

脈絡變數的多層次潛在變數模式分析： 口試評分者效應的多層次結構方程模式應用

邱皓政

國立中央大學企業管理學系

論文編號：06027；初稿收件：2006年7月5日；第一次修正：2007年4月14日；完成修正：2007年6月24日；
正式接受：2007年9月7日
通訊作者：邱皓政 桃園縣中壢市中大路300號 (E-mail: hawjeng@ncu.edu.tw)

潛在構念的估計是社會與行為科學研究的重要命題，結構方程模式取向是當代據以進行潛在構念估計與關係探究重要典範，但是，如果研究樣本涉及叢集抽樣而呈現階層結構，研究者即必須進行多層次結構方程模式分析 (MSEM)。階層資料的重要特性是低階層次變數在高階會形成脈絡變數，因此 MSEM 分析必需能夠處理潛在脈絡變數模型的萃取與模型評估。本文的目的，即在說明如何利用 MSEM 程序來進行潛在脈絡變數的因素模型與結構模型分析，並以一項口試實證資料 (76 位教師甄試口試委員針對 841 位考生的口試評分資料)，以 Mplus4 進行單層次與多層次結構方程模式分析。分析結果證實了 MSEM 程序可以有效分離考生共變矩陣與評審共變矩陣，並證實了個體層次與總體層次的潛在因素模型並不相同，為一非等值模型 (non-equivalent model)。個體層次測量變數在評審層次是一個單一完整的潛在構念，亦即方法效應或月暈效果的概念，但在考生層次則仍保有原先的二因素架構。本文的結論指出，MSEM 比傳統的 MLM 取向在處理脈絡變數上更具有彈性，在方法學上具有重要的價值，但是在技術層次仍有諸多有待突破的瓶頸，值得進一步深究。

關鍵詞：脈絡變數、潛在脈絡變數、多層次資料、多層次結構方程模式、評分者效應

潛在構念的測量是社會與行為科學研究最大的挑戰之一，尤其是當資料結構越趨複雜之時，潛在構念的定義與測量也就越加困難。傳統上，研究者多以因素分析 (factor analysis) 技術來萃取潛在構念，探索或確認潛在構念的結構，或進行構念效度的評估。近年來，則多以結構方程模式 (Structural Equation Modeling; SEM) (SEM, Bollen, 1989; Jöreskog, 1973) 來進行構念的估計與分析。SEM 的優勢是利用共變數矩陣同時估計技術，將潛在構念與變數關係在同一個模型架構下進行分析，但是特殊資料結構 (例如類別依變數、非線性模型、多層次資料) 的 SEM 分析的原理與技術問題，各種應用軟體如何有效運作，仍然有待探究與釐清。尤其是多層次資料的 SEM 分析，由於涉及多層次模式分析與潛在變數模型分析兩大技術系統的整合，在計量原理與分析實務上都是當代心理計量領域的重要議題 (Rabe-Hesketh, Skrondal, & Zheng, 2007)。

多層次資料與結構方程模式分析

多層次資料 (multilevel data) 是指研究樣本具有階層性 (hierarchical) 或叢集 (clustered) 的特徵，使得研究者所測量到的觀察值具有特殊的相依/隸屬/配對關係，造成樣本獨立性假設的違反與統計檢定的失效。常見的例子為家庭研究的子女夫妻嵌套 (nested) 在家庭之中，各家庭又嵌套在縣市地域之中；學生嵌套在班級之中，班級嵌套在學校之中；員工嵌套

¹ 本研究的資料數據係取自呂秋萍 (2005) 碩士論文中的部分資料，在此對於慨然提供論文數據進行分析示範表示感謝。

在部門之中，部門嵌套在組織之中；團隊成員嵌套在各團體中；縱貫研究的個體重複觀察嵌套在個體之中；或是本研究所檢驗的考生嵌套在評審（口試評分者）之中。

當研究數據具有多層次特性時，分析單位即存在著層次上的變化，不同層次間的變數對於依變數的影響亦涉及複雜的控制與調節關係，傳統的統計分析技術（例如迴歸、變異數分析）無法處理這類問題，必須採用多層次分析技術，否則將使分析數據遭到層次關係的混淆與研究結果的誤導。

基本上，多層次分析技術是延伸自線性迴歸的概念，將代表各階層的多組迴歸方程式組合成混合模型（mixed model），再以多元迴歸原理進行參數估計，稱為多層次線性模式（Multilevel Linear Modeling, MLM）（Tabachnick & Fidell, 2006; Curran, 2003; Snijders & Bosker, 1999）。經過了諸多學者的努力探究，近年來多層次資料的分析在原理與技術上都已有非常成熟的發展（參見溫福星，2006；Ferron, Dailey, & Yi, 2002; Ferron et al., 2004; Raudenbush & Bryk, 2002）。然而，在傳統的多層次模型分析方法學中，並沒有納入潛在構念的概念，在目前廣為流行的HLM軟體（Raudenbush, Bryk, Cheong, & Congdon, Jr. 2004）雖然可以處理因素的萃取，但是仍是受到諸多的限制（參見Tabachnick & Fidell（2006）的討論）。近年來多層次分析技術的主要焦點議題之一，就是在SEM的架構下來處理多層次的資料，進行多層次結構方程模式（Multilevel SEM, MSEM）（Heck, 2001; Jedidi & Ansari, 2001; Goldstein, 2003; Goldstein & Browne, 2001; Goldstein & McDonald, 2003; Hox, 2002; Jöreskog & Sorbom, 2004; Rabe-Hesketh, et al., 2007）。

基本上，在典型的SEM架構下，所有的樣本都被視為隨機取樣的獨立個體，經由研究者對於外顯變數背後所存在的理論上的潛在結構加以界定後，探討測量或潛在變數之間的結構關係。因此，如果樣本具有叢集結構，典型的SEM並沒有特殊的檢驗機制，使得多層次結構所反應的變數關係無法在假設模式中進行估計，也即因為資料本身的特徵沒有被納入模式設定中，可能導致重要解釋變數的疏漏，反應在誤差項上的特殊樣態（pattern），造成隨機常態分配的基本假設的違反或標準誤失真的情形發生。一般在SEM分析常會看到的誤差相關（correlated uniqueness），就是誤差間存有特殊樣態的隨機常態分配假設違反的典型現象。

由於傳統的MLM分析與SEM是兩套為了解決不同問題所獨立發展的統計技術，兩者各有優勢，但是當同一個資料庫，既具有多層次結構，又具有潛在

變數估計需求之時，兩種分析方法都只能解決部分問題，因此發展SEM的多層次分析技術（MSEM）可以讓這兩種重要的分析技術得以並用於同一個研究之中，同時解決多層次資料結構與潛在變數的估計問題（Bentler & Liang, 2003; Heck & Thomas, 2000; Muthén, 1989, 1994; Hox, 1995, 2002; Kaplan, 1995; Kalpan & Elliott, 1997）。本文的主要目的，即在檢視MSEM分析的原理，並進一步應用於潛在脈絡變數的估計與分析。

多層次模型與脈絡變數

在多層次資料結構中，最底層是由最小的分析單位所組成（例如個別的學生），稱為個體層次（micro level）。越高階的層次則分析單位越大，稱為總體層次（macro level），例如學生為第一層（個體層次），其所屬的「班級」屬於第二層（總體層次），班級所屬的「學校」屬於第三層（亦為總體層次）。在傳統的MLM模型中，依變數（或稱為效標變數或準則變數）是個體層次的觀察值，對於依變數進行解釋的預測變數（稱為解釋變數或自變數）可以存在於個體層次，也可以存在於總體層次，或同時存在兩個層次，用以探討不同層次解釋變數對於依變數的影響。

在MLM模型中，有一個特殊的變數形式稱為脈絡變數（contextual variables），亦即個體層次解釋變數透過組內聚合（aggregate）程序形成高階解釋變數時，稱為脈絡變數（Duncan, Curzort, & Duncan, 1966），例如學生IQ對於學業成績的影響，學生IQ雖作為個體層次解釋變數，但可聚合成為班級IQ（亦即求取全班學生IQ的平均數），此時的平均IQ即為脈絡變數，以「班級」為測量與分析單位。如果還有學校的區分，班級層次的脈絡變數可以再聚合成更高階的校級層次IQ。

脈絡變數的存在價值在於使研究者得以實證方法來進行脈絡分析（contextual analysis），避免生態謬誤（ecological fallacy）（Robinson, 1950; Snijders & Bosker, 1999）。因為脈絡變數反應了環境或背景的特徵，脈絡變數對個體的影響即是一種脈絡效果（contextual effects）。脈絡變數的一個方法上的限制，是個體層次的變數雖可以簡單的數學算則聚合成高階層次的脈絡變數，但是聚合後的變數背後的構念性質是否與個體層次測量構念仍然一致，還是產生了變化（構念偏移），則是一個根本的測量問題或構念效度問題。例如個體的成績聚合成總體層次的全班平均成績時，在個體層次與總體層次都是同樣的意義（學業能力指標），但個體層次的IQ與總體層次的IQ，可能

就反映了不同的構念。也就是說，同一個測量變數在不同階層自成不同分析單位的隨機變數，但未必反映相同的測量內容或構念。從數學的角度來看，高階變數觀察值為低階變數平均值，為高階變數的不偏估計數，但是從測量的觀點來看，總體層次的脈絡變數的構念性質卻必須重新經過構念效度的檢視。

過去文獻上對於脈絡變數的探討，多是在 MLM 的方法架構下，利用低階層次變數在高階所得到的平均值，形成脈絡變數後，檢驗其對於截距 (intercept) 或斜率 (slope) 造成影響 (Stephenson, 2006; De Silva, 2005)。國內最近的一篇論文 (邱皓政、溫福星, 2007)，即是以學校中的個別老師的組織創新氣氛知覺分數，聚合成學校層次的脈絡變數後，進而對於個別老師的績效表現造成影響，示範了脈絡變數的 MLM 分析。這些研究雖可利用 MLM 模型檢視脈絡變數對於個體的影響，但是卻未對於脈絡變數本身的潛在結構進行檢視。

基本上，MLM 模型中的變數為外顯變數 (manifest variables)，變數數值為沒有測量誤差的真實測量。脈絡變數由個體層次變數聚合成總體層次變數，本質上仍是外顯變數，變數所反應的內容無法以傳統心理計量中的效度檢驗程序來檢視，因此有學者質疑脈絡變數未必反應研究者所預設的環境或脈絡實體 (Macintyre & Ellaway, 2003; Macintyre, Ellaway, & Cummins, 2002; Raudenbush, 2003; Snijders & Bosker, 1999)。更進一步的，如果脈絡變數背後具有特殊的因素模型時，亦即脈絡變數背後存在特定的潛在構念時，脈絡變數更不宜以外顯變數來處理。

在心理計量領域，潛在構念的估計主要係透過因素分析模型來處理，當研究者對於因素結構抱持特定假設所進行的因素分析，稱為驗證性因素分析 (confirmatory factor analysis, CFA)，屬於 SEM 的一個次模型 (余民寧, 2006; 邱皓政, 2003; 黃芳銘, 2003)。若脈絡變數的構念本質或潛在結構遭到質疑時，最好的處理方法即是利用 SEM 來檢驗個體層次的因素模型與總體層次的因素模型的差異，但是前提是必須將 SEM 擴展到多層次架構，進行 MSEM 的分析。從測量的觀點來看，MSEM 取向的估計方法不但在理論層次符合脈絡變數的定義，亦即脈絡效應由個體層次累積產生，在總體層次形成另一個測量模式，得以進行另一層級 (學校、班級或組織) 的潛在變數萃取與因素結構分析，探究脈絡變數在總體層次所展現的構念是否仍為原始測量所定義的構念，是一種潛在脈絡變數模型 (latent contextual variable model) 的應用。以下，即對 MSEM 的統計原理進行說明。

多層次結構方程模式原理

多層次資料結構

單層次與多層次資料分析的主要差異，在於多層次資料結構係將低層次的個體，以某一個分組變數區分成不同的群體，然後分別對於分析變數 (Y) 的組間變異與組內變異進行分析。由於 SEM 與 MSEM 均以變數間的共變結構為分析材料，因此運算過程均以變數的離散分數 ($Y-\bar{Y}$) 來處理，亦即平減 (centering) 或離差分數 (deviated scores)，以小寫 y 表示 (Cronbach & Webb, 1979)。平減的重要功能是在進行量尺原點的平移，使變數的平均值為 0，令迴歸方程式的截距項能夠反映平均數。若將觀察值減去總平均數 ($Y-\bar{Y}_G$) 稱為總平減 (grand centering)，為不分組時，每一個觀察值的離散性，以 y_T 表示；若把觀察值減去各組平均數 ($Y-\bar{Y}_g$)，稱為組平減 (group centering)，反應同一組內的觀察值的個別差異 (以 y_w 表示)；各組平均數與總平均數的差 ($\bar{Y}_g-\bar{Y}_G$) 則為組間差異 (以 y_B 表示)。 y_T 、 y_B 、 y_w 三者均來自同一組 p 個觀察變數就組間與組內進行分割後的平減觀察值，各具有 p 個向量，其組合關係如下：

$$y_T = y_B + y_w \quad (1)$$

經過平減後的變數，量尺原點即平移至平均數位置，因此在 SEM 模式中進行聚合處理時得到的截距即等同於平均值。此外，由於組間與組內平減分數 (離差分數) 是正交分數，故可以導出兩個獨立且具有可加性的共變矩陣：組內共變矩陣 (S_w) 與組間共變矩陣 (S_B)：

$$S_T = S_B + S_w \quad (2)$$

由於從樣本觀察值所求出的 S_w 與 S_B 具有可加性，向量中各變數所求得的組內變異數總和與組間變異數總和分別是母體組內變異數 (σ_w^2) 與母體組間變異數 (σ_b^2) 的不偏估計數，下標為小寫的 w 與 b 表示為各變項自矩陣展開後累積所得之變項變異數純量，因此可計算各變項的組間變異數與總變異數的比值，亦即 ICC (組內相關係數, Intra-Class Coefficient)，且由於 ICC 是以觀察變數計算得出，因此本文特別將其稱之為外顯 ICC 係數 (manifest ICC; 以 ICC_M 表示)，反應觀察變數的組間異質性或組內同質性，公式如下：

$$ICC_M = \rho_M = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2} \quad (3)$$

ICC_M 代表測量變數的變異量中，組間差異的比例，亦即組間效果。當 ICC_M 甚小時，表示組間不明顯，多層次結構的影響可以忽略，以傳統方法即可處理，反之則表示組間差異不可忽略，必須以多層次分析技術來處理 (Roberts, 2002)。

多層次測量模型

因素變異拆解與潛在變數 ICC (ICC_L)

延續前述的變異數拆解與可加性原理，如果公式(2)當中的共變矩陣具有潛在變數的設定，測量分數的變異又可被切割成正交的兩部分：可被潛在變數解釋部分（潛在變數的變異數，以 Ψ 表示）與不可解釋部分（測量誤差，以 Θ 表示），兩者亦具有可加性；同理，因素萃取所剩餘的組間誤差 (Θ_B) 與組內誤差 (Θ_W) 亦具有獨立性與可加性：

$$\Psi_T = \Psi_B + \Psi_W \quad (4)$$

$$\Theta_T = \Theta_B + \Theta_W \quad (5)$$

由於各變數之間具有關連，因此相互獨立的組內矩陣與組間矩陣各自估計得出的組內與組間矩陣，可展開計算出各潛在變項累計總變異純量，進一步可計算潛在變數的變異數 Ψ_b 與 Ψ_w ，與佔總潛在變數變異的比例 (Muthén, 1994)。對於組間的因素變異 (Ψ_b) 佔因素全體變異 (Ψ_T) 的比例，即為潛在變數 ICC (ICC_L)：

$$ICC_L = \frac{\Psi_b}{\Psi_b + \Psi_w} = \frac{\Psi_b}{\Psi_T} \quad (6)$$

ICC_L 越大，表示潛在變數在組間（高層次）具有影響力，組間的因素結構是不可忽視的部分；反之， ICC_L 越小，表示潛在變數在組間的意義不大，因素結構在組內的意義較大。

樣本矩陣與估計矩陣

從樣本獲得的共變矩陣稱為樣本矩陣，由 SEM 假設模型所導出的共變矩陣稱為導出矩陣 (reproduced matrix, 以 Σ 表示)。在多層次的 MSEM 模型中，導出矩陣亦可拆解成組間估計矩陣 (Σ_B) 與組內估計矩陣 (Σ_W) 兩部分：

$$\Sigma_T = \Sigma_B + \Sigma_W \quad (7)$$

以典型的 SEM 矩陣關係式表述如下：

$$y_{gi} = \nu + \Lambda_B \eta_{Bg} + \varepsilon_{Bg} + \Lambda_W \eta_{Wgi} + \varepsilon_{Wgi} \quad (8)$$

其中 η_{Bg} 表示總體層次的因素結構， η_{Wgi} 表示個體層次的因素結構。與分別表示總體與個體層次的因素負荷量矩陣，為截距， ν 為誤差。 ε 各項下標當中的 i 表示各組內的個別觀察值，各組觀察值數目可能不相等； g 表示組數， $g=1$ 到 G 。

值得注意的是，由於公式(8)的誤差項被分割成組間與組內兩部分，因此可分別計算組間標準化殘差均方根指數 (SRMR_B) 與組內標準化殘差均方根指數 (SRMR_W)。過去學者主張 SRMR 是評估 SEM 模式適配的重要指標 (McDonald & Ho, 2002)，因此 SRMR_B 與 SRMR_W 可作為 MSEM 模型的組內與組間模式適配情形的重要參考指標。

從模式設定的原理來看，雖然公式(8)由組間與組內兩個部分組成，但總體與個體兩層次的潛在構念 (η_{Bg} 與 η_{Wgi}) 的結構可以相同，也可以不同。例如組內層次是三因素模型，組間層次可以是三因素或二因素模型；組內因素模型是斜交模型（因素間具有相關），而組間則為直交模型（因素間沒有相關）。利用 η_{Bg} 與 η_{Wgi} 兩個因素結構的相同與不同，以及 Λ_B 與 Λ_W 矩陣的比較，我們即可用來檢驗脈絡變數在總體層次的因素結構（構念恆定）是否與個體層次不同（構念偏移）。

值得注意的是，在多層次因素分析的估計過程中，組間與組內矩陣是分別由兩個樣本觀察矩陣 (S_B 、 S_W) 來分別估計。由於多層次資料具有叢集特性（個體層次觀察值被區分成不同的組），因此組內觀察矩陣 (S_W) 是由各叢集（組）的觀察矩陣合併，稱為整合組內觀察矩陣 (S_{PW}) (pooled within-group sample matrix)，用以估計組內導出矩陣 (Σ_W)：

$$S_{PW} = \frac{\sum_{g=1}^G \sum_{i=1}^{c_g} (y_{gi} - \bar{y}_g)(y_{gi} - \bar{y}_g)'}{N - G} \quad (9)$$

其中 c_g 表各組內的觀察值個數。另一方面，由組間資料估計得出的組間觀察矩陣，亦是由各組的 i 個個體，以分合 (disaggregated) 資料導出的平均數來計算得出，以 S'_B 表示：

$$S'_B = \frac{\sum_{g=1}^G c_g (\bar{y}_g - \bar{y})(\bar{y}_g - \bar{y})'}{G-1} \quad (10)$$

Muthén (1989, 1990) 指出，在多層次的資料結構中，總體層次的觀察資料是由個體層次聚合成平均數來表示（亦即脈絡變數），因此組間矩陣的估計數中帶有組內的資訊。此時， S_{PW} 矩陣是母體矩陣（ Σ_W ）的最大概似估計數，但 S'_B 矩陣不是組間矩陣（ Σ_B ）的最大概似估計數，而是組間（ Σ_B ）與組內母體矩陣（ Σ_W ）的線性整合加權最大概似估計數，如下列二式：

$$S_{PW} = \hat{\Sigma}_W \quad (11)$$

$$S'_B = \hat{\Sigma}_W + c_g \hat{\Sigma}_B \quad (12)$$

在多層次資料中，個體與總體層次的觀察單位不同，因此公式(12)中的各組人數 c_g 作為加權之用。當各組人數相同時， c_g 為固定常數，稱為平衡模式（balanced model）；如果各組人數不等的非平衡模式（un-balanced model）， c_g 為變數。換言之，當各組人數不同時，SEM 估計必須視每一組為不同的模型，才能利用完全訊息最大概似估計法（Full Information Maximum Likelihood; FIML）來估計參數，造成模式的極端複雜化。因此 Muthén (1989, 1990) 建議忽略各組人數差異的影響，改以非常接近平均組人數的事後估計組人數（ c^* ）來取代 c_g 權數，求取組間觀察矩陣，藉以簡化模型，達成模型收斂的目的。 c^* 定義如下：

$$c^* = \frac{N^2 - \sum_{g=1}^G c_g^2}{N(G-1)} \quad (13)$$

其中 N 為樣本總數， G 為組數。一旦使用公式(13)所進行的最大概似估計解並非 FIML 解，而為受限最大概似估計解（Limited Information Maximum Likelihood; LIML），此法 Muthén (1989, 1990) 稱之為 MUML 解（Muthén's ML），McDonald (1994) 則稱之為假性平衡解（pseudobalanced solution）。在組間與組內樣本數均很大的時候， S'_B 是非常接近 Σ_B 的估計數，但是樣本數小時，偏離情形越趨嚴重，參數估計數與標準誤越不真確。

根據 Hox 與 Maas (2001) 的模擬研究發現，在各組人數不等的情况下，假性平衡模式的組內觀察矩

陣估計數不會有偏誤的問題，但是組間矩陣的估計則受到組數大小的影響，在組數較低（50）且 ICC 偏低時，會發生不適當解與因素負荷量的估計偏差。而且不相等的組內樣本數會造成誤差變異數低估，標準誤縮小，導致過度拒絕虛無假設的型一錯誤膨脹問題。因此 Hox 與 Maas (2001) 建議各組人數應盡量相等，組數不宜在 50 以下，應維持在 100 以上。同時在各組人數不等時，應以更嚴格的型一誤差率（例如 $\alpha = .01$ ）來進行參數考驗。

多層次結構模型

在 SEM 分析架構下，CFA 模型的目的是在探討測量變數背後的潛在因素，一旦 CFA 模型證實了測量變數背後的潛在變數結構，即可納入迴歸分析的概念來進行潛在變數之間關係的探討，稱為結構模型分析。在 MSEM 模型中，學者對於多層次結構模型的定義方法各有不同的主張，但基本原理與前面所介紹的多層次測量模型的原理相似，亦是將模型分割成組間與組內兩部分，分別進行 Σ_B 與 Σ_W 的估計，不同點在於加入了結構模型的部分，分別從組間與組內兩層次來對 Λ_B 與 Λ_W 矩陣（迴歸係數）進行估計（Muthén, 1994; Lee & Shi, 2001）。此時由於參數增加，多層次模型又牽涉到不同組的組內估計，當組數越多，估計越趨複雜，因此，多層次的結構模型的分析，更必須建立在穩固的測量模型的基礎上，結構參數才有更適切的估計解，因此 MSEM 的分析仍應採 Anderson 與 Gerbing (1988) 所建議的兩階段程序，找出最適當的測量模型後，再接續進行結構模型分析。

根據典型的 SEM 模式設定，對於由外顯變數 X 與 Y 所構成的結構模型定義如下：

$$y = \alpha + By + \Gamma x + \zeta \quad (14)$$

公式(14)中， Y 變數為內衍變數（作為依變數）， X 變數為外衍變數（作為自變數）， B 為內衍變數之間的迴歸係數矩陣， Γ 為外衍變數對內衍變數的迴歸係數矩陣， α 為截距， ζ 為解釋誤差。如果資料具有多層次的特徵，公式(14)所表示的迴歸方程式可以進一步拆解成組間與組內方程式兩部分。Muthén 與 Muthén (2004) 將個體層次的組內模型變數以內衍變數來處理（各觀察變數由脈絡變數解釋），故可簡化成下式：

$$y_g = \alpha_g + By_g + \varepsilon_g \quad (15)$$

公式(15)稱為組內結構模式 (structural form of the within-group model; Kaplan, 1998; Kaplan & Elliott, 1997)，下標 g 表示參數係分就各組內的數據估計而得， ε_g 為組內迴歸的誤差項； α_g 為截距向量，由於資料在測量模型已經過平減處理，故反應各組在依變數上的組平均數。迴歸係數矩陣 (B) 在各組雖可能有不同的迴歸係數，但是在 MSEM 估計中以固定效果處理，即為單一的固定係數值，若以隨機效果處理將可能因為大量的疊代估計導致收斂不完全 (Rabe-Hesketh et al., 2007; Muthén & Muthén, 2004)。

如果模型中帶有潛在變數 (η)，組間與組內的結構模型定義如下 (組內模型的下標 g 省略)：

$$\eta_B = \alpha_B + B_B \eta_B + \zeta_B \quad (16)$$

$$\eta_W = B_W \eta_W + \zeta_W \quad (17)$$

公式(16)與(17)定義了組間與組內帶有潛在變數的矩陣關係，可以說是多層次結構模型的基本形式。值得注意的是，由於觀察變數經聚合程序成為組間的觀察值，因此組內結構模型係由平減數據 (組平均數為零) 估計而得，故為組內部分並沒有截距項。由此一模型出發，可以發展出其他不同的假設模型，有關的統計推導請參考 Hox (2002)、Muthén (1989, 1990)、Heck 與 Thomas (2000) 的介紹，本文則以實證的範例來說明多層次結構模型的設定與分析程序。

潛在脈絡變數模型的特性與模式設定

傳統 MLM 分析中的脈絡變數，是由個體層次變數聚合而成，在本質上是一種外顯變數 (亦即組平均數)，總體層次的脈絡變數的各觀察值 (各組平均數) 之間的變異即為組間變異。組間變異越大，ICC 越強。然而在 MSEM 的分析架構下，脈絡變數的定義方式則可能有外顯形式與潛在形式的不同。這是由於 MSEM 取向可以利用測量模型對於潛在構念的定義，以 CFA 策略來對脈絡變數的潛在結構進行檢驗，萃取出因素後進而對依變數進行解釋，本文將帶有潛在結構的脈絡變數模型稱為潛在脈絡變數模型 (latent contextual variable model)。

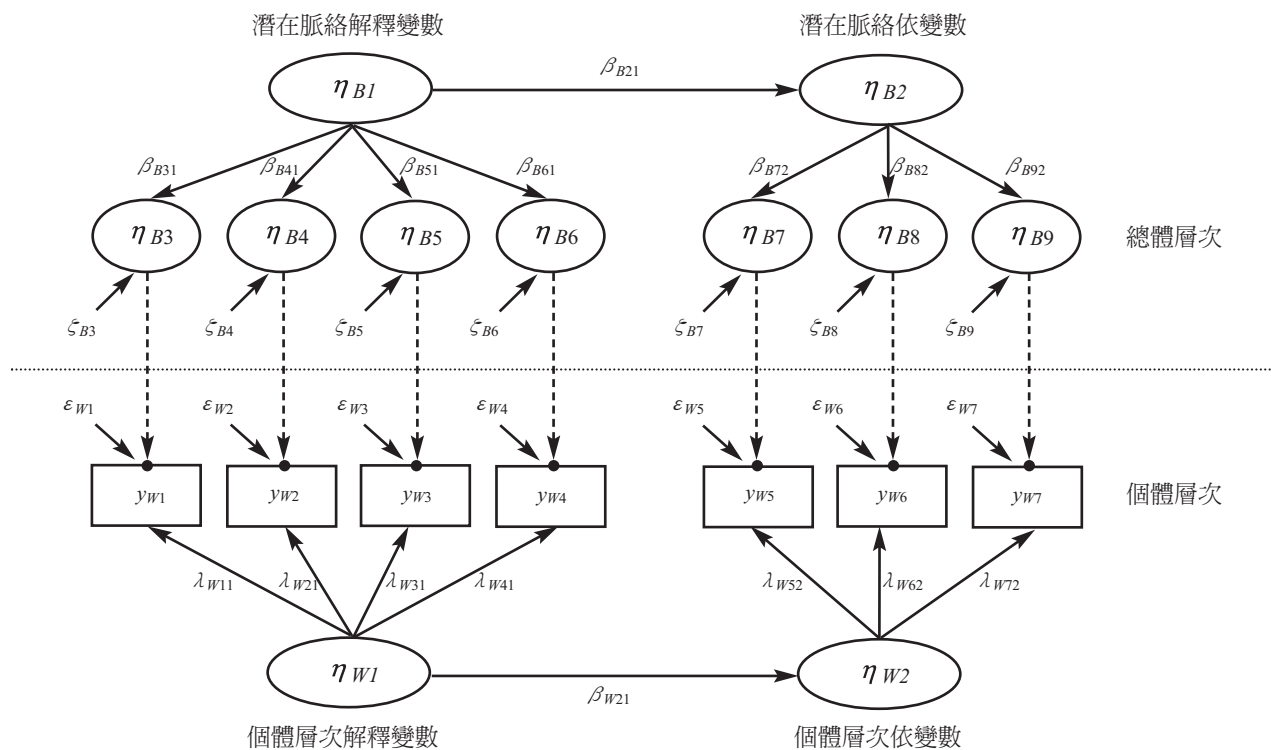
在傳統的 MLM 模型中，個體層次解釋變數聚合得到的總體層次變數的本質是組平均數，兩者的構念意義可以直接從測量的操作型定義來檢視，如果聚合平均數反應了脈絡的作用，即符合脈絡變數的定義。但是在 MSEM 模型中，由於帶有潛在變數的設定，因

此個體層次解釋變數的意義，必須透過背後的個體層次潛在變數來說明，同理，個體層次測量變數聚合成總體層次的脈絡變數後，也需透過測量模型來定義脈絡變數的意義，檢驗是否存在特定的因素結構，或是不具有潛在變數的單一指標變數，聚合後的變數結構若反應了脈絡的影響時，才符合脈絡變數的定義。由此可知，MSEM 取向對於脈絡變數的界定、解釋以及後續的預測分析，需透過驗證性因素分析程序來作為初始的確認程序，亦即因素效度 (factorial validity) 的檢驗。

MSEM 取向沿用 MLM 模型的定義方式，對於聚合平均數 (截距) 在組間的變異係以隨機效果處理，亦即隨機截距，也就是脈絡變數。利用 Muthén 與 Muthén (2004) 的圖示法 (見圖 1)，測量變數上所標示的實心黑點表示為隨機截距，為潛在變數。由此可知，在 MSEM 模型中，脈絡變數 (隨機截距) 的數學特性是帶有分配特性的潛在變數 (圖 1 中以虛線箭頭來對應觀察變數七個黑點的 η_{B3} 到 η_{B9} 七個潛在變數)，得以估計標準誤，並以 t 檢定來檢驗統計意義，如果隨機截距的變異數未達統計顯著水準 (或是 ICC_M 太小)，表示組間差異不明顯，不足以作為脈絡變數，亦無法在總體層次進行更高層次的因素分析萃取，因此，隨機截距的變異數顯著性考驗是脈絡變數檢驗的重要步驟。

如果隨機截距具有充分的變異，即可利用高階驗證性因素分析 (higher-order confirmatory factor analysis; HCFA) 原理，萃取出潛在高階因素 (Muthén & Satorra, 1989)。也就是說，各潛在脈絡變數的組間變化受到背後更高階的潛在構念的影響 (圖 1 中的 η_{B1} 與 η_{B2})，此時，高階潛在因素可能恰好對應了個體層次的潛在變數 (圖 1 中的 η_{W1} 與 η_{W2})，也是個體潛在變數在總體層次的潛在脈絡變數。然而，MSEM 模型的一個特性，是總體層次的因素結構不一定需與個體層次相同，各層的潛在變數數目與結構未必相等 (Hox, 2002; Heck & Thomas, 2000)，得以檢視總體層次與個體層次潛在因素結構的對應關係。如果兩個層次因素結構相等，此時可稱為是等值潛在脈絡變數模型 (equivalent latent contextual variable model)，如果兩個層次結構不相等，總體的脈絡效果與個體層次潛在變數無法直接對應，此時可稱為是非等值潛在脈絡變數模型 (non-equivalent latent contextual variable model)，亦即構念偏移現象的檢驗。

MSEM 取向的另一個特性，是不僅解釋變數得以進行總體層次的脈絡變數的定義與萃取，在依變數也同樣可以進行同樣的定義與萃取程序。當一個 MSEM 模型既有解釋變數的高層萃取，也有依變數的高層萃



圖一：多層次 SEM 脈絡變數模型的概念圖示

取，所形成的完整模型即可視為一般潛在脈絡模型 (general latent contextual variable model; GLCVM)。一個 GLCVM 的概念圖示與參數設定如圖 1 所示。

由圖 1 可知，左側的解釋變數在總體層次萃取出一個相對應的總體層次潛在脈絡變數，稱為潛在脈絡解釋變數 (latent contextual explanatory variable)，右側的依變數在總體層次同樣也萃取出一個相對應的總體層次潛在脈絡變數，因此可以稱為潛在脈絡依變數 (latent contextual dependent variable)。兩者可以作為結構模型的組成變數，進而進行潛在脈絡變數間的迴歸分析。

實證數據分析：評分者效應的潛在脈絡變數模型分析

研究樣本與工具

本研究所使用的資料庫為呂秋萍 (2005) 所蒐集的實證資料重新整理而得，該研究的原目的在探討教師甄選口試，口試評審委員與考生的背景特徵及口試評審委員心理因素如何影響教師甄選口試評價結果，尤其對口試評審委員類我程度在口試過程中所扮演的角色進行分析。研究對象是 93 年度北部三縣市國中

教師甄選口試評審委員 76 名與 841 位考生。76 位委員中，女性 31 人，男性 45 人，多為資深人員，20 年以上年資者 36 人 (佔 47.4%)，10 年以上者亦有 33 人 (佔 43.4%)，其中校長 5 人，主任 46 人，組長 11 人。考生部分共回收評量表 883 份，有效量表 841 份，其中女性考生 603 名 (佔 71.7%)，男性考生 238 名 (佔 28.3%)，年齡多為 30 歲以下 (佔 75.7%)；144 名 (17.1%) 為研究所畢業。實習結束初任教師者 294 名 (35%)，具有 1-2 年教學經驗者 349 名 (41.5%)，3-5 年者 140 名 (16.6%)，六年以上 58 名 (6.9%)。每位評審委員所面試的考生人數最少 3 人，最多 22 人，平均 11.07 人 (亦即 c^*)。

研究工具為「口試評量表」(Interview Assessment Form) (簡稱 IAF 量表)。此量表乃修訂自 Graves 與 Powell (1988) 所編製的評定表，原量表共包含 8 個因素 20 個題目，採李克特氏 (Likert-type) 五點量表記分。本研究取用資料庫當中的兩個因素「專業能力」(四題，Cronbach's $\alpha = .92$)，「類我程度」(三題，Cronbach's $\alpha = .89$)，依變數為三個有關「評價結果」的題目：「該考生表現很好」、「該考生適合當老師」、「可能錄取該考生」(Cronbach's $\alpha = .96$)，各分量表內部一致性均十分理想。

研究變數

本研究所分析的第一個潛在變數為考生的「專業能力」，亦即「具備擔任好老師的工作技巧和達成工作的專業能力」，係由口試之後，由口試委員主觀判斷給分，屬於口試評審委員對考生的主觀評價。一般而言，教師專業的評量是口試最重要的向度，內容包括教學方法好、具教學技巧、評分公平、教學清楚明白、生動活潑的教學與教材、課程安排有條理、幫助學生學習及自我成長、良好的溝通能力、口齒清晰表達能力好、對工作有熱誠、相關工作經驗及訓練、教師專業知識（謝臥龍，1997）。本研究取用「具相關工作經驗及訓練」、「具班級經營能力」、「具親師生溝通能力」、「具教師專業知識」等四題來評定。

第二個潛在變數為口試評審委員對於考生的「類我程度」的評估。文獻指出「類我程度」是為影響口試評價結果的重要因素（Graves & Powell, 1988）。本研究係以口試評審委員對於考生的「工作態度」、「對待學生的方式」、「教育理念」的相似程度的評定結果來表示，是解釋影響口試評審委員對考生口試評分的一個重要潛在因素（Anderson & Shackleton, 1990; Graves & Powell, 1988, 1995）。過去的研究發現，不論是履歷（Baskett, 1973）、錄影帶（Rand & Wexely, 1975）和真實口試情境（Anderson & Shackleton, 1990; Graves & Powell, 1988, 1995; Raza & Carpenter, 1987），都發現當口試評審委員和考生之間的態度、背景越相似，評審對於考生的口試評價越高。

本研究採用 Mplus4（Muthén & Muthén, 2004）進行 MSEM 分析，其優點是可以直接進行樣本共變矩陣的計算，不需要另行以其他軟體或計算程序來獲得 S_W 與 S_B 。本研究用於分析的組間與組內樣本共變矩陣列於附錄一，所使用的 MSEM 分析的 Mplus 語法附於附錄二。

結果與討論

變數描述統計與相關分析

表 1 列出了 76 位評審委員對於 841 位考生評定結果的描述統計量與相關係數。整體而言，評審委員對於 841 位考生在 10 個 IAF 評定量表的評分水準居中，平均數介於 3.40 至 3.63 之間，各項目評分的變異一致，標準差介於 .75 至 .99 之間。題目間相關均達 .01 顯著水準，係數介於 .49 至 .87，其中三個結果評定（R1 到 R3）的相關更高達 .80 以上，顯示這三個題目間的一致性頗高。

另一方面，針對 76 位評審委員的總體層次變數的關係來看，可以發現 IAF 各題的聚合平均數的相關更高，多數係數皆高於 .80 以上，甚至三個結果評定（R1 到 R3）的相關更達 .90 以上，顯示評審委員對於考生的評分，在評審之間的關聯強度更甚於個別考生間的關聯。由於每位評審委員口試人數不同，因此相關係數利用口試人數予以加權，以分合資料來進行相關係數的估計。

驗證性因素分析

為了驗證「專業能力」、「類我程度」與「評定結果」三個潛在構念的因素效度，首先進行 10 個觀察變數的 CFA。由於資料結構具有多層次特性，CFA 的檢驗將先針對個體層次的 841 考生評定資料，進行單層次 CFA，用以確認考生層次的因素結構，再以多層次 CFA 來檢驗三個潛在變數在總體層次的因素效度。

單一層次 CFA。單一層次的 CFA 模型中，共有 10 個觀察變數，獨立模型（SM0）沒有任何因素的設定，各觀察變數無相關。其他各 CFA 模型將依據潛在變數之間的相關有無，區分為四個模型：直交三因素模型（SM1），亦即三個構念相關為 0；斜交三因素模型（SM2），亦即三個構念間的相關自由估計；自變數單一因素模型（SM3），亦即「專業能力」與「類我程度」兩個構念合而為一，以及三個構念間具有完美相關（設定為 1.00），亦即為單一因素模型（SM4）。各模型的模式適配指標結果列於表 2。

由表 2 可知，四個假設模型中，以斜交三因素模型（SM2）的資料適配情形最好， $\chi^2(32)=114.065$, $p < .001$, $\chi^2/df=3.56$, $RMSEA=.055$, $CFI=.988$, $TLI=.983$, $SRMR=.022$ 。相對之下，直交模型（SM1）與單一因素模型（SM4）適配情形均不理想，表示三個構念並非獨立無關聯，也不是合而為一的構念。直交模型與斜交模型的卡方差異量達顯著水準， $\Delta\chi^2=1582$, $\Delta df=3$, $p < .001$ ，表示因素間的相關十分明顯。另一方面，單一因素模型與斜交模型的 $\Delta\chi^2$ 亦達顯著， $\Delta\chi^2=611.20$, $\Delta df=3$, $p < .001$ ，顯示因素之間的相關不至於過大而造成構念間的區辨效度過低。

最後，SM4 的模式設定將兩個自變數構念視為同一個構念，模型中僅有一個自變數因素與一個依變數因素，但模式適配指標並不理想， $\chi^2(34)=305.151$, $p < .001$, $\chi^2/df=8.98$, $RMSEA=.097$, $CFI=.960$, $TLI=.948$, $SRMR=.035$ 。因此，本節的結論是個體層次的因素結構以斜交三因素模式最佳，此一模型將作為多層次 CFA 的基礎。SM3 模型的參數估計結果列於圖 2。

表 1
「口試評定表」各項目描述統計量與相關矩陣

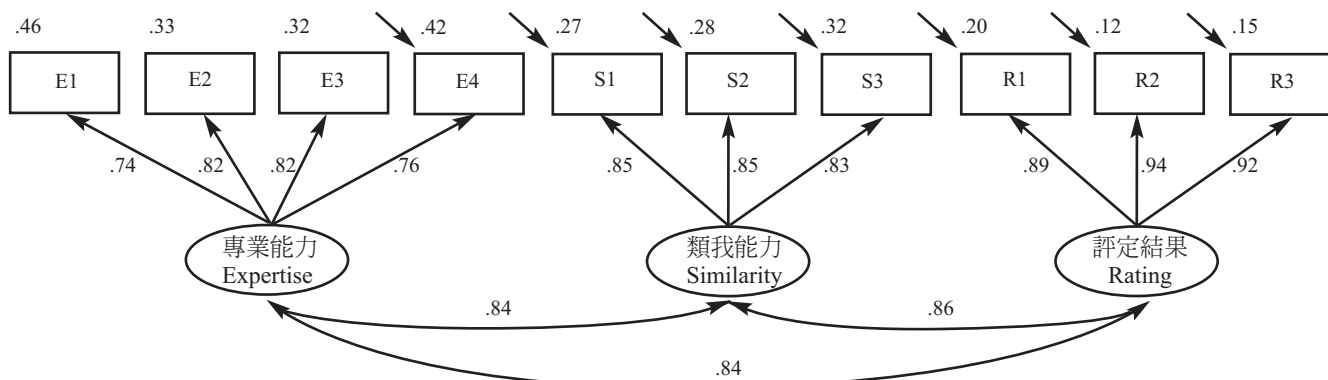
評量題項	描述統計			IV1：專業能力				IV2：類我程度			DV：評定結果		
	M	SD ₁	SD ₂	E1	E2	E3	E4	S1	S2	S3	R1	R2	R3
E1. 具相關工作經驗及訓練	3.55	.81	.46	1.00	.77	.81	.86	.75	.70	.71	.77	.77	.76
E2. 具班級經營能力	3.45	.77	.44	.63	1.00	.88	.76	.85	.86	.83	.79	.76	.75
E3. 具親師生溝通能力	3.53	.75	.45	.58	.72	1.00	.80	.85	.84	.84	.82	.81	.80
E4. 具教師專業知識	3.63	.78	.46	.60	.58	.61	1.00	.77	.73	.75	.82	.85	.83
S1. 工作的態度與我相似	3.45	.80	.45	.54	.60	.61	.61	1.00	.92	.89	.82	.80	.81
S2. 對待學生方式與我相似	3.40	.76	.43	.49	.58	.59	.56	.73	1.00	.88	.81	.76	.75
S3. 教育理念與我相似	3.41	.77	.44	.50	.59	.58	.57	.68	.71	1.00	.83	.80	.79
R1. 表現很好	3.50	.85	.47	.56	.63	.64	.61	.67	.65	.65	1.00	.93	.91
R2. 適當當老師	3.50	.89	.49	.57	.62	.62	.63	.66	.64	.65	.83	1.00	.96
R3. 可能錄用	3.49	.99	.53	.56	.61	.62	.62	.65	.62	.65	.81	.87	1.00

註：下三角為個體層次（考生）的各項相關係數，上三角形為總體層次（審查委員）加權相關係數（以各委員評審人次加權）。SD₁ 為個體層次 841 位考生求得標準差；SD₂ 為整體層次 76 位審查委員經過審查人數加權後（亦即分合資料）求得之標準差。

表 2
單層次與多層次 CFA 分析模式適配指標摘要表

組內	組間	χ^2*	df	χ^2/df	RMSEA	CFI	TLI	SRMR	
								組內	組間
單層次 CFA									
SM0 獨立模型	-	6890.687	45	153.13	.425	.000	.000	.572	-
SM1 三直交因素	-	1696.065	35	48.46	.238	.757	.688	.469	-
SM2 三斜交因素	-	114.065	32	3.56	.055	.988	.983	.022	-
SM3 自變數單因素	-	305.151	34	8.98	.097	.960	.948	.035	-
SM4 單一因素	-	725.267	35	20.72	.153	.899	.870	.057	-
多層次 CFA									
MM1 三斜交因素	獨立模型	423.479	77	5.50	.073	.941	.931	.086	.791
MM2 三斜交因素	三直交因素	584.100	70	8.34	.093	.913	.888	.053	.688
MM3 三斜交因素	三斜交因素	160.118	64	2.50	.042	.984	.977	.026	.043
MM4 三斜交因素	自變數單因素	170.190	66	2.58	.043	.982	.976	.021	.051
MM5 三斜交因素	單一因素	202.105	67	3.02	.049	.977	.969	.027	.066

* 所有的 χ^2 值均達 .001 顯著水準。



圖二：斜交三因素模型 (SM2) 參數估計結果圖示

由圖 2 可知，各潛在變數的測量變數的因素負荷量均大於 .70，顯示因素內的一致性頗高。自變數兩個潛在變數的相關為 .86 ($p < .001$)，顯示考生被評定具有專業能力者，同時與被評審委員的類我程度越高。兩個自變數與依變數的相關亦很高，「專業能力」與「評定結果」相關以及「類我程度」與「評定結果」相關均為 .84 ($p < .001$)。

多層次 CFA。延續前面的單層次 CFA，本節將個體層次（組內）與總體層次（組間）的因素結構同時納入分析，其中個體層次採前一節所保留的三因素斜交模型，總體層次則依照前一節的區分方式區分成五種不同模型：第一個多層次模型（MM1）是獨立模型，亦即總體層次並沒有特殊的因素結構設定，10 個觀察變數在組間層次的觀察矩陣，即是由 10 個觀察變數形成的 10 個總體層次的具單一觀察變數的潛在變數所組成，各潛在變數亦相互獨立，沒有相關的設定。此一模型並不具備研究者假設的因素結構，作為基準模型（baseline model）。總體層次的因素結構延續個體層次的模式設定，依序為直交三因素模型（MM2）、斜交三因素模型（MM3）、自變數單因子模型（MM4），以及三個潛在構念合而為一的單一因子模型（MM5）。這五個多層次 CFA 模型的模式適配結果列於表 2。

由表 2 可知，MM1 沒有任何因素結構的設定，因此模式適配最不理想， $\chi^2(77)=423.479$, $p < .001$ ，同時我們可以從組間 SRMR_B 指標 (.791) 看出，總體層次的標準化殘差非常大，個體層次的殘差相對較小 (SRMR_W=.086)，顯示 MM1 的適配不理想主要是發生在總體層次，符合此一模式設定的基本特性。總體層次若為直交三因素的模型 (MM1) 適配情形更不理想， $\chi^2(70)=584.1$, $p < .001$, RMSEA=.093, 表示三個潛在變數在總體層次並非無相關的獨立構念，由 SRMR_B=.688 可看出總體層次的模式設定不良。

關於 MM3 到 MM4 兩個模型，模式適配程度非常接近。雖然可以從各指標看出斜交三因素模型 (MM3) 仍較佳， $\chi^2(64)=160.118$, $p < .001$, $\chi^2/df=2.5$, RMSEA=.042, SRMR_B=.043, SRMR_W=.026；與 MM4 的適配差異的卡方差異量不大 ($\Delta\chi^2=10.07$, $\Delta df=2$)，達 .01 顯著水準，但未達 .001 顯著水準， χ^2/df 維持在小於 3 的 2.58，RMSEA=.043，顯示 MM4 的適配度尚佳，亦即自變數兩個構念在總體層次的區辨力很低。另外，從參數估計的結果可以看出，總體層次因素相關高達 .96，相關係數的 95% 信賴區間 (.911 到 1.005) 已涵蓋 1.00，亦即兩個因素相關係數可能是 1.00 的完美相關，因此自變數在總體層次有可能是單因子結構。儘管如此，前述結果無法明確論斷 MM3 與 MM4 的優劣，因此後續分析同時採用兩種設定方法（自變數為雙因子斜交或單因子結構），來檢驗脈絡變數在總體層次的狀態。

至於總體層次的三個潛在構念為單一因子的模型 (MM5)，模式適配指標為 $\chi^2(67)=202.105$, $p < .001$, $\chi^2/df=3.02$, RMSEA=.049, SRMR_B=.027, SRMR_W=.066，與 MM3 與 MM4 的參數數目僅分別相差 3 與 1，但是卡方數值則放大了 42 與 32，卡方差異量明顯達到 .001 顯著水準， χ^2/df 數值放大到 3.02，超過 3 的慣用標準，顯示 MM5 的適配程度較 MM3 與 MM4 為低，因此下文僅討論 MM3 與 MM4 兩個模型的參數估計結果，數據列於表 3。

表 3 列出了 MM3 與 MM4 兩個較佳模型的參數估計結果，兩者在個體層次的參數估計結果與先前單一層次 CFA 模型並無明顯差異，各測量變數的因素負荷量（介於 .666 至 .917）均十分理想。但是在總體層次部分，組間效果的 CFA 模型的因素負荷值都相當高，標準化 Lambda 係數均在 .85 以上，甚至超過 .95 的水準。根據 Bagozzi 與 Yi (1988) 的觀點，超過 .95 的因素負荷量屬於過高的係數，大於 1.00 者

表 3
多層次驗證性因素分析參數估計結果摘要表

模 型	MM4					MM3			
	總體層次自變數為單因素					總體層次自變數為斜交二因素			
	組內 (考生)			組間 (評審)		組內 (考生)		組間 (評審)	
	ICC	Coef.*	t	Coef.*	t	Coef.*	t	Coef.*	t
<u>IV1：專業能力</u>									
E1.具相關工作經驗及訓練	.243	.668	-	.885	-	.666	-	.907	-
E2.具班級經營能力	.269	.777	15.72	.956	9.97	.776	15.70	.947	9.87
E3.具親師生溝通能力	.298	.754	15.24	.971	11.34	.753	15.54	.973	11.47
E4.具教師專業知識	.265	.692	14.25	.885	11.63	.691	14.18	.906	12.10
<u>IV2：類我程度</u>									
S1.工作的態度與我相似	.246	.799	-	.993	8.55	.797	-	1.00	-
S2.對待學生方式與我相似	.257	.799	19.66	.977	8.09	.796	19.72	.985	16.92
S3.教育理念與我相似	.256	.774	24.87	.963	8.07	.772	24.83	.965	14.83
<u>DV：評定結果</u>									
R1.表現很好	.227	.864	-	.968	-	.864	-	.967	-
R2.適合當老師	.219	.917	30.38	1.00	13.90	.917	30.37	1.00	13.96
R3.可能錄用	.203	.906	24.64	.988	11.38	.906	24.67	.988	11.15
<u>因素變異數</u>									
二因素 IV1：專業能力		.215	7.14			.213	7.01	.157	4.49
二因素 IV2：類我程度		.308	8.03			.305	7.99	.131	4.17
單因素 IV：考生特質評價				.124	4.06				
依變數：評定結果		.415	7.10	.159	4.49	.415	7.11	.158	5.72
<u>因素相關</u>									
專業能力/類我程度		.790	7.73			.801	7.76	.957	5.64
專業能力/評定結果		.799	7.64			.799	7.63	.918	4.50
類我程度/評定結果		.817	7.35			.822	7.35	.864	5.15
考生特質評價/評定結果				.895	4.41				

* 除了變異數為未標準化估計值外，其餘參數均為完全標準化解之係數值。

則為不適當解 (improper solution)，顯然這是由於總體層次變數相關甚高 (造成過高負荷量的變數相關均高於.90)，以及總體層次觀察值僅有 76 人，在因素萃取過程中造成的數據標準化後超過 1.00 水準的異常現象。一般在樣本數小於 100 時，非常容易觀察到此一現象，在技術上可以利用限定誤差為特定值 (如.10) 等方式來進行調整 (Bentler & Lee, 1979; Loehlin, 2004)，但是考量本研究發生的超過 1.00 的係數數值超過量均在小數點第三位以下 (例如 1.007 與 1.001)，標準誤的估計仍屬穩定，參數的顯著性考驗值均在穩定合理的範圍，對於參數估計與潛在因素估計不利影響尚屬輕微，因此為了模型的比較與自由度的穩定，不以人為的方式進行參數限定，而維持原來的模式設定。

從 ICC 數值可以看出，各測量題項在組間差異的效果上非常明顯，10 個測量變數的 ICC_M 介於.203 至.298 之間，屬於高度的組內相關 (Cohen, 1988)。另外，以萃取後潛在變數變異數所計算出來的 $ICCL$ ，反應了潛在變數組間變異的強度，由於表 3 當中的個體層次與總體層次的因素數目不一定相同，因此對於潛在變數的 $ICCL$ 係以個體與總體層次所有的潛在變數變異數與共變數來計算，而不針對特定潛在變數來計算 $ICCL$ 。MM4 模型的 $ICCL$ 為.232，MM3 模型的 $ICCL$ 則為.323。很明顯的，由於 MM4 模型的自變數在總體層次僅抽取一個因素，萃取變異量較少，相對之下，MM3 模型的自變數在總體層次萃取了兩個因素，顯示 MM3 模型的潛在變數組間差異上較 MM4 模型強大許多。

結構模型分析

前一節的 CFA 確認了個體層次因素結構的最佳模型為斜交三因素模型，自變數在總體層次則可能是雙因素，亦可能是單因素結構。以下則以結構模型來探討「專業能力」與「類我程度」對於「評定結果」的影響。

在模式設定上，基於前一節的分析，由於總體層次測量模型在自變數部分有單因素 (MM4) 與雙因素 (MM3) 兩種不同的設定，因此結構模型的設定依測量基礎的不同區分成 MS1 (以 MM4 為基礎，總體層次測量模型的自變數為單因素) 與 MS2 (以 MM3 為基礎，總體層次測量模型的自變數為二斜交因素) 兩種。

結構模型檢驗亦以獨立模型為基準模型，若個體層次與總體層次的各因素 (包括解釋變數與依變數) 相關設定為 0，因素間完全獨立無關，為完全獨立模型 (以 00 表示)；如果個體層次各因素相關設定為

0，但總體層次因素相關自由估計，為部分獨立模型 (以 01 表示)；如果總體層次各因素相關設定為 0，但個體層次因素相關自由估計，為部分獨立模型 (以 10 表示)。相對於基準模型者，為潛在自變數對於潛在依變數進行迴歸分析的結構模型 (以 11 表示)。同時，合併前述三種基準模型，總計有八個模型有待分析。結果如表 4 所示。

由表 4 可知，各獨立模型與部分獨立模型的模式適配均不理想，其中更有三個模型 (MS1₀₀、MS1₀₁、MS2₀₁) 因為組內因素被設定為零相關而導致非正定問題無法收斂，顯見個體層次的模式設定錯誤會嚴重影響參數估計的正確性。另一方面，兩個結構模型的模式適配則相對良好。組間因素為單因素的結構模型 (MS1₁₁) 的 $\chi^2(66)=170.19, p < .001, RMSEA=.043, SRMR_B=.051, SRMR_W=.021$ ；組間因素為二因素的結構模型 (MS2₁₁) 的 $\chi^2(64)=160.12, p < .001, RMSEA=.042, SRMR_B=.043, SRMR_W=.026$ 。從適配指標來看，雖以 MS2₁₁ 的模式適配較佳，但是如同前一節的討論，組間二因素的區辨並不理想，亦即組間二因素的相關甚高，因此即使模式適配良好，反而可能在結構模型分析時造成嚴重的共線性問題，導致參數的估計產生扭曲現象。

由於 MS1₁₁ 與 MS2₁₁ 兩個模型的差異在於總體層次的因素結構設定，因此兩個模型的參數估計結果，在個體層次部分非常接近 (與表 3 的 CFA 數據一致)，因此不予以贅述。在迴歸係數部分，MS1₁₁ 模型的兩個潛在自變數對於潛在依變數的迴歸係數為「專業能力」的.409 ($t=5.14, p < .01$) 與「類我程度」的.493 ($t=6.32, p < .01$)；MS2₁₁ 模型的「專業能力」為.392 ($t=4.79, p < .01$) 與「類我程度」的.507 ($t=6.44, p < .01$)。MS1₁₁ 與 MS2₁₁ 兩模型在組內的迴歸解釋力 (R^2) 均為.73，顯示兩個自變數能夠有效的解釋依變數。

MS1₁₁ 與 MS2₁₁ 兩模型在組間層次的估計結果，在迴歸係數部分有明顯差異 (參見圖 3)。MS1₁₁ 與 MS2₁₁ 兩模型總體層次的 R^2 均分別為.80 與.85，顯示模型中的自變數能夠有效的解釋依變數，但是 MS2₁₁ 的兩個組間因素發生異常的迴歸係數，「專業能力」的迴歸係數為 1.091 ($t=2.52, p < .01$)，而「類我程度」的迴歸係數則變成不顯著的負數 (-.18, $t=-.43, ns$)，MS2₁₁ 的參數估計顯然是不適當解。由於 MS2₁₁ 模型中，總體層次兩個潛在自變數「專業能力」與「類我程度」的相關達.957 ($t=5.64, p < .001$)，顯示組間層次的雙因素模型的迴歸分析受到共線性的影響很大，是造成不適當解的主要原因。相對之下，MS1₁₁ 模型的總體層次以單一潛在自變數「考生特質評價」

表 4
多層次結構模式適配指標摘要表

模 型	χ^2 *	df	χ^2/df	RMSEA	CFI	TLI	SRMR	
							組內	組間
<u>總體層次自變數為單因子模型 (MS1)</u>								
MS1 ₀₀ 獨立模型 (全體因素無相關)	-	-	-	-	-	-	-	-
MS1 ₀₁ 部分獨立模型 (組內因素無相關)	-	-	-	-	-	-	-	-
MS1 ₁₀ 部分獨立模型 (組間因素無相關)	242.997	67	3.63	.056	.970	.960	.027	.527
MS1 ₁₁ 結構模型	170.190	66	2.58	.043	.982	.976	.021	.051
<u>總體層次自變數為二因子模型 (MS2)</u>								
MS2 ₀₀ 獨立模型 (全體因素無相關)	1520.259	70	21.72	.157	.755	.685	.403	.664
MS2 ₀₁ 部分獨立模型 (組內因素無相關)	-	-	-	-	-	-	-	-
MS2 ₁₀ 部分獨立模型 (組間因素無相關)	308.379	67	4.60	.065	.959	.945	.043	.667
MS2 ₁₁ 結構模型	160.118	64	2.50	.042	.984	.977	.026	.043

*所有的 χ^2 值均達.001顯著水準。

來反應 7 個解釋觀察變數，對於依變數的迴歸係數為.895 ($t=8.82$, $p<.001$)， $R^2=.80$ ，顯示「考生特質評價」可以有效解釋「評定結果」變異的 80%，是一個很強的預測變數。

討論與結論

脈絡變數的形成與影響，是社會科學研究當中非常值得重視的研究課題，只要資料的結構具有階層化叢集特徵，造成數據變異的成因除了個體間的差異，還有總體層次的組間效果，亦即脈絡的影響，即必須以多層次模型來進行分析。組間效果一方面可能是由個體層次匯聚而成的脈絡變數，也有可能是總體層次觀察值的固有特徵（例如組織的規模、學校的公私立性質等）。本文所關心的主要議題，在於如何利用結構方程模型來進行脈絡變數的因素分析，萃取出潛在脈絡變數，並進行結構模型分析來得到對效標變數的解釋情形。

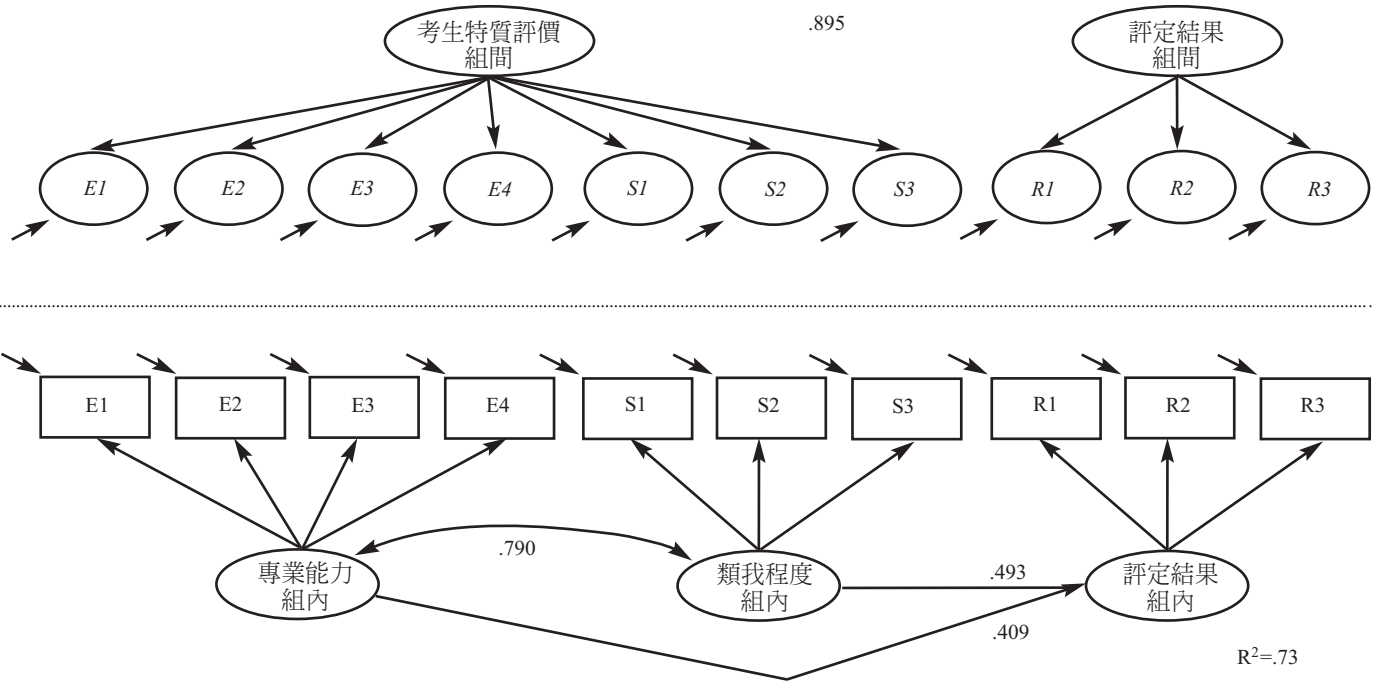
脈絡變數的概念意涵與分析

脈絡變數與脈絡效果的多元定義

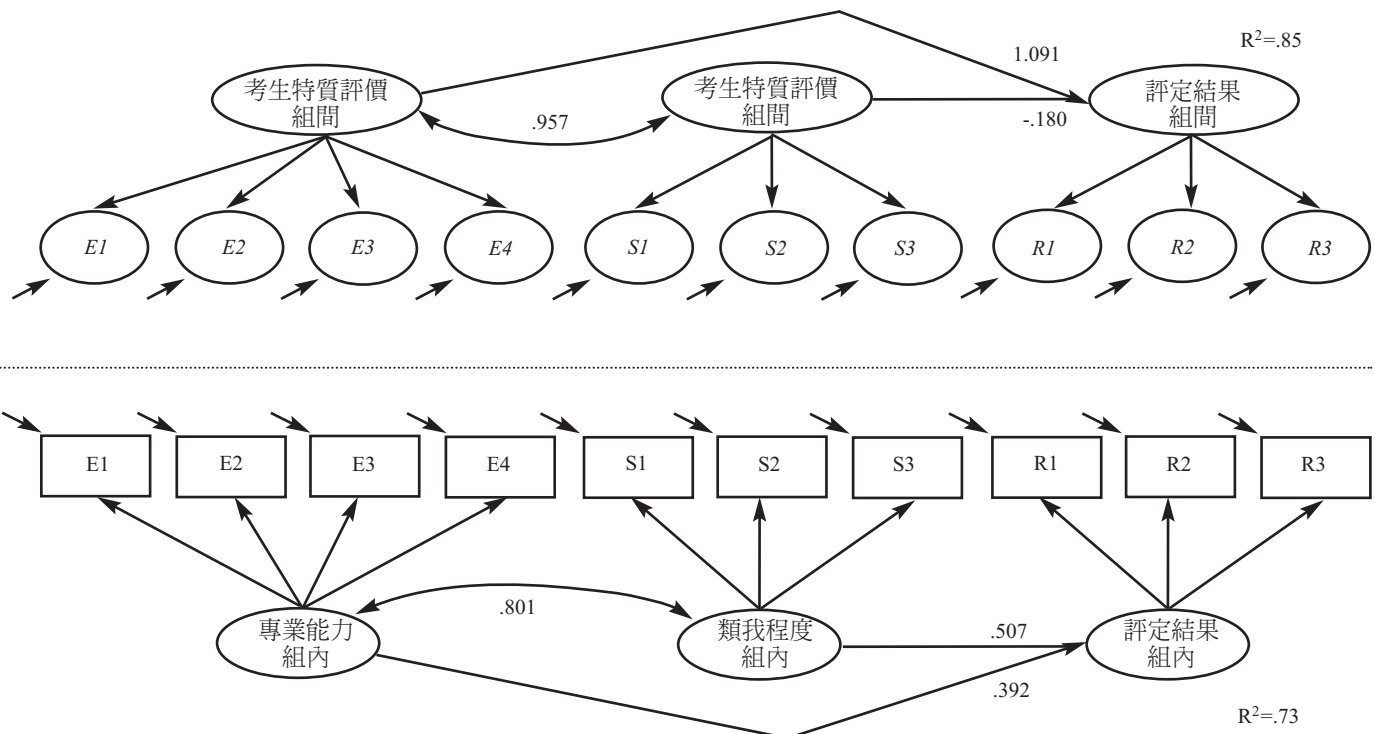
多層次模型在方法學上最特殊之處，在於能夠對脈絡的影響進行實證分析 (Snijders & Bosker, 1999)。Macintyre 與 Ellaway (2003) 認為，以個體

資料聚合成總體層次的脈絡變數作為脈絡的代理變數，反應了一群人的集合所產生的社會/人際影響，是一種社會瀰漫 (social miasma) 的現象，但直接對於脈絡進行測量得到的真實脈絡測量（例如公司的資產、規模等等）未必對個體產生任何作用。Raudenbush (2003) 也持相同的觀點，認為人們身存於個體共同建構的空間中，因此在研究團體歷程這類的課題時（例如團體凝聚力的影響），以個體層次資料的聚合變數來反應脈絡的影響力，比直接測量團體變數更有價值。我們可以把這種個體的行為受到個體自身所創造的脈絡變數的影響，稱之為集體性脈絡效果 (collective contextual effects)。另一種廣義的脈絡效果定義，認為只要是總體層次變數（不論是個體層次聚合得到的變數或總體層次固有變數），均可能對個體行為產生影響，這些影響都可稱為脈絡效果（例如 Courgeau, 2003; Kreft & de Leeuw, 1998），我們可以稱之為背景脈絡效果 (background contextual effects)。

本文利用口試歷程實證數據所進行的脈絡變數分析，反應了另一種類型的脈絡效果的存在。由於個體層次的數據是由總體層次的評審依其主觀感受進行評定，再以聚合程序形成本文所使用的脈絡變數，此舉與傳統脈絡變數的操作型態不同。當個別考生的個人特質與口試表現都直接由評審決定，而評審本身又帶有評分行為的個別差異（亦即評分者效應），此時考



(a) MS1₁₁ : 組間自變數為單因素模式



(b) MS2₁₁ : 組間自變數為二因素模式

圖三：多層次結構模式參數估計結果路徑圖示

生嵌套在評審之內，因此個別考生的狀態與評審特質所存在的交互作用，皆是由總體層次觀察單位所創造，而與個體層次的測量行為無關，本文將此種特殊型態的脈絡變數定義成一種評分者內設計的脈絡作用，或稱之為內在脈絡效果（intrapersonal contextual effects）

對於套用內在脈絡效果的觀念於評分者效應的探討，並且以 MSEM 模型來進行分析，可以說是人力資源與甄選評量領域進行類似研究的另一種替代觀點與分析策略。但在此必須特別指出，此舉與傳統對於脈絡變數（個體直接測量的聚合平均數）的定義方式不同，因此本文所主張的脈絡變數與其所產生的脈絡效果，自有其特定的意義，實證分析所獲得的結果僅適合於解釋評分者效應的內在化脈絡效果，未來的研究可以採用傳統的脈絡變數定義程序，以本文所建議的操作程序來進行 MSEM 分析，將更能接近一般所討論的集體性脈絡效果分析。然而，透過本研究分析範例，可以發現脈絡效果的存在形式，確實可能因為不同型態的多層次資料結構而有所不同，適度擴大脈絡變數的定義，相信能夠提升多層次模型的學理與實務價值。

脈絡變數的外顯性與潛在性

本研究對於脈絡變數的另一個特殊觀點，是以潛在構念的分析策略來探討脈絡變數的影響，試圖為脈絡變數是否能夠反應總體層次的脈絡現象的爭議，提出一套具體可行的分析方案。

過去在檢驗多層次結構的資料時，多利用外顯變數的觀點，將個體與整體層次的變數合併於一個混合迴歸方程式來進行估計，此一作法雖然可以簡捷有效地處理各層次解釋變數相互的影響，但卻沒有考慮脈絡變數可能是以一種潛在構念的形式存在，雖然研究者可以利用主成分分析（PCA）或是合成分數（composite score）將定義潛在變數的觀察變數加總成脈絡變數，但是此法忽略了測量誤差的影響，將導致統計檢定力的降低（Bollen, 1989）。本文利用驗證性因素分析來檢驗總體層次脈絡變數的潛在因素結構，突破傳統 MLM 取向的限制，將潛在構念方法學應用在脈絡效果的探究，未來將可延伸更多的學術議題。

構念偏移的計量證據

本文在具體利用 MSEM 模型來進行口試評分資料的潛在脈絡變數模型分析時，即觀察到多層次測量的構念偏移現象（Snijders & Bosker, 1999）。本文在進行研究變數的說明時，即已將個體層次的七個 IAF 自變數定義成「專業能力」與「類我程度」兩個潛在構

念。在個體層次 CFA 分析結果中可以看到，這兩個潛在構念即使關係很高，但是仍可有效分離；但是在總體層次，「專業能力」與「類我程度」的區辨性卻十分不理想，模式適配的評估結果顯示兩者在總體層次並非是分離的構念而是單一的構念。在評審層次的脈絡變數的構念內涵已經不再是反映考生「專業能力」與「類我程度」，既有專業能力的評定，也混雜了類我效應的影響，可以說是一種來自評分者個別差異的方法效應，在構念意義上已經產生了實質的變化。

本文將總體層次的潛在脈絡變數命名為「考生特質評價」，實證分析的結果顯示，此一構念與「對考生的評定」是兩個可區辨的構念，亦即「考生特質評價」並未混雜了作為潛在脈絡依變數的「評定結果」，因此可進行兩者迴歸關係的探討。值得注意的是，當我們將考生層次的「評定結果」聚合成總體平均數後，以因素分析萃取得到的潛在脈絡依變數，其構念意義也不再是反應考生的口試分數的「評定結果」，而是評分者對於評定分數寬嚴差異，顯然此一變數構念意義也產生偏移。本文將構念偏移現象以非等值因素模型來表現，並以 MSEM 分析程序來進行計量的分析估計，可以說是 MSEM 取向的一個重要應用。

評分者效應的探討

本文對於口試評分歷程的潛在脈絡變數模型分析，最直接的一個發現是總體層次的模型解釋程度（ R^2 ）高於個體層次，換言之，評分者層次的變數關係比考生層次的變數關係更有解釋上的效力。從模型中的高度組內相關係數（ICC），不論是潛在形式（ ICC_L ）或外顯形式（ ICC_M ）皆很高，即說明脈絡變數的外顯形式或潛在形式在總體層次的影響力均不可忽視。如果研究者忽略了組間效果的解釋，可能錯估了自變數與依變數間的關係。由本研究的數據發現，在個體與總體層次同時估計時，個體層次的解釋變異百分比雖然仍有 73%，但是整體層次的解釋力更高達 80%，顯示解釋變數除了從個體層次發揮影響力之外，更應從評審之間的差異影響來說明。也就是說，影響考生成績的因素，一方面是考生個別的「專業能力」或「類我程度」的高低，如果評審對於能力與類我程度評定分數越高者，最後打出的分數更高，此一關係強度尤甚於考生層次的狀況。更重要的是，本研究觀察到的評分者效應，在自變數（潛在脈絡解釋變數）與依變數（潛在脈絡依變數）各自反映了兩種不同的效應形式：評分者月暈效果（評分者偏誤）與評分寬嚴差異（定錨效應），說明如下。

評分寬嚴效應

基本上，本文的口試資料的各變數經聚合後成為脈絡變數（即 MLM 取向的隨機截距），反應了個別評分者的評定定錨點（anchor point）的個別差異，有的評分者分數打得高（定錨點高，評分較寬鬆），但有的評分者分數打得低（定錨點低，評分較嚴格）。經過因素分析的處理後，萃取出來的潛在變數反映了定錨效果。脈絡變數與潛在脈絡變數的標準差（或變異數）反應了評分者的變異程度，變異數越大，表示評分者間的不一致性越大，組間效果越強。同時，當變異數越大，在潛在構念的萃取過程中越能夠定義出獨立因素來解釋評分者在評分歷程上的個別差異，在過去的評分者效應研究中，把評分者的評分差異定義為共識高低或寬嚴差異的一種定錨效應（anchor effect）（Lord, 1977; Allen & Yen, 2002），利用潛在變數模式的分析策略，定錨效應得以利用 SEM 模型來進行潛在變數的估計，更重要的意義是把評分者個別差異（定錨的影響）從個體層次的影響中予以抽離，提升到總體層次來加以分析。本研究的實證資料分析發現，當定錨效應發生在本研究的總體層次依變數時，反應的是口試表現得分的寬嚴效果，而非考生個別的能力表現。Niu 與 Sternberg（2001）的一項研究發現，中國評分者較西方評分者在評定創意作品較為寬鬆，顯示文化因素是評分寬嚴效果的調節因素，此類型的研究，未來可以本文所提出的模型來進行更進一步的檢驗。

評分者偏誤

評分者效應的另一個問題是評分者偏誤（rater bias），亦即因為評分者的主觀因素，所造成測量分數的系統性偏誤（systematic bias）問題，在本研究的總體層次的潛在脈絡解釋變數，我們即觀察到月暈效果的存在。月暈效果是一種典型的共同方法變異（common method variance），一種基於某特定評量方法造成的系統變異，Campbell 與 Fiske（1959）提出了方法效應的概念，並提出了多重方法多重特質矩陣（MTMM matrix）來估計，後來的學者應用 SEM 取向來分析 MTMM 資料（例如 Kenny & Kashy, 1992; Lowe & Ryan-Wenger, 1992; Marsh & Hocevar, 1988），但是始終無法有效分離方法變異。本文在進行脈絡效果的分析過程中，發現了利用 MSEM 程序，可在總體層次觀察到脈絡變數間的高度相關，亦即評分者在某一個題目打高分，在另一個分數也會打高分，進而以潛在變數的形式來定義與估計，顯示多層次的 SEM 模型可以用來探討共同方法變異問題（Podsakoff, MacKenzie, Lee, & Podsakoff, 2003）。本研

究在總體層次萃取出的方法變異，亦即潛在脈絡解釋變數，命名為「考生特質評價」，可惜受限於研究目的與篇幅，未能深入探討方法效應的進一步內涵與成因，未來研究可以針對此一議題，設計適當的研究來深入討論。

估計技術上的限制與建議

對於多層次分析應用於脈絡變數構念萃取與影響的探討，本文僅以最簡單的兩因素模型來進行說明。事實上，實際的研究中仍可能要加入其他總體層次的解釋變數，才可以充分解釋依變數的變異情形。進一步的研究可以擴大到當模型中有其他總體層次解釋變數與脈絡變數同時存在模型中的分析，尤其是變數之間若具有交互影響或中介影響的複雜模型的應用，或是當模型中具有類別變數的處理方式。值得注意的是，不論是多層次或類別變數的 SEM 分析，在傳統的 LISREL、AMOS、EQS 軟體都有諸多限制。但是 Mplus 的估計則中，則充分考慮多層次資料特徵，並可考量特殊形式的變數，因此可以說是研究多層次模型的最佳工具。

應用 MSEM 模型來檢驗潛在脈絡變數雖然有諸多優點，但是由於有待估計的參數眾多，模型又區分為個體與總體層次，因此參數估計的穩定性與收斂性受到很大的挑戰。如果變數關係並不明確，或是樣本數不足，往往造成無法收斂的結果。本研究在分析過程，發現多層次模型受限於模型的複雜度，在技術上存有相當的瓶頸，以目前的技術尚無法克服。例如當模型中有隨機斜率時，疊代程序往往無法收斂完成，並造成記憶體超過負荷的窘境。因此，在操作實務上的建議是，MSEM 模型應越簡單越好，減少參數估計的負荷。由於脈絡變數的影響的探討，是透過結構模型參數估計來進行，如果測量變數過多，可以利用變數包裹（item parceling）的方法，減輕測量模型部分的負擔，以獲得結構模型部分的參數估計結果（Bandalos, 2001; Bandalos & Finney, 2001; MacCallum, Widaman, Zhang, & Hong, 1999）。

進一步的，本文對於脈絡變項的探討，僅針對個體層次觀察變數的聚合平均數，以及聚合平均數萃取得到的潛在脈絡變數進行實證分析，然而，個體層次觀察變數除了層次性的聚合集中量數得進行 SEM 分析之外，聚合過程中的變異量數（包括變異數與共變數），其心理計量特性與實務意義亦有加以探究之必要，因為平均數只反映了一部份的變數特性，其他變數特性的考量，是 MSEM 分析能夠完整捕捉多層次資料的變項分配特性的必要考量。從這個觀點來看，

本文僅是多層次潛在變數模型研究的一個起點，後續的研究具有相當開闊的研究空間。

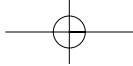
MSEM 估計另一個技術上的限制是樣本問題，由於多層次模型必須同時針對個體與總體層次進行 ML 估計，因此兩個層次的觀察資料都必須符合 SEM 分析的基本要求，Hox (2002) 指出多層次模型在個體層次的收斂與不偏性大致沒有問題，但是總體層次則因為研究實施的限制，樣本數多會不足，造成型一錯誤率的提高，以及標準誤低估的偏誤問題，一般而言，當模型越趨於複雜時，所需要的樣本數更高，當 ICC 越小時，總體層次的參數估計偏誤會趨於明顯 (Cheng & Au, 2005)。根據 Hox 與 Maas (2001) 的模擬研究結果發現，總體層次觀察值在 50 以下與 ICC 很小時，估計會有明顯的偏誤，直到 100 以上觀察值估計結果才趨於穩定，在本研究分析過程中，即因為總體層次雖有大於 50 的觀察值，但亦僅有 76，未能到達 100 的建議值，導致總體層次的參數估計發生非正定現象 (因素負荷量超過 1.0 的殘差值負數問題)，雖然情形輕微，尚不致影響其他參數的估計與解釋，但也說明了樣本數多寡對於 MSEM 模型估計的影響，未來要利用本文所建議的分析程序時，宜採納 Hox 與 Maas (2001) 的建議，擴大總體層次的樣本規模，才能得到穩定有效的估計結果。讀者必須謹記在心，SEM 分析的一個基本限制，是當模型越趨於複雜時，我們就要付出更大的代價。到了 Multilevel SEM，研究者將面對更嚴峻的挑戰。

參考文獻

- 余民寧 (2006)。《潛在變數模式：SIMPLIS 的應用》。台北市：高等教育出版公司。
- 呂秋萍 (2005)。《國中教師甄選口試決策歷程之研究：以結構方程模式檢驗》。國立政治大學教育研究所未出版之碩士論文。
- 邱皓政 (2003)。《結構方程模式：LISREL 的理論、技術與應用》。台北市：雙葉書廊有限公司。
- 邱皓政、溫福星 (2007)。脈絡效果的階層線性模式：以組織創新氣氛與創意表現為例。《教育與心理研究》，30(1)，1-35。
- 黃芳銘 (2003)。《結構方程模式理論與應用》。台北市：五南圖書公司。
- 溫福星 (2006)。《階層線性模式：原理、方法與運用》。台北市：雙葉書廊有限公司。
- 謝臥龍 (1997)。優良國中教師特質之德懷術分析。《教育研究資訊》，5(3)，14-28。
- Allen, M. J., & Yen, W. M. (2002). *Introduction to measurement theory*. Long Grove, IL: Waveland Press, Inc.
- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103, 411-423.
- Anderson, N., & Shackleton, V. (1990). Decision making in the graduate selection interview: A field study. *Journal of Occupational Psychology*, 63, 63-76.
- Bagozzi, R.P., & Yi, Y. (1988). On the Evaluation of Structural Equation Models. *Journal of the Academy of Marketing Science*, 16, 74-94.
- Bandalos, D. L. (2001). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling*, 9, 78-102
- Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides and R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 269-296). Mahwah, NJ: Lawrence Erlbaum Associates.
- Baskett, G. D. (1973). Interview Decision as Determined by Competency and Attitude Similarity. *Journal of Applied Psychology*, 57(3), 343-345.
- Bentler, P. M., & Lee, S. Y. (1979). A statistical development of three-mode factor analysis. *British Journal of Mathematical and Statistical Psychology*, 32, 87-104
- Bentler, P. M., & Liang, J. (2003). Two-level mean and covariance structure: Maximum Likelihood via an EM algorithm. In S. P. Reise & N. Duan (Eds.), *Multilevel modeling: Methodological advances, issues, and applications* (pp. 53-70). Mahwah, NJ: Lawrence Erlbaum Associates.
- Bollen, K. A. (1989). *Structural equation modeling with latent variables*. New York: John Wiley.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Cheng, M., & Au, K. (2005). Applications of multilevel structural equation modeling into cross-cultural research. *Structural Equation Modeling*, 12(4), 598-619.
- Cohen, J. (1988). *Statistical power analysis for the*

- behavioral sciences* (2nd Ed.). Hillsdale, NJ: Erlbaum.
- Courgeau, D. (2003). *Methodology and epistemology of multilevel analysis: Approaches from different social sciences*. Norwell, MA: Kluwer.
- Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research*, 38, 529-569.
- Cronbach, L. J., & Webb, N. (1979). Between class and within class effects in a reported aptitude \times treatment interaction: a reanalysis of a study by G. L. Anderson. *Journal of Educational Psychology*, 67, 717-724.
- De Silva, M. (2005). *Context and composition? Social capital and maternal mental health in low income countries*. Unpublished PhD. Dissertation, Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, UK.
- Duncan, O. D., Curzort, R. P., & Duncan, R. P. (1966). *Statistical geography: Problems in analyzing areal data*. Glencoe, IL: Free Press.
- Ferron, J., Dailey, R. F., & Yi, Q. (2002). Effects of misspecifying the first-level error structure in two-level models of change. *Multivariate Behavioral Research*, 37, 379-403.
- Ferron, J., Hess, M. R., Hogarty, K. Y., Dedrick, R. F., Kromrey, J. D., Lang, T. R., & Niles, J. (2004). *Hierarchical linear modeling: A review of methodological issues and applications*. Paper presented at the 2004 annual meeting of American Educational Research Association, San Diego.
- Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London: Arnold.
- Goldstein, H., & Browne, W. (2001). Multilevel factor analysis modeling using Markov Chain Monte Carlo (MCMC) estimation. In G.A. Marcoulides & I. Moustaki (Eds.), *Latent variable and latent structure models*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Goldstein, H., & McDonald, R. (2003). A general model for the analysis of multilevel data. *Psychometrika*, 53, 455-467.
- Graves, L. M., & Powell, G. N. (1988). An investigation of sex discrimination in recruiters' evaluations of actual applicants. *Journal of Applied Psychology*, 73, 20-29.
- Graves, L. M., & Powell, G. N. (1995). The effect of sex similarity on recruiters' evaluations of actual applicants: A test of the similarity-attraction paradigm. *Personnel Psychology*, 48, 85-98.
- Heck, R. H. (2001). Multilevel modeling with SEM. In G.A. Marcoulides & R.E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 89-127). Mahwah, NJ: Lawrence Erlbaum Associates.
- Heck, R. H., & Thomas, S. L. (2000). *An introduction to multilevel modeling techniques*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hox, J. J. (1995). *Applied multilevel analysis* (2nd Ed.). Amsterdam: TT-Publikaties. Available at: <http://www.fss.un.nl/ms/jh>.
- Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hox, J. J., & Maas, C. J. M. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling*, 8, 157-174.
- Jedidi, K., & Ansari, A. (2001). Bayesian structural equation models for multilevel data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 129-157). Mahwah, NJ: Lawrence Erlbaum Associates.
- Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social science* (pp. 85-112). New York: Academic.
- Jöreskog, K. G., & Sorbom, D. (2004). *LISREL 8.7*. Lincolnwood, IL: Scientific Software International, Inc.
- Kaplan, D. (1995). Statistical power in SEM. In R. H. Hoyle (Ed.), *Structural equation modeling: concepts, issues, and applications* (pp. 100-117). Newbury Park, CA: Sage.
- Kaplan, D., & Elliott, P. R. (1997). A didactic example of multilevel structural equation modeling applicable to the study of organizations. *Structural Equation Modeling*, 4, 1-24.
- Kenny, D. A., & Kashy, D. A. (1992). The analysis of the

- multitrait-multimethod matrix by Confirmatory factor analysis. *Psychological Bulletin*, 112, 165-172.
- Kreft, I., & de Leeuw, J. (1998). *Introducing multilevel modeling*. Newbury Park, CA: Sage.
- Lee, S. Y., & Shi, J.Q. (2001). Maximum likelihood estimation of two-level latent variable models with mixed continuous and polytomous data. *Biometrics*, 57, 787-794.
- Loehlin, J. C. (2004). *Latent variable models: An introduction to factor, path, and structural equation analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Lord, F. M. (1977). A broad-range tailored test of verbal ability. *Applied Psychological Measurement*, 14, 117-138.
- Lowe, N. K., & Ryan-Wenger, N. M. (1992). Beyond Campbell and Fiske: Assessment of convergent and discriminant validity. *Research in Nursing and Health*, 15, 67-75.
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4, 84-99.
- Macintyre, S., & Ellaway, A. (2003). Neighbourhoods and health: An overview. In I. Kawachi and L. Berkman, and F. Oxford (Eds.), *Neighbourhoods and health* (pp. 20-42). Oxford University Press.
- Macintyre, S., Ellaway, A., & Cummins, S. (2002). Place effects on health: How can we conceptualise, operationalise and measure them? *Social Science and medicine*, 55, 125-139.
- Marsh, H. W., & Hocevar, D. (1988). A new, more powerful method of multitrait-multimethod analysis. *Journal of Applied Psychology*, 73, 107-117.
- McDonald, R. P. (1994). The bilevel reticular action model for path analysis with latent variables. *Sociological Methods and Research*, 22, 399-413.
- McDonald, R. P., & Ho, M. R. (2002). Principles and practice in reporting structural equation analysis. *Psychological Methods*, 7, 64-82.
- Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, 54, 557-585.
- Muthén, B. O. (1990). *Means and covariance structure analysis of hierarchical data*. Los Angeles: UCLA Statistics series, #62.
- Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research*, 22, 376-398.
- Muthén, L. K., & Muthén, B. O. (2004). *Mplus user's guide* (3rd ed.). Log Angeles, CA: Muthén & Muthén.
- Muthén, B. O., & Satorra, A. (1989). Multilevel aspects of varying parameters in structural models. In Bock, R.D. (Ed.), *Multilevel analysis of educational data*. San Diego: Academic Press.
- Niu, W., & Sternberg, R. J. (2001). Cultural influences on artistic creativity and its evaluation. *International Journal of Psychology*, 36(4), 225-241.
- Podsakoff, P.M., MacKenzie, S.B., Lee, J.Y., & Podsakoff, N.P.(2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879-903.
- Rand, T. M., & Wexely, K. N. (1975). Demonstration of the effect, "similar to me", in simulated employment interview. *Psychological Reports*, 36, 535-544.
- Rabe-Hesketh, S., Skrondal, A., & Zheng, X. (2007). Multilevel structural equation modeling. In S.-Y. Lee (Ed.), *Handbook of Latent Variable and Related Models* (pp. 209-227). Amsterdam: Elsevier.
- Raudenbush, S. W. (2003). The quantitative assessment of neighbourhood social environments. In I. Kawachi, L. Berkman, and F. Oxford (Eds.), *Neighbourhoods and health* (pp. 43-69). Oxford University Press.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd Ed.). Newbury Park, CA: Sage.
- Raudenbush, S. W., & Sampson, R. J. (1999). Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighbourhoods. *Sociological Methodology*, 29, 1-41.
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon Jr., R. T. (2004). *HLM 6: Hierarchical Linear and Nonlinear Modeling*. Lincolnwood, IL: Scientific Software International, Inc.
- Raza, S. M., & Carpenter, B.N. (1987). A model of hiring decisions in real employment interviews. *Journal*

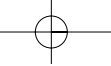


- of Applied Psychology*, 72, 596-603.
- Roberts, J. K. (2002). The importance of the intraclass correlation in multilevel and hierarchical linear modeling design. *Multiple Linear Regression Viewpoints*, 28, 19-31.
- Robinson, W. S. (1950). Ecological correlations and the behaviour of individuals. *American Sociological Review*, 15, 351-357.
- Snijders, T. & Bosker, R. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modelling*. Thousand Oaks, CA: Sage Publications.
- Stephenson, R. (2006). Contextual influences on the use of health facilities for childbirth in Africa. *American Journal of Public Health*, 96, 84-93.
- Tabachnick, B. G., & Fidell, L. S. (2006). *Using Multivariate Statistics* (5th Ed.). Boston, MA: Allyn and Bacon.

附錄一：IAF 量表題目的組內與組間樣本共變矩陣

	E1	E2	E3	E4	S1	S2	S3	R1	R2	R3
E1		.129	.141	.147	.127	.119	.122	.137	.138	.144
E2	.254		.151	.129	.148	.149	.141	.136	.132	.137
E3	.200	.258		.136	.149	.145	.146	.143	.144	.152
E4	.217	.207	.208		.128	.119	.124	.142	.151	.158
S1	.213	.217	.213	.241		.150	.148	.139	.138	.149
S2	.177	.191	.188	.209	.295		.141	.134	.129	.136
S3	.185	.205	.183	.213	.270	.276		.140	.136	.143
R1	.237	.271	.257	.255	.310	.287	.285		.161	.168
R2	.255	.283	.259	.273	.322	.299	.304	.461		.181
R3	.288	.319	.291	.302	.356	.325	.343	.506	.574	
W	.482	.432	.388	.434	.479	.432	.440	.558	.612	.766
B	.155	.159	.165	.156	.156	.150	.151	.164	.172	.195

註：下三角區域為整合組內觀察矩陣 (S_{PW})，上三角區域為組間觀察矩陣 (S'_B)。W 與 B 為各題的組內變異數與組間變異數。這些數據皆由 Mplus 估計得出



附錄二：本研究所使用的 Mplus 語法（以 MS211 模型為例）

```
TITLE:          2level MLSEM ANALYSIS (MS211)
DATA:           FILE IS g1d.dat;
VARIABLE:       NAMES ARE E1-E4 S1-S3 R1-R3 clus;
                 CLUSTER = clus;
ANALYSIS:       TYPE IS TWOLEVEL;
MODEL:

                %WITHIN%
                Exp1 BY E1-E4;
                Similar1 BY S1-S3;
                Rating1 BY R1-R3;
                Rating1 on Exp1 Similar1;

                %BETWEEN%
                Exp2 BY E1-E4;
                Similar2 by S1-S3;
                Rating2 BY R1-R3;

                Exp2 WITH Similar2;
                Rating2 on Exp2 Similar2;

OUTPUT:

                SAMPSTAT;
                STANDARDIZED;
                CINTERVAL;
```

Multilevel Latent Variable Modeling of Contextual Variables: Application of Multilevel Structural Equation Modeling on Rater Effects

Hawjeng Chiou

Department of Business Administration National Central University

The estimation of latent construct is the core issue of social and behavioral science. Structural equation modeling is the most important paradigm for estimation and for exploring the relationships among constructs. If the data for analysis involving multilevel or clustered structure, however, the multilevel SEM have to be applied for the special form of dataset. One of the features of multilevel data is that the observed variables in the individual level could be aggregated as a contextual variable into higher level. In terms of latent construct, contextual variable may behave as a latent form and should be estimated in a way of factorial analysis instead of manifest procedure. In a circumstance that the latent contextual variables have significant effects on dependent variable is called latent contextual effects. The purpose of present paper is to clarify the significance of the latent contextual effects in the social research. Rating data of oral examination

examined in the present study contains a total of 76 raters and 841 interviewees. A series of single and multiple level SEM models was proposed for examining the latent contextual effects. Results of Mplus indicated that the MSEM modeling can effectively extract the latent contextual variables and furthermore used to explain the dependent variables. Different factorial structure on latent contextual variables in both macro and micro level reveals that the contextual model is not equivalent across levels. The present study proved that the MSEM approach to examine contextual variables is more flexibility than traditional MLM paradigm. However, technical as well as methodological issues have to be solved.

Keywords: *contextual variables, latent contextual variables, multilevel data, multilevel structural equation modeling, rater effects*

