

Die Technik und Logik von linearen Strukturgleichungsmodellen

Empfehlenswerte Einführungen in die Arbeit mit Strukturgleichungsmodellen (“*structural equation modeling*”) finden sich in Byrne (1994) (spezifisch für das Programm EQS, von der Autorin liegen aber auch vergleichbare Einführungen für die Programme AMOS und LISREL vor) und Loehlin (1998) (hierin gibt es auch eine sehr gute Einführung in die der Technik der Strukturgleichungsmodelle vorausgehenden Methode der Pfadanalyse). Da die Regressionsanalyse auch das Fundament von Pfadanalysen und Strukturgleichungsmodellen ist, sei hier ebenfalls der für angewandte Forscher geschriebene Klassiker von Cohen & Cohen (2003) empfohlen. Für Anfänger wie auch Experten ist des weiteren die Internet-Diskussionsgruppe (“*mailing list*”) SEMNET zu empfehlen (<http://www.bama.ua.edu/archives/semnet.html>).

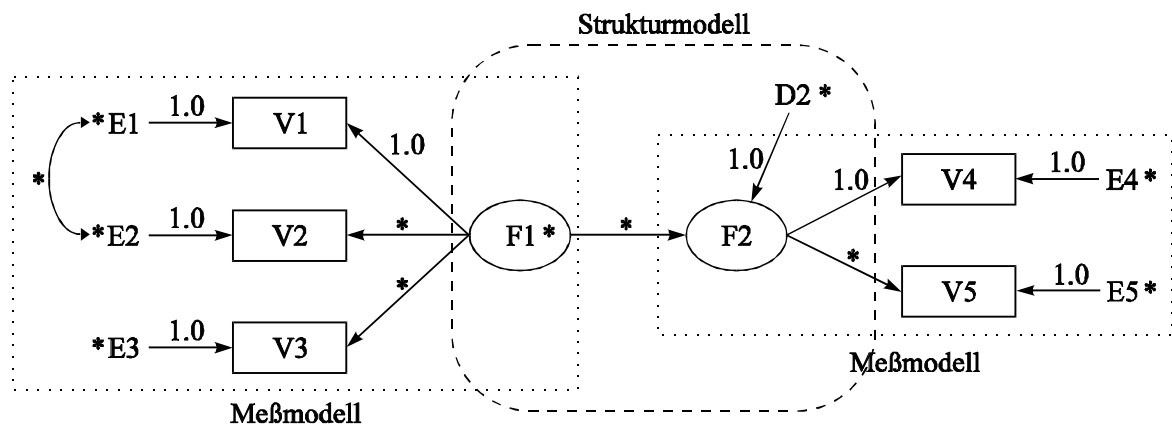
In der Logik von latenten Strukturgleichungsmodellen wird angenommen, daß ein latentes Konstrukt (ein Faktor) die beobachteten Werte bestimmter Items (d.h. gemessener Variablen) und damit auch deren beobachtete Kovarianz beeinflusst und daß ein Item somit ein Indikator für einen im Hintergrund wirksamen Faktor ist. Da aber davon ausgegangen wird, daß auch andere, unspezifische Einflüsse, die mit dem latenten Konstrukt nicht korrelieren, die konkreten Werte der Items bewirkt haben können, wird angenommen, daß die Varianz der beobachteten Werte eines Items sich aus der Varianz des latenten Faktors und einer itemspezifischen Residualvarianz zusammensetzt. Derjenige Teil eines Strukturgleichungsmodells, in dem Zusammenhänge zwischen latenten Faktoren und beobachteten Variablen derart als eine *faktorielle Struktur*¹ formuliert werden, wird auch als *Meßmodell* bezeichnet, da sich in ihm (wie im Rahmen der klassischen Testtheorie) die Varianz eines *gemessenen* Wertes aus der Varianz der "wahren Werte" einer (latenten) Variablen und einer nicht erklärten Restvarianz ergibt. In einfachen konfirmatorischen Faktorenanalysen wird also nur ein Meßmodell formuliert.

Im Rahmen von Strukturgleichungsmodellen können aber auch Wirkungszusammenhänge *zwischen den latenten Variablen selbst* angenommen und modelliert werden, hierbei wird in analoger Weise die Varianz eines Faktors (d.h. einer abhängigen latenten Variablen, z.B. emotionale Erschöpfung) durch die Varianz eines anderen, auf ihn wirkenden Faktors (d.h. einer unabhängigen latenten Variablen, z.B. Zeitdruck) und der Residualvarianz dieses abhängigen Faktors bestimmt. Je geringer die Residualvarianz des abhängigen Faktors ist, desto enger hängt er mit dem unabhängigen Faktor zusammen bzw. desto größer ist der Effekt des unabhängigen Faktors. Ein derartiges Modell formuliert Annahmen über die

¹ "Faktorielle Struktur" oder "Faktorstruktur" eines Instruments bezeichnet meist das Ladungsmuster der *Items* (inklusive möglicher Faktorinterkorrelationen), obwohl in der Terminologie von latenten Strukturgleichungsmodellen unter "Faktorstruktur" im engeren Sinne auch nur die Kovarianzstruktur der *Faktoren* verstanden wird.

Strukturzusammenhänge der latenten Variablen und wird daher auch als *Strukturmodell* bezeichnet. Ein Strukturmodell läßt sich also als ein Regressionsmodell mit meßfehlerbereinigten Variablen auffassen, da die latenten Variablen die "wahren Werte" der beobachteten Variablen darstellen. Insofern erlaubt die Methode der Strukturgleichungsmodelle die Prüfung von Annahmen über Zusammenhänge theoretischer Konstrukte.

Bezüglich des Meßmodells bedeuten Residualvarianzen der Items, daß die beobachteten Variablen nicht perfekt, d.h. nicht fehlerfrei gemessen wurden, bezüglich des Strukturmodells bedeuten Residualvarianzen der Faktoren, daß die abhängigen latenten Variablen durch die spezifizierten unabhängigen latenten Variablen nicht perfekt vorhergesagt werden können. Trotz meßfehlerbehafteter Beobachtungen kann mit latenten Strukturgleichungsmodellen der Zusammenhang von *geschätzten wahren Werten*, also ein *theoretisches* Modell getestet werden.



Anmerkungen: Beobachtete Variablen: V; Residuen beobachteter Variablen (Error): E; latente Variablen (Faktoren): F; Residuen latenter Variablen (Disturbances): D; Variablen, auf die *unidirektionale* Pfeile weisen, sind abhängige Variablen; Sterne kennzeichnen zu schätzende (freie) Parameter (bei Pfaden oder Doppelpfeilen Regressionskoeffizienten oder Kovarianzen, bei Variablen Varianzen); anstelle eines Sterns kann ein Wert die Größe eines fixen, nicht zu schätzenden Parameters kennzeichnen. In Strukturgleichungsmodellen nach Bentler & Weeks (1979, 1980) sind alle Modellvariablen entweder abhängige oder unabhängige Variablen, dabei können die Varianzen (und Kovarianzen) abhängiger Variablen niemals freie, zu schätzende Parameter sein, sondern sie werden durch solche Parameter erklärt. Residuen haben entweder eine fixierte Varianz (z.B. 1.0) und zu schätzende Regressionskoeffizienten oder fixierte Regressionskoeffizienten (z.B. 1.0) und zu schätzende Varianz. Für jede latente Variable muß die Skala (Maßeinheit) definiert sein, deshalb wird entweder der Regressionskoeffizient eines Indikators oder die Varianz der latenten Variablen auf 1.0 fixiert (bei abhängigen Variablen kann nur ein Regressionskoeffizient fixiert werden).

Abbildung 1: Ein hypothetisches Strukturgleichungsmodell (in Anlehnung an Byrne, 1994, S. 9)

Die Zusammenhänge von beobachteten Werten, latenten Konstrukten und Residuen lassen sich in Form eines Pfaddiagramms darstellen. In *Abbildung 1* sind (wie in Pfaddiagrammen üblich) direkt beobachtete Variablen mit rechteckigen Rahmen umgeben, latente Variablen

mit ovalen. Die Pfeile geben die im Modell angenommenen Wirkungsrichtungen in Form von Regressionskoeffizienten wieder, (gebogene) Doppelpfeile stellen Kovarianzen oder Korrelationen zwischen Variablen des Modells dar, bei denen eine bestimmte *Wirkungsrichtung* nicht spezifiziert wird. Im konkreten Modell wird angenommen, daß die beobachteten Variablen V1, V2 und V3 Indikatoren der latenten Variable F1 sind, ihre Residualvarianzen werden mit E1, E2 und E3 bezeichnet. V1, V2 und V3 werden also durch F1 sowie E1, E2 und E3 beeinflusst. Die Residualvarianzen (oder Meßfehler) E1 und E2 werden als korreliert angenommen. V4 und V5 sind dementsprechend die Indikatoren für die latente Variable F2, ihre Residualvarianzen sind E4 und E5. Des weiteren wird im Modell angenommen, daß die Variable F2 durch F1 beeinflusst wird, die durch diesen Zusammenhang nicht erklärte Residualvarianz von F2 wird mit D2 bezeichnet, d.h. auf F2 wirkt ein im Rahmen des Modells nicht spezifizierter Anteil.

Wenn die im Modell spezifizierte Struktur der Variablen in der Form von linearen Strukturgleichungen (z.B. $F2 = *F1 + D2$) formuliert wird (von daher der Name "Strukturgleichungsmodelle"), können unter bestimmten Voraussetzungen anhand der Varianzen und Kovarianzen der beobachteten Variablen die Parameter des Modells (z.B. mit der Maximum-Likelihood-Methode) simultan *geschätzt* werden, wobei die Parameter so gewählt werden, daß sie möglichst optimal mit den empirischen Daten übereinstimmen. Die Anwendung von Strukturgleichungsmodellen setzt also voraus, daß *a priori* theoretische Annahmen über die Zusammenhänge der Variablen formuliert worden sind (es sind allerdings auch Respezifikationen des Modells möglich, wodurch konfirmatorische in exploratorische Analysen übergehen). Will man hypothetische Zusammenhänge zwischen den Variablen eines Modells statistisch *testen*, ist es nötig, daß das Modell weniger freie (zu schätzende) Parameter hat, als die Anzahl der Elemente der Kovarianzmatrix der im Modell verwendeten beobachteten Variablen (zu den Problemen der theoretischen und empirischen Identifizierbarkeit von Modellen vgl. Bollen, 1989). Die Differenz zwischen der Anzahl der freien Parameter und der Anzahl der empirischen Datenpunkte ergibt die Anzahl der Freiheitsgrade des Modells. Hat ein Modell genau so viele freie Parameter wie die Kovarianzmatrix Elemente hat (also keine Freiheitsgrade), sind die Parameter eindeutig bestimmt; das Ergebnis ist dann nur noch eine *Beschreibung* der vorgefundenen empirischen Zusammenhänge. Zwar kann dann (wie in klassischen Regressionsanalysen oder Pfadmodellen) die Signifikanz der einzelnen Parameter bestimmt werden, die Übereinstimmung des *Modells* mit den Daten kann dann jedoch nicht getestet werden.

Da im gegebenen Beispiel nicht alle möglichen Zusammenhänge zwischen den Variablen spezifiziert worden sind und eine Varianz-Kovarianzmatrix von fünf Variablen 15 Elemente besitzt, in dem Modell aber nur 12 Parameter geschätzt werden müssen, hat das Modell drei Freiheitsgrade. Das bedeutet, daß die Passung von Daten und Modell nur dann perfekt ist, wenn es außer den spezifizierten Wirkungszusammenhängen tatsächlich *keine* weiteren Zusammenhänge zwischen den Variablen gibt: Gäbe es jedoch noch andere Wirkungs-

zusammenhänge zwischen den Variablen, würde ein bestimmter Anteil der Kovarianzen der Variablen durch das Modell nicht erklärt und die anhand der Modellparameter geschätzten Elemente der "reproduzierten" Kovarianzmatrix wären nicht mit den Elementen der empirischen Kovarianzmatrix identisch, die Elemente der Residualmatrix wären also nicht Null. Ob der nicht erklärte Anteil der Kovarianzen der Variablen signifikant ist, läßt sich approximativ über einen Anpassungstest (χ^2 -Statistik) prüfen. Der einfachste "Goodness-of-Fit"-Index ist der χ^2 -Wert, der anhand der Freiheitsgrade des Modells auf Signifikanz geprüft wird. Da er jedoch sehr sensitiv für die Stichprobengröße ist und mit wachsender Stichprobengröße schon geringe Modellabweichungen signifikant werden, mit kleinen Stichproben aber keine generalisierbaren und stabilen Ergebnisse gewonnen werden können, sind in den letzten Jahren alternative Fit-Indizes entwickelt worden, die diesem Problem Rechnung tragen sollen (vgl. Marsh, Balla & McDonald, 1988; Gerbing & Anderson, 1993; Tanaka, 1993). In Analogie zur Zerlegung der Varianz beobachteter Werte in die Varianz wahrer Werte und Fehlervarianz können die empirischen Daten in einen durch das Modell erklärten Anteil und einen nicht erklärten Anteil zerlegt werden, wobei dann das Modell um so besser zu den Daten paßt, je geringer der Residualanteil ist.

Wenn ein Modell noch Freiheitsgrade hat, erlaubt also die Methode der Strukturgleichungsmodelle, die Passung eines theoretischen Modells zu den empirischen Daten zu *testen*². Eine wichtige Implikation der Logik von Strukturgleichungsmodellen ist, daß mit dem Test eines theoretischen Modells geprüft wird, ob die Annahme von bestimmten, *nicht* existierenden Zusammenhängen zwischen den Modellvariablen empirisch haltbar ist. Deshalb (und weil ein Modell so gut wie nie perfekt zu den Daten paßt) folgt ebenfalls, daß es auch alternative Modelle geben kann, die ebenso gut oder besser zu den Daten passen, als das aktuell geprüfte Modell.

Strukturgleichungsmodelle ohne latente Variablen (Pfad- oder Regressionsmodelle)

Manchmal ist es nicht möglich (oder nicht sinnvoll), mit *latenten* Strukturgleichungsmodellen zu arbeiten. Die Formulierung eines Strukturgleichungsmodells, das *nur beobachtete* Variablen enthält (auch Pfad- oder Regressionsmodell genannt), ist jedoch ebenfalls möglich. Auch solche Modelle können getestet werden, wenn das Modell Freiheitsgrade besitzt. Im Unterschied zu Modellen latenter Variablen sind dann allerdings die Variablen des "Strukturmodells" meßfehlerbehaftet. Wenn ein derartiges Modell des weiteren auch keine Freiheitsgrade mehr besitzt, entspricht es einem klassischen (mehr oder weniger komplexen) Regressionsmodell, wie zum Beispiel bei einer multiplen Regression mit nur einer abhängigen Variablen (in dem die Korrelationen zwischen allen unabhängigen Variablen zwar angenommen, aber häufig nicht dargestellt werden, vgl. die SPSS-Prozedur

² Dies ist genau genommen nur eine notwendige, keine hinreichende Bedingung; praktisch werden jedoch nur Modelle geprüft, die auch empirisch identifizierbar sind.

Regression). Ist die Reliabilität (z.B. die interne Konsistenz) einer Variablen des Regressionsmodells bekannt, kann folgende Formel zur Schätzung der Fehlervarianz benutzt werden (vgl. Jöreskog & Sörbom, 1989, S.153):

$$\hat{\epsilon} = (1-\rho^2) \cdot s^2$$

mit $\hat{\epsilon}$ = geschätzte Fehlervarianz der beobachteten Variablen

ρ^2 = Reliabilität der beobachteten Variablen

s^2 = Varianz der beobachteten Variablen

Werden in Regressionsmodellen die Fehlervarianzen der beobachteten Variablen nicht gleich Null sondern gleich $\hat{\epsilon}$ gesetzt, kann auf diese Weise ein Strukturmodell mit den "wahren Werten" der beobachteten Variablen formuliert werden, obwohl nur jeweils eine beobachtete Variable pro "Faktor" existiert. Da in derartigen Modellen die Meßfehler nicht direkt aus den beobachteten Daten geschätzt werden und das Konzept der internen Konsistenz nicht identisch mit dem Konzept des Meßfehlers in latenten Strukturgleichungsmodellen ist, werden derartige Modelle auch als "quasi-latent" bezeichnet.

Kriterien für die Übereinstimmung von Modell und Daten

Da der χ^2 -Wert zur Bewertung des Modell-Fits (Passung von Modell und Daten) sehr stark von der Stichprobengröße abhängt (er wird insbesondere bei großen Stichproben leicht signifikant), sind eine Reihe von alternativen Fit-Indizes entwickelt worden, von denen vor allem der "Comparative Fit Index" (CFI; vgl. Bentler, 1990; für eine neuere und häufig zitierte Bewertung unterschiedlicher Fit-Indizes vgl. Hu & Bentler, 1999) die besten Eigenschaften zu haben scheint. Der CFI ist insbesondere dann vorzuziehen, wenn das Ergebnis des Modelltests auf die Gesamtpopulation generalisiert und seine Abhängigkeit von der Stichprobengröße reduziert soll (Tanaka, 1993). Er gehört zur Klasse der "relativen" Fit-Indizes, da die Passung des Modells zu den Daten relativ zu einem sogenannten "Nullmodell" evaluiert wird. In dem hypothetischen Nullmodell wird die völlige Unabhängigkeit aller Variablen angenommen (eine sehr restriktive Annahme). Der normalerweise sehr große χ^2 -Wert dieses Modells mit der maximalen Zahl von Freiheitsgraden dient als Basis, mit der das zu evaluierende (weniger restriktive) Modell verglichen wird (würde das Nullmodell zu den Daten passen, wären daraus abgeleitete, weniger restriktive Modelle, die einen Zusammenhang zwischen den Variablen annehmen bzw. zulassen, nicht mehr sinnvoll). Es ist dann möglich, daß das zu evaluierende Modell signifikant besser zu den Daten paßt, als das Nullmodell (also die Differenz der χ^2 -Werte beider Modelle mit entsprechender Differenz der Freiheitsgrade signifikant ist), obwohl der absolute χ^2 -Test des zu evaluierenden Modells zeigt, daß in den Daten noch ein signifikanter, unerklärter Rest bleibt (vgl. Bentler & Bonett, 1980). Im Gegensatz zu dem absoluten χ^2 -Wert des zu evaluierenden Modells (und zu anderen Fit-Indizes wie dem "Goodness of Fit"-Index (GFI) oder dem "Adjusted Goodness of Fit"-Index, vgl. Jöreskog

& Sörbom, 1989) bewertet der CFI nicht den absoluten Fit, sondern zeigt, wie gut ein Modell *relativ* zu dem hypothetischen Nullmodell zu den Daten paßt. Der CFI ist auf ein Intervall von 0 (keine Übereinstimmung zwischen Modell und Daten) und 1 (maximale Übereinstimmung) normiert. Da die Formel für den CFI im häufig gebraucht wird, um einen dem CFI entsprechenden Index für die Satorra-Bentler-Statistik (s.u.) zu berechnen (der wird im folgenden als SB-CFI bezeichnet wird), sei die Formel für den CFI hier angegeben (Bentler, 1992, S. 93):

$$\text{CFI} = 1 - \tau_k / \tau_i$$

mit: $\tau_k = \max [(\chi^2_1 - df_1), 0]$

$\tau_i = \max [(\chi^2_0 - df_0), (\chi^2_1 - df_1), 0]$

und $\chi^2_0 = \chi^2$ des Nullmodells

$df_0 =$ Freiheitsgrade des Nullmodells

$\chi^2_1 = \chi^2$ des zu evaluierenden Modells

$df_1 =$ Freiheitsgrade des zu evaluierenden Modells

Um von einer hinreichenden Übereinstimmung von Modell und Daten zu sprechen, sollte der CFI mindestens 0.90 betragen; ein CFI, der kleiner als 0.90 ist, zeigt eine ungenügende Übereinstimmung von Modell und Daten an, ab einer Größe von circa 0.95 wird von einer guten Übereinstimmung gesprochen.

Die statistische Überprüfung der Passung von Modell und Daten setzt voraus, daß die Daten multivariat normal verteilt sind. Wenn eine Abweichung von dieser Annahme nicht auf Extremwerte nur einiger weniger Personen zurückgeführt werden kann (also durch Ausschluß weniger Personen die Daten der übrigen Personen multivariat normal verteilt sind), kann als Alternative zum χ^2 -Wert die sogenannte robuste Satorra-Bentler-Statistik (SB- χ^2) verwendet werden (vgl. Bentler, 1992), die eine Korrektur des χ^2 darstellt und Abweichungen von der Normalverteilung kompensiert. Das Programm EQS berechnet jedoch diese Statistik nicht für Multi-Gruppen-Analysen. Auch der vom Programm ausgegebene CFI basiert nur auf dem unkorrigierten χ^2 -Wert. Allerdings läßt sich ein der Satorra-Bentler-Statistik entsprechender CFI (SB-CFI) berechnen, indem man den SB- χ^2 auch für das oben beschriebene Nullmodell berechnet und dann diesen Wert und den SB- χ^2 des zu evaluierenden Modells mit den entsprechenden Freiheitsgraden in die oben angegebene Formel für den CFI einsetzt.

Falls das Modell nicht hinreichend zu den Daten paßt, kann geprüft werden, ob es bisher nicht zur Schätzung freigesetzte Modellparameter gibt, deren Schätzung zu einer signifikant besseren Übereinstimmung mit den Daten führen würde, d.h. die Fehlspezifikationen des Modells in Form nicht angenommener Wirkungszusammenhänge können lokalisiert werden. Hierzu dient im Rahmen des EQS-Programms der Lagrange-Multiplier-Test (LM-Test, vgl. Bentler, 1992), der neben univariaten insbesondere multivariate χ^2 -Statistiken berechnet, die es erlauben, die Angemessenheit bisheriger Restriktionen des Modells zu überprüfen.

Der LM-Test kann also dazu benutzt, im Fall eines nicht passenden Modells mögliche Fehlspezifikationen zu analysieren. Allerdings sollte von einer nachträglichen Veränderung des Modells nur äußerst vorsichtig Gebrauch gemacht werden, da dies immer eine Anpassung des Modells an die Daten bedeutet und Modelltests, die ja dazu dienen, die Passung des Modells zu den Daten zu beurteilen, dann streng genommen nicht mehr gültig sind. Bei einer nachträglichen Anpassung des Modells an die Daten ist nicht mehr garantiert, dass die damit "entdeckten" Zusammenhänge (was ihre Größe und Signifikanz betrifft) nicht durch Eigentümlichkeiten der gerade vorliegenden Stichprobe bedingt sind – korrekterweise würde man dann eine Replikation des auf diese Weise konstruierten Modells anhand einer neuen, unabhängigen Stichprobe fordern müssen.

Literatur

- Bentler, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238-246.
- Bentler, P.M. (1992). *EQS: Structural Equations Program Manual*. Los Angeles, CA: BMDP Statistical Software.
- Bentler, P.M. & Bonett, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.
- Bentler, P.M. & Weeks, D.G. (1979). Interrelations among models for the analysis of moment structures. *Multivariate Behavioral Research*, 14, 169-185.
- Bentler, P.M. & Weeks, D.G. (1980). Linear structural equations with latent variables. *Psychometrika*, 45, 289-308.
- Bollen, K.A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Byrne, B.M. (1994). *Structural Equation Modeling with EQS and EQS/Windows. Basic Concepts, Applications, and Programming*. Thousand Oaks, CA: Sage.
- Cohen, J. & Cohen, P. (2003). *Applied multiple regression / correlation analysis for the behavioral sciences*. Mahwah, NJ (3rd ed.): Erlbaum.
- Gerbing, D.W. & Anderson, J.C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In K.A. Bollen & J.S. Long (Eds.), *Testing Structural Equation Models* (pp. 40-65). Newbury Park, CA: Sage.
- Hu, L. & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Jöreskog, K.G. & Sörbom, D. (1989). *LISREL 7: A Guide to the Program and Applications* (2nd ed.). Chicago, IL: SPSS Inc.
- Loehlin, J.C. (1998). *Latent variable models: An introduction to factor, path, and structural analysis*. Mahwah, NJ (3rd ed.): Erlbaum.
- Marsh, H.W., Balla, J.R. & McDonald, R.P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, 103, 391-410.
- Tanaka, J.S. (1993). Multifaceted conceptions of fit in structural equation models. In K.A. Bollen & J.S. Long (Eds.), *Testing Structural Equation Models* (pp. 10-39). Newbury Park, CA: Sage.