多點計分認知網路評量模式

—以國小自然科植物概念為例

國立台中師範學院教育測驗統計研究所

施慶麟　劉湘川　許天維　鄭富森

摘　　要

本研究的主要目的在於以 Marshall 的「基模評量模式」為基礎，結合平行分配處理模式(PDP)理論以及無參數試題反應理論，，改進 Marshall 基模評量模式，建立新的「認知網路評量模式」，從而獲得較合理的個別受試者之知識結構圖，藉以改善目前教學評量的缺失。

本研究的主要結論有下列幾點；

一、認知網路評量模式以得分期望值來取代 Marshall 的評量模式中全有全無的假定，如此更能符合實際資料，進而提高本模式的可行性及適用範圍。

二、Marshall 的評量模式以傳統總分做為學生的能力指標，卻容易發生總分相同，但知識結構差異很大的情形。反觀認知網路評量模式以 IRT 能力值作為學生的能力指標，在分組診斷上獲得較佳的分組結果，使得能力值接近的學生有相似的知識結構，如此可改進 Marshall 評量模式的缺失，並且有助於教師在有限的時間內進行教學安置、教學診斷及補救教學的工作。

三、認知網路評量模式可涵蓋 Marshall 基模評量模式所提供的訊息，且以強度的方式呈現，不僅較為合理，且更能看出各能力組學生的認知結構及進展情形，因此可提供更多訊息。

四、認知網路評量模式可進行不同試卷間的等化，進而深入瞭解學生的知識結構或不同層次學生間的差異情形，並能有效預測學生知識結構的學習或進展情形，藉而獲得相關訊息。

壹、緒論

傳統紙筆測驗在實際教學的應用上，常因試題編製技巧不足，導致只能測得零碎的、記憶的知識，無法真正測得學生的能力，也無法正確有效的發揮測驗的診斷功能。

為此，本所推動的評量多元化在哲學、社會學及心理學等各方面上的理念，乃是採用所謂的「第四代評量」的新評量觀，主張在評量的過程中以協商為取向(negotiation－oriented)，透過整個評量社群不斷辯證協商，形成共識，進而改進自己。在辯證的過程中，必須收集所有關於學生學習狀況的資料作為辯證的材料，以便提供評量社群適當的回饋。

另一方面，隨著認知理論蓬勃發展，許多學者均致力於探索如何以較佳且較適當的方法，以便能確實掌握學生的內在思考歷程及能力。根據國內學者余民寧(民 84)指出，認知理論對心理計量學的貢獻有三：

一、認知理論對既存測量的分析，可增進我們對測量背後所表徵的理論架構有更進一步的認識。

二、認知理論能夠針對教育測驗所要測量的目標(如學習成就、內容領域知識等)，提供創新的測量策略和改善現有測量工具的缺失，使教育測驗能針對不良學習和教學調適提供更正確的診斷及有效的訊息。

三、認知理論引導性向、學習及成就理論之改良，意味著教育測驗將有全新的發展方向、需求以及實施型態。

綜合學者之論述以及近數十年的研究發展，可發現認知理論與心理計量學的結合已成為新一代評量的主要發展趨勢之一。

在評量模式方面，國外已有部分關於此種結合認知理論與測驗理論的研究，如：Nichols(1994)將此種評量模式稱為「認知診斷評量」(cognitively diagnostic assessment，簡稱 CDA)；Tatsuoka(1983, 1986, 1990)及 Tatsuoka & Tatsuoka(1987)所發展出的「規則空間」模式；Langley, Wogulis, & Ohlsson (1990)結合人工智慧與統計方法，發展出「自動認知模式化」(Automated Cognitive Modeler)系統以及 Marshall(1993)的「文字題解決者」(Story Problem Solver，簡稱 SPS)系統等，在國內部分，則有施淑娟(民 86)「認知網路評量模式及其實例應用之研究－以分數的加法單元為例」。

從 Marshall 的 SPS 模式中，我們可以發現其基模評量模式引領出教育評量的一種新方向，此種新的評量取向最大的優點在於融合認知網路表徵知識的方式與古典測驗理論。綜言

之，Marshall 模式是將評量的重點放在統整的知識而非零碎的知識，評量方式轉變為評估特定的基模是否已經發展、基模的關鍵元素是否呈現，而如果已發展，其發展的程度又為何？亦即，學習評量可再被重新定義為評量個人基模的形成程度與使用情形。因此他主張基模的評量方式必須經由兩個向度進行，一是評估基模內容－測量節點數以估計知識結構(網路)的大小，二是評估基模結構－測量連結數以估計知識結構的關連程度。換言之，他的評量模式是將傳統總分式的單一向度指標擴展至二向度指標，學生的實際能力應可被更有效地加以描述。然而根據劉湘川等人(劉湘川、許天維、施淑娟、郭伯臣，民 84)的研究，卻發現 Marshall 模式有幾項重要的缺失，造成分析上之困難，這些缺失除部分是因理論假設與實際情況不十分相符外，也有許多部分是因採用古典測驗理論的計分方式所致，因此本研究將採用試題反應理論來加以克服，並且因為 Marshall 是以平行分配處理(PDP)之認知網路模式為基礎來發展其理論，故本研究企圖改進二者，試圖發展一更具合理性、可行性與適切性之認知評量方式，此評量方式本研究稱之為「多點計分認知網路評量模式」。

貳、文獻探討

一、試題反應理論

過去有關測驗試題的分析，最為人熟知並廣泛採用的理論，是為「古典測驗理論」(Classical test theory；CTT)，該理論根據弱勢假設(weak assumption)而來，植基於 $X = T + E$ 的真實分數模式上，其中 X 為觀察分數，T 為真實分數，E 為誤差。此種一次線性數學式簡單易懂，計算方便，再加上發展歷史最久，故在教育與心理學界流傳甚廣，但也因其模式之過於簡化，產生了以下主要缺失與限制：(余民寧，民 80；余民寧，民 82；許擇基、劉長萱，民 81；Allen & Yen, 1979)

(一)試題難度、信度、鑑別度是一種樣本相依(sample dependent)的指標，會隨樣本之不同而改變，缺乏客觀公正性。

(二)相同功能的不同測驗間不易比較不同受試者之能力差異。

(三)理論中只假設唯一的測量標準誤(standard error of measurement)，未考慮到實際情況中受試者的個別差異。

(四)以單一總分來表示受試者能力，忽略反應組型(response pattern)，實際上，總分相同，反應組型不一定相同，能力估計值亦應有所不同。

(五)無法提供測驗分數等化、試題偏誤及適性測驗建立題庫等問題滿意解決之道。

針對古典測驗理論之缺失，於是有試題反應理論之興起。試題反應理論根據強勢假設(strong assumption)而來，具有嚴謹的數理基礎，而且目前已發展了相當多的模式與電腦程式，可應用於不同的情境。這些試題反應理論模式，若依據試題特徵曲線的呈現方式，則可將其分成「參數型 IRT」及「無參數 IRT」兩大部分，其中，「參數型 IRT」又可依其適用之計分型態，區分為雙點計分(dichotomous score)模式與非雙點計分模式。然限於篇幅，以下僅對無參數 IRT 的部分加以介紹。

(一)無參數IRT

針對古典測驗理論與早期參數型試題反應理論的優點與缺失，國內學者鄭富森提出一將傳統古典測驗理論融入現代試題反應理論中之新式理論－無參數試題反應理論。

1.模式假設

無參數試題反應模式認為受試者在試題上的反應分數主要受(1)受試者的特性(person)、(2)試題的特性(item)、(3)評分的特性(scorer) 等三方面因素的影響。以下依照這三方面特性來闡明無參數試題反應模式的理論基礎與基本假設(assumption)。

(1)受試者的特性

影響受試者答題行為的因素主要是來自受試者的內在能力特質與外在測量誤差兩方面，其中受試者的能力特質才是測驗的主體。因此處理受試者的特性或受試者能力特質的量化過程中，有三個基本假設：

假設 1：能夠以未知基本單位的等距量尺來描述受試者能力。

假設 2：在任何的母群體下，受試者的能力只分布於有限的範圍內；

假設 3：影響受試者在測驗上的表現能力具單維性。

(2)試題的特性

試題做量化分析的目的是以計量來描述受試者能力與試題特徵的關係，探討不同能力受試者在試題上得分期望值(expected score)，因此有如下的假設：

假設 4：能夠用一條直線來表達受試者能力與其試題的表現之間的數量關係。

(3)評分的特性

試題的計分結果，是由評分者主觀的依受試者對試題的反應，決定量化成數值。若以 f_{is} 代表評分者量化數值，以 x_{is} 代表 f_{is} 的觀察值，e_{is} 是測量誤差。則

$$x_{is} = f_{is} + e_{is}$$

假設 5：x_{is} 是標準化分數。即：

$$\varepsilon(x_{is}) = f_{is}$$

則由上述假設可推導出：

$$\varepsilon(e_{is}) = 0$$

其次，f_{is} 與 g_{is} 的關係可分為兩種情況：

1) f_{is} 與 g_{is} 呈現線性關係 (linear relationship)，以數學方式可以表示為：

$$f_{is} = \left(a_{f,i} g_{is} + b_{f,i}\right)$$

其中 $a_{f,i}, b_{f,i}$ 為常數，化簡後可得

$$f_{is} = \left(a_i \theta_s + b_i\right)$$

其中 a_i, b_i 為常數

2) f_{is} 與 g_{is} 呈現非線性關係(nonlinear relationship)。雖然無參數試題反應模式不以任何參數型式來描述 f_{is} 與 g_{is} 或 θ_s 的非線性關係，可以減少一些難以辨別與真實情境是否符合的假設，但如果此非線性關係沒有一對一且連續 (one to one and continuous) 的數學性質，則無法藉 f_{is} 的觀察值 g_{is} 估計 θ_s 值，因此需要如下的假設：

假設 6：f_{is} 與 g_{is} 的非線性關係是一對一且連續的函數。在此假設的情況下，以數學方式可表為：

$$f_{is} = a_{f,i} g_{is} + b_{f,i} + \xi_{is}$$

在上式中，ξ_{is} 代表主觀決定的評分方式所造成 f_{is} 和 $a_{f,i} g_{is} + b_{f,i}$ 之間的差距。

換言之，ξ_{is} 代表主觀決定的評分方式所造成的誤差。為了簡化模式，可進一步假設

$$f_{is} = a_i \theta_s + b_i + \xi_{is}$$

由此可知ξ_{is}為主觀決定評分方式所造成f_{is}與$a_i\theta_s+b_i$的差距，ξ_{is}為0時則表示f_{is}與θ_s呈現線性關係。因此，不管f_{is}與θ_s呈現線性關係與否，任何資料均可以表為：

$$x_{is}=a_i\theta_s+b_i+\xi_{is}+e_{is}$$

假設 7：假設e_{is}和f_{is}之間的相關為零，e_{is}和g_{is}之間的相關為零，e_{is}和θ_s之間的相關為零，e_{is}和ξ_{is}之間的相關為零。

假設 8：假設「若$(i,s)\neq(i',s')$，則e_{is}和$e_{i's'}$之間的相關為零」。

假設 9：既然評分方式是由主觀所決定，而且ξ_{is}主要亦由評分方式所造成，所以本模式很自然地假設ξ_{is}和θ_s之間的相關為零，或ξ_{is}和g_{is}之間的相關為零。而且若$i\neq i'$，則ξ_{is}和$\xi_{i's}$之間的相關為零。

2.能力值之估計

在實際取樣或測驗施測中所獲得的x_{is}都是經過一種不可避免的「粗略化」(coarseness)結果。在理論上，本模式將這種「粗略化」過程所產生的影響歸諸於測量誤差。換言之，這種「粗略化」過程將增加測量誤差e_{is}的變異程度(magnitude of variance)。由於有「粗略化」的過程，因此可以假設對任何θ而言，x_i是一個有多項分配的變數。假設在第i個試題的得分有$r_1,r_2,....,r_q$個不同的分數，同時令 $p_{i,k}$代表$x_i=r_k$的機率，則

$$\varepsilon(x_i)=f_i=r_1p_{i,1}+r_2p_{i,2}+...+r_qp_{i,q}$$

基於上述假設，無參數試題反應模式的估計過程，主要是以直接得到的觀察值x_{is}來估計θ_S能力值，再以θ_S值估計其得分期望值f_{is}，估計方式是以最小平方和公式來運算：

$$Q_0=\sum_{i=1}^{n}\sum_{s=1}^{N}(x_{is}-f_{is})^2$$

以上運算過程，學者鄭富森已發展出無參數試題反應模式的電腦程式(ICCNP)可供個人電腦中執行(鄭富森，民 83)。在應用方面，亦有多篇論文引用，證實其確為有效的學說。(簡茂發、劉湘川、許天維、郭伯臣，民 82；劉兆文，民 83；陳怡欣，民 83)。

(二)無參數試題反應理論與CTT及參數型IRT的差異

根據以上對於無參數試題反應理論的介紹，可以發現該理論與 CTT 及各種 IRT 模式之間的有以下差異。(鄭富森，民 83)

1.無參數 IRT 與 CTT 之差異

(1)刪除 parallel forms 和 τ-equivalent forms 的假設。

(2)增加主觀計分方法引起「量尺誤差 ξ」的假設。

(3)模式方程式是以每一個試題的得分為基礎，而不是以整個測驗總分為基礎。

(4)對於未知參數的估計方法是以嚴格的數理統計技術為基礎，進行推演而來的。

(5)對參數估計值的統計特性有明確的探討。

2.無參數 IRT 與參數型 IRT 之差異

(1)沒有局部獨立(local independent)的假設。

(2)不是以 ICC 代表受試在試題上得到某種分數的機率，而是以它代表受試在試題上的得分期望值。

(3)沒有假設 ICC 的參數形式，只假設 ICC 是一對一而且連續的函數。

(4)增加主觀計分方法引起「量尺誤差 ξ」的假設。

(5)對於未知參數的估計方法是以嚴格的數理統計技術為基礎推演而來的，沒有直接借助數值分析技術。

(6)參數估計值的統計特性也有明確的探討。他們不只具有不偏性和最小的標準誤的特性，而且也具有所有 MLE 的特性。

基於以上差異，並考量測驗試題之特性，本研究採用無參數試題反應理論來分析較為合宜。

二、平行分配處理模式

PDP 理論主要是由 Rumelhart, Hinton & McClelland 以及 PDP 研究群所共同提出的。根據 Rumelhart, Hinton & McClelland(1986)所述，PDP 理論最早的研究可回溯至 1970 年代，當時只是對於平行機制(parallel mechanism)與關聯記憶(associative memory)的研究，而直至 1982 年，Rumelhart、McClelland 及研究小組，決定將此一理論命名為平行分配處理(Parallel Distributed Processing，簡稱 PDP)，至此才正式出現 PDP 這個名詞。

根據 Anderson(1995)及 Rumelhart, Hinton & McClelland(1986)的研究，發現上述學者均主

張目前的認知記憶理論依其表徵方式，大致可分為兩類，一為局部表徵(localist representation)模式，另一為分配表徵(distributed representation)模式，前者認為大腦中的單一神經元即可對應一特定記憶，因此在網路中將每個節點當作一種概念，後者則認為記憶的記錄並不屬於任何單一的神經元，而是由一組神經元的激發類型來呈現，因此在網路中概念不是儲存在某一節點，而是分散在系統中的節點。

除了上述分配本質的假定外，一般心理學家對於 PDP 模式的一般性描述如下：

(一)記憶由一群交互連結的單元(unit)所組成，每個單元皆可賦予激發值。

(二)概念由激發組型(pattern of activation)加以表徵，單元間的連結可被加權(weight)，加權值決定一單元的激發對另一單元激發之影響，此值可為＋、－或0。其絕對值也有一些差異，值愈大表示較多的激發或抑制。

(三)訊息的輸出或輸入皆以向量加以表示，例如(1 −1 −1 1)，相似的訊息在向量中有較多相同的值。

(四)輸入向量的短期作用顯示於激發類型，長期作用則經由單元間連結加權值之改變。此假定可表示為下式：

新激發值＝輸入向量和加權值之運算＋舊激發值

(五)知識的分配有兩種意義，一指知識不儲存於單一位置，而是由網路加權值來編碼。二指同樣的加權值集合，可用來編碼多種連結，如此可促進部份訊息的使用及概念的一般化。

(六)加權值集合隨學習而改變，表示長期記憶對新學習的事物編碼。

此外，根據 Rumelhart, Hinton & McClelland (1986)的歸納，他們也指出 PDP 模式有下列 8 個主要的特點：

(一)具有一群處理單元

此為 PDP 的基本要件，這些處理單元常被稱為節點(nodes)或認知單元(cognitive units)。每一個處理單元的工作為接收相鄰單元所傳來的刺激或訊息，經過本身的轉換(或稱為「運算(computation)」)後再將轉換過後的刺激或訊息傳送至相鄰的單元，PDP 假設許多處理單元可同時進行各自的運算。在 PDP 模式中，處理單元大致可被分為：輸入單元(input unit)、隱藏單元(hidden unit)及輸出單元(output unit)。

(二)具某一激發狀態

PDP 中以一 N 維向量 $\mathbf{a}(t)$ 來表示此一系統在時間 t 的狀態，向量的第 i 個元素 $a_i(t)$ 表第 i 個單元(u_i)在時間 t 的狀態，激發值可以是連續的或離散的。單元能被激發至不同狀態，並非全有全無。

(三)單元的輸出

單元從其它鄰近單元接收信號，同樣地，也輸出信號給鄰近的單元，而輸出的強度與單元所處的激發狀態有關

(四)連結的型態

PDP 認為單元間的連結決定了系統了解什麼，以及系統如何對輸入產生反應，而單元的連結可分成二種，一為激發性連結(excitatory connection)，另一為抑制性連結(inhibitory connection)，這些連結形成系統之所知且決定了系統對何種輸入會產生何種反應。

(五)擴散規則

所謂擴散規則是用來描述訊息由系統內各單元輸出經由彼此間的連結傳送至某一單元的變化情形。

(六)激發規則

所謂激發規則係指有關單元如何處理所接受之輸入的規則，亦即單元如何將所接收的訊息，轉化成節點的不同激發狀態，並將此種激發能量與目前的激發狀態結合。

(七)學習法則

PDP 的學者也發展了些學習的原則，試著來解釋人類的「學習」、「遺忘」、「未發生學習」…等等，要解釋這些現象通常需要了解單元間連結的變化情形，連結的改變可分成以下三種情形：

1.發展新連結

2..中止既有的連結

3.連結強度的改變

通常 1.、2.二種情形可由 3.來解釋。

(八)整體系統的環境

根據 Grossberg(1980)、Konorski (1967)和 Martindale(1981)的觀點：節點可依性質的不同成分許多的層(layers)，在不同層間的節點的連結通常是具刺激性的(excitatory)，而在同一層的節點的連結通常是具抑制性的(inhibitory)。

以上簡略說明 PDP 理論的概念。本研究採用 PDP 理論中「線性閾值處理單元」的概念，說明訊息傳遞與處理單元之間的關係。然由於閾值的訂定方式眾說紛紜，並無一個確定的標準，所以本研究改進以往的閾值決定方式，改以強度代替之。

三、Marshall 的「可評量基模」評量模式

Marshall(1993)在不違背基模理論抽象化的定義與特性的原則下，參考 PDP 理論之假設，提出可經由基模結構與內容兩個向度的估計，將基模的定義格式化，以利於進行基模的評量。

(一)基模的內容

Marshall(1993)認為基模以共同主題組織知識而後儲存在記憶中，成為個人的經驗，它對個人的價值最主要在於問題的解決，當個體面對不熟悉的環境時，便會搜尋記憶觸接幾個基模，並從中提取和情境最適合的一個，以使情境變得有意義，進而決定採取什麼行動。因此，他分析基模的必要成分，發現基模應包含四種不同類型的知識，分述如下：

1.特徵辨識(feature recognition)成分：是指關於基模的陳述性知識及例子。

2.限制(constraint)成分：是指一組管理基模促動的規則，將問題情境直接與基模的原型連結。

3.計畫/目標設定(planning/goal-setting)成分：設定目標、子目標，及達成目標的方法規劃。

4.執行(execution)成分：完成計畫的行動及程序訊息。

(二)基模的結構

僅僅定義四種成分 (component)的存在並不足以描述一個基模，這些成分間還必須是互相關連的才可成為有用的知識結構。每一種類型的知識形成一個網路，但不一定和其它類型的知識網路連結，須視個體基模發展的完整程度而定，一個發展完全的基模在其成分網路中將具有良好的內部及外部連結性。因此估計連結的強度對於基模知識的評量來說，是相當重要的工作之一。

此外，Marshall 進一步引伸出對某一簡單領域的基模結構之假定，其中連結分為

幾種層次，一是基模之間互相連結形成統合一致的領域知識結構，二是在每一個基模之中成分間的連結，三則是每一成分中節點的連結，且在此一部份，Marshall 以平面圖形表示基模的結構。然而，為求能更加明確具體地描述某一領域的基模結構，研究者認為最好能採用階層式的圖形來呈現。

(三)基模的評量方式

為將基模概念引入教學評量之中，以上述 Marshall 的基模理論為基礎，我們若要測量個體所具有特定基模之品質，一方面必須檢查基模的連結，連結愈好，提取的訊息量就愈大，另一方面也必須檢視互相連結的知識是哪些成分的知識，因為相同的節點數與連結數，可能有多種不同的基模結構。

綜合以上所述，可知以 Marshall 理論進行基模的評量必須注意以下三個重點：

1.區分四種成分加以評量。

2.基模內容測量－測量存在於基模中的節點。

3.基模結構測量－測量存在於基模中的連結線。

除以上重點外，Marshall「可評量基模」理論之評量模式尚有以下假定：

1.圖形的範圍限制在學生已有的節點上。

2.連結無方向性。

3.連結線起點與終點不在同一點。

4.二節點間至多存在一條連結線。

5.測量時，答對(得 1 分)，則視為該試題所測量之節點或連結存在，答錯(得 0 分)，則否。

因此，若試題只包含一個概念，視為測量單一節點的存否，若包含二個概念，可視為測量兩節點的存否與連結，因為必須在二節點皆存在的情況下才能論及二點之間有無連結，依此類推，若試題包含三個概念，則可視為測量三節點的存否與三節點之連結。

本研究採用 Marshall 理論的觀點，在評量基模時，測量受試者知識結構中各節點的強度，以及節點間的連結強度。唯改進 Marshall 理論全有全無的假定，以 ICCNP 程式進行相關運算，求得各節點的得分期望值，以代表節點及連結的實際強度。

四、國小自然科植物概念之知識結構分析

(一)關於植物的研究

根據 J. Piaget 及其夥伴的研究指出，學生需發展至形式運思期(約十一歲之後)，才認為動、植物具有生命、是活的；Richards & Siegler(1984)的研究發現兒童約至 8-9 歲時，才能判斷植物是生物，是活的；Stephan(1985)的研究也發現，兒童需至 11 歲後，才能將樹、花和植物認為是有生命的。(引自裘維鈺，民 84)

Wax 等人(1987)曾以紙筆測驗以及晤談的方式，針對以色列六至十五歲的學生進行研究，發現在植物方面，六至十一歲的學生中，僅有 30-60％的學生知道植物是屬於生物的一員，十二至十五歲的學生雖答對率較高，但仍未達 80％。

Stead(1980)以中小學生為對象，探討植物概念，發現部分學生對植物範圍的認定相當狹隘，因為這些學生認為種子、雜草、蔬菜並不是植物，而是各自成為一類，顯示部分學生對於植物類別的概念並不清楚。

黃達三(民 82)曾對 72 位國小一至六年級的學生，以晤談的方式瞭解學生對植物的界定，發現學生多以能否自由移動作為界定植物的首要特徵，其次是具有營養器官(根、莖、葉)及需種在土裡這些特徵作為辨認的指標，而只有少數學生指出植物能行光合作用。

根據裘維鈺(民 84)的研究指出，多數的生物教師均認為，光合作用是學生在生物學上問題最多的一個單元。研究指出(Amir & Tamir, 1991, Roth et al, 1987)學生對於光合作用的學習困難，不在於光合作用是一個較抽象的概念，而在於學生對於食物、能量、氣體及生物等基本概念，有許多迷思概念，更遑論將這些概念連結起來。所以要對光合作用達到科學的瞭解是一個非常緩慢的過程，不論哪一個年級、年齡，對於光合作用都存在許多普遍性的迷思概念。

根據裘維鈺(民 84)的研究指出，一般學生對於光合作用典型的看法如下：

1.光合作用就是綠色植物和光。

2.光合作用就是植物接受二氧化碳，並將其變為氧氣。

3.光合作用就是讓植物保持綠色。

4.光合作用就是植物將陽光和二氧化碳變為葉綠體。

由上述可知，學生在植物概念與光合作用的學習上，存在著許多困難與問題，

本研究希望能藉由發展植物概念的認知網路評量模式，對學生現存的迷思概念加以補救，並進而改善學生的學習成就。

(二)我國國民小學自然科學課程植物概念的內容

目前我國國小自然科學的目標，主要在於指導兒童接近自然，瞭解人與其周圍的環境和諧共存之重要，增進學科知能與科學情趣，熟練科學方法，以養成具有科學素養的國民。且自 85 學年度起，我國已開始進行國小教材多元化的工作，各版本之新教材至 87 學年度止已發行至第五冊，然由於本研究分析之植物概念，相關內容分佈於國小一至六年級的教材中，為避免教材差異過大及發生重複或缺漏的情形，故本研究分析的教材將完全採用國立編譯館之版本，其中第一至第五冊採用新教材，而第六至第十二冊則由於尚未有新版教材，故採用改版前之舊教材。

茲將現行自然科學教材，各年級有關植物概念的單元名稱及其主要內容，整理如表一。

表一　現行國小自然科學課程植物概念相關單元分析表

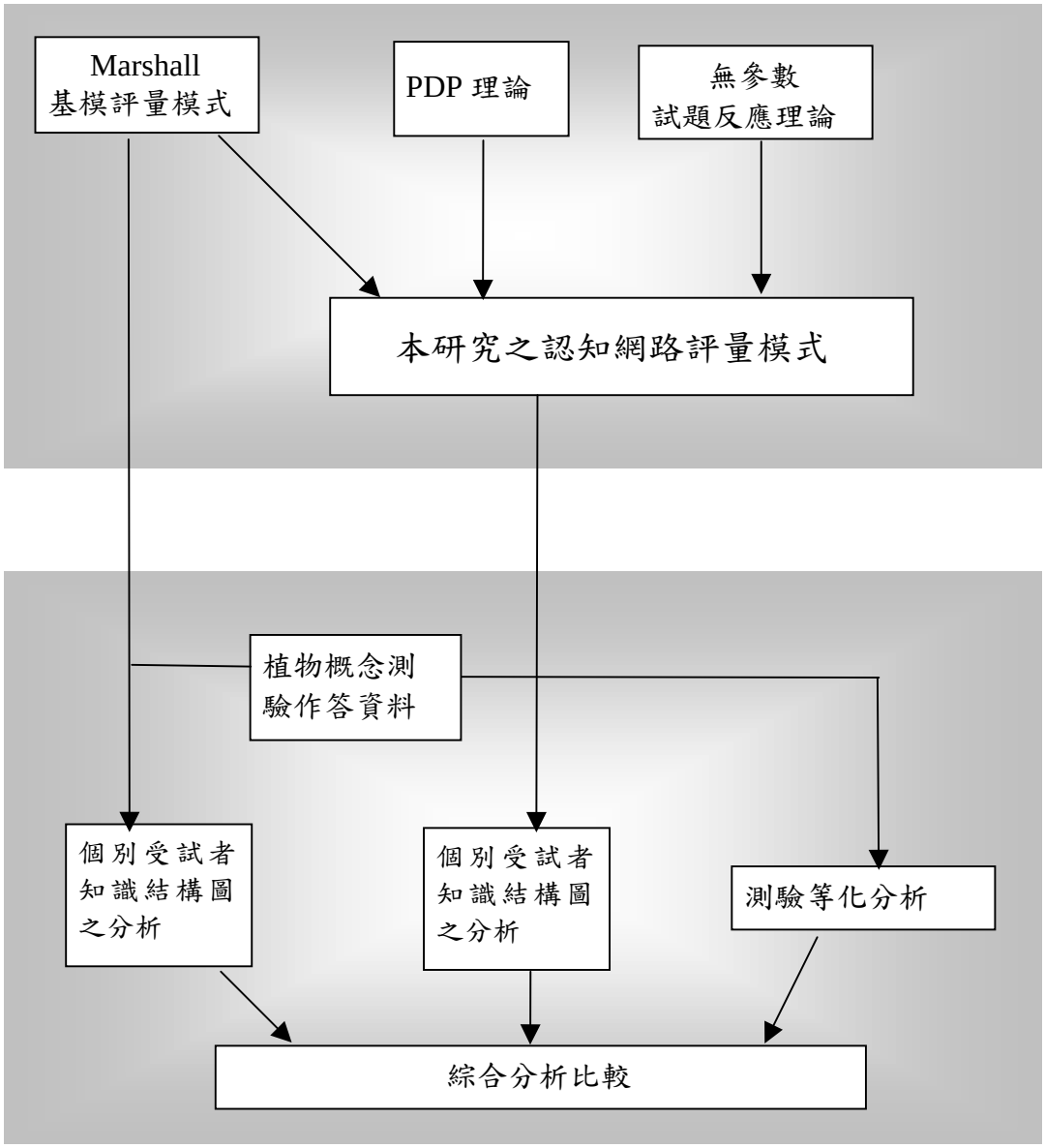
年級	冊別	單元名稱	主要內容
一	一	1、美麗的校園	介紹校園內的花的顏色、形狀、大小。
一	一	2、葉子	介紹葉子形狀、顏色、大小，並能加以比較。
一	二	2、豆子發芽	讓學生觀察豆子發芽的過程並加以記錄。
二	四	2、種花	藉由種花，讓學生體會陽光、空氣、水、土壤對植物生長的影響。
三	五	7、植物的身體	透過觀察校園植物，使學生認識植物身體的各個部位，並比較不同植物間的異同
四	七	3、植物的蒸散作用	讓兒童藉由實驗來觀察植物的蒸散作用，並瞭解影響蒸散作用的因素。
四	八	3、植物的繁殖	觀察不同植物繁殖的方式，並加以記錄。
五	九	4、細胞的觀察	觀察動植物的細胞，並比較其異同。
五	十	6、黴菌	介紹黴菌的構造，並瞭解影響黴菌生長的因素。
五	十	7、光合作用	介紹植物的光合作用，並藉由實驗來檢驗澱粉、氧氣、二氧化碳。
六	十一	1、種子的構造與發芽	觀察種子的構造並瞭解影響種子發芽的因素。
六	十一	3、自然界物質的循環	介紹自然界中水、空氣及礦物質的循環。

本研究將針對上述各單元進行知識結構分析，並據以編製本研究所需之試題。

參、研究設計

一、研究架構

本研究結合無參數試題反應理論、Marshall 的「可評量基模」理論及平行分配處理(PDP)理論，建立一個新的認知網路評量模式，並對現行國小自然科學教材進行分析，據以編製試題以及進行施測，作為本研究的實際資料。根據本研究之文獻及待答問題，將研究架構繪製如圖一。



圖一　本研究之研究架構

二、認知網路評量模式之建立

(一)PDP理論

本研究的部分理論依據乃是採用 PDP 理論中的線性閾值處理單元的概念，並根據教材分析的結果，將較基本的概念視為下層的處理單元，經激發後將激發值傳遞至上層的處理單元，當累積相當的激發值後，上層處理單元便會激發，亦即上層概念會出現；若是下層概念不足，激發值無法超過閾值，則上層概念將無法形成。

然而，由於決定閾值的方式眾說紛紜，莫衷一是，故本研究改進線性閾值處理單元之處在於，本研究並不決定閾值，而是依據受試者的實際受測情形，決定本模式中各處理單元間的連結強度，此一方式不僅較客觀，也較符合實際情形。

本研究將各概念及概念間的連結均以機率的方式表示，並以 ICCNP 程式分別估計之，如此可得到個別概念存在的機率以及各個連結激發的機率，將上述結果繪製成圖，如此便可獲得一個完整的認知診斷圖，藉以得知受試者的知識結構。

值得注意的是，本研究所得到的認知診斷圖，其每個節點均代表一個概念，而非一個試題，所以能更精確求得受試者的知識結構，進而達成認知診斷的功能。

(二)Marshall理論之缺失

本研究乃是以 Marshall 的基模評量模式為基礎，加以適度的修正。根據研究施淑娟(民 86)的研究指出，Marshall 評量方式在應用上有以下幾點缺失：

1.Marshall 假定節點的存在與否以及節點之間的連結為全有或全無，因此他使用二元計分(答對得 1 分，答錯得 0 分)的試題去測量節點的存否以及節點之間有無連結，然而，就認知心理學的觀點而言，概念節點的存在及節點之間的連結應有強弱之分，Martindale(1991)便曾在其「認知心理學－類神經網路取向」一書中提及，節點可呈現介於全有或全無的狀態，它可被部份促動，並非僅有全有全無兩種狀態。因為每一個節點都有一個抑制閾值，當輸入的刺激引起促動超越此閾值，則此節點才會完全促動。從日常生活中的實際情境，我們也可發現許多例子支持上述說法，故節點的存在與否以及節點之間的連結應並非如 Marshall 假定之全有或全無，而須呈現其強弱程度。

2.Marshall 雖將單一向度總分擴展至兩向度，但仍是以總分的型態來區分受試者能力的高低，因此在分析時發現同分組知識結構變異頗大(劉湘川、許天維、施淑娟、

郭伯臣，民 84)，增加診斷之複雜性與困難。

3.應用 Marshall 理論分析試題時，若有多個試題重複測量同一概念或同一概念連結，卻得到互相矛盾的結果(有的答對，有的答錯)(劉湘川、許天維、施淑娟、郭伯臣，民 84)，關於此種情形 Marshall 並未提出說明。

4.Marshall 假定連結無方向性，但從受試者的實際作答反應中，可發現某些節點之間的連結具有方向性(劉湘川、許天維、施淑娟、郭伯臣，民 84)，有關試題關聯結構的研究也顯示出某些試題概念間具有順序性的類似結果(劉湘川、許天維、郭伯臣、胡豐榮，民 83)。此外，PDP 的理論中也主張各個心理模組(如感覺、知覺、概念等等)皆為多層次系統(Martindale, 1991)，顯然不同層次間的節點應具方向性。因此 Marshall 的這項假定雖可簡化評量方式，但並不符合實際情況，有加以修正之必要。

(三)無參數試題反應理論與ICCNP

本研究將根據國內學者鄭富森所發展的無參數試題反應理論及其電腦程式 ICCNP 對施測資料進行分析，由於無參數試題反應理論的假設合乎實際施測的情況，並能真實反應受試者的能力，且程式在實際應用上沒有不合理的限制，故本研究在施測資料分析上，將採用無參數試題反應理論。

(四)教材分析

由於本研究的主題為國小自然科的植物概念，所以本研究將只專注於植物的相關教材內容，至於分析教材之版本，由於新教材僅出版至第五冊，故本研究選定國立編譯館自然科學課本第一、二、四、五、七、八、九、十以及十一等冊，其中第一至五冊為新教材，第七至十一冊為舊教材，主要的分析內容為各冊課本及習作。

在此需要說明的是，以往植物學家將黴菌視為植物，然而根據最新的分類方式，植物學家已將黴菌歸為真菌類，獨立於植物之外。但本研究認為在小學階段並無細分概念的必要，故仍將黴菌視為植物的一種，而納入本研究的研究範圍。

三、研究對象

本研究的主要研究對象為國小二、四、六年級學生，依據台灣省政府教育廳八十五學年度各國民小學學生數的統計資料，以中區國民小學二、四、六年級學生為母群體，分台中縣、台中市、彰化縣、南投縣四區，再依比例以隨機抽樣法抽取學校，被抽取學校之二、四、六年級學生各抽取一班接受施測。

四、研究工具

(一)植物概念測驗

本研究運用知識結構分析技術，將國小自然科學教材中關於植物概念的單元，進行知識結構分析，並據以編製植物概念測驗。

(二)SAS統計套裝軟體

本研究採用 SAS/GRAPH(6.12PC Windows 版)進行筆試資料的統計分析。

(三)ICCNP

此軟體為國內學者鄭富森(民 83)依「無參數試題反應理論」所發展的 PC 版軟體，本研究應用此軟體計算受試者總分、排序後的能力估計值以及試題特徵曲線等資料。

(四)EQUATICC

為李文忠、李慶峰(民 84)依無參數試題反應理論之等化模式所設計的試題特徵曲線等化程式，可用以計算出相同定錨試題間代表相同能力之等化迴歸方程式。

肆、研究結果與討論

以下將以三個部分來說明本研究之實例應用結果與討論，第一部分將分析二、四、六年級的受試者在植物概念測驗上的作答情形與等化後的相關訊息；第二部分探討本研究認知網路評量模式運用於國小學生自然科植物概念之結果，此部分僅抽樣五個能力群中共 5 位受試者進行分析；第三部分探討本研究認知網路評量模式與其他模式的優缺點，並說明本模式的可行性。

一、受試者施測資料之作答情形與相關分析

本研究抽樣的樣本是二、四、六年級，而施測的內容亦以該年級已教過的內容為主，故此，每一年級僅作該年級已教過之概念的試題。為獲得較多的訊息，本研究將三個年級的作答資料進行等化，將二、四年級的受試者能力值，等化至六年級的量尺上，以期所有受試者能在同一量尺上進行比較，也希望藉此獲得較多的訊息。

特別要加以說明的是，本研究所使用之植物概念測驗，在試題選項的安排上，是將所有試題的選項集中放在一個表格中(試題詳如附錄)，並把性質相同的選項安排在相鄰的位置，而非以往將選項放置於試題之後的形式，如此變成每一試題的選項群高達二十個以上，這樣的作法可有效的降低試題的猜對率。選項代碼以①、②、③、④、⑤等五個數字組合而成，如

此一來選項的個數便不限於五個。此種作答方式亦可應用至劃答案卡的考試上。

此外，在計分方式方面，本研究乃是採用三點計分：學生作答時如果選擇正確答案，則得 2 分；如未選擇正確答案，但選擇的選項是與正確答案性質相同者，得 1 分；其餘狀況則得 0 分。如此計分方式將可由得分看出受試者表現的差異，並能有效區分受試者的能力值。

本研究對於所有受試者能力進行相關分析，但以下僅呈現部分結果：

(一)各年級受試者在各概念上的比較

為便於在同一量尺上進行比較，所以本研究以下均採用等化後的能力值進行相關的分析工作，在此一前提下，應該摒除年段的區隔，僅就能力值加以分析，但由於本研究對於各年級在植物概念上的表現欲加以探討，故在資料呈現上仍以二、四、六年級的方式呈現。

根據圖二及表三可發現，四年級受試者在概念七、九及十一上的表現較差。根據表二，概念七是「果實的產生」，經查閱試題，發現試題是「花凋謝後，花的雌蕊部分就會生長成　　？」，之後經分析國小自然科第五冊第七單元「植物的身體」教材內容後，發現教材中確有提及並要學生觀察：「果實是從哪裡長出來的呢？」，由此推論學生對於果實的產生可能有學習上的困難或迷思的現象發生。

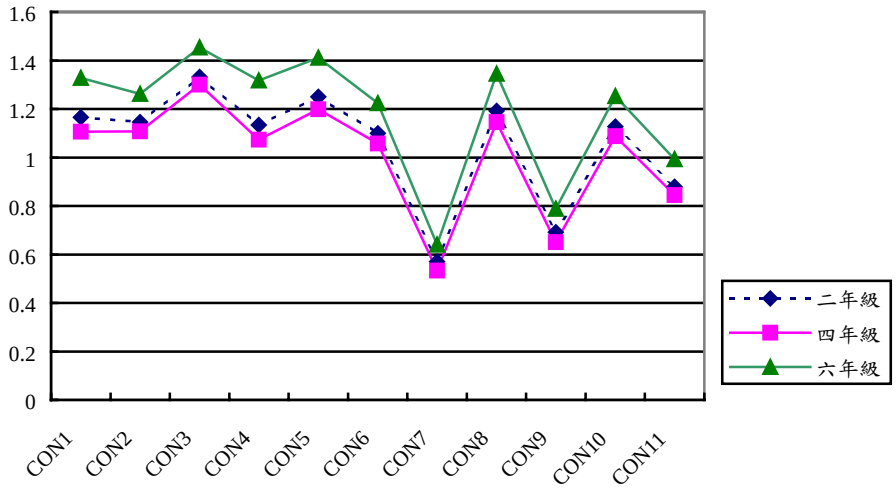
概念九「植物以葉子進行光合作用」及概念十一「如何檢驗水分」均是有關於蒸散作用的概念，均在國小自然科第七冊第三單元「植物的蒸散作用」中有所提及。經分析試題(該測驗之第 4、5、6、7、11 等題)之後，發現上述受試者表現不甚理想的試題，均是關於「以變色紙檢驗水分」的概念。由此可知學生對於「以變色紙檢驗水分」的概念，可能有學習上的困難。

表二　國小四年級教材植物概念表

概　　念	內　　容
概念一(con1)	影響花生長的因素有陽光、水分等
概念二(con2)	種花時應注意澆水、除草、施肥等事項
概念三(con3)	植物的根有多種功能
概念四(con4)	植物的莖有不同的外型(軟硬、高矮)
概念五(con5)	不同環境中的植物，葉片的形狀不同
概念六(con6)	花包含花瓣及花蕊
概念七(con7)	果實是從花朵的地方長出來的
概念八(con8)	植物以莖輸送養分
概念九(con9)	植物以葉子進行蒸散作用
概念十(con10)	影響植物蒸散作用的因素有陽光、葉數等
概念十一(con11)	變色紙可以檢驗水分

表三　各年級受試者在四年級植物概念上之期望得分比較表

	CON1	CON2	CON3	CON4	CON5	CON6	CON7	CON8	CON9	CON10	CON11
二年級	1.165403	1.146023	1.332112	1.134114	1.249715	1.099864	0.569915	1.192427	0.692155	1.12778	0.879691
四年級	1.10572	1.107426	1.300742	1.073615	1.199256	1.058027	0.533047	1.145956	0.650632	1.08759	0.844519
六年級	1.32835	1.262589	1.4535	1.317496	1.411027	1.223773	0.640453	1.344687	0.788471	1.252869	0.992883



圖二　各年級受試者在四年級植物概念上的期望得分比較圖

觀察本研究中各年級受試者在不同年級植物概念上的表現，發現在大多數概念上，二年級與四年級受試者的表現十分接近，由此可知學生在二年級到四年級階段，在植物概念上的學習進展情形並不顯著，甚至有些微學習停滯的情形；然而到了六年級的階段，受試者在各年級植物概念上的表現均有所提昇，且在大多數的概念上均呈現大幅成長的情形，可見在現行國小自然科學教材的植物概念中，學生在二年級至四年級的學習進展情形，遠不如四年級至六年級的學習進展情形。原因為何，有待後續研究探討。

由上述可知，本模式亦可有效診斷學生的學習進展情形，可提供日後進行學習進展指標研究時的參考。

(二)各能力群受試者在各概念上的比較

在上一段中，本研究探討各年級在各概念上的表現情形。本段則將直接以等化後的資料，將所有受試者依能力值的高低，區分為高能力組、中高能力組、中能力組、中低能力組、低能力組等五組，各組分組的依據如表四。

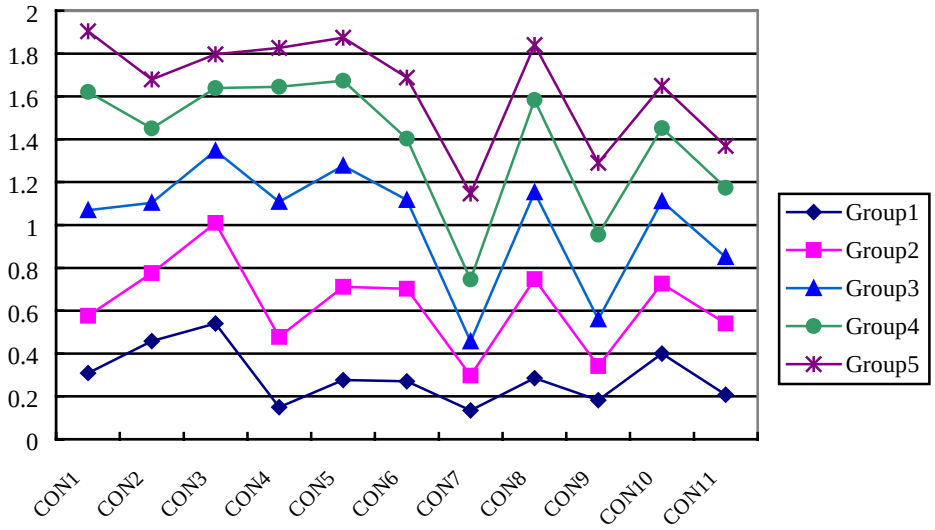
表四　各能力組能力值範圍

組　　別	能力值範圍
低能力組(Group1)	能力值＜－2
中低能力組(Group2)	－2≦能力值＜－1
中能力組(Group3)	－1≦能力值＜0
中高能力組(Group4)	0≦能力值＜1
高能力組(Group5)	能力值≧1

在資料分析部分，根據表五及圖三，發現各能力群受試者在概念七「果實的產生」、概念九「植物以葉子進行蒸散作用」及概念十一「如何檢驗水分」上的表現均差，經分析試題後，發現學生對於果實的產生方式可能有所迷思；此外，學生對於以變色紙檢驗水分的概念不熟悉，導致學生在關該於此概念的試題均表現不佳。

表五　各能力組受試者在四年級植物概念上之期望得分比較表

	CON1	CON2	CON3	CON4	CON5	CON6	CON7	CON8	CON9	CON10	CON11
Group1	0.308732	0.457518	0.539649	0.14912	0.275303	0.269365	0.13401	0.28443	0.182496	0.400016	0.206634
Group2	0.575536	0.775225	1.010262	0.476755	0.710936	0.701751	0.297504	0.746556	0.341435	0.726212	0.5394
Group3	1.068829	1.104162	1.346019	1.108105	1.276078	1.116409	0.455864	1.152353	0.559922	1.111019	0.849473
Group4	1.620969	1.450863	1.638331	1.645774	1.673208	1.403082	0.745799	1.584201	0.954727	1.452372	1.174193
Group5	1.905251	1.679993	1.796696	1.827215	1.874971	1.687648	1.146956	1.840922	1.290636	1.650168	1.368801



圖三　各能力組受試者在四年級植物概念上的期望得分比較圖

在概念一「影響花生長的因素」及概念二「種花的注意事項」的表現上，高能力受試者(能力值在 0 以上)及低能力受試者(能力值在 0 以下)有截然不同的情形。高能力受試者在概念一上的表現優於概念二，然而低能力受試者在概念一上的表現則較概念二為差，由此可知高能力受試者與低能力受試者在此二概念難易的認知上可能有所差異，若要進行補救教學，則所需加強的部分亦不相同。

同樣的，在概念三「根的構造」及概念四「莖的構造」的表現上，高能力受試者(能力值在 0 以上)及低能力受試者(能力值在 0 以下)亦有上述的情形。對高能力受試者而言，莖的構造似乎比根的構造稍微容易一些，然而對於低能力受試者而言，根的構造則是比莖的構造容易。

觀察各能力組在四年級植物概念上的學習進展情形，可發現在概念一「影響花

生長的因素」及概念四「莖的構造」上，當受試者由中低能力進展至中能力，以及從中能力進展至中高能力時，受試者的學習進展情形相當顯著，亦即能力值介於-2到 1 之間的學生只要能力稍微增加，其期望得分即可顯著增加。由於上述兩概念具有如此的特性，故可用以鑑別能力值介於-2 到 1 之間的受試者。

在概念三「根的構造」、概念五「葉的構造」及概念六「花的構造」等概念上，當受試者由低能力進展至中低能力，以及從中低能力進展至中能力時，受試者的學習進展情形相當顯著，亦即能力值小於 0 的學生只要能力稍微增加，其期望得分即可顯著增加。由於上述概念具有如此的特性，故可用以鑑別能力值小於 0 的受試者。

觀察受試者在概念八「植物以莖輸送養分」上的表現，發現受試者從低能力至中高能力間，每一階段的學習進展情形均相當顯著，顯示此概念對低能力學生是屬於較難懂的概念，但只要能力稍有進展，即可有不錯的學習進展。由於此概念具有如此的特性，故適合用於鑑別能力值小於 1 的受試者。

在其他概念方面，由於在各能力群之間的學習進展情形均相差不多，顯示學生在這些概念上呈現穩定發展的情形。在此需要特別說明的是，能力值小於 0 的受試者在概念七「果實的產生」及概念九「植物以葉子進行蒸散作用」上，似乎有學習進展緩慢的情形，值得教學者的注意。

二、認知網路評量模式的分析結果

本節將探討認知網路評量模式的適用性，主要是分析個別受試者的概念結構圖，並配合前一節的分析結果，進行診斷及相互印證。

個別受試者的知識結構圖主要分為節點與連結兩部分繪製，在節點的部分，乃是根據等化後的受試者能力值，以內插法計算其在各個植物概念(中概念)中，單一概念試題上的期望得分，並以大概念為節點，將所有相關的中概念的期望得分平均值，作為該大概念的存在機率(即強度)。

其次，在計算兩節點之間的連結強度時，首先找出此兩節點所涉及的所有中概念，再找出測量上述中概念的所有多概念試題，如果受試者答對某試題，則表示受試者在該題所涉及之中概念間的連結是存在的，此時中概念間連結的編碼記為 1；反之，如果答錯，則中概念間的連結不存在，編碼記為 0，可據此求得中概念間的連結強度。若同時有數個試題測量同一個連結，則本研究將所有涉及此連結之所有試題得分期望值的平均值加以呈現。

為使本研究繪製的知識結構圖清晰明瞭，所以本研究將以大概念作為知識結構圖的節點，並於節點內標示該節點的強度。在連結的線段方面，本研究依據連結的強度將線段區分為五個等級，如表六所示。

表六　連結線段的呈現型式

連結的強度範圍	線　段　型　式
0.8≦連結強度＜1	━━━━━━━━━━
0.6≦連結強度＜0.8	━ ━ ━ ━ ━ ━ ━
0.4≦連結強度＜0.6	————————
0.2≦連結強度＜0.4	━·━·━·━·━·━·━·━·━
0≦連結強度＜0.2	·····

由於本研究的抽樣人數達 1212 人，且分析的過程均相同，加上篇幅有限，所以本節僅自五個能力群中各抽取一人，共五人，呈現其知識結構圖，並做進一步的分析。本研究抽取的受試者編號及能力值如表七。

表七　受試者編號及能力值

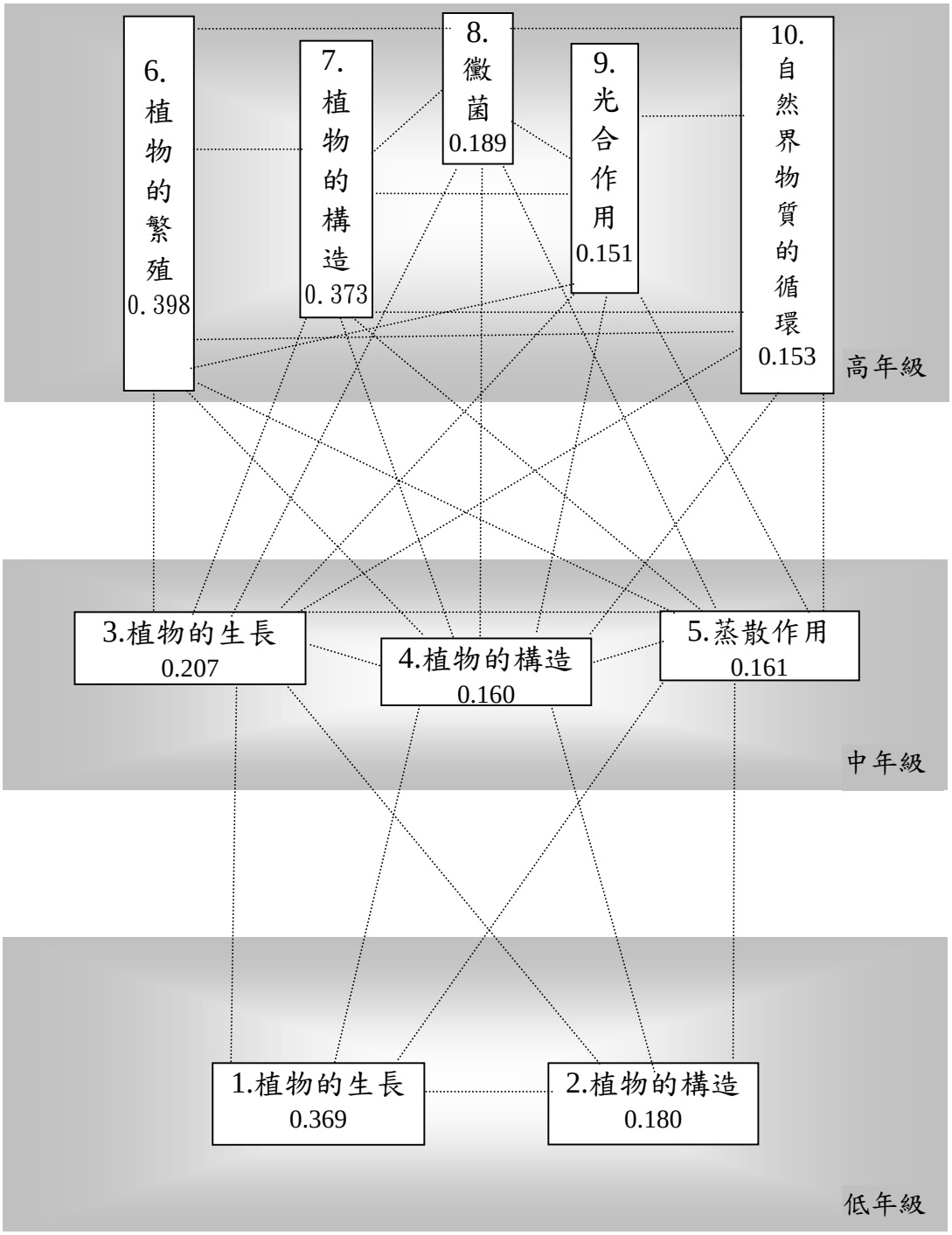
受試者編號	能力值
2642	– 2.19227
8623	– 1.10378
1225	– 0.03650
12429	0.81465
7225	1.52094

根據本研究的分析，國小自然科教材中，關於植物的中概念共有 46 個，所以以中概念作為節點繪圖時，繪製的圖形將相當複雜。為顧及圖形的可看性，所以本研究將上述以中概念分析的結果，再根據大概念及中概念的隸屬關係，將中概念轉化為大概念，並繪製成知識結構圖。

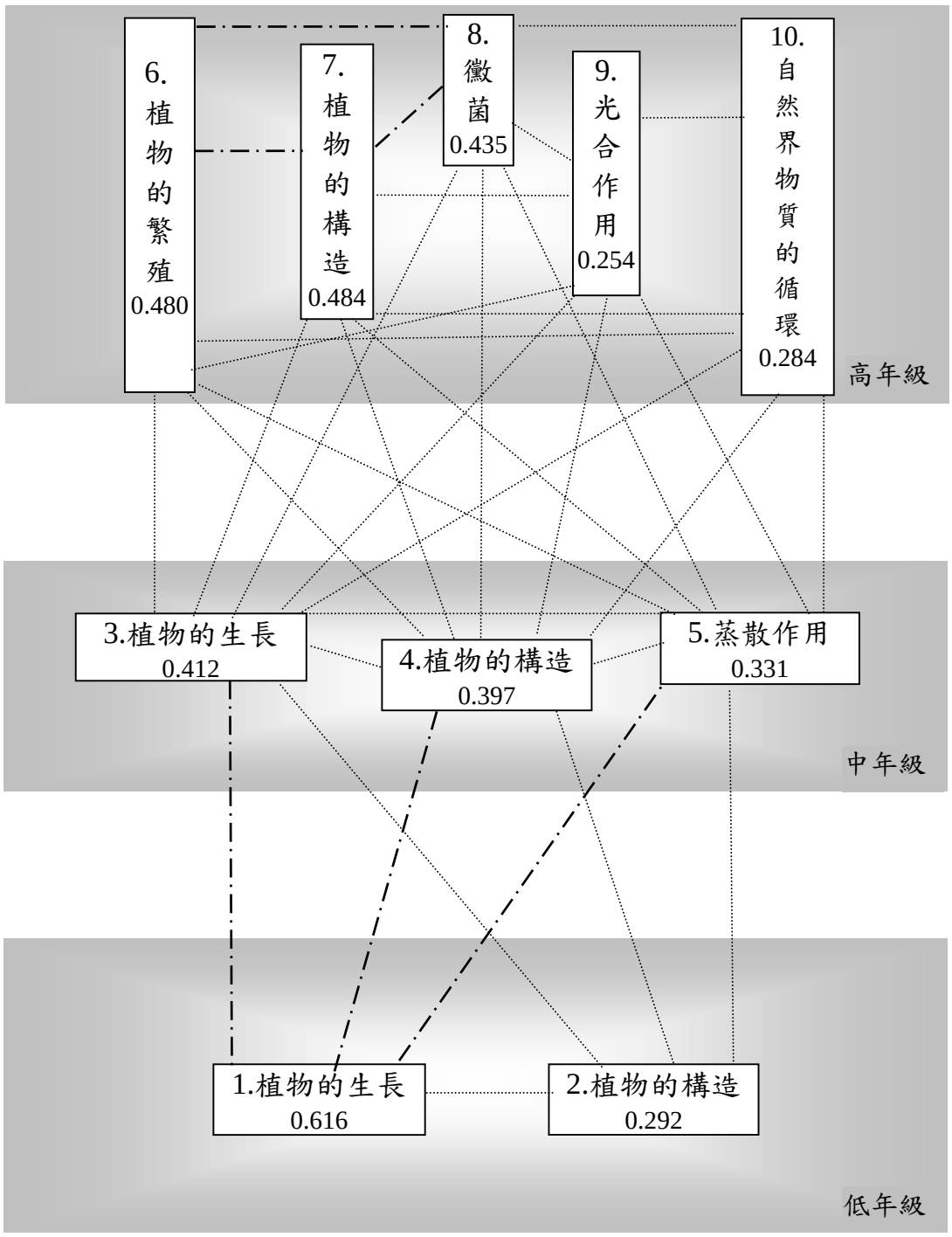
以下圖四至圖八將分別呈現 5 位受試者的知識結構圖，其排列順序是能力值由低至高，隨著能力值增加，各受試者知識結構圖的連結型式也有明顯的差異，能力值低的受試者，節

點與連結的強度均弱，但隨著能力值逐漸增加，節點以及節點間的連結強度均呈現增強的趨勢，而受試者的知識結構圖也就呈現越完整的情形。

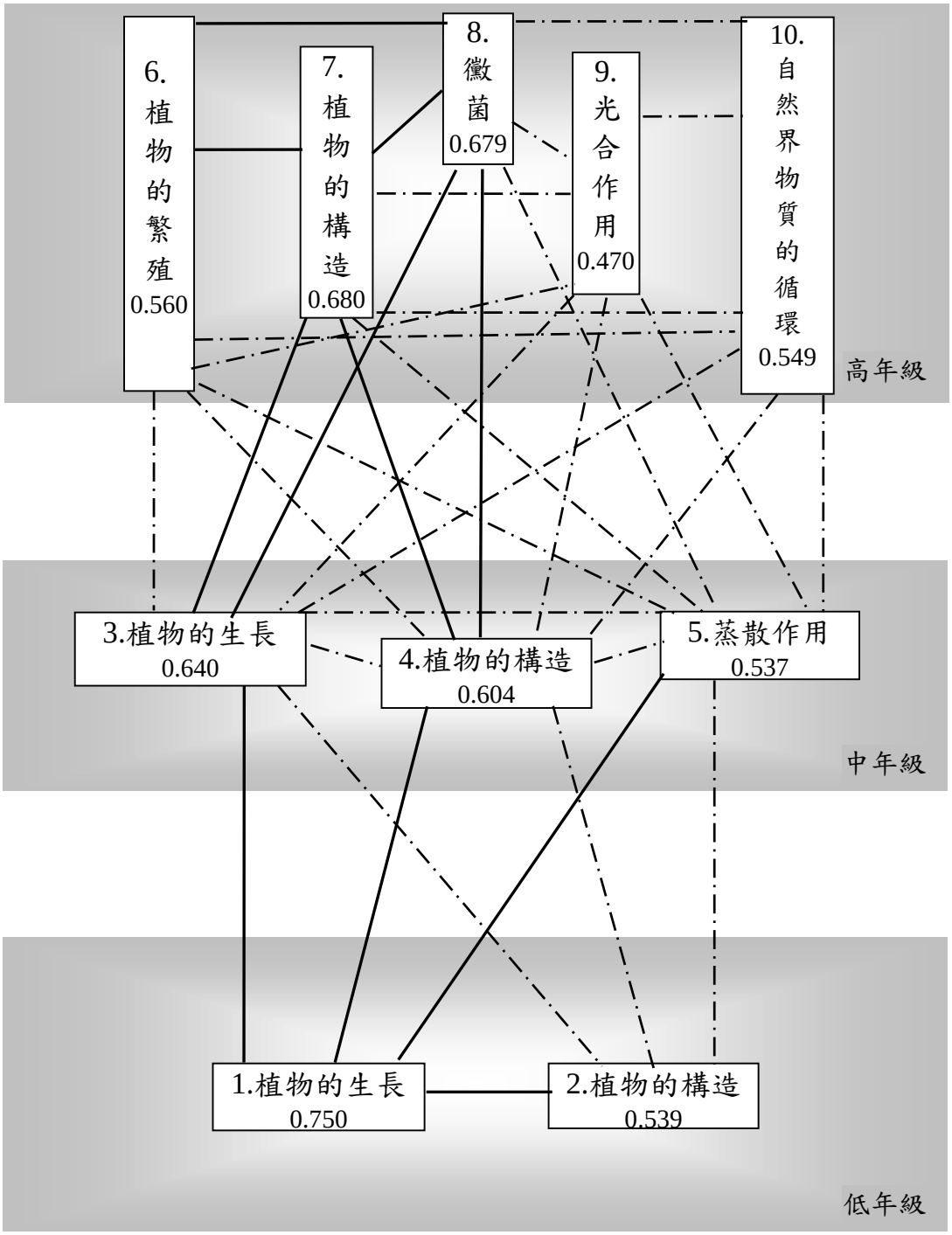
由上述分析可知，本模式以 IRT 能力值作為受試者的能力指標，可有效鑑別出不同能力的受試者，並能真實呈現出受試者的知識結構，協助教師瞭解學生的問題所在，以便迅速做最佳的補救。



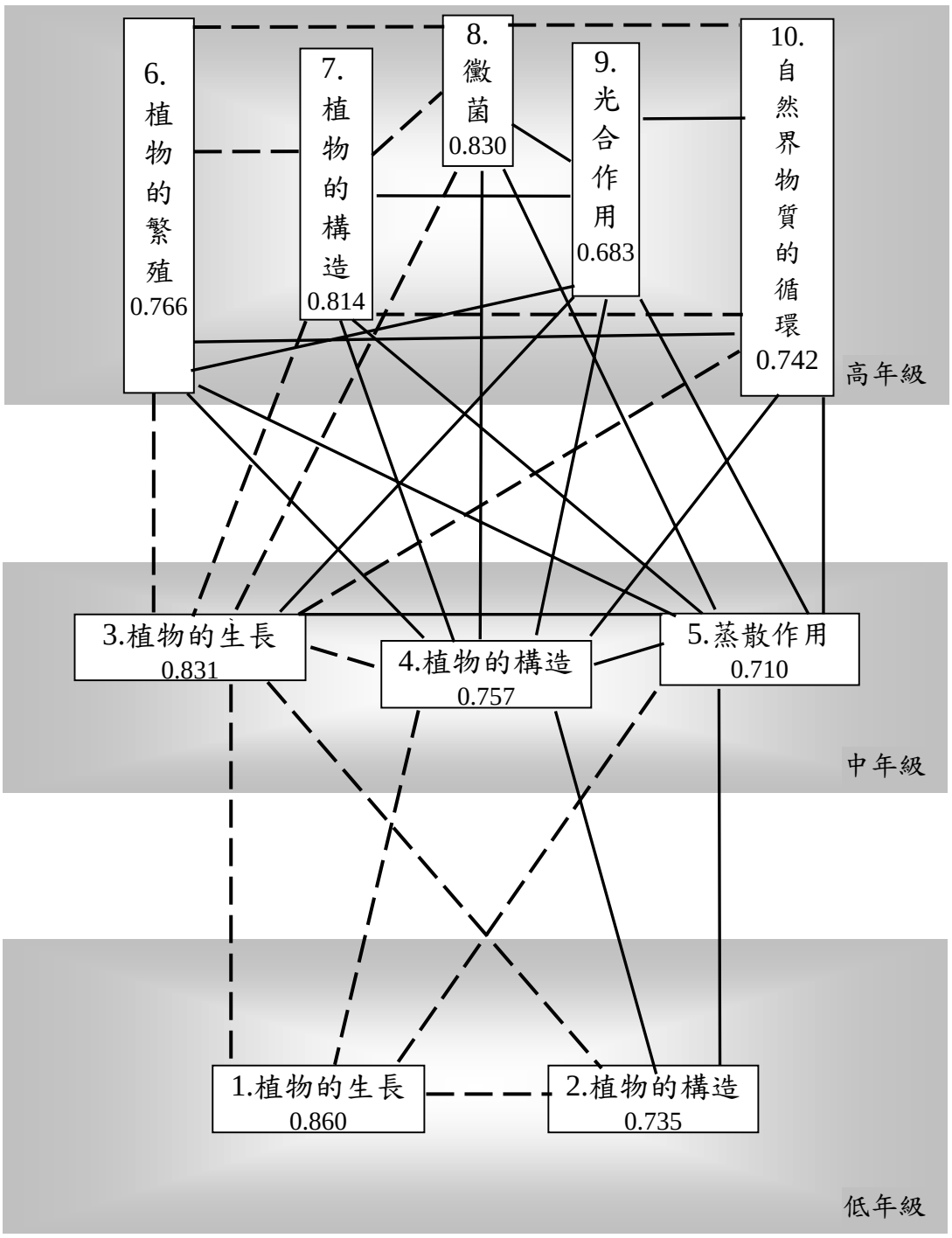
圖四　2642 號受試者(低能力組，能力值＝－2.19227)的知識結構圖



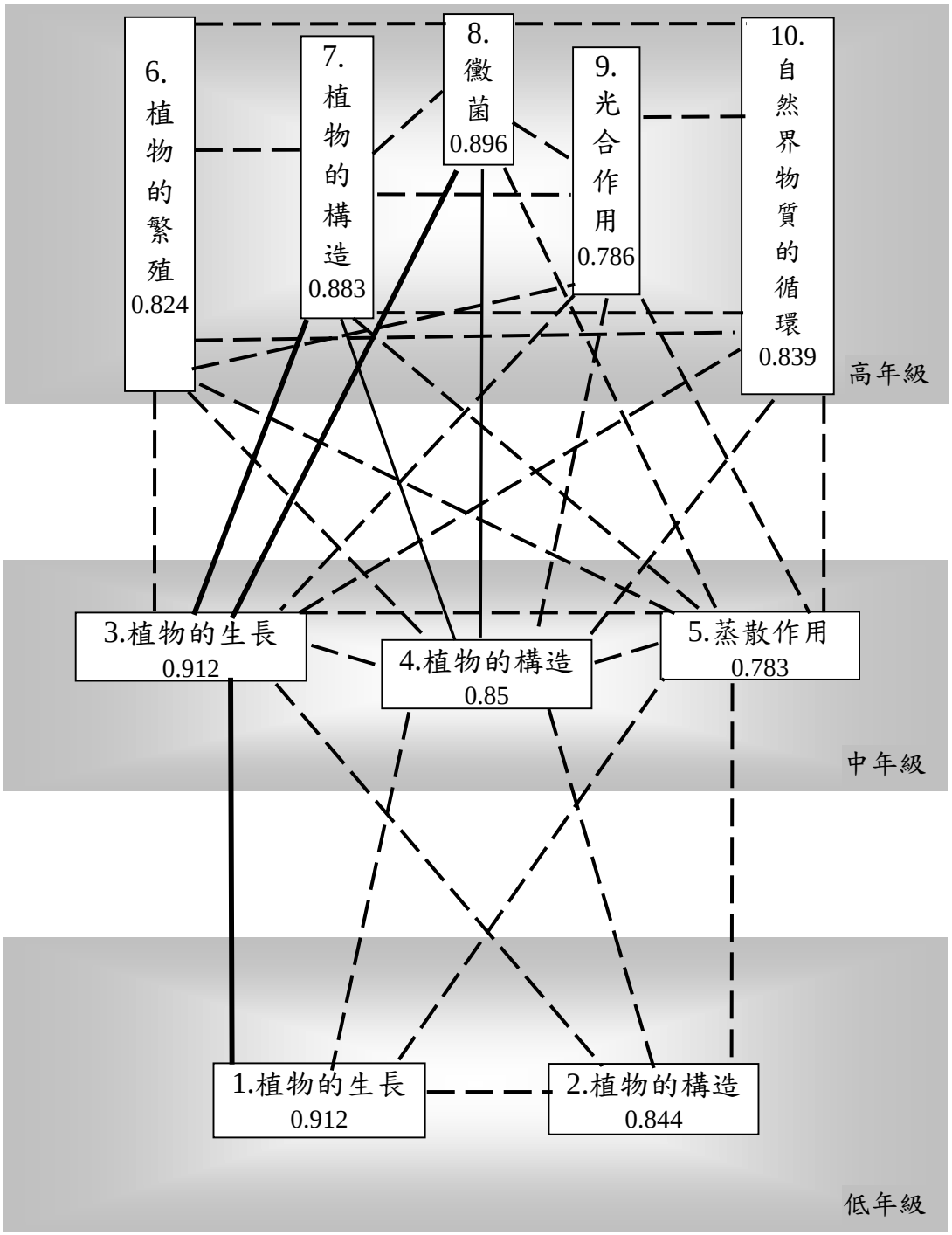
圖五　8623 號受試者(中低能力組，能力值＝－1.10378)的知識結構圖



圖六　1225 號受試者(中能力組，能力值＝－0.03650)的知識結構圖



圖七　12429 號受試者(中高能力組，能力值＝0.81465)的知識結構圖



圖八　7225 號受試者(高能力組，能力值＝1.52094)的知識結構圖

三、認知網路評量模式與 Marshall 模式之比較

綜合以上所述，本研究的認知網路評量模式與 Marshall 模式在理論假定及實例應用上的異同共有下列幾點：

(一)Marshall的模式假定節點及連結的強度是全有或全無的，如此雖可使基模圖簡潔明瞭，較能凸顯受試者的錯誤之處，然而此種假定太過簡化受試者複雜的行為，常常使得研究者在解釋實際資料時，無法得到合理的解釋。而本研究的認知網路評量模式是以ICCNP程式求得之得分期望值代表強度，摒除全有全無的限制，使本模式在應用上能更符合真實的情況。

(二)Marshall模式是以總分作為能力的指標，亦即總分相同的受試者，其能力值便被視為相同，事實上這是不正確的，因為以總分作為能力指標時，並未考慮到受試者的反應組型，如此將發生「總分相同、反應組型不同的受試者，其能力值卻相同」的不合理現象。而本研究的認知網路評量模式是以IRT能力值作為能力的指標，能力值相近的受試者，其知識結構圖亦相當接近，顯示無參數試題反應理論所估計出的能力值較能準確的代表學生的知識結構，因此以無參數試題反應理論所估計得到的能力值較傳統總分更適合作為能力的指標。

(三)Marshall的模式所能提供的訊息，在本研究的認知網路評量模式中亦能發現，且後者以強度的方式呈現，取代Marshall全有全無的方式，不僅合理且更能看出各能力組受試者的知識結構，顯示本研究的認知網路評量模式比Marshall的模式更為合理且能提供更多訊息。

(四)以IRT能力值作為能力指標時，由於能力值接近的受試者，其知識結構圖也會相似，所以具有準確分群的效果，如此將有利於教師進行教學安置、教學診斷及補救教學的工作。而由於Marshall模式以傳統總分作為能力指標，屬於同一群的受試者，其知識結構可能差異相當大，如此將增加教師進行教學診斷及補救教學工作時的複雜性。

(五)當前測驗統計理論中，只有試題反應理論提出完整且合理的測驗等化技術，認知網路評量模式是以無參數試題反應理論為理論基礎，因此能快速進行測驗等化的工作，進而深入瞭解受試者的知識結構。然而由於Marshall是以古典測驗理論為理論基

礎，使得測驗進行等化時，有許多限制不易解決(詳參李文忠，民84)，間接降低了模式的實用性。

經過上述比較，可發現本研究的認知網路評量模式的理論基礎較為完整，所能提供的訊息也較多，故適用性亦較 Marshall 模式為佳。

伍、結論與建議

經過以上的討論，我們可得到下列的結論與建議：

(一)由於Marshall評量模式的基本假定中有與實際情況不相符之處，為擴展評量模式的可行性及適用範圍，本研究在模式的建構上，即選擇以認知網路模式為基礎，並以無參數試題反應理論及PDP理論加以修正，因此本模式的基本假定較能符合實際的作答情況。而在受試者能力值的估計上，本模式採用無參數試題反應理論，所估計的能力值較精準，模式假定及限制亦較少，適用範圍較廣泛，程式的使用亦較方便。由此可知，認知網路評量模式可改善Marshall的基模評量模式的若干缺失。

(二)在本研究的文獻探討中曾提及，PDP理論假定節點的存在及節點間的連結，有部分促動的情形，然而若以古典測驗理論的技術而言，並無法具體呈現出節點及節點間連結的強弱程度。本研究結合無參數試題反應理論及PDP理論，利用無參數試題反應理論所估計的受試者得分期望值，作為節點及連結的強度，合理呈現受試者的知識結構圖。此外，應用無參數試題反應理論的另一項優點是能得到較佳的能力估計值，實際應用在分組診斷時，比用傳統總分來分群更為簡便正確。由此可知，以Marshall的評量模式為基礎，配合無參數試題反應理論及PDP理論所發展的「認知網路評量模式」，應是相當可行的模式。

(三)本模式根據PDP理論中「線性閾值處理單元」的概念，然改進其閾值的決定，並不決定特定的閾值，而是以ICCNP程式，計算處理單元的存在機率(即強度)，能更真實的呈現出受試者的知識結構。

(四)由於各年級受試者在關於人類感覺的概念，如「花的顏色、形狀、大小」、「葉的顏色、形狀、大小」、「花的氣味」等概念上的期望得分較高，顯示關於受試者親身體驗的概念，受試者較容易獲得良好的學習成果。因此，建議教師在進行教學前，應注意學生的先備知識，以協助學生學習新的教材內容。

(五)受試者在關於「以變色紙檢驗水分」的試題上，表現普遍不佳，由於此一概念在教材中是以實驗的方式呈現，推測應是實驗教學的問題，因為同樣的情形亦發生在「檢驗二氧化碳」及「影響黴菌生長的因素」等重視實驗的概念上，因此建議教師在進行上述概念的教學時，應加強實驗教學的部分。

(六)由於各年級受試者在「光合作用」概念的期望得分普遍均低，顯示對於國小學童而言，光合作用仍是相當難以理解的概念。所以關於此一單元是否適合放在國小教材，或是否該簡化該單元的教材內容，值得教材編製者加以考量。

(七)各年級受試者在「自然界物質的循環」概念上的表現均差，推測可能是由於該概念涉及的中概念，如「蒸散作用」、「蒸發作用」、「凝結作用」等概念的難度較高的原因。由於該單元進行教學時，並未安排適當的複習舊概念的活動，建議教師能在此方面加以注意，補教材內容之不足。

(八)由於在「影響花生長的因素」、「種花的注意事項」、「根的構造」及「莖的構造」等中概念的表現上，高低能力組的受試者有截然不同的情形，顯示高低能力的受試者在知識結構上應有相當大的差異，因此在進行補救教學時，對於高能力的受試者與低能力的受試者，應該分別提供不同的協助。

參考文獻

一、中文部分

余民寧(民 80)：試題反應理論的介紹。研習資訊，第 8 卷，6 期，13-17。

余民寧(民 82)：心理測驗的發展與應用。中國測驗學會主編，23-62。台北市：心理出版社。

余民寧(民 84)：認知診斷測驗的發展趨勢。教育研究雙月刊，45 期，14-22。

李文忠、李慶峰(民 84)：EQUATICC：無參數試題反應理論之等化程式(未出版)。國立台中師範學院測統中心技術報告NTTC-TSC-T-0010。

施淑娟(民 86)：認知網路評量模式及其實例應用之研究－以「分數的加法」單元為例。國立台中師範學院國民教育研究所碩士論文。

許擇基、劉長萱(民 81)：試題作答理論簡介。台北：中國行為科學社。

陳怡欣(民 83)：試題反應理論與知識結構分析在國小六年級社會科學習成就評量上之應用。市立台北師範學院初等教育研究所碩士論文。

黃達三(民 82)：國小學生分類能力初探。行政院國家科學委員會專題研究成果報告(NSC-81-0111-S-143-501-N)。

裘維鈺(民 84)：國小學童植物概念及其相關迷思概念之探究。國立台中師範學院國民教育研究所碩士論文。

劉兆文(民 83)：「無參數試題反應理論之等化模式」適用性之探討—國小數學科成就測驗之研究。市立台北師範學院初等教育研究所碩士論文。

劉湘川、許天維、郭伯臣、胡豐榮(民 83)：試題關聯結構分析法及其在國小數學科教學上的應用。初等教育研究集刊第二期，11-26。台中市：國立台中師範學院。

劉湘川、許天維、施淑娟、郭伯臣(民 84)：基模評量及實例應用。高雄市：第十一屆科學教育學術研討會彙編(付印中)。

鄭富森(民 83)：一個CTT和IRT之整合模式。初等教育學刊第三期，101-138。

簡茂發、劉湘川、許天維、郭伯臣(民 82)：試題反應理論在國民教育階段國小數學科基本學習成就評量上的運用。台北市：第一屆華文社會心理與教育測驗學術研討會。

二、外文部分

Allen, M. J., & Yen, W. M.(1979). Introduction to measurement theory. Monterey, CA：Brooks/Cole.

Amir, R., & Tamir, P.(1991). Proposition Generating Task(PGT) : A measure of meaningful learning and of conceptual change. The Annual meeting of the American Education Research Association. Chicago, IL.(ERIC Document Reproduction Service No. ED 334 222).

Anderson, J. R.(1995). Cognitive Psychology and its implications. New York：W. H. Freeman and Company.

Grossberg, S.(1980). How does the brain build a cognitive code？Psychological Review, 87, 1-51.

Konorski, J.(1967). Integrative activity of the brain. Chicago： University of Chicago Press.

Marshall, S. P.(1993). Assessing schema knowledge. In N. Frederiksen, R. Mislevy, & I. Bejar(Eds.). Test theory for a new generation of tests(pp.155-180). Hillsdale，NJ：Lawrence Erlbaum Associates.

Martindale, C.(1981). Cognition and consciousness. PacificGrove, CA：Brooks/Cole.

Martindale, C.(1991). Cognitive psychology-A neural-network approach. PacificGrove, CA：Brooks/Cole.

Nichols, P. D.(1994). A framework for developing cognitively diagnostic assessment. Review of Educational Research, 64, 575-603.

Roth, K. J., & Anderson, C. W. (1987). The power plant: Teacher's guide to photosynthesis. Occasional paper No. 112. Michigan State Univ., East Lansing. Inst. For Research on Teaching. (ERIC Document Reproduction Service No. ED 288 699)

Rumelhart, D. E., Hinton, G., & McClelland, J. L.(1986). A general framement for parallel distributed processing. In D. Rumelhart & J. L. McClelland(Eds.), Parallel distributed processing, Vol.1. Cambridge, MA：MIT Press.

Stead, B. (1980). Plants. Learning in science project. Working paper No. 24. Waikato Univ., Hamilton(New Zealand). (ERIC Document Reproduction Service No. ED 236 007)

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

Tatsuoka, K. K.(1986). Diagnosing cognitive errors：Statistical pattern classification and recognition approach. Behaviormetrika, 19, 73-96.

Tatsuoka, K. K.(1990). Toward an integration of item-response theory and cognitive error diagnosis. In N.Frederricksens, Glaser , Lesgold and Shafto (Eds.). Diagnostic monitoring of skill and knowledge acquisition. Lawrence Erlbaum, Hillsdale, New jersey.

Tatsuoka, K. K., & Tatsuoka, M. M.(1987). Bug distribution and pattern classification. Psychometrika, 52, 193-206.

Wax, N., & Stavy , R. (1987). Children's conceptions of plants as living things. Paper presented at biennial meeting of the International Society for the Study of Behavioral Development. Tokyo, Japan. (ERIC Document Reproduction Service No. ED 294 732).

附錄　本研究之植物概念測驗

光合作用

植物與動物有許多的差別，其中一個差別在於動物能夠移動，而絕大部分的植物則無法移動。因此，動植物在攝食方式上便有相當的不同。動物主要是以植物或其他動物為食，而植物則主要則是靠本身產生養分，以供應本身生存所需。

光合作用是植物用以產生養分的過程，沒有光合作用，大多數的植物都將無法維生。然而，光合作用是相當複雜的一個過程，牽涉到相當多的因素，例如：植物在何種環境下才能行光合作用？需要那些因素？又會產生哪些物質呢？關於這些問題，我們可以從以下的實驗中看出一些端倪。

《實驗一》：檢驗澱粉

收集植物的種子，並事先泡水。將泡過水的種子研磨成碎粉狀，放在培養皿中，滴幾滴稀碘液，結果碘液由__1__色變成__2__色。

《實驗二》：去除__3__

準備一些植物的葉子，用以測試是否含有澱粉，但是由於葉子表面光滑，又有顏色，加碘液之後不易觀察，故必須將葉子做某些處理。

1.首先，把葉片放進熱水中，浸泡約 2 分鐘。

2.將葉子移出，在放進一個裝有酒精的燒杯中，並確定酒精淹沒葉片。

3.將裝有酒精的燒杯放進將有熱水的燒杯中，以酒精燈加熱至葉片褪色為止。

4.熄滅酒精燈，在將葉片從酒精中取出。

5.在葉片上滴幾滴碘液，觀察葉片的顏色有何變化。

《實驗三》：__4__與光合作用的關係

1.在天氣良好的情況下，找一棵樹，用錫箔夾住幾片葉片，使其無法接受陽光的照射。

2. 2～3 天後，摘下照光及不照光的葉片，依照實驗二的方法，用熱水及酒精進行加熱。

3.分別加入幾滴碘液，並觀察二者顏色的差異。

《實驗四》：檢驗 5

1.取半杯清水，加入藍色指示劑，攪拌均勻，使水呈現藍色。

2.以吸管吹氣到容器的水中，觀察水溶液顏色，發現變成 6 色。

3.將上述的水溶液分別倒進四個小燒杯中，標明甲、乙、丙、丁，並在丙、丁二杯中各放進一枝水蘊草。

4.將甲、丙二杯放在陽光下，乙、丁放在陰暗處，觀察溶液的顏色會有什麼不同的變化。

《實驗五》：收集並檢驗氧氣。

《摘錄改寫自：國立編譯館自然科學課本第十冊第七單元》

請根據上述文章以及下列的選項及代碼，回答下列問題：

1. □① □② □③ □④ □⑤ □其他，_____。

2. □① □② □③ □④ □⑤ □其他，_____。

3. □① □② □③ □④ □⑤ □其他，_____。

4. □① □② □③ □④ □⑤ □其他，_____。

5. □① □② □③ □④ □⑤ □其他，_____。

6. □① □② □③ □④ □⑤ □其他，_____。

7.動物在進行呼吸作用時，是吸進何種氣體？

□① □② □③ □④ □⑤ □其他，_____。

8.植物在進行呼吸作用時，是吸進何種氣體？

□① □② □③ □④ □⑤ □其他，_____。

植物是以光合作用產生養分，以供應本身的能量。光合作用主要是靠葉片中的__9__，配合陽光，吸進二氧化碳，藉以產生__10__及澱粉。

9. □① □② □③ □④ □⑤ □其他，_____。

10. □① □② □③ □④ □⑤ □其他，_____。

11.在《實驗五》中，以試管收集到足夠的氧氣後，以線香檢驗之，此時線香會燃燒得更旺盛，這是因為氧氣有何種性質？

□① □② □③ □④ □⑤ □其他，_____。

《選項群與代碼》

代碼	答案	代碼	答案	代碼	答案	代碼	答案	代碼	答案
①	黃色	①②	紫黑色	②④	褐色	①②③	藍色	①④⑤	紅色
②	水蒸氣	①③	二氧化碳	②⑤	氧氣	①②④	氫氣	②③④	色素
③	脂肪	①④	礦物質	③④	澱粉	①②⑤	蛋白質	②③⑤	葉綠素
④	可燃性	①⑤	助燃性	③⑤	不燃性	①③④	根	②④⑤	莖
⑤	陽光	②③	紫外線	④⑤	錫箔	①③⑤	氣孔	③④⑤	葉片

Abstract

The purpose of this study is to develop a polytomous assessment model of cognitive network based on Marshall's assessment model of schema, parallel distributed processing (PDP) theory on cognitive and non-parametric item response theory. Consequently, the researcher also compares this new model with the assessment model of schema. In this study, researcher uses the data of a plant concept test to explores the advantages and disadvantages of the two models.

The main findings of this study are listed as following:

1. The polytomous assessment model of cognitive network uses the expected score to replace the total score in the assessment model of schema. The new model is better than the old one practically.
2. The polytomous assessment model of cognitive network improves the disadvantages of the assessment model of schema by using the ability parameter of IRT as the ability guideline of students. The assessment model of cognitive network helps teachers to diagnose the learning difficulties of students effectively and reasonably.
3. The polytomous assessment model of cognitive network can provides more information of students, such as cognitive diagnose, learning growth between different grades.
4. The polytomous assessment model of cognitive network has the advantage of test equating, which helps the researcher to understand the knowledge structure of students.

To sum up, the polytomous assessment model of cognitive network can provide more information than the assessment model of schema. It is a reasonable, feasible, and suitable model which can help teachers to understand the knowledge structure of students.