

## Formative Indikatoren: Einige Anmerkungen zu ihrer Art, Validität und Multikollinearität



Adamantios Diamantopoulos, Petra Riefler

### Überblick

- Ziel dieses Beitrags ist es, verbreitete Unklarheiten und Missverständnisse in Bezug auf die Beschaffenheit formativer Indikatoren und auf das Vorhandensein von Verfahren zur Prüfung ihrer Zuverlässigkeit zu klären.
- Im Besonderen wird dabei auf drei Kernthemen eingegangen, nämlich auf (a) die Spezifikation von Indikatoren als reflektiv oder formativ, (b) die Beurteilung der Validität von formativen Indikatoren, und (c) die Vorgehensweise bei Multikollinearität der Indikatoren. Es wird gehofft, die weitere Auseinandersetzung mit und Diskussion von reflektiven versus formativen Indikatoren zu stimulieren.

**Keywords** Formative measurement · multicollinearity · validity

Eingegangen: 19. Mai 2007

JEL: C10, M31

---

Prof. Dr. Adamantios Diamantopoulos (✉)

Lehrstuhl für Internationales Marketing, Universität Wien, Brünnerstr. 72, A-1210 Wien,

Email: [adamantios.diamantopoulos@univie.ac.at](mailto:adamantios.diamantopoulos@univie.ac.at), URL: [www.univie.ac.at/international-marketing](http://www.univie.ac.at/international-marketing)

Petra Riefler PhD (✉)

Lehrstuhl für Internationales Marketing, Universität Wien, Brünnerstr. 72, A-1210 Wien,

Email: [petra.riefler@univie.ac.at](mailto:petra.riefler@univie.ac.at), URL: [www.univie.ac.at/international-marketing](http://www.univie.ac.at/international-marketing)

## A. Einleitung

In einem kürzlich erschienenen Aufsatz beleuchteten Albers und Hildebrandt (2006) einige wichtige methodologische Aspekte der Verwendung von Strukturgleichungsmodellen in der empirischen Erfolgsfaktorenforschung. In diesem Bestreben galt ihr besonderes Augenmerk einerseits der Wahl eines adäquaten Operationalisierungsansatzes von Erfolgsfaktoren, das heißt der Wahl zwischen reflektiven (Effekt-) und formativen (kausalen) Indikatoren, sowie andererseits den Konsequenzen der Eliminierung formativer Indikatoren mittels konventioneller Bereinigungsverfahren (dem „Cronbach's  $\alpha$  – LISREL“ Paradigma). Ihre Analyse wurde durch eine Simulationsstudie gestützt, in welcher die Folgen einer Fehlspezifikation des Messmodells bei Verwendung von kovarianzstrukturbasierter (LISREL) und „soft-modelling“ (PLS) Verfahren untersucht wurden.

Das Ziel dieser Notiz ist es, einige Standpunkte von Albers und Hildebrandt (2006) aufzugreifen, die potentiell kontrovers erscheinen, und dadurch die weitere Auseinandersetzung mit und Diskussion von reflektiven versus formativen Indikatoren zu stimulieren. Zu Beginn soll ausdrücklich betont werden, dass diese Notiz *nicht* die Intention hat, die Vorgehensweise oder die Schlussfolgerungen der Autoren zu kritisieren. Wir stimmen den Autoren bezüglich der Vorgehensweise in der Simulationsstudie, als auch den darauf aufbauenden Schlussfolgerungen vollinhaltlich zu und schätzen ihre Forschungsanstrengungen. Ziel unseres Beitrags ist es vielmehr, den Beitrag von Albers und Hildebrandt (2006) zum Anlass zu nehmen, bestimmte Ansichten herauszugreifen und zu diskutieren, die mitunter auch von anderen Autoren geteilt werden, um so verbreitete Unklarheiten und Missverständnisse in Bezug auf die Beschaffenheit formativer Indikatoren und auf das Vorhandensein von Verfahren zur Prüfung ihrer Zuverlässigkeit zu klären. Im Besonderen wird dabei auf drei Kernthemen eingegangen, nämlich auf (a) die Spezifikation von Indikatoren als reflektiv oder formativ, (b) die Beurteilung der Validität von formativen Indikatoren, und (c) die Vorgehensweise bei Multikollinearität der Indikatoren. Unsere Diskussion dieser Themen bezieht sich bewusst nicht auf den Kontext der Erfolgsfaktorenmessung, den Albers und Hildebrandt (2006) für ihre Diskussion wählen, sondern ist angesichts der Relevanz der Themen für andere Forschungsrichtungen über die Erfolgsfaktorenforschung hinaus allgemein gehalten. Im Sinne der Konsistenz mit Albers und Hildebrandts Artikel wählen wir den Großteil der exemplarischen Beispiele dennoch aus dem Bereich der Erfolgsfaktorenmessung.

## B. Die Spezifizierung von Indikatoren als reflektiv oder formativ

Die möglicherweise kontroverseste Aussage von Albers und Hildebrandt (2006) in ihrer Diskussion zur Wahl eines adäquaten Messmodells für die Operationalisierung von Erfolgsfaktoren ist die folgende: „Man kann also entgegen der Ansicht von Jarvis, MacKenzie und Podsakoff (2003, S. 13) *letztendlich nicht prüfen, ob ein Ansatz der Konstrukt-Operationalisierung falsch ist*“ (Albers/Hildebrandt, 2006, S. 13, Hervorhebung hinzugefügt). Diese Argumentationslinie ist insofern problematisch, als dass sie (möglicherweise unbeabsichtigt) den Eindruck vermittelt, die „auxiliary theory“ (Blalock, 1969; Costner, 1969) eines Forschers, welche „the nature and direction of the relationship between cons-

tructs and measures“ (Edwards/Bagozzi, 2000, S. 156) spezifiziert, könnte nicht getestet werden. Die methodologische Literatur bietet jedoch eine Reihe von Hilfestellungen für den Forscher, um eine Entscheidung hinsichtlich der reflektiven oder der formativen Modellierung eines Indikators (oder eines Bündels von Indikatoren) zu treffen. Bollen (1989, S. 65-66) stellt zum Beispiel fest, dass „[t]emporal priority may be one means of establishing causal priority for measurement models“. Edwards und Bagozzi (2000, S. 156) verwenden währenddessen das Prinzip der Kausalität für die Formulierung von „guidelines for specifying the relationship between a construct and a measure in terms of (a) direction (i.e., whether a construct causes or is caused by its measures), and (b) structure (i.e., whether the relationship is direct, indirect, spurious, or unanalyzed“. Empirische Ansätze zur Bestimmung der Natur von Indikatoren wurden ebenfalls vorgeschlagen, diese umfassen Experimente (siehe bspw. Cermak, 1983), umfragebasierte Verfahren (siehe bspw. Fayers et al., 1997) und den bekannten TETRAD Test (Bollen/Ting, 2000). Eberl (2006) beschreibt eine strukturierte Vorgehensweise zur Prüfung einer Spezifikationshypothese, die, basierend auf der „auxiliary theory“ des Forschers, die Kausalität im Messmodell spezifiziert. Während eine Diskussion der Vorzüge und Einschränkungen dieser Ansätze über den Rahmen dieser Notiz hinausgehen würde, soll unterstrichen werden, dass (mehrere) Wege zur Beurteilung, ob ein Indikator formativ oder reflektiv spezifiziert werden soll, bestehen, die über die Liste an (teilweise überlappenden) Kriterien, die von Jarvis, Mackenzie und Podsakoff (2003; siehe auch Mackenzie/Podsakoff/Jarvis, 2005, S. 713)<sup>1</sup> vorgeschlagen wurde, hinausgehen.

Albers und Hildebrandt (2006, S. 13) machen in diesem Zusammenhang eine weitere zu diskutierende Aussage, nämlich folgende: „Bei Annahme reflektiver statt richtigerweise formativer Indikatoren hätte man also nicht unbedingt ein falsches, allerdings ein stark eingeschränktes Modell, nämlich einen engeren Bedeutungsinhalt für das Konstrukt“. Dieses Argument basiert offenbar auf der Annahme, dass reflektive Indikatoren einen engeren Umfang der Konstruktbedeutung erfassen als formative Indikatoren, denn in einem späteren Absatz wird konstatiert, dass „reflektiv gemessene Konstrukte [können] aber wegen der Forderung nach interkorrelierten und damit intern konsistenten Indikatoren immer nur eine Facette messen, während mit formativen Indikatoren ausdrücklich mehrere unterschiedliche Facetten eines abstrakten Sachverhaltes (Konstrukts) erfasst werden sollen, die dann das Konstrukt definieren“ (Albers/Hildebrandt, 2006, S. 25).

Dieser Ansicht der Autoren ist aus zwei Gründen zu widersprechen, da (a) die Frage, ob Indikatoren korrelieren, empirischer Natur ist, während die Kausalität und der Bedeutungsinhalt, der durch die Indikatoren (ob reflektiv oder formativ) gemessen wird, theoretische Fragen darstellen; und (b) die Frage, ob Elemente korrelieren, weil sie gemeinsam verändert werden, keine Auskunft über ihre Kausalität zu der latenten Variable gibt. Die entscheidende Frage ist vielmehr, wie man die theoretische Kausalität korrekt bestimmen kann, um einer Fehlspezifikation der „auxiliary theory“ (Blalock, 1969; Costner, 1969) entgegen zu wirken. Zu der Frage, wie viele Facetten durch Indikatoren erfasst werden, ist festzuhalten, dass die alleinige Tatsache, dass ein Bündel an Indikatoren hoch interkorreliert und intern konsistent ist, *nicht* notwendigerweise bedeutet, dass eine einzige Facette (oder ein eindimensionales Konstrukt) erfasst wird, was also weiter bedeutet, dass der Bedeutungsinhalt nicht per Definition enger werden muss. Auch ein Bündel an reflektiven Indikatoren, das intern konsistent ist (und somit hohe Interkorrelationen zei-

gen), vermag multiple (möglicherweise in Beziehung stehende) Facetten zu erfassen. Mögliche Beispiele in diesem Zusammenhang wären Indikatoren, welche die Ausgaben für unterschiedliche Elemente des Marketing-Mixes (wie Werbung, Verkaufsförderung oder Sponsoring), für die das Jahresbudget gemeinschaftlich festgelegt (d.h. zeitgleich erhöht oder gesenkt) wird, erfassen; oder auch Indikatoren, die verschiedene Facetten der fünf weitgefassten Persönlichkeitsmerkmale messen (Digman, 1990). Zudem ist es nicht zwingend, dass hohe Interkorrelationen ausschließlich bei reflektiven Indikatoren beobachtet werden. In ähnlicher Weise kann eine formative Indikatorengruppe hohe Interkorrelationen aufweisen, obwohl die individuellen Indikatoren klar unterschiedliche Facetten erfassen; Law und Wong (1999, S. 143) weisen in diesem Zusammenhang darauf hin, dass „having highly correlated causal indicators is a very common phenomenon for some composite variables“ und nennen das Beispiel der „Job Perception“, das sich aus den Facetten Skill Variety, Skill Significance, Task Autonomy, Identification und Feedback zusammensetzt.

Ein letztes Argument von Albers und Hildebrandt (2006, S. 33), das wir in diesem Zusammenhang aufgreifen möchten, ist das folgende: „Unterstellt man Reflektivität, auch wenn sich das Konstrukt aus unabhängigen gestalterischen Maßnahmen zusammensetzt, so könnte man dies damit verteidigen, dass man das Konstrukt als ganzheitliche Strategie [...] begreift. Versteht man die Strategie dann so, dass bei einer Veränderung der Strategieausprägung auch immer gleich alle Indikatoren zu diesem Konstrukt gleichermaßen verändert werden, so ist die Richtung der Kausalität zwischen Konstrukt und Indikatoren nicht bedeutsam. Allerdings ist die Aussagekraft des Modells eingeschränkt, da man nur die Auswirkungen von Strategie-Ausprägungen untersuchen kann [...]“. Wir stimmen den Autoren zu, dass es selbstverständlich Fälle geben mag, in welchen verschiedene Elemente (wie etwa die Ausgaben für Marketingmaßnahmen im Beispiel der Erfolgsfaktorenmessung) zeitgleich verändert werden. Dennoch möchten wir mit Nachdruck darauf hinweisen, dass unter einer (fälschlichen wie auch korrekten) reflektiven Spezifikation von Erfolgsfaktoren *niemals* eine Aussage über den Erfolgsbeitrag von Strategiebündel generiert wird, da dieses so genannte Strategiebündel nicht als *Ursache* (sondern als Folge) der Erfolgsvariable modelliert ist.

Zusammenfassend kann gesagt werden, dass die Beobachtung, ob Indikatoren hoch interkorreliert sind, keine eindeutigen Schlussfolgerungen darüber zulässt, ob diese Indikatoren eine einzelne Facette oder mehrere Facetten eines Konstruktes erfassen. Die vorangegangene Diskussion impliziert außerdem, dass in empirischen Anwendungen durchaus die Möglichkeit besteht, reflektive Indikatoren zu finden, die nicht hoch interkorreliert sind (da manche der Indikatoren „schlechte“ Items sind, die bspw. sehr geringe Varianzen aufweisen und nicht uni-dimensional sind) und formative Indikatoren, die eben dies tun. Somit deutet das Vorhandensein hoher Interkorrelationen nicht zwingend auf das Vorliegen reflektiver Indikatoren hin, und daher können die Konsequenzen eines Type I errors (d.h. einer inkorrekten reflektiven Spezifikation eines formativen Messmodells, siehe Diamantopoulos und Siguaw 2006) nicht simpel als das Hervorrufen „eine[s] engen Bedeutungsinhalt für das Konstrukt“ angesehen werden, wie von Albers und Hildebrandt (2006, S. 13) argumentiert. Letzteres *kann* geschehen, *muss* aber nicht geschehen. Vielmehr sollte dem Forscher bewusst sein, dass eine fälschliche reflektive Spezifikation missliche Folgen sowohl auf theoretischer als auch auf empirischer Ebene nach sich zieht. Auf theoretischer Ebene ist dies, wie oben dargestellt, die Fehlspezifikation der zugrunde

liegenden „auxiliary theory“; auf empirischer Ebene betrifft es die falsche Bereinigung von Indikatoren. Während wir Albers und Hildebrandt (2006) uneingeschränkt zustimmen, dass die empirischen Folgen schwerwiegend sind, möchten wir betonen, dass die theoretische Missspezifikation nicht minder wiegt. So kann eine falsche Spezifikation im Idealfall keinerlei Folgen auf der empirischen Ebene mit sich bringen, nämlich dann, wenn die Indikatoren hoch korreliert sind; dennoch würde die Fehlspezifikation auf theoretischer Ebene weiterhin vorliegen. Schließlich wollen wir uns Albers und Hildebrandt (2006) anschließen, dass den weiteren unerwünschten Folgen einer solchen Fehlspezifikation besondere Beachtung geschenkt werden soll, nämlich der verzerrten Parameterschätzung, den möglichen inkorrekten Schlussfolgerungen und der problematischen Modellgüte, wie von Albers und Hildebrandt in ihrer Simulationsstudie und andere Autoren in ähnlichen Studien eindrücklich belegt wird (siehe MacCallum/Browne, 1993; Law/Wong, 1999; Jarvis/MacKenzie/Podsakoff, 2003; MacKenzie/Podsakoff/Jarvis, 2005). Es ist daher von besonderer Wichtigkeit, Probleme dieser Art zu vermeiden, indem „the researcher should decide in advance which are effect- and which are cause-indicators“ (Bollen, 1984, S. 383).

### C. Die Beurteilung der Validität formativer Indikatoren

Albers und Hildebrandt (2006, S. 13) bringen ihre Ansicht zum Ausdruck, dass „man die Validität von Konstrukten mit formativen Indikatoren nie mit Hilfe irgendeiner Kennzahl beurteilen [kann], sondern die Indikatoren beeinflussen lediglich den Bedeutungsinhalt eines Konstrukts und damit den Aussagegehalt von Ergebnissen“. Bedauerlicherweise erläutern die Autoren weder die Argumente, die dieser Ansicht zugrunde liegen, noch belegen sie diese mit Literatur. Es ist jedenfalls anzumerken, dass die Autoren mit dieser – in unseren Augen nicht korrekten – Auffassung nicht alleine sind. So halten beispielsweise auch Scholderer und Balderjahn (2006, S. 65) fest, dass „aufgrund des definitiven Charakters des formativen Messmodells stellt sich auch nicht die Frage der Validierung“, während Rossiter (2002, S. 315, Hervorhebung im Original) der Ansicht ist, dass „all that is needed is a set of distinct components as decided by expert judgment ... once decided, these components must *all* be present in the scale because the items representing them are the *defining items* for the attribute ... items cannot be added or deleted from the scale“. In ähnlicher Weise meinen Homburg und Klarmann (2006, S. 731), “[d]as vollständige Fehlen von quantifizierbaren Qualitätsstandards im Falle formativer Messmodelle führt jedoch dazu, dass hier das Risiko theoretischer Beliebigkeit höher ist als bei der Verwendung reflektiver Messmodelle. Letztlich kann der Anwender bei der Spezifikation eines formativen Messmodells beliebig viele Variablen aggregieren, ohne dass eine Warnung erfolgt, wenn sein Vorgehen nicht sinnvoll ist.“

Die oben angeführten Auffassungen resultieren aus einem Missverständnis der zugrunde liegenden Charakteristika des formativen Messmodells. Dieses Missverständnis wird offensichtlich verursacht durch die eher unpräzisen und mehrdeutigen Definitionen für formative Indikatoren, die sich in der Literatur finden. So liegt zum Beispiel für Bagozzi und Fornell (1982, S. 34, Hervorhebung im Original) ein formatives Messmodell vor, wenn „a concept is assumed to be defined by, *or* to be a function of, its measure-

ments“; und Bagozzi (1994, S. 332, Hervorhebung im Original) meint in ähnlicher Weise „when a latent variable is defined as a linear sum of a set of measurements *or* when a set of measures of a dependent variable is determined by a linear combination of independent variables, the measures are termed formative indicators“.

Eine genaue Betrachtung der oben angeführten Definitionen lässt *zwei* unterschiedliche Perspektiven (siehe dazu auch Bollen, 2007) der Natur formativer Indikatoren erkennen (die bedauerlicherweise häufig miteinander vermischt werden, wie die oben angeführten Definitionen deutlich zeigen). Unter der ersten Perspektive wird das fokale Konstrukt einer Kombination von Indikatoren gleichgesetzt; diese Perspektive besitzt axiomatischen Charakter und lässt sich auf das „operational definition model“ (siehe Hempel, 1960) zurückführen, nach welchem „a concept becomes its measure and has no meaning beyond that measure ... [T]he entire meaning of a theoretical concept is assigned to its measurement and any theoretical concept has one and only one measurement“ (Bagozzi, 1982, S. 15). So könnte ein Forscher, in Anlehnung an das von Albers und Hildebrandt (2006) gewählte Beispiel, die Gesamtzufriedenheit mit einem Hotelaufenthalt ( $\eta$ ) als eine einfache lineare Summe aus der Zufriedenheit mit der Zimmerausstattung ( $x_1$ ), der Zufriedenheit mit der Gestaltung des Wellness-Bereiches ( $x_2$ ) und der Zufriedenheit mit dem Hotelpersonal ( $x_3$ ) *definieren*; wobei die Wahl von  $x_1 - x_3$  auf einer Literaturrecherche und/oder Expertenbewertungen – wie von Rossiter (2002) empfohlen – basieren könnte. In diesem Fall gilt also  $\eta = x_1 + x_2 + x_3$  (1) und die Frage der Indikatorenvalidität stellt sich angesichts der axiomatischen Natur von (1) nicht; die Indikatoren sind *per Definition* valide, weil  $\eta$  nur diese (und keine weiteren) Indikatoren umfasst. Es gilt in diesem Zusammenhang zu beachten, dass bei der Definition von  $\eta$  nicht alle Indikatoren zwingend gleich zu gewichten sind; es mag etwa der Fall sein, dass Experten die Zimmerausstattung für doppelt so bedeutsam für die Zufriedenheit mit dem Hotelaufenthalt ansehen als die beiden anderen Attribute, womit letzteres nun als  $\eta = 2x_1 + x_2 + x_3$  (2) definiert wäre. Generell kann jegliche Kombination an Gewichten ( $w_i$ ) den Indikatoren  $x_i$  auf einer *ex-ante* Basis zugeteilt werden, um ihren unterschiedlichen Beitrag zur Zusammensetzung von  $\eta$  zu berücksichtigen; eine generelle Spezifikation dieser „definitorischen“ Perspektive formativer Indikatoren wäre daher:

$$(3) \quad \eta = w_1x_1 + w_2x_2 + \dots + w_nx_n$$

wobei  $n$  = Anzahl an Indikatoren und  $w_i$  ( $i = 1, 2, \dots, n$ ) eine *vordefinierte* Gewichtung des Indikators  $x_i$  ist.

An dieser Stelle sei darauf hingewiesen, dass die oben angeführte Gleichung (3) eine *mathematische Identität* und kein *Messmodell*, das *unbekannte* Parameter enthält (die anschließend anhand empirischer Daten *geschätzt* werden), repräsentiert. Anders gesagt kann Gleichung (3) nicht durch *empirische* Überprüfung widerlegt und somit nicht als Repräsentierung einer *Messtheorie* angesehen werden. Falls eine überprüfbare „auxiliary theory“ (Blalock, 1969; Costner, 1969) postuliert werden soll, ist deshalb die definitorische Perspektive gemäß Gleichung (3) nicht adäquat. In diesem Fall ist hingegen eine Spezifikation nach dem zweiten Konzept formativer Messmodelle passend, welches beschrieben wird als

$$(4) \quad \eta = \gamma_1x_1 + \gamma_2x_2 + \dots + \gamma_nx_n + \zeta$$

wobei  $\gamma_i$  die zu den Indikatoren  $x_i$  gehörenden *unbekannten* Koeffizienten repräsentieren,  $\zeta$  sämtliche unberücksichtigte Gründe für  $\eta$  umfasst, die zusätzlich zu  $x_i$  ( $i = 1, 2, \dots, n$ ) existieren, und die Annahmen, dass  $\text{CORR}(x_i, \zeta) = 0$  (für alle  $i$ ) und  $E(\zeta) = 0$ , gelten. Unter dieser Spezifikation (siehe Bollen/Lennox, 1991 für weitere Details) ist die latente Variable  $\eta$  durch ihre Indikatoren nicht definiert, sondern (unvollständig) *determiniert*. Der spezifische Einfluss  $\gamma_i$  der Indikatoren auf die latente Variable ist hier a priori *nicht* bekannt, das gleiche gilt für den Anteil an nicht-erklärter Varianz in  $\eta$ . Greift man hier wieder das Beispiel der Zufriedenheit mit einem Hotelaufenthalt auf, würde nach dieser Perspektive die Gesamtzufriedenheit durch die Zufriedenheit mit der Zimmerausstattung ( $x_1$ ), der Gestaltung des Wellness-Bereiches ( $x_2$ ) und dem Hotelpersonal ( $x_3$ ) und mögliche weitere Faktoren, die nicht explizit modelliert sind ( $\zeta$ ), determiniert. Des Weiteren würden die Richtung und Stärke der individuellen Einflüsse von  $x_1 - x_3$  auf die Gesamtzufriedenheit empirisch bestimmt, das gleiche gilt für ihren kumulierten Beitrag zur Varianzerklärung der Gesamtzufriedenheit.

Wir vermuten, dass eine Verwechslung der Gleichungen (3) und (4) Albers und Hildebrandt (2006) – sowie andere Autoren – zu der Ansicht führt, dass die Validität formativer Indikatoren nicht empirisch beurteilt werden kann. Während dies unbestritten in Gleichung (3) der Fall ist, trifft das für formative Modelle, die gemäß Gleichung (4) konzeptionalisiert sind, eindeutig nicht zu. Für letztere gilt vielmehr „item validity is reflected in the significance and strength of the path from the indicator to the composite latent construct“ (MacKenzie/Podsakoff/Jarvis, 2005, S. 727). Diese Ansicht ist konsistent mit der Definition von Validität im Kontext von Strukturgleichungsmodellen, wo diese definiert wird als „the strength of the direct structural relationship between a measure and a latent variable“ (Bollen, 1989, S. 222). Präziser formuliert bedeutet dies, dass für einen formativen Indikator  $x_i$  „the unstandardized and standardized validity coefficients show the direct effect of  $x_i$  on the latent variable(s) it influences. The unique validity variance reveals the proportion of the variance in the latent variable that is uniquely attributable to the cause indicator“ (Bollen 1989, S. 222). Indem geprüft wird, in welchem Ausmaß sich die  $\gamma$ -Koeffizienten in Gleichung (4) signifikant von Null unterscheiden, lässt sich somit die Validität der Indikatoren belegen; wie Diamantopoulos und Winklhofer (2001, S. 273) klarstellen, „if the values of these [coefficients] are zero in the population, then arguably the indicators cannot be considered valid measures of the construct (and a non-significant t-statistic for  $\eta$  fails to reject the zero value hypothesis)“. In diesem Zusammenhang ist zusätzlich anzumerken, dass sich der Forscher im Fall, dass sich Indikatoren als nicht signifikant von Null unterschiedlich zeigen, mit der Entscheidung konfrontiert sieht, ob diese Indikatoren endgültig eliminiert werden sollen. Um dieser Unsicherheit zu begegnen, empfehlen wir, die Elimination nicht anhand eines einzelnen Datensatzes vorzunehmen, sondern die Elimination durch Replikationsstudien abzusichern. Zeigt sich ein Parameter über Replikationsstudien hinweg als nicht signifikant, sollte der entsprechende Indikator gemäß der Empfehlung von Diamantopoulos/Winklhofer (2001) eliminiert werden. Diese Empfehlung von Replikationsstudien wird in Anlehnung an Bollen (2007, S. 222) getan, der kürzlich im Kontext von instabilen Strukturparameterschätzungen folgendes vorschlägt: „if the coefficients ( $\gamma$ s) change depending on the outcome, this is evidence of a structurally misspecified model, and we should question the validity of the model and the appropriateness of the latent variable“.

Zur Rolle von MIMIC Modellen in der Validitätsbeurteilung – die Albers und Hildebrandt (2006, S. 25) ebenfalls kritisieren – ist vorauszuschicken, dass der vorrangige Beweggrund für die Verwendung eines MIMIC Modells die Identifikation des formativen Messmodells ist. Wie gemeinhin bekannt, ist das formative Modell gemäß Gleichung (4) statistisch unteridentifiziert (Bollen/Lennox, 1991), wodurch eine eindeutige Schätzung der Modellparameter nicht möglich ist. Aus diesem Grund müssen dem Modell *zusätzliche* Informationen zugeführt werden, um so die Schätzung der  $\gamma$ -Koeffizienten sowie der Varianz des Störterms  $\zeta$  zu ermöglichen. Wenngleich verschiedene Optionen zur Identifikation von formativen Modellen bestehen<sup>2</sup>, wird folgendes empfohlen: „the best option for resolving the identification problem ... is to add two reflective indicators to the formative construct when conceptually appropriate“ (Jarvis/MacKenzie/Podsakoff, 2003, S. 213). Durch dieses Vorgehen wird das Modell in Gleichung (4) zu einem MIMIC Modell (Jöreskog/Goldberger, 1975), wobei  $\eta$  „as a single latent construct with a mixture of formative and reflective indicators rather than a single reflective-indicator latent construct with multiple causes“ (MacKenzie/Podsakoff/Jarvis, 2005, S. 727) anzusehen ist.<sup>3</sup> Um eine solche Interpretation des MIMIC Modells sicherzustellen, sollten als reflektive Indikatoren *globale* Bewertungen von  $\eta$  gewählt werden, die von den proximalen Antezedenzen von  $\eta$ , also den formativen Indikatoren  $x_i$ , abhängen. In dem Beispiel der Hotelzufriedenheit von Albers und Hildebrandt (2006, S. 33) würden somit die individuellen Aspekte (Zimmerausstattung, Wellness-Bereich, Personal etc.) durch eine Reihe formativer Indikatoren erfasst, während die reflektiven Indikatoren als globale Items (z.B. „Gesamt gesehen bin ich mit dem Aufenthalt in diesem Hotel zufrieden“ oder „Dieses Hotel schätze ich sehr“) formuliert wären. Entgegen der Ansicht von Albers und Hildebrandt (2006) erfassen die gewählten Indikatoren nicht einen einzelnen spezifischen Aspekt des operationalisierten Konstruktes, sondern sind *beabsichtigt* weitgefasst formuliert, sodass sie *Gesamtbewertungen* des Konstruktes repräsentieren (siehe dazu Jarvis/MacKenzie/Podsakoff, 2003; Diamantopoulos/Siguaw, 2006). In diesem Zusammenhang wurde bereits darauf aufmerksam gemacht, dass „the selection of „external“ (i.e. additional) variables necessary for achieving identification is just as crucial in a formative measurement model as is the selection of the formative indicators themselves“ (Diamantopoulos, 2006, S. 15).

In ihrer Kritik an der Nützlichkeit von MIMIC Modellen zur Validierung behaupten Albers und Hildebrandt weiters: „Damit wollen Diamantopoulos/Winklhofer (2001) Indikatoren eliminieren, die nicht einem semantischen Verständnis der Konstruktes entsprechen“ (S. 25, Jahreszahl der zitierten Quelle aus dem Original korrigiert). Wir möchten an dieser Stelle festhalten, dass MIMIC Modelle – neben ihrer Nützlichkeit der Identifikation von formativen Modellen – für externe Validitätsprüfungen dienen, und nicht zur Indikatorelimination basierend auf semantischer Ähnlichkeit der Indikatoren. Die Interpretation von Albers und Hildebrandt überrascht, da Diamantopoulos und Winklhofer (2001 S. 272) in keinem Wort Bezug auf den semantischen Inhalt der Indikatoren nehmen, sondern stattdessen sagen: „If the overall model fit proves acceptable, this can be taken as supporting evidence for a set of indicators forming the index. In addition, by focusing on the  $\gamma$  parameters, researchers can assess the contribution and significance of the individual indicator“, was in weiterer Folge bedeutet, wie bereits oben zitiert, dass nicht-signifikante Parameter auf nicht-valide Indikatoren hinweisen. Somit ist der Grund der Elimi-



nation keine Bereinigung im Sinne von semantischer Kompatibilität, sondern Folge der negativen Validitätsprüfung.

Zusammenfassend ist festzuhalten, dass es angesichts der oben genannten Gründe sowohl möglich als auch notwendig ist, die Validität formativer Indikatoren zu beurteilen. Diese Validierung muss dabei über die Bewertung der Inhaltsvalidität, wie sie von Rossiter (2002) gefordert wird, hinausgehen. Wir unterstützen deshalb Edwards und Bagozzis (2000, S. 171) Sichtweise, nämlich „if measures are specified as formative, their validity must still be established. It is bad practice to ... claim that one's measures are formative, and do nothing more“.

#### D. Formative Indikatoren und Multikollinearität

Wenngleich das formative Messmodell in Gleichung (4) keine Bedingungen an die Korrelationen zwischen den Indikatoren stellt (Bollen, 1984; Bollen/Lennox, 1991)<sup>4</sup>, „muss darauf hingewiesen werden, dass unter besonderen Rahmenbedingungen ein hochgradig korreliertes Verhalten der Indikatoren eines formativen Messmodells nicht nur möglich, sondern auch zu erwarten ist“ (Fassott, 2006, S. 58). Da in formativen Modellen die latente Variable durch eine *multiple* Regressionsgleichung mit ihren Indikatoren verbunden wird, kann dies zu einem Multikollinearitätsproblem führen, was es „difficult to separate the influence of the individual xs on the latent variable  $\eta$ “ (Diamantopoulos/Winklhofer, 2001, S. 272)<sup>5</sup> macht.

Bedauerlicherweise besteht in der Literatur kein Konsens darüber, wie mit Multikollinearität in formativen Modellen vorzugsweise umgegangen werden soll. Manche Autoren (z.B. Diamantopoulos/Siguaw, 2006) rechtfertigen die Elimination von hoch interkorrelierten Indikatoren mit der Begründung, dass diese sehr wahrscheinlich redundante Information enthalten. Andere Autoren (z.B. Rossiter, 2002) lehnen eine Elimination mit dem Argument ab, dass dieses Vorgehen die Natur des Konstruktes verändern könnte, schlagen jedoch keine alternativen Lösungswege vor. Schließlich gibt es jene Autoren (z.B. Law/Wong, 1999), die vorschlagen, ausschließlich Indikatoren zu verwenden, von denen aus vorangegangenen Studien bekannt ist, dass sie mit der latenten Variable signifikant korrelieren; dieser Vorschlag ist selbstredend für die Entwicklung neuer Skalen und bei Fehlen vorangegangener Forschung wenig hilfreich.

Albers und Hildebrandt (2006, S. 13) schlagen einen anderen Lösungsweg bei vorliegender Multikollinearität vor und empfehlen „Indikatoren zu einem Index zusammenzufassen und dann den Index als Single-Item-Konstrukt in die Analyse einfließen zu lassen“. Konkret regen sie an, im Falle von (kollinearen) Indikatoren, die kompensatorische Effekte darstellen, das arithmetische Mittel (nach Normalisierung) zu bilden und dieses für die Analyse zu verwenden. Im Falle von nicht-kompensatorischen Indikatoren soll hingegen das geometrische Mittel (nach Normalisierung) verwendet werden. Auch wenn Albers und Hildebrandt (2006) keine detaillierte Erklärung geben, *weshalb* dieser Ansatz besser ist als andere, stellt das Zusammenfassen kollinearere Indikatoren eine von mehreren möglichen Optionen dar, die in der Literatur für den Umgang mit Multikollinearität in multiplen Regressionen vorgeschlagen werden. Berry und Feldman (1985, S. 48) erklären zum Beispiel, „one of these options is to combine two or more independent variables that are highly correlated into a single variable – such as a weighted or unweighted

average of the original variables – and then use the composite variable in place of the correlated variables in the regression.“ Dieser Lösungsansatz ist jedoch mit zwei potenziellen Schwierigkeiten behaftet.

Erstens muss es *konzeptionell* Sinn machen, die Indikatoren wie oben beschrieben zusammenzufassen, was bedeutet, dass die Indikatoren die *gleiche* Facette des Konstruktes erfassen müssen. Ist dies nicht der Fall, würde die Interpretation des resultierenden zusammengefassten Indikators (bzw. „Index“ in der Terminologie von Albers und Hildebrandt (2006)) unklar. Um ein Beispiel zu geben: „it is not sensible to combine the income and race variables, even if they are highly related“ (Lewis-Beck, 1980, S. 61). Bezogen auf das Modell in Gleichung (4) bedeutet dies, dass sich nur jene kollinearen  $x_i$ s potentiell zusammenfassen lassen, welche die gleiche Facette von  $\eta$  erfassen. Falls hingegen jedes  $x_i$  eine *unterschiedliche* Facette erfasst (ein Umstand, der beispielsweise eintreten kann, wenn Rossiter's (2002) C-OAR-SE Ablauf für die Item-Pool-Generierung für formative Attribute angewandt wurde), ist die Zusammenfassung von Indikatoren für die Reduktion von Multikollinearität keine Option.

Zweitens basiert die Zusammenfassung von Variablen auf der Annahme, dass die Ursache der Multikollinearität *bivariate* Korrelationen zwischen den Indikatoren sind. Es ist jedoch gemeinhin bekannt, dass „lack of any high correlation values,..., does not ensure a lack of collinearity“ (Hair et al., 1998, S. 191) und „it is possible,..., to find no large bivariate correlations, although one of the independent variables is a nearly perfect linear combination of the remaining independent variables“ (Lewis-Beck, 1980, S. 60). Unter solchen Umständen ist das Zusammenfassen von Indikatoren zu einem Index, wie von Albers und Hildebrandt (2006) befürwortet, keine realisierbare Option.

In Anbetracht der obigen Ausführungen stellt sich die Frage, wie dem Problem der Multikollinearität im formativen Messmodell begegnet werden soll. Ein guter Ausgangspunkt ist allen voran die Feststellung, dass die Effekte einer (auch starken) Multikollinearität größtenteils aufgehoben werden können, indem „the analysis is based on a large sample, a model specification that results in a high  $R^2$ , or a combination of the two“ (Mason/Perreault, 1991, S. 275). Zur Stichprobengröße sei angemerkt, dass angesichts der hohen Zahl an zu schätzenden Parametern in formativen Messmodellen die Verwendung großer Stichproben in jedem Fall sinnvoll ist, um ein adäquates Parameter-Fälle Verhältnis zu erreichen (Bentler/Chou, 1987); da „a total of  $k(k-1)/2$  covariance terms and one residual variance term have to be estimated for a composite variable with  $k$  causal indicators. In contrast, only  $k$  factor loadings and  $k$  error variances need to be estimated for a latent variable with  $k$  effect indicators“ (Law/Wong, 1999, S. 157). Zum  $R^2$  ist in ähnlicher Weise anzumerken, dass die formativen Indikatoren in einem Modell idealerweise einen erheblichen Teil der Varianz in der latenten Variable erklären, während der Anteil an nicht-erklärter Varianz gering ist (für eine detaillierte Darstellung siehe Diamantopoulos, 2006). Dies lässt sich wie folgt begründen: „as the amount of variance of the residual increases, the meaning of the construct becomes progressively ambiguous ... Certainly, the meaning of a construct is elusive when most of its variance is attributable to unknown factors“ (Williams/Edwards/Vandenberg, 2003, S. 908). Die sorgfältige Wahl der Indikatoren, die eine umfängliche Erfassung der konzeptionellen Konstruktbedeutung gewährleistet (siehe Diamantopoulos/Winklhofer, 2001 für Details), sollte zur Erhöhung der Erklärungskraft des Modells beitragen.

Während das Augenmerk auf die Stichprobengröße und die Indikatorenspezifikation das Multikollinearitätsproblem zum Teil aufheben könnte, mag es Fälle geben, in welchen die Elimination von Indikatoren die einzige reelle Option zur Behebung der Schwierigkeiten darstellt. Unsere persönliche Position zu dieser Angelegenheit ist von Bollen und Lennox (1991, S. 308, Hervorhebung hinzugefügt) geleitet, die argumentieren, dass „excluding an indicator changes the composition of the latent construct and *with few exceptions* each causal indicator is important to include“, sowie Diamantopoulos und Winklhofer (2001, S. 273), die davor warnen, dass „[i]ndicator elimination – by whatever means – should not be divorced from the conceptual considerations when a formative measurement model is involved“. Wir empfehlen, die Elimination eines Indikators für die Reduzierung von Multikollinearität nur dann in Erwägung zu ziehen, falls (a) das Ausmaß der Multikollinearität sehr hoch ist (das heißt wenn etwa 80 % der Varianz eines Indikators durch andere Indikatoren erklärt wird), *und* (b) der Pfad vom Indikator zur latenten Variable nicht signifikant von Null verschieden ist, *und* (c) die übrigen Indikatoren die Inhaltsbedeutung des Konstruktes ausreichend weit abdecken, das heißt dass die verbleibenden Indikatoren weiterhin den gesamten Bedeutungsinhalt des Konstruktes erfassen. Die letzte Bedingung ist von besonderer Wichtigkeit, denn „when constructing a measure, one has to reconcile the theory-driven conceptualization of the measure with the desired statistical properties of the items comprising the measure ... blindly eliminating items to improve reliability (in a reflective scale) or to reduce multicollinearity (in a formative index) may well have adverse consequences for the content validity of the derived measure“ (Diamantopoulos/Siguaw, 2006, S. 14).

## E. Zusammenfassung

In dieser Notiz haben wir einige Aussagen von Albers und Hildebrandt (2006) in ihrem Artikel zur Erfolgsfaktorenmessung herangezogen, um bestimmte potenzielle Missverständnisse in Bezug auf die Festlegung von Indikatoren als formativ, ihre Bewertung hinsichtlich Validität und den Umgang mit Multikollinearität aufzuklären. In diesem An-sinnen haben wir zuerst dargelegt, warum wir nicht Albers und Hildebrandts (2006) Ansicht teilen, dass die Kausalität in Messmodellen nicht eindeutig zu bestimmen ist, und die vordergründige Problematik einer Fehlspezifikation von Messmodellen in den Folgen eines fälschlich gewählten Bereinigungsverfahrens liegt, während auf konzeptioneller Ebene eine inadäquate reflektive Spezifikation als Einschränkung der Inhaltsbedeutung des Konstruktes angesehen und toleriert werden kann. Wir stimmen Albers/Hildebrandt (2006) vollumfänglich in dem Punkt zu, dass die Folgen einer falschen Bereinigung auf empirischer Ebene schwerwiegend sind, wie anhand verschiedener Studien demonstriert wurde, und in jedem Fall ein schlagendes Argument für die sorgfältige Spezifikation von Messmodellen darstellt. Wie wir jedoch ausführlich darlegen, stimmen wir den Autoren nicht in ihrem Punkt zur Tolerierbarkeit einer Fehlspezifikation auf konzeptioneller Ebene zu. Wir haben versucht, die zentrale Wichtigkeit einer korrekten Spezifikation auf theoretischer Ebene, das heißt einer korrekten auxiliary theory (Blalock, 1969; Costner, 1969), deutlich darzustellen. Denn die theoretischen Folgen einer Fehlspezifikation (wie etwa Schlussfolgerungen über theoretische Beziehungen zwischen Konstrukten basierend auf

über- oder unterschätzten Strukturparametern) ist von mindestens gleicher Relevanz wie die empirischen Folgen eines fälschlich gewählten Bereinigungsverfahrens. Danach wurde anhand von Literatur belegt, dass die Validität formativer Indikatoren sehr wohl formal beurteilt werden kann, sofern ein prüfbares Messmodell anstelle einer definitiven Darstellung des Konstruktes postuliert wird. Schließlich wurden potenzielle Einschränkungen für die Anwendbarkeit der Indikatorenzusammenfassung zur Reduzierung von Multikollinearität im Messmodell aufgezeigt.

Wir hoffen, dass unsere Diskussion der obigen Sachverhalte zur weiteren Auseinandersetzung mit dem Thema der formativen Messung anregt und ergänzende Erkenntnisse erlangt werden, wie Modelle mit formativen Indikatoren bestmöglich konzeptionalisiert und bewertet werden können.

## Anmerkungen

- 1 Hermann/Huber/Kressmann (2003) demonstrieren, dass Jarvis/MacKenzie/Podsakoff's (2003) Kriterienliste im Wesentlichen auf das erste Kriterium beschränkt werden kann. Alle weiteren Kriterien lassen sich auf das erste zurückführen.
- 2 Zu alternativen Sichtweisen siehe Bollen/Davis (1994), MacCallum/Browne (1993) sowie MacKenzie/Podsakoff/Jarvis (2005).
- 3 MIMIC Modelle können auf unterschiedliche Weise interpretiert werden, nämlich als: (a) einzelnes Konstrukt mit formativen und reflektiven Indikatoren, (b) exogene Variablen, die ein einzelnes, durch reflektive Indikatoren gemessenes, Konstrukt beeinflussen, sowie (c) ein formativ gemessenes Konstrukt, das die manifesten Variablen von zwei verschiedenen Konstrukten beeinflusst (siehe dazu Jarvis et al., 2003, S. 213). Entscheidend ist, dass, unabhängig von der gewählten Interpretation, diese Modelle statistisch ident sind: „Empirically, these interpretations are indistinguishable because they all produce identical estimates of the relationships between the measures and the constructs.“ (Jarvis et al., 2003, S. 213)
- 4 Es ist festzuhalten, dass in diesem Fall gilt: „there is no reason that a specific pattern of sign (i.e., positive versus negative) or magnitude (i.e., high versus moderate versus low) should characterize that correlations among formative indicators“ (Diamantopoulos/Winklhofer, 2001, S. 271). Im Gegensatz dazu gilt für reflektive Messmodelle die Bedingung, dass die Indikatoren positiv interkorrelieren müssen (für eine Beweisführung siehe Bollen, 1984).
- 5 In reflektiven Modellen stellt Multikollinearität kein Problem dar, da einfache Regressionen die erklärende latente Variable mit dem jeweiligen erklärten Indikator verknüpfen.

## Literatur

- Albers, S./Hildebrandt, L.(2006): Methodische Probleme bei der Erfolgsfaktorenmessung, in: zfbf, 58. Jg, S. 2–33.
- Bagozzi, R.P. (1982): The Role of Measurement in Theory Construction and Hypothesis Testing: Toward a Holistic Model, in: Fornell, Claes (Hrsg.): A Second Generation of Multivariate Analysis, Vol. 1, S. 5–23.
- Bagozzi, R.P./Fornell, C. (1982): Theoretical Concepts, Measurements, and Meaning, in: Fornell, Claes (Hrsg.): A second Generation of Multivariate Analysis, Vol. 1, S. 24–38.
- Bagozzi, R.P. (1994): Structural Equation Models in Marketing Research: Basic Principles, in: Bagozzi, Richard (Hrsg.): Principles of Marketing Research, S. 317–385, Blackwell Publishing, Oxford.
- Bentler, P.M./Chou, C.P. (1987): Practical Issues in Structural Modeling, in: Sociological Methods and Research, Vol. 16, S. 78–117.
- Berry, W.D./Feldman, S. (1985): Multiple Regression in Practice, Sage, Beverly Hills.
- Blalock, H.M., Jr. (1969): Multiple Indicators and the Causal Approach to Measurement Error, in: American Journal of Sociology, Vol. 75, S. 264–272.
- Bollen, K.A. (1984): Multiple Indicators: Internal Consistency or No Necessary Relationship, in: Quality and Quantity, Vol. 18, S. 377–385.
- Bollen, K.A. (1989): Structural Equations with Latent Variables, Wiley-Interscience, New York.

- Bollen, K.A./Lennox, R. (1991): Conventional Wisdom on Measurement: A Structural Equation Perspective, in: *Psychological Bulletin*, Vol. 110, No. 2, S. 305–314.
- Bollen, K.A./Davis, W.R. (1994): Causal Indicator Models: Identification, Estimation, and Testing, paper presented at the American Sociological Association Convention, Miami.
- Bollen, K.A./Ting, K. (2000): A Tetrad Test for Causal Indicators, in: *Psychological Methods*, Vol. 5, No. 1, S. 3–22.
- Bollen, K.A. (2007): Interpretational Confounding Is Due to Misspecification, Not to Type of Indicator: Comment on Howell, Breivik, and Wilcox (2007), in: *Psychological Methods*, Vol. 12, Nr. 2, S. 219–228.
- Cermak, G.W. (1983): An experimental Test of two Models of Attribute Integration, in: *General Motors Research Publication*, GMR-4386, Warren, Michigan.
- Costner, H.L. (1969): Theory, Deduction, and the Rules of Correspondence, in: *American Journal of Sociology*, Vol. 75, S. 245–263.
- Diamantopoulos, A./Winklhofer, H. M. (2001): Index Construction with Formative Indicators: An Alternative to Scale Development, in: *Journal of Marketing Research*, Vol. 38, S. 269–277.
- Diamantopoulos, A. (2006): The error term in formative measurement models: interpretation and modelling implications, in: *Modelling in Management*, Vol. 1, No. 1, S. 7–17.
- Diamantopoulos, A./Siguaw, J.A. (2006): Formative Versus Reflective Indicators in Organizational Measure Development: A Comparison and Empirical Illustration, in: *British Journal of Management*, Vol. 17, S. 1–20.
- Digman, J.M. (1990): Personality Structure: Emergence of the Five-Factor Model, in: *Annual Review of Psychology*, Vol. 41, S. 417–440.
- Eberl, M. (2006): Formative und reflektive Konstrukte und die Wahl des Strukturgleichungsverfahrens, *DBW*, Vol. 66, No. 6, S. 651–668.
- Edwards, J.R./Bagozzi, Richard P. (2000): On the Nature and Direction of Relationships between Constructs and Measures, in: *Psychological Methods*, Vol. 5, No. 2, S. 155–174.
- Fassott, G. (2006): Operationalisierung latenter Variablen in Strukturgleichungsmodellen: Eine Standortbestimmung, in: *zfbf*, 58. Jg., S.67–88.
- Fayers, P.M./Hand, D.J./Bjorndal, K./Groenvold M. (1997): Causal Indicators in quality of life Research, in: *Quality of Life Research*, Vol. 6, S. 393–406.
- Hair, J.F./Anderson, R.E./Tatham, R.L./Black, W.C. (1998): *Multivariate Data Analysis*, Prentice Hall, New Jersey.
- Hempel, C.G. (1960): Operationism, Observation, and Theoretical Terms, in: Danto, Arthur/Morgenbesser, Sidney. (Hrsg.): *Philosophy of Science*, The World Publishing Company, Cleveland, S. 101–120.
- Herrmann, A./Huber, F./Kressmann, F. (2006): Varianz- und kovarianzbasierte Strukturgleichungsmodelle – Ein Leitfaden zu deren Spezifikation, Schätzung und Beurteilung, in: *zfbf*, 58. Jg. S. 34–66.
- Homburg, C./Klarman, M. (2006): Die Kausalanalyse in der empirischen betriebswirtschaftlichen Forschung – Problemfelder und Anwendungsempfehlungen, in: *DBW*, Vol. 66, No. 6, S. 727–748.
- 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, in: *Journal of Consumer Research*, Vol. 30, S. 199–218.
- Jöreskog, K.G./Goldberger, A.S. (1975): Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable, in: *Journal of the American Statistical Association*, Vol. 10, S. 631–639.
- Law, K.S./Wong, C. (1999): Multidimensional Constructs in Structural Equation Analysis: An Illustration Using the Job Perception and Job Satisfaction Constructs, in: *Journal of Management*, Vol. 25, No. 2, S. 143–160.
- Lewis-Beck, M.S. (1980): *Applied Regression – An Introduction*, Sage, Beverly Hills.
- MacCallum, R.C./Browne, M.W. (1993): The Use of Causal Indicators in Covariance Structure Models: Some Practical Issues, in: *Psychological Bulletin*, Vol. 114, No. 3, S. 533–541.
- MacKenzie, S.B./Podsakoff, P.M./Jarvis, C.B. (2005): The Problem of Measurement Model Misspecification in Behavioral and Organizational Research and Some Recommended Solutions, in: *Journal of Applied Psychology*, Vol. 90, No. 4, S. 710–730.
- Mason, C.H./Perreault, W.D., Jr. (1991): Collinearity, Power, and Interpretation of Multiple Regression Analysis, in: *Journal of Marketing Research*, Vol. 28, S. 268–280.
- Rosstter, J.R. (2002): The C-OAR-SE procedure for scale development in marketing, in: *International Journal of Research in Marketing*, Vol. 19, S. 305–335.
- Scholderer, J./Balderjahn, I. (2006): Was unterscheidet harte und weiche Strukturgleichungsmodelle nun wirklich?, in: *Marketing ZfP*, 28. Jg., S. 57–70.
- Williams, L. J./Edwards, J.y R./Vandenberg, R.J. (2003): Recent Advances in Causal Modeling Methods for Organizational and Management Research, in: *Journal of Management*, Vol. 29, No. 6, S. 903–939.

## **Formative Indikatoren: Einige Anmerkungen zu Ihrer Art, Validität und Multikollinearität**

### **Zusammenfassung**

In dieser Notiz werden einige Aussagen von Albers und Hildebrandt (2006) in ihrem Artikel zur Erfolgsfaktorenmessung herangezogen, um bestimmte potenzielle Missverständnisse in Bezug auf die Festlegung von Indikatoren als formativ, ihre Bewertung hinsichtlich Validität und den Umgang mit Multikollinearität aufzuklären. In diesem An-sinnen wird zuerst dargelegt, warum die Autoren nicht Albers und Hildebrandts (2006) Ansicht teilen, dass die Kausalität in Messmodellen nicht eindeutig zu bestimmen sei, und gleichzeitig vehement auf die zentrale Wichtigkeit einer korrekten Spezifikation auf theoretischer Ebene hingewiesen. Danach wird anhand von Literatur belegt, dass die Validität formativer Indikatoren sehr wohl formal beurteilt werden kann, sofern ein prüfbares Messmodell anstelle einer definitorischen Darstellung des Konstruktes postuliert wird. Schließlich werden potenzielle Einschränkungen für die Anwendbarkeit der Indikatorenzusammenfassung zur Reduzierung von Multikollinearität im Messmodell aufgezeigt.

## **Formative indicators: Some remarks on their nature, validity and multicollinearity**

### **Summary**

In this article, we refer to some statements made by Albers and Hildebrandt (2006) in the context of formative measurement of critical success factors. Our underlying rationale is to clarify potential misconceptions regarding the choice of formative (versus reflective) indicators, the assessment of their validity, and the treatment of multicollinearity. In this vein, we first discuss our reasons for not sharing Albers and Hildebrandt's (2006) view that causality in measurement models cannot be specified. Instead, we point to the central importance of a sound theoretical specification of causality. Subsequently, we show based on literature that a formal validation of formative indicators is indeed possible, if a testable measurement model is postulated. Finally, we highlight a number of potential limitations of treating the problem of multicollinearity by combining highly correlated indicators.