

## 為何需要結構方程模式及如何建立潛伏變項？

侯傑泰

香港中文大學教育學院教育心理系

結構方程 (SEM) 已成多變量分析的重要技巧, 本文檢視 SEM 較傳統分析優越之處, 並討論如何用單指標變項、項目對 (或小組) 及單同源分子測量模式去建立潛伏變項; 此外亦簡略探討偏態序級量表、最少樣本數目及吻合指數等課題。

Structural equation modelling (SEM) has become a very important technique in multivariate analysis. The present article examines the advantages of SEM over traditional analytical methods. Construction of latent variables using single indicator, item pairs/parcels, and uni-factor congeneric factor model are discussed with concrete examples. Furthermore, issues related to skewed ordinal scales, minimum sample size, and choice of goodness of fit indexes are reviewed.

結構方程 (SEM, Structural Equation Modelling)、協方差結構模式 (covariance structure modelling、LISREL等類似名詞) 已漸流行, 並成爲一種十分重要的數據分析技巧; 在大學高等學位研究課程, 它是多變量分析 (multivariate analysis) 的重要課題; 比較重要的社會、教育、心理期刊, 也特闢專刊介紹 (如: Connell & Tanaka, 1987; Joreskog & Sorbom, 1982); 可見SEM在統計學中所建立的聲望及崇高地位, 無容置疑的。

### SEM 簡介

很多社會、心理等變項, 均不能準確地及直接地量度, 這包括智力、社經階層、學習動機等, 我們只好退而求其次, 用一些外顯指標 (observable indicators), 去反映這些潛伏變項。例如: 我們以學生父母教育程度、父母職業及其收入 (共六個變項), 作爲學生家庭社經地位 (潛伏變項) 的指標, 我們又以學生中、英、數三科成績 (外顯變項), 作爲學業成就 (潛伏變項) 的指標。

簡單來說SEM可分測量 (measurement) 及潛伏變項 (latent variable) 兩部份。測量部份就是求出六個社經指標與社經地位 (或三科成績與學業成就) (即外顯指標與潛伏

變項之間) 的關係; 而潛伏變項部份則指社經地位與學業成就 (即潛伏與潛伏變項間) 的關係。

指標 (外顯變項) 含有隨機 (或/及系統) 性的量度上誤差, 但潛伏變項則不含這些部份。SEM可用以下矩陣方程表示 (Joreskog & Sorbom, 1993a; McDonald 1985):

(a) 對於潛伏變項 (例如社經地位與學業成就) 的關係, 即潛伏變項部份:

$$\eta = \beta \eta + \Gamma \xi + \zeta$$

$\eta$  ——內生 (依變) (endogenous, dependent)

潛伏變項 (如: 學業成就)

$\xi$  ——外源 (自變) (exogenous, independent)

潛伏變項 (如: 社經地位)

$\beta$  ——內生潛伏變項間的關係 (如: 學業成績與其他內生潛伏變項的關係)

$\Gamma$  ——外源變項對內生變項的影響 (如: 社經地位對學業成就)

$\zeta$  ——模式內未能解釋部份 (即模式內所包含的變項及變項間關係所未能解釋部份)

(b) 對於指標與潛伏變項 (例如: 六個社經指標與社經地位) 間的關係, 即測量模式部份:

$$X = \Lambda_x \xi + \delta$$

$$Y = \Lambda_y \eta + \varepsilon$$

X, Y是外源（如：六項社經指標）及內生（如：中、英、數成績）指標。

$\delta$ ,  $\varepsilon$  是XY測量上的誤差。

$\Lambda_x$  是X指標與  $\xi$  潛伏變項的關係（如：六項社經地位指標與潛伏社經地位的關係）。

$\Lambda_y$  是Y指標與  $\eta$  潛伏變項的關係（如：中英、數成績與學業成就間關係）。

在一典型分析過程中，我們輸入：各指標變項的協方差矩陣（covariance matrix）、總受試人數、指標與潛伏變項的從屬關係（指標如何歸屬於各潛伏變項）、程式（如：LISREL）會估計指標與潛伏、潛伏與潛伏、模式未能解釋部份、指標測量上誤差等指定的各參數，其數值亦反映各關係的強弱。此外程式亦計算研究者所提出的模式，是否與樣本數據吻合（即數據是否可用模式表示）（詳見下面吻合指數一節）。

## 為何要SEM

傳統迴歸（regression）假設潛伏變項能準確地以指標變項測量，在實際情況，這假設罕能達致。測量上誤差對整個模式各項參數的估計，有極複雜的影響，下文將以簡化例子，說明忽略這誤差可能引致的錯誤（Bollen, 1989）。

### （一）變異量

假設對於 X 指標（無誤差變項）， $\xi$  為容許測量誤差的潛伏或指標變項， $\delta$  為誤差，可導出：

$$\text{VAR}(X) = \text{VAR}(\xi) + \delta^2$$

X 的變異量 =  $\xi$  變異量 + 誤差變異量

除非  $\delta$  等於零，否則容許誤差變項的變異量永遠小於無誤差變項的變異量，故在傳統迴歸方法（假設無測量誤差），我們是傾向高估變項的真正變異量。

### （二）相關及迴歸參數

假設我們找 XY 間之關係， $\xi$  及  $\eta$  為對

應容許測量誤差的潛伏變項，以傳統迴歸方法得 XY 關係為  $\gamma^*$ （即 beta 值），而用容許測量誤差的  $\xi$   $\eta$  間關係為  $\gamma$ ，可導得：

$$\gamma^* = \gamma (\text{var}(\xi) / \text{var}(X))$$

只有當  $\xi$  完全沒有測量誤差時（即  $\text{var}(\xi) = \text{var}(X)$ ）， $\gamma^*$  才等於  $\gamma$ ；否則右邊括號內的值小於 1，即  $\gamma^*$  小於  $\gamma$ ，換言之用傳統迴歸法是低估了兩潛伏變項的真正關係。

若  $\gamma_{xy}$  為 XY 間相關係數， $\gamma_{\xi\eta}$  為  $\xi$   $\eta$  之間的相關， $\gamma_{xx}$  及  $\gamma_{yy}$  為 X 及 Y 的信度（reliability），則

$$\gamma_{\xi\eta} = \gamma_{xy} / (\gamma_{xx} \gamma_{yy})^{0.5}$$

故傳統相關係數亦是低估潛伏變項的真正相關係數。假設 XY（無測量誤差）相關係數  $\gamma_{xy} = 0.5$ ，而 x,y 信度  $\gamma_{xx} = \gamma_{yy} = 0.7$ ， $\xi$  與  $\eta$  相關  $\gamma_{\xi\eta} = 0.5 / (0.7^2)^{0.5} = 0.71$  顯而易見， $\gamma_{\xi\eta}$  比  $\gamma_{xy}$  大很多。

由上述簡化例子可見，忽視測量誤差可導致高估變項變異量及低估相關係數。對於較複雜的模式，則較難簡單估計忽視測量誤差可能帶來的影響。總括而言，當測量誤差愈大，傳統方法導致的錯誤便愈大。

## 建立潛伏變項

假設研究者已依學理，建立了一個或數個指標，以量度某一潛伏概念，下述將以 LISREL 軟件為例，討論研究者所遇到的一些困難及解決方法，對於一般性測量模式，可參考 Joreskog & Sorbom (1993a)，在此不贅（亦可參考 Bagozzi & Heatherton, 1994）。

### （一）單指標變項

若果潛伏變項只得單一指標，我們無法同時估計潛伏變項與指標的關係，及指標的誤差值，當然我們可以假設  $\lambda = 1$  及  $\delta = 0$ ，但這表示指標並無測量誤差；較合宜的做法是粗略估計指標的信度，以此計算  $\delta$  值，要求程式將  $\delta$  值固定在這數值（ $\lambda$  值自然也決定了）（Joreskog & Sorbom, 1988）。

假設  $\text{VAR}(X) = 10$ ,  $X$  信度  $r_{xx} = 0.8$   
 $\delta^2 = \text{VAR}(X) * (1 - \text{信度}) = 10 * (1 - 0.8) = 2$   
 我們要求 LISREL 將  $\delta^2$  固定於 2。

有時（如縱貫性研究）我們應容許不同時間的相同變項誤差，有一定程度的相關，在單指標模式中，這相關量也只好固定於一估計值，一般是以單變項誤差中的一固定百分比計算。

例如：在中三、四數學成績縱貫性研究，每次成績均以期考分數為指標，設信度為 0.8，而誤差相關為 10%，則這誤差量為兩次分數協異量（covariance）的 2%〔即  $(1 - 0.8) \times 10\%$ 〕。

在計算整個模式時，我們宜做一敏感度測試（sensitivity test），查看不同信度值對各參數的影響。若各分析結果類同，則其結論可信大增；若不同信度衍生不同模式，則研究者應更小心討論及比較不同模式的異同（如：見 Marsh, 1988 對 Newman, 1984 研究的再分析）。

## (二) 項目對／小組 (item pair / parcel)

在一些研究，我們發覺每一潛伏變項有甚多（甚至過多）指標，一個有效可行做法，就是將指標分成項目對（兩題一組）或項目小組（數題一組），將眾多項目合併成較少（通常多於三）變項。

這方法的優點包括（Marsh, 1990; Marsh & O'Neill, 1984）：(a) 受試者與項目的比例倍增（因項目數目相對減少）；例如：原本是 400 受試者對 20 變項，用項目對後便成為 400 受試者對 10 變項）；(b) 項目對（或小組）的信度增加，令項目對的誤差（或獨特變異，unique variance）較獨立項目為小；(c)  $\Delta$  係數（潛伏變項與指標之間的關係）較為不受指標的獨特隨機變異（idiosyncratic variance）影響；(d) 採用項目對（或小組）令變項數目減少，減低運作成本；(e) 減低受軟硬件的可能限制（分別限制變項數目及記憶體容量）。

不過項目對亦有其缺點，部份項目的特性可能會在合併後消失，此外用於配對的項

目，必須有高度內在一致性（homogenous）。

將項目合併時，若不同項目與其他變項有明顯不同關係，則我們不應將項目隨便合併。假設我們研究學習態度與學業成績關係，若學習態度項目包括對中、英、數的態度，而學業成績亦包括各次中、英、數的測驗分數；不合宜的做法是將各科態度項目，隨機胡亂合併；適當方法是將項目依科目合併，那麼我們除了可以研究一般學習態度與成績的關係，亦可以分析中文態度與中文成績等的特定關係，將各類題目胡亂分組（如：Randhawa, Beamer & Lundberg, 1993），並不合宜。

## (三) 單同源分子測量模式 (one factor congeneric model)

另一個處理大量指標變項的方法，就是用因子數值（factor score）（Smith-Holmes & Rowe, 1994），將某潛伏變項的所有指標合併為一變項（因子數值），再以上述單指標變項的技巧處理。

假設潛伏變項校長領導才能由五指標（五題目：可親性、精明、支持、溝通、可靠）量度，首先我們用 LISREL 的同源測量模式（congeneric model）要求程式計算因子比重（factor weight）（在程式OU指令內加FS），假設比重分別為 .14, .07, .37, .21, .21；新的合併指標 領導 = .14可親性+.07精明+.37支持+.21溝通+.21可靠。

這方法將五指標變成單一合併指標，故可用前述單指標變項技巧處理，而新合併指標的估計信度，則採用 LISREL 的總決定係數（total coefficient of determination）（LISREL 7 有這係數的計算）。

顯而易見，相對於項目對（小組）辦法，單同源分子模式將指標合併的程度更大，前者是組成項目對，而後者則合併為一變項，故此項目特性在後者損失更大。

因合併後只得一變項，很多誤差的相關變異量（例如：在縱貫性研究中），只得倚靠研究者自行粗略估計，不能由 LISREL 程式計算。不過雖然有上述各項缺點，對於涉

及極多變項的模式，這方法能將模式大大簡化，因變項減少，相對而言，研究所要求的最低受試者人數，也得以相應減少。

順帶一提，在合併指標時，若只計算各題的簡單平均，即：領導 = (可親性+精明+支持+溝通+可靠) / 5，不用因子比重，是否更好？

傳統想法是簡單平均法與樣本特性無關 (sample-independent)，故更為可取，但為此可能要付出沉重代價，例如：簡單平均法假設每指標的測量是同等準確，誤差變異量相同，但若潛伏變項與各指標的關係並不一致 (即  $\lambda_1 \neq \lambda_2$  等)，則用傳統方法所得的信度 Cronbach alpha，只是真正信度的一個最低估計值 (lower bound)；再者，若各指標負重真的不同，用簡單平均法所得的潛伏變項，並不能作為真正潛伏特質的有效 (valid) 指標；即用平均法所得領導得分，並非真正量度領導才能。

### 最低樣本數目

一般來說，在傳統分析 (例如迴歸法)，我們希望受試者與變項數目的比例最少是 10；即有 10 變項，則應最少有 100 受試者；在 SEM 方程，這法則並不合宜 (Tanaka, 1987)。

首先，在迴歸方程，我們只估計各自變 (independent) 與依變 (dependent) 項的關係，但在 SEM 內，我們不單估計各潛伏變項間及潛伏變項與指標間關係，我們亦估計各誤差及模式未能解釋部份的數值，故可變動參數的數目，比變項數目本身，更能決定受試者數目的最低要求。

其次由於 SEM 模式估計參數用漸近 (asymptotic) 特性，若用傳統誤差法去衡量各參數，則需極大的樣本。研究者用蒙特卡羅 (Monte Carlo) 法，檢視樣本大小對參數穩定性的影響，一些學者認為樣本最少應為 200 名，亦有建議受試者應最少為變項 20 倍，即 20 指標變項需 400 受試者。不過在某些特殊情況，模式參數較多時，更易得到合宜解答；故此最低受試人數，變項及參數的數目

間關係，可能頗為覆雜，難有簡單絕對法則可循 (Gerbring & Anderson, 1985; Holmes-Smith & Rowe, 1994; Joreskog & Sorbom, 1993a; Marsh, 1989; Tanaka, 1987；亦見本文吻合指數一節)。

### 偏態序級 (skewed ordinal) 量表

在頗多心理、教育研究中，我們用三、五或七點黎克型 (Likert) 量表，這些數據很多時未能符合常態分佈的要求，序級量表常令分佈嚴重偏態，因而低估參數的標準誤差，導致用卡方 (chi-square) 檢定時高估模式的吻合性 (fitness)。為克服這些問題，學者 (Bernstein & Teng, 1989; Browne, 1984; Joreskog & Sorbom, 1988) 提出用不依漸近分佈 (asymptotic distribution free, ADF) 的估計法。

在用這方法作分析時，我們必先計算指標間的 polychoric 相關 (PM) (因序級量表，故不用 Pearson product moment)，及其 asymptotic covariance (ACM) (因偏態故用 ACM) 矩陣。上述兩矩陣均容易由 PRELIS (與 LISREL 一起用的套件) 獲得；隨著再用 LISREL 程式內的 General weighted least square (WLS) (不用 maximum likelihood ML) 方法估計各參數及模式的吻合指數。

雖然利用 PM、ACM 矩陣及 WLS 方法於偏態序級變項，我們可獲穩定的參數估計，不過這方法要求較大的樣本，任何少於 12 變項的模式，需要最少 200 名受試者，多於 12 項時 (假設為 k 項)，則需要  $1.5k \times (k+1)$  人。例如：20 指標變項，則需要  $1.5(20)(21)=630$  人。

值得一提的是 ACM 內元素 (不重複者)，隨 k 增長甚速，20 變項的 ACM 含 22155 個不同元素 (k=10 時只得 1540 元素)，故檔案所佔記憶體積頗大。

這方法需要較大的樣本，對一些多變項的研究，前述之同源測量模式便顯出其功用。首先我們先對每一潛伏變項，用 ACM 及 PM 將各指標合併成單一變項 (項目對或小組)，然後再分析潛伏變項間的關係。

例如：先用 PRELIS 求六種社經狀況指標的 ACM 及 PM，再用 LISREL 將六變項合成一

新指標（或項目對／小組），再用同一（或類似）方法處理三科學業成績，成一新的學業指標，然後再用LISREL求出學業與社經（及其他）潛伏變項的關係，此時若有需要，而樣本數目又足夠的話，亦可用ACM協助。

### 吻合指數（goodness of fit indexes）

當我們用LISREL測試某一模式時，其實我們在研究自己所提出的模式（即那些變項之間是有關，那些則沒有），是否與數據吻合。

LISREL所輸入的是指標變項的協方差矩陣（S, sample covariance matrix），而依我們指定（apriori）的模式，計算出一個最佳的衍生矩陣（E, reproduced/fitted covariance matrix）；E與S接近則表示我們建議的模式成立，若E與S差異大，則表示模式與數據不符；吻合指數是用於反映E與S差異的一個總指標。

例如：我們有A、B、C、D、E、F六潛伏變項，我們建議的模式是：A、B是有相關；而A、B引起C、D；C、D則導致E、F。假設S是所有指標變項（構成A、B、C、D、E、F的所有指標）的協方差矩陣（covariance matrix），而E則是LISREL依上述模式估計出的最佳衍生矩陣，若吻合指數高則表示E與S差異甚小，反之，則E與S差異甚大。

其實除了一些吻合指數外，我們還應檢視：

- (a) LISREL的解答（solution）是否適當的（proper, well defined），這包括：重複漸進估計法能匯集（iterated estimate converges），各參數是在合理範圍內（例如：相關係數在+1與-1之內，誤差的值是零或正數），標準誤差的值是合理等。
- (b) 參數與實際預定（substantive apriori）模式的關係是合常理的。當然數據分析可能出現一些我們預期以外的結果，但各參數絕不應出現一些互相矛盾，與我們先置的假設有嚴重衝突的現象。

(c) 將模式的卡方（chi-square）及其他吻合指數，與其他可能的模式比較。

簡單來說，卡方測試是將指定模式，與一個全無限制（unconstrained）模式作比較，卡方數值愈大，則指定模式愈不符合研究樣本的數據。不過卡方數值與樣本大小（N）有直接關係，當N足夠大時，所有建議模式均難以通過；反之當N足夠小時，任何建議模式均好像與數據吻合。

例如我們的研究有數千人，則無論我們建議何種模式，去解釋社經地位（及其他變項）與學業成績關係，用卡方檢視結果均指模式不吻合數據，但若我們的研究只有數十人，則無論我們提出任何模式，卡方檢視結果均指出模式吻合數據。

正因如此，學者提出各種不同吻合性指數，希望能用簡單方法，指示出建議模式是否與樣本的數據吻合。尤其是當我們依據LISREL分析結果（例如用修正係數，modification index）不斷增加參數項目，企圖令模式與數據更吻合時，這可能導致一些虛假及偶然性的吻合。

故此吻合指數應對複雜模式作某種程度的懲罰，簡單但與數據吻合的模式才算有高吻合度。總括來說，好的吻合指數，不單不應受樣本數目影響，更應依模式內參數的數目而作調整，不偏幫參數眾多的模式（即懲罰多參數模式。）

例如：當我們研究A、B、C、D、E、F的關係時，這吻合指數不應依研究人數而有太大變化，又不應因我們大量增加模式中各變項間的可能關係（各變項之間多有關係），而認為這是較佳的模式。

學者提出頗多吻合指數的方程式（Bentler, 1990; Bentler & Bonett, 1980; Bentler & Mooijaart, 1989; Bollen, 1989, 1990; Mulaik, James, Alstine, Bennett, Lind & Stilwell, 1989），亦有對其作實驗性及數學性的檢視（Marsh, Balla & McDonald, 1988; McDonald & Marsh, 1990），結果雖然顯示一些指數（如：RNI, TLI）確有較佳表現，不過較合宜的做法是對每一模式，同時檢查多個指數及上述其他參數的合理性，作一較全面的比較。

## 總結

SEM是一應用廣泛，有巨大影響力的分析方法，它不單可應用於指標與潛伏變項、潛伏與潛伏變項間的關係，這技巧亦可用於因素（factor）、多組比較（multiple group comparison, Watkins, 1989）、縱貫性數據、傳統迴歸、多特質多法（multitrait multimethod）等分析上；若結合多層分析法（Hau & Chung, 1994），更可處理有階層特質的數據。

本文只是簡單描述，處理測量模式指標變項的一些方法，並介紹吻合指數等部份與基本分析有關的問題。因這方法仍屬較新的技巧，就算是在重要的學術期刊，仍偶見有嚴重錯誤運用的例子（如在 Randhawa et al., 1993, 見Marsh, Hau, Roche, Craven, Balla, McInerney, in press 的批評）（Breckler, 1990; MacCallum, Wegener, Uchino & Fabrigar, 1993）。尤其是當軟件的設計日趨簡單易用時（如 LISREL 8 可用 SIMPLIS 較簡單的程式語言），使用 SEM 的研究者又豈能不加倍小心呢？

## 參考文獻

- Bagozzi, R. P., & Heatherton, T. F. (1994). A general approach to representing multifaceted personality constructs: Application to state self-esteem. *Structural Equation Modeling, 1*, 35-67.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238-246.
- Bentler, P. M., & Bonnett, D. G. (1980). Significant tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588-606.
- Bentler, P. M., & Mooijaart, A. (1989). Choice of structural model via parsimony: A rationale based on precision. *Psychological Bulletin, 106*, 315-317.
- Bernstein, I. H., & Teng, G. (1989). Factoring items and factoring scales are different: Spurious evidence for multidimensionality due to item categorization. *Psychological Bulletin, 105*, 467-477.
- Bollen, K. A. (1989). *Structural equations with latent variables*. NY: John Wiley & Sons.
- Breckler, S. J. (1990). Applications of covariance structure modeling in psychology: Cause for concern? *Psychological Bulletin, 107*, 260-273.
- Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research, 18*, 115-126.
- Connell, J. P., & Tanaka, J. S. (1987). Introduction to the special section on structural equation modeling. *Child Development, 58*, 2-3.
- Hau, K. T., & Chung, C. M. (1994). Multilevel analysis: Why and how. Manuscript submitted for publication.
- Gebring, D. W., & Anderson, J. C. (1985). The effect of sampling error and model characteristics on parameter estimation for maximum likelihood confirmatory factor analysis. *Multivariate Behavioral Research, 20*, 255-271.
- Holmes-Smith, P., & Rowe, K. J. (1994). The development and use of congeneric measurement models in school effectiveness research: Improving the reliability and validity of composite and latent variables for fitting multilevel and structural equation models. Paper presented at the 1994 International Congress for School Effectiveness and Improvement, Melbourne, Jan 3-6.
- Joreskog, K. G., & Sorbom, D. (1982). Recent developments in structural equation modeling. *Journal of Marketing Research, 19*, 404-416.
- Joreskog, K. G., & Sorbom, D. (1988). *LISREL VII: A guide to the program and applications*. Chicago, IL: SPSS.
- Joreskog, K. G., & Sorbom, D. (1993a). *Lisrel 8 user's reference guide*. Chicago, IL: Scientific software international.
- Joreskog, K. G., & Sorbom, D. (1993b). *PRELIS 2 user's reference guide*. Chicago, IL: Scientific software international.
- Joreskog, K. G., & Sorbom, D. (1993c). *Structural equation modeling with the SIMPLIS command language*. Hillsdale, NJ: Lawrence Erlbaum.
- MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin, 114*, 185-199.
- Marsh, H. W. (1988). Causal effects of academic self-concept on academic achievement: A reanalysis of Newman (1984). *Journal of Experimental Education, 56*, 100-103.
- Marsh, H. W. (1990). *Self-description questionnaire — 1 (SDQ) manual III*. Sydney, Australia: University of Western Sydney.
- Marsh, H. W. (1989). Confirmatory factor analyses of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement, 13*, 335-361.
- 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.
- Marsh, H. W., Hau, K. T., Roche, L., Craven, R., Balla, J., McInerney, V. (1994). Problems in the application of structural equation modeling: A critique of Randhawa, Beamer and Lundberg (1993). *Journal of Educational Psychology*, in press.
- Marsh, H. W., & O'Neill, R. (1984). Self Description Questionnaire III: The construct validity of multidimensional self-concept ratings by late adolescents. *Journal of Educational Measurement, 21*, 153-174.
- Martin, J. A. (1987). Structural equation modeling: A guide for the perplexed. *Child Development, 58*, 33-37.
- McDonald, R. P. (1985). *Factor analysis and related methods*. Hillsdale, NJ: Lawrence Erlbaum.
- McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin, 107*, 247-255.
- Mulaik, S. A., James, L. R., Alstine, J. V., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin, 105*, 430-445.
- Randhawa, B. S., Beamer, J. E., & Lundberg, I. (1993). Role of mathematics self-efficacy in the structural model of mathematics achievement. *Journal of Educational Psychology, 85*, 41-48.
- Tanaka, J. S. (1987). "How big is big enough?": Sample size and goodness of fit in structural equation models with latent variables. *Child Development, 58*, 134-146.
- Watkins, D. (1989). The role of confirmatory factor analysis in cross-cultural research. *International Journal of Psychology, 24*, 685-701.