

# Partial Least Squares Strukturgleichungsmodellierung: PLS- SEM

Hair / Hult / Ringle / Sarstedt / Richter / Hauff

2. Auflage 2024  
ISBN 978-3-8006-7145-8  
Vahlen

schnell und portofrei erhältlich bei  
[beck-shop.de](https://www.beck-shop.de)

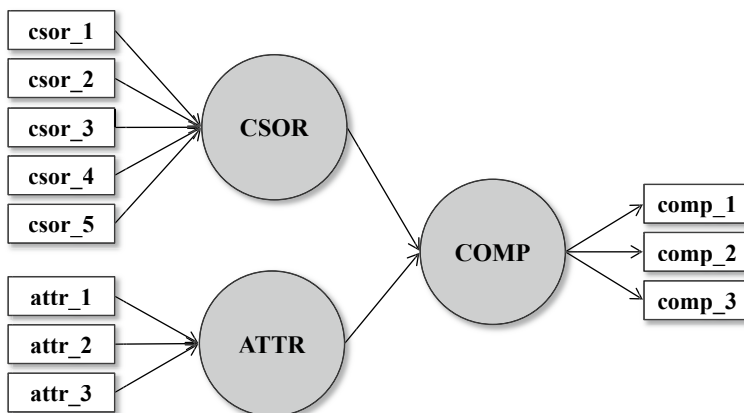
Die Online-Fachbuchhandlung [beck-shop.de](https://www.beck-shop.de) steht für Kompetenz aus Tradition. Sie gründet auf über 250 Jahre juristische Fachbuch-Erfahrung durch die Verlage C.H.BECK und Franz Vahlen. [beck-shop.de](https://www.beck-shop.de) hält Fachinformationen in allen gängigen Medienformaten bereit:

über 12 Millionen Bücher, eBooks, Loseblattwerke, Zeitschriften, DVDs, Online-Datenbanken und Seminare. Besonders geschätzt wird [beck-shop.de](http://beck-shop.de) für sein umfassendes Spezialsortiment im Bereich Recht, Steuern und Wirtschaft mit rund 700.000 lieferbaren Fachbuchtiteln.

Forscher können häufig zwischen verschiedenen etablierten Messansätzen wählen, die sich jeweils im Detail etwas unterscheiden. Die meisten Sozialwissenschaftler wählen heutzutage etablierte Messansätze, die in früheren Forschungsarbeiten oder Skalenhandbüchern (z.B. Bearden et al., 2011; Bruner et al., 2001) veröffentlicht wurden und gut funktionierten (Ramirez et al., 2013). Manchmal sind Forscher aber auch damit konfrontiert, dass es keine etablierten Messansätze gibt, so dass ein neues Instrument entwickelt oder ein bestehendes grundlegend verändert werden muss. Die Beschreibung des generellen Vorgehens zur Entwicklung von Indikatoren für die Messung eines Konstrukts bedarf umfangreicher Detailinformationen, die nicht im Fokus dieses Buches stehen. Wir empfehlen die Lektüre von Hair et al. (2019a), die die wesentlichen Grundlagen dieses Vorgehens diskutieren. Auch Diamantopoulos und Winklhofer (2001), DeVellis (2011), DeVellis (2017) sowie MacKenzie et al. (2011) bieten umfassende Erläuterungen zu verschiedenen Ansätzen zur Entwicklung von Messinstrumenten.

Abbildung 2.6 zeigt einen Ausschnitt des Pfadmodells, das wir als Beispiel im gesamten Buch verwenden. Das Pfadmodell hat zwei exogene Konstrukte – *Corporate Social Responsibility (CSOR)* und *Attraktivität (ATTR)* sowie ein endogenes Konstrukt *Kompetenz (COMP)*. Jedes dieser Konstrukte wird durch mehrere Indikatoren gemessen. Beispielsweise hat das endogene Konstrukt *COMP* drei Indikatorvariablen, *comp\_1* bis *comp\_3*. Unter Verwendung einer Skala von 1 (*stimme überhaupt nicht zu*) bis 7 (*stimme voll und ganz zu*) konnten Befragte ihre Zustimmung zu folgenden Aussagen ausdrücken: „[Das Unternehmen] ist ein Top-Wettbewerber in seinem Markt“, „Soweit ich weiß, ist [das Unternehmen] weltweit bekannt“ und „Ich glaube, dass [das Unternehmen] Spitzenleistungen bietet“. Die Antworten auf diese drei Aussagen repräsentieren die Indikatorvariablen bzw. Messungen des Konstrukts. Das Konstrukt selbst wird indirekt durch diese drei Indikatorvariablen gemessen, daher wird es auch als latente Variable bezeichnet.

Die anderen beiden Konstrukte im Modell, *CSOR* und *ATTR*, lassen sich in einer ähnlichen Weise beschreiben. Beide exogenen Konstrukte werden



**Abbildung 2.6** Beispiel eines Pfadmodells mit drei Konstrukten

indirekt durch Indikatoren gemessen, die jeweils Antworten zu bestimmten Fragen darstellen (und damit direkt gemessen werden). Allerdings sind die Beziehungen zwischen den Indikatoren und den korrespondierenden Konstrukten bei *COMP* im Vergleich zu *CSOR* und *ATTR* anders. Wenn wir uns das Konstrukt *COMP* anschauen, stellen wir fest, dass die Pfeile vom Konstrukt auf die Indikatoren zeigen. Dieser Typ eines Messmodells wird als *reflektiv* bezeichnet. Wenn wir uns dagegen die Konstrukte *CSOR* und *ATTR* anschauen, sehen wir, dass die Pfeile von den gemessenen Indikatorvariablen auf das Konstrukt zeigen. Dieser Typ eines Messmodells wird als *formativ* bezeichnet. In Kapitel 1 haben wir bereits ausgeführt, dass die einfache Integration von reflektiv und formativ spezifizierten Konstrukten eine wichtige Eigenschaft der PLS-SEM ist. Zudem kann die PLS-SEM auch leicht eingesetzt werden, wenn Konstrukte mit nur einem Item (Single-Item anstatt mehrerer Items) spezifiziert werden. Mit beiden Aspekten befassen wir uns in den folgenden Abschnitten.

### *Reflektiv und formativ spezifizierte Messmodelle*

Bei der Entwicklung von Konstrukten gilt es zwei Typen von Messmodellspezifikationen zu berücksichtigen: reflektiv und formativ spezifizierte Messmodelle. Das **reflektiv spezifizierte Messmodell** hat eine lange Tradition in den Sozialwissenschaften und basiert direkt auf der klassischen Testtheorie. Nach dieser Theorie repräsentieren Messungen die Effekte (oder Manifestationen) eines zugrundeliegenden Konstrukts. Daher ist die Beziehung auch vom Konstrukt auf seine Indikatoren gerichtet (*COMP* in Abbildung 2.6). Reflektive Indikatoren (in der psychometrischen Literatur manchmal als **Effekt-Indikatoren** bezeichnet) können als repräsentative Beispiele aller möglichen Items innerhalb der konzeptionellen Definition eines Konstrukts gesehen werden (Nunnally & Bernstein, 1994). Da alle reflektiven Indikatoren eines Konstrukts durch das Konstrukt „verursacht“ werden, sollten sie hoch miteinander korrelieren. Zudem sollten die einzelnen Indikatoren austauschbar sein und einzelne Indikatoren können – solange das Konstrukt genügend Reliabilität besitzt – ohne große Auswirkungen auf die Inhaltsvalidität der Konstruktmessung aus dem Messmodell herausgelassen werden. Die Tatsache, dass die Beziehung vom Konstrukt zu seinen Indikatoren geht, impliziert schließlich, dass Veränderungen des latenten Konstrukts zu einer simultanen Veränderung aller Indikatoren führen. Ein Set an reflektiven Indikatoren wird üblicherweise als **Skala** bezeichnet.

Im Vergleich dazu basieren **formativ spezifizierte Messmodelle** auf der Annahme, dass die Indikatoren das Konstrukt mittels einer Linearkombination bilden. Daher bezeichnen Forscher diesen Typ Messmodell auch als formativen **Index**. Eine wesentliche Eigenschaft formativer Indikatoren ist, dass diese im Gegensatz zu reflektiven Indikatoren nicht austauschbar sind. Vielmehr erfasst jeder Indikator eines formativ spezifizierten Konstrukts einen bestimmten Aspekt der Konstruktdefinition. Zusammengenommen bestimmen die Items die Bedeutung des Konstrukts, weshalb eine Elimination oder Vernachlässigung eines Indikators zu einer potenziellen Änderung der inhaltlichen Natur des Konstrukts führt. Somit muss unbedingt sichergestellt werden, dass der

Inhalt des Konstrukts in all seinen Facetten über die Indikatoren auch adäquat erfasst wird (Diamantopoulos & Winklhofer, 2001).

Einige Forscher unterscheiden im Kontext der **formativen Messung** zwei Typen von Indikatoren: Composite-Indikatoren und kausale Indikatoren. **Composite-Indikatoren** entsprechen weitestgehend der oben angeführten Definition formativ spezifizierter Messmodelle, wonach Indikatoren über eine Linearkombination die **Composite-Variable** bilden (Kapitel 1; Bollen, 2011; Bollen & Bauldry, 2011). Genau genommen wird die Composite-Variable vollständig durch die Indikatoren geformt (d.h. das  $R^2$  der Composite-Variable ist 1,0). Composite-Indikatoren wurden oft zur Messung von **Artefakten** verwendet, welche als von Menschen konstruierte Konzepte verstanden werden (Henseler, 2017b). Beispiele für solche Artefakte im Marketing sind der Einzelhandelspreisindex oder der Marketing-Mix (Hair et al., 2019d). Composite-Indikatoren können jedoch auch zur Messung von Einstellungen, Wahrnehmungen und Verhaltensabsichten verwendet werden (Sarstedt et al., 2016b; Rossiter, 2011; Rossiter, 2016), vorausgesetzt, die Indikatoren weisen eine konzeptionelle Konsistenz gemäß einer klaren theoretischen Definition auf. Aufgrund der Art und Weise, wie der PLS-SEM-Algorithmus formative Messmodelle schätzt, basiert dieser ausschließlich auf dem Konzept der Composite-Indikatoren (z. B. Diamantopoulos & Riefler, 2011).

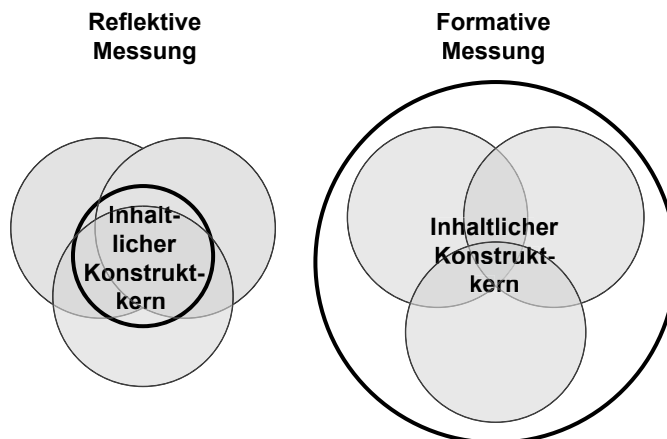
**Kausale Indikatoren** formen ebenfalls das latente Konstrukt. Diese Art der Messung erkennt aber an, dass es sehr unwahrscheinlich ist, dass irgendein Set kausaler Indikatoren jeden Aspekt eines latenten Phänomens voll erfassen kann (Diamantopoulos & Winklhofer, 2001). Daher haben die mit kausalen Indikatoren gemessenen latenten Variablen auch einen **Fehlerterm**, welcher alle anderen, nicht im Messmodell enthaltenen Ursachen erfasst (Diamantopoulos, 2006). In der CB-SEM ist die Verwendung von kausalen Indikatoren verbreitet, da das Verfahren (zumindest im Prinzip) die Modellierung und Schätzung der Fehlerterme eines formativ gemessenen latenten Konstrukts ermöglicht. Die Eigenschaften und Größe dieses Fehlerterms sind allerdings nicht zweifelsfrei bestimmbar, da seine Größe teilweise von anderen Konstrukten im Modell und deren Messqualität abhängt (Aguirre-Urreta et al., 2016).

Kurz gefasst liegt der Unterschied zwischen Composite-Indikatoren und kausalen Indikatoren in den unterschiedlichen Messphilosophien. Kausale Indikatoren nehmen an, dass ein bestimmtes Konstrukt prinzipiell vollständig durch ein bestimmtes Set an Indikatoren und einen Fehlerterm gemessen werden kann. Composite-Indikatoren machen keine derartige Annahme und sehen die Messung eher als eine Approximation eines theoretischen Konstrukts. Die Einbeziehung eines Fehlerterms in Modelle mit kausalen Indikatoren erscheint auf den ersten Blick naheliegend. Da jedoch die Größe des Fehlerterms von der Messqualität der nachgelagerten Konstrukte abhängt, ist dieser Wert für die Beurteilung der Qualität des formativen Messmodells nicht eindeutig (Rigdon et al., 2014). Durch die Aufnahme eines Fehlerterms in ein formatives Messmodell wird die formative Messung in der CB-SEM zudem so behandelt, als wäre diese ein **Faktormodell**. In der PLS-SEM werden formative Messmodelle dagegen mit Composite-Indikatoren geschätzt, was dem

composite-basierten Ansatz, der dem PLS-SEM-Algorithmus zugrunde liegt, völlig entspricht. Das heißt, unabhängig davon, ob reflektive oder formative Messmodelle geschätzt werden, verwendet die PLS-SEM Linearkombinationen zur Bildung von Composites, um die Konstrukte in einem Pfadmodell zu messen (Kapitel 3).

In Anbetracht dieser Ausführungen erscheint die Unterscheidung zwischen kausalen Indikatoren und Composite-Indikatoren eher künstlich und hat kaum Konsequenzen für die Methodenwahl. Zur Vereinfachung und Anlehnung an die einschlägige Forschung (z. B. Fornell & Bookstein, 1982) verwenden wir in den weiteren Darlegungen den Begriff formative Indikatoren und beziehen uns dabei auf Composite-Indikatoren (da diese in der PLS-SEM verwendet werden). Entsprechend benutzen wir den Begriff „formativ spezifizierte Messmodelle“, um Messmodelle zu beschreiben, bei denen aus messtheoretischen Überlegungen eine Wirkungsrichtung von den Indikatoren zum Konstrukt angenommen werden kann. Weitere Informationen zu Composite- und Faktormodellen sowie zu deren Unterscheidung finden sich bei Henseler et al. (2014), Rigdon et al. (2017) und Sarstedt et al. (2016b).

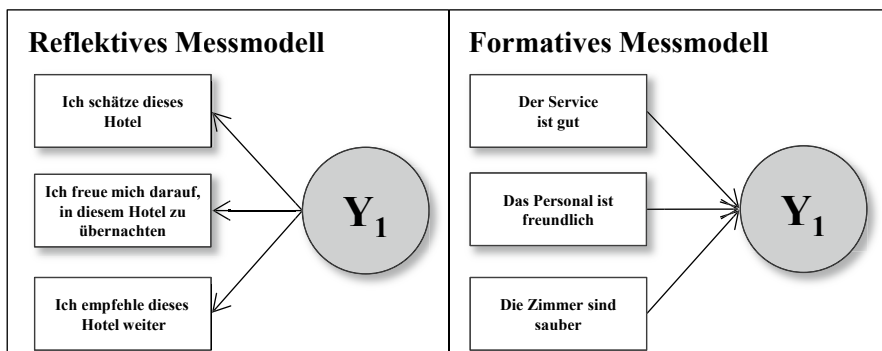
Abbildung 2.7 veranschaulicht die grundsätzlichen Unterschiede zwischen der reflektiven und formativen Messperspektive. Der schwarze Kreis veranschaulicht den inhaltlichen Kern, den das Konstrukt messen soll. Die grauen Kreise repräsentieren den durch die einzelnen Indikatoren erfassten Inhalt. Während der reflektive Messansatz darauf zielt, die inhaltliche Überschneidung zwischen den (austauschbaren) Indikatoren zu maximieren, versucht der formative Messansatz den Inhalt des untersuchten latenten Konstrukts (schwarzer Kreis) möglichst vollständig durch die verschiedenen formativen Indikatoren (graue Kreise) zu erfassen, wobei letztere eine möglichst geringe Überschneidung haben sollten.



**Abbildung 2.7** Unterschied zwischen reflektiven und formativen Messungen

Im Unterschied zum reflektiven Messansatz, dessen Ziel in der Maximierung der Überschneidung zwischen austauschbaren Indikatoren liegt, gibt es bei formativen Indikatoren keine spezifischen Erwartungen an die Richtung und die Höhe der Korrelationen zwischen den einzelnen Indikatoren (Diamantopoulos et al., 2008). Da es keine „gemeinsame Ursache“ für die Items eines Konstrukts gibt, müssen diese nicht notwendigerweise korrelieren und können sogar vollkommen unabhängig sein. Tatsächlich kann Kollinearität zwischen den formativen Indikatoren ein beträchtliches Problem darstellen, da die Gewichte, welche die Stärke des Zusammenhangs zwischen Indikatoren mit dem Konstrukt ausdrücken, instabil und nicht signifikant sein können. Darüber hinaus haben formativ spezifizierte Messmodelle mit Composite-Indikatoren keine individuellen Fehlerterme (d. h. es wird davon ausgegangen, dass sie fehlerfrei sind). Diese Eigenschaften haben wichtige Implikationen für die Evaluation von formativ gemessenen Konstrukten, weshalb im Vergleich zur Evaluation von reflektiven Messungen völlig andere Kriterien verwendet werden (Kapitel 5). So könnte beispielsweise eine Reliabilitätsanalyse auf Basis der Item-Korrelationen (interne Konsistenz) zur Elimination von inhaltlich wichtigen Items führen, wodurch die Validität des Index eingeschränkt würde (Diamantopoulos & Siguaaw, 2006). Allgemein gesagt müssen Forscher genauer auf die Inhaltsvalidität ihrer Messungen achten, indem sie ermitteln, wie gut ein Set Indikatoren den Inhalt eines zu untersuchenden Konstrukts (oder zumindest seine wesentlichen Facetten) erfasst (z. B. Bollen & Lennox, 1991).

Wann messen wir ein Konstrukt reflektiv und wann formativ? Es gibt keine eindeutige Antwort auf diese Frage, da Konstrukte nicht originär reflektiv oder formativ sind. Im Gegenteil hängt die Spezifikation von der Konzeptionalisierung des Konstrukts und den Zielen der Untersuchung ab. Schauen wir uns dazu Abbildung 2.8 an, in der verdeutlicht wird, wie das Konstrukt „Zufriedenheit mit Hotels“ ( $Y_1$ ) auf beide Arten operationalisiert werden kann (Albers, 2010).



**Abbildung 2.8** Zufriedenheit als reflektiv und formativ gemessenes Konstrukt

Quelle: In Anlehnung an Albers (2010). Nachdruck mit freundlicher Genehmigung durch Springer Science+Business Media.

Die linke Seite der Abbildung 2.8 zeigt eine reflektive Messmodellspezifikation. Diese Art der Spezifikation ist wahrscheinlich angemessener, wenn Forscher Theorien in Bezug auf die Zufriedenheit testen möchten. In vielen managementorientierten Untersuchungen liegt das Ziel aber darin, die wichtigsten Treiber der Zufriedenheit zu identifizieren, die dann letztendlich zu Kundenloyalität führen. In diesem Fall sollten Forscher die unterschiedlichen Facetten von Zufriedenheit berücksichtigen, wie etwa die Zufriedenheit mit dem Service oder dem Personal (Abbildung 2.8, rechte Seite). Eine solche formative Messmodellspezifikation ist hier vielversprechender, da sie die genaue Identifikation von Treibern der Zufriedenheit und damit die Ableitung spezifischer Handlungsempfehlungen ermöglicht. Dies trifft insbesondere zu, wenn die korrespondierenden Konstrukte exogen sind. Allerdings können formativ spezifizierte Messmodelle auch bei endogenen Konstrukten verwendet werden, sofern dies durch die Messtheorie gestützt wird.

Neben der Rolle, die ein Konstrukt innerhalb eines Modells spielt, und den Empfehlungen, die Forscher auf den Ergebnissen basierend geben möchten, hängt die Messperspektive insbesondere von der Spezifikation des Konstruktinhalts ab. Die Frage der geeigneten Spezifikation der Messmodelle wird in verschiedenen Disziplinen nach wie vor stark diskutiert und ist nicht vollständig geklärt. In Abbildung 2.9 präsentieren wir verschiedene Kriterien, die Forscher als Orientierung für ihre Entscheidung zwischen reflektiven oder formativen Messungen verwenden können. Darüber hinaus ist zu beachten, dass es auch

Kriterium	Entscheidung	Quelle
<b>Richtung der Kausalität zwischen Indikatoren und Konstrukt</b>	<ul style="list-style-type: none"> <li>Vom Konstrukt zu den Indikatoren: reflektiv</li> <li>Von den Indikatoren zum Konstrukt: formativ</li> </ul>	Diamantopoulos und Winklhofer (2001)
<b>Ist das Konstrukt eher ein Faktor, welcher die Indikatoren erklärt, oder eher eine Kombination der Indikatoren?</b>	<ul style="list-style-type: none"> <li>Falls Faktor: reflektiv</li> <li>Falls Kombination: formativ</li> </ul>	Fornell und Bookstein (1982)
<b>Repräsentieren die Indikatoren Auswirkungen oder Ursachen des Konstrukts?</b>	<ul style="list-style-type: none"> <li>Falls Auswirkungen: reflektiv</li> <li>Falls Ursachen: formativ</li> </ul>	Rossiter (2002)
<b>Ist es immer der Fall, dass sich bei einer anderen Bewertung des Konstrukts alle Indikatoren auf die gleiche Art und Weise verändern (angenommen sie sind gleich kodiert)?</b>	<ul style="list-style-type: none"> <li>Falls ja: reflektiv</li> <li>Falls nein: formativ</li> </ul>	Chin (1998)
<b>Sind die Items wechselseitig austauschbar?</b>	<ul style="list-style-type: none"> <li>Falls ja: reflektiv</li> <li>Falls nein: formativ</li> </ul>	Jarvis et al. (2003)

**Abbildung 2.9** Entscheidungskriterien für die Wahl zwischen einem reflektiven oder formativen Messmodell



empirische Möglichkeiten zur Bestimmung der Messperspektive gibt. Gudergan et al. (2008) schlagen die sogenannte **konfirmatorische Tetrad-Analyse für die PLS-SEM (CTA-PLS)** vor, mit deren Hilfe die Nullhypothese getestet werden kann, dass die Messmodelle reflektiver Natur sind. Wir stellen die CTA-PLS ausführlicher in Kapitel 8 vor. Natürlich sollte eine rein datengetriebene Perspektive durch theoretische Überlegungen, wie sie in den in Abbildung 2.9 gegebenen Richtlinien zusammengefasst werden, ergänzt werden.

### *Single-Item-Messungen und Summenwerte*

Forscher entscheiden sich manchmal für die Verwendung von **Single-Items** anstelle von multiplen Items zur Messung eines Konstrukts. Single-Items haben verschiedene praktische Vorteile, wie die einfache Anwendung, ihre Kürze und die mit ihrer Verwendung verbundenen geringeren Kosten. Im Vergleich zu langen und komplizierten Skalen, die oft zu Unverständnis und mentalen Herausforderungen bei den Befragten führen, fördern Single-Items höhere Antwortquoten, da die Fragen einfach und schnell beantwortet werden können (Fuchs & Diamantopoulos, 2009; Sarstedt & Wilczynski, 2009). Allerdings bieten Single-Items nicht nur Vorteile. Sollen die Daten beispielsweise in Gruppen eingeteilt werden, stehen bei der Verwendung von Single-Items weniger Freiheitsgrade zur Bestimmung der Gruppenzugehörigkeit zur Verfügung, da hierfür nur die Werte einer einzigen Variablen verwendet werden können. Auch bei der Behandlung von fehlenden Werten stehen nur die Informationen einer einzigen statt mehrerer Variablen zu Verfügung.

Gegen Single-Item-Messungen spricht besonders, dass sie aus psychometrischer Sicht nicht in der Lage sind, Messfehler zu korrigieren, wodurch sich ihre Reliabilität gegenüber Multi-Item-Messungen generell verringert. Die Reliabilität von Single-Items kann zwar entgegen weit verbreiteter Annahmen geschätzt werden (z. B. Cheah et al., 2018; Loo, 2002; Wanous et al., 1997) – siehe Abbildung 5.3 in Kapitel 5 für Details. Allerdings ist die Entscheidung für Single-Item-Messungen in den meisten empirischen Untersuchungen sehr riskant, da hierdurch die prädiktive Validität negativ beeinflusst wird. Genau genommen sind Situationen, in denen eine Single-Item-Messung einer Messung mit multiplen Items vorgezogen werden sollte, in der Praxis sehr unwahrscheinlich. Den Empfehlungen von Diamantopoulos et al. (2012) folgend, sollten Single-Item-Messungen nur in Situationen in Betracht gezogen werden, in denen (1) geringe Fallzahlen vorliegen (d. h.  $N < 50$ ), (2) Pfadkoeffizienten (also diejenigen Koeffizienten, welche die Konstrukte im Strukturmodell verbinden) von 0,30 und kleiner erwartet werden, (3) die Items der originalen Multi-Item-Skala sehr homogen sind (d. h. Cronbachs Alpha  $> 0.90$ ) und (4) die Items semantisch redundant sind (Abbildung 2.10; vgl. hierzu auch Kamakura, 2015).

Nichtsdestotrotz sollte diese rein empirische Perspektive bei der Spezifikation von Messmodellen durch praktische Überlegungen ergänzt werden. Einige Forschungssituationen erfordern die Verwendung von Single-Items. Befragte befürchten bei der Verwendung von Multi-Item-Skalen evtl. zu viel von sich preiszugeben, was zu geringen Antwortquoten führt. Zudem fehlt es häufig

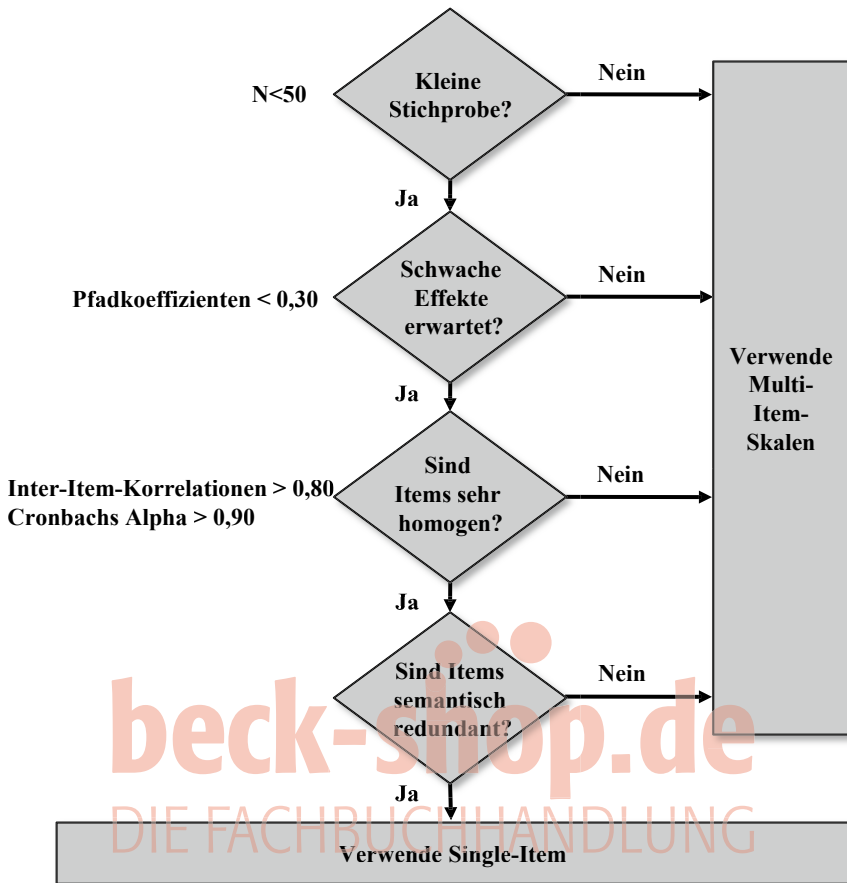


Abbildung 2.10 Richtlinien zur Verwendung von Single-Items

Quelle: Diamantopoulos et al. (2012).

auch an der Bereitschaft, sich die Zeit zum Ausfüllen eines Fragebogens zu nehmen, wodurch große Stichproben bei Umfragen schwierig werden. In der Konsequenz ist es oft notwendig, die Länge eines Fragebogens wo immer möglich zu reduzieren. Ist also die Population klein oder steht nur eine kleine Stichprobe zur Verfügung (z. B. durch Budgetrestriktionen oder Schwierigkeiten bei der Rekrutierung von Umfrageteilnehmern), kann die Nutzung von Single-Item-Messungen eine pragmatische Lösung sein. Sollten Forscher die Konsequenzen einer geringen prädiktiven Validität akzeptieren und Single-Item-Messungen trotzdem verwenden, bleibt dennoch eine grundlegende Frage: Wie oder was sollte dieses Item sein? Forschungsarbeiten weisen leider auf große Schwierigkeiten bei der Wahl eines Items aus einem Set möglicher Items, unabhängig davon, ob die Wahl auf statistischen Maßen oder Experteneinschätzungen beruht (Sarstedt et al., 2016a). Vor diesem Hintergrund sprechen wir uns klar gegen die Verwendung von Single-Items zur Messung von Konstrukten aus, es sei denn die Richtlinien von Diamanto-