

排序理論在概念結構評量上的應用

余民寧* 陳嘉成**

摘要

本研究目的旨在嘗試發展一套評量技術，以用來協助教師了解學生的學習困難與診斷學習缺陷所在，進而研擬有效的補救教學策略，增進學生的學習成果。

本研究採用隨機抽樣方式抽取研究對象，抽得台北市文山區萬興國民小學五年級學生 24 名，接受研究者彙編的測驗，該測驗已經過試題概念分析。接著，本研究提出實施本評量技術的過程說明，以幫助使用者達到分析測驗評量資訊的目的，增進對學生學習成果的了解，進而提出適當的補救措施。

本研究復根據上述評量技術的發展過程，擬定三項研究問題進行探討，並且獲致下列的研究結論：

1. 不同學習類型學生的概念結構品質確實有所不同。其中，「學習穩定型」學生比「學習異常型」學生的概念結構品質較為優良。

2. 不同學習類型學生的概念結構圖並不完全相似。相較之下，學習類型愈相近者，其概念結構較為相似；學習類型愈不相近者，其概念結構則愈不相似。

3. 不同學習類型學生的概念結構與（教師的）標準答案的概念結構之相似程度並不完全相同。其中，半專家型（即「學習穩定型」）學生比生手型（即「學習異常型」）學生的概念結構，更接近專家型（即教師的標準答案）教師的概念結構，或者說是與教師的概念結構更相似。

最後，本研究亦針對研究成果的涵義及限制，提出補充說明。

關鍵詞：概念結構、學生問題表分析、概念構圖、排序理論

* 本文作者感謝行政院國家科學委員會補助部份研究經費(NSC-85-2413-H-004-007)及兩位匿名評審提供寶貴的改進建議。

* 為本校教育學系教授

** 國立政治大學教育學系博士班研究生

Abstract

The purpose of this research is to develop an assessment technique to help teachers understand students' learning difficulties and learning defects, and to propose an efficient strategy to improve students' learning outcomes.

This research used 24 randomly selected fifth-grade elementary school children as subjects. One teacher-made and one researcher-made mathematical achievement are used in this research. The S-P table analysis is used in analyzing the first test to find out students' learning patterns. Ordering theory and order analysis technique is used in analyzing the second test to figure out students' concept structures. And the teacher's concept structure is analyzed in the same way.

The results are as following:

1. Students with different learning patterns will have different qualities in concept structure of math test. The quality of concept structure of Learning Stable Type is better than that of Learning Aberrant Type.
2. Students with different learning patterns will not always have the same concept structure maps of math test. In comparison, the more learning patterns are proximate, the more concept structure are similar. And vice versa.
3. The concept structure of students with different learning patterns will not be all the same as teachers' standard one. In comparison, the concept structures of the semi-expert type (i.e., the Learning Stable Type) students are more similar to teacher's than those of novice type (i.e., the Learning Aberrant Type) students are.

Finally, several implications, restrictions, and conclusions are proposed and discussed.

Key Words: concept structures 、 student-problem chart analysis
` concept mapping 、 ordering theory

壹、緒論

一、認知診斷在教學評量上的重要性

教室情境極為複雜，若欲增進教學效果，教師所應掌握之教學情境變項相當多，而教學

排序理論在概念結構評量上的應用

模式的設計即在藉由各種相關理論的整合，使教學效果發揮到極致，進而達到教學的目標。

一般而言，任何一種教學都在於期望學習者能學到教學者所要傳授的知識，而教學是否達到目的則需藉由「評量」來表達。過去在行為主義主導心理學發展，而教育心理學依附在心理學研究的時期，由這個典範所衍生出來的評量方式著重在學習「成果」的評量，而忽略了學生在知識建構過程中的「形成性評量」。注重成果評量所連帶產生的結果，就是使評量的功能只能停留在給予學生一個「標籤」的層次上，並進而影響學生的學習意願，致使評量的「目的」與評量的「方法」形成本末倒置的窘境。

Black (1993) 即指出，評量可以成為教學或學習的一個強而有力的工具，亦可能成為最大的致命傷，這其間的關鍵乃取決於學校決策者是注重形成性評量或總結性評量而定。注重總結性評量的後果，就是造成「評量引導教學」和「評量引導學習」；教師以知識最終的型態呈現給學生，學生將這些事實知識照單全收地背誦下來，在這樣的教學過程中，學生所學到的知識就是他們將來考試所會出現的知識形式。如此，經年累月下來，學生便喪失探索知識的能力與興趣，對於沒有標準答案的知識會失去窺其堂奧的信心，甚至在日常生活中凡事都依附於一個有形或無形之權威所給予的「標準答案」，因此在多元文化與價值解構的社會中，許多學生都因喪失批判性思考的能力，而沉淪於社會的盲流中（陳嘉成，民 85）。

由以上的論述看來，當我們在設計或發展一套課程或評量系統時，必須將「注重教學成果」的重點轉移至「如何達到這些目標」的方法或手段上；準此而言，評量本身除了評量學生的學習表現之外，還必須能扮演一個動機啟發者的角色（Linn, 1993）；即使無法啟發學生的學習動機，也至少應該要能夠提供學習者一個回饋的訊息，以作為下次學習或知識建構的參照。

本研究的目的即在以知識結構表徵的理論為基礎，提出並驗證一套評量的技術，它不僅可以作為一般教學評量用的技巧，更可以作為認知診斷測驗技術的先遣研究(pilot study)。必須一提的是，筆者認為：在認知診斷的歷程中，「質化」與「量化」的方法皆不可以偏廢，本文之所以提出量化的程序，並不意含著孰優孰劣的立場，而是因為以質化為主的認知診斷方式十分倚賴教師本身在學科知識與測驗理論及教學(認知)心理學方面的專業知識與經驗；學科知識對教師而言應該是沒有問題，但是後二者卻非一般教師都具備的專業知識。因此，

提供一套量化、客觀的診斷分析程序，卻可以彌補上述緩不濟急的情況，甚或將量化的程序當成是進行質化分析的先遣作業，如此將有助於教師能在短時間內，獲取到大部分學生的整體相關訊息，以利於安排後續的質化診斷分析。

二、知識結構、表徵與學習

(一) 「表徵」研究的重要性

自古以來，關於人類是如何知道他的環境，或是以什麼樣的方式來代表他們所知道的一切事物，一直是哲學家、心理學家甚或教育學家所關切的問題。而傳統的自然科學教學深受經驗主義的影響較深，認為知識存在於學習者的經驗之外，與學習者獨立無關；而教師的任務就是將知識「注入」學生的腦袋裡，再利用傳統的紙筆測驗來檢查學生是否習得這些知識。在這種教育模式下，學生或許真的是學到了一些知識，不過，也真的「只有」學到這些「考試會考」的知識而已（陳嘉成，民 85）。

過去在行為主義主導心理學發展的時期，教育心理學家在了解人類的學習歷程上強調：學生是否發生學習，端視行為是否產生改變而定。因此，評量上是重視學生的測驗分數是否提高（Goldsmith, Johnson & Acton, 1991）。但是，這樣的評量取向卻無法讓教師得知學生是如何組織這些知識的訊息；此外，若是學生的知識中存有「錯誤概念」(misconception)，亦難以得知。

自七十年代，認知科學在學術界蓬勃發展以來，它的目的在深入了解人類的學習歷程；亦即，有關人類如何「獲取知識」、如何「處理與儲存知識」，以及如何「提取既存知識」等議題，都是這門學問所關心的重點（Mayer, 1981, 1987）。

認知心理學對人類學習行為的種種解釋中，特別強調「先備知識」(prior knowledge) 所扮演的角色（黃文俊，民 83）。先備知識的俱足，可以協助學生做外在聯結，使學生在一定的架構上將新、舊知識作一番統合，並使得新知識的細節部份，具有更多的提取線索。通常，我們稱學生此時的先備知識為「認知結構」。

排序理論在概念結構評量上的應用

所謂的認知結構，係指個人在知覺及理解客觀現實的基礎上，在腦中所形成的一種心理結構，係由個人過去的認知經驗所組成的（朱智賢，1989）。Piaget 認為它是個體認知活動的產物；Bruner 則稱其為「表徵」(representation)。而所謂的表徵（朱智賢，1989）即是指將外在現實世界轉換為內在的心理事件；在這段以概念代表實物的歷程，即稱之為表徵（張春興，民 78）。在學習的過程中，一個新的訊息或觀念不是被個人現有的認知結構所同化，就是改變現有的認知結構，或是接納新的經驗，以產生新的認知結構。

在研究知識結構的相關文獻中，有些著重在「命題」(proposition)，有些則將焦點放在「概念」(concept)的分析上；但是，不論知識結構的基本單位是「命題」或「概念」，我們在探索一個知識結構時，通常都是經由「知識表徵」來得知。認知心理學家在研究概念形成、語文理解、思考運作與記憶等主題時，常會涉及探討「知識結構的內在表徵」的問題。所謂「知識表徵」，即是指外在知識結構存在人類心理的內在表徵，是人類對知識的記憶或貯存型式（郭姿媚，民 75；Solso, 1995）；換句話說，知識表徵乃是個體經過「認知活動」後的產物，係由心理活動過程中所產生的各種內在形象。

Solso (1995) 認為知識表徵就是我們對知識編碼、儲存、及提取的形式；也就是說，人們是透過內在表徵的形式來處理訊息。因此，在教學或評量時若能由學生在學科知識的表徵上著手，將有助於我們了解：學生是如何組織知識，以及在學生的認知結構裡知識（概念）與知識（概念）之間的關係為何？

(二) 專家與生手的比較

在談到知識結構與知識表徵時，有個貫穿此二者間的研究 ---- 專家與生手 (experts and novices) ---- 是不可不提的。綜觀專家與生手的研究中，我們可以發現二者之間的差異，大致包括下列各點（林清山，民 80；Behr, Goldin, Pierce, & Threadgill-Soweder, 1985; Chi, Feltovich, & Glaser, 1981; Larkin, Mcdermont, Simon, & Simon, 1980; Shavelson, 1972, 1974）：

(1) 在知識組織方面：Larkin 等人發現，生手在知識表徵上傾向於做「小規模的程式單位」的組織，而專家則傾向於做「大規模的程式單位」的組織。因此，很顯然的，後者

在問題解決的自動化上比前者來得好。

(2) 在解題策略方面：Larkin 等人發現，生手在解題時傾向使用「逆向運作」的方式，即從最終的目標條件往前回溯去尋找達成目標條件的變項；相對的，專家則傾向於使用「順向運作」的方式，即從現有已知的條件來找未知的條件。因此，在解題時，只需要將已知的條件置入適當的公式，便可以求出答案。

(3) 專家有較多的概念節點。

(4) 專家使用較多的關係來定義每一個概念節點。

(5) 專家在節點與節點之間，有較多相互聯結的關係。

(6) 專家在概念節點間的聯結較強，知識檢索所需的時間也較短。

(7) 專家在問題解決時，傾向於使用「由上至下」的過程，即由抽象（理論）到具體（資料、訊息）的過程。而生手則傾向於使用「由下至上」的程序。

(8) 若只針對「陳述性知識」的知識結構來說，專家在知識結構的深度、廣度、抽象性、和諧性上都比生手來得好。

過去有關專家與生手的研究，大都已經證實二者在訊息的處理上有所不同，但是卻缺乏一套具體可行的方式，來比較專家與生手之間在知識結構上的差異到底在哪裡？而本研究即是以此為主要的著眼點。

三、知識結構的評量與認知診斷

由以上分析可知，近來教育心理學受到認知學派的衝擊，學習內容與教學歷程的內涵亦重新界定，其主要概念包括下列三點（洪碧霞，民 84）：(1)學習是知識的建構而非累積，(2)學習的發生不在於記錄訊息而是在於解釋訊息，(3)教學是一種介入，介入學生知識建構的過程而不是知識的直接傳遞。而 Ault(1985)亦指出：知識或概念的保留絕非單獨、個別的存在，而是以整體的方式保留在記憶當中。但是，過去的評量方式以古典測驗理論為基礎，最常用的是非、選擇題與簡答題，所測量的通常是許多彼此獨立的訊息，因此它們似乎不足以用來說明知識是一種統整的(intergrative)事實。

此外，過去古典測驗理論係以「試題」為最基本的分析單位，評量結果所能提供給老師

排序理論在概念結構評量上的應用

的「訊息」，不是全對就是全錯，因而忽略了學生在某些概念的學習上並非「全有或全無」(all or none)的可能，所以缺乏認知診斷的積極性意義。

余民寧（民 86a, 民 86b）指出：傳統的紙筆測驗方式，太過重視統計技術的分析而忽略掉測量背後所具有的心理學建構，如果未來的認知心理學與心理計量學的發展，能夠針對測量背後之理論建構提出認知上的診斷與分析，將可以解決或改進當前教學或評量上的諸多問題。

就在這些迫切的需要下，目前測驗的發展正朝向這種科際整合的方向進行，而獨自發展成為一門新興學科領域，稱為「認知診斷評定」(cognitively diagnostic assessment，簡稱 CDA) (Nichols, 1994)。這個領域有別於傳統上探討認知診斷測驗的古典測驗理論，和植基於單向度潛在特質理論的試題反應模式。它是以當代測驗理論為基礎，同時結合認知心理學對學習理論的研究與發現，用更精緻的統計模式來探索學生的認知結構（余民寧，民 86a）。

典型的研究例子，如 Tatsuoka (1983, 1986)、Tatsuoka & Tatsuoka (1987, 1988) 所發展出一種叫做「規則空間」(rule space) 的測量模式，用來診斷、偵測、及解釋小學生在四則運算的問題解決時，會因為他們使用了「錯誤規則」，而產生系統性的認知失誤 (cognitive bugs) 而答錯，因而出現怪異的反應組型。

然而，這個模式背後涉及艱深的數理統計學知識，並不容易被中小學教師所接受，而且使用起來亦頗為複雜，推廣上亦不容易。基於這樣的理由，S-P 表是另外一種可行的工具。

S-P 表的分析技術係由日本學者佐藤隆博 (Takahiro Sato) 於 1970 年代所創 (Sato, 1969, 1971)，是一種將學生在試題上的作答反應予以「圖形化」的分析技術，其目的在於獲得學生的學習診斷資料，以提供教師實施有效的學習輔導之參考。

S-P 表所關心的課題是學生在測驗試題上的作答「反應組型」，嘗試以幾個指標化數據做為診斷或判讀該反應組型是否為不尋常(unusual)或異常(aberrant)的一種測驗分析方法。這種分析測驗資料的方法，特別適用於以班級為單位的少數人資料，尤其更適合用於形成性評量的測驗資料分析，它是一種屬於不對母群體特性設定任何假設值的統計推論方法 ---- 無母數統計方法 (nonparametric method)。S-P 表分析可用來協助教師診斷學生表現、測驗品質、及教學成果的有效工具，以作為改進教學、命題、與輔導學生之參考；近年來，更有學者從及

事中文化 S-P 表分析軟體程式的研究和出版(余民寧，民 86a;吳信義、吳錫修，民 86)，預期更方便教師們的使用。

四、排序理論在認知診斷上應用的可能性

由於 S-P 表的注意係數係以 S-P 表之邊際次數為基準變異量來表示(余民寧，民 86a)。因此，若兩組學生在邊際次數的得分都一樣時，所計算出來的注意係數亦相同，但是此二組學生的試題反應卻未必相同。當然，不一樣的試題反應組型所表徵的學生的知識結構亦不相同。因此，若能將 S-P 表加以改良的話，將可以達到更進一步的診斷功能。

由 Airasian 等人提出的「排序理論」(ordering theory)，正可以克服過去統計方法在分析階層作業網路 (hierarchy task network) 資料時所面臨的限制 (Airasian & Bart, 1973, 1975; Bart & Krus, 1973)。排序理論是以「量表圖分析法」(scalogram analysis) (Guttman, 1944, 1950) 為基礎所發展出來的一種測量模式，但是它打破了量表圖技術只能用在線性作業網路的限制。

Perkins & Brutten(1990)認為：排序理論是一組用來決定某一技能或能力是否為學習其它更高階技能或能力時的「必要條件」(prerequisite) 之程序。

進一步說，排序理論是以受試者在施測時的「作業反應組型」來決定作業間的先後順序關係 (Airasian, 1975; Bart & Airasian, 1975)。譬如，若 task A 要成為學習 task B 時的必要條件，那麼在作業反應組型 (或稱為試題反應組型) 上，便不能有 (0, 1) 的反應組型出現；括號內的「 0 」是表示 task A 答錯了 (fail)。而括號內的「 1 」是表示 task B 答對了 (success)。因為，既然 task A 是學習 task B 時的必要條件，那麼就沒有理由說：簡單 (或低階) 的問題答錯了，卻能答對較為困難 (或高階) 的 task B 。

當然這牽涉到排序理論的幾個基本假設，茲分述如下：

- (1) 受試者對每個試題都作答。
- (2) 試題的記分方式必須是「二分化的」(dichotomous)。所謂二元化的記分，就是受試者若是在該題上答對，就登錄為「 1 」；若是答錯就登錄為「 0 」。而這種 0 與 1 的組合可分為四種，Airasian 等人將 (0, 1) 的反應組型稱之為「不確定的」(disconfirmatory)，

排序理論在概念結構評量上的應用

而 (1,0) 、 (1,1) 、 (0,0) 則稱之為「確定的」(confirmatory) 反應組型。

這樣的假設，否認了受試者在較簡單的 task A 答錯，是由於誤填或忘了填答的可能。至於 (0,1) 組型出現的另一原因 ---- 受試者是否是用猜的？排序理論則定了一個「容忍水準」(tolerance level) 來考驗整體誤差的機率大小 (Bart & Airasian, 1974) 。

在排序理論當中又包含兩大模式 (Krus, 1975) ，茲分述如下：

(1) 決定性的次序分析模式 (deterministic order-analytic model) 。其中又包括： A 模式 (nonmetric model) 和 B 模式 (a unit test space model) 。這兩個模式較適用於小樣本的情況，否則會產生複雜而難以解釋的排序。

(2) 機率性的次序分析模式 (probability order-analytic model)。又稱為 z 模式。此模式係以「不確定的」反應組型發生之機率是否超過 Z 值來決定第 i 題是否是第 j 題的必要條件。

一個教學活動是否達成教學目標，必須經由評量活動才能給予回饋；而依據「教學目標」來設計「教學活動」時，傳統的方法就是由教師依照自己的經驗，來做教學分析 (instructional analysis) ，以擬定教學活動的前後順序，但是教師若依照自己直覺所建構的次序，未必能符合學生發展的實際需求。認知心理學的興起在教學上特別強調先備知識的功能，係在於使每一次的學習，都能以先前的知識為基礎，並使得每一次學習成為下一次學習的先備知識，從而使學習活動更有意義。因此，透過排序理論的次序分析，來分析每個學習作業間的最佳學習次序，將可以提供教師一個更客觀的教學參照與評量的根據。

貳、方法

一、研究假設

本研究擬自樣本中分析辨認受試學生的幾種不同學習類型，再找出其中的兩組類型（即學習穩定型和學習異常型）學生，進行其概念結構差異之比較，並與教師的標準答案概念結構做比較，以確定這些不同學習類型學生間的概念結構及其與教師的標準答案型概念結構間之差異。

基於緒論之說明，本研究擬提出幾個待答問題的虛無假設如下：

1. 不同學習類型學生間的概念結構品質沒有顯著差異存在。
2. 不同學習類型學生間的概念結構相似性沒有顯著差異存在。
3. 半專家類型(即學習穩定型)學生比生手類型(即學習異常型)學生的概念結構，較不類似專家型教師(即標準答案)的概念結構。

二、研究樣本與工具

(一) 研究樣本

本研究擬以隨機抽樣方式，自臺北市十三個行政區一百多所國民小學中，隨機抽取一所學校，再從學校裡的五年級班級中隨機抽得一個班級學生為樣本，並以這些樣本的兩次數學科成就測驗的試題反應作為分析資料。

根據實際抽樣結果，計抽得文山區萬興國小五年二班 24 名同學，為本研究資料分析樣本。由於該校屬於小型學校，每個年級只有三班，每班人數均低於 30 人。因此，本研究針對該班全部學生進行研究。

(二) 研究工具

本研究擬採用的研究工具，計有二種：

1. 教師自編測驗

係採用受試者班級的數學科教師自編數學成就測驗。本測驗的用途，主要作為進行 S-P 表分析之用，以找出不同學習類型的學生。

2. 研究者彙編測驗

係由研究者根據已經過試題分析後的數學科成就測驗中，抽取出單元進度一致的題目所彙整而成的試題。本測驗的用途，主要是作為進行概念結構的排序分析之用，以找出不同學習類型學生的概念結構。本測驗的彙編，乃根據許天維（民 83）已進行標準化過的數學科成就測驗乙卷，抽取其原始試卷的前 21 題彙編而成。

收集到受試者在上述二種測驗的得分後，即根據第一份測驗結果進行 S-P 表分析，以找出不同學習類型的學生。接著，再根據第二份測驗結果進行概念結構圖的分析，以確定不同

學習類型學生的概念結構之品質良窳，並且分析排序不同學習類型學生的概念結構圖，並以二次式分配準則（余民寧，民 81；Yu, 1992）方法比較這兩類概念結構圖的相似性。

三、實施程序及資料分析

本研究擬以下列步驟及統計分析方法來進行：

- (1) 針對受試學生在第一份教師自編數學成就測驗上的資料，進行 S-P 表分析，利用 S-P 表中的學生注意係數和答對試題的百分比兩項指標，找出所有受試者的學習類型，並從中挑選出 A 組（即「學習穩定型」，在本研究中當成是「半專家」類型）和 C' 組（即「學習異常型」，在本研究中當成是「生手」類型）學生，作為後續分析和討論的主要對象。
- (2) 本研究所採用的第二份研究者彙編測驗，其試題與試題間具有階層的次序關係，亦即，學生若要能答對某些特定試題之前，他們必定能先答對某些其他試題（例如：能夠答對測量乘法概念的試題，必定也能夠答對測驗連加法概念的試題），換句話說，試題之間存在有隸屬的階層次序關係。因此，本研究聘請目前任教於萬興國小的數學教師（可以被視為學科專家）2 人，針對上述研究者彙編測驗，進行試題的概念結構分析，亦即是先分析每一個試題各是測量到什麼概念。
- (3) 接著，將這些受試者在第二份測驗試題上的原始作答反應組型，經過批改計分之後，轉化成一個 $N \times n$ 階的二元化（即以「0」表示答錯、「1」表示答對）作答反應資料矩陣（其中，N 表示學生數，n 表示試題數），簡稱為 A 矩陣。
- (4) 將上述研究者彙編測驗試題及其所測量到的概念二者，合併表示成一個 $n \times k$ 階的「試題概念表」（其中，「列」表示試題，共有 n 個試題；「行」表示概念，共有 k 個概念），它亦是一個二元化資料矩陣（不過，其中的元素「1」表示某試題可以測量到某個概念，「0」表示某試題無法測量到某個概念），並簡稱為 C 矩陣。
- (5) 再依序將每位學生的二元化作答反應向量，轉化成一個對角線矩陣（簡稱為 S 矩陣），並且與上述的試題概念表矩陣（即 C 矩陣）相乘，得到一個試題 x 概念的矩陣（簡稱為 X 矩陣）。X 矩陣的意思，即表示每位學生在本次研究者彙編的 k 個試題測驗上所測得的概念數。當他在某試題答錯（即以「0」表示）時，即反應出該試題所預擬測量的概念數全部落空，

他不具有該試題所欲測量的概念，所以無法從上述個別的試題概念矩陣中表示出來，因此，通常也是以「0」來表示；而當他在某試題答對（即以「1」表示）時，即反應出該試題所預擬測量的概念數全部可以被測量到，他擁有該試題所欲測量的概念，因此，可以從上述個別的試題概念矩陣中表示出來，通常是以「1」來表示。所以，上述個別的試題概念矩陣資料，正好反應出在某次測驗之後，個人所具有該測驗原先預擬測量的概念總數；概念總數愈多，即表示個人可能答對題數愈多，他的概念結構可能愈接近教師的標準答案之概念總數，反之，概念總數愈少，即表示個人可能答對題數愈少，他的概念結構可能愈偏離教師的標準答案之概念總數。

接著，再拿上述的 X 矩陣進行排序分析，其計算過程如下所述：

$$D = (X') \cdot (\sim X)$$

上述的 $\sim X$ ，即表示「否定矩陣」(negation matrix) 的意思，亦即將原矩陣中的「0」元素改變成「1」，而將「1」元素改變成「0」的意思；而 X' 是 X 矩陣的轉置矩陣，即將原矩陣的「列」改變成「行」，「行」改變成「列」的意思；而矩陣 D 即稱為「支配矩陣」(dominance matrix)，即經過排序分析後所顯示出概念與概念間的階層結構關係（即概念間具有支配性的前後次序或隸屬關係）之矩陣。每位學生各有一個「支配矩陣」，所以共有 24 個支配矩陣。

(6) 根據上述所計算出的支配矩陣資料，將每個元素值除以總概念數 21，獲得一個百分比值的矩陣，再以對角線為軸，凡對映於下三角形矩陣元素為 0 的上三角形矩陣元素為非 0 者，即可標上星號，其數值可以作為畫出概念結構圖中概念與概念間之距離指標 (distance indicators) 值，然後依序將轉換成百分比值的支配矩陣，其中的每個概念及其與其他概念間的距離指標值畫出。

重複上述步驟，依序將每位學生的概念支配矩陣計算出來，轉換成百分比值的矩陣後，即可畫出每位學生的概念結構圖。由於篇幅所限，本研究僅從中挑選學習穩定型和學習異常型的學生各一人作為代表，以概念結構圖作為分析比較的對象，其他學習類型學生的概念結構圖，則可以參考余民寧（民 85）研究中的詳細說明。

排序理論在概念結構評量上的應用

(7) 根據上述同樣過程，將教師的標準答案（即全對的二元化作答反應資料向量），轉化成一個對角線矩陣，並與試題概念表矩陣相乘，再進行排序分析，以求得一個標準的支配矩陣，並且畫出及確定標準的概念結構圖（在本研究中即視為「專家」類型）。

(8) 將這 24 名學生和標準答案的概念結構圖依據 Novak & Gowin (1984) 的記分原則，分別進行記分。並以標準答案的得分當作百分之百，作為比較的標準，再分別拿每位學生的概念結構圖得分與之相比，化成一個百分比值，當成是概念結構的品質指標，以比較兩組不同學習類型學生的概念結構間是否有差異存在，以回答假設一的研究問題。

(9) 然後，再分別將學生兩兩之間的支配矩陣進行二次式分配準則（余民寧，民 81；Yu, 1992）分析，並且也分別和（教師的）標準答案的支配矩陣進行二次式分配準則分析，以分別找出這兩組不同學習類型學生兩兩之間、及其與教師間概念結構圖的相似性，以回答並且考驗假設二和假設三的研究問題。

參、結果與討論

首先，根據 24 名受試學生在第一份教師自編數學成就測驗之結果，進行 S-P 表分析。茲依照 S-P 表輸出結果檔的排列方式，將每位學生學習類型的判斷資料呈現於表 1。

表 1 24 名受試學生的學習類型判斷資料

學生 座號	判斷 類型	學生 座號	判斷 類型	學生 座號	判斷 類型
04	A	15	A	02	B
19	A	12	A'	10	B
21	A'	24	A'	11	B
01	A'	03	A	18	B
09	A	13	A	22	B
16	A	14	A	05	B'
23	A'	25	A'	20	B
07	A	26	A	17	C'

從表 1 所示可知，除座號 05 號的學生屬於「欠缺充份型」，與座號 17 號的學生屬於「學習異常型」外，大致而言，這班學生的程度均相當整齊，大約為中上程度。

接著，再根據 24 名受試學生在第二份研究者彙編測驗之結果，進行排序分析，並繪製其概念結構圖。茲先將這份原始作答反應資料以二元化資料矩陣形式呈現，該矩陣是一個 24 (學生) \times 21(試題)的長方形矩陣資料（即前述 A 矩陣）；並將經過概念分析後的研究者彙編測驗之測量內容，表示成「試題 - 概念矩陣」（即前述 C 矩陣），該矩陣為一個 21(試題) \times 21(概念)的正方形矩陣資料。

再依據「方法」一節中所述的研究程序，計算每位學生與教師的概念結構圖資料。

一、不同學習類型學生的概念結構品質之比較

為了進一步了解這些概念結構圖之間的差異情形，乃仿照 Novak & Gowin (1984) 所述的「概念構圖的記分」原則：(1)關係（記 1 分）、(2)階層（記 5 分）、(3)交叉聯結（記 10 分）、及(4)舉例（記 1 分）等方式來記分。但由於本研究均採用相同的方法對每位學生的支配矩陣進行排序分析，所以所得的概念結構形狀大致上會呈現類似的架構，只是概念構圖法中的「交叉聯結」和「舉例」兩項記分項目在本研究中未曾出現，這是由於排序分析所得的概念結構圖與概念構圖法所得的概念圖，是使用不同的分析方法與建構方式的緣故。因此，本研究以標準答案的概念結構為準，採用修正的記分方式如下：

(1) 關係：分成正確的關係與錯誤的關係兩種。每位學生的概念結構圖與標準答案的概念結構相比，如果出現相同的關係，則表示該關係是正確的，每個正確的關係記 1 分；反之，如果出現不相同的關係，不論該關係是多餘的或欠缺的，均表示該關係是不正確、錯誤的、或異常的，則每個錯誤的關係扣 1 分。

(2) 階層：分成正確的階層與錯誤的階層兩種。每位學生的概念結構圖與標準答案的概念結構相比，如果出現相同的階層，則表示該圖的概念階層是正確的，每個正確的階層記 5 分；反之，如果出現不相同的階層，不論該階層是多餘的或欠缺的，均表示該階層是不正確、錯誤的、或異常的，則每個錯誤的階層扣 5 分。

茲依照上述記分原則，分別針對每位學生的概念結構圖加以記分，並將結果陳列於表 2。

表 2 各概念結構圖的記分

記 分 項 目						
座號 (類型)	正確 關係 $\times 1$	錯誤 關係 $\times (-1)$	正確 階層 $\times 5$	錯誤 階層 $\times (-5)$	總分	概念結 構得分 百分比
標準答案	23	0	4	0	43	100
01 (A' 型)	22	1	4	0	42	97.67
02 (B 型)	15	23	4	1	7	16.28
03 (A 型)	20	22	4	2	8	18.60
04 (A 型)	23	0	4	0	43	100
05 (B' 型)	18	33	4	1	0	0.0
07 (A 型)	23	4	4	0	39	90.70
09 (A 型)	23	3	4	0	40	93.02
10 (B 型)	19	23	4	1	11	25.58
11 (B 型)	16	43	4	1	-12	-27.91
12 (A' 型)	21	13	4	1	23	53.49
13 (A 型)	23	4	4	0	39	90.70
14 (A 型)	23	7	4	0	36	83.72
15 (A 型)	23	6	4	0	37	86.05
16 (A 型)	19	6	4	0	33	76.74
17 (C' 型)	14	66	4	1	-37	-86.05
18 (B 型)	15	34	4	0	1	2.33
19 (A 型)	23	2	4	0	41	95.35
20 (B 型)	14	40	4	1	-11	-25.58
21 (A' 型)	23	1	4	0	42	97.67
22 (B 型)	22	10	4	1	27	62.79
23 (A' 型)	23	9	4	1	29	67.44
24 (A' 型)	22	10	4	1	27	62.79
25 (A' 型)	22	10	4	1	27	62.79
26 (A 型)	16	31	4	0	5	11.63

註：百分比值愈小，表示其概念結構與標準答案愈不相似。

從表 2 可知，若以標準答案的概念結構圖之總分當成 100%，以作為進行比較的基準值，並以每位學生的概念結構圖總分除以此基準值，則可以分別求出如表 2 中其他學生之概念結構得分百分比。

這些概念結構得分百分比，反映出不同學習類型學生在本次測驗中所顯示的不同學習成果。由表 2 可知，愈是偏向「學習穩定型」(A 型)的學生，其概念結構愈與標準答案的概念結構相似；反之，愈是偏向「學習異常型」(C' 型)的學生，其概念結構愈不似標準答案的概念結構，甚至呈現負值的分數，表示其概念結構已明顯偏離標準答案的概念結構。

由表 2 的統計資料可以看出，與標準答案的概念結構相比，其百分比值的大小排列，似乎可以歸納出一個概括的規則類型，亦即是： $A(\text{學習穩定型}) \geq A'(\text{粗心大意型}) > B(\text{努力不足型}) \geq B'(\text{欠缺充份型}) > C'(\text{學習異常型})$ ；若該百分比值能夠反應出與標準答案的概念結構間之相似(或優良)程度的話，則可以由此驗證假設一：「學習穩定型」學生比「學習異常型」學生的概念結構品質優良的初步結果。

二、不同學習類型學生的概念結構圖之相似性比較

上述表 2 所示，僅呈現每位學生的概念結構圖得分百分比資料，雖然可以由表中明顯看出「學習穩定型」學生比「學習異常型」學生的概念結構品質優良，但仍缺乏客觀的統計考驗證據，因為屬於「學習異常型」者僅有一人，樣本數過少，並且缺乏變異數可茲分析，因此，表 2 所示結果缺乏客觀的統計考驗證據，實為一大缺憾。

為了彌補上述的缺憾，本研究乃進一步採用余民寧（民 81）和 Yu (1992)所提的「二次式分配準則」(quadratic assignment paradigm，簡稱 QAP)方法，來進行統計分析與考驗，以補實上述表 2 資料之不足。「二次式分配準則」技術是一種資料分析的通式 (general form)，它可以計算兩個不同矩陣中對映元素分佈所形成的組合次數分配間之相關係數。當該相關係數高時，即表示該二矩陣元素的分佈趨勢極為相似或一致；反之，當該相關係數低時，即表示該二矩陣元素的分佈趨勢極不相似或不一致。此外，本技術同時會提供 z 值，用來考驗該相關係數是否達到 $\alpha = .05$ 或 $\alpha = .01$ 的顯著水準，以[†] 為研究者判斷這兩個矩陣是否相似；但是，當矩陣的元素較多時，兩個矩陣相對映元素的組合次數將會相當龐大，因此使得 z 值

排序理論在概念結構評量上的應用

趨大，可能造成即使是很小的相關係數值，也都很容易達到顯著水準。雖然，二次式分配準則技術有這種現象的缺失，但它背後的方法學理論卻比較嚴謹、合理，尤其可以適用在當樣本人數僅有一人時的矩陣資料。

因此，本研究引用這 24 名受試學生在研究者彙編測驗之結果資料，將每位學生的作答反應組型轉換成一個對角線矩陣，再乘以研究者彙編測驗的「試題概念矩陣」，即可獲得一個「個別的試題概念矩陣」（即 X 矩陣），以作為「二次式分配準則」分析法的基本輸入矩陣資料。在計算過程中，我們採用兩兩的 X 矩陣進行分析，其中一個矩陣當作結構矩陣（structure matrix），另一個當作資料矩陣（data matrix），即可求出所要計算的兩兩矩陣間的相關係數值及其 z 值，當 z 值大於 1.96 時，即表示該相關係數達到 $\alpha = .05$ 的顯著水準，當 z 值大於 2.58 時，即表示該相關係數達到 $\alpha = .01$ 的顯著水準。由於原有 24 名學生，兩兩成對分析的話，共可計算出 276 個相關係數值及其 z 值，並且每個 z 值均大於 2.58，顯示每個相關係數值均達顯著水準，均是有意義的存在。

經由整理，我們可以將每類型間之相關係數的最大值與最小值，歸納呈現於表 3 中：

表 3 不同學習類型學生間相關係數的最大值與最小值

類型	A	A'	B	B'	C'
A	.99 .70	.97 .70	.96 .45	.83 .68	.69 .51
A'		.99 .70	.87 .55	.77 .60	.66 .51
B			.93 .39	.85 .51	.70 .28
B'					.54

註：表中元素上列為最大值，下列為最小值。

由表 3 可以很明顯看出，不同學習類型學生間的概念結構是不完全相似的。相較之下，學習類型愈相近者，其概念結構間有較相似之趨勢；而學習類型愈不相近者，其概念結構間則有愈不相似之趨勢。

三、不同學習類型學生與教師間概念結構圖之相似性比較

茲以標準答案的 X 矩陣當成結構矩陣，以每位學生的 X 矩陣當成資料矩陣，並進行二次式分配準則的分析，以分別求出標準答案的概念結構與每位學生的概念結構間的相關係數值及其 z 值，並將結果呈現於表 4。

表 4 標準答案與每位學生概念結構間之相似性比較

學生 座號	學習 類型	相關 係數	z 值	學生 座號	學習 類型	相關 係數	z 值
04	A	1.00	14.30	13	A	.87	12.20
19	A	.95	13.54	14	A	.83	11.60
21	A'	.96	13.66	25	A'	.97	13.80
01	A'	.94	13.35	26	A	.91	12.82
09	A	.90	12.77	02	B	.81	11.16
16	A	.94	13.35	10	B	.81	11.38
23	A'	.93	13.15	11	B	.83	11.51
07	A	.92	13.03	18	B	.81	11.38
15	A	.90	12.71	22	B	.84	11.68
12	A'	.94	13.26	05	B'	.75	10.40
24	A'	.92	12.97	20	B	.62	8.60
03	A	.89	12.54	17	C'	.62	8.60

註：z 值大於 2.58，表示達到 $\alpha = .01$ 的顯著水準。

由表 4 所示可知，所有的相關係數值均達到 $\alpha = .01$ 的顯著水準並且各不相同，顯示不同學習類型學生的概念結構與(教師的)標準答案的概念結構之相關程度各不相同。但是，很明顯的可以看出來，「學習穩定型」(A 型)學生和「粗心大意型」(A' 型)學生與標準答案間之相關係數值最高，其次才是「努力不足型」(B 型)學生和「欠缺充份型」(B' 型)學生，而「學習異常型」學生與標準答案間之相關係數值則是最低(相關係數僅 .62)。由此可見，半專家型(即「學習穩定型」)學生比生手型(即「學習異常型」)學生的概念結構，更接近專家型(即教師的標準答案)教師的概念結構，亦即，「學習穩定型」學生比「學習異常型」學生的概念結構更接近教師的概念結構，或說是與教師的概念結構更相似。

四、綜合討論

若觀察從全對到全錯的作答反應組型中，我們可以分別排列出每個概念結構圖的不同形狀，從最完整的概念結構圖到最不完整（或說是異常的）概念結構圖(參見圖 1 至圖 6 所示)，並拿它們與標準答案的概念結構圖(即圖 1 所示者，由於座號 04 學生的答案完全正確之故)做對映比較 (mapping comparison)，以進一步找出個別的概念結構圖與標準答案的概念結構圖間不相同（或差異）之處(本例圖以粗黑線條來顯示)，這些不相同（或差異）之處常會出現在「多餘的聯結關係」或「欠缺的聯結關係」上，也就是說，與標準答案的概念結構圖做比較，凡學生的概念結構圖中，有「多餘的聯結線」或「缺少的聯結線」，即表示該地方的概念結構並非本次測驗所應有的表現，換句話說，該地方即是產生學習缺陷的所在。當學習缺陷之處愈多時，除了表示該學生的答對題數較少，原始得分較低，其概念結構是呈現愈偏離理想的概念結構形狀，該生愈需要列入優先進行補救教學的對象之一，而實施補救教學的重點，即針對上述發現的學習缺陷所在來進行的。

由表 2 到表 4 的結果顯示，「學習穩定型」學生的概念結構品質很明顯地比「學習異常型」學生的概念結構品質為優，這點可從表 2 的數據及其概念結構圖的比較中看出這個趨勢。這是由於前者以一種比較精簡、有效率的方式來組織自己的概念，所以所呈現出來的概念結構自然就很接近標準答案的概念結構；而後者以一種頗為紊亂、不規則性的方式來組織自己的概念，所以所呈現出來的概念結構，很自然地就偏離標準答案的概念結構。

其次，每位學生的概念結構形狀，隨著其不同的學習類型而有不同的變化。其中，學習類型愈接近者，其概念結構就愈相似；學習類型愈不相同者，其概念結構則就愈不一致。這是由於學習類型近似的學生，他們多半使用較為類似的學習方法學習，採用較為類似的方式來組織自己的概念，因此，獲得較為一致的概念結構圖；而學習類型較不相同者，他們多半使用差異頗大、且比較無效率、無結構性的方式來組織自己的概念，因此，所獲得的概念結構圖則較為凌亂、不規則、且不一致。

而與標準答案的概念結構圖比較起來，「學習穩定型」學生的學習成果（可以被看成是「半專家」的程度），他的概念結構當然會比「學習異常型」學生（可以被看成是「生手」的

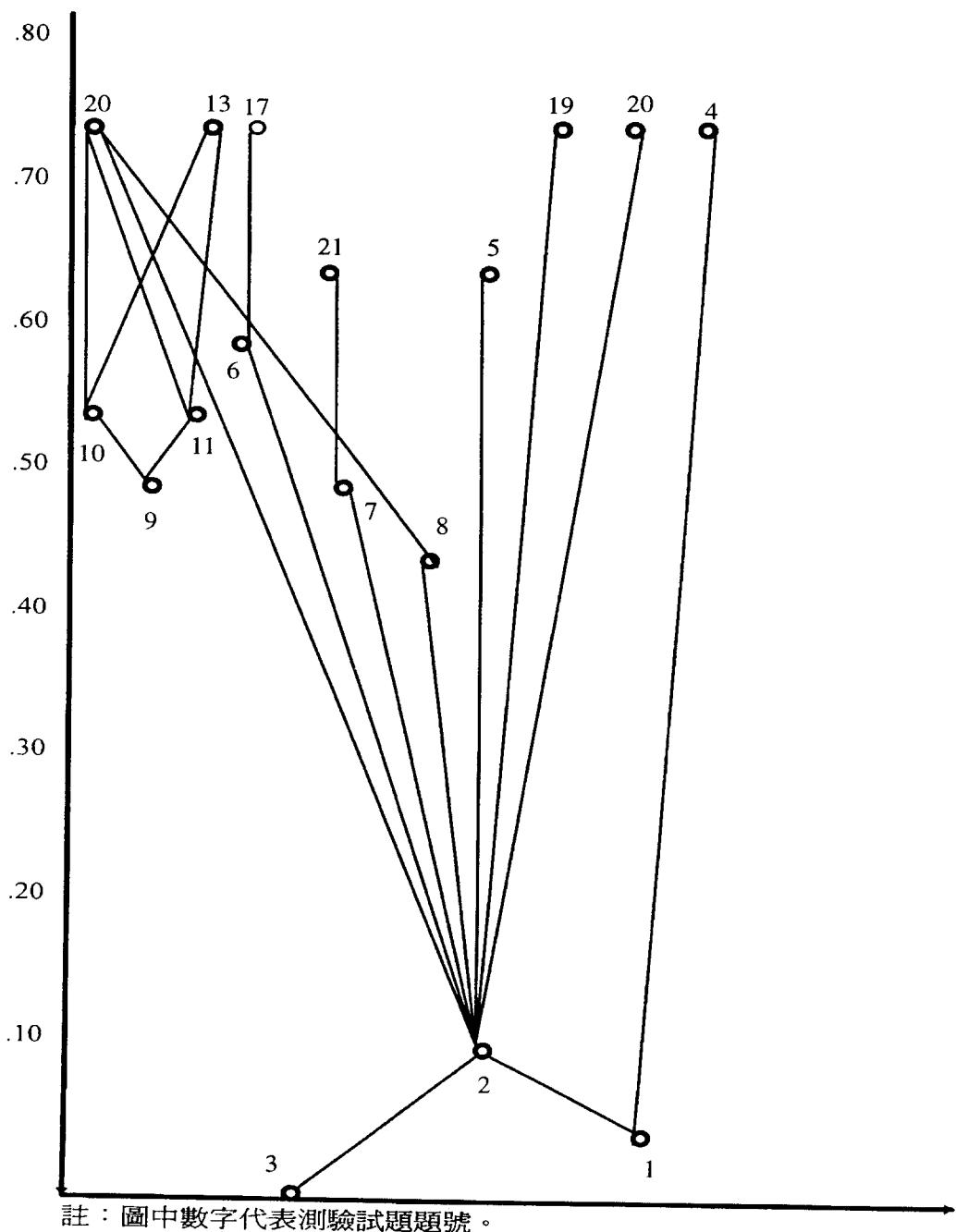


圖 1 座號 04 學生的概念結構圖(A型)

排序理論在概念結構評量上的應用

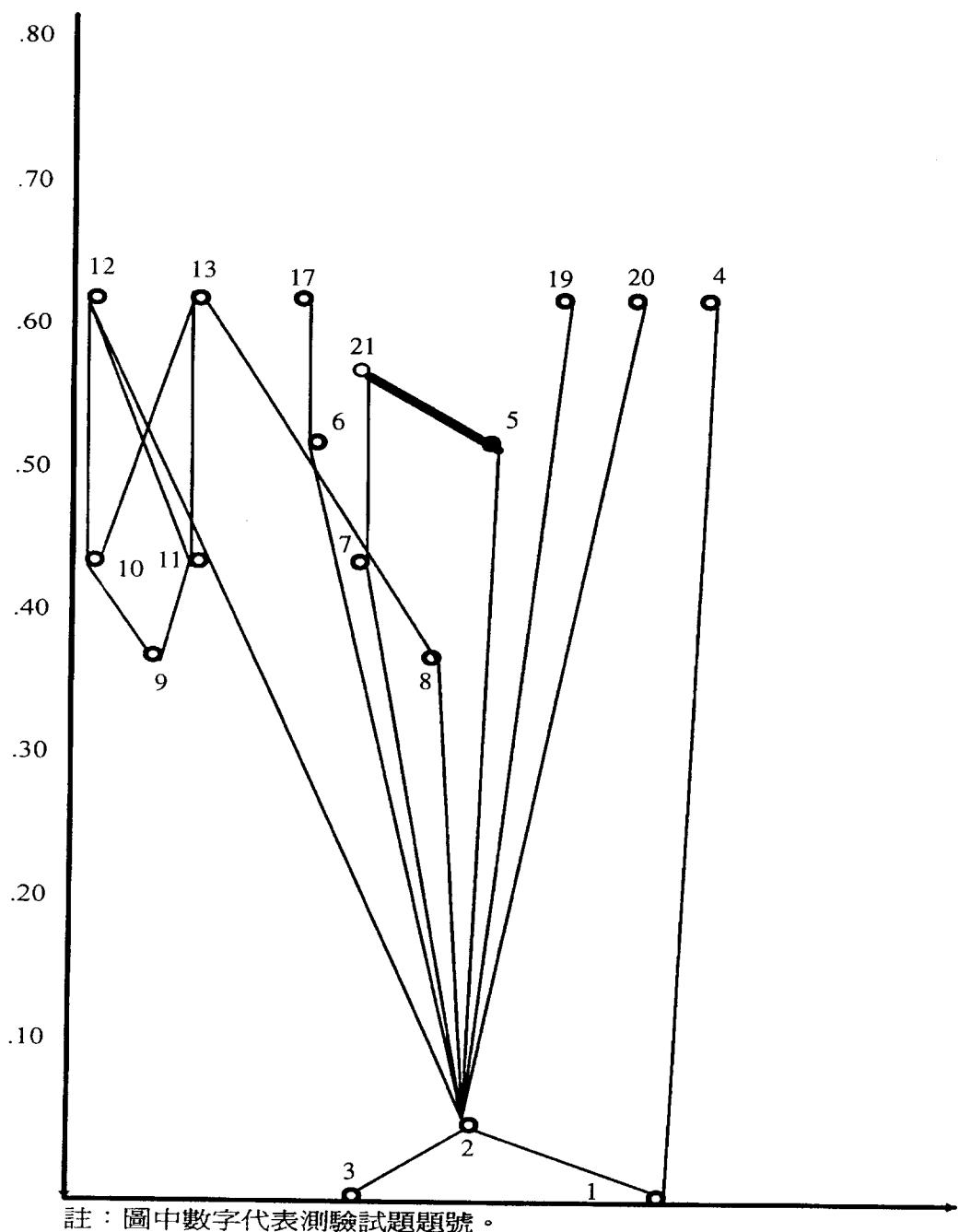


圖 2 座號 01 學生的概念結構圖(A'型)

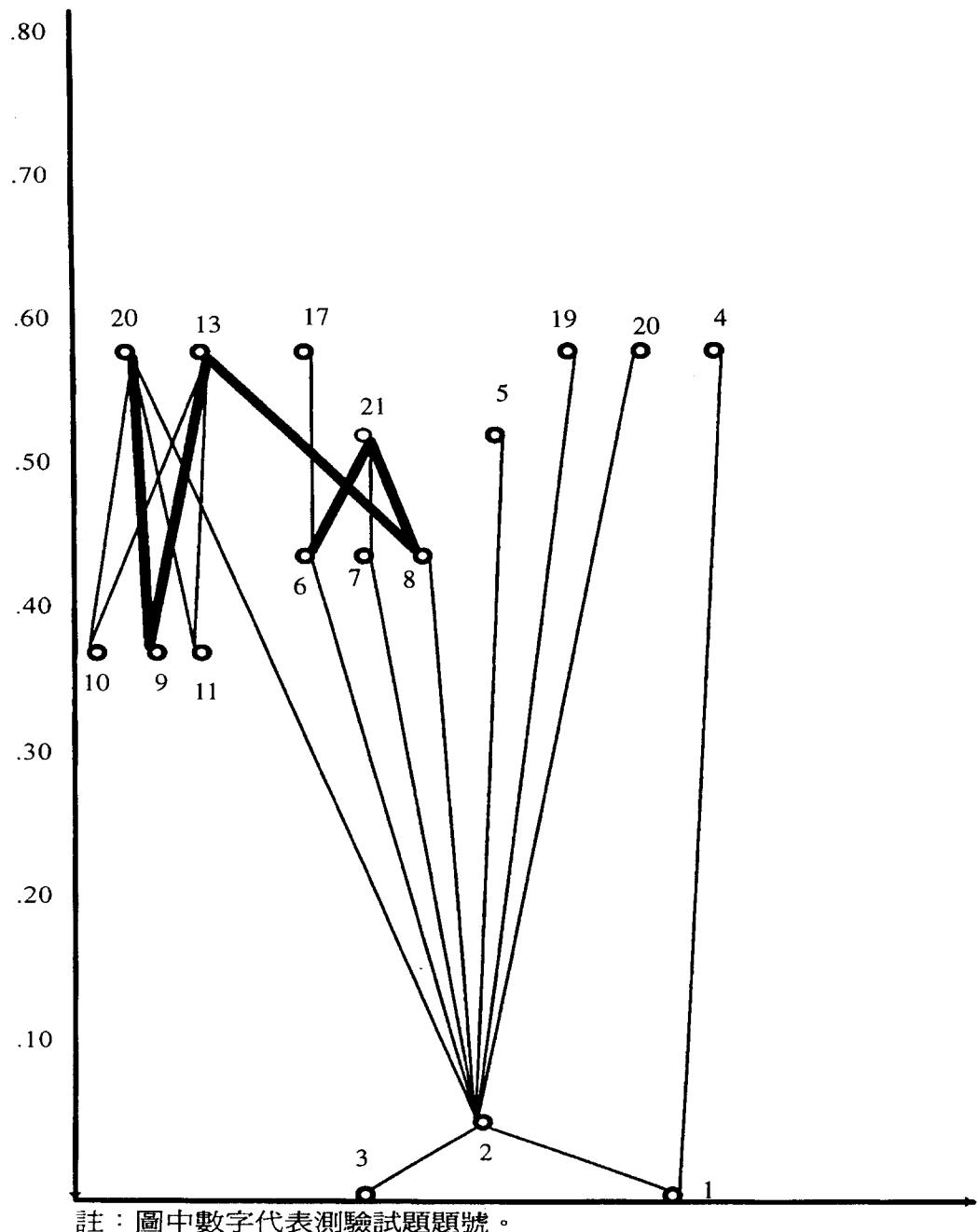
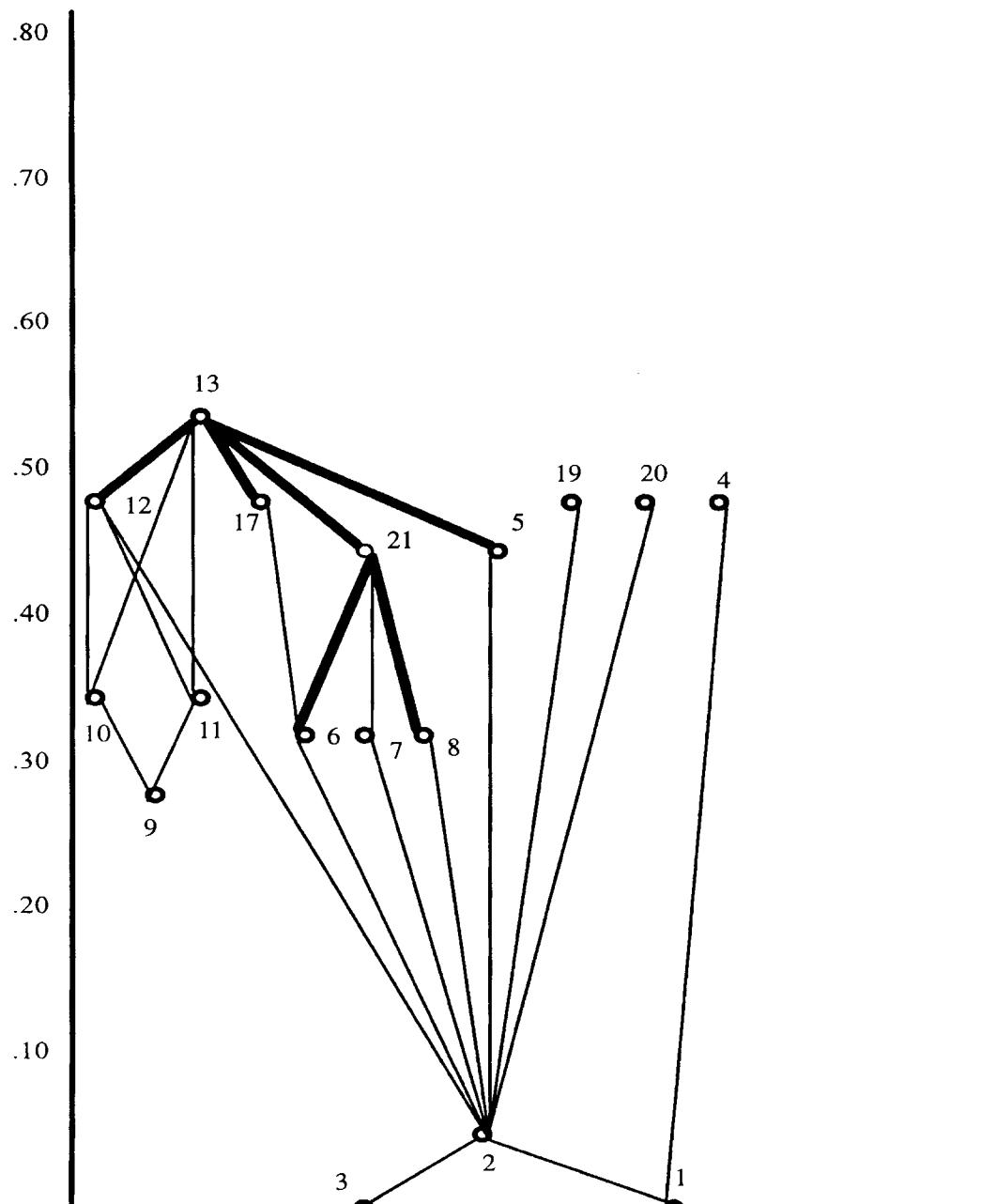


圖 3 座號 15 學生的概念結構圖(A型)

排序理論在概念結構評量上的應用



註：圖中數字代表測驗試題題號。

圖 4 座號 22 學生的概念結構圖(B型)

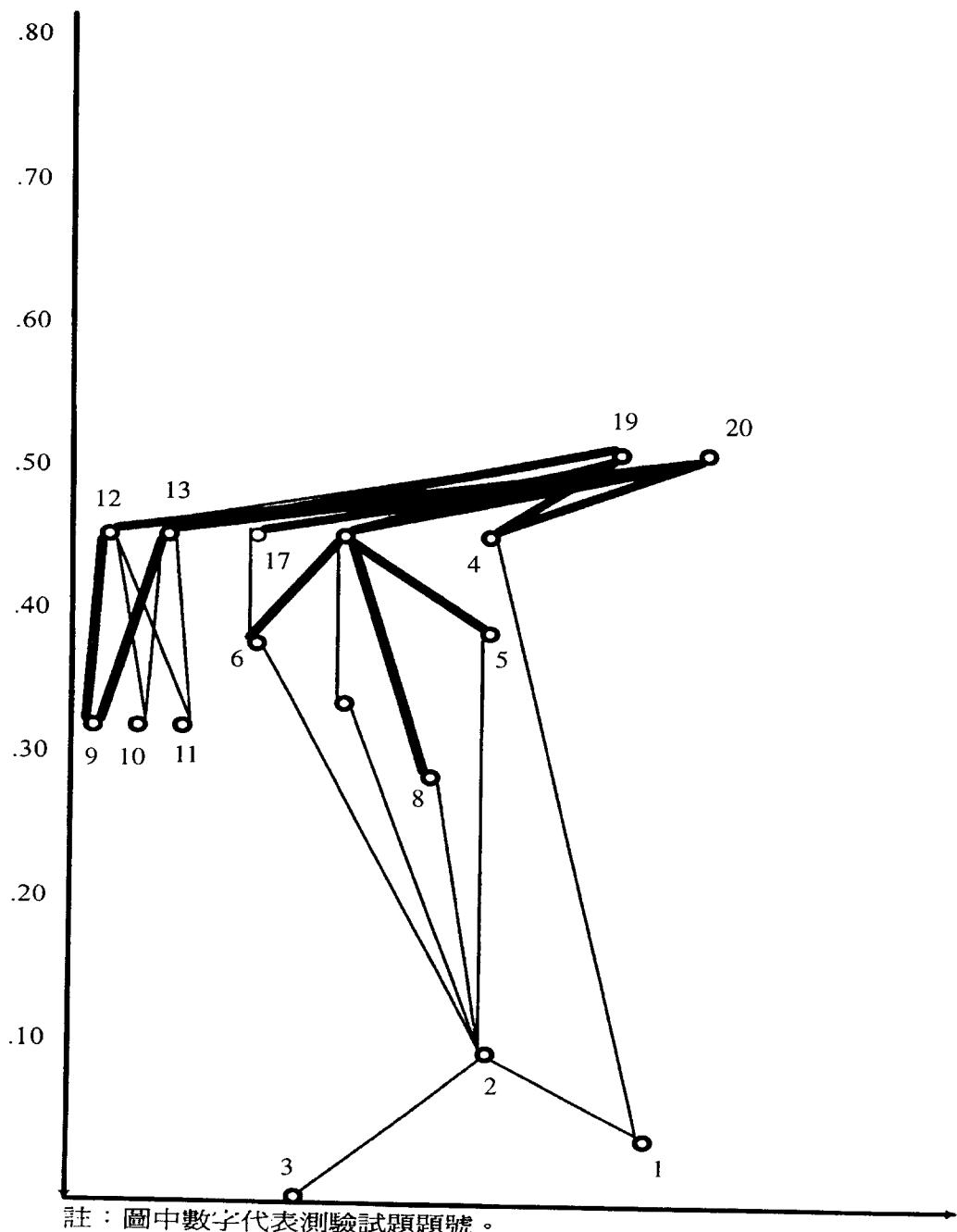


圖 5 座號 05 學生的概念結構圖(B'型)

排序理論在概念結構評量上的應用

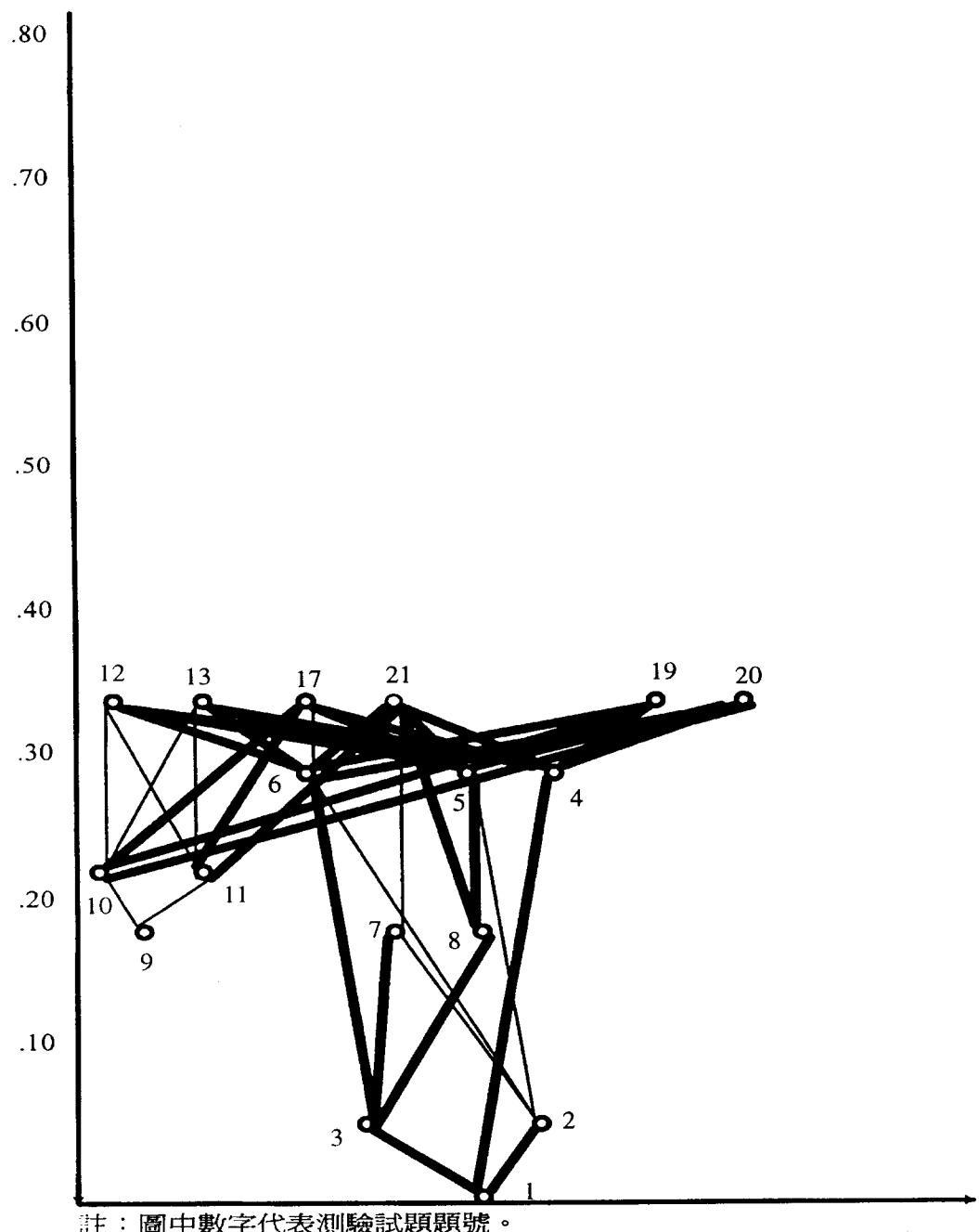


圖 6 座號 17 學生的概念結構圖(C'型)

程度）的概念結構，還要接近標準答案的概念結構（可以被看成是「專家」的程度），這是因為他用來組織自己概念的方式比後者較為精簡、有效率，並且更接近或類似教師所使用的方法之故。

肆、結論與涵義

一、結論

傳統的教學評量方法，過度重視紙筆測驗方式的總結性評量，殊少創新對評量資訊的獲取，甚至對評量資訊背後的意義進行分析與檢討；但本研究所企圖開發的評量技術卻不同，它不僅能夠達到傳統的評量目的，更能夠進一步告訴教師：哪一位學生在哪些概念上的學習是不正確的？哪些概念間的聯結是多餘的或是欠缺的？以及如何透過網路結構的表徵方式來增進對學習結果的了解等，這些企圖才是達到改善學習成果或進行補救教學的真正目的。

因此，本研究針對研究目的，即嘗試發展一套評量技術，以用來協助教師了解學生的學習困難與學習缺陷所在，進而研擬補救教學策略，增進學生的學習成果。

因此，本套評量技術的使用者，只要遵守下列評量技術的實施過程，即可達到分析教學評量資訊的目的，增進對學生學習成果的了解，進而提出適當的補救措施：

- (1) 針對教師自編成就測驗進行概念分析，列出自編測驗的每道試題所能測量到的概念，將它列成一份「試題概念矩陣」表。
- (2) 開始進行教師自編成就測驗的評量，並將學生的原始作答反應資料，經批改後，轉換成二元化的作答反應資料矩陣。
- (3) 依序將每位學生的二元化作答反應資料向量，轉變成一個二元化的「對角線矩陣」。
- (4) 將每位學生的二元化「對角線矩陣」乘上「試題概念矩陣」，以獲得一個「個別的試題概念矩陣」。
- (5) 將「個別的試題概念矩陣」代入公式，求出每位學生的支配矩陣，並將支配矩陣中的元素除以本測驗的概念總數，以轉換成一個百分比值的距離指標矩陣。
- (6) 根據上述的距離指標矩陣資料，畫出每位學生的概念結構圖。

(7) 依照上述(3)到(6)的步驟，畫出標準答案的概念結構圖。

(8) 將每位學生的概念結構圖，拿來與標準答案的概念結構圖做對映比較，以檢討及分析每位學生在概念結構上的學習缺陷。

本研究復根據上述評量技術的發展過程，擬定三項研究問題進行探討，大致上，均獲得圓滿的證實。茲歸納本研究的結論如下：

1. 不同學習類型學生的概念結構品質確實有所不同。其中，「學習穩定型」學生比「學習異常型」學生的概念結構品質較為優良。

2. 不同學習類型學生的概念結構圖並不完全相似。相較之下，學習類型愈相近者（例如：「學習穩定型」（即A型）與「粗心大意型」（即A'型）、或「努力不足型」（即B型）與「欠缺充份型」（即B'），其概念結構較為相似；學習類型愈不相近者（例如：A（或A'）型與B（或B'）型、或A（或A'）型與「學習異常型」（即C'型）、或B（或B'）型與C'型），其概念結構則較不相似。

3. 不同學習類型學生的概念結構與（教師的）標準答案的概念結構之相似程度並不相同。其中，半專家型（即「學習穩定型」）學生比生手型（即「學習異常型」）學生的概念結構，更接近專家型（即教師的標準答案）教師的概念結構，或者說是與教師的概念結構更相似。

二、對教育的涵義

綜合上述，本研究的初步成果至少對教育具有下列幾項涵義：

（一）知識乃是一個個別化的概念組織整體

傳統評量法之所以為人所詬病的地方，主要是來自於下列的兩個假定：

1. 專家與生手之間的差異係來自於知識「量」的不同。因此表現在教學評量上就是以提昇學生的考試成績為目的，當成績越高就表示學生的知識量越多。但是這樣的評量方式可能只對於每次都能拿高分的學生才有意義，對於那些真正需要幫助的低成就學生卻無法發揮認知診斷的功能。

2. 此外，傳統的評量法是以個別的試題作為分析的單位。評量的結果所能提供給老師

的訊息不是全對就是全錯，因而忽略了部份學生在某些概念的學習上並非是「全有或全無」的情況；甚至，學生的問題並不是在於概念的有無，而是在概念與概念之間的不當聯結。而概念之間的整體聯結情況如何？卻是傳統評量方法所無法勝任表徵的地方。

認知心理學家認為：任何一個概念都無法單獨存在，而是與其他相關的概念一起組織成一個整體的知識結構而保留在記憶當中(Ault, 1985)。因此，本研究所提出的分析技術不僅是心理計量理論的更迭，更是一種對知識論上的典範變遷的呼應；不是一種更新的學生分類技術，而是一種幫助學生、協助老師的認知診斷技術。

(二) 評量應是一種概念結構的測量

教學評量不應該只重視總結性評量而已，反而應該重視平時即需進行的形成性評量，而形成性評量也不光是以獲得一個足以代表精熟程度的總分為滿足，而是應該進一步著重學生在形成性測驗中所呈現的概念結構，以便找出其中的可能學習缺陷所在，對症下藥提出匡正與補救教學的措施。因此，在評量技術上，教師應該先分析自編成就測驗到底在測量什麼樣的概念？接著，並從學生的作答反應組型資料中，進一步分析排序，以找出學生的概念結構，再進一步檢討可能的學習缺陷所在，並針對學習缺陷謀求對症下藥的改進之道。

(三) 具體的概念表徵有助於補救教學的進行

傳統式的教學評量僅能獲得一個單一數據（通常都是以測驗總分來代表），教師也僅能憑此數據來判斷學生學習的好壞，他無從得知學生到底是在哪些概念結構上有學習缺陷？評量結果能否反應出學生的學習品質？或應該從何處實施補救教學？但是，從本研究中可知，概念結構圖是一種優良的圖解式知識表徵方法，它不僅可以將每次的評量結果，以圖畫式的概念結構圖，把學生在本次評量所得的概念結構表徵出來：一來，方便教師按圖索驥，診斷出每位學生的概念結構中有出差錯的地方，以謀求補救措施；二來，更方便了解學生的學習成果與教師期望的（或理想的）成果間之差距，以尋求縮短差距的有效教學方法。同時，概念結構圖亦可以作為一種教學與學習或學習與評量間的溝通橋樑，達到表徵知識時對認知資源需求的簡潔性要求，並且促進溝通的正確性。

(四) 發展認知診斷評量技術以作為診斷學習缺陷的工具

本研究試圖發展一種具有診斷分析功能的評量技術，是一套客觀的評量分析方法，可以作為診斷或評定學生概念結構與標準概念結構間是否一致或相符合的工具，亦是未來心理計量學界和教育學界突破教學瓶頸的希望所在。雖然，本研究目前的發展已略見成果，但仍未到達技術成熟的階段，在可見的未來尚且需要結合各種教學法、教學策略、或認知科學等其他領域知識的發展與科際整合，才能形成一套確實可行的認知診斷評量技術，以作為學校教育診斷學生學習缺陷的有效工具。

三、本研究的限制

雖然，本研究初步證實所研提的診斷評量技術，具有方法論與實用方面的價值，教師只要遵照上述評量技術的實施程序，即可從自編成就測驗中實施這套概念結構的分析技術，以獲致學習診斷的訊息。但是，這並不表示它就是一套完美無缺的評量技術；其實，本研究尚有許多技術瓶頸尚待克服，例如：

1. 針對教師自編測驗所進行的概念分析工作，尚缺乏標準化的程序。
2. 尚缺乏一套可以連貫執行的電腦程式，以節省分析和繪製概念結構圖的時間，便利推廣使用。
3. 尚缺乏對概念節點間的聯結關係（包含聯結方向與聯結距離之長短等）所具之意義，適用一套明確且強而有力的詮釋方法。
4. 尚缺乏對概念結構圖進行客觀且標準化的計分程序，以作為傳統紙筆評量的另類評量替代品。
5. 本分析技術尚缺乏經由大樣本施測所建立的外在推論效度。

也就是說，本分析技術目前尚處於技術開發與測試的階段，尚未臻致技術成熟的階段。上述尚待克服的技術瓶頸，恐怕還是需要學術界的通力合作發展，才能建立起更嚴謹的實證理論證據，達到便利分析與推廣應用的成熟發展階段。

參考書目

中文部份

- 朱智賢（1989）。心理學大辭典。北平：北京師範大學。
- 余民寧（民 81）。二次式分配準則在估計再測信度上的應用。政大教育與心理研究，15 期，21-38 頁。
- 余民寧（民 85）。不同學習類型學生的概念結構之評量研究。行政院國家科學委員會專題研究計畫成果報告(NSC-85-2413-H-004-007)。
- 余民寧（民 86a）。教育測驗與評量：成就測驗與教學評量。台北：心理。
- 余民寧（民 86b）。有意義的學習：概念構圖之研究。台北：商鼎。
- 林清山（民 80）。教育心理學 -- 認知取向。臺北：遠流。
- 吳信義、吳錫修(民 86)。學生問題圖表分析應用軟體之設計。行政院國家科學委員會專題研究計畫成果報告(NSC-86-2511-S-252-001-CL)。
- 許天維（民 83）。數學試題分析法。高雄：大漢唐。
- 黃文俊（民 83）。國中生物理壓力迷思概念及概念改變教學可行性之研究。淡江大學教育資料科學研究所碩士論文。
- 陳嘉成（民 85）。以概念構圖為學習策略之教學對小學生自然科學習之成效研究。國立政治大學教育研究所碩士論文。
- 郭姿媚（民 75）。知識結構的計量分析。國立臺灣大學心理研究所碩士論文。
- 張春興（民 78）。張氏心理學辭典。臺北：東華。

外文部分

- Airasian, P. W., & Bart, W. M. (1973). Ordering theory: A new and useful measurement model. *Educational Technology*, 5(13), 13-24.
- Airasian, P. W. (1975). A method for validating sequential instructional hierarchies. In D. J. Krus, W. M. Bart, & P. W. Airasian (Eds.), *Ordering theory and methods*. MI: Theta Press.
- Airasian, P. W., & Bart, W. M. (1975). An ordering-theoretic measurement model. In D. J. Krus, W. M. Bart, & P. W. Airasian (Eds.), *Ordering theory and methods*. MI: Theta Press.
- Airasian, P. W., & Bart, W. M. (1975). Validating a priori instructional hierarchies. *Journal of Educational Measurement*, 12(3), 163-173.
- Ault, C. R. (1985). Concept mapping as a study strategy in earth science. *Journal of College Science Teaching*, 15, 38-44.
- Bart, W. M., & Airasian, P. W. (1975). Determination of the ordering among the seven Piagetian tasks by an ordering theoretic method. *Journal of Educational Psychology*, 66(2), 277-284.
- Bart, W. M., & Krus, D. J. (1973). An ordering-theoretic method to determine hierarchies among items. *Journal of Educational and Psychological Measurement*, 33, 291-300.

排序理論在概念結構評量上的應用

- Behr, M. J., Goldin, G. A., Pierce, J. W., & Threadgill-Soweder, J. (1985). Knowledge representation: A foundation for education research and practice. In M. J. Behr, (Ed.). *Center for the study of learning* (20-27). Washington, DC: National Institute of Education.
- Black, P. J. (1993). *Assessment and feedback in science education*. Paper presented at the International Conference "Science education in developing countries: From theory to practice". Jerusalem, Israel, January 3-7.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121-152.
- Goldsmith, T. E., Johnson, P. J., & Acton, W. H. (1991). Assessing structural knowledge. *Journal of Educational Psychology*, 83, 88-96.
- Guttman, L. A. (1944). A basis for scaling qualitative data. *American Sociological Review*, 9, 13.
- Guttman, L. A. (1950). A basis for Scalogram Analysis. In S. A. Stouffer et al (EDs.), *Studies in social Psychology in World War II, Volume 4, Measurement and Prediction*. Princeton: Princeton University Press.
- Krus, D. J. (1975). Construction of logical item hierarchies. In D. J. Krus, W. M. Bart, & P. W. Airasian (Eds.), *Ordering theory and methods*. MI: Theta Press.
- Larkin, J., Mcdermont, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, 208, 1335-1342.
- Linn, R. L. (1993). Educational assessment: Expanded expectations and challenges. *Educational Evaluation and Policy Analysis*, 15(1), 1-16.
- Mayer, R. E. (1981). Frequency norms and structural analysis of algebra story problems into families categories and templates. *Instructional Science*, 10, 135-175.
- Mayer, R. E. (1987). *Educational psychology: A cognitive approach*. Boston, MA: Little Brown and Company.
- Nichols, C. (1994). A framework for developing cognitively diagnostic assessment. *Review of Educational Research*, 64, 575-603.
- Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. London: Cambridge University Press.
- Perkin, K., & Brutten, C. (1990). In S. A. Biggs, & T. Bullock (Eds), *Innovative learning strategies*. (ERIC Document Reproduction Service. NO. ED 321-239).
- Sato, T. (1969). *A method of analyzing data gathered by the Response Analyzer for diagnosis of student performance and the quality of instructional sequence*. Proceedings of IECE of Japan annual conference S12-1.
- Sato, T. (1971). *Analysis of student's performance score data*. In K. Hirata, Tokyo: Kyoiku-Kogakusha.
- Shavelson, R. J. (1972). Some aspects of the correspondence between content structure and cognitive structure in physics instruction. *Journal of Educational Psychology*, 63(3), 225-234.
- Shavelson, R. J. (1974). Methods for examining representations of subject-matter structure in a student's memory. *Journal of Research in Science Teaching*, 11, 231-249.
- Solso, R. (1995). *Cognitive Psychology* (4nd ed.). Boston, MA: Allyn & Bacon.

- Solso, R. (1995). *Cognitive Psychology (4nd ed.)*. Boston, MA: Allyn & Bacon.
- Tatsuoka, K. K. (1983). Rule space: An approach of dealing with misconception based on item response theory. *Journal of Educational Measurement*, 20, 345-354.
- Tatsuoka, K. K. (1986). Diagnosing cognitive errors: Statistical pattern classification and recognition approach. *Behaviormetrica*, 19, 73-86.
- Tatsuoka, K. K., & Tatsuoka, M. M. (1987). Bug distribution and pattern classification. *Psychometrika*, 52, 193-206.
- Tatsuoka, M. M., & Tatsuoka, K. K. (1988). Rule space. In S. Kotz, & N. L. Johnson, (Eds.), *Encyclopedia of Statistical Science*, 8, 217-220. New York: John Wiley & Sons.
- Yu, M. (1992). A comparison of two proximity matrices. *Journal of National Chengchi University*, 65, 387-408.