studie nicht ohne Zusatzannahmen auf spätere Zeitpunkte übertragen werden kann. Wenn zu einer Messgelegenheit eine hohe Konvergenz von Lehrer- und Schülerratings festgestellt werden konnte, können wir nicht automatisch davon ausgehen, dass sich ein halbes Jahr später keine Veränderungen der Übereinstimmung der beiden Ratings zeigen werden.

Die Entwicklung der konvergenten und diskriminanten Validität über die Zeit hinweg kann nur mithilfe *längsschnittlicher MTMM-Modelle* untersucht werden. Ebenso kann die konvergente und diskriminante Validität auf der Ebene von Traits und messgelegenheitsspezifischen Einflüssen untersucht werden. Dabei geht es um die stabilen Anteile der konvergenten Validität, die auf das überdauernde Merkmal

Längsschnittliche MTMM-Modelle

zurückzuführen sind und um variable Anteile der konvergenten Validität, die dadurch zustande kommen, dass sich Effekte der Situation und der Interaktion von Situation und Person auf alle Methoden gleichmäßig auswirken.

27.1.5 Verbindung beider Ansätze

Längsschnittliche Analyse der konvergenten und diskriminanten Validität

Bislang sind wenige Modelle formuliert worden, die eine Verbindung der beiden Ansätze ermöglichen. Aber gerade eine *längsschnittliche Analyse der konvergenten und diskriminanten Validität* kann von sehr großem Nutzen sein, da Querschnittmodelle immer nur eine Momentaufnahme sein können. Es ist plausibel anzunehmen, dass situative Schwankungen oder situative Merkmale zu einer erheblichen Erhöhung oder Verringerung der Koeffizienten der konvergenten und diskriminanten Validität führen.

Beispiel aus der Persönlichkeitspsychologie

So könnte die konvergente Validität zwischen zwei Beurteilern und die diskriminante Validität zwischen den Konstrukten „Extraversion“ und „Verträglichkeit“ bei einer einzigen Messgelegenheit verzerrt sein, z. B. bei einer studentischen Stichprobe, bei der kurz vor der MTMM-Untersuchung die Big-Five-Persönlichkeitsfaktoren (Costa und McCrae 1998) Gegenstand in der Vorlesung zur Persönlichkeitspsychologie waren. Nach der Veranstaltung haben die studentischen Selbst- und Fremdbeurteiler die Konstrukte „Extraversion“ und „Verträglichkeit“ besonders gut verstanden und evtl. mit Kommilitonen darüber diskutiert. Sie haben sich vielleicht sogar im Hinblick auf diese beiden Konstrukte untereinander verglichen. In diesem Moment wird die Studie im Vergleich zu „normalen“ Bedingungen verzerrte konvergente und diskriminante Validitäten aufweisen, da die Studierenden die Konstrukte gerade kognitiv besser verfügbar haben und feiner zwischen ihnen unterscheiden können, als dies üblicherweise der Fall wäre.

Beispiel aus der Entwicklungspsychologie

Ein anderes Beispiel kommt aus der Entwicklungspsychologie: Kinder reifen sehr schnell, lernen viele neue Verhaltensweisen und entwickeln in kurzer Zeit neue Fähigkeiten. Cole und Martin (2005) berichten, dass die Selbsteinschätzung der Depressivität bei Kindern zunächst stark von situativen Faktoren abhängt. Bei älteren Kindern im Übergang zur Pubertät werden die Einschätzungen jedoch stabiler und das Konstrukt der Depressivität bekommt stärker den Charakter eines Traits. Eltern hingegen schätzen die Depressivität ihrer Kinder von Beginn an eher wie einen stabilen Trait ein, sodass die Konvergenz der Ratings im Laufe der Entwicklung zunehmen sollte.

Variable Zustände

Mit der Kombination von MTMM- und LST-Modellen können Analysen der konvergenten und diskriminanten Validität über die Zeit mit systematischen Veränderungen auch für *variable Zustände* vorgenommen werden. Variable Zustände wie Stimmungen zeichnen sich dadurch aus, dass sie über die Zeit schwanken. Sie variieren jedoch nicht beliebig, sondern um ein „mittleres Niveau“, das für jede Person unterschiedlich sein kann. Für Stimmungsforscher ist es daher interessant, die konvergente und diskriminante Validität auf der Ebene der Traits und auf der Ebene der Messgelegenheiten zu untersuchen. Die Frage, ob sich dieselben situativen Effekte auf mehrere Methoden homogen auswirken, ist dabei von besonderem Interesse.

Im diagnostisch-therapeutischen Kontext könnten folgende Fragen von hoher Relevanz sein: Wie stark konvergieren Selbst- und Lehrereinschätzungen in Bezug auf die Ängstlichkeit und Depressivität? Generalisieren Lehrer im Sinne eines Halo-Effekts (Thorndike 1920), kommen sie also zu nahezu identischen Einschätzungen von Ängstlichkeit und Depressivität? Wie sehr hängen die Einschätzungen der Lehrer von stabilen oder situativen Faktoren ab? Können Lehrer die Schwankungen der Ängstlichkeit ihrer Schüler nachvollziehen?

Diese Fragen können weder mit LST- noch mit MTMM-Modellen umfassend beantwortet werden, sondern müssen in Kombinationen der beiden Ansätze unter-

sucht werden. In ► Abschn. 27.2 werden die folgenden drei Modelle vorgestellt, die eine Beantwortung dieser Fragen ermöglichen:

1. Multioccasion-MTMM-Modell (z. B. Burns und Haynes 2006)
2. Multiconstruct-LST-Modell (Schermelleh-Engel et al. 2004)
3. Multimethod-LST-Modell (Courvoisier 2006; Courvoisier et al. 2008)

Die drei Ansätze unterscheiden sich darin, wie sie den Verlauf über die Zeit und die Einflüsse unterschiedlicher Methoden in die Modellierung aufnehmen. Da das LST-Modell strukturell einem CTUM-Modell entspricht (Marsh und Grayson 1995), werden im folgenden Abschnitt auch Messgelegenheiten als Methoden aufgefasst.

27.2 Längsschnittliche MTMM-Modelle

Längsschnittuntersuchungen, in denen mehrere Konstrukte mit mehreren Methoden wiederholt gemessen werden, sind aufwendig erhobene Datensätze, die gewissen Anforderungen genügen müssen. Die Konstrukte sollten zu allen Messgelegenheiten mit den gleichen Methoden erhoben werden. Darüber hinaus sollten identische Indikatoren zur Messung der einzelnen Trait-Methoden-Einheiten erhoben werden. Aufgrund dieser recht hohen Anforderungen an die Daten und die Komplexität der zu analysierenden Strukturgleichungsmodelle ist es kaum verwunderlich, dass bislang kaum Längsschnitt-MTMM-Modelle vorgestellt wurden.

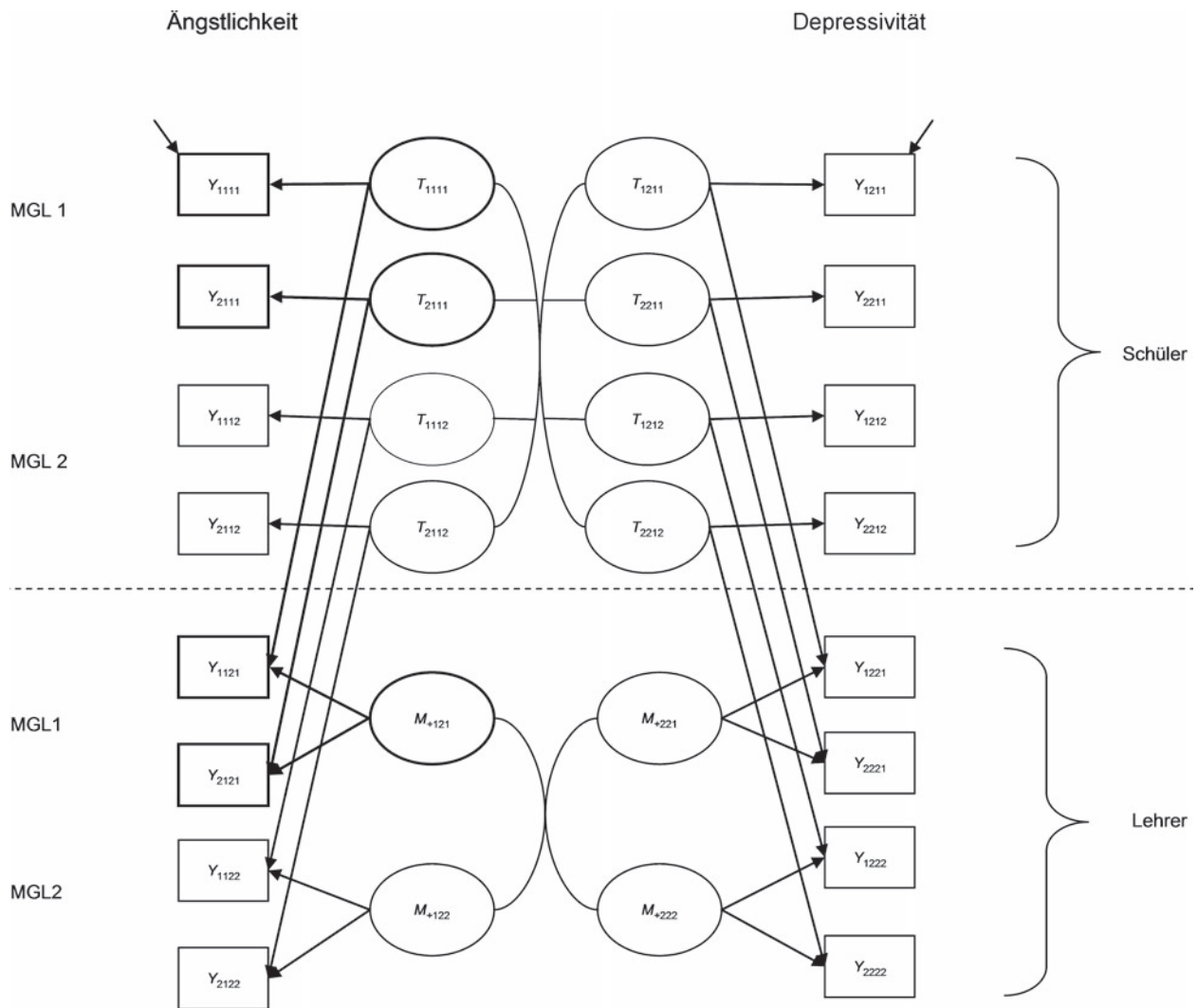
Anforderungen an die Daten

27.2.1 Multioccasion-MTMM-Modell

Burns und Haynes (2006) stellen eine längsschnittliche Erweiterung des CTCM-Modells (Marsh und Grayson 1995) vor, in dem zu jeder der Messgelegenheiten (Occasions) ein MTMM-Modell geschätzt wird. In ihrem Multioccasion-MTMM-Modell kann die *Stabilität von States und Methodeneffekten* überprüft werden. Da zu jeder Messgelegenheit ein MTMM-Modell geschätzt wird, sind die „Trait-Variablen“ dieser Modelle State-Variablen im Sinne der LST-Theorie. Im Folgenden werden wir deshalb von States sprechen. Die Korrelationen derselben State-Variablen zu unterschiedlichen Messzeitpunkten gibt die Stabilität der Rangplätze der Testpersonen auf dem Merkmal wieder. Korrelationen zwischen den Methodenvariablen zu verschiedenen Messgelegenheiten zeigen an, ob sich die Einflüsse der Methoden über die Zeit verändern. Überträgt man den Ansatz des CTC($M - 1$)-Modells auf den Vorschlag von Burns und Haynes (2006), erhält man ein *Multioccasion-Related-States-Related-(Method-minus-1)-Modell*, kurz Multioccasion-CSC($M - 1$)-Modell (s. auch Geiser et al. 2008).

Multioccasion-CSC($M - 1$)-Modell

Das Multioccasion-CSC($M - 1$)-Modell in ■ Abb. 27.2 kann dazu genutzt werden, die Stabilität von Schülern im Hinblick auf ihre Depressivitäts- und Ängstlichkeitszustände zu analysieren. Die bivariate Korrelation zwischen den Depressivitätsvariablen (oder Ängstlichkeitsvariablen) zur ersten und zweiten Messgelegenheit spiegelt diese Stabilität wider. Die Methodenfaktoren stellen die Abweichungen der Lehrer- von den Schülerratings dar. Korrelationen zwischen den Methodenfaktoren zeigen somit an, ob Lehrer, die die Depressivität einer Schülerin oder eines Schülers zur ersten Messgelegenheit überschätzten (unterschätzten), dies auch zur zweiten Messgelegenheit tun. Die trait-spezifischen Methodenfaktoren der Lehrer (M_{+kmt} ; „+“ zeigt an, dass sich der Methodenfaktor auf beide Indikatoren auswirkt) dürfen auch über die Konstrukte hinweg korrelieren (in ■ Abb. 27.2 nicht dargestellt). Diese Korrelationen geben den Grad der Generalisierbarkeit von Methodeneffekten an. Je höher diese Korrelation ausfällt, desto homogener ist der Einfluss der Methode auf die Verzerrung der Lehrerratings. Sie



■ **Abb. 27.2** Multioccasion-CSC($M - 1$)-Modell für zwei Traits, gemessen mit zwei Methoden zu zwei Messgelegenheiten. T_{ikm+} : Trait-Faktor; M_{+km+} : Methodenfaktor; MGL: Messgelegenheit. Fehlervariablen sind nur für die ersten Indikatoren abgebildet

gibt an, ob Lehrer die Depressivität und die Ängstlichkeit der Kinder in gleichem Maße über- bzw. unterschätzen. Die Methodenfaktoren können auch über Traits und Messgelegenheiten hinweg korrelieren. Diese Korrelationen geben z. B. an, ob Lehrer, die zu einer Messgelegenheit die Depressivität überschätzen, die Ängstlichkeit zu einer späteren Messgelegenheit ebenfalls überschätzen.

27.2.2 Multiconstruct-LST-Modell

Multiconstruct-LST-Modell

Das *Multiconstruct-LST-Modell* (Dumenci und Windle 1998; Eid et al. 1994; Majcen et al. 1988; Schermelleh-Engel et al. 2004; Schmitt 2000; Steyer 1989; Steyer et al. 1990) wird zur Erfassung mehrerer Konstrukte im zeitlichen Verlauf eingesetzt. Im Gegensatz zu dem von Burns und Haynes (2006) vorgeschlagenen Modell werden hier nicht zu jeder Messgelegenheit MTMM-Modelle geschätzt, sondern für jede Trait-Methoden-Einheit ein LST-Modell (■ Abb. 27.3).

Wie in den klassischen LST-Modellen gibt es eine indikatorspezifische Trait-Variablen (T_{ikm+}) für jede Kombination von Indikator, Trait und Methoden, d. h., es gibt beispielsweise zwei indikatorspezifische Trait-Variablen für die Einschätzung

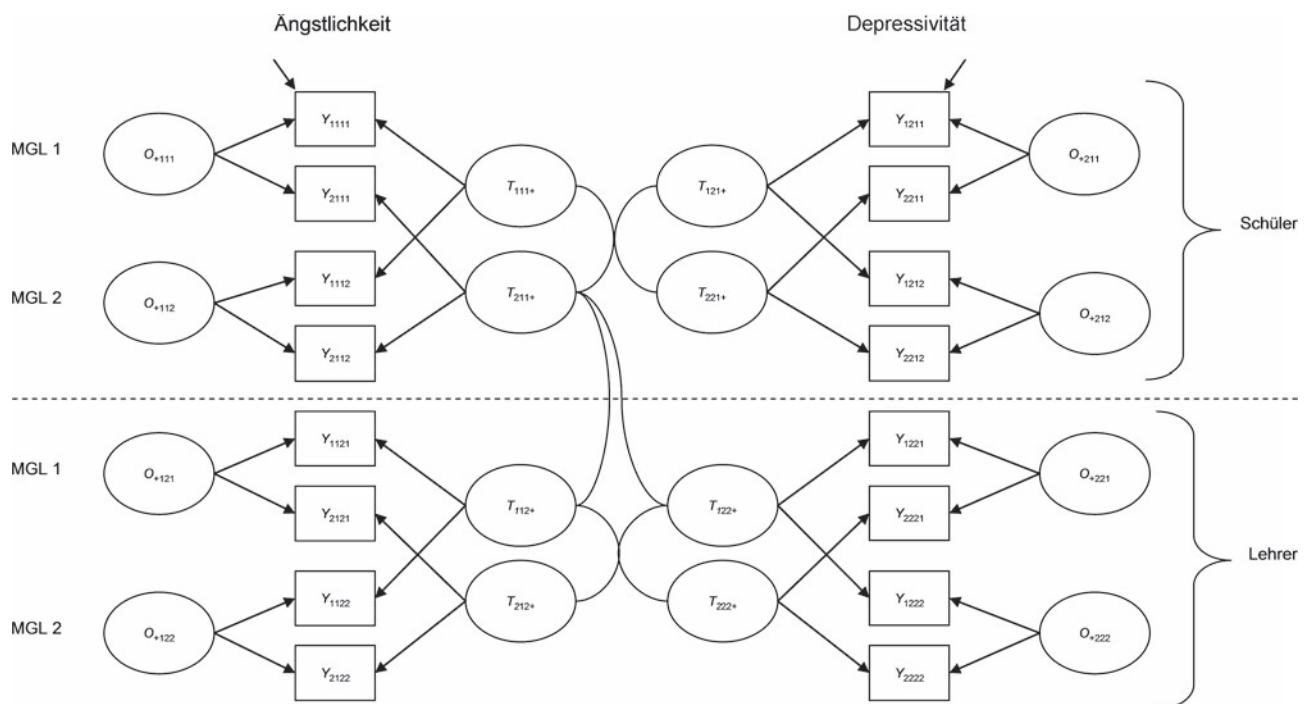


Abb. 27.3 Multiconstruct-LST-Modell. T_{ikm+} : Trait-Variablen; O_{+kmt} : Messgelegenheitsvariablen; MGL: Messgelegenheit. Die zulässigen Korrelationen zwischen den Messgelegenheitsvariablen sind aus Gründen der Lesbarkeit nicht alle abgebildet. Korrelationen zwischen Trait-Variablen sind nur exemplarisch eingezeichnet. Prinzipiell korrelieren alle Trait-Variablen miteinander. Fehlervariablen sind nur für die ersten Indikatoren dargestellt

der Ängstlichkeit durch die Schüler und zwei indikatorspezifische Trait-Variablen für die Einschätzung durch die Lehrer. Die Korrelationen dieser Trait-Variablen eines Konstrukts, gemessen mit unterschiedlichen Methoden, geben in diesem Modell die konvergente Validität an, so z. B. $Corr(T_{1k1+}, T_{1k2+})$, die Korrelation der ersten Trait-Variablen ($i = 1$) für ein beliebiges Konstrukt (k), gemessen anhand des Schülerratings ($m = 1$) und des Lehrerratings ($m = 2$). Die Korrelationen verschiedener Trait-Variablen für unterschiedliche Konstrukte (Ängstlichkeit und Depressivität) geben die diskriminante Validität an, so z. B. $Corr(T_{111+}, T_{121+})$, die Korrelation der ersten Trait-Variablen ($i = 1$) der beiden Konstrukte ($t = 1$ und $t = 2$), gemessen anhand des Schülerratings ($m = 1$). Korrelationen zwischen einer Trait-Variablen und der Trait-Variablen eines anderen Konstrukts, das mit einer anderen Methode gemessen wurde, spiegeln ebenfalls die diskriminante Validität wider, wobei in diese zusätzlich die Unterschiede der Methoden einfließen, so z. B. $Corr(T_{111+}, T_{122+})$, die Korrelation der ersten Trait-Variablen ($i = 1$) der beiden Konstrukte ($t = 1$ und $t = 2$), gemessen anhand des Schülerratings ($m = 1$) und des Lehrerratings ($m = 2$).

Die Korrelation der messgelegenheitsspezifischen Variablen einer Methode und einer anderen Methode desselben Konstrukts zu einer Messgelegenheit spiegelt dabei wider, ob die beiden Methoden zu dieser Messgelegenheit in gleicher Weise vom jeweiligen stabilen Trait abweichen, so z. B. $Corr(O_{+111}, O_{+121})$, die Korrelation der situationsspezifischen Abweichungen vom stabilen Trait eines Konstrukts ($k = 1$), gemessen anhand des Schülerratings ($m = 1$) und des Lehrerratings ($m = 2$) zur gleichen Messgelegenheit ($k = 1$). Dies entspricht einem Effekt der Messgelegenheit, der sich auf die Messungen beider Methoden auswirkt und gleichzeitig ein weiteres Kennzeichen für die konvergente Validität darstellt. Die Korrelation zweier messgelegenheitsspezifischer Variablen einer Methode, die zu zwei Konstrukten gehören, zeigt den Einfluss der Messgelegenheit auf beide Konstrukte an, so z. B. $Corr(O_{+111}, O_{+211})$, die Korrelation der situationsspezifischen

Abweichungen vom stabilen Trait zweier Konstrukte ($k = 1$ und $k = 2$), gemessen anhand des Schülerratings ($m = 1$) zur gleichen Messgelegenheit ($k = 1$); sie ist ein Maß für die diskriminante Validität der Konstrukte auf messgelegenheitsspezifischer Ebene. Die Korrelationen zwischen den messgelegenheitsspezifischen Variablen einer Methode mit den messgelegenheitsspezifischen Variablen der anderen Methode sind auch über die Zeit hinweg erlaubt, jedoch nicht einfach zu interpretieren. Sie geben keinen Hinweis auf die konvergente oder diskriminante Validität und werden deshalb hier nicht weiter erläutert (s. dazu Courvoisier 2006).

Anwendung auf das Beispiel

Wendet man das Multiconstruct-LST-Modell erneut auf das Beispiel der Ängstlichkeits- und Depressivitätsbeurteilungen bei Schulkindern an, so müssten vier LST-Modelle geschätzt werden:

1. Depressivität im Schülerrating
2. Depressivität im Lehrerrating
3. Ängstlichkeit im Schülerrating
4. Ängstlichkeit im Lehrerrating

Findet sich eine hohe Korrelation der Schüler- und Lehrerratings für die Einschätzungen der Depressivität, ist dies ein Beleg für die Übereinstimmung der Methoden und somit für die konvergente Validität. Die Korrelationen der Trait-Variablen für Ängstlichkeit und Depressivität des Schülerratings (Lehrerratings) gibt die diskriminante Validität an. Die Korrelation der messgelegenheitsspezifischen Variablen der Schüler und der Lehrer eines Konstrukts zu einer Messgelegenheit spiegelt wider, ob die beiden Methoden in gleicher Weise vom jeweiligen stabilen Trait abweichen (konvergente Validität). Schätzt sich der Schüler momentan erhöht auf der Skala der Depressivität ein, so ist dies auch tendenziell für den Lehrer der Fall (bei positiver Korrelation).

Im Gegensatz zum Multioccasion-MTMM-Modell können in diesem Modell die stabilen Varianzkomponenten von den situativ bedingten Varianzkomponenten und den Residualkomponenten getrennt werden. Eine Erweiterung des Multiconstruct-LST-Modells zum hierarchischen LST-Modell findet sich bei Schermelleh-Engel et al. (2004).

27.2.3 Multimethod-LST-Modell

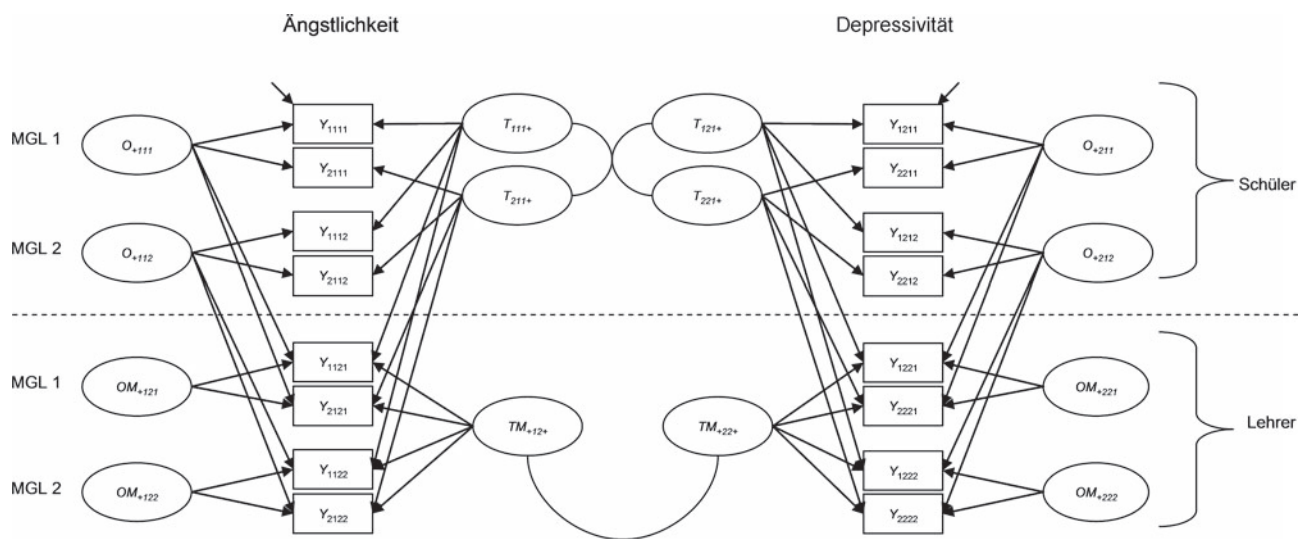
Verknüpfung von LST- und MTMM-Modellen

Das Multioccasion-MTMM- und das Multiconstruct-LST-Modell ermöglichen die Analyse unterschiedlicher Aspekte der Daten. Beim ersten Modell steht der MTMM-Charakter, beim zweiten der LST-Charakter stärker im Vordergrund. Je nach wissenschaftlicher Fragestellung kann eines dieser Modelle ausgewählt werden und entsprechend aussagekräftige Resultate liefern. Liegt das Augenmerk jedoch auf der Analyse der konvergenten und diskriminanten Validität über die Zeit und möchte man die unterschiedlichen Varianzkomponenten identifizieren, die durch den Trait, die Methode, die Messgelegenheit und den Messfehler bedingt sind, so muss die Verknüpfung von LST- und MTMM-Modellen noch stärker erfolgen als in den vorangegangenen Modellen. Courvoisier (2006) und Courvoisier et al. (2008) stellen das Multimethod-LST-Modell vor, bei dem die longitudinalen und multimethodalen Bestandteile der einzelnen Modelle nicht nebeneinander gestellt, sondern ineinander verschränkt werden (■ Abb. 27.4).

Obere Hälfte des abgebildeten Modells

Das hier vorgestellte Modell verbindet die Eigenschaften des LST-Modells mit denen des CTC($M - 1$)-Modells. Das Modell lässt sich in sinnvolle Submodelle unterteilen. In der oberen Hälfte des abgebildeten Modells befinden sich zwei klassische LST-Modelle für die Standardmethode (die Schüler).

Die indikatorspezifischen Trait-Variablen (T_{ik1+}) entsprechen den stabilen Komponenten der Standardmethode, also ihren stabilen Trait- und Methodeneinflüssen in Bezug auf einen Indikator. Wie im klassischen CTC($M - 1$)-Modell ist



■ **Abb. 27.4** Multimethod-LST-Modell. T_{ikm+} : Trait-Variable; TM_{+km+} : Methodenvariable; O_{+kmt} : messgelegenheitsspezifische Variable; OM_{+kmt} : messgelegenheitsspezifische Abweichungsvariable; *MGL*: Messgelegenheit. Die zulässigen Korrelationen zwischen sämtlichen messgelegenheitsspezifischen Variablen sind aus Gründen der Lesbarkeit nicht abgebildet. Fehlervariablen sind nur für die obersten Indikatoren dargestellt

es wichtig, die Standardmethode anhand theoretischer Überlegungen zu wählen, da sämtliche Methodeneinflüsse und deren Interaktionen mit dem Trait in die Trait-Variable einfließen.

Die messgelegenheitsspezifischen Variablen (O_{+kt}) entsprechen den situationsspezifischen Abweichungen der Standardmethode (Schüler) von den stabilen Trait-Variablen.

Die Methodenfaktoren der Nichtstandardmethoden (TM_{+km+} , Lehrer) entsprechen dem zeitlich stabilen Bias der Nichtstandardmethoden gegenüber dem Standard, d. h. der stabilen positiven oder negativen Abweichung vom erwarteten Rating.

Die messgelegenheitsspezifischen Abweichungsvariablen (OM_{+kmt}) entsprechen den situationsspezifischen Abweichungen der Nichtstandardmethoden (Lehrer) von den durch die ersten drei Variablen (T_{ik1+} , O_{+k1t} und TM_{+km+}) vorhergesagten Werten. Hierdurch kann geprüft werden, ob das Lehrerrating stärker oder schwächer abweicht, als es aufgrund des Schülerratings (stabiler Anteil und situationsspezifischer Anteil) und der stabilen Abweichung des Lehrers zu erwarten ist.

Die Korrelationen zwischen den Variablen der Standardmethode haben praktisch dieselben Bedeutungen wie im Multiconstruct-LST-Modell. Korrelationen zwischen Trait-Variablen spiegeln die konvergente oder diskriminante Validität, Korrelationen der messgelegenheitsspezifischen Variablen zu einer Messgelegenheit den trait-übergreifenden Einfluss der Messgelegenheit wider.

Die Korrelationen in der unteren Hälfte des abgebildeten Modells sind aus der Perspektive des CTC($M - 1$)-Modells zu interpretieren. Die Korrelationen zwischen den Methodenfaktoren zeigen den generellen Methodenbias, so z. B. $Corr(TM_{+12+}, TM_{+22+})$. Korrelationen zwischen den messgelegenheitsspezifischen Abweichungsfaktoren zu einer Messgelegenheit (z. B. $Corr(OM_{+121}, OM_{+221})$), in ■ **Abb. 27.4** nicht eingezeichnet) geben den merkmalsübergreifenden Einfluss der Messgelegenheit, der sich ausschließlich auf die Nichtstandardmethode auswirkt, wieder (die gemeinsame Auswirkung auf die Standard- und Nichtstandardmethode steckt bereits in dem messgelegenheitsspezifischen Faktor O_{+k1t} der Standardmethode).

Bezogen auf das Beispiel der Ängstlichkeit und der Depressivität von Schulkindern bietet es sich an, das Schülerrating als Standardmethode auszuwählen, um die Lehrerratings vorherzusagen. Die Interpretation der Variablen für die Schüler

Untere Hälfte des abgebildeten Modells

Anwendung auf das Beispiel

entspricht der Interpretation derselben Variablen im klassischen LST-Modell. Ihre manifesten Variablen werden in einen stabilen Anteil (T) und einen messgelegenheitsspezifischen Anteil (O) zerlegt. Die Ratings der Lehrer werden ebenfalls durch diese Variablen vorhergesagt. Darüber hinaus wird die stabile Abweichung der Lehrer in der Methodenvariable abgebildet (TM). Dies ist die stabile Abweichung (Über- oder Unterschätzung) der Lehrerratings von den Erwartungen aufgrund der stabilen Schülerratings, sie ist unabhängig von situativen Einflüssen auf die Lehrerratings. Situationsspezifische Abweichungen der Lehrerratings werden durch die messgelegenheitsspezifische Abweichungsvariable abgebildet (OM). Diese Variable zeigt Veränderungen des Bias der Lehrer von ihren stabilen Abweichungen an. Sie kann als interner Zustand der Lehrer aufgefasst werden, der nicht mit dem internen Zustand der Schüler geteilt wird. Beispielsweise wären gemeinsame Einflüsse auf Schüler- und Lehrerratings durch die bevorstehenden Ferien in der messgelegenheitsspezifischen Variablen (O) erfasst; Einflüsse, beispielsweise die zusätzliche Arbeit, die fälligen Zeugnisse auszustellen, beeinflussen nur die Lehrer und werden in der messgelegenheitsspezifischen Abweichungsvariablen (OM) abgebildet.

Die Konvergenz der Lehrer- und Schülerratings ergibt sich aus dem Anteil der Varianz der Lehrerratings, der durch den Trait *und* die messgelegenheitsspezifische Variable der Schüler vorhergesagt werden kann. Sie besteht also aus stabilen und variablen Anteilen. Ob die Lehrer die Schwankungen der Schüler in Bezug auf die Depressivität (Ängstlichkeit) richtig nachvollziehen können, zeigt sich in der Größe des Einflusses der messgelegenheitsspezifischen Variablen der Schüler (O_{+k1t}) auf die beobachteten Variablen der Lehrer (dieser Einfluss wäre bei optimaler Übereinstimmung genau so hoch wie der Einfluss dieser Variablen auf die Schülerratings). Die Unterschiede zwischen den beiden Methoden werden durch zwei Faktoren abgebildet. Zunächst wirkt sich der stabile Bias der Lehrer aus, d. h., ob ein Lehrer die Depressivität (Ängstlichkeit) eines Schülers konsistent höher oder niedriger einschätzt als erwartet. Da sich aber auch der Lehrer zu jeder Messgelegenheit in einer bestimmten Situation befindet, wird noch eine messgelegenheitsspezifische Abweichungsvariable in Betracht gezogen, die den momentanen Bias der Lehrer darstellt (OM_{+kmt}). Die diskriminante Validität der beiden Konstrukte „Ängstlichkeit“ und „Depressivität“ zeigt sich in der Höhe der Korrelation der beiden Trait-Variablen, so z. B. in $Corr(T_{111+}, T_{121+})$.

Generalisierbarkeit eines messgelegenheitsspezifischen Einflusses

Die Generalisierbarkeit eines messgelegenheitsspezifischen Einflusses kann über die Korrelation der messgelegenheitsspezifischen Variablen zu einer Messgelegenheit ermittelt werden ($Corr(O_{+111}, O_{+211})$, homogener Einfluss der Messgelegenheit auf die Schülerratings). Ob der stabile Anteil des Bias der Lehrer über beide Konstrukte generalisiert, kann an der Korrelation $Corr(TM_{+12+}, TM_{+22+})$ der beiden Methodenvariablen abgelesen werden. Inwiefern der instabile (momentane) Bias der Lehrer für die Ängstlichkeit und die Depressivität identisch ist, kann mit der Korrelation $Corr(OM_{+121}, OM_{+221})$ der beiden messgelegenheitsspezifischen Abweichungsvariablen zu einer Messgelegenheit überprüft werden.

Im Multimethod-LST-Modell sind noch weitere Korrelationen latenter Variablen zulässig, die wir hier aber nicht besprechen werden, da sie für die Bestimmung der konvergenten und der diskriminanten Validität im Längsschnitt von nachgeordneter Priorität sind. Nicht erlaubt sind Korrelationen von Variablen, die in derselben Modellgleichung für eine bestimmte beobachtete Variable vorkommen. So sind die Trait-Variablen (T), die messgelegenheitsspezifischen Variablen (O), die Methodenvariablen (TM) und die messgelegenheitsspezifischen Abweichungsvariablen (OM) unkorreliert, wenn sie zur Erklärung einer bestimmten beobachteten Variablen herangezogen werden. Folglich sind alle latenten Variablen einer Trait-Methoden-Einheit zu einer Messgelegenheit unkorreliert (das sind fast alle latenten

Variablen, die zu einem Konstrukt gehören)⁴. Hingegen können alle latenten Variablen einer Trait-Methoden-Einheit mit allen latenten Variablen einer anderen Trait-Methoden-Einheit korrelieren. Zu beachten ist, dass die oben aufgeführten Korrelationen am leichtesten und sinnvoll zu interpretieren sind. Andere Korrelationen sind testtheoretisch erlaubt, jedoch schwierig zu interpretieren.

27.2.4 Vergleich der drei längsschnittlichen MTMM-Modelle

In der folgenden Box sind die Modellgleichungen und Varianzdekompositionen der drei longitudinalen MTMM-Modellen übersichtlich zusammengestellt.

Modellgleichungen und Varianzdekompositionen der längsschnittlichen MTMM-Modelle (für Abweichungsvariablen, s. z. B. Bollen 1989)

Für alle Modelle gilt:

α_{ik1t}	Interzept des Indikators
λ_{Tikmt}	Ladungsparameter auf dem Traitfaktor
λ_{Oikmt}	Ladungsparameter auf dem messgelegenheitsspezifischen Faktor
λ_{Mik2t}	Ladungsparameter auf dem Methodenfaktor (λ_{TMik2t} im Multi-method-LST-Modell)
λ_{OMikmt}	Ladungsparameter auf dem messgelegenheitsspezifischen Methodenfaktor
i	Indikator;
k	Trait;
m	Methode;
t	Messgelegenheit

Multioccasion-MTMM-Modell

Standardmethode zu allen Messgelegenheiten:

$$Y_{ik1t} = \alpha_{ik1t} + \lambda_{Tik1t} S_{ik1t} + E_{ik1t}$$

$$\Rightarrow \text{Var}(Y_{ik1t}) = \lambda_{Tik1t}^2 \text{Var}(S_{ik1t}) + \text{Var}(E_{ik1t})$$

Nichtstandardmethode zu allen Messgelegenheiten:

$$Y_{ik2t} = \alpha_{ik2t} + \lambda_{Tik2t} S_{ik1t} + \lambda_{Mik2t} M_{+k2t} + E_{ik2t}$$

$$\Rightarrow \text{Var}(Y_{ik2t}) = \lambda_{Tik2t}^2 \text{Var}(S_{ik1t}) + \lambda_{Mik2t}^2 \text{Var}(M_{+k2t}) + \text{Var}(E_{ik2t})$$

Multiconstruct-LST-Modell

Für alle Methoden zu allen Messgelegenheiten:

$$Y_{ikmt} = \alpha_{ikmt} + \lambda_{Tikmt} T_{ikmt} + \lambda_{Oikmt} O_{ikmt} + E_{ikmt}$$

$$\Rightarrow \text{Var}(Y_{ikmt}) = \lambda_{Tikmt}^2 \text{Var}(T_{ikmt}) + \lambda_{Oikmt}^2 \text{Var}(O_{ikmt}) + \text{Var}(E_{ikmt})$$

4 Lediglich die messgelegenheitsspezifischen Variablen dürfen mit den messgelegenheitsspezifischen Abweichungsvariablen zu einem anderen Messzeitpunkt korrelieren, so z.B. $\text{Corr}(O_{+111}, OM_{+112})$. Diese Korrelation ist jedoch theoretisch nicht einfach zu interpretieren. In vielen Anwendungen bietet es sich an, diese Korrelationen nicht zuzulassen.

Multimethod-LST-Modell

Standardmethode zu allen Messgelegenheiten:

$$Y_{ik1t} = \alpha_{ik1t} + \lambda_{T_{ik1t}} T_{ik1t} + \lambda_{O_{ik1t}} O_{ik1t} + E_{ik1t}$$

$$\Rightarrow \text{Var}(Y_{ik1t}) = \lambda_{T_{ik1t}}^2 \text{Var}(T_{ik1t}) + \lambda_{O_{ik1t}}^2 \text{Var}(O_{ik1t}) + \text{Var}(E_{ik1t})$$

Nichtstandardmethode zu allen Messgelegenheiten:

$$Y_{ik2t} = \alpha_{ik2t} + \lambda_{T_{ik2t}} T_{ik1t} + \lambda_{O_{ik2t}} O_{ik1t} + \lambda_{TM_{ik2t}} TM_{+k2t}$$

$$+ \lambda_{OM_{ik2t}} OM_{+k2t} + E_{ik2t}$$

$$\Rightarrow \text{Var}(Y_{ik2t}) = \lambda_{T_{ik2t}}^2 \text{Var}(T_{ik1t}) + \lambda_{O_{ik2t}}^2 \text{Var}(O_{+k1t})$$

$$+ \lambda_{TM_{ik2t}}^2 \text{Var}(TM_{+k2t}) + \lambda_{OM_{ik2t}}^2 \text{Var}(OM_{+k2t})$$

$$+ \text{Var}(E_{ik2t})$$

Vergleich der Modellgleichungen und der spezifischen Varianzkomponenten

Die drei hier vorgestellten längsschnittlichen MTMM-Modelle unterscheiden sich im Wesentlichen in Bezug auf ihre Komplexität. Das *Multioccasion-MTMM-Modell* und das *Multiconstruct-LST-Modell* ermöglichen die Varianzzerlegung in drei Komponenten: den Anteil des Traits, den Anteil einer Methode oder der Messgelegenheit sowie den Anteil des Fehlerterms. Das *Multimethod-LST-Modell* zerlegt die Varianz in fünf Bestandteile, und zwar den Anteil des Traits, der Nichtstandardmethode, des messgelegenheitsspezifischen Einflusses auf die Standardmethode, des messgelegenheitsspezifischen Einflusses auf die Nichtstandardmethode und des Fehlerterms.

Der State im *Multioccasion-MTMM-Modell* enthält sowohl stabile Bestandteile als auch messgelegenheitsspezifische Bestandteile. Die Methodenvariable enthält die momentane Abweichung der Nichtstandardmethode, die sich theoretisch aus stabilen und situativen Komponenten zusammensetzen könnte.

Die Trait-Variable im *Multiconstruct-LST-Modell* umfasst die zeitlich stabilen Anteile des Ratings mit der jeweiligen Methode. Sie beinhaltet damit sowohl Anteile, die auf die wahre Ausprägung zurückzuführen sind, als auch Anteile des Methodenbiases. Die messgelegenheitsspezifischen Variablen setzen sich ebenfalls aus Einflüssen der Messgelegenheit und Methodeneinflüssen zusammen und spiegeln die momentanen Abweichungen vom stabilen Anteil wider.

Lediglich das komplexeste Modell, das *Multimethod-LST-Modell*, ermöglicht es, die stabilen Einflüsse von Trait und Nichtstandardmethode von den variablen Einflüssen der Messgelegenheiten und messgelegenheitsspezifischen Methodeneffekte zu trennen. In diesem Modell ist die feinste Zerlegung der Varianz einer beobachteten Variablen möglich.

In **Tab. 27.1** sind die Verfahren zur Bestimmung der Stabilität, der Messgelegenheitsspezifität, der konvergenten Validität sowie der Methodenspezifität in den drei longitudinalen MTMM-Modellen übersichtlich zusammengestellt.

Beim Einsatz des *Multioccasion-MTMM-Modells* entspricht die Anzahl der Traitfaktoren dem Produkt von Traits und Messgelegenheiten pro Methode. So entstehen schon bei nur drei Messgelegenheiten eine Vielzahl von Korrelationen, die die zeitliche Stabilität des Konstrukts, aber auch die diskriminante Validität zwischen den Traits kennzeichnen. Ein Maß für die mittlere Konvergenz der Methoden oder die mittlere diskriminante Validität ist nicht einfach zu berechnen (s. Bortz und Schuster 2010, zur allgemeinen Verrechnung von Korrelationen). Aus diesem Grund werden im folgenden Beispiel nur das *Multiconstruct-LST-Modell* und das *Multimethod-LST-Modell* eingesetzt.

Vorteile des Multimethod-LST-Modells

Vergleich der Verfahren zur Bestimmung der Stabilität, der Messgelegenheitsspezifität, der konvergenten Validität sowie der Methodenspezifität

Table 27.1 Bestimmung der Stabilität, der Messgelegenheitsspezifität, der konvergenten und diskriminanten Validität sowie der Methodenspezifität in den drei longitudinalen MTMM-Modellen

	Multioccasion-MTMM-Modell	Multiconstruct-LST-Modell	Multimethod-LST-Modell
Stabilität	Korrelation der State-Variablen über die Zeit: $Corr(S_{ik1t}, S_{ik1t'})$, $t \neq t'$	$\frac{\lambda_{Tik1L}^2 Var(T_{ik1L})}{Var(Y_{ik1L})}$	Für die Standardmethode: $\frac{\lambda_{Tik1L}^2 Var(T_{ik1L})}{Var(Y_{ik1L})}$ Für die Nichtstandardmethode: $\frac{\lambda_{Tikmt}^2 Var(T_{ik1L}) + \lambda_{Tikmt}^2 Var(TM_{+kmt})}{Var(Y_{ikmt})}$
Messgelegenheitsspezifischer Einfluss	–	$\frac{\lambda_{Oikmt}^2 Var(O_{+kmt})}{Var(Y_{ikmt})}$	Für die Standardmethode: $\frac{\lambda_{Oik1L}^2 Var(O_{+k1L})}{Var(Y_{ik1L})}$ Für die Nichtstandardmethode: $\frac{\lambda_{Oikmt}^2 Var(O_{+k1L}) + \lambda_{Oikmt}^2 Var(OM_{+kmt})}{Var(Y_{ikmt})}$
Konvergente Validität	Für Nichtstandardmethoden: $\frac{\lambda_{Tikmt}^2 Var(S_{ik1L})}{Var(Y_{ikmt})}$	–	Für Nichtstandardmethoden: $\frac{\lambda_{Tikmt}^2 Var(T_{ik1L}) + \lambda_{Oikmt}^2 Var(O_{+k1L})}{Var(Y_{ikmt})}$
Diskriminante Validität	Korrelation der State-Variablen zweier Konstrukte zu einer Messgelegenheit: $Corr(S_{ik1t}, S_{ik'1t})$, $k \neq k'$	Korrelationskoeffizienten der Trait-Faktoren: $Corr(T_{ikmt}, T_{ik'mt})$, $k \neq k'$	Korrelationskoeffizienten der Trait-Faktoren: $Corr(T_{ik1L}, T_{ik'1L})$, $k \neq k'$
Methodenspezifität	Für die Nichtstandardmethode: $\frac{\lambda_{Mikmt}^2 Var(M_{+kmt})}{Var(Y_{ikmt})}$	–	Für die Nichtstandardmethode: $\frac{\lambda_{TMikmt}^2 Var(TM_{+kmt}) + \lambda_{OMikmt}^2 Var(OM_{+kmt})}{Var(Y_{ikmt})}$ Davon stabil: $\frac{\lambda_{TMikmt}^2 Var(TM_{+kmt})}{Var(Y_{ikmt})}$ Davon messelegenheitsspezifisch: $\frac{\lambda_{OMikmt}^2 Var(OM_{+kmt})}{Var(Y_{ikmt})}$
Generalisierbarkeit der Methodeneffekte	Korrelation der Methodeneffekte: $Corr(M_{+kmt}, M_{+k'mt})$, $k \neq k'$	–	Stabil: $Corr(TM_{+kmt}, TM_{+k'mt})$, $k \neq k'$ Abhängig von der Situation: $Corr(TM_{+kmt} + OM_{+kmt}, TM_{+k'mt} + OM_{+k'mt})$, $k \neq k'$
Reliabilität	$Rel(Y_{ikmt}) = \frac{\lambda_{Tikmt}^2 Var(S_{ik1L}) + \lambda_{Mikmt}^2 Var(M_{+kmt})}{Var(Y_{ikmt})}$	$Rel(Y_{ikmt}) = \frac{\lambda_{Tikmt}^2 Var(T_{ikmt}) + \lambda_{Oikmt}^2 Var(O_{ikmt})}{Var(Y_{ikmt})}$	Für die Standardmethode: $Rel(Y_{ik1L}) = \frac{\lambda_{Tik1L}^2 Var(T_{ik1L}) + \lambda_{Oik1L}^2 Var(O_{+k1L})}{Var(Y_{ik1L})}$ Für die Nichtstandardmethode: $Rel(Y_{ik2L}) = \frac{\lambda_{Tik2L}^2 Var(T_{ik1L}) + \lambda_{Oik2L}^2 Var(O_{+k1L}) + \lambda_{TMik2L}^2 Var(TM_{+k2L}) + \lambda_{OMik2L}^2 Var(OM_{+k2L})}{Var(Y_{ik2L})}$

Anmerkung: Leere Zellen zeigen an, dass es keinen direkten Koeffizienten gibt. Oft lassen sich allerdings Korrelationen oder Vergleiche von Korrelationen im Sinne dieser Koeffizienten interpretieren.

27.3 Multiconstruct-LST- und Multimethod-LST-Modell in der empirischen Anwendung

Cole und Kollegen (Cole und Martin 2005; Cole et al. 1996; Cole et al. 1997) befragten u. a. 375 Schüler und deren Lehrer einer amerikanischen Grundschule („elementary school“) zu vier Messgelegenheiten im Hinblick auf ihre Depressivität und Ängstlichkeit. Die Depressivität der Schüler wurde mit dem Child Depression Inventory (Kovacs 1981, 1982) und ihre Ängstlichkeit mit der Revised Children’s Manifest Anxiety Scale (Reynolds und Richmond 1978) erfasst. Die Ratings der Lehrer wurden mit dem Teacher Report Index of Depression (Cole und Jordan 1995) und dem Teacher Report Index of Anxiety (Cole und Jordan 1995; Lefkowitz und Tesiny 1980) erfasst.

Das Multiconstruct-LST-Modell (Dumenci und Windle 1998; Eid et al. 1994; Majcen et al. 1998; Schermelleh-Engel et al. 2004; Schmitt 2000; Steyer et al. 1989, 1990) und das Multimethod-LST-Modell (Courvoisier 2006; Courvoisier et al. 2008) wurden in einer Reanalyse mit dem robusten Maximum-Likelihood-Schätzer (MLR, *Maximum Likelihood Robust Estimator*) und der *Complex-Option* von *Mplus* (Muthén und Muthén 2017) geschätzt⁵.

27.3.1 Ergebnisse mit dem Multiconstruct-LST-Modell

Das Multiconstruct-LST-Modell zeigte in der ersten Schätzung einen nur mäßigen Datenfit ($\chi^2 = 842.2$, $df = 396$, $p = .00$, Comparative Fit Index [CFI] = .964, Root Mean Square Error of Approximation [RMSEA] = .055). Ein möglicher Grund für die Fehlanpassung kann darin liegen, dass die Lehrer von der zweiten zur dritten Messgelegenheit mit dem Schuljahr wechselten. Dadurch ist die Annahme eines stabilen Traits für die Lehrer zu stark. Teilt man die Submodelle für die Schüler und Lehrer in je zwei neue LST-Modelle (entsprechend dem Schuljahr) und setzt man zusätzlich alle Trait-Ladungen gleich⁶, verbessert sich die Modellanpassung erheblich und weist einen guten Modellfit auf ($\chi^2 = 362.4$, $df = 312$, $p = .03$, CFI = .996, RMSEA = .021; Teilmodell für die Ängstlichkeit in ■ Abb. 27.5).

In ■ Tab. 27.2 präsentieren wir nur einen Auszug der Ergebnisse der Stabilität, Variabilität (Messgelegenheitsspezifität) und Reliabilität für die Ängstlichkeit der Schüler, da an diesem Auszug bereits die wesentlichen Merkmale des Multiconstruct-LST-Modells für die Varianzzusammensetzung deutlich gemacht werden können. Die vollständigen Ergebnisse (einschließlich der Korrelationstabellen) sind in detaillierter Form bei Courvoisier et al. (2008) beschrieben.

Zunächst können wir feststellen, dass alle Indikatoren hoch reliabel gemessen wurden. Sowohl die Schüler als auch die Lehrer zeigten eine recht stabile Einschätzung der Ängstlichkeit, da ein großer Teil der aufgeklärten Varianz auf die Trait-Variablen zurückgeführt werden kann (52 bis 87 %). Auffallend ist weiterhin, dass mit zunehmender Dauer der Untersuchung die Stabilität der Ratings zunimmt. Bei den Schülern sinkt die Variabilität von der ersten bis zur dritten Messung und steigt dann wieder an. Bei den Lehrern ist die jeweils zweite Messgelegenheit eines Schuljahres durch eine geringere Variabilität gekennzeichnet (Messgelegenheiten 2 und 4), wenngleich die Lehrer im zweiten Schuljahr der Untersuchung etwas variablere Einschätzungen zeigen.

Stabilität, Variabilität und Reliabilität

5 Diese Spezifikation berücksichtigt die Homogenität der Schüler in den einzelnen Schulklassen (Schachtelung von Ratings, Multilevelstruktur) und korrigiert die Standardfehler der Parameterschätzungen und den χ^2 -Wert entsprechend.

6 Durch diese Restriktion wird sichergestellt, dass in beiden Modellen die gleiche Beziehung zwischen Indikatoren und Traits besteht, also das identische Konstrukt gemessen wird.

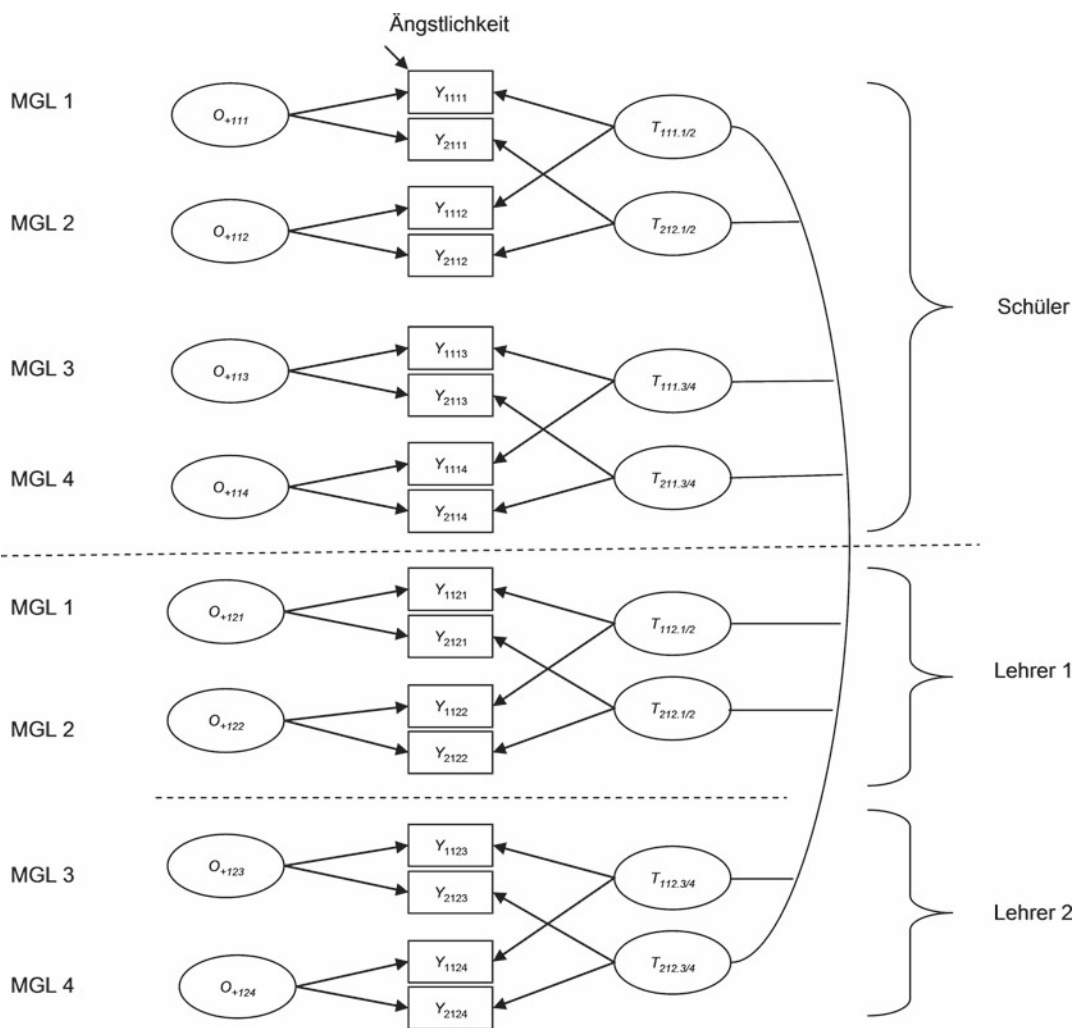


Abb. 27.5 Darstellung einer Trait-Einheit des Multiconstruct-LST-Modells für vier Messgelegenheiten. T_{ikm+} : Trait-Variablen; O_{+kmt} : Messgelegenheitsvariable; T_{ikmt} : $t = 1/2$ kennzeichnet die Trait-Variablen des ersten Schuljahres (MGL 1 und 2), $3/4$ die Trait-Variablen des zweiten Schuljahres (MGL 3 und 4); MGL : Messgelegenheit. Die Korrelationen zwischen den Messgelegenheitsfaktoren sind aus Gründen der Lesbarkeit nicht abgebildet. Die Fehlervariable ist nur für den ersten Indikator angegeben

Wie sehr hängen nun die Schüler- und Lehrerratings zusammen? Eine Analyse der Korrelationen der indikatorspezifischen Trait-Variablen zeigt, dass diese nur zu .19 bis .29 zwischen den Schülern und den Lehrern korrelieren. Lehrer und Schüler stimmen offensichtlich in ihren Ratings nicht sehr hoch überein. Darüber hinaus stimmen die Lehrer der unterschiedlichen Schuljahre in ihren Einschätzungen auch nur zu einem sehr geringen Teil überein (alle Korrelationen sind kleiner als .17). Die testhälftenspezifischen Traits der Schüler und der beiden Lehrer sind jedoch sehr hoch miteinander korreliert, was auf die Homogenität der Testhälften hinweist (alle Korrelationen sind höher als .86 für die Schüler und .94 für die Lehrerratings innerhalb eines Schuljahres).

Die in der Ergebnisliste nicht wiedergegebenen Korrelationen zwischen den Traits „Ängstlichkeit“ und „Depressivität“ der Schüler sind alle größer als .68, ein deutliches Zeichen für die Verwandtschaft und mangelnde diskriminante Validität beider Konstrukte. Bei den Lehrern wird die mangelnde Trennbarkeit der beiden Konstrukte noch stärker sichtbar. Die Korrelationen der Trait-Variablen der beiden Konstrukte in einem Schuljahr sind alle größer als .83.

Konvergente und diskriminante Validität

■ Tabelle 27.2 Stabilitäten, Messgelegenheitsspezifitäten und Reliabilitäten der Indikatoren im Multiconstruct-LST-Modell für die Ängstlichkeit

Testhälfte	Stabilität	Messgelegenheitsspezifität	Reliabilität
<i>Ängstlichkeit im Schülerrating</i>			
Y_{1111}	.59	.31	.90
Y_{2111}	.56	.32	.88
Y_{1112}	.86	.07	.93
Y_{2112}	.82	.08	.90
Y_{1113}	.83	.07	.90
Y_{2113}	.87	.06	.93
Y_{1114}	.69	.21	.90
Y_{2114}	.73	.19	.92
<i>Ängstlichkeit im Lehrerrating</i>			
Y_{1121}	.60	.31	.91
Y_{2121}	.58	.32	.90
Y_{1122}	.87	.04	.91
Y_{2122}	.85	.04	.89
Y_{1123}	.55	.37	.92
Y_{2123}	.52	.38	.90
Y_{1124}	.62	.29	.91
Y_{2124}	.61	.31	.92

Wie gut können die Lehrer Schwankungen der Ängstlichkeit und der Depressivität der Schüler nachvollziehen? Wirken sich die Situationen in gleichem Ausmaß auf die Einschätzungen der beiden Konstrukte für die Schüler und die Lehrer aus? Zur Beantwortung dieser Fragen können die Korrelationen der messgelegenheitsspezifischen Variablen herangezogen werden. Die Schwankungen der Einschätzungen der Schüler und der Lehrer sind sehr unterschiedlich. Schätzen die Schüler ihre momentane Ängstlichkeit höher ein als gewöhnlich, so schätzen die Lehrer die Schüler tendenziell sogar weniger ängstlich ein als gewöhnlich (zwei der vier möglichen Korrelationen der Messgelegenheitsfaktoren sind negativ $[-.09$ und $-.25]$, eine ist 0, lediglich eine Korrelation ist positiv mit $.19$). Ein ähnliches Muster zeigt sich für die messgelegenheitsspezifischen Variablen der Depressivität.

Die messgelegenheitsspezifischen Variablen der Schüler für Ängstlichkeit und Depressivität sind zwischen $.41$ und $.74$ korreliert, was für einen gleichartigen Einfluss der Messgelegenheit auf beide Konstrukte spricht. Für die Lehrer zeigen sich sehr hohe Korrelationen der messgelegenheitsspezifischen Variablen für die beiden Konstrukte zur jeweils ersten Messgelegenheit ($.80$ und $.84$). Die situativen Einflüsse generalisieren somit sehr stark über die beiden Konstrukte hinweg. Zur zweiten Messgelegenheit fallen diese Korrelationen auf $.40$ und $.58$. Dies spricht dafür, dass die Lehrer (zumindest in ihrer Wahrnehmung) besser zwischen den beiden Konstrukten differenzieren, wenn sie die Schüler am Ende des Schuljahres besser kennen.

27.3.2 Ergebnisse mit dem Multimethod-LST-Modell

Das Multimethod-LST-Modell mit den Selbsteinschätzungen der Schüler als Standardmethode (Teilmodell für die Ängstlichkeit in **Abb. 27.6**) passt ebenfalls gut auf die Daten ($\chi^2 = 338.4$, $df = 288$, $p = .02$, CFI = .996, RMSEA = .022).

Die Reliabilitäten und Varianzkomponenten für Ängstlichkeit sind in **Tab. 27.3** wiedergegeben. Im Gegensatz zum vorherigen Modell können hier die gemeinsamen Varianzkomponenten direkt bestimmt werden. Wieder sind die Ratings der Schüler relativ stabil über die Zeit. Auch die Ratings der Lehrer zeigen sich fast ebenso stabil. Analysiert man die Stabilität etwas genauer, so zeigt sich, dass sie fast ausschließlich auf die Stabilität des Methodeneffekts zurückgeführt werden kann und nicht an einer erhofften Übereinstimmung mit den Schülern liegt (die maximale Stabilität, die auf den Trait der Schüler zurückzuführen ist, liegt bei .06). Die Ratings der Lehrer sind in keiner Weise von den situativen Schwankungen der Schüler beeinflusst, sondern die Schwankungen der Lehrer hängen ausschließlich von den internen Zuständen der Lehrer ab (alle auf die situativen Variablen der Schüler zurückzuführenden Varianzkomponenten sind kleiner als .01).

Ergebnisse für Ängstlichkeit

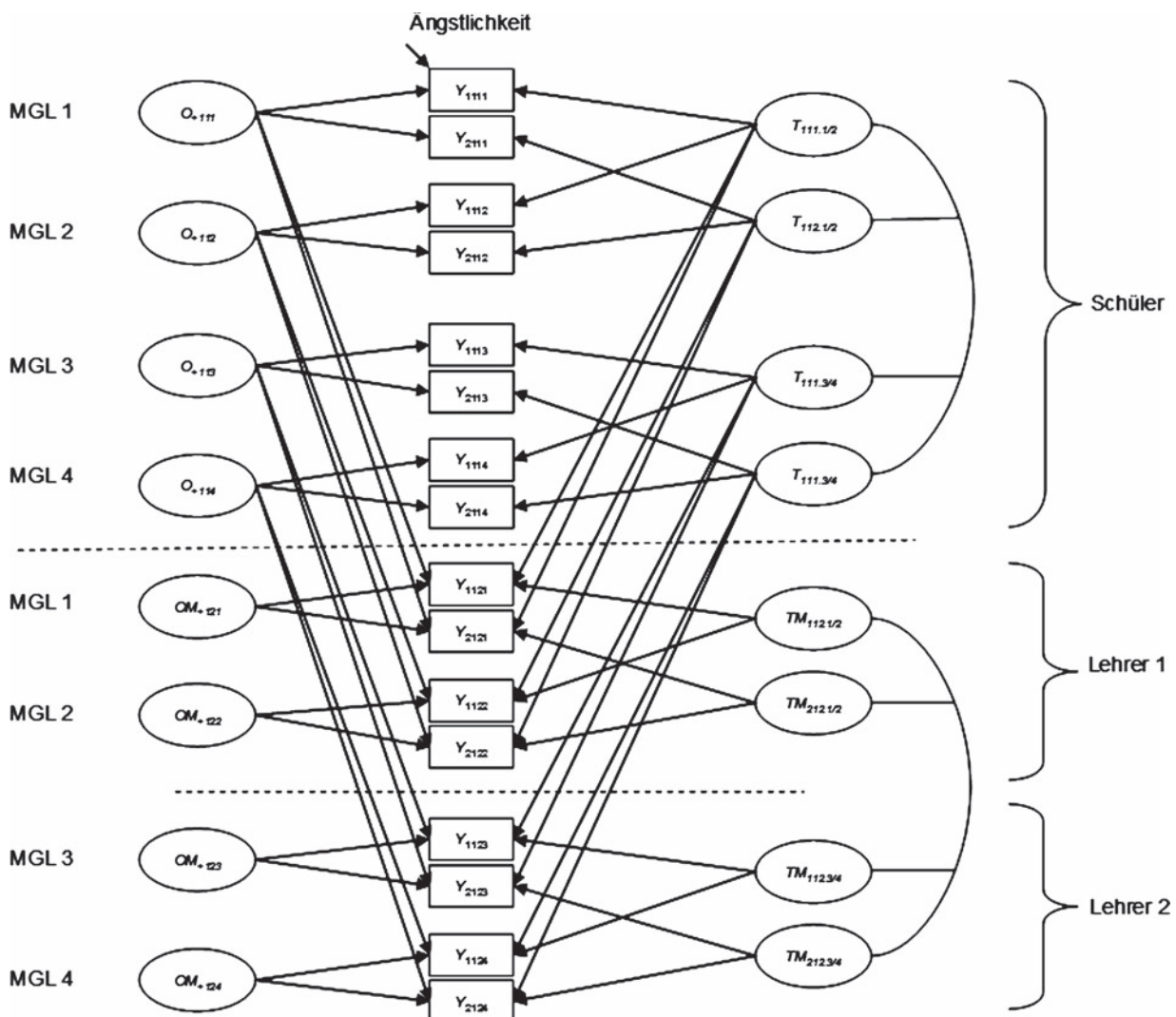


Abb. 27.6 Multimethod-LST-Submodell zur Ängstlichkeit für das erste Schuljahr. $T_{ikm.1/2}$: Trait-Variable; $TM_{ikm.1/2}$: MethodenvARIABLE; O_{+kmt} : messgelegenheitsspezifische Variable; OM_{+kmt} : messgelegenheitsspezifische Abweichungsvariable; MGL: Messgelegenheit. Korrelationen sind aus Gründen der Lesbarkeit nicht abgebildet. Die Fehlervariable ist nur für den ersten Indikator angegeben

■ Tabelle 27.3 Stabilitäten, Variabilitäten und Reliabilitäten der Indikatoren im Multimethod-LST-Modell für die Ängstlichkeit

Testhälfte	Stabilität	Stabilität des Traits	Stabilität der Methode	Messgelegenheitsspezifität	Mit den Schülern geteilte Messgelegenheitsspezifität	Messgelegenheitsspezifität der Lehrer	Reliabilität
<i>Ratings der Schüler</i>							
Y ₁₁₁₁	.59			.28			.87
Y ₂₁₁₁	.55			.35			.91
Y ₁₁₁₂	.86			.06			.92
Y ₂₁₁₂	.80			.12			.92
Y ₁₁₁₃	.84			.03			.87
Y ₂₁₁₃	.86			.14			1.00
Y ₁₁₁₄	.68			.24			.92
Y ₂₁₁₄	.74			.17			.91
<i>Ratings der Lehrer</i>							
Y ₁₁₂₁	.76	.06	.70	.16	.00	.16	.92
Y ₂₁₂₁	.74	.03	.71	.17	.00	.17	.91
Y ₁₁₂₂	.69	.05	.64	.21	.01	.20	.89
Y ₂₁₂₂	.67	.04	.63	.21	.00	.21	.88
Y ₁₁₂₃	.62	.01	.61	.30	.00	.30	.92
Y ₂₁₂₃	.60	.02	.58	.31	.00	.31	.91
Y ₁₁₂₄	.58	.04	.54	.33	.00	.33	.91
Y ₂₁₂₄	.56	.05	.52	.35	.00	.35	.91

Ergebnisse für Depressivität

Die Ergebnisse der Depressivitäts- weichen kaum von den Ergebnissen der Ängstlichkeitsratings ab. Das Multimethod-LST-Modell verdeutlicht somit unverkennbar die starke Divergenz der Schüler- und Lehrerratings. Die Korrelationen zwischen den Trait-Faktoren und zwischen den Methodenvariablen zeigen ein nahezu identisches Bild wie im Multiconstruct-LST-Modell. Es findet sich eine hohe Korrelation der Trait-Variablen der Schüler, die auf mangelnde diskriminante Validität hindeutet, und eine hohe Korrelation der Methodenvariablen der Lehrer, die auf einen starken Generalisierungseffekt der Methode hindeutet. Das heißt: Überschätzt ein Lehrer die Ängstlichkeit der Schüler, so überschätzt er tendenziell auch deren Depressivität. Zur besseren Anpassung des Modells wurden für jede Testhälfte separat Methodenvariablen geschätzt, wobei sich die Einflüsse der Methode als recht homogen auf beide Testhälften erwiesen.

Die messgelegenheitsspezifischen Effekte der Schüler wirken sich auf beide Konstrukte auf einem ähnlich hohen Niveau wie im vorhergehenden Modell aus. Die messgelegenheitsspezifischen Abweichungsvariablen der Lehrer generalisieren ebenfalls über die Konstrukte hinweg. Die messgelegenheitsspezifischen Variablen der Schüler können nicht herangezogen werden, um die Abweichungen der Lehrer bei der Einschätzung des anderen Konstrukts zu erklären; d. h., Schüler, die sich selbst ängstlicher als sonst erleben, werden von den Lehrern nicht depressiver als sonst eingeschätzt (die Korrelationen zwischen den messgelegenheitsspezifischen Variablen der Schüler für die Ängstlichkeit mit den messgelegenheitsspezifischen Abweichungsvariablen der Lehrer für die Depressivität und vice versa liegen alle nahe 0).

27.3.3 Fazit der Anwendungen der beiden multimethodalen LST-Modelle

Die Anwendungen der beiden multimethodalen LST-Modelle illustrieren den Nutzen dieser Modelle und ermöglichen interessante Einblicke in den Zusammenhang von Schüler- und Lehrerratings: Lehrerratings können *nicht* dazu herangezogen werden, die von den Schülern berichtete Ängstlichkeit und Depressivität angemessen widerzuspiegeln. Vielmehr zeigen Lehrer sowohl im Querschnitt als auch im Verlauf eines Schuljahres einen stabilen Bias in Bezug auf ihre Einschätzungen. Sie sind nicht nur nicht in der Lage, die stabile Ängstlichkeit oder Depressivität der Schüler einzuschätzen, sie können auch die jeweiligen situativen Schwankungen der Ängstlichkeit oder der Depressivität der Schüler nicht nachvollziehen. Gesteht man den Schülern der vierten und fünften Klasse in US-amerikanischen Schulen die Fähigkeit zu, ihre eigene Ängstlichkeit und Depressivität angemessen einschätzen zu können, können sich Psychologen diesen Ergebnissen zufolge keinesfalls auf die Einschätzungen der Lehrer verlassen. Die landläufige Annahme, dass Lehrer nach einem Schuljahr in der Lage sein sollten, die Gefühle der Schüler ihrer Klasse einschätzen zu können, wird zumindest mit dieser Studie widerlegt.

**Querschnittlicher und
längsschnittlicher Bias im
Lehrerrating**

27.4 Praktische Hinweise zur Analyse longitudinaler multimodaler Modelle

Im folgenden sollen praktische Hinweise zur Analyse der komplexen Modelle gegeben werden. Da in diesen Modellen eine Vielzahl von Parametern geschätzt werden, kann erst ab relativ hohen Stichprobengrößen mit verlässlichen Ergebnissen gerechnet werden. Die Bestimmung der benötigten minimalen Stichprobengrößen ist Gegenstand gegenwärtiger Forschung.

Es empfiehlt sich, longitudinale multimethodale Modelle schrittweise in den empirischen Anwendungen aufzubauen. Wir wollen den Aufbau dieser Modelle in drei Schritten skizzieren und auf mögliche Probleme eingehen:

■ ■ Schritt 1: Ein Konstrukt und eine Methode

Zunächst empfiehlt es sich, für jedes Konstrukt und jede Methode getrennte Analysen vorzunehmen. Für jede dieser Trait-Methoden-Einheiten kann ein Modell geschätzt werden, in dem für jede Messgelegenheit nur eine State-Variable eingeführt wird. Die Korrelationen dieser (Single-Method-)State-Variablen sollten signifikant sein, damit im nächsten Schritt überprüft werden kann, ob ein LST-Modell diese Daten angemessen repräsentiert (s. auch Geiser et al. 2010). Korrelieren die State-Variablen nicht miteinander, d. h., es handelt sich um ein vollständig zeitlich instabiles Merkmal, so kann auch kein LST-Modell angepasst werden. Eine Möglichkeit, konvergente und diskriminante Validität im zeitlichen Verlauf festzustellen, bietet dann das Multioccasion-MTMM-Modell. Es sei ausdrücklich darauf hingewiesen, dass Methoden gut übereinstimmen können, auch wenn keine zeitliche Stabilität vorliegt. Psychologische Konstrukte haben nicht immer einen stabilen zeitlichen Charakter (Stimmungen, Hormonlevel, Ärgererleben etc.).

**Korrelationen der State-Variablen
sollten signifikant sein**

■ ■ Schritt 2: LST-Modell

Im zweiten Schritt kann für jede Trait-Methoden-Einheit ein (Mono-Method-)LST-Modell geschätzt werden. Mit diesem Modell kann die zeitliche Stabilität eines Konstrukts anhand einer einzigen Methode überprüft werden. Mögliche Gründe für eine schlechte Anpassung des Modells an die Daten können hier neben mangelnder Stabilität noch weitere Faktoren sein:

Gründe für mangelnden Modellfit

- Der sog. „Sokrates-Effekt“: Testpersonen kennen nicht alle psychologischen Konstrukte und verfügen folglich nicht über eine kognitive Repräsentation der

Sokrates-Effekt

- Trait-Ausprägung. Durch die Befragung können Testpersonen eine Vorstellung darüber entwickeln, welches Konstrukt gemessen werden sollte, sodass sie in weiteren Befragungen eventuell konformer im Hinblick auf das von ihnen vermutete Konstrukt antworten. Bei Modellen, in denen die Annahme der Messinvarianz getroffen wurde, kann dies zu Problemen führen. Liegt ein Sokrates-Effekt vor, können die Ladungen der Indikatoren des ersten Messzeitpunkts von der Annahme der Messinvarianz ausgenommen werden (d. h. frei geschätzt werden).
- Mangelnde messgelegenheitsspezifische Varianz**
- Eine oder mehrere messgelegenheitsspezifische Variablen haben keine Varianz: Dieses Phänomen tritt bei sehr stabilen Merkmalen (wie der Intelligenz) auf, die keinen oder nur sehr geringen situativen Schwankungen unterliegen; es kann aber auch bei weniger stabilen Konstrukten auftreten. Durch Eliminierung der betreffenden Messgelegenheitsvariablen kann das LST-Modell an die Daten angepasst werden. Im extremen Fall, wenn alle Messgelegenheitsvariablen eliminiert werden, bleibt ein reines Latent-Trait-Modell bestehen.
 - Die verschiedenen Indikatoren eines Traits sind nicht perfekt homogen: Dieses Problem wurde bereits in ► Abschn. 27.2 behandelt.
- Konstrukt unterliegt Veränderung**
- Das Konstrukt unterliegt einem Veränderungsprozess und ist nicht stabil über die Zeit: In diesem Fall müssen andere longitudinale Modelle wie etwa das Latent-Curve-Model oder ein Modell mit autoregressiven Strukturen (s. dazu Bollen und Curran 2006) herangezogen werden.

■ ■ Schritt 3: Longitudinales multimethodales Modell

Je nach inhaltlicher Fragestellung kann in einem dritten Schritt eines der vorgestellten MTMM-LST-Modelle analysiert werden.

27.5 Zusammenfassung

Ziel dieses Kapitels war es, zu verdeutlichen, dass Merkmalsausprägungen von Individuen über die Zeit schwanken können, und dass somit auch die konvergente und diskriminante Validität verschiedener Methoden und Konstrukte zeitlichen Veränderungen unterworfen sind. Die Analyse konvergenter und diskriminanter Validität ist Basis jeder diagnostischen Entscheidung. Nur bei gesicherter Qualität der eingesetzten Verfahren können zuverlässig Indikationen für mögliche Interventionen getroffen werden. Besonders bei Kindern, die sich in einem Entwicklungsprozess befinden, aber auch bei Erwachsenen ist es notwendig, die zeitliche Stabilität der gefundenen Testscores zu untersuchen. Nur bei gegebener Stabilität der Messungen kann von einem stabilen Trait ausgegangen werden. Darüber hinaus ist es wichtig, zu analysieren, wie sich die konvergente Validität verschiedener Messmethoden über die Zeit entwickelt.

Drei longitudinale multimethodale Modelle für mehrere Traits wurden vorgestellt, die es erlauben, die Konvergenz verschiedener Methoden und die diskriminante Validität von Traits und States zu untersuchen. Die empirischen Anwendungen zeigen deutlich, dass implizite Annahmen zur Übereinstimmung verschiedener Methoden prinzipiell überprüft werden müssen.

27.6 EDV-Hinweise

Die hier vorgestellten longitudinalen multimethodalen Modelle können mit gängigen EDV-Programmen zur Analyse von Strukturgleichungsmodellen vorgenommen werden (s. hierzu ► Abschn. 25.9).

27.7 Kontrollfragen

? Die Antworten auf die folgenden Fragen finden Sie im Lerncenter zu diesem Kapitel unter ► <https://www.lehrbuch-psychologie.springer.com> (Projekt Testtheorie und Fragebogenkonstruktion).

1. Was wird unter austauschbaren, strukturell unterschiedlichen und gleichwertigen Methoden im Zusammenhang mit MTMM-Modellen verstanden? Was sind die wesentlichen Unterschiede zwischen diesen Methoden?
2. Worin liegt der Vorteil trait-spezifischer Methodeneffekte in MTMM-Modellen?
3. Worin besteht der Unterschied zwischen LST- und MTMM-Modellen?
4. Bei welchen wissenschaftlichen Fragestellungen sollte das Multioccasion-MTMM-, das Multiconstruct-LST- oder das Multimethod-LST-Modell eingesetzt werden?

Literatur

- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.
- Bollen, K. A. & Curran, P. J. (2006). *Latent curve models: A structural equations perspective*. Hoboken, NJ: Wiley.
- Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler*. Berlin, Heidelberg: Springer.
- Burns, G. L. & Haynes, S. N. (2006). Clinical Psychology: Construct validation with multiple sources of information and multiple settings. In M. Eid and E. Diener (Eds.), *Handbook of Multimethod Measurement in Psychology* (pp. 401–418). Washington, DC: American Psychological Association.
- Cole, D. A. & Jordan, A. E. (1995). Competence and Memory: Integrating psychosocial and cognitive correlates of child depression. *Child Development*, *66*, 459–473.
- Cole, D. A. & Martin N. C. (2005). The longitudinal structure of the children's depression inventory: Testing a latent trait-state model. *Psychological Assessment*, *17*, 144–155.
- Cole, D. A., Martin, J. M., Powers, B. & Truglio, R. (1996). Modeling causal relations between academic and social competence and depression: A multitrait-multimethod longitudinal study of children. *Journal of Abnormal Psychology*, *105*, 258–270.
- Cole, D. A., Truglio, R. & Peeke, L. (1997). Relation between symptoms of anxiety and depression in children: A multitrait-multimethod-multigroup assessment. *Journal of Consulting and Clinical Psychology*, *65*, 110–119.
- Costa, P. T. & McCrae, R. R. (1998). Trait theories of personality. In D. F. Barone, M. Hersen & V. B. Van Hasselt (Eds.), *Advances Personality* (pp.103–121). New York: Plenum Press.
- Courvoisier, D. S. (2006). *Unfolding the constituents of psychological scores: Development and application of mixture and multitrait-multimethod LST models*. Unpublished doctoral dissertation, University of Geneva, Switzerland.
- Courvoisier, D. S., Eid, M. & Nussbeck, F. W. (2007). Mixture distribution latent state-trait analysis: Basic ideas and applications. *Psychological Methods*, *12*, 80–104.
- Courvoisier, D. S., Nussbeck, F. W., Eid, M., Geiser, C. & Cole, D. A. (2008). Analyzing the convergent and discriminant validity of states and traits: Development and applications of multimethod latent state-trait models. *Psychological Assessment*, *20*, 270–280.
- Dumenci, L. & Windle, M. (1998). A multitrait-multioccasion generalization of the latent trait-state model: Description and application. *Structural Equation Modeling*, *5*, 391–410.
- Eid, M. (1996). Longitudinal confirmatory factor analysis for polytomous item responses: Model definition on the basis of stochastic measurement theory. *Methods of Psychological Research*, *1*, 65–85.
- Eid, M. (2006). Methodological approaches for analyzing multimethod data. In M. Eid & E. Diener (Eds.), *Handbook of psychological measurement: A multimethod perspective* (pp. 223–230). Washington, DC: American Psychological Association.
- Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika*, *65*, 241–261.
- Eid, M., Lischetzke, T. & Nussbeck, F. W. (2006). Structural equation models for multitrait-multimethod data. In M. Eid and E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 283–299). Washington, DC: American Psychological Association.
- Eid, M., Lischetzke, T., Nussbeck, F. W. & Trierweiler, L. I. (2003). Separating trait effects from method-specific effects in multitrait-multimethod models: A multiple indicator CTC(M-1) model. *Psychological Methods*, *8*, 38–60.
- Eid, M. Notz, P., Steyer, R. & Schwenkmezger, P. (1994). Validating scales for the assessment of mood level and variability by latent state-trait analysis. *Personality and Individual Differences*, *16*, 63–76.

- Eid, M., Nussbeck, F. W., Geiser, C., Cole, D. A., Gollwitzer, M. & Lischetzke, T. (2008). Structural equation modeling of multitrait-multimethod data: Different models for different types of methods. *Psychological Methods, 13*, 230–253.
- Geiser, C., Eid, M., Nussbeck, F. W., Courvoisier, D. S. & Cole, D. A. (2008). Analyzing the convergent and discriminant validity of change: Structural equation modeling of multitrait-multimethod-multioccasion data. *Psychological Assessment, 20*, 270–280.
- Geiser, C., Eid, M., Nussbeck, F. W., Courvoisier, D. S. & Cole, D. A. (2010). Analyzing true change in longitudinal multitrait-multimethod studies: Application of a multimethod change model to depression and anxiety in children. *Developmental Psychology, 46*, 29.
- Kovacs, M. (1981). Rating scales to assess depression in school-aged children. *Acta Paedopsychiatrica, 46*, 305–315.
- Kovacs, M. (1982). *The children's depression inventory: A self-rating depression scale for school-aged youngsters*. Unpublished test manual.
- Lefkowitz, M. & Tesiny, E. (1980). Assessment of childhood depression. *Journal of Consulting Clinical Psychologists, 48*, 43–50.
- Little, T. D., Cunningham, W. A., Shahar, G. & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling, 9*, 151–173.
- Majcen, A.-M., Steyer, R. & Schwenkmezger, P. (1988). Konsistenz und Spezifität bei Eigenschafts- und Zustandsangst. *Zeitschrift für Differentielle und Diagnostische Psychologie, 9*, 105–120.
- Marsh, H. W. & Grayson, D. (1994). Longitudinal confirmatory factor analysis: Common time-specific, item-specific, and residual-error components of variance. *Structural Equation Modeling, 1*, 116–145.
- Marsh, H. W. & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. E. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 117–187). Thousand Oaks, London: Sage.
- Muthén, L. K. & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.
- Reynolds, C. R. & Richmond, B. O. (1978). What I think and feel: A revised measure of children's manifest anxiety. *Journal of Abnormal Child Psychology, 6*, 271–280.
- Schermelleh-Engel, K., Keith, N., Moosbrugger, H. & Hodapp, V. (2004). Decomposing person and occasion-specific effects: An extension of the latent state-trait (LST) Theory to hierarchical LST models. *Psychological Methods, 9*, 198–219.
- Schmitt, M. J. (2000). Mother-daughter attachment and family cohesion: Single and multi-trait latent state-trait models of current and retrospective perceptions. *European Journal of Psychological Assessment, 16*, 115–124.
- Steyer, R. (1987). Konsistenz und Spezifität: Definition zweier zentraler Begriffe der Differentiellen Psychologie und einfaches Modell zu ihrer Identifikation. *Zeitschrift für Differentielle und Diagnostische Psychologie, 8*, 245–258.
- Steyer, R. (1989). Models of classical test theory as stochastic measurement models: Representation, uniqueness, meaningfulness, and testability. *Methodika, 3*, 25–60.
- Steyer, R., Majcen, A.-M., Schwenkmezger, P. & Buchner, A. (1989). A latent state-trait anxiety model and its application to determine consistency and specificity coefficients. *Anxiety Research, 1*, 281–299.
- Steyer, R., Ferring, D. & Schmitt, M. (1992). States and traits in psychological assessment. *European Journal of Psychological Assessment, 8*, 79–98.
- Steyer, R., Mayer, A., Geiser, C. & Cole, D. A. (2015). A theory of states and traits-revised. *Annual Review of Clinical Psychology, 11*, 71–98.
- Steyer, R., Schwenkmezger, P. & Auer, A. (1990). The emotional and cognitive components of trait anxiety: A latent state trait anxiety model. *Personality and Individual Differences, 11*, 125–134.
- Thorndike, E. L. (1920). A constant error in psychological rating. *Journal of Applied Psychology, 4*, 25–29.
- Yousfi, S. & Steyer, R. (2006). Latent-State-Trait-Theorie. In F. Petermann & M. Eid (Hrsg.), *Handbuch der Psychologischen Diagnostik* (S. 346–357). Göttingen: Hogrefe.