臺大管理論叢 2009/6
第19卷 第2期 263-294

多層次模型方法論：階層線性模式的關鍵議題與試解

Methodology of Multilevel Modeling: The Key Issues and Their Solutions of Hierarchical Linear Modeling

溫福星 ／東吳大學國際經營與貿易學系助理教授
Fur-Hsing Wen, Assistant Professor, Department of International Business, Soochow University

邱皓政 ／國立中央大學企管系副教授
Haw-Jeng Chiou, Associate Professor, Department of Business Administration, National Central University

Received 2007/10, Final revision received 2008/9

摘要

組織與管理研究的資料多涉及階層特性，因此多層次模式是當代組織與管理領域的重要研究典範之一，而階層線性模式則是多層次研究主要分析技術。本文針對階層線性模式在多層次研究上有關於技術、測量與方法論所遭遇的問題，進行文獻整理並討論其實務意涵，並依據使用時機、抽樣議題、資料彙總、分析方法、與模式評估與估算等五方面整理出十二項重要議題與解決之道，包括：組內相關係數的意義與判斷準則、多層次研究各層的樣本數問題、變數聚合成組織層次問題、中心化的意義、固定效果與隨機效果的設定、估計法的選擇、適配度的比較、解釋變異量的計算、多元共線性的分析、強韌性標準誤、第一層誤差項變異數與實證貝氏估計值的運用。文中除了進行原理討論之外，並試圖就實務應用的解決方法提出說明。

【關鍵字】階層線性模式、多層次模式、組內相關係數

Abstract

Multilevel study is one of the important contemporary research paradigms in organizational and management field, due to the fact that the data collected in organizational and management research frequently involve nested structure. Hierarchical linear modeling is the most frequently used technique for analyzing multilevel data. In this paper, the twelve key issues in multilevel research and/or using HLM software were reviewed, and provided their exact meanings and feasible solutions. The 12 key issues included intraclass correlation coefficient, sample sizes, formation of organizational constructs, centering problems, fixed effect and random effect, estimation methods, goodness of fit index, explained variation, multicollinearity, robust standard error, level one variance equation, and empirical Bayes estimates. In addition to the discussions on principles of the issues, implications and solutions to research practices were illustrated in the present paper.

【Keywords】hierarchical linear modeling, multilevel modeling, intraclass correlation coefficient

壹、緒論

多層次研究 (Multilevel Research) 可以說是當代組織與管理領域最熱門的新興研究典範 (Bliese, Chan, & Ployhart, 2007; Kozlowski & Klein, 2000)，而階層線性模式 (Hierarchical Linear Modeling；HLM) (Bryk & Raudenbush, 1992) 則是一種將迴歸擴展到階層資料結構的統計分析技術。雖然目前學界已有多種不同的方法與軟體來分析多層次資料，並快速擴展到縱貫資料、非線性模型、潛在變數模型的應用，但若非 Goldstein (2003)、Snijders 與 Bosker (1999)、Bryk 與 Raudenbush (1992) 等學者在階層線性模式的原理進行深入探究並持續開發便捷的軟體，多層次研究或許未有今日的成熟發展。

事實上，社會科學研究者並非無視於多層次分析的重要性，例如早在一甲子之前，Robinson (1950) 就已經指出不同層次的資料若無適當的分析與推論將會犯下嚴重的錯誤。但階層線性模式遲至九十年代才獲得發展，主要的原因之一是受限於一般線性模式的傳統分析框架。在過去，傳統一般線性模式 (General Linear Models；GLM) 的迴歸分析或是變異數分析 (註[1])，僅能處理單一層次變項間的關係，而無法同時處理包含個體與總體等不同層次的跨層級資料 (註[2])。如果要將不同層次的分析單位整合在一個線性模式中，必須將傳統的 GLM 擴充到 HLM。在此同時，模型的提出必須要有充分的理論依據或邏輯推論，再加上資料分析方法及結果解釋均要與研究設計有關，因此階層線性模式不僅是統計分析技術，更是一種統計方法學 (Statistical Methodology) (Courgeau, 2003)。

關於 HLM 的原理已經有相當多的文獻介紹 (Goldstein, 2003; Hofmann, 1997; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999)，本文不再贅述這些基本原理與技術，而是著眼於使用 HLM 背後的思考脈絡與所潛藏的各種問題。在使用這技術進行研究之前，研究者除了瞭解分析程序與解釋方法之外，更必須熟悉多層次研究與傳統研究的差異，並進一步瞭解新方法的限制與可能帶來的新問題。例如，HLM 與 GLM 的最大差異在於 HLM 模式引進了不同層次的誤差項，而 GLM 模式只有一個誤差項，而且是屬於個體層次。過去，我們若要研究個體與總體等不同層次的跨層級資料，共有兩種類型的迴歸分析可以處理，一個是將總體層次的資料解構 (Disaggregated) 或是進行虛擬化處理 (Dummy)，使成為個體層次資料來進行個體層次的迴歸分析；其次，是將個體層次的資料聚合 (Aggregated) 成總體層次的變項，進行總體層次的迴歸分析。這樣的分析方法是取決於研究目的，如果我們想研究的是個體

註[1] 指固定效果模式的變異數分析。

註[2] 指個體與總體等不同層次的誤差項。

與總體等不同層次的跨層級資料對個體層次結果變項的影響時，解構方式的個體層次迴歸分析就會遇到資料獨立性與同質性假設被違反的可能，得到所估計的迴歸係數標準誤被低估的情形發生，導致容易拒絕虛無假說的型 I 錯誤膨脹結果。這是因為忽略了相同總體層次下的個體間存在高度的相關，亦即個體與總體層次的資料彼此間具有「內屬、巢套、叢集、鑲嵌 (Nested、Clustered、Embedded)」的結構特性，例如員工內屬於公司、團隊成員內屬於團隊等，由於多這一層關係，相同公司下的員工、相同團隊下的成員，因組織文化、組織氣氛等因素的潛移默化，使得員工或成員在同一組織 (公司或團隊) 內較為相似，在不同的組織間較相異。這樣的資料有別於迴歸分析的資料來自於獨立且同質母體的假設，更有別於變異數分析的隨機分派。HLM 透過隨機效果的設計，在迴歸模式引進總體層次的誤差項，透過其變異數來捕捉內屬關係的組內相關進行參數的估計與檢定，因此所得到的結果較為正確，表 1 為 GLM 與 HLM 主要差異的比較。

表 1 GLM 與 HLM 差異比較表

項目	GLM	HLM
樣本組成	所有樣本相互獨立	樣本非獨立
誤差項	單一層次誤差	個體層次誤差
		總體層次誤差
交互作用項	同一層級	同一層級
		跨層級
模式適配指標	R square	Pseudo R square
	Adjusted R square	Deviance

以下，本文將簡述 HLM 的模式特性與基本假設，並根據使用時機、抽樣議題、資料彙總、分析方法、與模式評估與估算等五個方面來探討應用 HLM 的一些關鍵議題，並提出我們的看法。

貳、HLM 的模式與假設

一、階層線性模式

前面緒論已經說明了，當研究資料具有內屬或鑲嵌的特性時，對個體層次結果變項的解釋效果可能會來自不同層次，再加上資料間的相似性，此時，傳統的 GLM 無法處理多層次的資料結構。例如社經地位 (X) 對於學業成績 (Y) 的影響，社經地位作為解釋變項，研究者的假設為社經地位的高低影響學業成績 (方程式 (1))。但是，如果樣本來自於不同的學校，每一個學校的整體社經地位可能有所不同。此時，社經

地位就有個體 (level 1) 與總體 (level 2) 兩個層次的影響，總體層次的社經地位以 Z 表示，Z 對結果變項的影響以方程式 (2) 與 (3) 表示，方程式 (1)、(2) 與 (3) 的組合即稱為階層線性模式：

$$\text{Level 1: } Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + \varepsilon_{ij} \quad (1)$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + u_{0j} \quad (2)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} Z_j + u_{1j} \quad (3)$$

方程式 (1) 代表第一層 (level 1) 的迴歸模式，也就是個體層次解釋變項與個體層次結果變項的關係，而方程式 (2) 與 (3) 代表第二層 (level 2) 的迴歸模式，u_{0j} 與 u_{1j} 表示為第二層迴歸模式的誤差項，假設服從以 0 為平均數、 τ_{00} 與 τ_{11} 為變異數、 τ_{01} 為共變數的二元常態分配。值得注意的是，(2) 與 (3) 式的結果變項是第一層迴歸模式的參數，亦即以各校個體層次的迴歸分析截距項與斜率項作為結果變項進行高階迴歸分析，而不是個體層次的結果變項數據。若將這三式結合，將 (2) 與 (3) 式代回 (1) 式得到整合方程式，稱為混合模型 (Mixed Model)：

$$\text{Mixed: } Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} Z_j + \gamma_{11} Z_j X_{ij} + u_{0j} + u_{1j} X_{ij} + \varepsilon_{ij} \quad (4)$$

(4) 式中，迴歸係數 γ_{00} 為平均截距、 γ_{01} 就是總體層次解釋變項對結果變項的直接影響或稱為脈絡效果 (Contextual Effect)、 γ_{10} 為個體層次解釋變項對結果變項的影響、而 γ_{11} 為跨層級交互作用效果(Cross-level Interaction Effect)。若將 (4) 式倒數第二項與第三項拿掉，則就是我們一般的多元迴歸分析，有總體層次的解釋變項 Z_j、個體層次的解釋變項 X_{ij}，以及交互作用項 $Z_j \times X_{ij}$，不過這是跨層級兩個變項的乘積項，不是傳統 GLM 所處理的交互作用項。

二、階層線性模式的基本假設

如果我們將(1)、(2)與(3)式的解釋變項 X_{ij} 與 Z_j 都拿掉，則這三條方程式縮減為兩條：

$$\text{Level 1: } Y_{ij} = \beta_{0j} + \varepsilon_{ij} \quad (5)$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + u_{0j} \quad (6)$$

這就是最簡單的 HLM 模型，稱為零模型 (Null Model) 或是具隨機效果的變異數分析模型，所有的 HLM 分析都是從零模型開始。零模型除了有 GLM 的假設外，因

為它多出了不同層次的誤差項，其假設：

(1) ε_{ij} 為獨立且服從以 0 為平均數、σ^2 為變異數的常態分配；

(2) u_{0j} 為獨立且服從以 0 為平均數、τ_{00} 為變異數的常態分配；

(3) ε_{ij} 與 u_{0j} 相互獨立。

讓模型稍微複雜一點，以最常用的簡單迴歸為例，它比零模型多一個個體層次的解釋變項 X_{ij} (方程式 (7))。此外，第二層的兩條迴歸方程式，一條有誤差項 (方程式 (8))，另一條則無 (方程式 (9))，其設定如下：

$$\text{Level 1: } Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \varepsilon_{ij} \tag{7}$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + u_{0j} \tag{8}$$

$$\beta_{1j} = \gamma_{10} \tag{9}$$

(7) 式為一般的簡單迴歸分析，差別只在於迴歸係數有下標符號。(8) 式稱為隨機效果模式，(9) 式稱為固定效果模式，這三條方程式在 HLM 稱為隨機截距模型或具隨機效果的共變數分析模型。

我們就以 (7)、(8)、(9) 為例，為了討論一項特殊效果的檢定 (相對 (5)、(6) 模式)：γ_{10} 效果的統計檢定力，我們需要對於理論模型與真實資料的適配 (Goodness of Fit) 或離異數 (Deviance)、樣本數、以及 γ 係數的估計方法加以定義。因為統計檢定力是以發現一個具有顯著性效果的機率來定義，此時該效果必須存在於模型中，且利用顯著性考驗來檢驗之。顯著性考驗非常倚重被估計的 γ 係數標準誤的計算，因此 γ 係數的估計抽樣變異決定了迴歸模型的檢定力。如果抽樣變異估計出來的數值很小，檢定力就會很大；反之，檢定力就會很小。除此之外，如同迴歸分析一樣，必須估計誤差變異數。在 HLM 中，除了估計個體層次誤差項的變異數 σ^2 外，尚必須估計第二層誤差項 u_{0j} 的變異數 τ_{00}，有時還必須估計 τ_{11} 與 τ_{01}，這是和 GLM 不一樣的地方。

參、階層線性模式的關鍵議題與試解

HLM 是一種從線性迴歸模型所衍生出來的統計方法，換言之，所有影響迴歸分析的因素都會影響到 HLM。因此，當我們從單層次提升到多層次分析的同時，也必須有一個新的解答或作必要的修正，例如解釋變數的中心化 (Centering)，正因為 HLM 具有階層結構，因而有特殊的跨層級交互作用，因此對於解釋變數的中心化處理也就特別的重要。而面對不同層級的多重變異來源 (σ^2與 τ_{00})，使得我們在定義解釋變異 (Explained Variance) 時顯得格外複雜，而且要以不同以往的方法來界定。以下我們就分別從五個構面來探討十二個 HLM 的重要議題。而在下面的討論中，我們泛

指的是結果變項為連續變項的討論，並且以兩層結構的 HLM 為例；此外，我們以 MLM (Multilevel Modeling) 來稱呼多層次模式，有時也與 HLM 交替使用。

一、使用時機

Hofmann (1997) 指出，使用 HLM 的第一個議題就是什麼時候可以或一定要使用 HLM？關於這個問題，除了研究所引用的理論架構與待檢假設涉及多層次結構之外，當研究者所蒐集的資料來自於集群抽樣 (Cluster Sampling) 時將具有鑲嵌特性，資料鑲嵌則會造成資料獨立性的違反，導致 GLM 不能使用。而用來偵測資料是否違反獨立性，多以「組內相關係數」(ρ Intraclass Coefficient；ICC) 來判斷，其公式如下：

$$ICC = \rho = \frac{\tau_{00}}{\tau_{00} + \sigma^2} \tag{10}$$

上式是以零模型為基礎所計算而得，一方面是因為 HLM 的分析都是從零模型開始，也就是模式 (5) 與 (6)，更重要的是 ICC 的計算必須在模型中沒有任何解釋變數的情況下，來估計組間變異數佔總變異數的大小。ICC 所代表的是結果變項的總變異數中可以被組間解釋的百分比，如果 ICC 很大，則代表存在組間差異，組的效果不能並忽略。

ICC 的真義：組內相關重要嗎？

ICC 被稱之為「組內相關係數」，其目的在衡量資料違反獨立性的程度，它代表任一群體 (Class 或 Group) 內任兩位受試者，其結果變項間相關的期望值，是用來捕捉組內資料的相似性或是資料的非獨立性 (Hox, 2002)。換言之，組內相關係數即是個體間相依程度的測量。當個體間因為時間與空間因素導致有較高的相同經驗與共通性時，彼此的相似性提高，甚至相似性高到了某個程度時，每一個人可以說是完全相似的複製品。所以，計算零模型的 ICC，就是用來檢視是否進行 HLM 的關鍵條件。

組內相關會影響 GLM 的誤差變異數，誤差變異數代表了所有遺漏的解釋變數與測量誤差的影響，而且這些誤差具有相互獨立的基本假設。在 GLM 中，解釋變數的遺漏被假設為隨機、非結構性的現象。但是當資料具有鑲嵌特性時，此一假設是有問題的。例如，在學校效能的研究中，學校氣氛或同儕壓力可能是一個沒有被測量但具結構性影響的變數，此時同一個學校或班級學生的誤差項即會以組內相關係數的型態浮現。此時，GLM 對觀察值需為獨立的假設即遭到違反。在文獻上，當存在組內相關時，迴歸分析的基本假設被嚴重違反，其影響會反映在型 I 錯誤率的增加。例如 30 個在同一個班級的學生已非 30 個獨立的個體，學生之間獨立性的消失，取決於班級內個體的相似性或是組內同質性的高低，因此組內相關的強度決定了觀察值間真實的獨立性與否。

零模型的 ICC 要多大呢？

在 HLM 中，是透過檢視零模型的 ICC 來判斷是否必須以多層次統計技術來分析多層次資料。ICC 在平衡設計下有其機率分配，因此可以計算其估計值的標準差 (Donner, 1986)。但在社會科學的研究，各組的樣本數很難都相等，因此檢定 ICC 的顯著性是相對不易。Snijders 與 Bosker (1999) 建議透過檢定固定效果下變異數分析的 F 考驗，或是檢定隨機效果下 τ_{00} 的卡方檢定，據以判斷組間變異數是否有效果，以便進行多層次分析。除了以 F 或卡方檢定考驗組間變異數的效果外，Cohen (1988) 認為 ICC 在不同的研究領域其差異很大，因此在不同領域有不同的 ICC 判斷值。不過，Cohen (1988) 認為當 ICC 小於 0.059 時，算是相當小的組內相關，而介於 0.059 與 0.138 則算是中度相關，至於高於 0.138 則算是高度的組內相關。他認為中度程度的組內相關就不能忽略其相似性的存在，因此當 ICC 大於 0.059，則必須考慮多層次的統計分析。而 Bliese (2000) 研究發現 ICC 一般在 0.05 與 0.20 的範圍內，而 James (1982) 的 ICC 研究發現中位數為 0.12，雖然要判斷多大的 ICC 是不能被忽略沒有很精確的門檻，但約略差不多。

不管如何，Luke (2004) 認為是否必須以 HLM 來分析可以從三個角度來檢視，第一是理論層次 (Theoretic View)，是否所研究的就是屬於多層次的議題；第二是統計層次 (Statistical View)，資料結構在分析上是否違反了迴歸分析對誤差項的獨立性與同質性假設；第三是實證層次 (Empirical View)，是否資料的 ICC 夠大到不能忽略其對迴歸分析結果的影響，讀者可依據上述來判斷是否使用 HLM。

二、抽樣議題：樣本大小

因為 HLM 涉及到多層次，因此要考慮到不同層次的抽樣問題。在機率抽樣中，與 HLM 最相關的就是集群抽樣。以兩層結構的 HLM 為例，就必須考慮從許多公司中抽出公司，再由這些公司內抽出員工，這樣的抽樣設計就存在鑲嵌、內屬的特性。至於抽多少公司、公司內抽多少員工就涉及到樣本數的問題。

在社會科學的研究中，樣本數的大小影響到統計推論的有效性。我們都知道在民意調查中，假設您會投票的結果是服從二項分配，在 95% 的信心水準下，若抽樣誤差要在正負 3% 以內，則樣本數必須為 1067，所以這個樣本數的決定取決於抽樣分配、效果量、顯著水準與檢定力。所以，有關實證研究樣本的大小是一個理論層次的問題。

到底樣本數指的是誰？多大的樣本數才是適合？

而在 HLM，由於涉及跨層次，因此樣本規模的決定更為複雜，甚至牽涉到 ICC。由於資料鑲嵌特性而喪失獨立性，因此 ICC 與樣本大小的關係相對完全獨立樣本可以以下式的設計效果來表示：

$$\text{design effect}\ (\eta)=1+(\bar{n}_j-1)\times ICC \tag{11}$$

(11) 式中的 $\bar{n}_j$ 為平均的組內樣本數，其設計效果的涵義是當在集群抽樣設計且存在組內相關時，在所抽取的組內樣本數 $\bar{n}_j$ 要達到完全與 $\bar{n}_j$ 個獨立樣本有相同的效果，必須將樣本增加到 η 倍。換言之，ICC 使得 $\bar{n}_j$ 個樣本所提供的資訊必須打折扣，這些折扣是因為樣本之間的相似性所造成的，它對於迴歸分析的估計，是對迴歸係數估計值的標準誤直接產生影響，使得真正的標準誤是忽略組內相關時的 η 倍 (Kish, 1995)。

除了第一層樣本數與 ICC 的關係外，Kreft (1996) 建議採用 30/30 準則，亦即總體層次不少於 30 組，每組不少於 30 人來決定樣本規模。但是如果研究者偏重於個體層次與總體層次交互作用時，可以調整比率為 50/20；如果重視隨機效果的檢驗，甚至可以調整為 100/10。Heck 與Thomas (2000) 認為在 HLM 中，關於模式中的迴歸係數和變異數可以利用最大概似法或是貝氏估計法來估計，以兩層的 HLM 為例，當第二層的 J 很大時，且各個 j 內的樣本數也很大，具有一致性的估計法都適用，但是若組內樣本數太少時，則參數的估計會有問題。

Snijders 與 Bosker (1994) 計算了增加組數與檢定力的消長關係，以及所耗費的成本問題。研究結果指出，增加組數比增加組內樣本數要耗費許多倍的成本。Mok (1995) 的模擬研究發現，較多的學校較少的學生會比較少的學校較多的學生的偏誤要小、效能更好。Bassiri (1988) 和 Van der Leeden 與 Busing (1994) 分別對於跨層級交互作用的檢定力進行探討，這兩個研究均指出，為了獲得理想跨層級交互作用的檢定力，至少要有 30 個組，每組要有 30 個觀察值。同時，研究也發現當有 60 個組，每一組 25 個觀察值時，會得到相當理想的檢定力。在較少組數的時候，例如 30 組時，每一組需要更多的樣本以獲得 0.90 的檢定力。當組數較大的時候，例如 150 組時，每一組只要 5 個觀察值就可獲得 0.90 的檢定力。 Bassiri (1988) 發現多選擇一些組，會比在各組內多選擇一些觀察值來得有幫助。而 Kreft 與 de Leeuw (1998) 認為，一般來說為了獲得足夠檢定力，觀察值要多，除非研究者所探究的現象有非常強且容易被偵測到的效果。然而，研究最理想的樣本數為何，每一個研究都有所不同。當組數很少時，隨機成分會被低估，或有較大的標準誤。對於跨層級效果，要有足夠檢定力，組樣本數不能太少，且組數要大於 20。而 Maas 與 Hox (2005) 的研究則指出，為了能夠得到固定效果的標準誤估計值能不偏，至少要有 50 組的樣本數。

值得注意的是，這些文獻都是在既有的設定下以模擬的方式所進行的結果推論，他們都是在探討不同主題下例如固定效果、隨機效果或是標準誤不偏情況的檢定力要求，必須的樣本數限制，因此並沒有一個判斷標準。不過從這些文獻結果來看，對第二層樣本數的要求是越多越好，一來是估計法，二來是檢定力。主要原因在於第二層誤差項變異數與共變數的估計，依最大概似法的定義，隨機效果的變異數估計

值，其抽樣分配是在大樣本下，極限分配才服從常態分配，因此樣本數是越大越好。而且在 (11) 式中，當第一層的平均樣本數越大，則很小的 ICC 都會造成很大的設計效果，使得有效樣本數驟減，所以第二層樣本數在 MLM 中是比較重要的因素。此外，有關於追蹤資料或是對偶數據的分析，因第一層樣本數常受到限制，考慮檢定力緣故，增加第二層樣本數是解決問題的方法。不管如何，當模式越複雜時，兩層的樣本數都是要隨之增加。

三、資料彙總：組織構念

多層次與單一層次的研究，最大差異引進了總體層次的解釋變項與誤差項到迴歸模式中。因此，總體層次解釋變項的取得就成為另一個研究重點。在教育與心理研究上，脈絡模型被定義成包含有兩類變數的模型：個體層次變數與聚合脈絡變數 (Aggregated Context Variables)，這個聚合脈絡變數就是總體層次解釋變數，是由個體層次解釋變數經計算而來。例如當我們蒐集學生的資料，而學生內屬學校，則學生學業成績被社經地位所影響，同時也被其所屬學校的平均社經地位所解釋。此時社經地位使用了兩次：學生社經地位與學校平均社經地位。在文獻中，迴歸分析若包含了此種聚合脈絡變數則稱為脈絡模型 (Duncan, Curzort, & Duncan, 1966)。

Pedhazur (1997) 將脈絡效果定義成控制住某一個變項在個體層次的差異後在團體層次的淨效果，例如社經地位對於總統大選投票行為的影響，每一個人除了自己的社經地位，在其居住地區也會有一個社經地位。此時，個體層次的社經地位與總體層次的社經地位是兩個的數據 (雖然總體層次的社經地位是取該地區居民社經地位的平均)，若將這兩個數據放入迴歸模型，得到地區社經地位的迴歸係數即為脈絡效果。除了以聚合方式將個體層次變項提升至總體層次變項外，整體特徵 (Global Characteristics) 也可以是總體層次變項。整體特徵被定義成對於脈絡特性直接加以測量所得的變數 (Lazarsfeld & Menzel, 1970)，而不是從個體層次計算而來。

另外在組織研究中，對於總體層次解釋變項的蒐集有幾種型式，Chan (1998)、Chen、Mathieu 與 Bliese (2004)、Kozlowski 與 Klein (2000) 和林鉦棽與彭台光 (2006) 都將這幾種型式整理的相當完整。類似心理與教育的研究，用來捕捉組織研究中組織層次構念方式主要有兩種：composition 與 compilation (Chan, 1998; Kozlowski & Klein, 2000)；而同心理與教育的整體特徵，組織研究中有整體法 (Global Approach) (Kozlowski & Klein, 2000)；林鉦棽與彭台光 (2006) 針對 Kozlowski 與 Klein (2000) 所提的 global、composition與compilation方式所獲得的變數分別稱為共通單位變數 (Global)、共享單位變數 (Shared) 與共構單位變數 (Configural)。除此之外，他們提出另一種稱為共塑單位變數 (Formative)，詳細的定義與之間的差異請參閱他們的文章。

本研究有關於變數聚合問題是聚焦於共享單位變數的形成 (Kozlowski & Klein,

2000)，稱為組織構念或是組織層次構念。由於組織不是個人，無法填答問卷，因此組織構念的產生是透過組織成員資料的取得，因此要形成有效的組織構念必須要有嚴謹的程序。如同脈絡變數一樣，聚合後的脈絡變數是否和個體層次變數具有相同的意義，取決於 composition 與 compilation。換言之，從另一個角度來看就是個體構念與組織構念的恆等性 (Van de Vijver & Poortinga, 2002)，或是意義的遷移問題 (邱皓政，2007; Snijders & Bosker, 1999)。所謂的 composition 與 compilation 分別是指構成組織構念是否來自同質 (Isomorphic) 或是來自相異 (Discontinuity) 的個體層次構念。這裡的共享就是組合 (Composition) 的意思，因為要形成有效的共享單位變數，必須探討組內變異與組間變異的問題，意即這些個體層次的變數聚合成總體層次變數的效度與信度問題。有關此多層次的效度與信度，將牽涉到 r_{wg} 與 *ICC* (2) 的計算。所謂的 r_{wg} 是指共識程度 (Agreement)，而 *ICC* (2) 是指信度 (Reliability)。 r_{wg} 是指針對總體層次構念的測量指標，在相同的組織下其組織成員回答這些測量題目的變異程度，當這個變異程度與在組織成員在隨機或是均等分布的假設下作答的變異程度所計算的比值，這就是 r_{wg} (或 $r_{wg(j)}$)。因此，有多少組織就有多少個 r_{wg}，是用來計算組織內的成員針對總體層次的構念是否具有高度的共識與相同的看法，此指標可以作為個體層次變項聚合成組織構念的先決條件。由於個體層次的題項與組織層次的構念牽涉到問題的陳述，在組合 (Composition) 或聚合 (Aggregated) 過程有兩種形式用來編制組織構念題項，分別為直接共識法 (Direct Consensus / Self-referenced) 與參考遷移共識法 (Reference Shift Consensus / Group-referenced) (Chan, 1998)。至於，*ICC* (2) 則來自 *ICC* 相似的概念，它是計算各組織成員在各題項的得分，經求平均數作為組織的分數，然後計算組間變異數佔組織平均數變異數的比值，代表的是這些題項平均數在組織間的一致性程度，換言之也就是信度的意思。各組平均數的變異數公式如下：

$$Var\left(\overline{Y}_j\right) = \tau_{00} + \frac{\sigma^2}{n_j} \tag{12}$$

而第 j 組的信度為：

$$\lambda_j = \frac{\tau_{00}}{\tau_{00} + \dfrac{\sigma^2}{n_j}} \tag{13}$$

經計算 λ_j 的平均數就是 *ICC* (2)。如果信度夠大，意思是用組織成員的平均數代表組織的分數是可以被信賴，其可靠性是夠的。至於 r_{wg} 與 *ICC* (2) 的差異， r_{wg} 是在計算組內變異數，而 *ICC* (2) 則是計算組間變異數。這兩個指標的判斷準則，一般都以 0.7 為切割標準，當研究的個體層次變項經計算其 r_{wg} 與 *ICC* (2) 都高於 0.7 時，我們是有足夠的證據認為這些個體層次的變項可以聚合成組織構念；反之，則代表這些

題項是屬於個體層次，不能經計算組織平均數做為高層解釋變項。

過去在進行組織構念的檢驗時，由於遷就於統計理論與技術，而以 r_{wg} 與 *ICC* (2) 作為判斷依據。但在現今的多層次技術下，可以利用多層次驗證性因素分析 (Multilevel Confirmatory Factor Analysis；MLCFA) 方法來確認個體層次構念與組織層次構念 (黃芳銘、溫福星，2007; 邱皓政，2007; Hox, 2002; Muthen, 1994)。MLCFA 是考慮在多層次下組織內個體層次資料間的相關性，透過將總變異數共變數矩陣拆解為組內與組間兩個矩陣，利用 CFA 分別求算個體層次與組織層次的因素結構，再檢驗適合度指標與因素負荷量，判斷是否可以形成個體層次與組織層次的構念，但這個方法適用於直接共識法。

四、分析方法

關於多層次資料分析方法，就是 HLM 的設定問題，本節分成三個部份來探討，分別是總平減與組平減、固定效果與隨機效果、最大概似法與受限最大概似法。

(一) 中心化問題：總平減與組平減的影響為何？

在簡單迴歸分析中，如果解釋變項未經任何變數變換，我們將會遇見所估計出來的截距項會沒有意義。因為截距項是在解釋變項皆為 0 的情況下的結果變項，此時截距項是數學方程式下的一個輔助項，要能讓截距項在解釋上有意義，不是解釋變項要包含 0 外，就是要經中心化 (Centering，平減或平移) 或是標準化處理。

Aiken 與 West (1991) 曾整理傳統迴歸模型中心化的影響，中心化是指每個人的解釋變數都減去同一個數值，通常是總平均數，但也可能是其他數值。基本上，每一個分數減去同一個數值並不會改變數據間的關係。Aiken 與 West (1991) 指出，如果模型中僅有個體層次變數，對其加減某一常數並不會對此變數的變異數產生影響，至於與其他變項的共變數與相關都不會有任何改變。中心化的作用是在使變數改以離均差的形式，僅影響截距的數值以及其標準誤，這樣的處理另一個目的是在估計斜率上較容易。

實務上，傳統 GLM 之所以要進行解釋變數的中心化，目的在改變截距的意義以便於解釋。在這種情況下，解釋變數的中心化使得截距具有意義，是指當解釋變數為平均數時反應變數的數值。在 HLM 中，總平減 (Grand Mean Centering) 比組平減 (Group Mean Centering) 簡單許多，因為大家都減去同一個數值。而組平減是每個人的解釋變數減去各組平均數，因各組的組平均並不相同，部分人所減掉的數值也不相同。使用此一新的變數來取代原始分數形式的解釋變數，將使模型產生變化，包含固定效果與隨機效果。但是組平減有個好處，就是當所有第一層解釋變項都採組平減時，第一層迴歸方程式的截距項估計值就是各組結果變項的平均數。

至於使用原始資料不做任何平減，與總平減進行估計，雖然在截距項與截距誤

差項變異數的估計上不同，但並沒有改變其他解釋變項的估計，Kreft、de Leeuw 與 Aiken (1995) 稱這兩種模型為等值線性模型 (Equivalent Linear Models)。等值模型會有相同的適配度、相同的預測值、相同的殘差，至於兩者模型的參數估計值，則可利用數學方法來證明其間的關係。另外，Anderson (2004) 以統計相等性 (Statistical Equivalence)、參數相等性 (Parameter Equivalence)、參數穩定性 (Parameter Stability) 與迴歸係數的解釋性來比較解釋變項的不同平減處理發現，不平減的處理與以總平減，基本上是具有參數與統計的相等性。同樣地，Hofmann 與 Gavin (1998) 在脈絡效果與跨層級交互作用的研究上，發現對解釋變項不做平減和總平減結果可視為相同的模式。

Hofmann 與 Gavin (1998)、Mathieu 與 Taylor (2007) 的研究認為以總平減進行 HLM，一來可以避免共線性問題，二來其模式與不平減是統計等價模式，可以用來偵測脈絡效果與跨層級交互作用。而 de Leeuw 與 Kreft (1995) 認為在隨機係數模型下，若進行組平減，建議要把平減用的組平均數置回模型中，如果沒有這麼做，那麼研究者所估計得到的效果並沒有控制住組間的差異。因此，Raudenbush 與 Bryk (2002) 建議除非研究者有很清楚的理論，否則不適合配適以組平減的隨機斜率模型。

此外，有關於截距項和截距項標準誤的解釋，甚至與截距項相關的共變數等解釋也要特別小心。Kreft 與 de Leeuw (1998) 認為，有關於平減方法的選擇，除了對資料特性的瞭解外，分析的目的也很重要，如果要對第一層進行組平減，則最好能夠將各組平均數納入分析模式中，以利於對於組平均數效果的修正。Snijders 與 Bosker (1999) 建議：當在配適隨機斜率模型時，最好不要採用組平減，除非有清楚的理論，說明相對分數 ($X_{ij}-\overline{X}_{.j}$) 與結果變項有關，否則不建議使用。

什麼時候適合組平減呢？除了研究欲探討組內相對位置而不考慮組間的影響，例如池中蛙模式 (Frog Pond Model) (Hofmann & Gavin, 1998)，除此之外，Bickel (2007) 以美國總統大選研究為例，美國總統大選的選舉人團投票是採贏者通吃 (Winner-take-all) 的制度，各州的選舉人投票相當於組內迴歸一樣，不受到其他州結果的影響，這就是非常適合組平減。同時 Hofmann 與 Gavin (1998) 和 Bickel (2007) 都認為第一層變項一旦採用組平減時，其解釋上的意義與其他兩者不盡相同，在結果變項上的解釋是屬於平均的意思。

至於不平減、總平減或是組平減哪一個模型才是正確？這個問題並沒有辦法單從技術層次來回答，因為這三個模型都是「正確的」。特別是在複雜的 HLM 模型，如何選擇，必須就研究者本身對於資料本身的理解與理論知識，以及研究目的來考量。如果研究者感興趣的是模型能夠解釋反應變數變異量的多寡，而不是在於第二階層的效果，那麼利用原始資料來適配 HLM 是最簡單的作法。研究者不必去擔心如何平減的問題，因為一開始平均數就沒有被處理。同時，如果研究者較關心的是個體層

次解釋變項的表現，而不是組織層次間的差異時，不平減的模型也是最好的選擇 (Kreft et al., 1995)。

從以上的討論，我們可以發現平減的處理，尚牽涉到各種不同 HLM 的模型設定，以及研究欲探討的主題：脈絡效果、個體層次的關係、亦或是跨層級交互作用，甚至與固定效果或是隨機效果的選擇有關。不過我們可以從這些研究結果發現，透過平減確實可以減少多元共線性問題，也讓截距項較容易解釋，不過要注意的是其隨機效果的意義可能要加以注意。除此之外，不平減與總平減是等價的模式，唯一的差異就是截距項、以及與截距項有關的變異數與共變數，至於其它迴歸斜率的數值與意義則沒有改變。而 Enders 與 Tofighi (2007) 指出，若單獨研究脈絡效果時，則是要選擇總平減，如果只是關心個體層次解釋變項對結果變項的影響，則建議以組平減，至於研究跨層級交互作用效果是組平減較好。而 Hofmann 與 Gavin (1998) 也提到過當探討第二層解釋變項對結果變項的遞增效果 (Incremental) 時，是採總平減較適合，研究第二層解釋變項對結果變項的中介 (Mediational) 時，採用固定斜率以及總平減處理，或是以組平減外帶將組平均數置於截距方程式中；當考慮跨層級第二層解釋變項的調節作用時 (Moderational)，則以組平減較適當。同時，Wu 與 Wooldridge (2005) 也得到相似的結果。

我們以下面模式為例，說明平減方法的選擇：

$$\text{Level 1: } Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j} + X_{2ij} + \beta_{3j}X_{1ij}X_{2ij} + \varepsilon_{ij} \quad (14)$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + \gamma_{01}Z_{1j} + \gamma_{02}Z_{2j} + \gamma_{03}Z_{1j}Z_{2j} + u_{0j} \quad (15)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_{1j} + u_{1j} \quad (16)$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}Z_{2j} + u_{2j} \quad (17)$$

$$\beta_{3j} = \gamma_{30} \quad (18)$$

當 (14) 式的 β_{0j} 要有意義，則必須要所有解釋變項 X_{1ij}、X_{2ij}、Z_{1j} 與 Z_{2j} 都為 0 時，如果這些變項的全距不包含零時，連帶方程式 (15) 也沒有意義，因此必須對這些解釋變項進行平減。同樣地， γ_{00}、 γ_{10} 與 γ_{20} 要有意義，則必須 Z_{1j} 與 Z_{2j} 的全距包含 0。此外，將方程式 (15) 至 (18) 代回方程式 (14)，可以獲得混合模式，則 $X_{1ij}X_{2ij}$ 稱為組內交互作用，因為這兩個變項都是個體層次，$Z_{1j}Z_{2j}$ 稱為組間交互作用，因為都是組織層次，而 X_{1ij} Z_{1j} 與 $X_{2ij}Z_{2j}$ 皆稱為跨層級交互作用，因為這兩個解釋變項一個是個體層次，一個是組織層次。這些交互作用都是對個體層次的結果變項產生影響，至於這些交互作用的選擇設定，除了理論的依據外，實證的角度是去檢視上述方程式誤差項的分布是否同質性來加以考慮。基本上，牽涉到第二層的解釋變項是總平減，牽涉到第一層解釋變項的主效果也是總平減，若以組平減時必須將其組平均數置於第二

層截距方程式內；若是跨層級交互項，因為此項的效果是設為固定效果，因此個體層次解釋變項是要組平減。

不管如何，兩種平減方法皆會改變所有解釋變項的變異數共變數的資料結構，連帶所估計的迴歸係數與誤差項變異數皆與不平減的意義有些不同 (溫福星，2006)，Enders 與 Tofighi (2007) 建議在文章中去說明選擇該項平減方法的理由。

(二) 隨機與固定問題：不同的作法應如何選擇？不同的選擇之間有何差異？

在變異數分析中，類別自變項對於結果變項的影響有兩種不同的形式：固定效果模式 (Fixed-effect Models) 與隨機效果模式 (Random-effect Models)。固定效果模式是指當一個研究的類別自變項的水準個數 (*k*)，就是該變項所有可能的水準數 (*K* 組)，也就是樣本的水準數等於母體的水準數 (*k*=*K*)，此時類別自變數對於結果變項的影響，無須推論到其他的情境。而隨機效果模式是指研究所取用的類別自變項的 *k* 個水準，是從具有 *K* 個水準的母體中所抽取，亦即樣本的水準數小於母體的水準數，此時類別自變數對於結果變項的影響，是「隨機」取樣的結果，完整的自變項效果必須從樣本的 *k* 個水準推論到母體的 *K* 個水準，增加了一項抽樣與推論程序。

HLM 與 GLM 的最大差異，在於資料具有階層結構，低階的分析單位「嵌套」在高階的單位中。更具體來說，總體層次的分析單位是由個體層次的分析單位匯集或分組而得，因此，在 HLM 分析中，高階的分析單位的組間差異，即可能具有固定效果與隨機效果之別：當高階的組數 (*k*) 等於母體的組數 (*K*)，組間差異無須推論到其他情況，因此沒有抽樣問題，稱為固定效果模式；相對的，當高階的組數 (*k*) 是從母體的 *K* 組中隨機取樣，那麼高階的組間差異即有抽樣問題，在檢驗高階效果時，必須估計抽樣誤差，稱為隨機效果模式。Kreft 與 de Leeuw (1998) 認為模式的誤差項是隨機變數，隨機變數的意義是指從某一機率分配隨機抽樣的結果，所以在 HLM 中結果變項也是屬於隨機變數，隱含就是一種隨機效果。如果將 HLM 模型第二層以上的誤差項變異數設定為零，則各層誤差項的隨機變數效果消失，只存在迴歸係數這一固定部分，等於迴歸模式沒有隨機變數存在，這就是固定效果模型。在有隨機效果下的固定效果的檢定中，其檢定統計量為 t 統計量，自由度為第二層樣本數減去該方程式的迴歸係數參數個數；如果不存在該隨機效果下的固定效果的檢定，檢定統計量仍為 t 統計量，但自由度為第一層樣本數總和減去所有固定效果迴歸係數參數總數。而第二層隨機效果變異數是否為 0 的檢定，檢定統計量為卡方統計量，自由度為第二層樣本數減去該方程式的迴歸係數參數個數 (註[3])。

有關於 HLM 模型是否設計為固定效果或是隨機係數模型，Snijders 與 Bosker (1999) 提出一個建議法則，也就是依據第二層的組數、每組內的樣本數、第一層與第

註[3] 感謝審查委員的提出指正。

二層殘差項的分布、第二層各組如何被抽出的假設、研究結果推論的一般化與我們分析的關注焦點為何等來進行判斷。因此他們的建議如下：(1) 當第二層每一組的樣本有特殊的特徵，而我們欲在每一組建構同樣模式時，建議採用固定效果模型；(2) 相反地，當樣本來自母群抽樣的結果，而我們欲推論回母群特徵時，建議採用隨機係數模型；(3) 如果是關心第二層解釋變項每一組樣本數都過小時，建議採用隨機係數模型，因為參數的估計量必須藉助 (Borrowing Strength) 所有資料的資訊，利用實證貝氏估計量的縮動 (Shrinkage) 特性，否則採用固定效果其參數估計將不穩定。當然，隨機係數模型是假設在第二層模式的誤差項必須服從常態分配的條件下進行，Snijders 與 Bosker (1999) 曾提出一個經驗法則，這個法則是取決於第二層 J 的數目，如果 J 太少則使用固定效果的設定；如果 J 不會太少且每一組內的樣本數是多時，建議採用隨機效果模式。Longford (1993) 則建議當第二層組數少時，採固定效果模式，當組數多時則採隨機效果模式。

以上的原則只是一種建議法則，並不是最後決定因素，其固定效果與隨機效果的選擇還是端賴其研究的理論基礎，不過唯一可以確定的是，在 MLM 下至少必須採用隨機截距模型，意即第一層迴歸模式的截距項 β_{0j}，在第二層迴歸模式時必須帶有 u_{0j}，方能表達組內相關的意涵。如果要還要捕捉誤差異質性時，可以在斜率方程式中加入誤差項；當要探討第二層解釋變項與第一層解釋變項對結果變項的跨層級交互作用時，則必須將誤差項加入到斜率項方程式中。有關隨機效果誤差項變異數顯著的解釋，一般而言就是個別差異的意思，在 MLM 中則是指組織之間在控制解釋變項之後仍存在組織間的差異。

(三) 估計法問題：偏與不偏估計？FML 與 REML 的差異重要嗎？

在多層次研究中，主要用來分析多層次資料的專業軟體有 HLM 與 MLwiN 兩種軟體，在其參數估計方面，包含了固定效果與隨機效果的參數估計，而主要估計法，在 HLM 軟體有 FML 與 REML 法的選擇。所謂的 FML 是「完全最大概似估計法 (Full Maximal Likelihood)」，而 REML 是「受限最大概似估計法 (Restricted Maximal Likelihood)」。之所以有這兩種估計法的產生，來自於多層次迴歸分析中，牽涉到有固定效果的迴歸係數與隨機效果的第一層誤差項和第二層各個誤差項變異數、與之間的共變數要估計。在一般迴歸分析中，所要估計的參數除了迴歸係數外，只有一個誤差項的變異數。而最大概似法在估計一般迴歸分析的變異數時，是用樣本數作為自由度來計算殘差變異數：

$$\hat{\sigma}^2_{ML} = \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{n} \tag{19}$$

因此最大概似估計法所估計出來的不是母體的不偏估計值，誤差變異數會有低估的現象，不偏的母體誤差變異數的估計值為：

$$\hat{\sigma}^2_{REML} = \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{n-p-1} \tag{20}$$

(20) 式中分母的 p 為解釋變項個數，換言之最大概似法並未對要估計的迴歸係數所用掉的自由度進行調整，而 REML 則是考慮了估計 $\hat{Y}_i$ 所要估計的迴歸係數參數個數。

在 MLM 架構下，估計誤差項變異數與誤差項間的共變數就更複雜。REML 在估計迴歸係數與變異數成分時，相當於先估計固定效果迴歸係數，再求觀察值與固定效果估計預測值之間的整體殘差項，然後再從整體的殘差項去估計各層誤差項的變異數與之間的共變數，因此 REML 的 R 亦可以稱為殘差。一般而言，REML 所估計的隨機效果的變異數成分會是較 FML來得大，在 Bryk 與 Raudenbush (1992) 的軟體中，REML有時表現比 FML 來得好，尤其是在組數較少時。Goldstein (2003) 也討論這兩種方法的差異，但也沒有清楚指出哪種方法應該使用在哪種情況。從這些資訊中，我們得知所有的估計方法，對於變異成分估計最重要的條件是要有足夠數量的組數。

如果根據統計不偏估計的原則，則是較傾向使用 REML，但是在大樣本的情況下，當第二層組數是很大時，這兩種估計法在估計變異數成分上的差異應該很小。由於這兩種估計法在估計過程中，REML 是考慮到隨機效果的設定，在計算迴歸模式的參數個數與 FML 不同，因此也會影響到固定效果與隨機效果設定下模式的適配度比較。因此在探討不同估計法的同時，也要考慮到模式的設定以及在競爭模式之間的比較問題。

五、模式評估與估算

有關於多層次分析的模式評估與估算，本節以兩個部份來介紹。第一部分是有關競爭模式的比較，模式適配、變異數解釋量與多元共線性三方面來檢視模式的優劣；第二部分則是探討特殊的 HLM 議題，是關於一些特殊參數的意義與估算，用來增加對 HLM 的瞭解。

(一) 模式評估

1. 模式適配 (Goodness-of-fit) 如何使用？模式都可以互相比較？

在 HLM 的適配是以離異數 (Deviance) 來表示，而這個離異數是負兩倍的對數概似值 (Log Likelihood)，以符號表示為 -2LL。事實上，離異數是用來衡量模型與觀察資料間的不吻合程度，因此稱為「離異數」。所以當離異數越小，代表模型的適配是較佳的。但是，由於單一模式的離異數是個絕對數值，無法說明好壞，唯有透過模式

之間的比較，方能用來比較離異數的差異。

由於離異數用到概似值，亦即將所假設的參數估計值代入概似函數，因此離異數與概似函數有關，而概似函數又與結果變項或誤差項的分配有關。當兩個模型是屬於巢套 (Nested) 關係時，就可以計算兩個模型的離異數以及所估計的參數個數，透過兩個模型離異數的差服從自由度為兩個模型參數個數差的卡方分配，檢定離異數差異是否顯著以判斷模型的優劣。若以符號來表示，假設模型 A 包含於模型 B，Δ代表離異數的差異，p 代表模式的參數個數，則

$$\Delta = -2LL_A - (2LL_B) \sim \chi^2_{p_B - p_A} \tag{21}$$

若 Δ 顯著，則代表模式 B 優於模式 A。

由於離異數的計算牽涉到參數的估計法，因此有關於模式的比較與模式的設定、以及所用的估計法有關。當我們研究所要比較的是固定效果的不同(牽涉到不同的解釋變項)，在巢套下是以 FML 來進行，此時所估計出來的離異數，其所計算的參數個數是包含固定效果與隨機效果的個數。當我們要比較的是在相同的固定效果下不同的隨機效果設定時，同樣在巢套下是以 REML 來進行，此時離異數所計算的參數個數只包含隨機效果的個數。

在模式的比較上，是比較其適配度的好壞，但前提條件是兩個模式是巢套情況下方能以離異數來比較。換言之，當要比較的模型是不具巢套關係時，離異數的差異檢定則無法派上用場。此時，必須利用不同的適配度指標，一般是採用 AIC (Akaike Information Criterion) 或 BIC (Bayesian Information Criterion)，其判斷準則為選擇有較小 AIC 或 BIC 的模式。AIC 的公式為：

$$AIC = -2LL + 2p \tag{22}$$

這裡的為模式所需估計的參數個數，此指標與模式的精簡程度 (Parsimonious) 有關，當模式越精簡則所計算出來的 AIC 越小；而 BIC 的公式為：

$$BIC = -2LL + p\ln(N^*) \tag{23}$$

上述 BIC 的公式中，除了與模式所需估計的參數 p 個數有關外，尚與樣本數有關，這個 N^* 是指有效樣本數，但在多層次架構下這個 N^* 並不是很清楚。Goldstein (2003) 認為在 MLM 下，這個樣本數是第二層的組數較為恰當。

2. 效果量：解釋變異量 (Explained Variance) 問題，R^2？

在迴歸分析中，最常用來說明解釋變數的貢獻，是利用結果變項的誤差變異程度來定義。這個誤差變異程度是計算結果變項實際觀測值與預測值的差異，再求平方和。因此先計算在沒有任何解釋變數下的離均差平方和，再與引進解釋變項後的預測誤差平方和的差異，計算一個稱為「解釋平方 (R^2)」，亦即計算迴歸變異佔總變異的百分比來說明解釋變項可以解釋結果變項的變異程度。由於這個計算公式是解釋變項個數的非遞減函數，只要不斷增加解釋變項，不管其對結果變項有解釋力否，其計算出來的 R^2 不會減少。為了校正因為一直增加解釋變項而使得自由度持續減少，因此設計了「R_a^2」，亦即將自由度考慮到原先的公式上：

$$R_a^2 = 1 - \frac{\sum_i^n \left(Y_i - \hat{Y}_i\right)^2 / (n-p-1)}{\sum_i^n \left(Y_i - \bar{Y}\right)^2 / (n-1)} \tag{24}$$

當不斷增加解釋變項，則第二項分子的 $(n-p-1)$ 越來越小，當殘差變異減少幅度不及 $(n-p-1)$ 變小的程度，可能讓整個分子變大，而在分母不變情況下，所計算出來的 R_a^2 可能會變小，不再是解釋變項個數的非遞減函數。透過這樣的調整，可以在解釋變項的挑選與解釋變異的變化上取得調和的程度。事實上，R_a^2 可以再經整理成為我們所熟悉的變異數形式：

$$R_a^2 = 1 - \frac{\sum_i^n \left(Y_i - \hat{Y}_i\right)^2 / (n-p-1)}{\sum_i^n \left(Y_i - \bar{Y}\right)^2 / (n-1)} = 1 - \frac{S_e^2}{S^2} = \frac{S^2 - S_e^2}{S^2} \tag{25}$$

從最後一個等式可以看出，R_a^2 可以視為變異數的改善幅度。當沒有任何解釋變數時，迴歸模式就是一條水平線，代表平均數的意思，因此 S^2 是原始資料的變異數，而 S_e^2 是在迴歸模式下的殘差變異數。根據同樣的概念，R_a^2 可以應用在 MLM 上稱為「pseudo R^2」。因為 MLM 計算誤差是包含多個層次的誤差項和，無法計算單一的變異指標，因此無法計算出 R^2。以零模型為例：

$$\text{Level 1: } Y_{ij} = \beta_{0j} + \varepsilon_{ij} \tag{26}$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + u_{0j} \tag{27}$$

第一層與第二層誤差項變異數的估計值分別為 $\hat{\sigma}^2_{Null}$ 與 $\hat{\tau}_{00Null}$，當引進第一層與第二層解釋變項 X_{ij} 與 Z_j 後，將模式設定如下，稱為「Intercept as Outcome」模型：

$$\text{Level 1: } Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + \varepsilon_{ij} \tag{28}$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + u_{0j} \tag{29}$$

$$\beta_{1j} = \gamma_{10} \tag{30}$$

此時第一層誤差項變異數的估計值為 $\hat{\sigma}^2_{Inter}$，第二層則為 $\hat{\tau}_{00Inter}$。我們可以計算在第一層迴歸模式中引進 X_{ij} 後 $\hat{\sigma}^2$ 的減少程度，以及在第二層模式中引進 Z_j 後 $\hat{\tau}_{00}$ 的減少程度：

$$R_1^2 = 1 - \frac{\hat{\sigma}^2_{Inter}}{\hat{\sigma}^2_{Null}} = \frac{\hat{\sigma}^2_{Null} - \hat{\sigma}^2_{Inter}}{\hat{\sigma}^2_{Null}} \tag{31}$$

$$R_2^2 = 1 - \frac{\hat{\tau}_{00Inter}}{\hat{\tau}_{00Null}} = \frac{\hat{\tau}_{00Null} - \hat{\tau}_{00Inter}}{\hat{\tau}_{00Null}} \tag{32}$$

但這種削減百分比的形式或是 R_a^2 有個缺點，當在進行模式比較時，可能會得到負的解釋力，這是不合理的現象，但是在 HLM 中會常發生。特別是在第二層加入解釋變數後，其 $\hat{\tau}$ 不降反升。一般而言，增加第一層的解釋變項會使得 $\hat{\sigma}^2$ 減少，增加第二層的解釋變項會使得 $\hat{\tau}$ 變小。為了克服因增加解釋變項而使 $\hat{\tau}$ 增加的這種違反直覺現象，Snijders 與 Bosker (1999) 利用預測誤差的概念，整合所有不同層次解釋變項對結果變項的解釋力，以整體的變異數定義個體層次的 R_1^2 與總體層次的 R_2^2。根據上述模型，R_1^2 與 R_2^2 的定義為：

$$R_1^2 = 1 - \frac{\hat{\sigma}^2_{Inter} + \hat{\tau}_{00Inter}}{\hat{\sigma}^2_{Null} + \hat{\tau}_{00Null}} \tag{33}$$

$$R_2^2 = 1 - \frac{\dfrac{\hat{\sigma}^2_{Inter}}{\tilde{n}} + \hat{\tau}_{00Inter}}{\dfrac{\hat{\sigma}^2_{Null}}{\tilde{n}} + \hat{\tau}_{00Inter}} \tag{34}$$

R_1^2 第二項的分子是「Intercept as Outcome」個體層次平均預測誤差，亦即 $Var(Y_{ij}^{Inter})$，而分母為零模型下個體層次平均預測誤差為 $Var(Y_{ij}^{Null})$。而 R_2^2 第二項的分子是「Intercept as Outcome」總體層次的平均預測誤差，亦即 $Var(\bar{Y}_j^{Inter})$，而分母為零模型下總體層次的平均預測誤差為 $Var(\bar{Y}_j^{Null})$，這個總體層次預測誤差變異數即為平均數的變異數，而 $\tilde{n}$ 為平均的組內樣本數。這個定義有兩個優點，第一是將個體層次與總體層次解釋變項一同考慮進來，雖然不同層次的解釋變項對不同的誤差項產生影響，但是以整體的形式來看解釋變項的解釋力；第二個是 R_1^2 與 R_2^2 不會是負值，如果一旦產

生負值的 R_1^2 或 R_2^2 ，即代表模式設定有問題。

3. 多元共線性 (Multicollinearity) 在多層次分析當中有多麼嚴重？

在 GLM 中，多元共線性一直是非常重要的議題，而在 HLM 中，多元共線性也有一些非常值得注意之處。在一個完整的多層次模型中會有三種形式的預測變項：第一層解釋變數、第二層解釋變數，以及跨層級交互作用變項，如方程式 (4)。因為方程式 (4) 等號右邊第四項是第二項與第三項變數的乘積，因此這三種預測變項間可能存在相關，如同單一層次的迴歸分析，當存在交互作用項時，容易造成多元共線性問題。在單一層次迴歸分析中，為克服交互作用項造成共線性問題，一般會先將解釋變項進行平減，以減少共線性問題。(Aiken & West, 1991; Tabachnick & Fidell, 2006)

在多層次模式中，一旦存在多個個體層次解釋變項與總體層次解釋變項時，在完整模式下關係就更加複雜，更增添多元共線性的機會。如同一般迴歸分析，檢測解釋變項間是否有嚴重的多元共線性問題，可以採取變異數膨脹因子 (Variance Inflation Factor；VIF) 與條件數 (Condition Number)，因為 HLM 在估計參數時是採用混合模式，是將第二層變項解構為個體層次進行分析。

在 HLM 中，主要平減方法有兩種：總平減與組平減。總平減是透過每個變項都減去其總平均數來減少共線性 (Aiken & West, 1991)，而組平減則是每個觀察分數減去所屬的組平均數，將解釋變項拆解為組間與組內兩個部分的和：

$$X_{ij} = \bar{X}_j + (X_{ij} - \bar{X}_j) \tag{35}$$

上式中的組平均數 $\bar{X}_j$ 與離均差 $(X_{ij} - \bar{X}_j)$ 彼此獨立，這就是變異數分析中的正交拆解。如果第一層解釋變數都經過了組平減，則此一變數與任何一個第二層解釋變數的跨層級交互作用之間都是獨立的，因此透過組平減可以減少共線性。但是透過組平減來減低共線性問題，同時也引發另一個問題，就是整個模式不再和以原始單位分析的模式恆等 (Equivalence)，一來截距項的解釋變為較有意義外，其次是透過組平減會消除組間的差異。為了能讓模式能夠維持組間差異，必須將組平均數放回第二層的迴歸模式以還原原先模式的意涵。

(二) 模式估算

1. 強韌性標準誤的估計

由於 HLM 在估計參數時是採用 ML 法，而 ML 一般而言是較強韌性，亦即當資料的分配與假設是不吻合，但不吻合程度不大，ML 所估計的參數仍然具有一致性。但是，這些參數估計值的抽樣標準誤卻沒有這樣的特性，使得所計算的檢定統計量以及假設檢定的結果會與所假設分配的離異程度有關。

在迴歸分析中，最常被違反的假設是常態性與同質性，而獨立性因資料是來自

橫斷面抽樣結果比較沒有問題。而常態性被違反常伴隨變異數異質性，因此常態分配的假設在迴歸分析中也是相當重要。而 HLM 亦屬於迴歸分析，所以常態分配也是 HLM 很重要的假設，特別是第二層的誤差項。在 GLM，針對誤差項的同質性與獨立性被違反，可以一般化最小平方法 (GLS) 來估計迴歸係數，因此所估計出來的迴歸係數如下式：

$$\hat{\beta}_{GLS} = \left(X^T V^{-1} X\right)^{-1} X^T V^{-1} Y \tag{36}$$

上式中的 V 為 Y 的條件變異數，而 $\hat{\beta}_{GLS}$ 的抽樣變異誤為：

$$Var\left(\hat{\beta}_{GLS}\right) = \left(X^T V^{-1} X\right)^{-1} X^T V^{-1} Var(Y) V^{-1} X \left(X^T V^{-1} X\right)^{-1} \tag{37}$$

當 V 以估計值 $\hat{V}$ 代入，則變異誤化簡為：

$$V\hat{a}r\left(\hat{\beta}_{GLS}\right) = \left(X^T \hat{V}^{-1} X\right)^{-1} \tag{38}$$

若 Y 或誤差項違反常態分配，則以 $\hat{V}$ 代入所估計的變異誤則不具一致性，換言之即使樣本夠大，其估計值亦未必是母體的不偏變異數。所以在 Var (Y) 的計算上不用估計值 $\hat{V}$，而改用實際資料所計算出來的殘差 ($\tilde{Y}$) 來估計，亦即：

$$Var(Y) = E\left(\tilde{Y}\tilde{Y}^T\right) \tag{39}$$

並將 (39) 式代回 (37) 式：

$$Var\left(\hat{\beta}_{GLS}\right) = \left(X^T V^{-1} X\right)^{-1} X^T V^{-1} E\left(\tilde{Y}\tilde{Y}^T\right) V^{-1} X \left(X^T V^{-1} X\right)^{-1} \tag{40}$$

此時，所計算出來迴歸係數估計值的標準誤稱為強韌標準誤 (Robust Standard Error)，這個標準誤就會與當初假設的誤差項分配關係較小，而達到強韌性功用。仔細觀察 (40) 式，這個實際殘差所計算的部分是在中間，因此也被稱為「三明治估計值 (Sandwich Estimate) (Huber, 1967)」。

而在 HLM 中，因為用來估計迴歸係數是利用混合模式，基本精神與 GLM 一樣，但牽涉到誤差項或殘差則包含了各個層次，因此所計算的公式相形之下更複雜。至於什麼時機該選擇強韌性標準誤來檢定迴歸係數的顯著性，文獻上並沒有很明確的指出，因為迴歸分析的誤差項是否違反常態性，必須利用估計之後的殘差來檢視。至

於稱為強韌性理由應該隨時都可以使用，因為不管有無違反，它的公式都是使用實際估計的殘差計算，比較貼近實際的資料。不管如何，當所估計的一般標準誤與強韌性標準誤有明顯的差異時，則說明了資料與模式所假設的分配是不吻合的，這個不吻合包含模式的設定錯誤 (Hox, 2002)，研究者必須小心解釋分析結果，最好的方式是重新檢視資料，去發掘可能的問題，再透過適當的模式設定進行分析。換言之，這個強韌性標準誤和一般標準誤的差異可以視為一種模式錯誤設定的警告訊號。

2. 第一層誤差項變異數異質性的估計

HLM 中透過第二層的隨機效果設定，可以補捉在巢套資料結構下誤差項非獨立性與異質性問題。在 HLM 軟體中可以針對多層次模式下的第一層誤差項的同質性進行檢定，對模式誤差項假設違反可以提供更進一步的資訊。除了可以利用同質性檢定來檢視第一層誤差項，亦可以利用圖示將第一層殘差對各解釋變項做圖來判斷。一旦發現第一層殘差的同質性檢定顯著，或是殘差圖有系統性的類型時，我們可以進一步來配適第一層誤差項的變異數方程式，來解釋第一層誤差項的可能影響因素。

傳統上對於誤差項變異數同質性假設一旦被違反，過去的修正做法都是對結果變項進行變數變換，透過檢視殘差圖來選擇適當的轉換函數，例如取對數或開根號等，再重新配適新的迴歸方程式。如果新的迴歸方程式殘差項符合假設，則接受新的迴歸方程式，但此做法變項間的關係可能因變數變換而改變意義。在財務計量中的 GARCH 模式，其假設誤差項的變異數可以不具同質性，可以透過變異數方程式加以估計。HLM 軟體亦提供第一層誤差項變異數的估計，如下所示：

$$\ln\left(\sigma_{ij}^{2}\right)=\alpha_{0}+\alpha_{1}X_{1ij}+\alpha_{2}X_{2ij} \tag{41}$$

方程式 (41) 中，假設第一層誤差項的變異數存在異質性，且可能與第一層個體解釋變項 X_{1ij} 與 X_{2ij} 有關，則可以利用迴歸來檢視之間的關係。上式等號的左邊是對變異數取對數，其原因在於變異數不可能為負，因此透過對數的轉換後可以是正也可以是負，因此在方程式的右側參數估計就可以為負。HLM 軟體可以透過檢視方程式 (41) 在HLM 模式下，和第一層誤差項變異數同質性的假設進行卡方檢定，進行異質性與同質性的適配度比較。

3. 貝式估計與實證貝式估計

在HLM中，我們是透過個體層次結果變項對個體層次解釋變項的迴歸，得到第一層迴歸係數後，再將迴歸係數對第二層解釋變項進行迴歸建構多層次迴歸模式，如方程式 (1)、(2) 與 (3) 所示。如果將第二層迴歸模式代回第一層迴歸方程式，第一層迴歸係數則消失，如方程式 (4) 所示，這時的 β 係數如同潛在變項一樣是觀察不到的。如果我們研究關心的是 β 係數，則如何估計這些係數呢？為了簡化說明，我們仍

以零模型來示範，其模型設定如方程式 (5) 與 (6) 所示，其混合模式為：

$$Y_{ij} = \gamma_{00} + u_{0j} + \varepsilon_{ij} \tag{42}$$

由於 γ_{00} 是個常數，則 β_{0j} 分配的變異數與 u_{0j} 一樣，其分配為：

$$\beta_{0j} \overset{iid}{\sim} N(\gamma_{00}, \tau_{00}) \tag{43}$$

而第一層為：

$$Y_{ij} \overset{iid}{\sim} N(\beta_{0j}, \sigma^2) \tag{44}$$

要估計 β_{0j} 相當於貝氏估計，而 N (γ_{00} , τ_{00}) 相當於事前分配 (Prior Distribution)。由於 γ_{00} 已由混合模式估計而得，所以 N ($\hat{\gamma}_{00}$, $\hat{\tau}_{00}$) 為已知，因此稱 β_{0j} 的貝氏估計值為實證貝氏估計值，因為事前的參數不是給定，而是透過資料估計而得。在 $\hat{\gamma}_{00}$, $\hat{\tau}_{00}$ 與 $\hat{\sigma}^2$ 已知條件下， β_{0j} 的事後分配期望值或實證貝氏估計值為：

$$\hat{\beta}_{0j}^{EB} = \frac{\hat{\tau}_{00}}{\hat{\tau}_{00} + \hat{\sigma}^2 / n_j} \bar{Y}_j + \frac{\hat{\sigma}^2 / n_j}{\hat{\tau}_{00} + \hat{\sigma}^2 / n_j} \hat{\gamma}_{00} \tag{45}$$

因為 $Y_{ij} = \beta_{0j} + \varepsilon_{ij}$，而 $\varepsilon_{ij} \overset{iid}{\sim} \mathrm{N}(0, \sigma^2)$， β_{0j} 的最佳不偏估計值則為 $\bar{Y}_j$，其 $\bar{Y}_j$ 分配的變異數為 σ^2/n_j；此外 $\beta_{0j} = \gamma_{00} + u_{0j}$，所以 β_{0j} 的估計值為 $\hat{\gamma}_{00}$，其分配變異數為 τ_{00}，在估計 σ^2 與 τ_{00} 之後，就可以利用這兩個資訊來估計 β_{0j}^{EB}。而上式方程式等號右邊第一個分式即為信度的估計值，而 $\bar{Y}_j$ 又是第 j 樣本下 β_{0j} 的最小平方估計值，所以 β_{0j}^{EB} 又可以表示成：

$$\hat{\beta}_{0j}^{EB} = \hat{\lambda}_j \hat{\beta}_{0j}^{OLS} + (1 - \hat{\lambda}_j)\hat{\gamma}_{00} \tag{46}$$

從上式可以發現， β_{0j} 的事後分配期望值是 $\hat{\beta}_{0j}^{OLS}$ 與 $\hat{\gamma}_{00}$ 的加權平均數，而 $\hat{\beta}_{0j}^{OLS}$ 是第 j 樣本的平均數，其資訊的提供只有來自第 j 群樣本；而 $\hat{\gamma}_{00}$ 則為總樣本的加權平均數，代表資訊是來自所有的資料。而 $\hat{\lambda}_j$ 所扮演的角色就是權重，因此 $\hat{\beta}_{0j}^{BE}$ 又被稱為「縮收變動估計值 (Shrinkage Estimate)」。當第 j 群的樣本數 n_j 很大時，則 $\hat{\lambda}_j$ 接近於 1，意即 $\hat{\beta}_{0j}^{OLS}$ 可信度高，因此 $\hat{\beta}_{0j}^{EB}$ 往 $\hat{\beta}_{0j}^{OLS}$ 移動。當第 j 群的樣本數 n_j 很小而 $\hat{\sigma}^2$ 又很大，則 $\hat{\lambda}_j$ 接近於 0，意即 $\hat{\beta}_{0j}^{OLS}$ 的資訊不可靠，其信度低，因此 $\hat{\beta}_{0j}^{EB}$ 往 $\hat{\gamma}_{00}$ 靠。換言

之，當第 j 群樣本少時，其資訊的準確性低，則β_{0j} 的估計值是透過所有樣本資料的資訊來填補 (Filling in) 自己資訊的不足，因此有向全體樣本借力 (Borrowing Strength) 之意 (Kreft & de Leeuw, 1998)。此外，$\hat{\beta}_{0j}^{EB}$ 不是β_{0j} 的不偏估計值，但卻有最小的平均估計誤差(MSE)。所以，HLM 在各組的樣本數可以不一樣多，在考慮信度的情況下，當然是組內樣本數越多越好。貝氏估計的使用時機在於想要知道各組第一層迴歸方程式為何，以及瞭解各組之間迴歸模式的比較。

肆、結論

從第參部分的探討，可以發現在 MLM 中，這些議題都是相互關聯，是 GLM 中較難發現的獨特地方。換言之，HLM 的正確使用必須要有整體的考量，每個細節必須能夠有所掌握之外，更需瞭解彼此間的關係。以兩層結構的多層次樣本數為例，由於牽涉到第二層組數與第一層組內樣本數的權衡，在設計效果中，當平均組內樣本數越多時，即使 ICC 很小，設計效果也可能會很大，造成有效樣本數的減少，以及檢定力的減弱，因此不能有太多的組內樣本數。但同時檢視信度公式，當組內樣本數越大，該組的信度就越高，在估計第一層的迴歸係數效果就更加準確。從這兩個第一層組內樣本數的結果來看剛好互相矛盾，但事實不然，因為他們是探討不同的課題，因此在不同的前提條件，所得到的結論也就不同。茲將上述所探討十二個議題的涵義與建議整理成表二所示：

表 2 十二個關鍵議題的涵義建議

範疇	關鍵議題	建議
一、使用時機	ICC	1. 大於 0.059 2. F 或卡方檢定顯著
二、抽樣議題	樣本數	1. 30/30 原則 2. 第二層樣本數越多越好
三、資料彙總	組織構念	1. 遷移共識法設計問卷 2. r_{wg} 與 ICC (2) 大於 0.7
四、分析方法	總平減或組平減	1. 第二層解釋變項總平減 2. 第一層解釋變項主效果組平減，但將組平均數置於截距方程式。 3. 跨層級交互作用項，第一層以組平減 4. 脈絡效果與中介效果以總平減
	固定效果或隨機效果	1. 第二層樣本數太少則採固定效果 2. 涉及抽樣推論則採隨機效果 3. 固定效果的解釋是平均迴歸係數的概念 4. 變異成分的變異數是個別差異的概念
	FML 或 REML	1. 關心的是迴歸係數採 FML 2. 關心的是變異成分採 REML 3. 牽涉所有參數估計的模式比較採 FML
五、模式評估	模式配適	1. 巢套下採離異數差異的卡方檢定，固定效果的比較採 FML 估計法、隨機效果的比較採 REML 估計法。 2. 非巢套模式比較採 AIC 或 BIC
	解釋變異量	1. pseudo R squares 2. 若上式為負採 Snijders 與 Bosker (1999) 的公式，若為負則模式設定有誤。
	共線性	1. 以混合模式資料結構進行 VIF 或條件數檢測 2. 以平減處理
	強韌性標準誤	1. 隨時可使用，但與一般標準誤有極大差異時，顯示模式的假設被違反。
	第一層變異數方程式	1. 第一層誤差項同質性被違反，可以用來估計其變異數方程式。
	貝式估計法	1. 估計第一層方程式的迴歸係數 2. 用來比較組間方程式的差異

在實際研究中，這彼此關連的十二個重要議題可以從測量 (資料的特性與品質)、統計 (分析的方法與策略)、方法論 (研究問題的解決與研究發現的解釋應用) 等三方面來進行有意義的區分，本文基於議題的性質區分為五個構面，但關於 HLM 還有其他有待深入的部份，例如本文只侷限對兩層鑲嵌結構資料的連續結果變項進行整理，至於三層結構或是其他類型資料更值得去探討。在這裡，我們只針對在 MLM 的研究中，所經常會遭遇到的技術與方法論上的問題進行整理，絕大多數是作者使用 HLM 所遭遇問題而整理的解決方法，至於細節部份，可能必須回到文獻的源頭去詳讀。在這相關的十二個議題上，作者有個心得可以以這句話來表達：「如果研究者忽視脈絡效果的分析，錯以個體層次的效果去解釋總體層次的效果 (Atomic Fallacy)，或是錯以總體效果去解釋個體效果 (Ecological Fallacy)，在方法學上都是屬於推論上的謬誤，由此可知多層次資料的探討，方法學上的考量更勝於統計的操作與分析。」

誠如許多文獻所述，HLM 作為一個近代崛起的新興統計技術，必須兼顧方法學與技術層次的考量。但是從 HLM 可以完成的分析任務與彈性，以及 2000 年之後實證論文累積的速度，HLM 確實是當代最具發展潛力的統計分析技術，在國外如此，在國內亦然。在可預見的數年內，國內將會有越來越多有關 HLM 的深入研究與廣泛應用。因此，本文如同林鉦棽與彭台光 (2006) 兩位學者的結論建議，期許本文能對國內 HLM 或 MLM 提供一個較為明確的應用原則。

參考文獻

林鉦棽、彭台光，2006，「多層次管理研究：分析層次的概念、理論和方法」，管理學報，23 卷6 期：頁649-675。

邱皓政，2007，「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實證應用」，中華心理學刊，49 卷4 期：頁383-405。

黃芳銘、溫福星，2007，「學習型學校量表之發展：多層次驗證性因素分析取向」，測驗學刊，54 卷1 期：頁197-222。

溫福星，2006，初版，階層線性模式：原理、方法與應用，台北：雙葉書廊圖書公司。

Aiken, L. S., & West, S. G. 1991. ***Multiple regression: Testing and interpreting interaction***. Newbury Park, CA: Sage.

Anderson, C. 2004. ***Multilevel analysis/hierarchical linear modeling***. http://www.ed.uiuc.edu/courses/EdPsy490CK/. Accessed Sep. 28, 2007.

Bassiri, D. 1988. ***Large and small sample properties of maximum likelihood estimates for the hierarchical linear model***. Unpublished doctoral dissertation, Department of Counseling, Educational Psychology and Special Education, Michigan State University.

Bickel, R. 2007. ***Multilevel analysis for applied research: It's just regression***. New York, NY: Guilford Press.

Bliese, P. D. 2000. Within-group agreement, nonindependence, and reliability: Implications for data aggregation and analysis. In K. J. Klein, & S. W. J. Kozlowski (Eds.), ***Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions***: 349-381. San Francisco, CA: Jossey-Bass.

Bliese, P. D., Chan, D., & Ployhart, R. E. 2007. Multilevel methods: Future direction in measurement, longitudinal analyses, and nonnormal outcomes. ***Organizational Research Methods***, 10 (4): 551-563.

Bryk, A. S., & Raudenbush, S. W. 1992. ***Hierarchical linear models: Applications and data analysis methods***. Newbury Park, CA: Sage Publications.

Chan, D. 1998. Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. ***Journal of Applied Psychology***, 83 (2): 234-246.

Chen, G., Mathieu, J. E., & Bliese, P. D. 2004. A framework for conducting multilevel construct validation. In F. J. Dansereau, & F. J. Yammarino (Eds.), ***Research in multi-level issues: The many faces of multi-level issues***: 273-303. Oxford, UK:

Elsevier Science.

Cohen, J. 1988. ***Statistical power analysis for the behavioral sciences*** (2nd ed.). Hillsdale, NJ: Eribaum.

Courgeau, D. 2003. ***Methodology and epistemology of multilevel analysis: Approaches from different social sciences***. Norwell, MA: Kluwer.

de Leeuw, J., & Kreft, I. G. G. 1995. Questioning multilevel models. ***Journal of Educational and Behavioral Statistics***, 20 (2): 171-189.

Donner, A. 1986. A review of inference procedures for the intraclass correlation coefficient in the one-way random effects model. ***International Statistical Review***, 54 (1): 67-82.

Duncan, O., Curzort, R., & Duncan, R. 1966. ***Statistical geography: Problems in analyzing a real data***. Glencoe, IL: Free Press.

Enders, C. K., & Tofighi, D. 2007. Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. ***Psychological Methods***, 12 (2): 121-138.

Goldstein, H. 2003. ***Multilevel statistical models*** (3rd ed.). London, UK: Edward Arnold.

Heck, R. H., & Thomas, S. L. 2000. ***An introduction to multilevel modeling techniques***. Mahwah, NJ: Lawrence Erlbaum Associates.

Hofmann, D. A. 1997. An overview of the logic and rational of hierarchical linear models. ***Journal of Management***, 23 (6): 723-744.

Hofmann, D. A., & Gavin, M. B. 1998. Centering decisions in hierarchical linear models: Implications for research in organizations. ***Journal of Management***, 24 (5): 623-641.

Hox, J. J. 2002. ***Multilevel analysis: Techniques and applications***. Mahwah, NJ: Erlbaum.

Huber, P. J. 1967. ***The behavior of maximum likelihood estimates under non-standard condition***. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1: 221-233. Berkeley, CA: University of California Press.

James, L. R. 1982. Aggregation bias in estimates of perceptual agreement. ***Journal of Applied Psychology***, 67 (2): 219-229.

Kish, L. 1995. ***Surveying sampling***. New York, NY: Wiley.

Kozlowski, S. W. J., & Klein, K. J. 2000. A multi-level approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. Klein, & S. W. J. Kozlowski (Eds.), ***Multilevel theory, research, and methods in organizations***:

3-90. San Francisco, CA: Jossey-Bass.

Kreft, I. G. G. 1996. Are multilevel techniques necessary? An overview, including simulation studies. Manuscript, California State University at Los Angeles.

Kreft, I. G. G., & de Leeuw, J. 1998. ***Introducing multilevel modeling***. London, UK: Sage.

Kreft, I. G. G., de Leeuw, J., & Aiken, L. S. 1995. The effect of different forms of centering in hierarchical linear models. ***Multivariate Behavioral Research***, 30 (1): 1-21.

Lazarsfeld, P. F., & Menzel, H. 1970. On the relation between individual and collective properties. In A. Etzioni (Ed.), ***A sociological reader on complex organizations***: 499-516. New York, NY: Holt, Rinehart and Winston.

Longford, N. T. 1993. ***Random coefficient models***. Oxford, UK: Oxford University.

Luke, D. 2004. ***Multilevel modeling***. Newbury Park, CA: Sage.

Maas, C. J. M., & Hox, J. J. 2005. Sufficient sample sizes for multilevel modeling. ***Methodology***, 1 (3): 86-92.

Mathieu, J. E., & Taylor, S. R. 2007. A framework for testing meso-mediational relationships in organizational behavior. ***Journal of Organization Behavior***, 28 (2): 141-172.

Mok, M. 1995. Sample size requirements for a 2-level designs in educational research. ***Multilevel Modelling Newsletter***, 7 (2): 11-15.

Muthen, B. 1994. Multilevel covariance structure analysis. ***Sociological Methods & Research***, 22 (3): 376-398.

Pedhazur, E. J. 1997. ***Multiple regression in behavior research: Explanation and prediction***. Forth Worth, TA: Harcourt.

Raudenbush, S. W., & Bryk, A. S. 2002. ***Hierarchical linear models: Applications and data analysis methods*** (2nd ed.). Newbury Park, CA: Sage.

Robinson, W. S. 1950. Ecological correlations and the behaviour of individuals. ***American Sociological Review***, 15 (3): 351-357.

Snijders, T. A. B., & Bosker, R. J. 1994. Modeled variance in two-level models. ***Sociological Methods & Research***, 22 (3): 342-363.

________. 1999. ***Multilevel analysis: An introduction to basic and advanced multilevel modeling***. Thousand Oaks, CA: Sage Publications.

Tabachnick, B. G., & Fidell, L. S. 2006. ***Using multivariate statistics*** (5th ed.). Boston, MA: Allyn and Bacon.

Van de Vijver, F. J. R., & Poortinga, Y. H. 2002. Structural equivalence in multilevel research. ***Journal of Cross-Cultural Psychology***, 33 (2): 141-156.

Van der Leeden, R., & Busing, F. M. T. A. 1994. First iteration versus IGLS/RIGLS estimates in two-level models: A monte carlo study with ML3. Manuscript, Department of Psychometrics and Research Methodology, Leiden University.

Wu, Y. W. B., & Wooldridge, P. J. 2005. The impact of centering first-level predictors on individual and contextual effects in multilevel data analysis. ***Nursing Research***, 54 (3): 212-216.

作者簡介

溫福星

國立政治大學企管博士，目前為東吳大學國際經營與貿易學系助理教授。主要研究領域為風險管理與心理計量，專長於多層次研究，研究興趣包含多層次調節中介效果、多層次潛在類別分析與追蹤資料的分析。文章主要發表於亞太管理評論、管理學報、教育與心理研究與測驗學刊等，目前擔任台灣統計方法學學會副理事長。

邱皓政

美國南加大心理計量博士，目前服務於國立中央大學企管系，擔任副教授職務。主要研究領域為創造力與組織研究，專長於統計分析與測驗編製，研究興趣包含結構方程模式、多層次結構方程模式、潛在類別分析與組織構念的形成等。文章主要發表於應用心理研究、教育與心理研究、管理學報與中華心理學刊、測驗學刊等，目前擔任中國測驗學會常務理事、台灣統計方法學學會監事等。