

從認知診斷模式探究比例式解題的學習 路徑

蔡易達* 鄒慧英**

中文摘要

本研究旨在探索學生在比例推理中解未知數問題時的學習路徑。根據文獻回顧，本研究之比例推理測驗使用五個特質，建立了三個階層模型：一個非結構化模型和兩個發散模型。然後利用認知診斷模型（CDMs）及其對應的 Q 矩陣估計學生組型，及各個組型在每個試題的期望值。此外，該測驗根據比值內/比值間和整數/非整數將試題分為四種類型，以了解不同學生組型和題型之間的表現差異。研究結果顯示，三種階層模型與高分組的作答反應具有高度一致性；比較 DINA 模型與特質階層模型（DINA-AHM）和階層診斷分類模型（HDCM）後發現，飽和 HDCM 模型可提供較多訊息；透過學生組型之預期分數所構建的學習路徑，可知悉精熟那個特質有助於提高學生的表現；四種題型中以比值內和比值間均為非整數的題型難度最高。

關鍵字：認知診斷模型、特質階層模型、觀察分數、學習路徑

*國立臺南大學教育系測驗統計博士班

**國立臺南大學教育系教授

通訊作者：蔡易達，email: charlex@gmail.com

Constructing the Learning Path of Proportional Reasoning by Using Cognitive Diagnosis Models

Yida Tsai* Hueying Tzou**

ABSTRACT

This study aims to explore the learning paths of students in proportional reasoning when solving problems with unknowns. Based on the literature review, the proportional reasoning test employed five attributes and three hierarchical models were established: an unstructured model and two divergent models. Cognitive Diagnosis Models (CDMs) and their corresponding Q-matrices were then utilized to estimate student profiles and the expected values on each item for each profile. Additionally, the test categorizes items into four types based on within/between-ratio and integer/non-integer. Thus, it allowed us to understand performance differences among different student profiles and item types. The results indicated high consistency between three hierarchical models and responses from high-performing groups. Comparing the DINA with Attribute Hierarchy Model (DINA-AHM) and the Hierarchical Diagnostic Classification Model (HDCM), the saturated HDCM model provided more information. The learning path constructed from the expected scores of student profiles allowed us to understand which attribute mastery was beneficial for improving student performance. Among the four item types, those involving non-integer ratios in both within and between ratios are the most difficult.

Keywords: Cognitive Diagnosis Models(CDMs) 、 Attribute Hierarchy Models(AHMs) 、 Observed Score 、 Learning Path

* Ph.D. Student, Department of Education, National University of Tainan

**Professor, Department of Education, National University of Tainan

Corresponding Author: Yida Tsai, e-mail:charlex@gmail.com

壹、緒論

一、研究背景與動機

認知診斷模式(Cognitive Diagnosis Models, CDMs)的目的是分析學生在教學內容的精熟程度，且給予師生立即的回饋，讓學生了解自身在每項學習技能上的優劣勢。學生在認知領域的能力可視為有許多技能(skill)或者特質(attribute)所組成。每個試題所量測的特質可組成 Q 向量，因此整份評量的每個試題向量，再合併一個 Q 矩陣(Tatsuoka, 1983)，在 CDMs 中，試題與特質要先在 Q 矩陣中設定，然後針對學生作答反應進行分析，得到學生的是否具備該項特質，進而將整體受試學生分類成不一樣的特質精熟組型(attribute mastery patterns, AMPs; Tatsuoka, 1983)或稱為學生組型(profile)，藉此教師能為學生進行針對性的教學加強，以提高學生的進步幅度(Duckworth et al., 2009)。

學者們致力於 CDMs 的研究，發展出不同的模式，去探討認知過程中，技能或特質如何影響學生對試題的作答：像是 DINA (the deterministic inputs, noisy “and” gate model) 模式(Haertel, 1989; Junker & Sijtsma, 2001; Maris, 1999) 模式是假定學生若無具備答對該試題所測之所有特質時，學生就不會答對此題；而有別於 DINO (the deterministic inputs, noisy “or” gate) 模式(Templin & Henson, 2006) 是假定學生只需具備該試題的其中一個特質就預期能答對該題，由這兩種最常使用 DINA 及 DINO 的模型來說，可知特質之間是否相關或者互補，使 CDM 模式可分為非補償型(non-compensatory)或補償(compensatory)模型(Henson et al., 2009)，然而避免特質之間是存在補償性或非補償性的假設，Henson 等人在 2009 年提出 Log-linear CDM(LCDM)。與 de la Torre 在 2011 年發展外所謂的飽和的 CDM 模式 GDINA，可經過參數調整可以轉換不同的 CDM 模式，而用 logit 函數作為連結的 GDINA 模式等價於 LCDM(Ma & de la Torre, 2020)。

CDMs 應用是否成功需要根據認知診斷評量的原則來設計評量，迄今在 CDM 框架下開發的診斷測驗為數不多，其中 Tjoe 和 de la Torre 在 2014 年針對中學生的比例推理測驗(the Proportional Reasoning Assessment, PRA)，根據此測驗進行研究，利用單向度 IRT 去估計學生的能力值，可得到相同能力值的學生，卻精熟不同的特質，導致其學生的 profile 不盡相同。另外探討特質之間是否有階層，衍生出特質階層模型(Attribute Hierachy Models, AHMs)，透過階層結構來解釋特質相互依賴的可能性，掌握某一個特質為是另一個特質或多個特質的先決條件，因此可讓特質的組合數量減少(Leighton et al., 2004)。以 PRA 為例，量測的特質數有 8 個如表 1，代表學生組型可能會有 的組合，以常見的階層結構分成線性、發散、收斂及未結構化，進而根據結構大幅減縮可能的學生組型，而以本研究會採用未結構化與發散的結構進行探討，其結構圖如圖 1，特質 A1 為所有特質的先備條件，未結構化模型可看出，從特質 A1 可以發展出其他特質，其學生組型變成 129 個；發散模型中，另外發現特質 A2a 是特質 A2b 的先決條件、A3a 及 A3b 是 A4 的先決條件，經過計算可知學生組型則減少至 61 個。討論特質之間是否有階層，如掌握 A3a 及 A3b 為是 A4 的先決條件，代表學生在學習比例推理時是有其順序性，代表學生須先精熟 A3a 及 A3b，才會精熟 A4。本研究想了解國中學生在比例推理的學習路徑(Learning path)，根據認知診斷理論(Leighton & Gierl, 2007; Tatsuoka, 2009; Rupp et al., 2010)，學習的起點跟終點是相同的，如同在 CDMs 分析的結果，學生分類到全

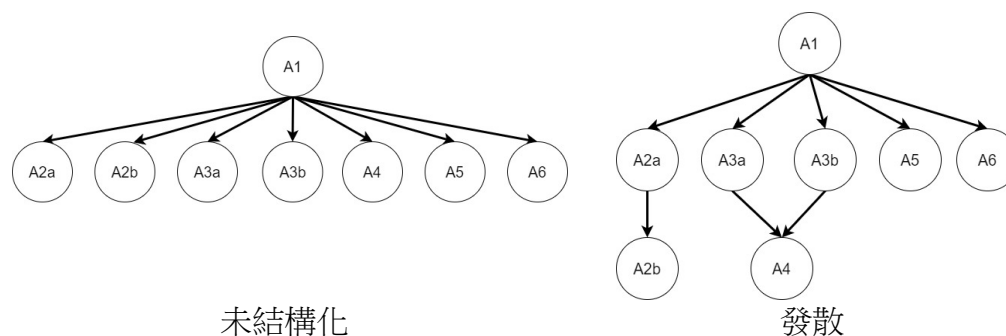
部精熟跟全部不精熟的比例是相對最高的，由於每個特質難易度不同，每個學生精熟特質的順序均有不同，促使每個人學習的路徑卻不竟相同(Gotwals and Songer, 2010)。

表 1
比例推理測驗(PRA)之的特質描述

編碼	特質內涵
A1	進行比例推理中所需要的先備知識
A2a	比較分數
A2b	排序分數
A3a	建構比值
A3b	建構比例式
A4	辨識數組數值間乘法關係
A5	區辨比例關係或非比例關係
A6	運用算則解比例推理問題

資料來源：Tjoe and de la Torre(2014)

圖 1
特質階層模型範例



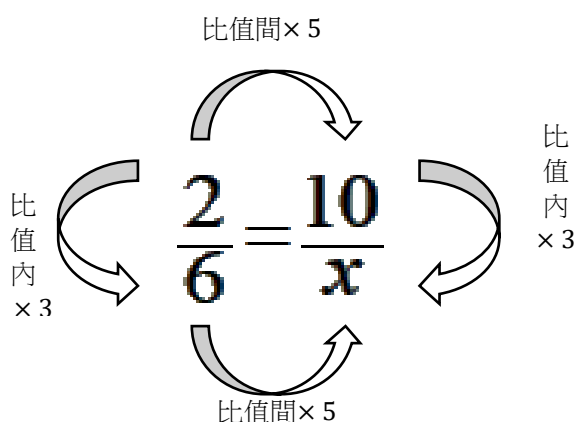
說明：左為未結構化結構，右為發散結構

Templin and Bradshaw(2014)將 LCDM 與 AHM 兩者連結發展出階層診斷分類模型(Hierarchical Diagnostic Classification Model, HDCM)，而在模擬研究中的屬性是線性(Linear)階層模型，而研究結果發現在 DINA 和 DINO 模式下分類的準確性明顯不如 LCDM 及 HDCM；由於相較於 HDCM 等飽和模式來說，DINA 模式較為簡單易懂，因此有研究利用 DINA 模式與 AHM 進行研究，也利用未結構化或發散模型及伴隨 Q 矩陣去估計學生組型的分布，再進行繪製樹狀圖以了解學生的學習路徑，稱為 DINA-AHM(Cai et al., 2018; de La Torre et al., 2010; Köhn & Chiu, 2019)；而 Yamaguchi 和 Templin 於 2022 年提出用以 LCDM 為基礎發展其計算公式，以各學生組型的測驗分數之期望值作為學習路徑的依據，掌握越多的特質在此份診斷評量的分數更高，另對學生、家長來說用分數解釋其分析結果較為容易(Yamaguchi & Templin, 2022)。本研究欲設定特質的階層關係，如未結構化與發散，利用 DINA-AHM 及 HDCM 兩種模式進行分析學生組型及其測驗之期望值，透過觀察分數之差距去建構學生的學習路徑，也能知道那些特質會影響此份測驗的期

望值，可作為教學或學習的參考，另透過測驗分數瞭解來學生組型，知道影響學生的精熟特質之情形。

圖 2

比值內(within)及比值間(between)的乘法關係示意圖



資料來源：Steinthorsdottir and Sriraman(2009)

以 PRA 的特質進行對照，A3a 及 A3b 為建構比值與建構比例式，而解比例式的未知數，學生進行比例式題，最常使用的策略有累加、倍數乘法、約分或擴分及單位分數等，而其中特質 A4 辨識數組數值間乘法關係，即為學生最常使用「內項乘積等於外項乘積」來計算出比例式中的未知數(Lobato et al., 2010)，然而學生精熟特質 A5「區辨比例關係或非比例關係」，進而列出正確比例關係，在利用比例推理所需運算法則進行解題，此為特質 A6。而根據比例式時比值內或比值間為整數(integer, I)或非整數(Non-integer, N)，題型可分成 4 種，分成 A、B、C 及 D，均影響其該題的難度(Steinthorsdottir et al., 2007)，如表 2 所示。

表 2

比例解題之未知數題型

		比值內	
		整數(I)	非整數(N)
比值間	整數(I)	A : II	C : IN
	非整數(N)	B : NI	D : NN

資料來源：Riehl and Steinthorsdottir(2019)

以 A-II 題型為例，比值內與比值間均為整數，如 $\frac{120}{3} = \frac{x}{6}$ ，學生利用累加的策略，即為 $3+3=6$ ，故 $x=120+120=240$ ；或是 $120 \div 3 = 40$ ，故 $x=6 \times 40 = 240$ 。而 D-NN 題型為比值內與比值間均為非整數，如 $\frac{5}{4} = \frac{7}{x}$ ，因為分數間為非整數，所以無法用累加的方式去運算，只能用非整數的倍數關係，故涉及到分數的乘除，難度較高，然而較佳的解法為「內項乘積等於外項乘積」， $5x=4 \times 7, x = \frac{4 \times 7}{5} = \frac{28}{5}$ ，B-NI 及 C-IN 題型之示例題及解法請見表 4。

過往的研究探討不同模型去看學生組型或特質的精熟程度，較少研究將學生組型與不同題型之間的答題表現或解題策略進行分析，因此本研究欲利用 PRA 中解未知數的題目，以 AHMs 之未結構化及發散結構且利用 DINA-AHM 及 HDCM 兩個模式進行估計學生在所測量的特質及四種題型的學習表現，提供教師更多的資訊來提升學習成效。另一藉由測驗分數的期望值去建構學生在比例式解題上各個特質之學習路徑，及解未知數的四種題型之答題表現；另探究不同學生組型的學生在四種題型可加強哪些特質有助於提升學習表現。

二、研究目的與待答問題

基於上述的研究背景與動機，具體而言，研究目的有二：

- (一) 探討在 DINA-AHM 及 HDCM 模式探討學生在比例推理解題策略的學習路徑。
- (二) 根據學生組型探討 4 種未知數題型的答題表現。

根據上述研究目的，本研究提出下列兩個待答問題：

- (一) 在 DINA-AHM 及 HDCM 模式上，設定特質階層是否符應學生的比例推理解題策略的學習路徑。
- (二) 各學生組型在四種題型上之答題表現是否有差異，是否能找出關鍵的特質提升學習。

貳、文獻探討

一、認知診斷模式

認知診斷模式是一系列受限制的潛在類別模型(restricted latent class models)，用來了解學生在知識、思維過程和解決問題的策略。從建模觀點來看，Templin 和 Henson(2006)指出 CDMs 是將學生的作答反應建模成潛在變量的函數，即一組給定的特質或技能；從評量目的的角度來看，CDMs 用於以考生分類形式提供一組特質的診斷資訊，而不是如 CTT 或 IRT 模型的描述性和總結性的能力值(de la Torre & Minchen, 2014)；另外 CTT 和 IRT 模型，學生的分數視為學科表現，然而分數相同的學生在不同技能方面的成就水平可能會有所不同，為了獲得可用於促進教學與學習的診斷資訊，我們可使用 CDMs 進行分析，是否掌握該測驗所量測的特質，是否有達到精熟標準。

Q 矩陣是 CDMs 的關鍵組成部分，反映了每個試題(行)和給定特質(列)之間的關係。例如，item 1 的 q-vector(行向量)是[10000]，表示第一個特質是在答對這個試題上的唯一必需特質。PRA 是根據認知診斷模型所開發的認知診斷測驗，而 PRA 的 Q 矩陣可參考 Likun 等人(2020)之表 1，試題數為 31 題及量測 8 個特質。31 題中有量測到單一特質，或二至三個特質，而 Q 矩陣的設定應由測驗學者、學科專家及教學工作者共同設定，為求達到此診斷目的。

考慮第 j 題量測到兩個特質： A_1 (先備知識)和 A_2 (比較分數大小)，假定第 i 位學生的真實組型(true profile) $\alpha_i = (A_{i1}, A_{i2})$ 。因為每個特質會假設若達到精熟(1)或未

達精熟(0)，故可能會有 4 種可能的組型。例如：當 $\alpha_i = (1,1)$ 代表第 i 位學生具備精熟兩個特質； $\alpha_i = (0,0)$ 代表此學生不具備精熟任何一個特質。CDMs 利用試題反應函數，將作答反應映射至特質上。GDINA 模式，是文獻中最通用的 CDMs 模式之一，並包含幾種廣泛使用的模型，對於有第 i 位學生的組型為 α_i 答對第 j 題的機率函數可寫成

$$P_j(\alpha_i) = \delta_{j0} + \delta_{j1}A_{i1} + \delta_{j2}A_{i2} + \delta_{j1,2}A_{i1}A_{i2} \quad (1.1)$$

在(1.1)中的截距項，即 $\delta_{j0} = P_j(00)$ 所表示為代表不精熟任一特質的答對此題之機率，另 $\delta_{j1} = P_j(10) - P_j(00)$ 表示為 A_1 的主要效果會因為精熟第一個特質 A_1 而增加，相同的 $\delta_{j2} = P_j(01) - P_j(00)$ 表示 A_2 的主要效果會因為精熟第二個特質 A_2 而增加，最後參數 δ_{j12} 是因為精熟這兩個特質而導致答對此題的機率而變化的交互效果。

根據 GDINA 模型的作答反應函數包含試題中每一個特質的主要效果以及所有特質間可能的交互效果，當某一試題 j 量測到 K 個試題時，GDINA 下就會產生 2^k 的試題參數，因此 GDINA 又稱為飽和模式，其作答反應函數可以經過簡化，轉換成其他簡約模式，另外還有常見的簡約模型如 A-CDM、R-RUM 或者 LLM 都能加上一些限制條件的調整，若需要更多的細節，建議請閱讀 de la Torre(2011)。

近年關於 CDM 的文獻中，因為 DINA 與 DINO 模式是容易理解及操作，為最廣泛應用的模型，兩者之主要差別在於特質之間是否有補償性，然而檢驗特質之間具有補償性或沒有補償性的假設是不容易的，因此 Henson 等人(2009)提出 Log-linear CDM(LCDM)，此模型在試題層次上可讓特質之間如 GDINA 模式在參數設定上，可以有補償或非補償的參數，亦可視為 LCDM 等價於用 logit 函數等化的 GDINA 模型，在 GDINA 套件中，其模型的指令為 logitGDINA(Ma & de la Torre, 2020)，其函數表示為

$$P_j(\alpha_i) = \frac{\exp(\lambda_{j0} + \lambda_{j1}A_{i1} + \lambda_{j2}A_{i2} + \lambda_{j1,2}A_{i1}A_{i2})}{1 + \exp(\lambda_{j0} + \lambda_{j1}A_{i1} + \lambda_{j2}A_{i2} + \lambda_{j1,2}A_{i1}A_{i2})} \quad (1.1)$$

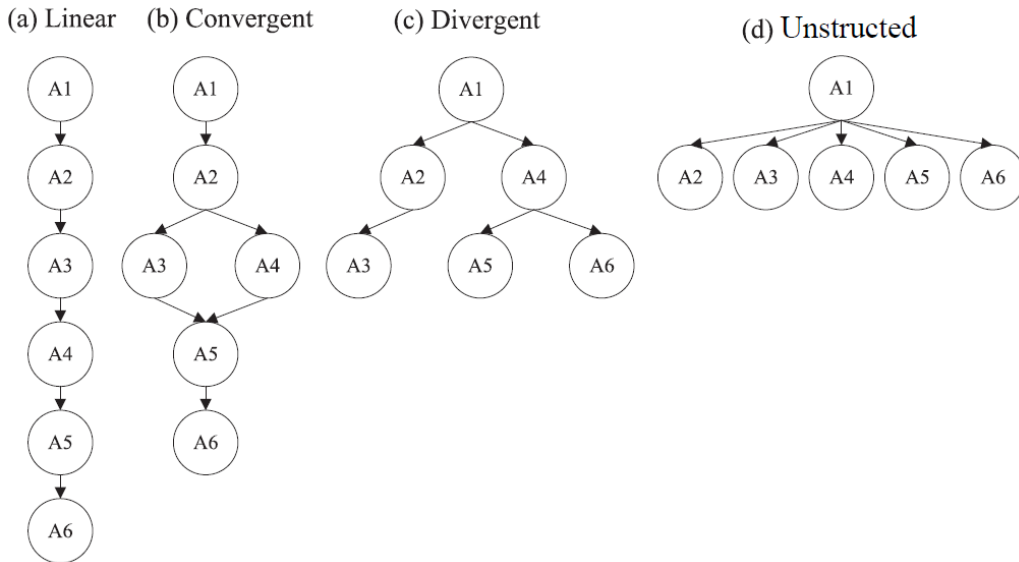
其中(1.2)的截距項 λ_{j0} 代表受試者完全不精熟第 j 題所測量的兩個特質，所答對之對數勝算比(log-odds)； λ_{j1} 及 λ_{j2} 為其主要效果，表示各別精熟特質 1 或特質 2 所增加的答對第 j 題之對數勝算； $\lambda_{j1,2}$ 為特質 1 及特質 2 的雙向交互效果，表示受試者均精熟第 j 題所量測的兩個特質而增加答對第 j 題之對數勝算。由此可知，LCDM 與 GDINA 相同均屬於飽和模式，若該題測量 k 個特質，則試題參數會有 2^k 個，以往都是利用 Mplus 進行估計，現可用 R 語言的 GDINA 套件進行估計及資料分析(Ma & de la Torre, 2020)。

二、階層特質模式與學習路徑

探討特質間是否存在階層關係，倘若特質 A1 是 A2 的先決條件，代表受試者得先精熟特質 A1，才有可能精熟特質 A2。此特質階層應用在教育現場，代表教學的先後次序，以數學科教師來說，教學的內容包含許多的概念與解題技能，而這些概念與解題技能都是彼此相關且往往都是建立在另外一個概念或技能之上。例如加減乘除，大部分的學生都是先熟練加法，進而熟練減法；乘法建立在連續加法，除法有時候學生會有減法的概念去建構觀念等。而常見的階層模式分別為線型

(Linear)、收斂(Convergent)、發散(Divergent)及未結構化(Unstructured)，以圖 3 為例，有 6 個特質的階層模型為線型，代表學生須熟練特質 A1，才能熟練特質 A2，以此類推，故學生僅有 7 種學生組型，即[000000]、[100000]、[110000]、[111000]、[111100]、[111110]及[111111]等 7 種，相同的，收斂、發散及未結構化模型下各有 8 種，16 種及 33 種的學生組型，而在每個模型下，這些學生組型的比例總和依舊為 1。

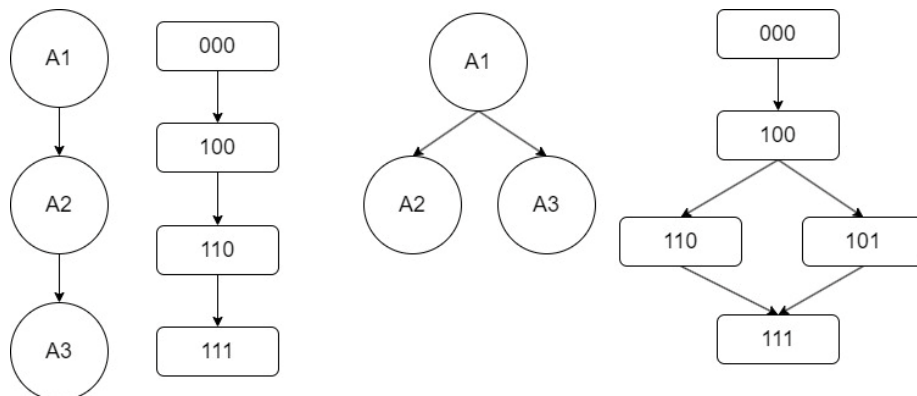
圖 3
四種常見的特質階層模型



資料來源：Leighton et al. (2004)

欲利用 AHM 想了解學生的學習路徑，可依照特質之間的階層模型來設定的學生組型的發展，若以特質數為 3 來討論，如圖 4 所示，如果三個特質是線性，代表學生會先精熟 A1，依序 A2、A3，配合學生組型的學習路徑為[000] → [100] → [110] → [111]；若三個特質的階層模型為發散，即為精熟 A1 後，才有可能精熟 A2 或 A3，因此配合學生組型的學習路徑[000] → [100]，而後面發展就可能為[110] 或 [101]，最後達到精熟三個特質[111]。

圖 4
特質階層模型與學習路徑示例圖



資料來源：Köhn & Chiu(2019)

當特質數越多，階層模型的不同，其學習路徑就會相對複雜，學生的起點(全

無)跟終點(全有)是相同的,每個特質之難易度不同,每個人學習路徑卻不盡相同, Jia et al. (2021)利用大型評比 PISA2012 的資料,進行多種 CDM 模式,包含 DINA、DINO、GDINA、ACDM、LLM、RRUM 及 LCDM,探討 14 個國家或地區之 15 歲學生在統計知識的學習路徑,此研究之 4 個特質並無階層關係,各種模型利用適配指標進行比較,發現 LLM 表現最佳,而學習路徑的繪製是以受試學生之學生組型按所占比例依序排列,其精熟數逐步增加,反映出能力提升,進而建構其學習路徑,學生組型的數量影響分類結果,若用 DINA 等模式而沒有利用階層模型進行分析,全無(0)或全有(1)佔了大部分的比例,無法詳細資訊建構出學習路徑, Templin 和 Bradshaw 在 2014 年發展出 HDCM,透過階層模型的設定,進行 Q 矩陣的調整進行模擬研究,結果表明在線性階層下,DINA 與 DINO 的分類準確度明顯低於 LCDM 或 HDCM;由於 DINA 模式在試題參數僅有猜測參數(g_j)及失誤參數(s_j),因此有研究利用伴隨特質階層的 Q 矩陣進行 DINA 分析,稱之為 DINA-AHM(Cai et al., 2018; de La Torre et al., 2010; Köhn & Chiu, 2019),若受試者具備該題的所有特質,則答對該題的機率為 $1 - s_j$,反之則猜對機率為 g_j ,再從階層關係去建構期學習組型之樹形圖,以圖三為例,假設 Q 矩陣(特質數為 3,試題數為 4),各學習組型之間跟試題答對機率如表 3,可見 DINA-Linear 模式下,學生組型有 4 種,而 DINA-Diverge 模式下,學生組型多了(101),在第三題的答對機率為 $1 - s_3$,以兩種結構進行比較,精熟 A2 或 A3 特質,何者影響其答題表現。

表 3
DINA-AHM 分析之試題參數對照表

試題	Q 向量			線型(DINA-Linear)				發散(DINA-Diverge)				
	A ₁	A ₂	A ₃	000	100	110	111	000	100	110	101	111
1	1	0	0	g_1	$1 - s_1$	$1 - s_1$	$1 - s_1$	g_1	$1 - s_1$	$1 - s_1$	$1 - s_1$	$1 - s_1$
2	1	1	0	g_2	g_2	$1 - s_2$	$1 - s_2$	g_2	g_2	$1 - s_2$	$1 - s_2$	$1 - s_2$
3	1	0	1	g_3	g_3	g_3	$1 - s_3$	g_3	g_3	g_3	$1 - s_3$	$1 - s_3$
4	1	1	1	g_4	g_4	g_4	$1 - s_4$	g_4	g_4	g_4	g_4	$1 - s_4$

資料來源：Köhn & Chiu(2019)

在 2022 年 Yamaguchi 和 Templin 提出用以 LCDM 為基礎發展其計算公式,以各學生組型(α_c)在測驗每一題分數之期望值(ω_{ci})之總和作為學習路徑的依據,掌握越多的特質在此份診斷評量的分數更高,其期望值及變異數公式表示為

$$E[X_c | \alpha_c] = \sum_{i=1}^I E[X_{ci} | \alpha_c] = \sum_{i=1}^I \omega_{ci} \quad (2.1)$$

$$V[X_c | \alpha_c] = \sum_{i=1}^I V[X_{ci} | \alpha_c] = \sum_{i=1}^I \omega_{ci} (1 - \omega_{ci}) \quad (2.2)$$

而條件測量標準誤(conditional standard error of measurement, CSEM)是應試者觀察到的分數的標準偏差,在 2014 年出版的教育與心理測驗標準(the Standards for educational and psychological testing)一書提及 CSEM 在對具有固定、不變的真實分數為其重要指標, Yamaguchi 和 Templin 將兩個學生組型之間的期望分數的差距與 CSEM 之平均做比較,進而看出組型間是否有改變,畫出哈斯圖 (Hasse diagram)來建構其學習路徑,對照 DINA-AHM,若能將學生組型及其各題的猜測參數及失

誤參數轉換成各題的期望分數進行加總且計算 CSEM，便可比較兩種模型之所建構的學習路徑與預先設定的階層模型能否符合，亦能透過觀察分數可判別學生在特質精熟程度之差異。

在 CDMs 中欲檢驗設定之特質階層與學生作答反應是否一致，可由 2009 年 Cui 與 Leighton 所提出階層一致性指標(Hierarchy Consistency Index, HCI)，其公式為：

$$HCI_i = 1 - \frac{2 \cdot \sum_{j \in S_{\text{正確}_i}} \sum_{g \in S_j} X_{ij} (1 - X_{ig})}{N_{c_i}} \quad (2.3)$$

其中 $S_{\text{正確}_i}$ 代表學生 i 答對試題形成的集合、 X_{ij} 為學生 i 在試題 j 的得分(0 或 1)，其試題屬於 $S_{\text{正確}_i}$ 、 S_j 代表試題 j 所測量的特質的集合所形成的集合、 X_{ig} 為學生 i 在試題 g ，且試題 g 是在 S_j 這個集合內，以及分母的 N_{c_i} 是代表學生 i 答對試題的比較總數，從公式可得知計算每一個學生的 HCI_i ，可用來解釋特質階層是否能解釋學生表現，其值介於 -1 至 +1 之間，若 HCI_i 接近 -1，代表學生 i 作答反應與特質階層並不適配，即為特質階層無法提供學生 i 在進行測驗時的表現，亦或有猜測等異常作答行為，換言之，若 HCI_i 接近 +1，特質階層可以解釋學生 i 的學習表現或解題表現(Cui & Leighton, 2009)。因此透過 $HCI_{\text{平均}}$ (學生 HCI 之平均)評估 CDMs 與測驗資料的整體模式適配度，由於測量的特質數越多，其標準也隨著 CDMs 的複雜而進行調整，多以 .60 作為良好適配的標準，大於 .80 視為非常適配；Roberts, Alves, Chu, Thompson, Bahry 與 Gotzmann(2014)利用測量 8 個特質的數學診斷評量，利用專家認知模型(線性)跟學生認知模型(發散)與學生作答反應進行 $HCI_{\text{平均}}$ 計算，以將 $HCI_{\text{平均}}$ 分成三個類別，一是 $HCI_{\text{平均}} > .80$ 非常適配，其次是良好適配，其 $HCI_{\text{平均}}$ 介於 .60 至 .80 之間，第三類為適配不佳，其 $HCI_{\text{平均}} < .60$ (Cui & Leighton, 2009)；Roberts 等人的研究根據學生的能力分群進行 $HCI_{\text{平均}}$ 之比較，發現能力佳的學生，兩種發散模型的 $HCI_{\text{平均}}$ (.85 和 .87)大於線型模型(.80)，達非常適配之標準，而對於中等能力的學生結果相似，可是兩種發散模型的 $HCI_{\text{平均}}$ (.48 和 .51)大於線型模型(.47)，均未達 .60；而低能力組 $HCI_{\text{平均}}$ (-.26 和 -.25)大於線型模型(-.32)，數值均小於 0，可見學生能力也會影響 $HCI_{\text{平均}}$ 之適配情形，因此本研究使用 PRA 測量為 8 個特質，且利用兩種不同之階層模型， $HCI_{\text{平均}}$ 以 .60 作為良好適配之標準，作為學生作答組型是否與設定階層模型能否一致性之標準。

另古典測驗理論以觀察分數為真實分數(true score)加上誤差分數為理論架構，發展出信度，其公式為

$$\rho_{\text{CTT}} = \frac{\text{真實分數之變異數}}{\text{觀察分數之變異數}} = 1 - \frac{\text{誤差分數之變異數}}{\text{觀察分數之變異數}} \quad (2.4)$$

Yamaguchi 和 Templin(2022)利用上式，配合 CDMs 所估計的各學生組型的觀察分數提出信度指標 ρ_{DCM} ，其公式如下：

$$\rho_{DCM} = 1 - \frac{\sum_{c=1}^{2^4} \pi_c V[X|\alpha_c]}{V[X]} \quad (2.5)$$

其中 $V[X]$ 為整份測驗觀察分數之變異數； $\sum_{c=1}^{2^4} \pi_c V[X|\alpha_c]$ 是所有學生組型之條件變異數的加權總和，從 ρ_{DCM} 可得知學生組型之間的總分差異之變異數比例。

本研究探討學生在比例推理解題之學習路徑，以 GDINA 套件進行分析，以非結構化及發散模型，利用 DINA-AHM、HDCM 兩種模式進行估計學生之期望分數，去建構其學習路徑，進行比較。

三、PRA 與未知數題型

以 Tjoe & de la Torre(2014)根據認知診斷模式所發展出的 PRA，是屬於認知診斷測驗，此測驗共有 31 題，量測 8 個特質，其中 A1 是先備知識，為其他 7 個特質的先決條件；A2 細分兩個特質：分數比較大小(A2a)及分數排序(A2b)，可知 A2a 是 A2b 的先決條件；A3 也細分為兩個特質，分別為建立比值(A3a)及建立比例(A3b)；A4 是確定一組數字之間的乘法關係，可知 A3a、A3b 皆為 A4 的先決條件；另外兩個特質分別為區辨比例關係或非比例關係(A5)及運用算則解比例推理問題(A6)。關於 CDMs 的文獻多為模式的理論架構的發展，利用模擬資料進行分析，觀察其模式是否適配、學生組型分類之準確度、各特質精熟的比例、發展 Q 矩陣的驗證方法、使用不同的估計方法等，佐以 PRA 資料進行 CDMs 分析，或以古典測驗理論、試題反應理論之結果進行試題難度、信度、鑑別度之比較，得到相同能力值的學生，卻精熟不同的特質。換言之，其學生組型不一定相同(Ma et al., 2020)，此外 PRA 中，測驗題型大致上可以分成三類，一是分數的比較大小、其二是解未知數題型，第三種為辨認是否為比例關係，其中解未知數的題型題數最多，多為生活應用題且辨認兩組數量之間是否存在比例關係，過往相關文獻多為模式估計方法之比較，探究特質精熟之情形，較少研究會利用測驗之題型與學生組型進行比較分析，因此本研究欲針對 PRA 之解未知數題型進行分析，探討其答題表現。

表 4
四種解未知數之題型示例題

題型	示例題	比例式	解題策略
A-II	Q03：3 個檸檬賣 120 元，那 6 個檸檬賣多少元？	$\frac{120}{3} = \frac{x}{6}$	1.先算出單價 2.累加策略 3.內項乘積等於外項乘積
B-NI	Q19：迦旭有一張長 36 英吋、寬 24 英吋的照片，他想要沖洗一張寬 6 英吋的縮小版，試問縮小版的照片長度有多長？	$\frac{36}{24} = \frac{x}{6}$	1.利用長寬比 2.約分到最簡分數，再擴分 3. $24 \div 6 = 4$ ， $x = 36 \div 4 = 9$ 4.內項乘積等於外項乘積

表 4
四種解未知數之題型示例題(續)

題型	示例題	比例式	解題策略
C-IN	Q18: 義傑 7 趟路程共走了 14 公里。如果 10 趟路程，義傑將會走了多少公里？	$\frac{14}{7} = \frac{x}{10}$	<ol style="list-style-type: none"> 1. 計算一趟多少公里 2. $14=7 \times 2$, $x=10 \times 2=20$ 3. 內項乘積等於外項乘積
D-NN	Q07: 等式 $\frac{5}{4} = \frac{7}{x}$ 中的「 x 」值為多少？	$\frac{5}{4} = \frac{7}{x}$ $\therefore 5x = 4 \times 7$ $= 28$ $\therefore x = \frac{28}{5}$	<ol style="list-style-type: none"> 1. 內項乘積等於外項乘積 2. 因為分數間為非整數，所以無法用累加的方式去運算，只能用非整數的倍數關係。 3. 涉及到分數的乘除，難度較高。

備註：示例題選自 PRA；解題策略參考 Lobato 等人(2010)

Lesh, Post, and Behr(1988)提出比例推理是中學數學相當重要的學習內容，學生能利用兩個數量之間建立乘法關係，進而建立比例式，區辨兩個數量之間是否有比例關係，且能應用在生活情境中進行解題(Lobato et al., 2010)，大型國際評比如 TIMSS 及 PISA 均將比例推理視為關鍵的數學能力之一。根據 Lobato 等人(2010)整理學生進行比例推理解未知數的題型時，最常使用的解題策略有累加、倍數乘法、約分或擴分及單位分數等，或者是使用「內項乘積等於外項乘積」來計算出比例式中的未知數，就以 PRA 有 14 題解未知數的題型，根據如圖 2 所呈現兩個比值的乘法關係，可分為比例式的比值內(within)是整數(Integer, I)或非整數(Non-Integer, N)，兩個比值間(between)的關係是整數(Integer, I)或非整數(Non-Integer, N)，會有 4 種未知數的題型，如表 4 所示，分為 A、B、C、D 四種題型，根據示例題列出其比例式，且列出學生可能的解題策略。

如以 Riehl 和 Steinhorsdottir (2019)的研究指出 A 題型最簡單，而 D 題型對學生來說相對困難，其中 B 和 C 題型就要看學生對於比例推理的特質之精熟程度而定，隨著學生對於比例推理的數學概念有深入的了解，即使題型的困難度增加，學生能使用的解題策略隨之越多(Lobato, Ellis, & Zbiek, 2010; Steinhorsdottir & Sriraman, 2009)。加上過往 CDMs 研究以不同模式去分析學生組型(profile)或特質的精熟程度，較少研究去分析學生組型與題型之答題表現。因此本研究，從 PRA 的 14 題未知數題目中所量測 6 個特質中去了解各學生組型的期望分數，來建構學生在比例推理之解未知數題型的學習路徑，另外了解各學生組型在這 4 種題型的答題表現，進而探討出關鍵特質影響其答題表現，提供教師進行調整教學之參考，也提升學生之學習表現。

綜觀上述，採用 PRA 中 14 題比例式解未知數題型所量測的特質，以未結構化及發散兩種特質階層模型，利用 DINA-AHM 和 HDCM 兩種模式進行研究，了解學生組型之分布情況及各題之期望值，並計算其該測驗的觀察分數(答對題數)，畫出哈斯圖，學生組型之答對題數的差距與平均 CSEM 進行比較，建構其比例推理

解未知數之學習路徑；另外利用根據 4 種比例式解未知數題型去分析學生組型之答題表現，欲找出那一個關鍵特質會影響 4 種題型之答題表現，提供給教師作為參考。

參、研究方法

本研究欲利用 CDMs 及 PRA 來探討台灣學生在比例推理解題策略之學習路徑，使用工具、研究對象、資料分析分述如下。

一、比例推理測驗

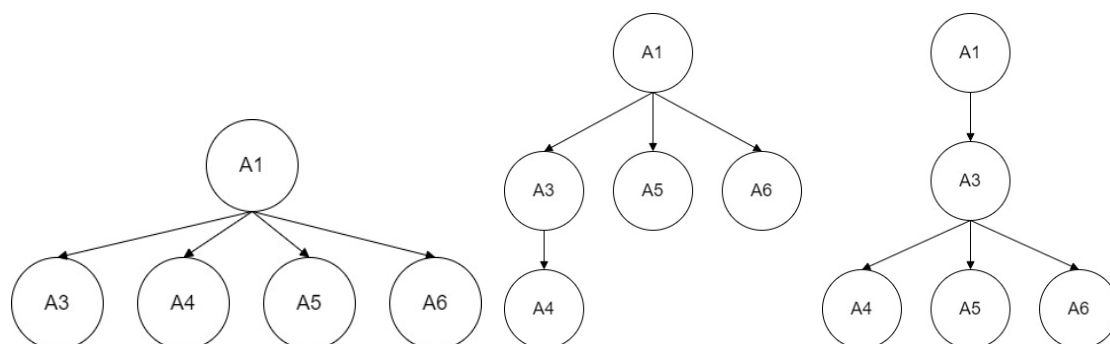
此測驗主要用於瞭解國中生在比例推理的技能(Skill)，因此結合數學研究人員、數學教育學者、中學數學科教師以及心理計量的研究人員根據認知診斷評量的編製原則，完成這份測驗(Tjoe & de la Torre, 2014)。本份測驗量測 8 個關於比例推理的技能，這裡稱之為特質(attribute)，共有 31 題，本測驗採選擇題四選一的題型。本研究針對解未知數的題型進行研究，PRA 有 14 題是屬於此類，其中 A(II)型、B(NI)型、C(IN)型各有 4 題，D(NN)型有 2 題；測量的特質有 A1、A3a、A3b、A4、A5 及 A6 等 6 個，考量題數及階層模型的關係，將其 A3a 及 A3b 兩個特質合併成 A3，特質數改為 5 個。

二、特質階層模型

採未結構化及發散模型，根據 PRA 的特質描述，A1 是比例推理的先備知識，若以非結構化模型如圖 5 進行研究，學生組型可能會有 17 種情況；發散模型採兩種模型，第一種 A1 為 A3、A5 和 A6 之先備特質，另 A3 是 A4 的先備特質，即要先精熟 A3，才會精熟 A4，此階層模型之可能學生組型數為 13 種；其二是除 A1 是先備知識，A3 為 A4、A5 和 A6 為先備特質，此模型之學生組型數為 10 個。欲利用三種特質階層模型去分析，以此建構學生的學習路徑是否符應特質之階層關係。

圖 5

三種特質階層模型



備註：由左而右分別為未結構化、發散模型 1、發散模型 2

三、研究對象

比例推理是七年級第二學期的學習內容，因此本研究從 2021 年五月份寄發測

驗試題至有意願參與此研究的班級，共 11 間學校，公立國中有 9 間，1 間私立中學(1 班)，1 間附屬中學(1 班)。七年級的班級數 14 班，其中台南佔 8 班，其餘班級分布於雲林、屏東及高雄等地區，小計 380 人；八年級有 9 班，其中宜蘭佔 4 班，其餘班級分布於高雄、彰化、屏東，小計 234 人，合計 614 筆資料，其中女學生佔 46.58%。

四、認知診斷部分

利用 R 的 GDINA 套件，利用三種不同階層模型進行 DINA-AHM 及 HDCM 兩種模式進行分析，根據以下結果討論：

- (一) 階層一致性指標 HCI：判別預設特質階層模型與學生作答反應一致性，可知那一個階層模型與實徵資料較為適合。
- (二) 適配度指標：比較六組之 deviance、AIC、BIC 進行比較。
- (三) 信度：觀察者分數信度(Yamaguchi & Templin, 2022)， ρ_{DCM} 了解學生組型與觀察分數之間的變異數比例進行計算。
- (四) 透過 DINA-AHM 及 HDCM 估計學生組型之比例、每種學生組型之測驗期望值作為觀察分數及 CSEM。
- (五) 依據 DINA-AHM 及 HDCM 兩種不同模式下之各學生組型之答對題數差距與其 CSEM 之平均進行比較，建構其比例式解未知數的學習路徑。
- (六) 依據 PRA 中解未知數的題型分成 4 種題型 A(II)型、B(NI)型、C(IN)及 D(NN)，分別依照 DINA-AHM 及 HDCM 兩種模式估計各學生組型之答對題數進行討論，了解各學生組型與不同題型之答題表現，討論哪一個特質會讓答題表現有顯著增加，即為差距大於平均 CSEM。

肆、研究結果

一、設定特質階層是否符應學生的學習路徑

- (一) 檢驗學生作答反應與預先設定的三個階層是否一致？如利用學生作答反應與三個預設的階層模型進行比較，可以整體平均均達.50 以上，2009 年 Cui 與 Leighton 之標準為 $HCI_{平均} > .80$ 非常適配，介於.60 至.80 之間是良好適配，而 $HCI_{平均} < .60$ 為適配不佳，學生作答反應與預設的三個階層模型均未達適配標準；整份測驗為 14 題，將 614 位學生進行 K-means 方法依據答對題數分成三組，其中高分組的學生有 383 位，且與三個階層模型的 HCI 均達.80 以上，代表預先設定的模型與高分組達到非常適配，結果如同 Roberts 等人(2014)的研究發現高分組的學生與專家設定模型達到非常適配，代表專家學者設定的模型是根據學習內容及其脈絡進行設定，而高分組的學生較能掌握學習內容，故能進行解題上能按照預設的階層模型進行學習及發展；反之其他學生對於部分學習內容並未達精熟，在進行解題時會根據試題內容及所學到的學習內容進行連結，或直接猜測，故 HCI 小於.60，達適配不良；另考慮中高分組的學生，其 HCI 之平均為.60，達適配良好，故此三種模型對於中高分組學生是一致性，可作為後續研究。

表 5
特質階層(AHM)與階層一致性指標(HCI)之對照表

學生表現	N	答對 平均 題數	AHM		
			未結構化 階層一致性指 標(HCI)	發散模型 1 階層一致性指 標(HCI)	發散模型 2 階層一致性指 標(HCI)
低分組	51	3.373	-.526	-.512	-.511
中分組	180	8.511	.173	.199	.201
高分組	383	12.635	.801	.814	.818
中高組	563	11.316	.600	.617	.621
全部	614	10.656	0.507	0.524	0.526

備註：中高組 563 人為中分組 180 人與高分組 383 人合併之。

(二) 利用三種階層模型進行 DINA-AHM 與 HDCM 估計之適配情形：如表 6 所示，在未結構化及發散模型 1 下，僅有 DINA-AHM 的 BIC 低於 HDCM，其餘均是 HDCM 適配情形較佳；而在發散模型 2 下，HDCM 之適配指標均優於 DINA-AHM。而在觀察者信度來看，由於 HDCM 所需要估計試題參數的數量較 DINA-AHM 模型下的猜測參數與失誤參數來的多；另 HDCM 所估計學生組型之間的期望值分數變異較大，故觀察分數信度較高。

表 6
DINA-AHM 及 HDCM 之模型適配指標

	AHM					
	未結構化		發散模型 1		發散模型 2	
	DINA- AHM	HDCM	DINA- AHM	HDCM	DINA- AHM	HDCM
AIC	7589.041	7384.650	7582.627	7368.450	7577.650	7378.941
BIC	7783.520	7817.811	7759.430	7775.089	7741.189	7714.860
觀察分數 信度	0.791	0.849	0.790	0.848	0.787	0.842

(三) 三個階層模型之 DINA-AHM 與 HDCM 估計結果

1. 依照未結構化模型，其學生組型會有 17 種：

如表 7 比例推理測驗中解未知數的題目有 14 題，而 DINA-AHM 與 HDCM 估計之平均答對題數均為 10.656，而 CSEM 分別為 2.845 與 3.085；在學生組型的期望答對題數進行比較，因為 A1 為其他特質之先備條件，因此就 DINA-AHM 模型的全無[00000]和僅精熟 A1[10000]的學生之答對題數為 6.237(1.669)，各佔全部學生的 5.723%，而 HDCM 估計結果全無學生的答對題數為 3.476(1.498)、僅精熟 A1 的學生答對題數為 6.875(1.676)，符合多精熟一個特質，答對題數應較高；可見兩種模型下估計結果有所差異。

表 7

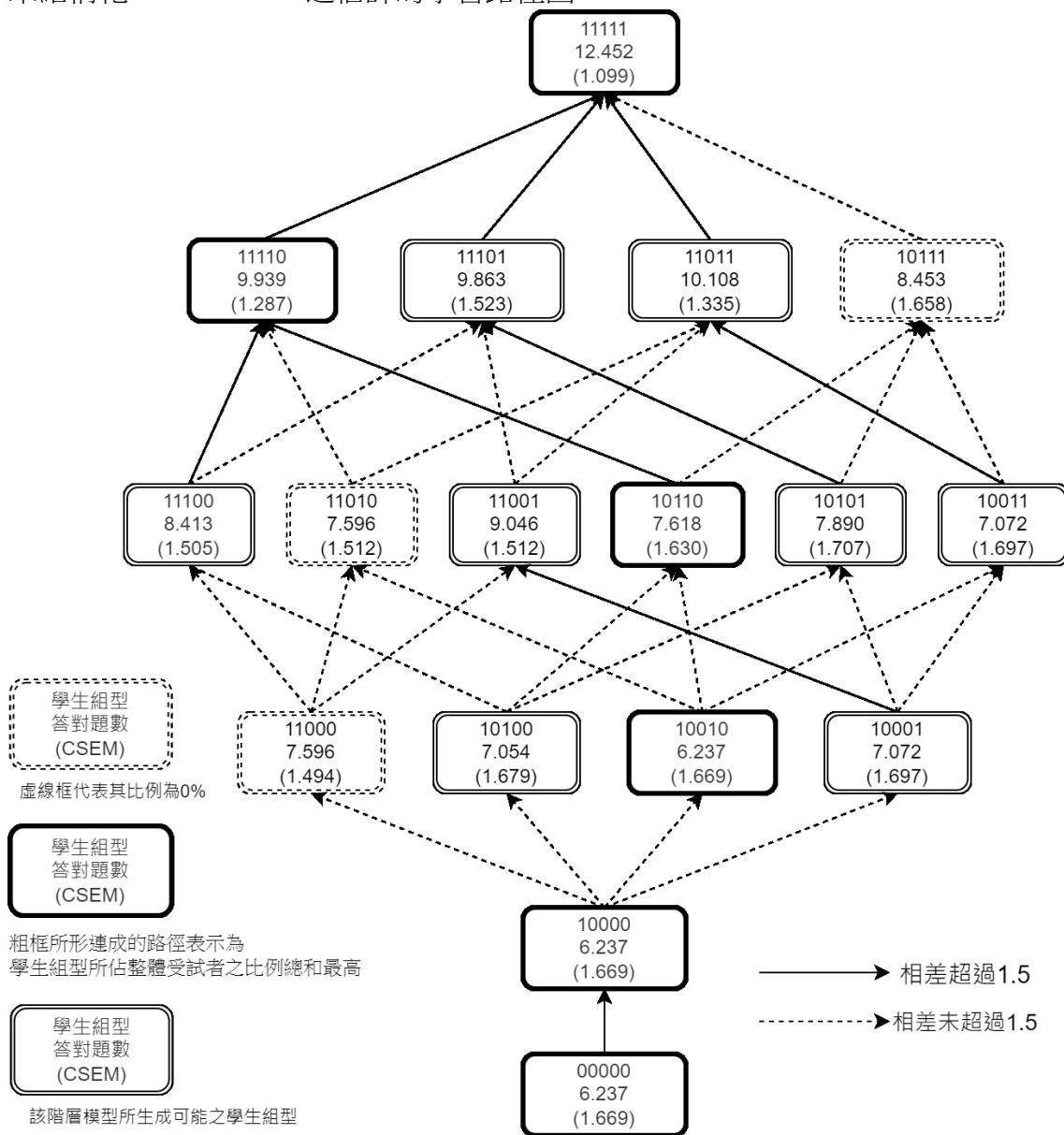
未結構化特質階層模型之 DINA-AHM 與 HDCM 之估計結果

DINA-AHM				HDCM			
學生組型	比例	答對題數	CSEM	學生組型	比例	答對題數	CSEM
00000	5.723%	6.237	1.669	00000	7.736%	3.476	1.498
10000	5.723%	6.237	1.669	10000	6.817%	6.875	1.676
10010	5.723%	6.237	1.669	10010	11.872%	8.963	1.536
10110	0.083%	7.618	1.630	11010	6.934%	10.662	1.203
11110	6.022%	9.939	1.287	11011	11.944%	12.045	1.015
11111	63.078%	12.452	1.099	11111	33.931%	13.223	0.815

備註：呈現主要路徑之學生組型的比例、答對題數及 CSEM。

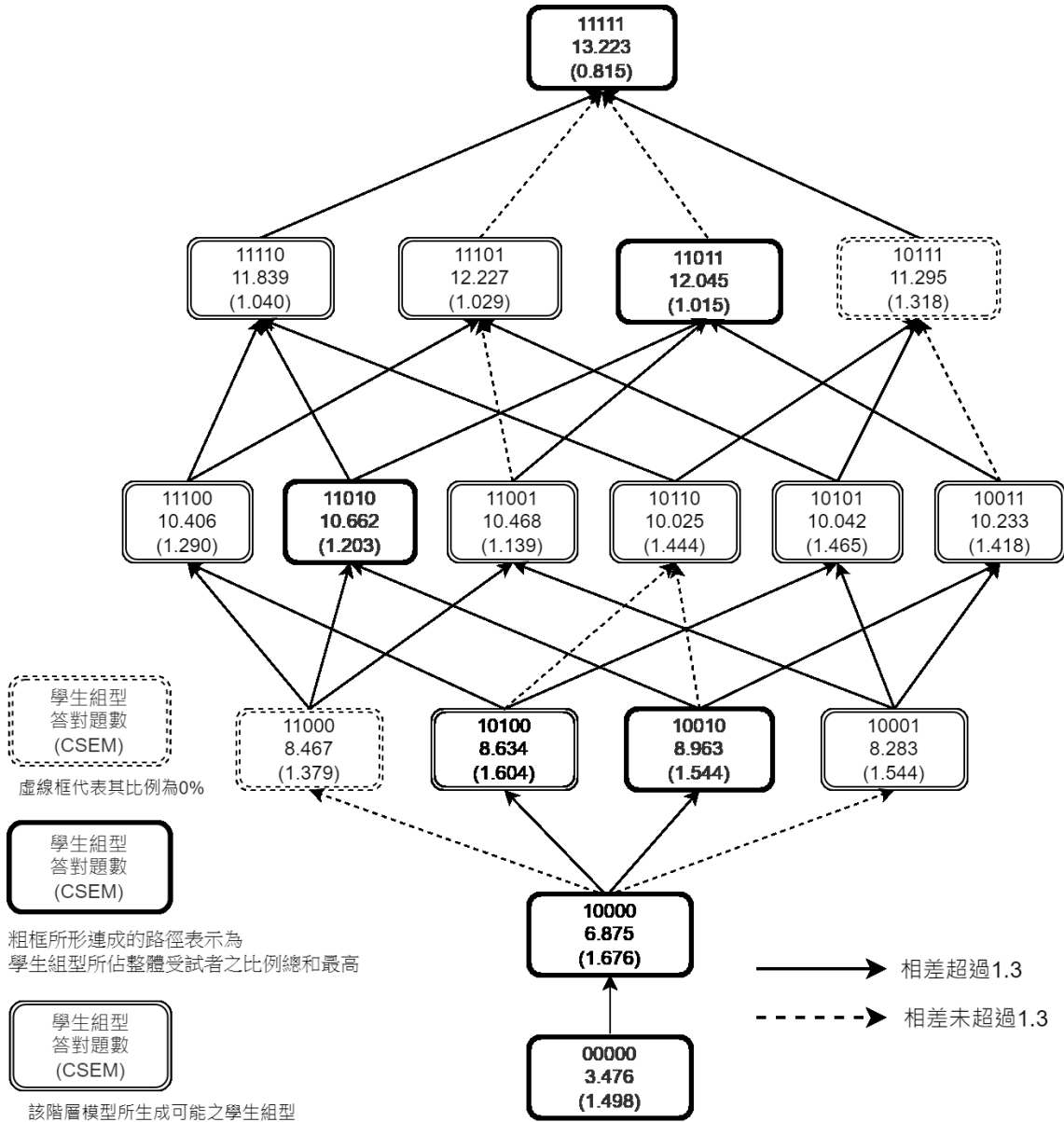
利用學生組型及期望答對題數繪製學習路徑，DINA-AHM 及 HDCM 之平均 CSEM 分別為 1.500 及 1.300，作為學生組型之答對題數的比較基準，若兩個組型之差超過平均 CSEM，代表有顯著的進步，以粗線標記，另外路徑連結各學生組型，其中粗框所表示為此路徑之學生所佔整體受試者之比例最高，作為主要學習路徑，如圖 6 所示，路徑為[00000]→[10000]→[10010]→[10110]→[11110]→[11111]，佔全部學生之 86.350%，依照此路徑可得學生精熟比例推理的先備知識 A1 後，學習 A5 區辨比例關係，前三種學生組型之答對題數均為 6.237(1.669)，而多精熟 A4 確定一組數字之間的乘法關係，即為內項乘積等於外項乘積之後，其答對題數略增 1.381，而精熟 A3 建立比與比值後，其答對題數增加至 9.939(1.287)，可見 A3 為關鍵特質，最後在精熟 A6 利用運算法則解題，全部特質均達精熟之答對題數可達 12.452(1.099)。

圖 6
未結構化 DINA-AHM 之估計的學習路徑圖



採用飽和模型 HDCM 進行估計，除了各特問之間的主要效果外，也會估計各特質之間的交互效果，依照圖 7 所示其主要學習路徑，未精熟任何特質之學生答對題數為 3.476(1.498)，而精熟 A1 先備知識後，增加至 6.875(1.676)，再學習到 A5 區辨比例關係後，答對題數增加至 8.963(1.554)，增加了 2.088，大於平均 CSEM 值，[00000]→[10000]→[10010]這個路徑與 DINA-AHM 之結果相同，然而不同的是，精熟 A1、A5 後，學生再精熟 A3 建立比值，可增加 1.699，達到 10.662(1.203)，然後再精熟 A6 運用算則解比例推理問題，可增加 1.383，最後精熟 A4，達到全部特質精熟，其答對題數增加至 13.233(0.815)，由 HDCM 所估計之主要學習路徑為 [00000]→[10000]→[10010]→[11010]→[11011]→[11111]，此路徑之學生組型佔整體之 79.234%。

圖 7
未結構化 HDCM 之估計的學習路徑圖



2. 用發散模型 1 進行估計：

與未結構化模型的差異為 A3 為 A4 之先備條件，代表學生要精熟建立比值這個特質，才會精熟 A4 確定一組數值之乘法關係，故學生組型有 13 種。如表 8 所示，利用 DINA-AHM 及 HDCM 估計之平均答對題數均為 10.656，其 CSEM 各為 2.845 及 3.085，如未結構化之結果相同，另學生組型[00000]、[10000]、[10010]在 DINA-AHM 所估計的答對題數均為 6.032(1.684)，而 HDCM 估計之答對題數各為 3.254(1.467)、6.545(1.669)及 8.554(1.609)，而兩種模型下，精熟 A1、A3 兩個特質的學生 11000 在佔整體受試者 614 位學生的比例為 0.000%及 0.302%，而在 DINA 模式下，全有[11111]之學生佔整體的 64.217%，反之 HDCM 精熟 4 個特質之學生，如[11110]佔 6.405%、11011 佔 19.685，以及全有[11111]佔 28.935，透過 HDCM 之估計，可以知道更細緻的資訊，作為教學及學習之參考。

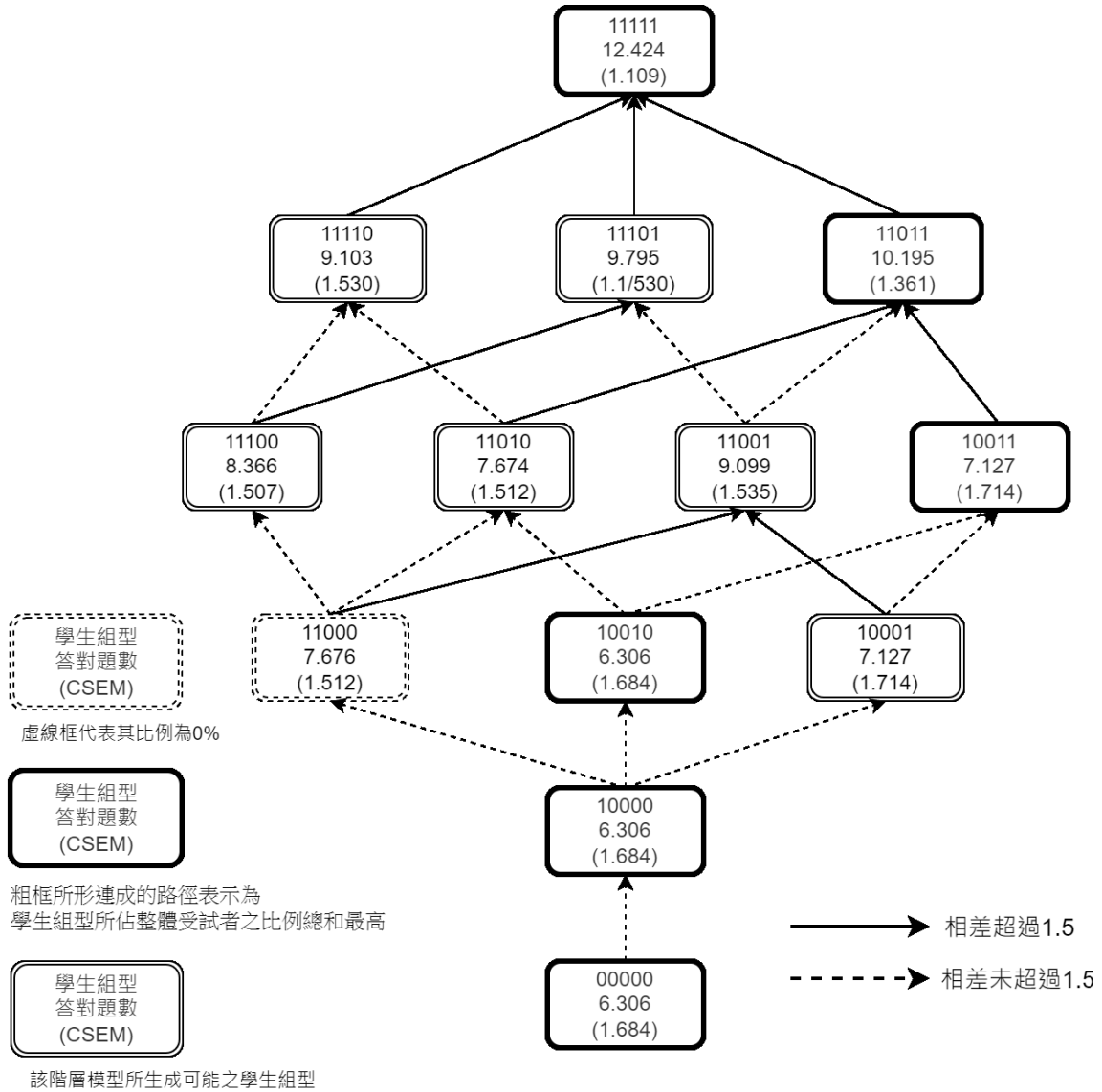
依照發散模型 1 及 DINA-AHM 估計學生組型及答對題數，平均 CSEM 為 1.526，因此若兩學生組型之間之答對題數差大於 1.500，利用粗線表示，另外以路徑所連結之學生組型佔整體比例最高者，為主要學習路徑，其結果如圖 8 所示，可以發現路徑為[00000]→[10000]→[10010]→[10011]→[11011]→[11111]，與之前未結構化 DINA-AHM 之結果不同，學生精熟特質的順序為 A1、A5、A6 及 A3，最後再 A4；以 DINA 模型去分析比例推理解未知數的題目，僅精熟 A1 或 A5 兩個特質的學生與全部不精熟的學生皆為 6.306(1.684)，表示並無法有效答對試題，若能精熟 A1 後，再精熟 A3 或 A6，其答對題數分別增加 1.370 及 0.821，若 A1、A3 及 A6 皆達精熟，即為學生組型 11001，其答對題數為 9.099(1.536)，為精熟 3 個特質的學生組型中答對題數最高者，而主要路徑之學生組型佔整體之 86.863%。

表 8
發散結構 1 之 DINA-AHM 與 HDCM 之估計結果

DINA-AHM				HDCM			
學生組型	比例	答對題數	CSEM	學生組型	比例	答對題數	CSEM
00000	6.491%	6.302	1.684	00000	6.963%	3.254	1.467
10000	6.491%	6.302	1.684	10000	5.201%	6.545	1.669
10010	6.491%	6.302	1.684	10010	15.449%	8.544	1.609
10011	2.395%	7.127	1.714	11010	9.486%	10.781	1.200
11011	0.778%	10.193	1.361	11011	19.685%	12.231	1.033
11111	64.217%	12.422	1.109	11111	28.935%	13.237	0.809

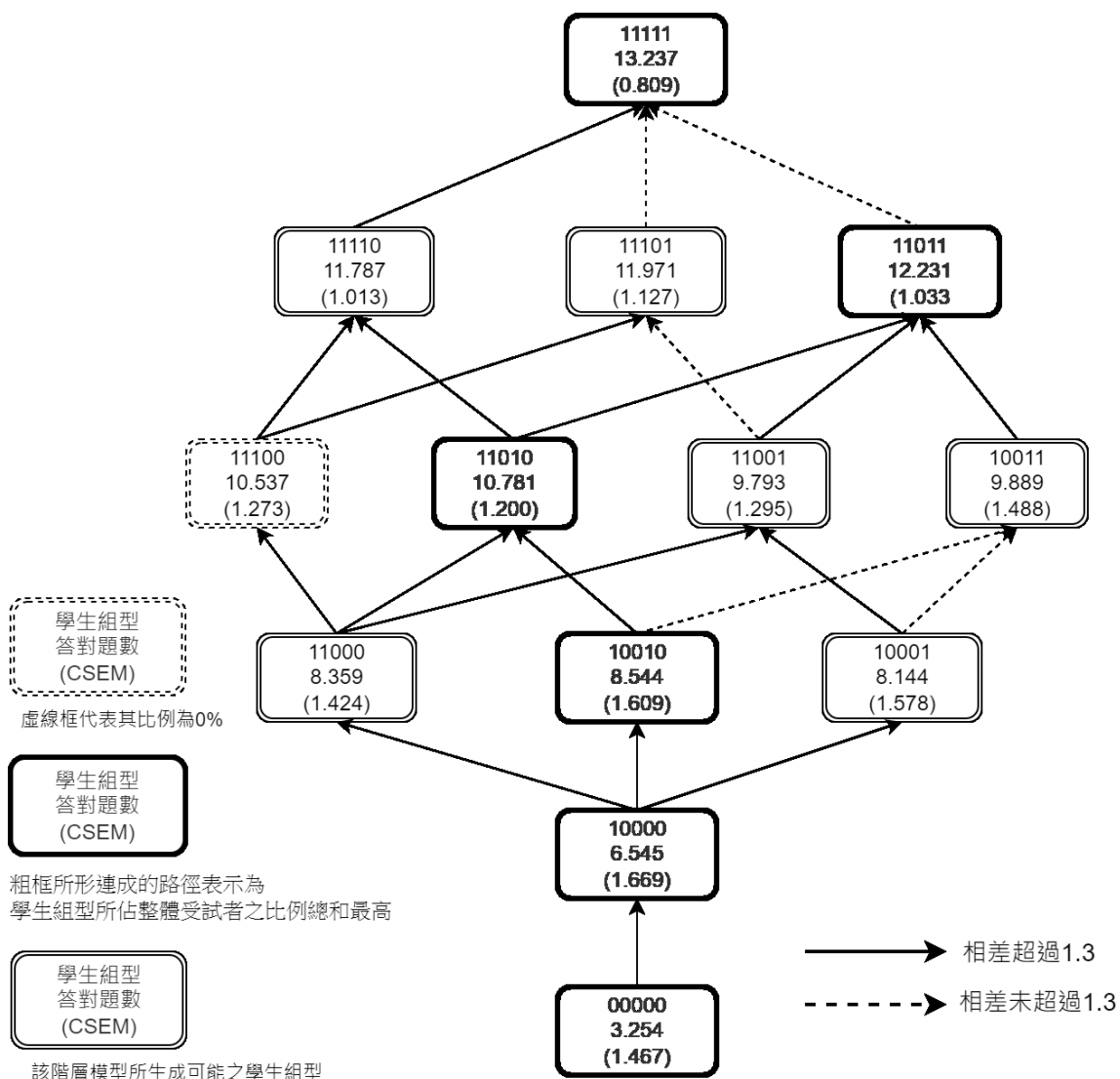
備註：呈現主要路徑之學生組型的比例、答對題數及 CSEM。

圖 8
發散模型 1 DINA-AHM 之估計的學習路徑圖



發散模型 1 改以 HDCM 進行估計，其學習路徑與未結構化之路徑相同，其特質精熟之順序為 A1、A5、A3 及 A6，最後才精熟 A4，其答對題數從全部不精熟的 2.004(1.227)增加至全部精熟的 13.237(0.809)，如圖 9 所示，精熟 A1 後，若能精熟 A3 或 A6 的話，答對題數分別增加 1.814 及 1.599，然而學生組型[11000]及 [10001]所佔比例為 0.302%和 2.890%，代表 614 位學生中，低分組的學生並未精熟 A3 或 A6，故僅能辨認其比例關係，並無法在解題上獲得高分，此結果是值得注意；以主要路徑來看，當學生組型[10010]變成[11010]，即為精熟 A3 特質，其答對題數增加 2.237，再精熟 A6，又再增加 1.45，最後精熟 A4，此學習路徑之連結得學生組型佔整體 85.719%。

圖 9
發散模型 1 HDCM 之估計的學習路徑圖



3. 預先設定的發散模型 2，與發散模型 1 的差異為 A3 是 A4、A5 及 A6 的先決條件，代表學生精熟 A3 後，才會精熟 A4、A5 及 A6，可能的學生組型共有 10 種，估計結果如表 9 所示；

表 5
發散模型 2 之 DINA-AHM 與 HDCM 之估計結果

DINA-AHM				HDCM			
學生組型	比例	答對題數	CSEM	學生組型	比例	答對題數	CSEM
00000	11.906%	6.386	1.721	00000	7.456%	3.352	1.484
10000	11.906%	6.386	1.721	10000	20.953%	7.885	1.719
11000	0.045%	7.774	1.554	11000	0.057%	9.373	1.496
11100	0.608%	8.470	1.550	11100	1.304%	10.479	1.356
11010	0.045%	7.774	1.554	11010	14.178%	10.901	1.226
11001	2.953%	9.016	1.541	11001	4.334%	10.676	1.333
11110	4.736%	10.022	1.341	11110	6.956%	11.921	1.058
11101	1.717%	9.712	1.537	11101	0.000%	11.782	1.174
11011	0.935%	10.146	1.366	11011	16.853%	12.265	1.001
11111	65.147%	12.395	1.117	11111	27.910%	13.285	0.788

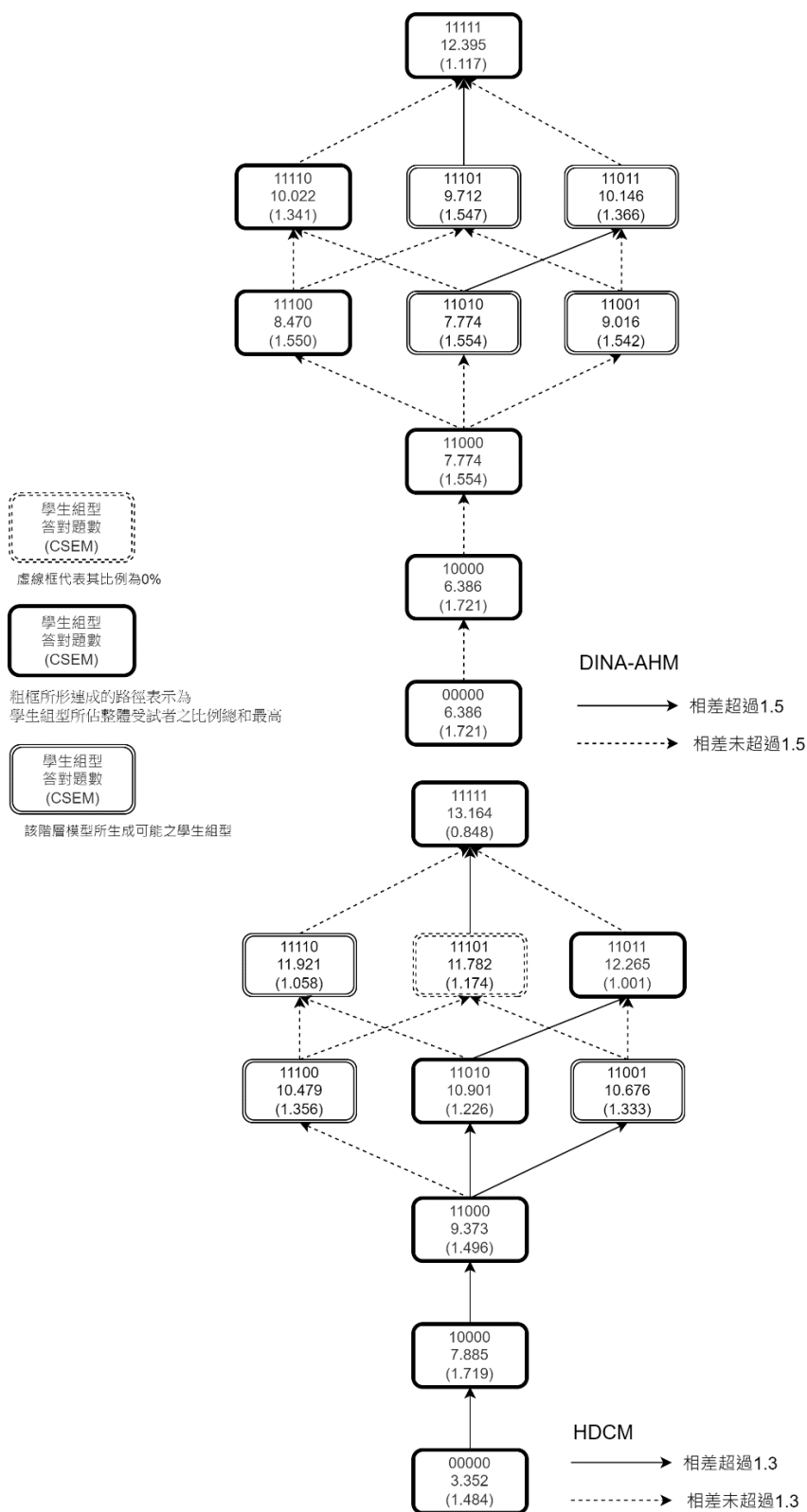
備註：主要路徑之學生組型的比例、答對題數及 CSEM 用粗體表示之

以 DINA-AHM 進行估計，可以發現學生組型[00000]、[10000]答對題數均為 6.386(1.721)，佔整體學生的 23.812%，以 DINA 估計來說，6.386 等於各題之猜測參數的總和，DINA 所估計的結果為全有或全無佔多數，全有[11111]佔 65.174%，故此三類學生即佔 88.986%；其餘的 11.014%的學生再進行分析，可知多精熟 A3 一個特質，而增加 1.388，再多精熟 A5 這個特質，其答對題數不變，為 7.774(1.554)，若是精熟 A3，則是增加至 8.470(1.550)，而比例最高的是精熟 A6 特質的[11001]有 2.950%，答對題數可增加至 9.016(1.541)；而從精熟 3 個特質之學生組型，變成精熟 4 個特質的路徑中：以[11010]→[11110]，答對題數增加 2.248，或是[11010]→[11011]，會增加 2.372，兩者均大於平均 CSEM，如前面討論，就以解未知數題型來說，精熟 A1、A3、A5 這三個特質，答對題數無明顯增加，而精熟 A3 或 A6，會有進步較多；而從四個特質變成全有[11111]的學生，其答對題數均大於平均 CSEM，均有明顯進步，故 [00000]→[10000]→[11000]→[11100]→[11110]→[11111]為主要學習路徑，佔整體受試學生之 94.33%。繪製的學習路徑如圖 10 所示。

以 HDCM 進行估計，可以看到全無[00000]佔全部的 7.456%，答對題數為 3.352(1.484)，而精熟 A1 特質的學生組型 10000 佔全部的 20.953%，其答對題數為 7.885(1.719)，兩學生組型之間的差大於平均 CSEM；而精熟 A3，增加至 9.373(1.496)，增加幅度也是大於平均 CSEM；精熟 A1 和 A3 後，再精熟 A4，其增加幅度是小於平均 CSEM，有別於精熟 A5、A6，尤其是精熟 A5 的學生組型 11010，佔整體學生之 14.178%，從[11010]→[11110]或者[11010]→[11011]，其答對題數分別增加至 11.921(1.058)及 12.265(1.001)，其中精熟 A1、A3、A5 及 A6 的學生組型[11011]佔受試學生的 16.853%，而此類學生多精熟 A4 特質，其答對題數增加 0.899，若是精熟 A1、A3、A5 及 A4 的學生組型[11110]多精熟 A6 特質，其答對題數增加至 13.164(1.058)，因此發散模型 2 以 HDCM 估計的結果來看，如圖 10 所示 [00000]→[10000]→[11000]→[11010]→[11011]→[11111]為主要學習路徑，佔整體

學生之 87.350%。

圖 10
發散模型 2 估計的學習路徑圖



備註：上為 DINA-AHM、下為 HDCM

綜合上述，利用因為 AHM 模型設定的不同，其學生組型的數量隨之減少，而

透過每一個學生組型的答對題數作為學習路徑的繪製，兩者差距大於平均 CSEM，用實線表示，從 6 個哈斯圖進行比較，可發現 HDCM 分析的結果，學習組型間較多用實線連結，代表多精熟一個特質，會增加其答對題數。另以連接起來的學生組型佔受試者之比例總和最高者作為主要學習路徑，從 6 條主要的路徑得知其精熟特質之順序整理如表 10，由於 AHMs 預先設定的關係，學生必先精熟 A1，代表學生須具備比例推理的先備知識，而未結構化及發散結構 1 的學習路徑中精熟 A5，後續發展的學習路徑若以 DINA-AHM 分析會先精熟 A6、再者是 A3、A4，而 HDCM 則會先精熟 A3 和 A6，最後才是 A4。發散結構 2 預定學生得先熟練 A1 和 A3，而 DINA-AHM 之結果顯示，特質精熟的順序為 A4、A5，最後是 A6，而 HDCM 的結果是 A5、A6，然後是 A4。

就以比例推理之特質的教學順序為 A3→A4→A5→A6，具備比例推理的先備知識的國中生學習建立比值、比例式、利用乘法關係解未知數，進而學習兩組數值之間是否有比例關係(正比、反比或非比例關係)，再運用算則解比例推理問題，然而從學生的學習路徑來看，學生精熟的特質與教學順序並未一致，依據發散結構 2 用 HDCM 估計之結果發現有助於提升答題表現，首要是 A1 及 A3，若能建構比值，可能採取的解題策略為單位分數、累加、擴分、約分等，然而學生應學會用乘法關係進行解題，卻是最後精熟的特質，此代表學生進行數學解題時，並無遵循其順序，遇到相對應的數學問題就使用熟悉的解題策略，故教學時應在教導學生進行比例推理的解題訓練，先是建立正確的比值或比例式，進而理解數量關係，除了內項乘積等於外項乘積之策略外，活用其他解題策略。

表 6

以不同階層模型估計其主要學習路徑的精熟特質之順序

AHMs	CDMs	學生組型連結主要學習路徑 精熟特質之順序
未結構化	DINA-AHM	[00000]→[10000]→[10010]→[10110]→[11110]→[11111] A1→A5→A4→A3→A6
	HDCM	[00000]→[10000]→[10010]→[11010]→[11011]→[11111] A1→A5→A3→A6→A4
發散結構 1	DINA-AHM	[00000]→[10000]→[10010]→[10011]→[11011]→[11111] A1→A5→A6→A3→A4
	HDCM	[00000]→[10000]→[10010]→[11010]→[11011]→[11111] A1→A5→A3→A6→A4
發散結構 2	DINA-AHM	[00000]→[10000]→[11000]→[11100]→[11110]→[11111] A1→A3→A4→A5→A6
	HDCM	[00000]→[10000]→[11000]→[11010]→[11011]→[11111] A1→A3→A5→A6→A4

二、各學生組型在四種題型上之答題表現是否有差異，是否能找出關鍵的特質提升學習。

解未知數題型的 A、B 和 C 題型各有 4 題和 D 題型有 2 題，共 14 題，根據上述發散模型 2 之階層一致性指標(HCI)較另外兩個模型高，因此採用 HDCM 進行估計，得知 10 種學生組型在 4 個題型之解題表現，詳見表 11。仿前面哈斯圖進行繪製，了解哪些特質有助於提升答題表現。

表 7

以發散結構 2 模型 HDCM 估計解未知數的 4 種題型之答對題數

學生組型	題型							
	A-II 4 題	CSEM	B-NI 4 題	CSEM	C-IN 4 題	CSEM	D-NN 2 題	CSEM
00000	1.476	0.871	0.966	0.822	0.620	0.721	0.291	0.498
10000	2.539	0.822	2.451	0.952	2.234	0.965	0.662	0.665
11000	3.209	0.616	3.248	0.695	2.251	0.966	0.665	0.666
11100	3.226	0.603	3.703	0.485	2.404	0.931	1.146	0.611
11010	3.226	0.603	3.311	0.669	3.700	0.498	0.665	0.666
11001	3.831	0.391	3.261	0.686	2.251	0.966	1.333	0.471
11110	3.226	0.603	3.827	0.396	3.721	0.476	1.146	0.611
11101	3.847	0.370	3.716	0.472	2.404	0.931	1.814	0.389
11011	3.847	0.370	3.324	0.659	3.761	0.458	1.333	0.471
11111	3.847	0.370	3.841	0.379	3.782	0.435	1.814	0.389

在 A 題型的答題表現如圖 11 之左上圖所示，學生組型[00000]、[10000]及[11000]的答對題數分別為 1.476、2.539 和 3.209，兩者之間的差距均大於其平均 CSEM(.562)，用實線表示，另外發現多精熟 A6，如[11000]→[11001]，[11010]→[11011]和[11110]→[11111]，均用實線表示其差距大於.562，代表精熟 A6 有助於提升 A 題型的答題表現，代表學生在 A-II 題型上，可活用累加、單位分數、約分或擴分的方式，未必僅使用乘法關係進行解題。B 題型的答題表現如圖 11 所示，可以發現跟上述結果相同，精熟 A1 和 A3 兩個特質之學生，答對題數達 3.248，可知兩個比值間是整數關係，可利用約分擴分進行計算，而[11011]之答對題數為 3.324，較[11100]、[11110]之答對題數略低，若能精熟 A4，使用乘法關係進行解題，則答對題數可達 3.703 以上。

在 C 題型之答題表現如圖 11 之左下圖所示，從未精熟任何特質到精熟 A1 和 A3、A4 的特質，答對題數從.620 至 2.404，若加上 A5，則答對題數增加至 3.721，另[11000]→[11010]、[11001]→[11011]、[11101]→[11111]均有相同的情況，代表學生必須從題目中辨識兩組數值的比例關係，因比值間不是整數關係，無法使用直接利用累加，必須利用其他策略，得先確認比值內的比例關係才能進行解題，表示解生活應用題，A5 辨識兩組數值的比例關係是重要特質。

PRA 的 D 題型有 2 題，如圖 11

之右下圖所示，[00000]、[10000]及[11000]各答對 0.291、0.662 及 0.665，除了精熟 A1、A3，若精熟 A6，其增加量大於平均 CSEM(.544)，由於這兩題均已列比例式直接解未知數，故可發現精熟 A5 的[11010]之答對題數僅.665，而精熟 A4 的[11100]可答對 1.146 題，代表此種題型利用內項乘積等於外項乘積進行解題。由此可知不同的學生組型在這 4 種題型的答題表現均有差異，即代表不同的題型使

用不同的解題策略。

圖 11

發散結構 2 的學生組型與 A-II 和 B-NI 答對題數的哈斯圖

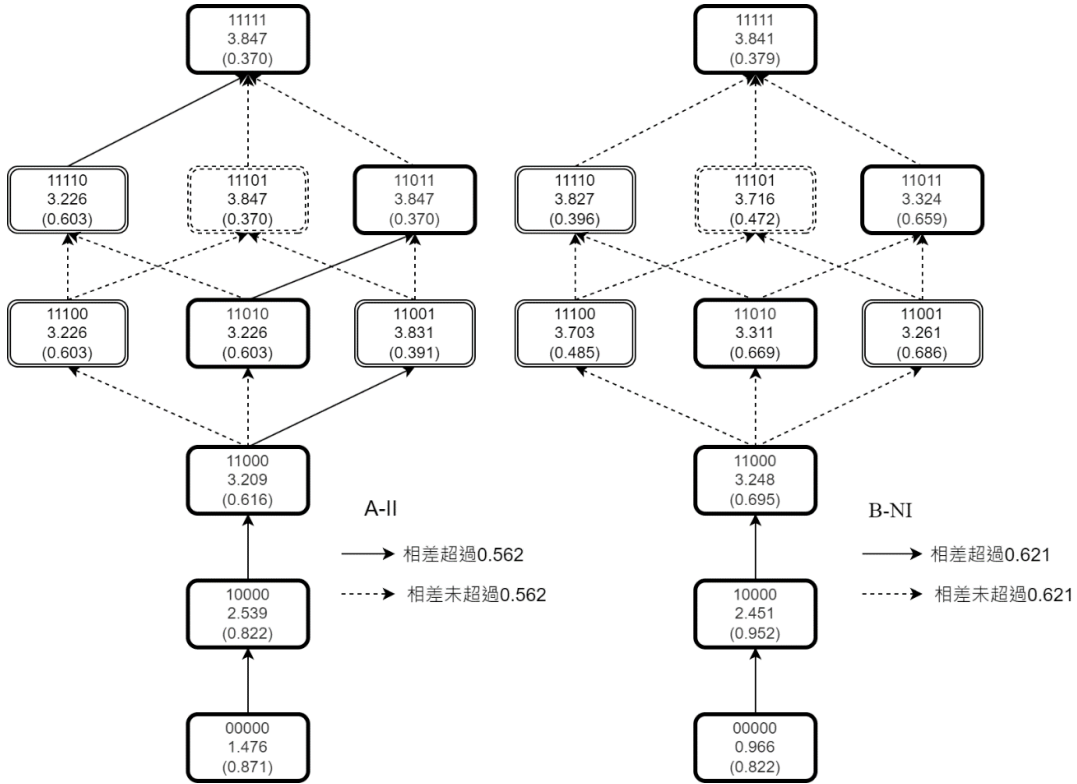
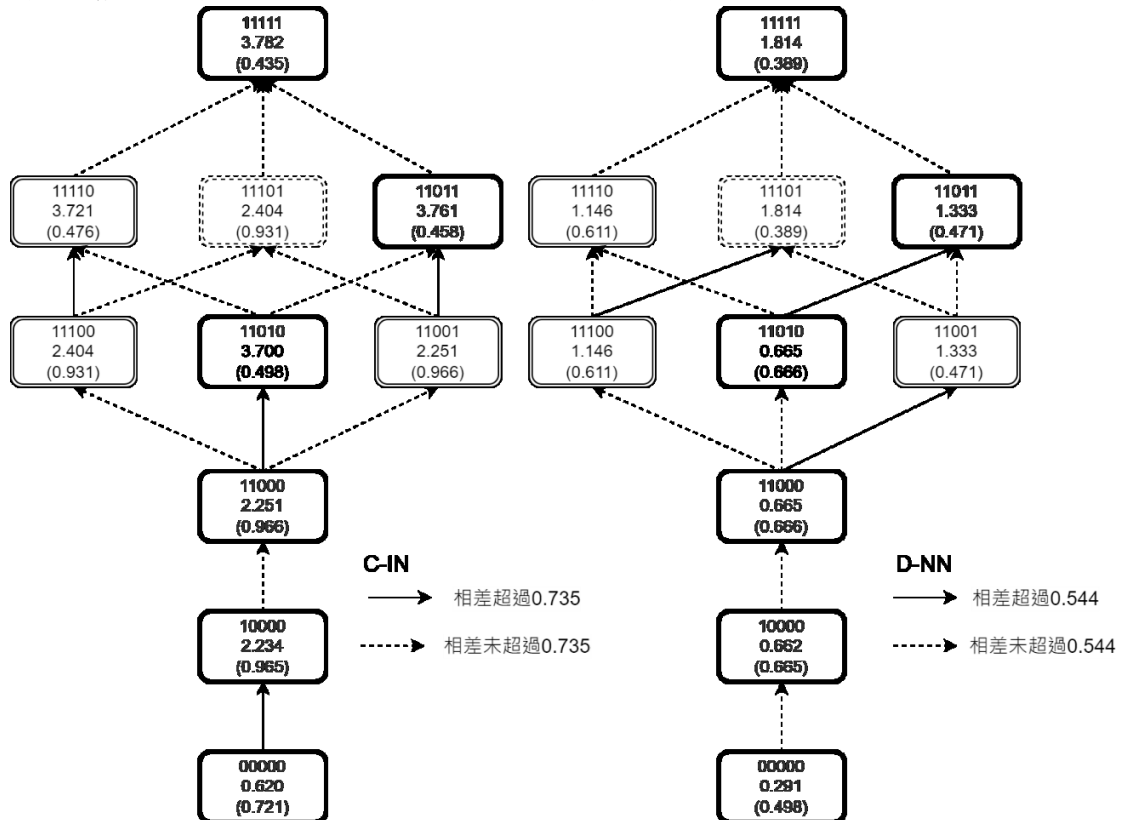


圖 12

發散結構 2 的學生組型與 C-IN 和 D-NN 答對題數的哈斯圖



伍、討論與建議

一、討論

利用 CDMs 估計學生在 PRA 解未知數題型的答題表現，去建構學生的學習路徑，了解預先設定的未結構化及兩種發散模型與學生的作答反應是否一致，其解果發現高分組學生之作答反應與所設定的 AHMs 達到非常適配，一方面可知道高分組學生精熟特質的順序與預設的結構吻合，另高分組的學生有按照教學的順序去掌握每個特質，反觀中低分組的同學，對於 PRA 所測量的特質並沒有全部精熟，因此解未知數題型時，會採取其他的策略、甚至僅是猜測作答，故與預訂的模型並不一致，而中高分組的學生在三種階層模型之平均 HCI 達.60，屬適配良好，亦代表這三種階層模型可作為後續研究使用。DINA 模式由於它的簡單易懂，為 CDMs 研究中最常使用的一種，伴隨特質階層模型進行估計每題的猜測及失誤參數，以此建構學習路徑來看，多精熟一個特質，增加的答對題數往往不會大於平均 CSEM；而飽和模式的 HDCM 估計學生組型的答對題數時，估計每一個特質的主要效果以及特質之間的交互效果，從中可得到更多且細緻的資訊，比較三個不同階層模型，發現其適配情況均較 DINA-AHM 佳，另外表 6 之觀察分數信度(Yamaguchi & Templin, 2022)進行比較 DINA-AHM 與 HDCM，可以發現 HDCM 之值亦較高，也代表各學習組型間的變異較大，由各學生組型的答對題數之間的差距了解，精熟哪幾個特質有助於提升學習表現。

如同研究結果，三種階層模型與 DINA-AHM、HDCM 分析之結果進行比較，發現 DINA-AHM 估計全有或全無的學生比例最高，加上各學生組型間之答對題數差距大於平均 CSEM 的情況較少，因此很難看出多哪一個特質可對答題表現有所助益；用 HDCM 進行估計，不精熟全部任一特質的學生之答對題數為 3.254~3.476，對照本研究採用的 14 題 4 選 1 的選擇題之答對期望值，加上單調性的限制，即為多一個特質，其答對題數會增加，因此可以看到每個學生組型是多精熟那個特質，以及各學生組型間答對題數的差距，繪製哈斯圖可知比例推理的學習路徑，以未結構化或發散模型 1 來看，多數學生可以精熟特質順序為 A1→A5→A3→A6→A4，若從先預設學生首先要先精熟 A1 和 A3 的發散模型來看，DINA-AHM 順序為 A1→A3→A4→A5→A6，而 HDCM 的順序則為 A1→A3→A5→A6→A4，與前面兩個階層較為類似。就以教學來說，章節安排順序為 A1→A3→A4→A5→A6，根據 PRA 所估計的學習路徑未依照教學順序而發展，較為合理的解釋是學生再進行解題時，所採取的策略是根據所精熟的特質進行，因此並未按照所謂的專家模型(線性)發展(Roduta Roberts et al., 2014)，這也是 CDMs 一開始的目的，透過學生作答反應進行估計，其結果提供個別化的學習資訊及細緻的診斷資訊，據此教師及學生都能透過評量及估計結果，在教師能精進其教學且促進學生學習的效果。

本研究根據比例式中比值內及比值間是整數或非整數的關係，將 PRA 的 14 題解未知數的題目分成 4 種題型：A-II、B-NI、C-IN 及 D-NN，比對學生組型的答對題數進行比較，與 Riehl 和 Steinhorsdottir(2019)研究結果一致，A、B 兩種題型較為容易，C、D 較為困難，而利用不同模型進行估計，發現由易至難的順序都相同，此外，利用 HDCM 估計在 4 個題型上每個學生組型的答對題數及其 CSEM，知道學生精熟哪些特質，可在那些題型上表現較好，如精熟 A1 和 A3 的學生即可在 A、B 兩題型，可獲得高分，而 C 題型上學生僅精熟 A1 和 A3 是不夠的，若多精熟 A5，答對題數增加量大於平均 CSEM，代表 A5 可提升 C 題型的答題表現，就以

解題策略來說，C 題型所列出的比例式，其比值內是整數，而比值間是非整數，因此學生在解題，並無法用累加的策略，需區辨這兩組數字的是否成比例關係，進而在利用比例關係進行解題；D 題型難度最高，這兩題的題目設計均為從已知的比例式求出未知數，等號兩邊的比值內或兩個比值間的關係是非整數，並非生活應用題，然而學生得利用運算法則進行解題，而 CDMs 估計的結果顯示精熟 A6 有助於提升其答對題數是相呼應的。透過 CDMs 分析，從答對題數或者是觀察分數去比較各學生組型精熟哪些特質，或是補足哪些特質有助於答題表現，亦或是從精熟特質中去分析學生解題所用的策略，這些診斷資訊可作為教學及學習補強之用。

二、建議

本研究利用比例推理測驗，設定三種不同的階層模型，利用兩種不同的 CDMs 模型：一是最多使用屬非補償型的 DINA 模型加上特質階層，稱之為 DINA-AHM，另一個是由 LCDM 加上 AHM 的 HDCM，估計受試者在每一題的答對率，作為觀察分數，以 PRA 中解未知數的題型有 14 題，觀察分數即為答對題數，透過 DINA-AHM 及 HDCM 估計各學生組型之答對題數及 CSEM，若每多精熟一個特質的學生組型可增加的答對題數是否大於平均 CSEM，評斷是否符合單調性的假設，即為每多精熟一個特質，其答對題數會增加些許，因此第一個建議是階層模型的設定，本研究使用研究工具原本並沒有設定特質之間階層關係，研究者為探討未知數題型的答題表現，依據文獻及教材編排所設定的三種不同的階層模型進行研究，雖在高分組的平均 HCI 達 .80 以上，代表此階層設定與學生作答反應是一致，然而對於中等程度的學生之平均 HCI 卻僅有 .20，低分組甚至達到 -.50，可見階層模型與中低能力學生的作答反應並不一致，較難得知他們學習上路徑的資訊，而本研究中高分組學生佔受試者的 62.38% 以上，因此根據階層模型所估計的學習路徑是有參考價值。另外從文獻裡都會有 Q 矩陣驗證的，根據作答反應與每一題之 Q 向量的設定是否合理，是否有調整的空間；如前面所提 Q 矩陣的設定由內容學者及編制測驗專家等共同討論而定，其階層模型的設定理應如此，也影響估計之結果。

本研究探討學生在比值內與比值間為整數或非整數的解未知數題型之答題表現，透過 CDMs 估計，得知學生組型所精熟的特質對於四種題型的答題表現有助益，或者是根據學生組型來了解學生如何解這四種題型，由於本研究僅收集到學生的作答反應，僅能從教學經驗或解題技巧去思考學生的解題策略，往後研究可建議不僅針對選項進行分析，也蒐集學生的計算過程，從中了解學生會使用那些解題策略，會提供更多資訊供教學及學習參考之用。

參考文獻

一、西文部份

- Cai, Y., Tu, D., & Ding, S. (2018). Theorems and methods of a complete Q matrix with attribute hierarchies under restricted Q-matrix design. *Frontiers in psychology*, 9, 1413. <https://doi.org/10.3389/fpsyg.2018.01413>
- Cui, Y., & Leighton, J. P. (2009). The hierarchy consistency index: Evaluating person fit for cognitive diagnostic assessment. *Journal of Educational Measurement*, 46(4), 429-449. <https://doi.org/10.1111/j.1745-3984.2009.00091.x>
- de La Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76(2),

- 179-199. <https://doi.org/10.1007/s11336-011-9214-8>
- de La Torre, J., Hong, Y., & Deng, W. (2010). Factors affecting the item parameter estimation and classification accuracy of the DINA model. *Journal of Educational Measurement*, 47(2), 227-249. <https://doi.org/10.1111/j.1745-3984.2010.00110.x>
- Duckworth, A. L., Quinn, P. D., & Seligman, M. E. (2009). Positive predictors of teacher effectiveness. *The Journal of Positive Psychology*, 4(6), 540-547. <https://doi.org/10.1080/17439760903157232>
- Gotwals, A. W., & Songer, N. B. (2010). Reasoning up and down a food chain: Using an assessment framework to investigate students' middle knowledge. *Science Education*, 94(2), 259-281. <https://doi.org/10.1002/sce.20368>
- Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of educational measurement*, 26(4), 301-321. <https://doi.org/10.1111/j.1745-3984.1989.tb00336.x>
- Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, 74, 191-210. <https://doi.org/10.1007/s11336-008-9089-5>
- Jia, B., Zhu, Z., & Gao, H. (2021). International comparative study of statistics learning trajectories based on PISA data on cognitive diagnostic models. *Frontiers in Psychology*, 12, 657858. <https://doi.org/10.3389/fpsyg.2021.657858>
- Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25(3), 258-272. <https://doi.org/10.1177/01466210122032064>
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of educational measurement*, 50(1), 1-73. <https://doi.org/10.1111/jedm.12000>
- Köhn, H.-F., & Chiu, C.-Y. (2019). Attribute hierarchy models in cognitive diagnosis: identifiability of the latent attribute space and conditions for completeness of the Q-matrix. *Journal of Classification*, 36(3), 541-565. <https://doi.org/10.1007/s00357-018-9278-6>
- 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. <https://doi.org/10.1111/j.1745-3984.2004.tb01163.x>
- Leighton, J., & Gierl, M. (2007). *Cognitive diagnostic assessment for education: Theory and applications*: Cambridge University Press. <https://doi.org/10.1017/cbo9780511611186>
- Lesh, R., Post, T. R., & Behr, M. (1988). Proportional reasoning. In M. Behr, & J. Hiebert (Eds.), *Number concepts and operations in the middle grades* (Vol. 2, pp. 93-118). National Council of Teachers of Mathematics, Lawrence Erlbaum Associates.
- Likun, H., TERZİ, R., & de la Torre, J. (2020). Wald test formulations in DIF detection of CDM data with the proportional reasoning test. *International Journal of Assessment Tools in Education*, 7(2), 145-158. <https://doi.org/10.21449/ijate.689752>
- Lobato, J., Ellis, A., & Zbiek, R. M. (2010). *Developing Essential Understanding of Ratios, Proportions, and Proportional Reasoning for Teaching Mathematics: Grades 6-8*. National Council of Teachers of Mathematics. 1906 Association Drive, Reston, VA 20191-1502.
- Ma, W., & de la Torre, J. (2020). GDINA: An R package for cognitive diagnosis modeling. *Journal of Statistical Software*, 93, 1-26. <https://doi.org/10.18637/jss.v093.i14>
- Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, 64, 187-212. <https://doi.org/10.1007/BF02294535>
- Riehl, S. M., & Steinhorsdottir, O. B. (2019). Missing-value proportion problems: The

- effects of number structure characteristics. *Investigations in Mathematics Learning*, 11(1), 56-68. <https://doi.org/10.1080/19477503.2017.1375361>
- Roduta, Roberts, M., Alves, C. B., Chu, M.-W., Thompson, M., Bahry, L. M., & Gotzmann, A. (2014) "Testing expert-based versus student-based cognitive models for a grade 3 diagnostic mathematics assessment." *Applied Measurement in Education* 27(3): 173-195. <https://doi.org/10.1080/08957347.2014.905787>
- Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. Guilford Press. ISBN: 978-1606235270
- Steinthorsdottir, O. B., & Sriraman, B. (2009). Icelandic 5th-grade girls' developmental trajectories in proportional reasoning. *Mathematics Education Research Journal*, 21(1), 6-30. <https://doi.org/10.1007/BF03217536>
- Steinthorsdottir, O. B., Sriraman, B., & Steinthorsdottir, O. B. (2007). Gender and strategy use in proportional situations: an Icelandic study. *Nordic Studies in Mathematics Education*, 12(3), 25-56. https://doi.org/10.1007/978-94-007-4978-8_38
- Tannenbaum, R. J. (2018). Validity aspects of score reporting. In *Score reporting research and applications* (pp. 9-18): Routledge. <https://doi.org/10.4324/9781351136501-2>
- Tatsuoka, K. (1996). Use of generalized person-fit indexes, zetas for statistical pattern classification. *Applied measurement in education*, 9(1), 65-75. https://doi.org/10.1207/s15324818ame0901_6
- Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of educational measurement*, 345-354. <https://doi.org/10.1111/j.1745-3984.1983.tb00212.x>
- Tatsuoka, K. K. (1995). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach. *Cognitively diagnostic assessment*, 327-359. <https://doi.org/10.4324/9780203052969-18>
- Tatsuoka, K. K. (2009). *Cognitive assessment: An introduction to the rule space method*: Routledge. <https://doi.org/10.4324/9780203883372-11>
- Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological methods*, 11(3), 287. <https://doi.org/10.1037/1082-989X.11.3.287>
- Templin, J., & Bradshaw, L. (2014). Hierarchical diagnostic classification models: A family of models for estimating and testing attribute hierarchies. *Psychometrika*, 79(2), 317-339. <https://doi.org/10.1007/s11336-013-9362-0>
- Tjoe, H., & de la Torre, J. (2014). The identification and validation process of proportional reasoning attributes: An application of a cognitive diagnosis modeling framework. *Mathematics Education Research Journal*, 26(2), 237-255. <https://doi.org/10.1007/s13394-013-0090-7>
- Yamaguchi, K., & Templin, J. (2022). Observed score reliability indices in diagnostic classification models. *Behaviormetrika*, 49(1), 47-68. <https://doi.org/10.1007/s41237-021-00153-9>

投稿日期：2023 年 04 月 12 日

修正日期：2023 年 09 月 10 日

接受日期：2023 年 09 月 22 日