

# Formative Messmodelle und Möglichkeiten ihrer Anwendung im empirisch-pädagogischen Kontext

Datengeleitete Indexbildung nach der MARI-  
Methode

Dr. phil. habil. Julia Fluck, Dipl.-Psych.  
RWTH Aachen University  
Institut für Erziehungswissenschaft  
LFG Schulpädagogik und empirische Bildungsforschung  
Eilfschornsteinstraße 7, Raum 112  
52062 Aachen  
Tel.: +49 241 80 96297  
[julia.fluck@rwth-aachen.de](mailto:julia.fluck@rwth-aachen.de)  
<http://www.ezw.rwth-aachen.de>

---

## Zitiervorschlag:

Fluck, J. (2020). *Formative Messmodelle und Möglichkeiten ihrer Anwendung im empirisch-pädagogischen Kontext – Datengeleitete Indexbildung nach der MARI-Methode*. RWTH Aachen: Institut für Erziehungswissenschaft.

## Zusammenfassung

Das übliche Messmodell, das Strukturgleichungsmodellen zugrunde liegt, das so genannte reflektive Messmodell, basiert auf der Grundannahme, dass das Antwortverhalten der Probanden und Probandinnen in Bezug auf einzelne Items durch das Konstrukt determiniert wird. Die Items messen alle das Gleiche und sind untereinander austauschbar. Dem steht mit dem formativen Messmodell eine Alternative gegenüber, die in den Wirtschaftswissenschaften einige Beachtung erfährt, im pädagogisch-psychologischen Kontext hingegen weitgehend unbeachtet ist. Beim formativen Messmodell wird die Bedeutung des Konstruktes erst durch die Items generiert, die jeweils distinkte Aspekte des Konstruktes abbilden. In diesem Beitrag wird zunächst die Logik des formativen Messmodells dargestellt und auf kritische Aspekte eingegangen. Im Anschluss wird mit der MARI-Methode eine auf qualitativen und quantitativen Daten basierende Prozedur vorgestellt, anhand derer die Angemessenheit und Güte formativer Modelle empirisch geprüft werden kann.

Schlagworte: Messmodellierung; formative Modelle; Strukturgleichungsmodelle; Forschungsmethoden; Indexbildung

## 1. Formative Messmodellierung

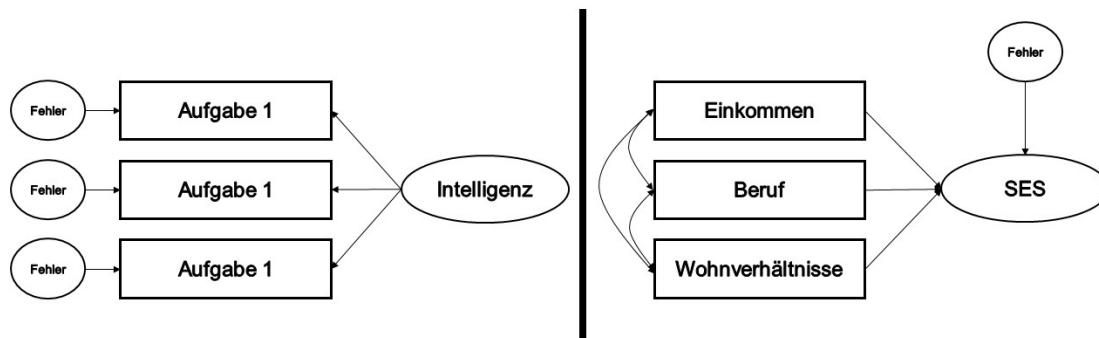
In der erziehungswissenschaftlichen Forschung hat man es häufig mit latenten Konstrukten (bspw. Intelligenz, Motivation, Kompetenzen, Selbstkonzept etc.) zu tun, die sich der direkten Beobachtung entziehen und nur mithilfe von manifesten Indikatoren annäherungsweise erhoben werden können.

Intelligenz ist beispielsweise eine latente Variable, die man nicht direkt messen kann. Was man hingegen messen kann, ist die Fähigkeit, bestimmte Aufgaben (Indikatoren) lösen zu können, was wiederum einen Rückschluss auf die Intelligenz erlaubt (Bollen, 1989). Wenn in Strukturgleichungsmodellen Zusammenhänge zwischen latenten Konstrukten untersucht werden, so wird im Messmodell angegeben, aus welchen Indikatoren sie sich jeweils zusammensetzen. Das Messmodell ist eine konfirmatorische Faktorenanalyse, die den Zusammenhang zwischen latentem Konstrukt und Indikatorvariablen modelliert.

Diese Art der Modellierung wird als reflektives Messmodell bezeichnet, weil die Indikatoren ein nicht direkt beobachtbares Konstrukt widerspiegeln. Wie im linken Teil von Abbildung 1 dargestellt, geht der Pfeil vom latenten Konstrukt zu den Indikatoren. Diese Konvention kommt daher, dass das Konstrukt den Indikatoren als kausal<sup>1</sup> vorgeordnet verstanden wird, was sich leicht veranschaulichen lässt, wenn man von einer Veränderung ausgeht. Erhöht sich beispielsweise die Intelligenz eines Kindes, so steigt damit die Wahrscheinlichkeit, dass Indikatoritems richtig gelöst werden. Umgekehrt ist es jedoch nicht schlüssig, dass eine Veränderung im Antwortverhalten kausal eine Erhöhung der Intelligenz verursachen kann.

---

<sup>1</sup> Edwards und Bagozzi (2000) sprechen sich entgegen der häufig vorzufindenden Nomenklatur deutlich dagegen aus, den Begriff der Kausalität für die Beziehung zwischen Konstrukt und Indikator zu verwenden, da die Bedingungen für Kausalität nicht oder nur sehr eingeschränkt, gegeben sind. Der Kausalitätsbegriff wird im Folgenden demnach nur für die Beschreibung der Modelle, nicht aber bei der Interpretation von Ergebnissen herangezogen.



Im reflektiven Modell (links), ist die Varianz jedes Indikators erklärt durch das latente Konstrukt und einen indikatorspezifischen Messfehler. Im formativen Modell (rechts) ist das latente Konstrukt die zu erklärende Variable, weshalb weder die Varianz der Indikatoren noch die Kovarianz zwischen den Indikatoren durch das Modell erklärt werden. Die Kovarianzen müssen also gesondert spezifiziert werden.

Abbildung 1 Reflektives (links) und formatives (rechts) Messmodell

Neben dem reflektiven Messmodell ist aber noch eine andere Art von Zusammenhang zwischen Konstrukt und Indikatoren möglich, die in der pädagogischen und psychologischen Forschung weitgehend ignoriert wird, im letzten Jahrzehnt in den Wirtschaftswissenschaften aber ein „Revival“ erfährt (Simonetto, 2012; Weiber & Mühlhaus, 2014). Formative Messmodelle, bereits 1962 vorgeschlagen von Curtis und Jackson, bezeichnen Beziehungen, in denen das Konstrukt nicht unabhängig von den Indikatoren existiert, sondern erst durch diese „geformt“ wird (Edwards & Bagozzi, 2000). In der Literatur wird häufig der Sozioökonomische Status (SES) als Beispiel für ein formatives Messmodell angeführt. Der SES ist ein latentes Konstrukt, das sich aus mehreren Indikatoren zur Beschreibung von Lebensumständen zusammensetzt, z. B. Einkommen, Beruf, Ausbildung und Wohnverhältnisse (Hauser, 1971). Wie in Abbildung 1 (rechte Seite) dargestellt, zeigen hier die Pfeile von den Indikatoren zum Konstrukt, denn „[im formativen] Modell konstituiert sich also das Konstrukt aus den es beeinflussenden Indikatoren. Damit stehen diese der latenten Variablen auch kausal vor“ (Eberl, 2004, S. 6). Formative Indikatoren werden deshalb in der englischsprachigen Literatur mitunter als „cause indicators“ bezeichnet, reflektive als „effect indicators“.

Beim reflektiven Messmodell würde man erwarten, dass eine Person mit hoher Intelligenz, die Item A lösen kann, mit hoher

Wahrscheinlichkeit auch Item B und C lösen wird<sup>2</sup>. Im formativen Messmodell gibt es keine sachlogische Begründung für eine solche Korrelation, wie an einem weiteren Beispiel deutlich wird: In der Stressforschung werden verschiedene Arten von Belastungen abgefragt, um die Gesamtbelastung einer Person abbilden zu können. Für die Gesamtbelastung können Lebensumstände wie hohe Anforderungen im Beruf, kürzlicher Tod eines Angehörigen etc. sorgen. Hier gibt es im Gegensatz zum reflektiven Modell nicht die Erwartung, dass eine Person mit hohem Belastungsniveau, die gerade unter starken Anforderungen im Beruf leidet, auch mit hoher Wahrscheinlichkeit angeben wird, dass gerade ein Angehöriger verstorben ist (siehe Edwards & Bagozzi, 2000).

### 1.1 Abgrenzung zum reflektiven Modell

Neben dem zentralen Unterschied bezüglich der Kausalbeziehungen zwischen Konstrukt und Indikatoren gibt es einige weitere Unterschiede zwischen den Modellierungsvarianten und daraus resultierende Konsequenzen. In Tabelle 1 sind die im Folgenden diskutierten Unterschiede im Überblick dargestellt.

---

<sup>2</sup> Vorausgesetzt die Items weisen eine ähnliche Schwierigkeit auf.

Tabelle 1 Unterschiede zwischen reflektivem und formativem Messmodell

Eigenschaften	Reflektives Messmodell	Formatives Messmodell
Modellgleichung	$x_i = \lambda_i \eta + \varepsilon_i$	$\eta = \sum \gamma_i x_i + \zeta$
Bedeutung der Indikatoren	Stichproben aus einem Itemuniversum	Distinkte Aspekte des Konstruktes
Bedeutung der Entfernung eines Indikators	Verringerung der Reliabilität	Veränderung der Bedeutung des Konstruktes
Korrelation der Indikatoren	nur um den Messfehler verringerte perfekte Interkorrelationen zwischen Items	Indikatoren können positiv, negativ oder zu null korrelieren
Skalenbereinigung	Reliabilitätsanalysen	Expertenurteile, Externe Kriterien, Kollinearitätsanalyse
Messfehlers	Messfehler auf Itemebene	Messfehler auf Konstruktebene
Modellidentifikation	Isolierte CFA ab 3 Indikatoren identifiziert	Isolierte CFA grundsätzlich unteridentifiziert
Überprüfung des Messmodells	CFA	MIMIC-Modell

### Modellgleichung

Im reflektiven Messmodell sind die Indikatoren die abhängigen Variablen, da sich die Antworten auf ein Item durch das dem Item zugrunde liegende latente Konstrukt erklären lässt. Die Reaktion eines Individuums auf ein Item  $x_i$  lässt sich als lineare Funktion des latenten Konstrukts  $\eta$ , der jeweiligen Ladung  $\lambda_i$  und eines Messfehlers  $\varepsilon_i$  ausdrücken:

$$x_i = \lambda_i \eta + \varepsilon_i$$

Im formativen Messmodell hingegen ist das latente Konstrukt die zu erklärende Variable, da dieses von den Indikatoren abhängig ist, die das Konstrukt formen. Da die Indikatoren  $x_i$  durch das Modell nicht erklärt werden, weisen sie keinen Messfehler auf. Stattdessen wird der latenten Variable  $\eta$  ein Fehler  $\zeta$  zugewiesen (s. u.). Für die Bildung des latenten Konstruktes wird die Regression als Modell gewählt, weil bei der Regression jeder Indikator mit seinem jeweiligen Regressionskoeffizienten  $\gamma$  gewichtet in die latente Variable einfließt (Diamantopoulos, Riefler, & Roth, 2008).

$$\eta = \sum \gamma_i x_i + \zeta$$

#### Bedeutung der Indikatoren und Konsequenzen ihrer Entfernung

Laut Grundannahmen der klassischen Testtheorie<sup>3</sup> (Lord & Novick, 1968) setzt sich der Testwert einer Person bei der Beantwortung eines Items aus dem wahren Wert zusammen, der durch das latente Konstrukt determiniert ist, und aus einem itemspezifischen Messfehler. Geht man von einem eindimensionalen Konstrukt aus, so sind die Items in einem Test Stichproben aus einem endlichen oder unendlichen Itemuniversum. Allein der itemspezifische Messfehler ist dafür verantwortlich, dass die Antworten eines Individuums auf zwei Indikatoren unterschiedlich ausfallen und folglich, dass die Korrelation zwischen zwei Items derselben Skala nicht perfekt ist. Vor diesem Hintergrund macht es im Beispiel eines Intelligenztestes theoretisch keinen Unterschied, welches von zwei möglichen Items in den Test aufgenommen wird, da beide Ausdruck desselben latenten Konstruktes sind, dieses jedoch womöglich unterschiedlich gut messen. Auch das Entfernen von Items – weil etwa aus ökonomischen Gründen die Skala gekürzt werden muss – vermindert lediglich die Reliabilität der Messung und ist somit weitgehend unbedenklich sofern die verbleibenden Items eine zufriedenstellende Reliabilität aufweisen.

Im Gegensatz dazu stellen Indikatoren im formativen Modell distinkte Aspekte des Konstruktes dar und sind nicht untereinander austauschbar. Einen Indikator wegzulassen, würde die inhaltliche

---

<sup>3</sup> Auch die probabilistische Testtheorie (Embretson & Reise, 2000) ist mit der Annahme des reflektiven Messmodells kongruent, da ein latenter Fähigkeitsscore dem Antwortverhalten bezüglich einzelner Items zugrunde liegt und dieses beeinflusst.

Bedeutung des Konstruktes verändern. Dies wird am o. g. Beispiel des sozioökonomischen Status deutlich: Lässt man den Aspekt „Beruf“ weg, so resultiert eine andere Bedeutung des Konstruktes als wenn auf den Aspekt „Wohnverhältnisse“ verzichtet würde.

#### Korrelation der Indikatoren und Skalenbereinigung

Im reflektiven Modell ist die Korrelation der Indikatoren eine Voraussetzung, die auf der o.g. Grundannahme beruht, dass die einzelnen Items Stichproben aus einem imaginären Itemuniversum darstellen und alle dasselbe messen. Maßnahmen zur Skalenbereinigung im Sinne der Verbesserung der Reliabilität einer Skala (und in Konsequenz, deren Validität) sehen daher vor, Items zu entfernen, die mit dem zu latenten Konstrukt – gemessen an den anderen Items – nicht in hohem Zusammenhang stehen. So gelten Trennschärfen von  $< .30$  (Brown, 2006) als Hinweise darauf, dass man es mit einem „schlechten“ Item zu tun hat, das von der Skala entfernt werden sollte.

Hingegen gibt es im formativen Modell aufgrund der Tatsache, dass die Items distinkte Aspekte eines Konstruktes erfassen, keine Voraussetzung, dass die Indikatoren miteinander korrelieren. Laut Grundannahmen des formativen Modells dürfen sie jedoch korrelieren, was in der Praxis – und sei es nur aufgrund von common-method-variance – auch vorkommt (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Daher eignen sich für das formative Modell die üblichen Prozeduren zur Konstruktvalidierung nach der klassischen Testtheorie (z. B. innere Konsistenz als Maß der Reliabilität; Cronbachs Alpha; Analyse der Kovarianzstruktur durch CFA) nur sehr bedingt (Bollen, 1989). Um Skalen im formativen Modell zu bereinigen, muss daher auf andere Maßnahmen zurückgegriffen werden (s. u.): Kollinearitätsanalysen, Befragungen von Experten sowie externe Validierungskriterien (Eberl, 2004).

#### Bedeutung des Messfehlers

Während im reflektiven Modell für jeden Indikator ein Messfehler geschätzt wird, wird der Fehlerterm  $\zeta$  im formativen Modell auf der Ebene der latenten Variable modelliert (s.u.).



## Modellidentifikation und Überprüfung des Messmodells

Ein formatives Messmodell ist – isoliert betrachtet – grundsätzlich nicht identifiziert.<sup>4</sup> Um das Identifikationsproblem zu lösen, sind mehrere Möglichkeiten gegeben (Diamantopoulos et al., 2008; Wilcox, Howell & Breivik, 2008), die aber gemeinsam haben, dass mindestens zwei weitere Pfade in das Modell mit aufgenommen werden müssen (Wilcox et al., 2008):

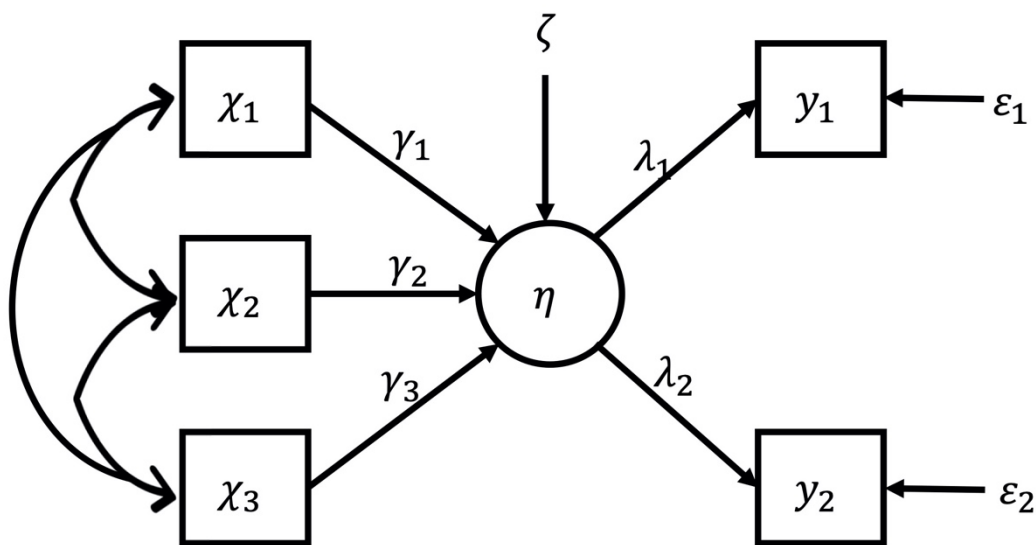


Abbildung 2 MIMIC-Modell - neben drei formativen Indikatoren  $x_1$ - $x_3$  enthält das Modell auch zwei reflektive Indikatoren  $y_1$  und  $y_2$ . Dadurch ist das Modell identifiziert.

Im sog. MIMIC-Modell (multiple indicators, multiple causes, Jöreskog & Goldberger, 1975) wird das latente Konstrukt nicht nur durch die formativen, sondern zusätzlich durch mindestens zwei reflektive Indikatoren bestimmt (siehe Abbildung 2). Das MIMIC-Modell stellt also eine (teil)formative Entsprechung zur CFA dar, da das latente Konstrukt isoliert betrachtet wird.

<sup>4</sup> Beispiel: Das in Abbildung 1 dargestellte reflektive Modell ist mit sechs zu schätzenden (drei Ladungen und drei Messfehlern) und mit sechs vorliegenden Informationen (je drei Varianzen und Kovarianzen) in der Regel gerade identifiziert (Eid, Gollwitzer, & Schmitt, 2015). Beim formativen Modell müssen hingegen sieben Parameter geschätzt werden (drei Interkorrelationen der Indikatoren, drei Regressionsgewichte und ein latenter Fehlerterm).

Ein anderer Ansatz findet im Rahmen eines Strukturgleichungsmodells Anwendung: Das formativ gemessene Konstrukt wird entweder per Regression mit mindestens zwei manifesten Variablen oder mit zwei reflektiv gemessenen latenten Konstrukten in Verbindung gebracht (Bollen, 1989). Strukturgleichungsmodellen liegt bei den meisten Programmen (z. B. AMOS, LISREL, Mplus) ein Schätzverfahren zugrunde, das versucht, den Unterschied zwischen der empirischen und der durch das Modell implizierten Kovarianzmatrix zu minimieren. Während dieser Ansatz für die reflektive Messung sehr gut geeignet ist, können sich bei formativen Modellen mit geringer Interkorrelation der Items Probleme ergeben. Für formative Modelle wird daher mitunter empfohlen, Strukturgleichungsmodelle aufgrund der partial-least-squares (PLS) Methode zu schätzen, bei welcher der Schätzalgorithmus die Varianz der (endogenen) latenten Variable zu maximieren sucht, Interkorrelation der Items also keine Voraussetzung ist (Simonetto, 2012).

### 1.2 Güteprüfung

Zur Güteprüfung formativer Messmodelle werden in der Literatur eigene Verfahren vorgeschlagen, da das Prozedere zur Skalenbereinigung für reflektive Modelle aufgrund der mangelnden Korrelationsannahme ungeeignet ist.

Bei reflektiven Modellen werden beispielsweise Items entfernt, wenn sie mit den anderen Items der Skala nicht in ausreichend hohem Zusammenhang stehen (Trennschärfe  $< .30$ ). Im formativen Modell messen die Indikatoren jedoch distinkte Aspekte des Konstruktes. Zur Feststellung von Indikatorsignifikanz und -relevanz bedarf es daher eines Außenkriteriums, das zur Validierung der formativen Indikatoren herangezogen wird. Im oben beschriebenen MIMIC-Modell dienen beispielsweise die reflektiven Indikatoren als Validierungskriterien. Indikatoren sollten eine Relevanz (im Sinne einer Korrelation mit einem oder mehreren Außenkriterien) von mindestens  $r = .20$  aufweisen und die Indikatoren sollten in ihrer Gesamtheit dazu führen, dass das MIMIC-Modell ein Bestimmtheitsmaß von mindestens  $R^2 = .20$  aufweist (Chin, 1998).

Das Entfernen von Indikatoren aus der Skala kann weitreichende Konsequenzen haben, da dadurch womöglich die Bedeutung des Konstruktes beschnitten wird. Skalenbereinigung muss bei formativen Konstrukten daher mit Bedacht realisiert werden. Aus diesem Grund weisen Diamantopoulos et al. (2008) darauf hin, dass formative Skalenbereinigung nicht ausschließlich auf statistischen Erwägungen gegründet sein darf, sondern dass Entscheidungen über Itemselektion stets auch sachlogische Überlegungen beinhalten müssen.

Neben statistischen Analysen wird daher empfohlen (Eberl, 2004), die Itemselektion aus Expertenurteilen zu begründen. Experten sollen dabei eine Entscheidung dahingehend treffen, ob alle Items einer Skala tatsächlich zum Konstrukt gehören und ob weitere Aspekte des Konstruktes fehlen, die in zusätzlichen Items operationalisiert werden sollten.

Zu beachten gilt, dass formative Modelle von mehrdimensionalen reflektiven Modellen zu unterscheiden sind. Auch bei Faktoren höherer Ordnung ist ebenso wie bei Faktoren erster Ordnung die Frage nach der Richtung der Kausalität zu beantworten und kann zugunsten der reflektiven oder zugunsten der formativen Modellierung zu beantworten sein.

## 2. Bedeutung des formativen Modells im empirisch-pädagogischen Kontext

Während in den Wirtschaftswissenschaften das formative und das reflektive Messmodell parallel gelehrt und angewandt werden (Weiber & Mühlhaus, 2014), scheint in den empirischen Humanwissenschaften ein geringes Bewusstsein für die formative Modellierungsvariante zu existieren. Einige wenige Themenbereiche bilden eine Ausnahme, beispielsweise im Bereich der Messung von Mobbing und Cybermobbing wird von einigen Autoren die These vertreten, dass die Konstrukte formativ zu modellieren seien<sup>5</sup> (Festl, 2015; Savage, 2012; Teimouri et al., 2015; für eine ausführliche Diskussion siehe Fluck, 2020).

---

<sup>5</sup> Streng genommen ist die Frage nach der formativen oder reflektiven Modellierung allerdings immer nur für eine konkrete Art der Operationalisierung beantwortbar und nicht generell auf der Ebene des Konstrukts anzusiedeln.

Der sozioökonomische Status ist nicht das einzige Konstrukt in der pädagogischen Forschung, bei dem eine formative Modellierung naheliegt.

- Gesundheitsrelevante Konstrukte wie soziale Unterstützung, Bewältigungsstrategien und Emotionsregulation erfordern laut Smith (2011) einen genaueren Blick auf die Beziehung zwischen Indikatoren und Konstrukt. Sind solche Strategien zunächst unabhängig voneinander und ist erfolgreiche Stressbewältigung, Emotionsregulation, soziale Unterstützung etc. bereits gegeben, wenn eine von mehreren Strategien eingesetzt wird, ohne dass die Nutzung einer Strategie auch mit der Nutzung einer anderen einhergeht, so liegt die formative Spezifikation nahe.
- Wenn im Marketingbereich der Abbruch von Kundenbeziehungen (Siehe z.B. Bruhn, Lucco, & Wyss, 2008) mit formativen Modellen untersucht wird, so ist auch eine Übertragung auf pädagogische Kontexte denkbar (z. B. Abbruch von Ausbildung und Studium).
- In der Forschung zu beruflichen Handlungskompetenzen (Marquardt, 2019) findet sich ein weiteres Beispiel: Verschiedene Facetten der Personalkompetenz (z. B. Reflexion, Motivationsregulation) setzen sich additiv zu einem Gesamtfaktor zusammen, der jedoch nicht gleichermaßen diese Einzelfacetten bewirkt. Bei anderen Kompetenzkonstrukten dürften sich weitere Anwendungsfelder des formativen Modelles bieten.

Auch aus einer wissenschaftsphilosophischen Perspektive ist die formative Modellierung interessant. Borsboom, Mellenbergh, and Van Heerden (2003) analysieren verschiedene Konzepte latenter Variablen und weisen darauf hin, dass das reflektive Modell beispielsweise in der Psychologie schon allein deshalb präferiert wird, weil es einer vom operationalistischen Realismus geprägten Sichtweise entspricht. Das Konstrukt per se existiert irgendwo im Individuum als „wahrer Wert“ (s.o.), der durch die Wahrnehmung und Messung anhand vorhandener Instrumente jedoch nur

annäherungsweise – und daher messfehlerbehaftet – bestimmt werden kann.

Das formative Modell, welches keinen Messfehler beinhaltet, wird daher von Psychometrikern als kontraintuitiv empfunden (Edwards, 2011). Das formative Modell entspricht hingegen eher einer konstruktivistischen Sichtweise (Borsboom et al., 2003; Edwards, 2011). Das latente Konstrukt existiert hier nicht unabhängig vom Beobachter, ergo nicht unabhängig von seiner Messung anhand konkreter Indikatoren und wird durch diese erst geformt.

Während für den wirtschaftswissenschaftlichen Bereich Daten vorliegen, dass etwa ein Drittel aller Messmodelle als reflektiv fehlspezifiziert sind, obwohl ein formatives Modell vorliegt (Jarvis, MacKenzie, & Podsakoff, 2003) gibt es für den sozial- und humanwissenschaftlichen Bereich keine vergleichbaren Informationen. Simulationsstudien zeigen zwar, dass bei fehlerhafter Modellierung Fit-Indizes und Ladungsparameter fehlgeschätzt werden (Diamantopoulos et al., 2008). Die Wahl der falschen Modellierungsvariante wird jedoch selten als solche erkannt, da einerseits diese statistische Effekte – insbesondere bei hohen Interkorrelationen der Items – nur marginal sind und andererseits bei den meisten Anwendern und Anwenderinnen kein Bewusstsein für die Existenz der formativen Modellierungsvariante existiert.

Die Konsequenzen von Fehlmodellierungen sind eher auf inhaltlicher Ebene zu verorten. Geht der/die Anwender/in davon aus, dass ein reflektives Modell vorliegt, so wird er oder sie „schlechte Items“ im Sinne einer geringen Item-Skala-Korrelation entfernen. Da formative Items jedoch häufig distinkte Aspekte des Konstruktes abbilden, wird dessen Bedeutung durch die Selektion dieser Items beschnitten. Dies ist bemerkenswerterweise gerade bei denjenigen Items besonders deutlich der Fall, die nur gering mit den anderen Items korrelieren. Eine geringe Interkorrelation mit anderen Indikatoren erhöht die Wahrscheinlichkeit, dass das Item einer Selektion zum Opfer fällt und macht zugleich die Bedeutung dieser Selektion besonders folgenswer, da die Validität des Konstruktes womöglich beschränkt wird.

### 3. Kritik am formativen Ansatz und Lösungsmöglichkeiten

Neben Kritik an konkreten Modellierungsvarianten<sup>6</sup> findet man in der Literatur auch einige Kritikpunkte am formativen Modell per se (für einen umfassenden Überblick siehe Edwards, 2011). Zunächst betrifft diese das Konzept des Messfehlers.

Der auf Konstruktebene angesiedelte Messfehler wird entweder als kumulierter Messfehler über alle Indikatoren interpretiert (Jarvis et al., 2003) oder als derjenige Varianzanteil der latenten Variable, der nicht durch die Indikatoren erklärt wird (Diamantopoulos, 2006). Dass der Messfehler nicht auf Itemebene angesiedelt ist liegt daran, dass hier nicht der Anspruch besteht, die Varianz im Indikator durch das Konstrukt zu erklären. Dadurch, dass der Indikator die unabhängige Variable in der Gleichung ist, bleibt ihre Varianz unbekannt, so dass sich darin der Anteil des wahren Wertes nicht vom Fehleranteil separieren lässt. Edwards (2011) macht deutlich, dass die Lesart des formativen Modells damit aber zugleich den Eindruck erweckt, dass Indikatoren perfekt reliabel seien<sup>7</sup>. Mit dem Ansatz des formativen Modells ist jedoch auch die Deutung konform, dass die Indikatoren zwar nicht fehlerfrei messen, der Anteil des Fehlers jedoch nicht quantifiziert werden kann.

Schwerwiegender ist die Kritik, die u. a. von (Markus, 2014) hervorgebracht wird. Er argumentiert, dass das formative Modell Tür und Tor zu missbräuchlicher Datennutzung öffne, indem die Wahl zwischen reflektivem und formativem Modell dem Anwender die Möglichkeit lasse, eine aufgrund geringer Interkorrelationen schlechte reflektive Skala a posteriori als formativ zu deklarieren und somit deren Einsatz zu rechtfertigen. Markus (2014) rät, von formativen Modellierungen abzusehen, da diese Wegbereiter für „bad science“ seien. Mit dieser Kritik wird tatsächlich ein problematischer Aspekt der formativen Modellierung

---

<sup>6</sup> Beispielsweise wird das MIMIC Modell von Edwards (2011) und von Simonetto (2012) als kontraintuitiv empfunden, da hierbei formative Items anhand reflektiver validiert werden. Reflektive Items seien jedoch – so die Kritik – die Folgen eines Konstruktes und es sei nicht schlüssig, die Ursachen des Konstruktes aus diesen abzuleiten.

<sup>7</sup> Ein messtheoretisch eleganter Vorschlag zur Lösung des Problems findet sich bei Edwards und Bagozzi (2000), wo die formativen Indikatoren als latente Konstrukte aufgefasst werden und durch ein je eigenes Item determiniert sind. Der Ladungsparameter für dieses Item kann jeweils vorab fixiert werden, je nach angenommener Reliabilität des Indikators.

angesprochen. Die Entscheidung für eine formative Modellierung muss zwingend a priori – bereits bei der Konstruktion der Skalen fallen.

Autoren wie Bainter und Bollen (2015) oder (Eberl, 2004) versuchen diesem Punkt entgegenzutreten, indem sie Möglichkeiten aufzeigen, wie die Angemessenheit des formativen Modells empirisch überprüft werden kann.

Bollen und Ting (1993, 2000) empfehlen die Verwendung des konfirmatorischen TETRAD-Tests. Dieser prüft, ob die aus der Modellgleichung des reflektiven Modells abgeleiteten Voraussetzungen<sup>8</sup> Gültigkeit besitzen. Die praktische Anwendung des TETRAD-Tests ist mithilfe eines SAS-Makros (Hipp, 2008) oder seit geraumer Zeit auch mithilfe des PLS-Strukturgleichungsmodellierungsprogrammes smartPLS (Ringle, Wende & Becker, 2015) möglich.

Eberl (2004) beschreibt einen komplexeren Algorithmus, der aus mehreren aufeinanderfolgenden Tests besteht. Zunächst soll der/die Anwender/in auf Basis der Theorie eine Spezifikationshypothese aufstellen, ob das Konstrukt auf Basis der vorliegenden Operationalisierung reflektiv oder formativ zu modellieren ist. Die Skalenbereinigung erfolgt bei formativer Spezifikationshypothese dann bereits mit den Methoden, die beim formativen Modell üblich sind: Expertenurteile und Analysen des Zusammenhangs der Items mit anderen Variablen, die ebenfalls das Konstrukt messen. Eberl (2004) sieht analog zu Bollen und Ting (2000) den TETRAD-Test vor, um die reflektive Modellierungsvariante zu bestätigen. Ebenfalls soll die reflektive Modellierung mithilfe einer konfirmatorischen Faktorenanalyse im LISREL-Modell verworfen werden. Ist die reflektive Variante anhand beider Prüfverfahren verworfen worden, so empfiehlt Eberl (2004) abschließend die Überprüfung<sup>9</sup> des formativen Messmodells,

---

<sup>8</sup> Wenn das reflektive Messmodell gilt, dann muss bei jeder denkbaren Tetrade aus vier Indikatoritems gelten, dass die Differenz zwischen dem Produkt aus zwei Kovarianzen und dem Produkt zweier weiterer Kovarianzen von Items null ist. Für eine Herleitung siehe (Fluck, 2020).

<sup>9</sup> Streng genommen handelt es sich dabei, wie der Autor ebenfalls betont, nicht um eine Überprüfung, sondern um ein Nicht-Verwerfen des formativen Modells.

etwa im Rahmen eines MIMIC-Modells oder einer umfassenderen Strukturgleichungsmodellierung mit weiteren Konstrukten, wobei er in beiden Fällen die Verwendung des PLS-Schätzalgorithmus empfiehlt.

Zunächst scheint die Möglichkeit einer empirischen Prüfung der richtigen Modellvariante eine gute Lösung zu sein. Zum einen suggeriert eine empirische Fundierung eine höhere Zuverlässigkeit als reine Gedankenexperimente und könnte dem/der Anwender/in bestätigen, die richtige Entscheidung getroffen zu haben. Zum anderen dienen empirische Daten auch gegenüber Dritten als bessere Rechtfertigung gegenüber dem „bad-science“-Argument. Dennoch sind im Zusammenhang mit solcherlei empirischen Verfahren zur Prüfung der Modellierungsvariante einige Schwierigkeiten zu bedenken.

Erstens stellt sich bei der Angemessenheitsprüfung anhand mehrerer Teilschritte die Frage, wie mit widersprüchlichen Ergebnissen aus mehreren Tests umzugehen ist. Was ist zu tun, wenn weder die formative noch die reflektive Modellierungsvariante verworfen werden kann? An dieser Stelle muss letztlich doch der/die Anwender/in eine Entscheidung treffen, wie die Richtung der Kausalität zwischen Item und Konstrukt zu bewerten ist.

Zweitens ist darauf hinzuweisen, dass sowohl der TETRAD-Test als auch die konfirmatorische Faktorenanalyse die Angemessenheit des reflektiven Modells dadurch prüfen, dass die Abweichung von der modellimpliziert perfekten Erklärbarkeit der Indikatoren durch das Konstrukt quantifiziert wird. Bei einem Modell wird demnach die reflektive Variante dann nicht verworfen, wenn zwischen den Items hohe Interkorrelationen bestehen. Faktisch bedeutet dies, dass ein Beibehalten des reflektiven Modells aufgrund dieser Tests keine Grundlage für die Ablehnung der formativen Variante darstellt, sofern es sich um ein formatives Modell handelt, bei dem – aus welchem Grund auch immer – die Items hoch miteinander korrelieren.



Drittens ist am TETRAD-Test ebenso wie an der PLS-Modellierung problematisch, dass beides mit Standardsoftware wie SPSS nicht durchgeführt werden kann. Ganz allgemein am formativen Modell kritisieren West and Grimm (2014), dass diese schwer zu rechnen und damit wenig praktikabel seien.

#### 4. Anwendung des formativen Modells im empirisch-pädagogischen Kontext

Ausgehend von der Annahme, dass es formative Messmodell gibt, lässt sich aus den bisherigen Ausführungen folgende Ausgangslage feststellen: Einerseits ist es aufgrund der Konsequenzen von Fehlmodellierungen wichtig, formative Operationalisierungen als solche zu erkennen und korrekt zu spezifizieren. Andererseits muss gewährleistet werden, dass die Umsetzung des formativen Modells den Kriterien der Wissenschaftlichkeit entspricht und es wäre wünschenswert, dass eine methodische Umsetzung auch auf Basis von Standardsoftware realisiert werden kann.

In der Praxis werden formative Modelle häufig nicht im Rahmen von MIMIC- oder Strukturgleichungsmodellen spezifiziert und eigens Programme zur PLS-Modellierung bemüht. Formative Konstrukte lassen sich deutlich weniger komplex modellieren, zum Beispiel durch einen einfachen Summenscore über die Indikatoren oder als gewichtete lineare Funktion der Indikatoren (Edwards & Bagozzi, 2000), mit denen dann auch in komplexeren Modellen weitergerechnet werden kann (siehe z. B. Festl, 2015).

Dabei grenzt es jedoch sicherlich an die von Markus (2014) formulierte Unwissenschaftlichkeit, wenn ein Konstrukt ohne theoretische oder empirische Begründung als formativ bezeichnet wird und ein Index gebildet wird, ohne dass eine Auseinandersetzung mit den Fragen nach Güte und Angemessenheit des formativen Modells erfolgt ist<sup>10</sup>. Im Idealfall erfolgt die Indexbildung nicht nur auf Basis theoretischer Reflektionen, sondern datengeleitet. Im Folgenden wird hierfür ein pragmatischer Ansatz vorgeschlagen, der zum Ziel hat, die Güte der

---

<sup>10</sup> Es sei denn, solche Erwägungen lassen sich für die gewählte Operationalisierung an anderer Stelle der Forschungsliteratur finden.

formativen Modellierung zu gewährleisten und die Frage nach der Angemessenheit nur im ersten Teilschritt thematisiert.

Die MARI-Methode (siehe Abbildung 3 und Erläuterungen in den folgenden Teilkapiteln) ermöglicht eine datenbasierte Indexbildung und ist an jeder Stelle auf (qualitativen oder quantitativen) empirischen Daten begründet. Eine detaillierte Darstellung der Vorgehensweise illustriert mit zahlreichen Beispielen findet sich bei Fluck und Lichtenberg (in Druck).

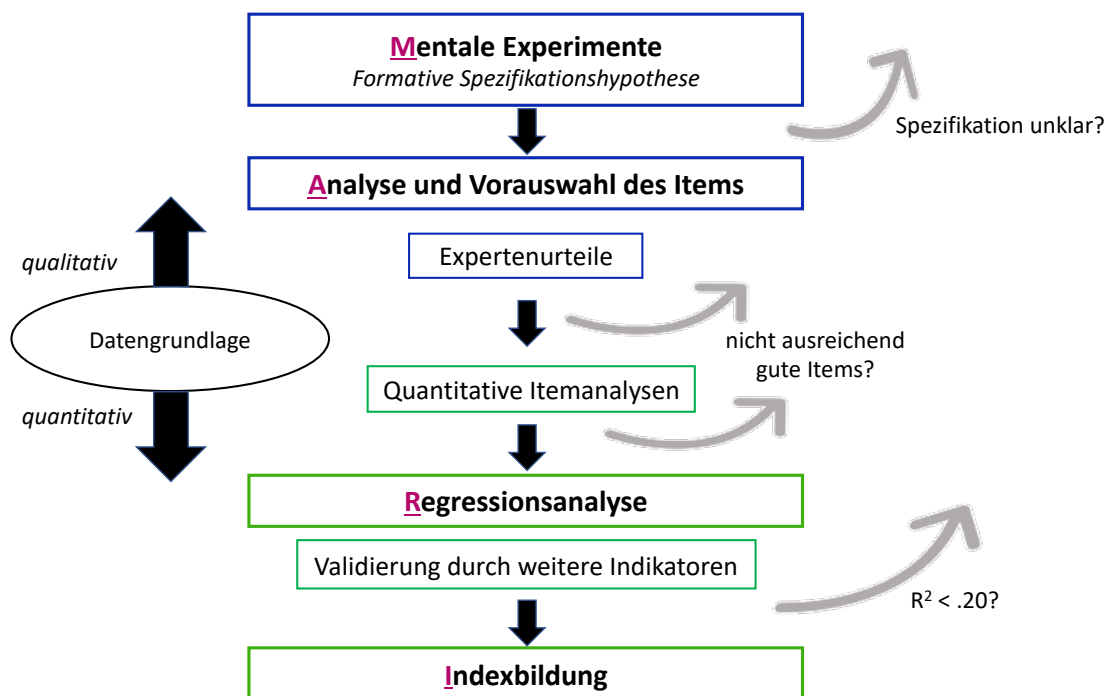


Abbildung 3 MARI - Methode zur datengeleiteten Indexbildung bei formativen Modellen

### Erster Schritt: Mentale Experimente

Damit das formative Modell nicht a posteriori missbraucht wird, um schlechte Skalen (d.h. solche mit geringer interner Konsistenz) zu rechtfertigen, muss die Spezifikation a priori im ersten Schritt festgelegt werden. Die weiteren (empirischen) Teilschritte der Prozedur fokussieren dann lediglich auf die Güte der formativen Modellierung.

Zu den mentalen Experimenten, wie sie Bollen (1989) vorschlägt, finden sich in der Literatur drei Aspekte der Unterscheidung

zwischen formativen und reflektiven Modellen, die möglichst unabhängig voneinander beantwortet werden sollen. Da Konstrukte nicht per se formativ oder reflektiv sind (Simonetto, 2012), müssen bereits Items oder zumindest Itementwürfe vorliegen, für die diese Fragen zu beantworten sind. Ist dies nicht der Fall und besteht der Wunsch, bereits bei der Konstruktion von Items eine spätere formative Modellierung vorzubereiten, so können die Leitfragen als Konstruktionshilfe genutzt werden.

1. Messen die Items eines Konstruktes alle „das Gleiche“ bzw. sind sie untereinander austauschbar (Bagozzi, 1984; Jarvis et al., 2003)?
2. Wie ist die Richtung der Kausalität, d. h. ergibt sich die Bedeutung des Konstruktes aus der Bedeutung der Indikatoren oder umgekehrt (Bagozzi, 1984; Bollen, 1989)?
3. Wenn sich auf einem Item eine Veränderung ergeben würde, müssten sich dann schlusslogisch auch auf den anderen Indikatoritems Veränderungen in die gleiche Richtung ergeben (Rossiter, 2002)?

Das mit den Mängeln der Introspektion verbundene Prinzip der mentalen Experimente lässt sich zu einer empirischen und damit der Prüfung von Objektivität zugänglichen Fragestellung machen, indem diese Experimente von mehreren Personen – idealerweise Expert/innen im Bereich des Themas – durchgeführt werden. Dies ist insbesondere vor dem Hintergrund hilfreich, dass die Antworten auf die drei Fragen womöglich nicht einheitlich ausfallen (Rossiter, 2002) und eine eindeutige Entscheidung nicht möglich ist. Dann müssen die Fragen aufgrund sachlogischer Erwägungen gewichtet werden, um zu einer Entscheidung zu gelangen. Diese Entscheidung kann dadurch validiert werden, dass mehrere Experten und Expertinnen bezüglich der Spezifikationshypothese zum gleichen Schluss gelangen.

Die grauen Pfeile in Abbildung drei deuten in jedem Schritt Abbruchprozesse an, da es je nach Resultat des jeweiligen Arbeitsschrittes angezeigt sein kann, zu einem früheren Schritt zurückzukehren. Im ersten Schritt müsste das Prozedere dann abgebrochen werden, wenn zwischen den Experten und Expertinnen keine Einigung bezüglich der Frage erzielt werden

kann, ob das Konstrukt anhand der angedachten Operationalisierung formativ zu modellieren ist oder nicht. Der Weg führt dann zurück zur konzeptionellen Ebene.

Andernfalls steht am Ende des ersten Schrittes entweder die Entscheidung, dass das Konstrukt reflektiv ist, dann sind die üblichen Maßnahmen zur Skalenbereinigung nach klassischer oder probabilistischer Testtheorie angezeigt; oder dass das Konstrukt formativ ist – in diesem Fall können die Items in einem zweiten Schritt ersten qualitativen und quantitativen Analysen unterzogen werden.

#### Zweiter Schritt: Analyse und Vorauswahl der Items

Der zweite Schritt untergliedert sich in zwei Teilschritte. Die Items werden zunächst qualitativ analysiert, bevor sie einer quantitativen Untersuchung bezüglich der Multikollinearität unterzogen werden.

#### Teilschritt 2a: Qualitative Analyse

Die inhaltliche Bedeutung eines formativ gemessenen Konstrukts wird vorab festgelegt. Diese Festlegung kann aus der bisherigen Theorie und Empirie zu einem Konstrukt abgeleitet oder unter Experten und Expertinnen ausgehandelt werden. Die Bedeutung des Konstruktes ist aufgrund der Logik des formativen Modells eng mit der Itemkonstruktion verbunden (Diamantopoulos & Winklhofer, 2001). Da die Items distinkte Aspekte des Konstruktes abbilden ist es entscheidend, dass dieses in größtmöglicher Breite abgedeckt wird, die Items also zum Zwecke der Maximierung von Validität möglichst exhaustiv sind. Im Zuge der Skalenkonstruktion sollten die Items vor ihrer Erprobung an der Zielgruppe wiederum Experten und Expertinnen vorgelegt werden (Eberl, 2004). Die Experten und Expertinnen sind dann gefordert, die vorläufige Skala in zweierlei Hinsicht zu bewerten:

Erstens ist für jedes Item die Frage zu beantworten, ob es einen (wichtigen) Aspekt des Konstruktes abbildet und somit in die Skala aufgenommen werden sollte oder nicht. Zweitens sind die Experten und Expertinnen angehalten, weitere Aspekte des Konstruktes zu nennen, die bisher nicht durch vorhandene Items operationalisiert sind. Die Expertenbefragung kann schriftlich und einzeln erfolgen,

alternativ ist eine gemeinsame und auch mehrstufige Befragung mittels Delphi-Methoden möglich. Letzterer Ansatz kann gewinnbringend sein, wenn neue Phänomene untersucht werden, zu denen es wenig theoretischen Hintergrund und/oder Forschungsliteratur gibt.

### Quantitative Studie

Auf Basis der Expertenurteile lässt sich eine Multi-Item-Skala konstruieren, die sodann an der Zielgruppe in einer ersten größeren Studie erprobt werden kann. Hierbei sind nicht nur die Items selbst den Probanden und Probandinnen vorzulegen, sondern möglichst weitere Kriterien, die (siehe Schritt 3) zur Validierung hinzugezogen werden können.

### Teilschritt 2b: Quantitative Analyse

Für die quantitative Analyse steht die Prüfung auf Multikollinearität im Fokus (Diamantopoulos & Winklhofer, 2001). Zusätzlich kann eine Inspektion von Korrelationstabellen oder Trennschärfen erfolgen.

Zwar dürfen Items im Rahmen formativer Modelle theoretisch sowohl positiv oder negativ als auch gar nicht miteinander korrelieren, jedoch stellen Indikatoren, die sehr hoch korrelieren, manchmal annähernd perfekte Linearkombinationen voneinander dar. Solche kollinearen Items sind nicht nur inhaltlich redundant, sondern führen auch zu Problemen bei regressionsanalytischen Verfahren. Regressionsgewichte werden systematisch unterschätzt und die Identifizierbarkeit von Strukturgleichungsmodellen kann beeinträchtigt sein (Weiber & Mühlhaus, 2014). Um die Kollinearität der Indikatoren zu prüfen, kann der Varianzinflationsfaktor (Backhaus, Erichson, Plinke, & Weiber, 2003) herangezogen werden. Dieser wird berechnet, indem zunächst für jedes Item eine multiple Regression gerechnet wird, bei der das zu betrachtende Item als abhängige Variable auf die restlichen Items als unabhängige Variablen regrediert wird. Der VIF eines Items errechnet sich dann wie folgt aus dem jeweiligen Bestimmtheitsmaß:

$$VIF = \frac{1}{1 - R^2}$$

Je nach Quelle werden Grenzwerte von  $VIF > 3$  bis  $VIF > 10$  angenommen, jenseits derer eine Selektion in Erwägung gezogen werden muss, jedoch bei formativen Modellen erst nach einer inhaltlichen Examination des Items (Diamantopoulos & Riefler, 2008; Diamantopoulos et al., 2008; Weiber & Mühlhaus, 2014). Da eine sehr hohe Varianzinflation darauf hindeutet, dass ein Item keine Aspekte des Konstruktes misst, die nicht von den anderen Items mit abgedeckt werden, dürfte die Entfernung von Items mit hohem VIF nur geringe Konsequenzen bezüglich der Validität der Messung haben.

Nachdem ggf. Items aufgrund von Multikollinearität entfernt wurden können die (verbleibenden) Indikatoren im nächsten Schritt anhand von regressionsanalytischen Verfahren validiert werden. Ist nach den qualitativen und quantitativen Analysen keine ausreichende Anzahl an Items übriggeblieben, so muss einen Schritt zurück gegangen werden, um neue Items zu konstruieren. Auch hier können Experten und Expertinnen hinzugezogen werden.

#### Dritter Schritt: Regressionsanalysen

Wie Eberl (2004) betont, sind formative Skalen neben Expertenurteilen auch durch externe Validierungen zu bereinigen. Hierfür können die oben beschriebenen MIMIC-Modelle eingesetzt werden, wenn es beispielsweise möglich ist, das fragliche Konstrukt mit mindestens zwei weiteren, reflektiven Indikatoren zu messen. Eine einfacher zu realisierende Variante besteht darin, nur eine einzelne Validierungsvariable hinzuzuziehen. Dies kann ein prognostisch validierendes Kriterium sein oder aber ein Globalmaß des Konstruktes. Bei Fluck (2020) wird beispielsweise das Erleben von Cybermobbing auf zweierlei Art erfragt, einerseits anhand einer formativ zu modellierenden Multi-Item-Skala, die verschiedene Arten von Cybermobbing erfasst, andererseits anhand einer singulären Globalfrage, die das Ausmaß der Erfahrungen mit Cybermobbing erfragt. Eine multivariate Regressionsanalyse mit der Globalfrage als abhängige Variable und den formativen Items als Prädiktoren zeigt auf, welche Items signifikant mit dem Konstrukt in Verbindung stehen und welche nicht. Laut Rossiter (2002) sind solche Einzelitems dann hinreichend zur Messung eines Konstruktes, wenn dieses von den Probanden und Probandinnen

einfach und einheitlich verstanden werden kann sowie nur eine einzelne Facette umfasst.

Dabei ist zu beachten, dass derartige Globalurteile teilweise nicht exakt das gleiche Konstrukt erfassen. Bei Varianzaufklärung in der abhängigen Variable ist daher nur mit moderaten Effekten zu rechnen. Analog zum MIMIC-Modell kann auch hier ein  $R^2$  von mindestens .20 als untere Grenze angenommen werden und neben der (stichprobenabhängigen) Indikatorsignifikanz auch die Indikatorrelevanz in Form der Höhe des  $\beta$ -Gewichtes erwogen werden. Ob auch Indikatoren mit  $\beta < .20$  (das entspricht dem Grenzwert für MIMIC Modelle bei Chin (1998)) beibehalten werden obliegt dabei wiederum der Entscheidung des Anwenders oder der Anwenderin. Möglicherweise kann dies sinnvoll sein, wenn durch solche Items ein nur kleiner Teil des Konstruktes erklärt wird, dieser Aspekt jedoch wesentlich ist.

Auch am Ende des dritten Schrittes kann es angezeigt sein, den Prozess abubrechen und zu einem früheren Stadium zurückzugehen. Wenn beispielsweise nach der Itemselektion nur noch wenige signifikant mit einem oder mehreren relevanten Kriterien in Verbindung stehende Items verbleiben und/oder diese in ihrer Gesamtheit nicht genug Varianz aufklären. In diesem Fall müssen ggf. weitere Items konstruiert (s.o.) und erneut empirisch erprobt werden.

Im Idealfall ist jedoch eine ausreichende Anzahl Items mit zufriedenstellender Messgüte übriggeblieben, die zur leichteren Arbeit mit dem formativen Konstrukt zu einem Index kombiniert werden.

#### Vierter Schritt: Indexbildung

Der Index kann als einfacher Summen- bzw. Mittelwert über diejenigen Indikatoren gebildet werden, von denen angenommen werden kann, dass sie valide und reliable Messungen des Konstruktes darstellen. Die qualitativen und quantitativen Daten sowie die sachlogischen und theoriebasierten Erwägungen der Anwender bilden die Basis für die datengeleitete Indexbildung. Durch einen Summen- oder Mittelwert wird jedem Indikator das

gleiche Gewicht zugesprochen. Die Gleichwertigkeit der Indikatoren mag sich etwa aus der Theorie heraus begründen. Alternativ kann auch ein gewichteter Summenwert gebildet werden, bei dem entsprechend der Modellgleichung des formativen Ansatzes jeder Indikator mit einem spezifischen  $\beta$ -Gewicht versehen wird. Diese Gewichte können theoretisch begründet sein oder empirisch aus Regressionsanalysen (siehe oben), einfaktoriellen Hauptkomponentenanalysen gewonnen worden sein oder durch Expertinnen und Experten festgelegt werden.

Die bisherigen Schritte bezogen sich allesamt auf sog. „within-construct issues“, bei welchen das Konstrukt per se untersucht wurde. Eine weitere Validierung im Sinne von konvergenter oder diskriminanter Validität der Skala kann nun ggf. unter Zuhilfenahme von Fragestellungen vorgenommen werden, die man als „between-construct issues“ bezeichnen kann (Craven, Marsh, & Parada, 2013). Der Index kann als manifeste Variable bspw. in Pfadanalysen mit anderen manifesten Variablen Verbindung gebracht werden oder als manifeste Variable in Strukturgleichungsmodelle eingehen, in denen andere Konstrukte latent modelliert werden. Darüber hinaus bestehen auch Möglichkeiten, das formative Messmodell in Strukturgleichungsmodelle zu integrieren. Hierfür können je nach Höhe der Interkorrelation zwischen den Indikatoren und abhängig von der Fragestellung auch PLS-Modelle favorisiert werden (Weiber & Mühlhaus, 2014).

## 5 Fazit und Ausblick

Die Angemessenheit des formativen Modells wird bei der MARI-Methode nur im ersten Schritt untersucht, womit sie sich wesentlich vom Vorgehen nach Eberl (2004) unterscheidet. Es kann davon ausgegangen werden, dass eine quantitativ-empirische Prüfung der Angemessenheit durch die Konfundierung mit der Höhe von Interkorrelationen suboptimal ist. Daher scheint es sinnvoller, auf Basis von Expertenurteilen zu einer von der Subjektivität des einzelnen Anwenders/der einzelnen Anwenderin befreiten Validierung zu kommen und das Hauptaugenmerk auf die Güte der Modellierung zu legen.



Der Mehrwert der vorgeschlagenen Methode liegt in der Tatsache, dass die Indexbildung datengeleitet erfolgt. Damit wird die Forderung Borsbooms (2006) aufgegriffen, in der sozialwissenschaftlichen Forschung Fragestellungen unter Ausschöpfung des vorhandenen Methodenrepertoires zu beantworten, jedoch unter Verwendung von Standardmethoden. Zugleich wurde in Zusammenhang mit den bisherigen Schritten deutlich gemacht, dass Entscheidungen bei der Skalenbereinigung jedoch nicht allein aus statistischen Ergebnissen begründet sein sollen. Durch die Kombination von Expertenurteilen, sachlogischen Erwägungen der Anwender und statistischen Ergebnissen wird sowohl der Komplexität und Besonderheit formativer Konstrukte Rechnung getragen als auch sichergestellt, dass durch bewussten und reflektierten Umgang mit dieser Besonderheit dem Vorwurf unwissenschaftlichen Missbrauchs der formativen Methode vorgebeugt wird.

Mit diesem Beitrag soll ein Bewusstsein für das Vorhandensein einer alternativen Modellierungsart für latente Konstrukte geschaffen werden und den Lesern und Leserinnen zugleich eine Methode an die Hand gegeben werden, formative Modelle zu realisieren. Es scheint in der empirisch-pädagogischen Forschung eine Reihe an Konstrukten zu geben, für die formative Operationalisierungen denkbar und angemessen sind. Nicht zuletzt, weil das formative Modell dem Umstand Rechnung trägt, dass Theorien mitunter aus der Analyse der Empirie erst entstehen und sich konkretisieren.

#### Weiterführende Literatur:

Fluck, J. & Lichtenberg, H. (in Druck). *Formative Modellierung in Psychologie und Erziehungswissenschaft – Datengeleitete Indexbildung nach der MARI-Methode*. Wiesbaden: Springer VS Verlag.

Weiber, R., & Mühlhaus, D. (2014). *Strukturgleichungsmodellierung: Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS* (2. Auflage). Berlin: Springer.

## Literatur

Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2003). *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung* (10., neu bearb. und erw. Aufl ed.). Berlin: Springer.

Bagozzi, R. P. (1984). Expectancy-value attitude models an analysis of critical measurement issues. *International Journal of Research in Marketing*, 1(4), 295-310.

Bainter, S. A., & Bollen, K. A. (2015). Moving Forward in the Debate on Causal Indicators: Rejoinder to Comments. *Measurement: Interdisciplinary Research and Perspectives*, 13(1), 63-74. doi:10.1080/15366367.2015.1016349

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bollen, K. A., & Ting, K.-F. (1993). Confirmatory Tetrad Analysis. *Sociological Methodology*, 23, 147-175.

Bollen, K. A., & Ting, K.-F. (2000). A tetrad test for causal indicators. *Psychological Methods*, 5(1), 3-22. doi:10.1037//1082-989x.5.1.3

Borsboom D. (2006). The attack of the psychometricians. *Psychometrika*, 71(3), 425-440. doi:10.1007/s11336-006-1447-6

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological review*, 110(2), 203-219.

Brown, T. (2006). *Confirmatory factor analysis for applied research*. New York: The Guilford Press.

Bruhn, M., Lucco, A., & Wyss, S. (2008). Beendigung von Kundenbeziehungen aus Anbietersicht. *Marketing ZFP*, 30(4), 221-238.

Chin, W. W. (1998). The Partial Least Squares Approach for Structural Equation Modeling. In Marcoulides, G. A. (Hrsg.), *Modern Methods for Business Research* (S. 295-336). London.

Craven, R. G., Marsh, H. W., & Parada, R. H. (2013). Potent ways forward: new multidimensional theoretical structural models of cyberbullying, cyber targetization and bystander behaviors and their potential relations to traditional bullying constructs. In S. Bauman, D. Cross, & J. L. Walker (Eds.), *Principles of cyberbullying research* (pp. 68-86). New York, NY: Routledge.

Diamantopoulos, A. (2006). The error term in formative measurement models: interpretation and modeling implications. *Journal of modelling in management*, 1(1), 7-17.

Diamantopoulos, A., & Riefler, P. (2008). Formative Indikatoren: Einige Anmerkungen zu ihrer Art, Validität und Multikollinearität. *Zeitschrift für Betriebswirtschaft*, 78(11), 1183-1196.

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research*, 61(12), 1203-1218. doi:10.1016/j.jbusres.2008.01.009

Diamantopoulos, A., & Winklhofer, H. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research*, 38(2), 269-277.

Eberl, M. (2004). *Formative und reflektive Indikatoren im Forschungsprozess: Entscheidungsregeln und die Dominanz des reflektiven Modells*. Ludwig-Maximilians-Universität München: Schriften zur Empirischen Forschung und Quantitativen Unternehmensplanung.

Edwards, J. R. (2011). The fallacy of formative measurement. *Organizational Research Methods*, 14(2), 370-388.

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, 5(2), 155-174. doi:10.1037//1082-989x.5.2.155

Eid, M., Gollwitzer, M., & Schmitt, M. (2015). *Statistik und Forschungsmethoden: Lehrbuch. Mit Online-Material* (4., Originalausgabe, 4., überarbeitete und erweiterte Aufl ed.). Weinheim, Bergstr: Beltz, J.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah: Lawrence Erlbaum.

Festl, R. (2015). *Täter im Internet: Eine Analyse individueller und struktureller Erklärungsfaktoren von Cybermobbing im Schulkontext*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Fluck, J. (2020). *Cyberbullying – Theoretische und empirische Analysen zur Konstruktklärung und Messmodellierung eines Gewaltphänomens*. Hamburg: Dr. Kovac.

Fluck, J. & Lichtenberg, H. (in Druck). *Formative Modellierung in Psychologie und Erziehungswissenschaft – Datengeleitete Indexbildung nach der MARI-Methode*. Wiesbaden: Springer VS Verlag.

Gudergan, S. P., Ringle, C. M., Wende, S., & Will, A. (2008). Confirmatory tetrad analysis in PLS path modeling. *Journal of Business Research*, 61(12), 1238-1249.

Hauser, R. M. (1971). *Socioeconomic background and educational performance* (Vol. 2): American Sociological Association.

Hipp, J. R. (2008). *Performing Vanishing Tetrads Tests Using CTANEST1*. Verfügbar unter:  
[https://www.google.de/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCIQFjAAahUKEwiK0afTx4\\_IAhWmh3IKHZTKByS&url=https%3A%2F%2Fwebfiles.uci.edu%2Fhippj%2Fjohnhipp%2FCTANEST1\\_documentation.doc&usg=AFQjCNGAWdHQs-2y4\\_2Rs\\_kCPh-oZ2lR6g&sig2=FtmYFo79tGlQLiMuiaGqQQ&bvm=bv.103627116,d.bGQ&cad=rja](https://www.google.de/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCIQFjAAahUKEwiK0afTx4_IAhWmh3IKHZTKByS&url=https%3A%2F%2Fwebfiles.uci.edu%2Fhippj%2Fjohnhipp%2FCTANEST1_documentation.doc&usg=AFQjCNGAWdHQs-2y4_2Rs_kCPh-oZ2lR6g&sig2=FtmYFo79tGlQLiMuiaGqQQ&bvm=bv.103627116,d.bGQ&cad=rja) [28.10.2018].

Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model

misspecification in marketing and consumer research. *Journal of consumer research*, 30(2), 199-218.

Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, 70(351a), 631-639.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading: Addison-Wesley.

Markus, K. A. (2014). Unfinished Business in Clarifying Causal Measurement: Commentary on Bainter and Bollen. *Measurement: Interdisciplinary Research and Perspectives*, 12(4), 146-150. doi:10.1080/15366367.2014.980106

Marquardt, C. (2019). *Personalkompetenz Modellierung und Erfassung intrapersonaler Facetten beruflicher Handlungskompetenz*. Dissertationsschrift. Fachbereich 8. Universität Koblenz Landau.

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(5), 879-903.

Ringle, C. M., Wende, S. & Becker, J.-M. (2015). *SmartPLS 3*. Boenningstedt: SmartPLS GmbH. Verfügbar unter: <http://www.smartpls.com> [28.10.2018].

Rossiter, J. R. (2002). The C-OAR-SE procedure for scale development in marketing. *International journal of research in marketing*, 19(4), 305-335.

Savage, M. (2012). *Developing a Measure of Cyberbullying Perpetration and Victimization*. Tempe: Arizona State University.

Simonetto, A. (2012). Formative and reflective models: State of the art. *Electronic Journal of Applied Statistical Analysis*, 5(3), 452-457.

Smith, T. W. (2011). Measurement in health psychology research. In H. S. Friedman (Ed.), *The Oxford Handbook of Health Psychology* (pp. 42-72). New York: Oxford University Press.

Teimouri, M., Hassan, M. S., Griffiths, M., Benrazavi, S. R., Bolong, J., Daud, A., & Adzharuddin, N. A. (2015). Assessing the Validity of Western Measurement of Online Risks to Children in an Asian Context. *Child Indicators Research*. doi:10.1007/s12187-015-9316-4

Weiber, R., & Mühlhaus, D. (2014). *Strukturgleichungsmodellierung: Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS* (2. Auflage). Berlin: Springer.

West, S. G., & Grimm, K. J. (2014). Causal Indicator Models: Unresolved Issues of Construction and Evaluation. *Measurement: Interdisciplinary Research and Perspectives*, 12(4), 160-164. doi:10.1080/15366367.2014.980108

Wilcox, J. B., Howell, R. D. & Breivik, E. (2008). Questions about formative measurement. *Journal of Business Research*, 61(12), 1219-1228.