DINA 與 G-DINA 模式參數不變性探討

楊智為 卓淑瑜 郭伯臣 陳亭宇

國立臺中教育大學 教育測驗統計研究所

摘要

所謂的「因材施教」即是說明教師必須先瞭解每一位學生的長處與短處，才能夠設計教學方針，實施補救教學。然而，一般傳統的紙筆測驗，僅提供學生在團體中的量尺分數，並無法顯現出學生是否精熟某種概念的訊息。

為了進一步幫助學生或教師對於測驗的結果有更多的瞭解，進而施行更有效率的學習，認知診斷模型可以提供解決的方案，而其中又以 DINA 模式最簡單也最容易解釋，目前國外已有許多學者投入模式的開發與實際應用的研究。同樣地，參數不變性在認知診斷評量的測驗設計上著實具有相當重要地位，也就是可以研究如何以認知診斷模型來估計試題參數的潛在基本特徵與其分佈變化。

本研究透過模擬樣本資料的實驗設計，探討在不同樣本人數、不同認知屬性分佈之下，分別以 DINA 模式與 G-DINA 模式估計來進行探討，當資料符合 DINA 模型時，有很明顯的參數不變性，若資料型態為未知時，建議採用 G-DINA 模式來進行分析。

關鍵字：認知診斷模型、DINA 模式、G-DINA 模式、參數不變性。

Explore the invariance of DINA model and G-DINA model parameters

Chih-Wei Yang　Shu-Yu Cho　Bor-Chen Kuo　Ting-Yu Chen

Graduate Institute of Educational Measurement and Statistics, National Taichung University of Education

Abstract

"Teaching students in accordance with their aptitude" means that teachers must understand each student's strengths and weaknesses so that they can design principles of teaching and remedial education. However, the general tradition written test is only supply scale score of each student in a group, but it cannot display if the students good at some kinds of skills.

In order to help students and teachers to know about the test results for having more efficiency learning, the "cognitive diagnosis models" could provide the solutions. The DINA model is the simplest and easiest to explain among these models. Many foreign scholars are involved in the development and practical application research at the present time. Similarly, invariance of parameter is really important to the designing of cognitive diagnosis assessment, it means we can research how to use cognitive diagnosis models to estimate the potential basic features and the distributive changes of the item parameters.

This paper adopt simulated data from design of experiment to explore that to estimate the parameters with DINA model and G-DINA model respectively so that we can compare with the existence of invariant in different sample size, and cognitive attribute distributions. The results show that the DINA and G-DINA model are invariant when the DINA model fits the data and the G-DINA model estimations get better accuracies when we don’t know about the data.

Key words: cognitive diagnosis models, DINA model, G-DINA model, invariance of parameter

壹、緒論

教育測驗與評量的目的，除了能夠測量出受試者的學習表現，還要能夠同時診斷出受試者學習時的缺失，以方便教師根據這些診斷訊息進行有效的補救教學，因此，認知診斷模型（cognitive diagnosis models, CDMs）成為一個令人興奮的心理計量研究的新領域。認知診斷模型被使用來更深入地分析試題反應資料（de la Torre & Douglas, 2004; DiBello, Stout, & Roussos, 1995; Tatsuoka, 1983; von Davier, 2005），針對受試者能力診斷精熟的狀態，在這類模型中，常用離散的形式來表示潛在的能力或技能，能提供關於受試者強項及弱點等詳細的訊息。

認知診斷模型以發展出許多不同的模式，其中以 DINA 模式（Deterministic Inputs, Noisy “And” Gate Model, DINA; Junker & Sijtsma, 2001）的簡單性及容易理解，近年許多的研究皆以此模式進行探討，如測驗的組卷、參數的估計方法等。

參數不變性在測驗評量的發展上亦是一個重要的議題，代表參數估計結果的穩定性，de la Torre & Lee(2010)利用不同能力分佈的受試者來探究在 DINA 模式下，試題參數估計的結果是否具有不變性。

de la Torre(2011)提出 DINA 模式的一般化模型，稱為 G-DINA 模式（generalized deterministic inputs, noisy “and” gate），能夠針對不同的認知屬性有更彈性的估計，可以更符合一般情境下的測驗，然而 G-DINA 模型是否保有 DINA 模型下的參數不變性仍有待確認，因此，本研究將以 G-DINA 模式為基礎，以模擬樣本資料的實驗方式來進行探討，在不同樣本大小下，不同能力水準在是否對試題參數的估計造成影響。

根據上述，本研究的問題如下：

一、以 DINA 模式與 G-DINA 模式來探討不同能力分佈的受試者對試題參數不變性估計的影響。

二、探討 DINA 模式與 G-DINA 模式對試題參數估計的不變性與受試者認知屬性之診斷辨識率的影響。

貳、文獻探討

本節分為兩部分，第一部分為認知診斷模型探討及本研究相關的模式簡介，第二部份為參數不變性相關文獻。

一、認知診斷模型

認知診斷模型是一種結合認知科學與心理計量學的方法，對於受試者潛在能力狀態的分類最具特色，認知診斷模型在近年受到越來越多的重視，主要原因來自於「沒有落後的孩子」法案（No Child Left Behind, 2002），其要求美國全國3-8年級的學生每年必須接受各州政府的閱讀與數學會考，而考試的目的在於能夠診斷出學生在閱讀與數學的各項概念或認知屬性是否達到精熟狀態，當然也提供了學生優缺點分析的訊息（Huebner, 2010），因此，認知診斷模型可應用於協助教師進行個別化的診斷，也可以提供能力較佳的學生自我學習的方向與目標。

在認知診斷模型中，受試者的認知屬性的分數向量常以 $\alpha=\{\alpha_1,\alpha_2,\ldots,\alpha_K\}$ 表示在 K 個概念上的狀態，每個概念可以對應到精熟與不精熟兩種狀態，也就是說受試者共有 2^K 種認知組型。若概念數 $K=4$，則受試者 $\alpha=(1,0,0,1)$ 表示精熟第 1 個與第 4 個概念，而對第 2 個與第 3 個概念尚未精熟。

為了達到診斷的目的，需要先清楚定義試題和概念之間的關係，大多數的認知診斷模型需要建立一個 Q 矩陣（Tatsuoka, 1985），通常是由學科專家定義的，Q 矩陣由數值 0 與 1 所組成的，表示試卷中的試題所測量的特定概念，如有 J

個試題與K個概念，則 Q 矩陣的大小為$J \times K$，Q_{jk}代表要解答第j個試題是否需具備概念k，若需要則Q_{jk}為 1，反之則為 0，舉例來說，假設 Q 矩陣如下：

$$Q = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1)$$

可以發現要解答試題一需要概念 2 與 3；解答試題二需要概念 1、2 與 3，而試題三只需要具備概念 4，依此類推。

認知診斷模型已發展出相當多的模型，包括統一模型（unified model）、融合模型（fusion model; Hartz, 2002; Hartz, Roussos, & Stout, 2002）、DINA 模式、NIDA 模式（Noisy Inputs, Deterministic “and” Gate model, Junker & Sijtsma, 2001）以及 G-DINA 模式等，以下將介紹本研究使用的兩種模式：DINA 模式以及 G-DINA 模式。

（一）DINA 模式

DINA 模式是許多認知診斷模型的基礎，適用於二元計分試題來進行認知診斷評量的測驗，而其創建與流行則是開始於 Junker & Sijtsma(2001)的研究，在 DINA 模式中，依據受試者是否完全具備試題所測量的認知屬性，將受試者分為兩個類別，在理想的反應模式下，若完全具備時，應當答對該題；若缺少任一個認知屬性則會答錯，但是在實際的作答情形中，會受到雜訊（noise）干擾，在具備應有的認知屬性時，可能會因粗心（slip）而答錯；而猜測（guessing）答對可能會發生於缺乏某些認知屬性的情況下，假設Y_{ij}代表受試者i在試題j的反應，η_{ij}代表受試者i是否完全具備試題j所測量的概念，則 DINA 模式答對試題的機率模型如公式（2）。

$$P(Y_{ij}=1|\eta_{ij}) = (1-s_j)^{\eta_{ij}} g_j^{(1-\eta_{ij})} \quad (2)$$

其中，

$$\eta_{ij} = \prod_{k:Q_{jk}=1} \alpha_{ik} = \prod_{k=1}^{K} \alpha_{ik}^{Q_{jk}}$$

$$s_j = P(Y_{ij}=0|\eta_{ij}=1)$$

$$g_j = P(Y_{ij}=1|\eta_{ij}=0)$$

DINA 模式定義簡單，僅包含粗心及猜測兩參數，是 CDMs 中最容易理解的模型之一，有關 DINA 模式相關的研究及應用可以參閱 Cheng (2009)、DeCarlo (2011)、de la Torre & Lee (2010)、de la Torre & Douglas (2008)、de la Torre (2009a、2009b)、Henson & Douglas (2005)、Rupp & Templin (2008)。

（二）G-DINA 模式

DINA 模式將受試者分類為兩個群體，探討不同答題反應的情形，在受試者缺乏某一個或多個概念時，答對機率皆屬於猜測作答，在這樣的分類方式下，將完全不會任何概念的受試者與缺乏部分受試者的答對機率視為相同，在一般情境中較為少見，因此 de la Torre(2011)提出 DINA 模式的一般化模型，稱為 G-DINA 模式，在對受試者分類群體時，細分為$2^{K_j^*}$個組別，K_j^*代表試題所測量的最大概念數，試題區分的組別數量不盡相同，測量的概念數越多，涵蓋的認知狀態組型就越多，其方程式如（3）。

$$P(\alpha_{ij}^{*}) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk}\alpha_{lk} + \sum_{k'>k}^{K_j^*} \sum^{K_j^*-1} \delta_{jkk'}\alpha_{lk}\alpha_{lk'} + \delta_{j12...K_j^*} \prod_{k=1}^{K_j^*} \alpha_{lk} \quad (3)$$

其中，

δ_{j0}：第 j 題試題的截距。

δ_{jk}：對 α_k 的主要影響。

$\delta_{jkk'}$：對 α_k 與 $\alpha_{k'}$ 交互的影響。

$\delta_{j12\cdots K_j^*}$：由 α_k 到 $\alpha_{k'}$ 的交互影響。

這些參數在模式上各有不同的意義，δ_0 代表答對機率的底線，也就是不具備任何所需的概念時的答對機率；δ_k 代表精熟單一的概念 α_k 時，所能影響的答對機率；$\delta_{kk'}$ 代表 1 階層的交互作用效果，也就是同時具備 α_k 及 $\alpha_{k'}$ 時，所影響的答對機率；同理，$\delta_{12\cdots K^*}$ 代表精熟全部所需的概念時，所影響的答對機率，且影響的程度應較其他項次更加顯著。

以測量 2 個概念的試題為例，假設該題的 Q 矩陣為(1,1)，則受試者可能的認知狀態為{(0,0),(1,0),(0,1),(1,1)}，在 DINA 模式中僅將受試者分為{(1,1)}及{{(0,0),(1,0),(0,1)}兩個群體，而在 G-DINA 模式中，將依 4 種認知狀態各別計算答對機率，受試者的認知狀態不同，則可能有不同的答對機率，而 DINA 模式為 G-DINA 模式的其中一種特定情形，因此，G-DINA 為一般化的模型。

二、參數不變性

在使用一份試題之前，我們必須先檢定模式與資料之間是否具有滿意的適合度（goodness-of-fit），以確定所選用的模式能夠正確地適用於研究分析的資料。而檢定模式是否能適用於所研究分析資料的方法有許多種（Hambleton & Swaminathan, 1985）。

例如：（1）模式對資料所具有的基本假設是否能夠滿足？（2）模式所具有的特性是否能如期獲得？（3）在使用實徵資料與模擬資料下，模式預測力的正確性為何？其中，本節針對模式所具有的特性－參數不變性來做探討。

最常用來檢驗模式所具有特性的方法可分為兩種：能力參數的不變性與試題參數的不變性（余民寧，2009）。

（一）能力參數的不變性

當研究者欲瞭解受試者能力估計值是否穩定時，我們可以選用不同的測驗試題樣本來進行測驗，例如：比較同一範圍內簡單試題與困難試題的答對情形、選擇同一內容範圍中不同題型的測驗試題、以及由題庫中抽取不同內容範圍的試題來比較。當其估計值所相對應的估計精準度差異不大時，便算是符合了能力參數不變性的特性。

（二）試題參數的不變性

在對不同組型（例如：男生與女生、高分組與低分組、北部地區與南部地區……等）的受試者進行測驗後，所獲得該測驗試題的參數估計值（例如：猜測機率與粗心機率）被畫成分佈圖來表示時，除了因為樣本本身大小所造成的分散誤差外，這張圖應該會呈現出直線分佈的型態，而且其基準線是由兩個隨機的相等樣本所建立的。故此，便符合了試題參數不變性的特性。

總而言之，當能力參數（同一群受試者在兩種以上的測驗試題之反應資料，所求得的能力參數估計值）與試題參數（同一份測驗試題讓兩組以上的受試者施測之反應資料，所求得的試題參數估計值）被畫成分佈圖時，研究者便可用來判

斷該分佈圖是否呈現直線分佈情形，此直線的斜率近似為 1，截距近似為 0，則可說明某估計模式適用於該份測驗資料，且具有模式的參數不變性之特性。

參、研究方法

一、研究流程

本研究以認知診斷模型為基礎，探討不同能力分佈的受試者在模擬實驗設計與實徵資料的情境下，對試題參數的估計與受試者認知屬性的診斷辨識率之影響。透過軟體進行模擬樣本資料的產生，使用 DINA 模式與 G-DINA 模式進行參數估計，探討估計的準確性與穩定性。

二、實驗設計

本研究參考 de la Torre & Lee (2010)實驗設計，設定題數為 20 題、概念數為 5 個、樣本數各別為 1000、500、100 人以及試題參數 $s=g=0.1$，每個實驗重複 25 次，模擬實驗所使用的 Q 矩陣設計如表 1 所示，其中包含每個試題測量的 2 個概念及 3 個概念的類型，題數共 20 題。

表 1 **模擬實驗之 Q 矩陣設計**

試題	概念 1	概念 2	概念 3	概念 4	概念 5
I1	1	1	0	0	0
I2	1	0	1	0	0
I3	1	0	0	1	0
I4	1	0	0	0	1
I5	0	1	1	0	0
I6	0	1	0	1	0
I7	0	1	0	0	1
I8	0	0	1	1	0
I9	0	0	1	0	1
I10	0	0	0	1	1
I11	1	1	1	0	0
I12	1	1	0	1	0
I13	1	1	0	0	1
I14	1	0	1	1	0
I15	1	0	1	0	1
I16	1	0	0	1	1
I17	0	1	1	1	0
I18	0	1	1	0	1
I19	0	1	0	1	1
I20	0	0	1	1	1

而在模擬作答反應時，為了探討不同群體的估計結果，使用 Higher-Order DINA(de la Torre & Douglas, 2004)模式及均勻分配來產生認知屬性，其內容分述如下。

（一）以 Higher-Order DINA 模式模擬認知屬性

在完整的認知屬性分佈上共有 2^k 個狀態，若將認知屬性加上高階層的架構時，其複雜度則可被大大降低（Leighton, Gierl, & Hunka, 2004），de la Torre & Douglas(2004)提出 Higher-Order DINA 模式，結合在高階層的能力來控制認知屬性的發生機率，本研究使用來模擬產生認知屬性的邏輯迴歸模型如公式（3）所式，給定一個高階的潛在能力值 θ，並假定元素 α_k 為獨立條件。

$$P(\alpha_k \mid \theta) = \frac{\exp \lambda_1(\theta - \lambda_{0k})}{1 + \exp \lambda_1(\theta - \lambda_{0k})} \quad (3)$$

上式是一個類似雙參數 Logistic 模式，λ_1 代表認知屬性的鑑別度參數，而 λ_{0k} 表示為認知屬性的難度參數，其值愈高代表要精熟此認知屬性越困難。本研究設定概念數 $K=5$，高階層試題迴歸參數 $\lambda_{0k}=(-1,-0.5,0,0.5,1)$ 與 $\lambda_1=1$，受試者的能力分佈為 $\theta \sim N(-1,1)$、$\theta \sim N(0,1)$、$\theta \sim N(1,1)$ 三種，以公式（3）產生受試者 α_k 的狀態。

（二）以均勻分配模擬認知屬性

受試者掌握概念的機率從均勻分佈中抽取，如發生機率大於 0.5，則代表概念為精熟，反之則為不精熟。

表 2 為不同能力分佈模擬樣本之平均概念精熟人數比率，隨者能力高低及概念的難易，精熟的比率最低為 0.08，最高為 0.92，均勻分配下精熟比率皆為 0.5，最後依 DINA 模式計算在每一題的答對機率，再透過與 0 到 1 之間的隨機值比較判定該題的答對與否，因而產生作答反應矩陣。

表 2 **平均概念精熟人數比率表**

能力分佈	K1	K2	K3	K4	K5
theta~ *N*(-1,1)	0.50	0.36	0.25	0.15	0.08
theta~ *N*(0,1)	0.76	0.64	0.50	0.36	0.24
theta~ *N*(1,1)	0.92	0.85	0.76	0.64	0.50
Uniform	0.50	0.50	0.50	0.50	0.50

三、實徵資料

本研究以洪祥堯（2009）於亞洲大學碩士論文「資訊科技融入國小六年級圓形圖單元教學與評量之行動研究」所收集之測驗結果作為實徵資料，該研究以數學科「圓形圖」單元為主，測驗試題為 33 題，概念數為 13 個，以台中縣某國小六年級的四個班級進行教學實驗，編製以知識結構為基礎的教學教材與補救教材，探討教材使用後的成效，進而以貝氏網路為推論工具，應用於診斷受試者的概念與錯誤類型的精準度預測，有效樣本為 290 人，測驗平均通過率為 0.66，難易度屬中間偏易，洪祥堯的研究資料中，有請兩位具豐富教學經驗國小現職教師，依據作答反應及解題歷程，判斷受試者是否具備所測量的概念，作為貝氏網路分析的基準，因此，本研究以此判斷結果作為本研究的真值，並以 DINA 模式與 G-DINA 模式估計其試題參數與概念，以探討 DINA 模式與 G-DINA 模式在實際教育測驗上的運用效果。

該研究的概念數與測驗試題對照，如表 3 所示。

表 3 **實徵資料概念與試題對照表**

概念內容	試題
【概念 01】能判斷出正確的次數分配表	1、14、26
【概念 02】能將分數化成小數	2、15、27
【概念 03】能把分數擴分為分母是 100 的分數	3、16
【概念 04】能把分母是 100 的分數記成百分率	4
【概念 05】能把分數換算為百分率	5、17

表 3 **實徵資料概念與試題對照表（續）**

【概念 06】能用四捨五入法求取百分率的近似值	6、18、28
【概念 07】各分量百分率總和不是 100%時，能調整分量的百分率，使其總和剛好為 100%	7、19
【概念 08】能將各個統計資料換算成所占的百分率	8、20、29、31
【概念 09】能報讀長條百分圖	9、21
【概念 10】能報讀圓形圖	10、22
【概念 11】能依據統計資料判斷出正確的圓形圖	11、23
【概念 12】能從圓形圖判斷出正確的統計資料	12、24
【概念 13】能解決圓形圖的各種應用問題	13、25、30、32、33

四、評估指標

本研究以平均絕對誤差（mean absolute bias, MAB）與診斷辨識率（diagnosis accuracy rate, ACC）作為評估指標，藉以獲得模式估計時的精準度，估計誤差越小，則代表估計越準確，其計算方法如下：

（一）MAB

$$MAB = \frac{\sum_{j=1}^{n} | \hat{p}_j - p_j |}{n} \qquad (4)$$

其中 p_j 為第 j 題試題參數的真值；$\hat{p}_j$ 為第 j 題試題參數的估計值；n 代表總題數。

當受試者不完全具備所需要的認知屬性時可認定為猜測作答，在 G-DINA 模式中，不同的認知屬性狀態具有不同的答對機率，因此，在比較猜測參數的估計上，則以不完全具備認知屬性的答對機率平均值，作為判斷參數不變性的標準；在受試者完全具備所需的認知屬性時的答對機率，就等同於 DINA 模式下 $\eta = 1$ 的答對機率 $(1-s)$，以此計算出粗心參數來進行比較。

（二）診斷辨識率

表 4 **辨識率計算方法表**

		診斷結果分佈	
		答對(1)	答錯(0)
專家判斷（真值）	答對(1)	n_{11}	n_{10}
	答錯(0)	n_{01}	n_{00}

因此，其診斷辨識率的計算方法如下

$$\text{診斷辨識率} = \frac{n_{11} + n_{00}}{N} \qquad (5)$$

其中，

N ：測驗資料的總數。

n_{ij} ：則是將專家判斷的結果 i 與診斷結果 j 的事件組合數。

四、研究工具

（一）MATLAB

本研究使用 MATLAB2008a 來模擬產生受試者之認知屬性，配合試題參數

設定，計算其答對機率值，進而模擬作答反應，並利用來計算估計誤差。

（二）Ox

本研究使用 Ox 軟體（Doornik, 2003）進行 DINA 模式與 G-DINA 模式估計，來估計作答反應組型內的試題參數與受試者認知屬性狀態。

肆、研究結果

本研究結果分為兩部分，首先是模擬樣本資料分析之估計結果，最後為實徵資料之驗證分析。

一、模擬資料分析

本研究模擬實驗結果以不同的估計模式區分，模式下的括號代表樣本大小，估計結果如表 5。

表 5 **模式的估計結果**

估計模式	能力分佈	試題參數估計		辨識率(%)
		MAB(g)	MAB(s)	ACC
DINA (1000)	低能力組	0.0048	0.0024	84.62
	中能力組	0.0022	0.0019	89.99
	高能力組	0.0018	0.0014	95.90
	Uniform	0.0011	0.0028	86.30
DINA (500)	低能力組	0.0064	0.0031	83.03
	中能力組	0.0035	0.0030	89.19
	高能力組	0.0035	0.0015	95.68
	Uniform	0.0012	0.0033	85.92
DINA (100)	低能力組	0.0155	0.0254	79.15
	中能力組	0.0101	0.0053	86.97
	高能力組	0.0087	0.0055	94.49
	Uniform	0.0046	0.0060	84.81
G-DINA (1000)	低能力組	0.0119*	0.0029	78.10
	中能力組	0.0137*	0.0020	84.44
	高能力組	0.0169*	0.0014	93.29
	Uniform	0.0046*	0.0027	83.58
G-DINA (500)	低能力組	0.0158*	0.0054	76.83
	中能力組	0.0170*	0.0028	83.32
	高能力組	0.0203*	0.0017	92.78
	Uniform	0.0064*	0.0033	82.66
G-DINA (100)	低能力組	0.0293*	0.0324	0.7326
	中能力組	0.0288*	0.0072	0.8263
	高能力組	0.0411*	0.0078	0.9147
	Uniform	0.0075*	0.0068	0.8252

*表示 g 參數為多個認知屬性狀態答對機率的平均值

()中代表樣本大小

從表 5 可以發現在試題參數的估計上，估計誤差皆非常的小，最大的估計誤差為 0.0324，顯現 DINA 與 G-DINA 模式在參數估計上具有參數不變性；以平

均概念辨識率的估計結果發現，不同能力分組會影響模式診斷的正確率，能力越高，診斷正確的比率越高，最高可達 95%的辨識率，而低能力組的辨識率大約在 73%~84%左右，皆略低於均勻分配的估計結果；而人數對估計結果的影響，人數越少則估計誤差越大，但變動幅度不是很大，以模擬結果而言，建議樣本數 500 以上較佳。

二、實徵資料分析

根據模擬樣本資料研究的分類方式，並參考 de la Torre & Lee(2010)於實徵資料的能力區分方式，先計算出平均答對題數為 21.89 分，將受試者分為低分組（133 人，45.9%）與高分組（157 人，54.1%），並隨機抽取低分組 132 人與高分組 88 人為低能力組；中能力組從兩組中隨機抽取相同人數；高能力組抽取人數與低能力組人數相反，每一組大約包含了全體 75%的樣本數，使用軟體進行估計並與全體受試者的估計結果比較，三組能力組的平均數與標準差，如表 6 所示。

表 6 **實徵資料－平均數與標準差**

受試者	平均數	標準差
低能力組	20.08	6.95
中能力組	21.35	6.89
高能力組	22.65	6.81
全體	21.89	7.07

表 7 與表 8 為實徵資料在 DINA 模式與 G-DINA 模式下第 1~10 題的試題參數估計結果，完整估計結果在附錄 1，在 DINA 模式下，以表 7 的第 3 題來看，全體受試者所估計出來的 g 參數（0.357）不包含在低、中、高能力組受試者的估計範圍（0.328~0.354）內，低估了猜測參數，依此判斷可以發現，第 3、7、9 題低估了猜測參數；而第 2、5、10 題則是高估了估計結果。在 G-DINA 模式下也可以發現相似的情況，第 9 題低估了猜測參數；第 1、2、5 題則高估了參數。

表 7 **實徵資料的猜測參數估計比較**

ITEM	DINA				G-DINA			
	低	中	高	全體	低	中	高	全體
1	0.219	0.451	0.318	0.321	0.226	0.382	0.324	0.217
2	0.576	0.560	0.645	0.559	0.575	0.571	0.638	0.535
3	0.354	0.328	0.346	0.357	0.356	0.333	0.363	0.360
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.351	0.352	0.391	0.341	0.350	0.352	0.383	0.342
6	0.208	0.220	0.258	0.237	0.215	0.204	0.270	0.239
7	0.221	0.194	0.211	0.240	0.000	0.000	0.000	0.001
8	0.175	0.179	0.203	0.190	0.169	0.175	0.207	0.189
9	0.706	0.726	0.725	0.747	0.711	0.731	0.727	0.752
10	0.528	0.536	0.521	0.517	0.532	0.553	0.522	0.531

類似的結果也可以在粗心參數的估計結果裡發現，如表 8，在 DINA 模式中，第 2、7、8、10 題的參數高估；在 G-DINA 模式中，第 1、2、4、9、10 題也是高估，顯示猜測及粗心參數可能會隨者受試者能力變化而有不同，不一定具有參數不變性的特性。

表 8 **實徵資料的粗心參數估計比較**

ITEM	DINA				G-DINA			
	低	中	高	全體	低	中	高	全體
1	0.111	0.122	0.000	0.000	0.053	0.060	0.056	0.046
2	0.119	0.085	0.101	0.069	0.054	0.050	0.050	0.049
3	0.000	0.000	0.000	0.000	0.069	0.077	0.086	0.068
4	0.000	0.000	0.000	0.000	0.107	0.110	0.117	0.097
5	0.000	0.000	0.000	0.000	0.043	0.047	0.052	0.043
6	0.148	0.116	0.101	0.117	0.037	0.039	0.045	0.037
7	0.262	0.264	0.268	0.228	0.062	0.047	0.061	0.050
8	0.144	0.114	0.132	0.109	0.033	0.035	0.040	0.033
9	0.000	0.000	0.000	0.000	0.069	0.074	0.075	0.062
10	0.014	0.014	0.009	0.008	0.065	0.072	0.075	0.062

若以學科專家診斷受試者所具備的概念為依據，可以獲得 DINA 模式與 G-DINA 模式之診斷平均辨識率如表 9，在 DINA 模式下，低能力組辨識率為 0.7294，中能力組為 0.7406，高能力組為 0.7612，而全體受試者因樣本數較多，平均為 0.7822，雖然辨識率隨者受試者能力提升而增加，但變動幅度不大，可能是因粗心參數估計造成的影響，此結果與模擬實驗相符合；而在 G-DINA 模式下，低能力組的辨識率為 0.7510，中能力組為 0.7458，高能力組為 0.7587，全體辨識率平均為 0.7948，變動幅度較小，亦無隨能力變化有固定的趨勢，除了高能力組之外，辨識率皆高於 DINA 模式估計的結果，但也僅高於 1~2%，且整體來說辨識率皆不到八成，對照模擬實驗結果，可能是樣本數較少所造成的影響。

表 9 **實徵資料之診斷辨識率**

模式	低能力組	中能力組	高能力組	全體
DINA	0.7294	0.7406	0.7612	0.7822
G-DINA	0.7510	0.7458	0.7587	0.7948

就實徵資料結果來看，DINA 模式與 G-DINA 模式的參數估計可能不具有不變性，但是在模擬研究的結果上，卻有非常好的參數不變性，代表資料型態可能不符合 DINA 模式，導致估計結果不穩定；而在實徵資料的辨識率比較上，G-DINA 模式略比 DINA 模式表現佳，此結果與模擬實驗的結果恰好相反，其原因在於模擬實驗中作答反應是以 DINA 模式產生，故 DINA 模式有較好的成效。

伍、結論與建議

由模擬實驗結果得知，DINA 模式與 G-DINA 模式在參數估計上皆有參數不變性的性質，但在辨識率的估計上，能力越高的受試者，認知屬性的估計越準確。

當樣本數較小為 500 人與 100 人時，會出現部分資料估計無法收斂的狀況，顯示 DINA 模式在樣本數較小時，估計可能較不穩定，除此之外，皆符合樣本數愈多，試題參數估計的準確性愈高，概念辨識率也有較佳的結果。

不論在試題參數與辨識率的估計上，DINA 模式在模擬研究的表現都比 G-DINA 模式良好，除了資料是由 DINA 模式產生外，G-DINA 模式估計的參數有 $2^{K_j^*} \times n$ 個，因此需要較多的樣本數，方可達到較佳的估計精準度。

在實徵資料的分析結果上，DINA 模式與 G-DINA 模式的試題參數估計可能都不具有參數不變性的特性，在 de la Torre & Lee(2010)研究中，使用實徵資料的參數，再進行模擬實驗，結果 DINA 模型的參數即具有不變性；而在辨識率的比較，G-DINA 模式辨識率表現較 DINA 模式佳。故在實徵資料形態未知的情境下，較適用於一般化模型的 G-DINA 模式來進行估計，即使資料屬於 DINA 模式的型態，其估計結果也不至於太差。

參考文獻

中文部分

余民寧（2009）。**試題反應理論（IRT）及其應用**。臺北：心理。

洪祥堯（2009）。**資訊科技融入國小六年級圓形圖單元教學與評量之行動研究。**亞洲大學資訊工程學系碩士論文，未出版，台中縣。

英文部分

Cheng, Y. (2009). When cognitive diagnosis meets computerized adaptive testing: CD-CAT. *Psychometrika, 74(4)*, 619–632.

de la Torre, J., & Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika, 69(3)*, 333-353.

DiBello, L. V., Stout, W. F., & Roussos, L. A. (1995) Unified congnitive/ psychometric diagnostic assessment liklihood-based classification techniques. In P. D. Nichols, D. F. Chipman, & R. L. Brennan (Eds.) Cognitively diagnostic assessment. (pp. 361-389). Hillsdale, NJ : Erlbaum.

de la Torre, J., & Lee, Y.-S. (2010). A note on the invariance of the DINA model parameters. *Journal of Educational Measurement, 47(1)*, 115-127.

de la Torre, J., & Douglas, J. (2008). Model evaluation and multiple strategies in cognitive diagnosis: An analysis of fraction subtraction data. *Psychometrika, 73(4)*, 595-624.

de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika, 76(2)*, 179-199.

de la Torre, J. (2009a). A cognitive diagnosis model for cognitively based multiple-choice options. *Applied Psychological Measurement, 33(3)*, 163-183.

de la Torre, J. (2009b). DINA model and parameter estimation: A Didactic. *Journal of Educational and Behavioral Statistics, 34(1)*, 115-130.

DeCarlo L.T. (2011) On the Analysis of fraction subtraction data: the DINA model, classification, latent class sizes, and the Q-matrix. *Applied Psychological Measurement, 35(1)*, 8-26.

Doornik, J. A. (2003). *Object-oriented matrix programming using Ox*(Version 3.1). [Computer software]. London, England: Timberlake Consultants Press.

Hartz, S. (2002). *A Bayesian framework for the Unified Model for assessing cognitive abilities: blending theory with practice.* Unpublished doctoral thesis, University of Illinois at Urbana-Champain.

Hartz, S. Roussos, L., & Stout, W. (2002). *Skills diagnosis: Theory and practice* [User manual for Arpeggio software]. Princeton, NJ: Educational Testing Service.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications.* Boston: Kluwer.

Henson, R. A., & Douglas, J. (2005). Test construction for cognitive diagnosis. *Applied Psychological Measurement, 29(4)*, 262-277.

Huebner, Alan, (2010). An Overview of Recent Developments in Cognitive Diagnostic Computer Adaptive Assessments. *Practical Assessment, Research & Evaluation, 15(3)*. Available online: http://pareonline.net/getvn.asp?v=15&n=3.

Junker, B., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement, 25(3)*, 258-272.

Leighton, J. P. Gierl, M. J., & Hunka, S. M. (2004). The attribute hierarchy method for cognitive assessment: a variation on Tatsuoka's rule space approach. *Journal of Educational Measurement, 41(3)*, 205-237.

No Child Left Behind Act of 2001, Pub. L. No. 107-110 Stat. 1425 (2002).

Rupp, A. & Templin, J. (2008). The effects of q-matrix misspecification on parameter Estimates and classification accuracy in the DINA model. *Educational and Psychological Measurement, 68(1)*, 78-96.

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconception based on item response theory, *Journal of Education Measurement, 20(4)*, 345-354.

Tatsuoka, K. K. (1985). A probabilistic model for diagnosing misconceptions by the pattern classification approach.*Journal of Educational Statistics, 10*, 55-73.

von Davier, M. (2005). A general diagnostic model applied to language testing data (ETS Research Rep. No. RR-05-16). Princeton, NJ: Educational Testing Service.

附錄 1

表 10 **實徵資料的猜測參數估計值**

ITEM	DINA				G-DINA			
	低	中	高	全體	低	中	高	全體
1	0.219	0.451	0.318	0.321	0.226	0.382	0.324	0.217
2	0.576	0.560	0.645	0.559	0.575	0.571	0.638	0.535
3	0.354	0.328	0.346	0.357	0.356	0.333	0.363	0.360
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.351	0.352	0.391	0.341	0.350	0.352	0.383	0.342
6	0.208	0.220	0.258	0.237	0.215	0.204	0.270	0.239
7	0.221	0.194	0.211	0.240	0.000	0.000	0.000	0.001
8	0.175	0.179	0.203	0.190	0.169	0.175	0.207	0.189
9	0.706	0.726	0.725	0.747	0.711	0.731	0.727	0.752
10	0.528	0.536	0.521	0.517	0.532	0.553	0.522	0.531
11	0.691	0.571	0.594	0.557	0.549	0.568	0.598	0.557
12	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
13	0.443	0.466	0.507	0.467	0.444	0.480	0.523	0.475
14	0.152	0.350	0.496	0.436	0.160	0.293	0.493	0.368
15	0.529	0.526	0.645	0.507	0.528	0.455	0.512	0.514
16	0.236	0.303	0.317	0.284	0.239	0.309	0.334	0.263
17	0.236	0.241	0.262	0.275	0.234	0.242	0.263	0.276
18	0.115	0.139	0.106	0.133	0.117	0.140	0.112	0.132
19	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20	0.314	0.325	0.325	0.340	0.316	0.326	0.320	0.337
21	0.201	0.203	0.174	0.223	0.216	0.219	0.180	0.240
22	0.366	0.397	0.403	0.394	0.346	0.365	0.430	0.395
23	0.000	0.258	0.278	0.262	0.253	0.253	0.285	0.257
24	0.402	0.434	0.460	0.432	0.000	0.432	0.459	0.431
25	0.261	0.300	0.295	0.290	0.268	0.308	0.324	0.302
26	0.159	0.138	0.333	0.295	0.153	0.149	0.333	0.340
27	0.000	0.083	0.000	0.080	0.000	0.313	0.363	0.181
28	0.152	0.162	0.155	0.140	0.142	0.165	0.162	0.138
29	0.438	0.463	0.429	0.448	0.439	0.469	0.432	0.446
30	0.224	0.173	0.141	0.220	0.229	0.179	0.136	0.225
31	0.170	0.162	0.193	0.178	0.170	0.164	0.199	0.181
32	0.207	0.264	0.287	0.232	0.199	0.255	0.290	0.222
33	0.238	0.228	0.223	0.258	0.234	0.231	0.240	0.252

表 11 **實徵資料的粗心參數估計值**

ITEM	DINA				G-DINA			
	低	中	高	全體	低	中	高	全體
1	0.111	0.122	0.000	0.000	0.053	0.060	0.056	0.046
2	0.119	0.085	0.101	0.069	0.054	0.050	0.050	0.049
3	0.000	0.000	0.000	0.000	0.069	0.077	0.086	0.068
4	0.000	0.000	0.000	0.000	0.107	0.110	0.117	0.097
5	0.000	0.000	0.000	0.000	0.043	0.047	0.052	0.043
6	0.148	0.116	0.101	0.117	0.037	0.039	0.045	0.037
7	0.262	0.264	0.268	0.228	0.062	0.047	0.061	0.050
8	0.144	0.114	0.132	0.109	0.033	0.035	0.040	0.033
9	0.000	0.000	0.000	0.000	0.069	0.074	0.075	0.062
10	0.014	0.014	0.009	0.008	0.065	0.072	0.075	0.062
11	0.132	0.000	0.000	0.000	0.054	0.059	0.063	0.053
12	0.000	0.000	0.000	0.000	0.073	0.071	0.076	0.072
13	0.145	0.118	0.106	0.087	0.048	0.051	0.053	0.045
14	0.180	0.193	0.243	0.246	0.047	0.057	0.059	0.051
15	0.193	0.144	0.165	0.147	0.054	0.051	0.052	0.049
16	0.000	0.000	0.000	0.000	0.062	0.076	0.084	0.064
17	0.000	0.010	0.001	0.021	0.039	0.042	0.047	0.040
18	0.362	0.336	0.308	0.309	0.029	0.033	0.032	0.029
19	0.000	0.000	0.000	0.000	0.088	0.059	0.069	0.052
20	0.061	0.053	0.044	0.043	0.040	0.043	0.046	0.039
21	0.000	0.000	0.000	0.000	0.067	0.078	0.077	0.069
22	0.000	0.000	0.000	0.000	0.063	0.071	0.075	0.062
23	0.000	0.000	0.000	0.000	0.048	0.053	0.059	0.048
24	0.418	0.364	0.317	0.364	0.090	0.047	0.050	0.042
25	0.049	0.065	0.043	0.063	0.043	0.048	0.051	0.042
26	0.151	0.000	0.042	0.050	0.046	0.046	0.055	0.050
27	0.000	0.000	0.000	0.018	0.077	0.048	0.051	0.040
28	0.274	0.255	0.240	0.233	0.032	0.036	0.038	0.030
29	0.052	0.044	0.036	0.052	0.043	0.045	0.048	0.041
30	0.334	0.333	0.311	0.319	0.040	0.040	0.037	0.038
31	0.196	0.207	0.184	0.176	0.033	0.034	0.039	0.032
32	0.197	0.217	0.203	0.212	0.039	0.045	0.049	0.038
33	0.349	0.303	0.297	0.287	0.041	0.044	0.046	0.040