國　立　臺　南　師　範　學　院
「南師學報」第 37 卷第二期教育類（民國 92 年）：67　97

認知診斷評量的探究

涂金堂

國立高雄師範大學教育學程中心

摘　要

教育評量的目的，除了測量出學習者的學習現況外，同時也應該提供學習者學習成敗的診斷訊息，以利教學者根據診斷訊息，進行有效的補救教學。國外已有許多教育學者主張，教育評量的實施，應該以認知心理學對學習歷程的研究成果，作為編製測驗的實質理論依據。此種將教育評量與認知心理學的實質理論相結合的評量方式，即是所謂的認知診斷評量(cognitively diagnostic assessment)。

鑑於國內在教育評量的實務上，對於認知診斷評量，仍未有充分討論。因此，本文旨在探究認知診斷評量的理論基礎，以及分析認知診斷測驗的編製歷程，並且，舉認知診斷評量的應用實例，希望讓讀者對認知診斷評量有較清楚的瞭解。

關鍵字：認知診斷評量、反應組型、知識空間

壹、緒論

近幾年來，國內教育界對於單一紙筆評量的強烈批判，促使目前中小學的教育評量，逐漸採用各種不同的評量方式，來評量學生的學習成效。雖然這一波多元評量的風潮，解放了制式的評量內容與評量方法。然而透過評量歷程，提供老師足夠訊息，用來診斷學習者的學習成效，是教育評量相當重要的一個目標。以此重要目標，來檢視目前多元評量的實施，顯然多元評量尚未達到此一重要的評量目標。若探究其中的緣由，可能是目前多元評量所隱含的測驗與評量理論，並未結合當前認知心理學與教學心理學的研究成果。Mislevy(1993)即認為今日的教育評量面臨一個可能危及基礎的危機，這個危機是隱藏在標準測驗理論(standard test theory)關於人類能力的觀點，並無法配合認知心理學與教育心理學的快速發展。

傳統評量主要的目的在於選擇的功能，其測驗理論的建構，主要是想估計出個人在某種潛在變項中的位置。在古典測驗理論中，此種潛在變項即是所謂的真分數，而在試題反應理論中，即是單向度的潛在特質。傳統評量主要是根據邏輯分類與內容細目來進行評量設計，但這些評量設計缺乏對該領域知識結構與歷程的詳細描述(Nichols, 1994)。例如 Bloom et al.(1956)提出教學目標可分認知、情意和動作技能等三類領域，其中認知領域的教學目標可分成六個階層：知識、理解、應用、分析、綜合與評鑑。

傳統測驗的編製常根據上述六個認知教學目標，然而以這種方式所評量出的結果，只是一種統括性的描述，並無法顯現受試者在該領域的知識結構。Anastasi(1967)即認為傳統測驗似乎比較強調統計技術，而較忽略所欲評量的能力或特質，是否具有心理學上的意義。傳統教育測驗的結果常是一些測驗分數的集合，這些測驗分數反映了學生答對與答錯的題數，如此的分數可提供一種可靠且穩定的訊息，這種訊息是關於學生的能力在團體中所佔的相對位置。但這種訊息卻無法由學生的作答反應組型中，顯現出學生是否精熟某種技能的訊息。而這些額外的訊息可以幫助學生或老師更加瞭解分數所代表的涵義，以及哪一類的的學習可以增進學習成效(Sheehan, 1997)。

其實，當前的認知心理學對於學習歷程已具備相當豐富的研究成果，這些認知心理學的研究成果，可以對教育評量或心理測驗理論有下列幾點顯著的貢獻(余民寧，民 84)：

1.認知心理學對認知歷程的研究成果，有助於我們進一步瞭解測量背後所表徵的理論建構。

2.認知心理學能針對教育測量所要測量的教學目標，提供創新的測量策略，以及改善現有測量工具的缺失。透過認知心理學的協助，教育測驗可以針對教學與學習歷程，供更多的診斷和有用的訊息。

3.教育學術有許多新創的性向、學習與成就理論，都是由認知心理學的相關研究所建構得來的。因此，將認知心理學的研究結果，與教育評量相結合，意謂著教育測量將有全新的發展趨勢。

從上述認知心理學對教育評量的貢獻中可知，以認知心理學的研究成果為理論基礎，所建構的教育評量，較能正確的評量出學習者真實性的認知能力，也較能提供教學者進行診斷教學時所需的訊息。

鑑於教育評量與心理學有關學習與表現的理論仍存有許多的鴻溝，因此，許多學者一再地呼籲希望能將心理計量學與心理學整合，讓心理計量模式不單只是描述試題的統計特質，也能根據認知心理學所提出的認知模式，提出個體表現的質化特質(qualitative characterization)，使教育與心理評量能對教學目標提供更多的協助(Frederiksen, 1986; Lohman & Ippel, 1993; Snow & Lohman, 1989; Yamamoto & Gitomer, 1993)。

Nichols(1994)即主張傳統評量理論並無法提供有效的訊息，讓教師對學生的錯誤學習進行診斷的評量，因此，他提倡將認知科學(cognitive science)與心理計量學(psychometrics)結合，發展新的診斷評量方法，以幫助教學目標的達成。Nichols 將這種新的診斷評量方法，稱為認知診斷評量(cognitively diagnostic assessment,簡稱為 CDA)。他認為認知診斷評量應該有明顯

的實質假定(substantive assumptions)，就是探討受試者在答題時，產生何種認知歷程與知識結構？這些認知歷程與知識結構如何發展？以及能力高的受試者與能力低的受試者在這些方面上有何差異？這種認知診斷評量可以透過受試者對試題的作答反應型態，而推論出其認知歷程與知識結構的可能狀態。例如 Tatsuoka(1983)所發展的規則空間，可以推測出受試者所採用的解題策略；而 Goldsmith, Johnson, & Acton(1991)所使用的徑路搜尋法，可以推測出受試者的知識結構等。

Nichols 認為測驗的編製必須以心理學所發展出的理論為基礎，如此才能編製一份可以反映受試者知識狀態的測驗，進而達成認知診斷評量所追求的目標。因此，他提出編製以心理學為導向的認知診斷測驗，須經歷五個重要的編製歷程。

貳、認知診斷測驗的編製歷程

Nichols(1994)主張編製以心理學為導向的認知診斷測驗，須經歷實質理論的建構(substantive theory construction)、設計的選擇(design selection)、測驗管理(test administration)、反應計分(response scoring)與設計修正(design revision)等五個重要的編製歷程，如表 1 所示。

表 1　發展以心理學為導向的認知診斷測驗之五個編製步驟

步驟一	**實質理論的建構** 實質理論關注於描述知識與技能的模式或理論的發展，這些知識與技能被假定在認知表現時所產生的，而測量的試題必須能描述個體表現時的知識與技能。
步驟二	**設計的選擇** 在這個步驟，測驗的編製者必須選擇觀察與測量設計，選擇的標準必須以步驟一的實質理論為基礎，所建構的試題必須能預測受試者所可能產生的特定知識與技能。建構測量的程序是一種測量設計的操作化過程。
步驟三	**測驗管理** 測驗管理包括測驗內容的每個部分：試題的形式、反應的種類、計分的工具、施測的環境等。
步驟四	**反應計分** 這個步驟的目標是根據受試者的反應組型，給予某個數值，並將其反應組型與實質理論所建構的策略或錯誤規則相聯結。
步驟五	**設計修正** 設計修正是一個蒐集支持模式或理論的過程，即透過證據的蒐集，可以獲知理論是被支持或被挑戰的。在這個步驟，測驗施測的結果將用來修正實質理論的建構。

(Nichols,1994)

Nichols 曾舉 Gitomer and Van Slyke(1988)讓航空電子工程師進行航空電子儀器的錯誤偵測研究為例，來說明如何透過上述的五個步驟，編製出可以評量航空電子工程師能否正確偵測出錯誤的認知診斷測驗。由於 Nichols 所舉的例子，是針對航空電子工程師的評量，該例子並非特別針對教室情境的教學評量。為了讓讀者更清楚瞭解 Nichols 所主張的五個測驗編製步驟，如何應用在教室的教學情境，研究者乃另外自己舉「國小低年級數學加法與減法的文字題」的例子，來說明如何將 Nichols 所提的五個測驗編製步驟，實際應用於教學評量。

近年來認知心理學與教學心理學對於國小低年級數學加法與減法文字題的研究，已累積許多相關的研究成果(Nesher, 1986; Riley, Greeno & Heller, 1983; Resnick & Ford, 1984)。Fuson(1992)綜合對國小低年級加法與減法文字題的相關研究後，歸納出加法與減法的文字題可分成改變(change)、組合(combine)與比較(compare)三種不同的問題題型，其中改變問題是屬於動態的問題類型，而組合與比較問題是屬於靜態的問題型態。雖然這三種問題題型的解法算則完全相同，但卻對低年級學生產生不同的難度，這三種不同的問題題型分別為：

1.改變的問題：這是屬於動態情境的問題類型，整個問題包括三個語意關係，即起始的狀態、數量改變的大小與最後獲得的結果。通常題目會給兩個已知條件的語意關係，然後求第三個語意關係的數量多寡，如此可得到六種不同語意關係的加法和減法文字題(如表 2 所示)。例如最後的結果量未知，而起始狀態與改變量已知時，就可產生如「喬有 3 顆彈珠，後來湯姆給他 5 顆彈珠，現在喬有多少顆彈珠？」這類的加法問題。

2.組合的問題：這是屬於靜態情境的問題型態，只是集合之間的組合關係，沒有任何行動的改變。學生在解這組合問題時，必須具備「部分－整體」(part-whole)的概念，知道整體等於部分的總和。例如「喬有 3 顆彈珠，湯姆有 5 顆彈珠，兩人共有多少顆彈珠？」。

3.比較的問題：這是屬於兩個集合相互比較差異的關係，例如「喬有 3 顆彈珠，湯姆比喬多 5 顆彈珠，湯姆有多少顆彈珠？」。

雖然上述三道題目的解法算則都是「3＋5＝8」，但因以不同的語意關係呈現，就會對低年級的學生產生了不同難度，根據研究的結果顯示，國小低年級的學生最會解答改變的問題，比較問題對他們而言是最難的。

表 2　加、減法文字題的類型

動態	靜態
第一類「改變問題」	**第二類「組合問題」**
1.結果量未知	1.組合量未知
(1).喬有 3 顆彈珠, 後來湯姆給他 5 顆彈珠, 現在喬有多少顆彈珠？	(1).喬有 3 顆彈珠, 湯姆有 5 顆彈珠, 兩人共有多少顆彈珠？
(2).喬有 8 顆彈珠, 後來他給湯姆 5 顆彈珠, 現在喬有多少顆彈珠？	2.次集合未知 (2).喬和湯姆兩人共有 8 顆彈珠, 喬有 3 顆彈珠, 湯姆有多少顆彈珠？
2.改變量未知	**第三類「比較問題」**
(3).喬有 3 顆彈珠, 後來湯姆給他一些彈珠, 現在喬有 8 顆彈珠, 湯姆給他幾顆彈珠？	1.差異量未知 (1).喬有 8 顆彈珠, 湯姆有 5 顆彈珠, 喬比湯姆多幾顆彈珠？
(4).喬有 8 顆彈珠, 後來他給湯姆一些彈珠, 現在喬有 3 顆彈珠, 他給湯姆幾顆彈珠？	(2).喬有 8 顆彈珠, 湯姆有 5 顆彈珠, 湯姆比喬少幾顆彈珠？
3.起始量未知	2.比較量未知
(5).喬有一些彈珠, 後來湯姆給他 5 顆彈珠, 現在喬有 8 顆彈珠, 喬一開始的時候有多少顆彈珠？	(3).喬有 3 顆彈珠, 湯姆比喬多 5 顆彈珠, 湯姆有多少顆彈珠？ (4).喬有 8 顆彈珠, 湯姆比喬少 5 顆彈珠, 湯姆有多少顆彈珠？
(6).喬有一些彈珠, 後來他給湯姆 5 顆彈珠, 現在喬有 3 顆彈珠, 喬一開始的時候有多少顆彈珠？	3.參照量未知 (5).喬有 8 顆彈珠, 他比湯姆多 5 顆彈珠, 湯姆有多少顆彈珠？ (6).喬有 3 顆彈珠, 他比湯姆少 5 顆彈珠, 湯姆有多少顆彈珠？

(Riley, Greeno, & Heller, 1983)

以下就以上述國小低年級數學加法與減法文字題為例，來說明如何透過五個測驗編製的

步驟，來編製以心理學研究成果為基礎的認知診斷測驗。

一、實質理論的建構

實質理論的建構就是要尋找由心理學的研究所發展出的理論，作為發展測驗的理論依據。由心理學的研究所得到的實質理論，可以協助測驗編製者，藉由編製的試題推測受試者所具有的知識結構。因此，有用的實質理論最好能提供受試者在答題時，可能會使用到哪些認知機制(cognitive mechanisms)，或是不同能力的受試者，其可能的差異點為何。例如 Resnick & Ford(1984)認為表 2 加法和減法的數學文字題，高數學解題能力與低解題能力的學生，其數學知識結構有明顯的不同，高數學解題能力的學生會將加法與減法統合成一個相互關聯的運算系統，如圖 1 所示；而低數學解題能力的學生通常將加法與減法視為兩個獨立的運算系統，如圖 2 所示。因此，高數學解題能力學生對於加法與減法的數學問題，比低數學解題能力的學生，能形成較多不同的解題資源。

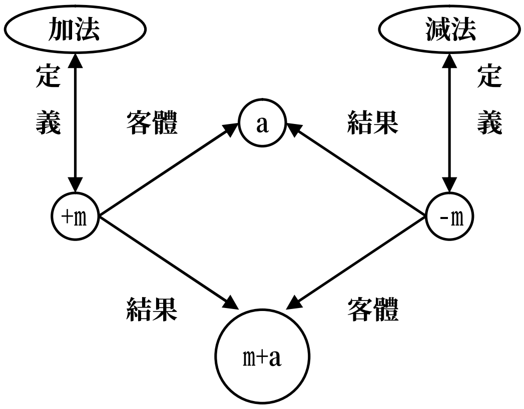

圖 1　高數學解題能力學生對加減法的知識結構(Resnick & Ford, 1984)

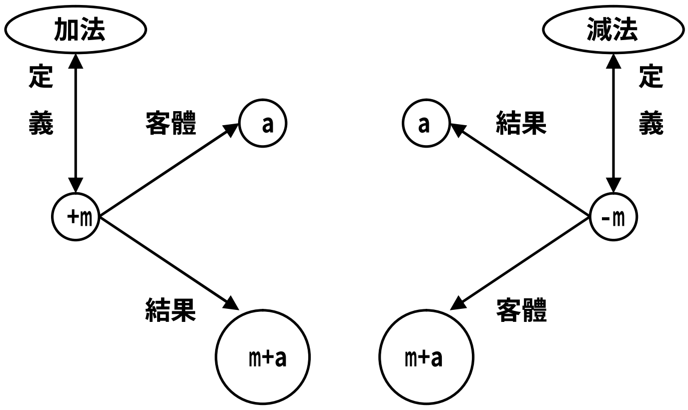

圖 2　低數學解題能力學生對加減法的知識結構(Resnick & Ford, 1984)

二、設計的選擇

Nichols(1994)認為設計的選擇主要包括觀察設計(observation design)與測量設計(measurement design)兩個部分。

觀察設計是描述測量活動的特質，例如試題的建構與組織、答題的類型等。觀察設計的目的在於建構適當的測量試題，使之能有效的推論出受試者於答題過程中，所經歷的認知歷程與知識結構狀態。而理想的測量試題，必須能夠區別出不同能力組別的受試者，其認知歷程與知識結構的差異處。因此，觀察設計包含了演繹推理，亦即由心理學對某學科領域的知識結構所發展出的理論模式，來推論受試者可能的不同行為反應。亦即觀察設計隱含著測驗的編製者必須能夠瞭解受試者在測量的過程中，所可能遭遇的認知需求(cognitive demands)，以及所可能產生的反應類型。這些重要的訊息，對試題編製的成功與否，扮演著關鍵性的角色，而這些重要的訊息，有賴於心理學的研究結果，提供作為編製試題的參考依據。

表 2 加法與減法的數學文字題，有些低年級學生會採用關鍵字的解題策略，針對特定的字句，來選擇計算的方法。例如對第二類的組合問題「喬有 3 顆彈珠，湯姆有 5 顆彈珠，兩人共有多少顆彈珠？」，採用關鍵字的解題策略的學生，看到題目中的關鍵字「共有」兩個字，就會很快的將 3 和 5 這兩數字相加。為了想瞭解學生是否採用關鍵字的解題策略，老師可以將題目改為「喬有 3 顆彈珠，湯姆比喬多 2 顆彈珠，兩人共有多少顆彈珠？」，如果某學生的解法是將 3 和 2 相加，則可以推論該學生可能採用關鍵字的解題策略。

測量設計是描述欲測量的個體(object)，以及如何將欲測量的個體，分配某個數值或某個類別的程序。例如常見的測量設計是將受試者答對的分數加總起來，即得到受試者的得分情形；而潛在類別分析(latent class analysis)則是將受試者分配為某個類別的一種測量程序。測量設計的目的，透過蒐集與整理受試者的作答情形，以確定受試者的反應類型，進而推論其所擁有的認知機制。因此，測量設計包含了歸納推理，亦即由受試者所觀察到的行為反應，來推論其認知歷程與知識狀態。

在測量設計的部分，如果只重視受試者總分的高低，則得分越高代表能力越好。然而只從分數的高低考量，可能會失去許多有用的訊息。例如有甲、乙兩位學生分別計算表 2 的 10 道數學加法和減法的文字題，前 5 題是屬於「改變問題」的題目，後 5 題是屬於「比較問題」的試題。反應組型中的「1」代表答對，「0」代表答錯。

作答反應組型

甲生：[1111100000]　　總分 5 分

乙生：[0000011111]　　總分 5 分

由作答反應組型可看出，甲生答對所有「改變問題」的類型，而「比較問題」的類型全答錯，顯示甲生對「改變問題」的類型學的比較好，「比較問題」的類型則有必要進行補救教學。乙生雖然的得分與甲生同為 5 分，但對不同問題類型的學習成效，卻恰與甲生相反。而通常大部分學生較會計算「改變問題」的類型，較不會計算「比較問題」的類型，據此，對於乙生作答反應組型所呈現的訊息，教師有必要再更深入瞭解其中的緣由。

三、測驗管理

測驗管理主要的目的是讓受試者在測驗的歷程中，能將真實能力完全的表現出來，亦即將影響受試者實際表現的干擾因素，進行適當的控制。這些干擾因素包括測驗歷程的每個細節，例如答題的指導語、測驗的題型(是非題、選擇題或問答題)、試題的配分、試題的語句、施測的環境

等等。

這些干擾因素常會影響受試者的實際表現，以試題敘述的語句為例，對表 2 第二類的組合問題「喬和湯姆兩人共有 8 顆彈珠，喬有 3 顆彈珠，湯姆有多少顆彈珠？」，根據 Verschaffel and Corte(1993)的研究結果發現，有許多小學生將「喬和湯姆兩人共有 8 顆彈珠」，解讀成「喬有 8 顆彈珠，湯姆也有 8 顆彈珠」，因而得到錯誤的答案。但如果將題目改為「喬和湯姆兩人共有 8 顆彈珠，喬有 3 顆彈珠，其餘的彈珠是湯姆的，湯姆有多少顆彈珠」，則大部分的小學生可以算出正確的答案。由此可知，不同語句的敘述，會影響學生的實際表現。

四、反應計分

傳統測驗會對每道試題，計算其難度(P)與鑑別度(D)，而這兩個指標，用於診斷學習者的學習歷程，並無法提供有效的訊息。因此，認知診斷評量的計分方式，強調測驗的計分能夠提供老師足夠的診斷訊息，來評估學生的學習歷程。為了讓試題能提供施測者較多診斷訊息，Bart and Williams-Morris(1990)曾提出兩個重要指標，用來判斷試題計分方式，能否提供較多的診斷訊息。這兩個指標分別是：反應解釋度(response interpretability)與反應區別度(response discrimination)。

反應解釋度用以判斷每道試題中的每個反應選項，是否至少含有一個規則。試題的反應解釋度其最小值為 0，顯示該道試題的每個選項，都沒有包含一個或一個以上的規則。試題的反應解釋度其最大值為 1，顯示該道試題的每個選項，都包含至少一個規則以上。反應解釋度的數值大小，能提供施測者許多重要的診斷訊息，因為根據反應解釋度的數值大小，施測者可藉由受試者所挑的選項，來推論受試者答題時，可能經歷的各種認知運作歷程。反應解釋度數值越大，越有機會顯現出受試者答題時，可能經歷的各種心理運作歷程。

反應區別度用以判斷每道試題中的每個反應選項，是否只包含一個規則。試題的反應區別度其最小值為 0，顯示該道試題的每個選項，都沒有包含一個或一個以上的規則。試題的反應解釋度其最大值為 1，顯示該道試題的每個選項，都只包含一個規則。反應區別度的數值大小，會影響到施測者能否由獲得的相關訊息，正確的推論出受試者答題時，經歷過哪一種認知運作歷程。反應解釋度數值越大，越有機會推論出受試者答題時，真正經歷哪一種心理運作歷程。

我們再以表 2 加法和減法的文字題為例，採用選擇題的題型，來說明如何計算反應解釋度與反應區別度。例如有道選擇題「喬有 2 顆彈珠，湯姆有 2 顆彈珠，兩人共有多少顆彈珠？①0 ②1 ③4 ④6」。

這道選擇題的第一個反應選項為 0，其包含的運算規則，只有一種可能，亦即減法的規則：2－2＝0。

這道選擇題的第二個反應選項為 1，其包含的運算規則，只有一種可能，亦即除法的規則：2÷2＝1。

這道選擇題的第三個反應選項為 4，其包含的運算規則，有兩種可能，亦即加法的規則與乘法的規則：2＋2＝4；2 ×2＝4。

這道選擇題的第四個反應選項為 6，這個選項並沒有包含任何運算規則。

綜合上述的分析可知，該道選擇題含有 4 個反應選項，以及 4 個可能的運算規則(加法、減法、乘法和除法)。接著我們就以反應選項與運算規則，建立一個矩陣，如表 3 所示。矩陣的元素 1，代表該選項具備該種運算規則，矩陣的元素 0，代表該選項不具備該種運算規則，

表 3　反應選項 ×規則的矩陣

反應選項	規則			
	1(加法)	2(減法)	3(乘法)	4(除法)
1(選項一)	0	1	0	0
2(選項二)	0	0	0	1
3(選項三)	1	0	1	0
4(選項四)	0	0	0	0

反應解釋度的計算方式為，將具有一個或一個以上規則的選項，相加總後，再除以所有的選項。表 3，第一選項(具一個規則)、第二選項(具一個規則)與第三選項(具兩個規則)，都具有一個或一個以上運算規則，所以相加總後共得到 3，再將 3 除以所有選項(共有 4 個選項)，得到 3 ÷4 ＝0.75。

反應區別度的計算方式有兩個步驟，第一步驟先將每個選項所具有的規則數目，求其倒數，再相加總起來。表 3，第一個選項有 1 個規則，其倒數為 1/1，第二個選項有 1 個規則，其倒數為 1/1，第三選項有 2 個規則，其倒數為 1/2。但遇到選項沒有任何規則時，則直接將其倒數設為 0。所以第四選項有 0 個規則，則直接將其倒數設為 0。因此，四個選項所具規則數目的倒數，再相加總起來，就得到 1/1+1/1+1/2+0=5/2。然後再將第一步驟所獲得的數字，除以所有選項(共有 4 個選項)，得到(5/2) ÷4＝0.625。

由上述反應解釋度為 0.75，反應區別度為 0.625，此道題目的兩個指標數值都蠻高的，顯示此道題目可以提供不錯的診斷訊息。例如選擇第一個選項的受試者，可能使用減法的運算方式。選擇第二個選項的受試者，可能使用除法的運算方式。選擇第三個選項的受試者，可能使用加法或乘法的運算方式。至於到底受試者採取加法，或是採用乘法，並無法有效確定，所以第三個選項的試題區別能力，比前兩個選項的試題區別能力還低；選擇第四個選項的受試者，並無法有效解釋其可能使用的運算方式。

五、設計修正

設計修正是在發展測驗時，透過相關資料的蒐集，來判斷所獲得的資料是否支持觀察設計與評量設計的一種歷程。在這個步驟，測驗施測的結果將用來修正實質理論的建構。這個步驟是一種持續不斷的歷程，一旦有新的資料出現，就必須針對整個設計進行調整，而藉由對整個設計的不斷修正，就能逐漸提高測驗的效度。

綜合上述的探討可知，透過上述五個步驟所編製而成的認知診斷測驗，比傳統測驗的編製歷程，更強調須以心理學研究成果的實質理論為編製試題的依據，同時對於試題也都特別經過周詳的設計，以期能從受試者答題反應組型中，推論受試者可能具備的知識結構狀態。Nichols(1994)即認為認知診斷評量與傳統測驗評量至少有兩點的不同：1.認知診斷評量所發展的評量指標，並不強調將受試者的得分以線性方式加以排名，而古典測驗評量的鑑別度與信度等指標，都是為瞭解受試者在團體中的相對位置。2.認知診斷的指標，可以在施測之前即可獲得，而古典測驗評量的指標，則必須等到施測後才能取得。這些相異點反映了認知診斷評量與古典測驗評量所重視的評量內容不同：認知診斷評量關注於受試者因認知歷程與知識結構不同，而產生不同的診斷訊息；而古典測驗理論則重視如何在特定的教育環境中，選擇較有可能學習成功的受試者。

雖然認知診斷測驗比傳統測驗能提供較多的診斷訊息，但由於認知診斷測驗是屬於新興的理論，因此，並未普遍的被應用到教室的教學評量中。為了讓讀者對於認知診斷評量有更清楚的瞭

解，茲舉幾個實際應用的例子，供讀者參考。

參、認知診斷評量的實際應用

認知診斷評量與心理學的許多分支都有密切的關聯性，這些心理學的分支包括職業心理學、軍事心理學、工業心理學、諮商心理學、教育心理學、發展心理學、認知心理學等(Bart and Williams-Morris, 1990)。而認知診斷評量也已被實際應用到這些相關的學科領域中，為了讓讀者對於認知診斷評量的實際應用，有較清晰的瞭解，茲針對 Tatsuoka(1983)的規則空間(rule space)、Fischer(1973)的線性邏輯測驗模式(linear logistic test model，簡稱 LLTM)、Mislevy & Verhelst(1990)的混合策略模式(mixed strategies model)、Doignon & Falmagne(1985)的知識空間(knowledge space)等四種比較受到重視的認知診斷評量，先探究其理論基礎，再介紹其實際應用例子。

一、規則空間模式

規則空間(rule space)是由 Tatsuoka(1983)所發展出的一種認知診斷評量，它是藉由試題評量的方式，找出受試者的試題反應組型(item response pattern)，進而推論受試者所擁有的潛在知識狀態(latent knowledge state)。在獲得受試者的潛在知識狀態後，即能瞭解受試者的知識結構，哪些部分是已經具有良好的聯結關係，哪些部分是需要再補強的。教師可藉由規則空間所診斷出的學習結果，對受試者進行補救教學。

(一)、理論模式

使用規則空間的評量方法，其所採用的試題必須經過特別的設計，如此才能從受試者的試題反應，診斷出受試者的知識狀態。每道試題必須包含幾個認知屬性(cognitive attributes)，這些認知屬性需要能反應出知識的向度。必須注意的是，試題所包含的認知屬性並非由規則空間的方法所產生的，而是透過工作分析(task analysis)的方式，分析出能代表該知識領域的表現行為。

規則空間的評量方法，通常包括五個步驟：定義試題的認知屬性、將認知屬性組合成試題、決定出各種的知識狀態、形成分類的空間、對受試者的反應進行分類(Katz, Martinez, Sheehan, & Tatsuoka, 1998)。茲將這五個步驟說明如下：

1.定義試題的認知屬性

使用規則空間的第一步驟就是界定所要評量的認知屬性，這些認知屬性可能是陳述性知識、程序性知識或是解題的策略等，它是構成認知診斷評量的基礎。藉由評量受試者是否擁有該認知屬性，施測者才能推論受試者可能的知識狀態。在決定試題的認知屬性時，通常是採用工作分析法，選擇該知識領域的重要成分，作為試題的認知屬性

2.將認知屬性組合成試題

確定好欲評量的認知屬性後，接著就是將認知屬性組合成試題，每道試題至少必須包含一個認知屬性。試題的編製過程中，並非任意的將認知屬性組合成試題，必須考量認知屬性的相似程度與難易程度。試題與認知屬性的關係，可藉由關聯矩陣(incidence matrix，通常以 Q 表示)顯現出來，關聯矩陣的階數(order)是由認知屬性的個數(k)乘以試題的數目(n)，若第 j 道試題包含第 k 個認知屬性，則 $q_{kj}=1$，否則 $q_{kj}=0$。例如有三道試題 j_1、j_2、j_3，有兩個認知屬性 k_1、k_2，其中 j_1與 j_3題各含有認知屬性 k_1，j_2題則包含認知屬性 k_2，亦即若想答對 j_1或 j_3題，需具備認知屬性 k_1的知識；若想答對 j_2題，需具備認知屬性 k_2的知識。則該關聯矩陣 Q 為(2×3)矩陣，如圖 3 所示。

$$\begin{array}{cc} & \text{試題} \\ & \begin{array}{ccc} j_1 & j_2 & j_3 \end{array} \\ \text{認知屬性} \begin{array}{c} k_1 \\ \\ k_2 \end{array} & \begin{bmatrix} 1 & 0 & 1 \\ & & \\ 0 & 1 & 0 \end{bmatrix} \end{array}$$

圖 3 三道試題與兩個認知屬性所構成的關聯矩陣
(Katz, Martinez, Sheehan, & Tatsuoka, 1998)

3.確定各種知識狀態

受試者的知識狀態，並無法經由直接的觀察得知，必須由試題反應組型，進行推估。因此，它是屬於一種潛在的特質。藉由各種認知屬性的排列組合，可以獲得許多種不同的知識狀態。知識狀態的類型是透過關聯矩陣 Q 來決定的。例如圖 3 的例子，j_1、j_2、j_3這三道試題，可能有八種不同的試題反應組型，分別為(0,0,0)、(1,0,0)、(0,1,0)、(0,0,1)、(1,1,0)、(1,0,1)、(0,1,1)、(1,1,1)，其中 1 代表答對，0 代表答錯。而 j_1題與 j_3題皆包含認知屬性 k_1，j_2題包含認知屬性 k_2，則由此三道試題與兩個認知屬性，構成了四種的可能知識狀態，分別為：

知識狀態一：受試者具備認知屬性 k_1的知識，而不具備認知屬性 k_2的知識，則其知識狀態為(0,1,0)。
知識狀態二：受試者具備認知屬性 k_2的知識，而不具備認知屬性 k_1的知識，則其知識狀態為(1,0,1)。
知識狀態三：受試者同時不具備認知屬性 k_1與 k_2的知識，則其知識狀態為(0,0,0)。
知識狀態四：受試者同時具備認知屬性 k_1與 k_2的知識，則其知識狀態為(1,1,1)。

若受試者的知識狀態是屬於上述四種知識狀態，則屬於典型試題反應組型(ideal item-response pattern)，若受試者的知識狀態屬於另外的四種類型：(1,0,0)、(0,0,1)、(1,1,0)、(0,1,1)，則屬於非典型試題反應組型。由典型試題反應組型，施測者可以清楚掌握受試者具有或缺乏哪些認知屬性的知識，至於非典型試題反應組型的出現，常導因於猜題或不小心等干擾因素。因此，施測者不易推估受試者具有或缺乏哪些認知屬性的知識。

4.形成分類的空間

分類的空間是採用兩維的笛卡兒座標，橫座標是以能力值(θ)代表，此θ值即為試題反應理論中的能力參數；縱座標則是以非典型反應組型(ζ)表示。如果以 R 代表典型的反應組型，則在規則空間中，任何的知識狀態皆可以表示為座標(θ_R, ζ_R)的型態。Tatsuoka(1985)曾舉單參數試題反應模式(one-parameter logistic IRT model)為例，來說明進行分類空間的演算歷程。單參數試題反應模式為公式(1)。P_j(θ)代表 logistic function，其公式為公式(1)。

$$P_j(q) = \frac{1}{1+\exp[-1.7(q_i - b_j)]} \qquad (1)$$

公式(1)的 P_j(θ)代表 logistic function，θ代表能力參數，b_j 代表試題 j 的難度參數。若以向量的型態表示，則 n 道試題的反應組型為 $x=(x_1,x_2,\ldots,x_n)$，而 n 道試題的答對機率為 P(θ)＝$[P_1(q), P_2(q),\ldots., P_n(q)]$，測驗的答對機率為 T(θ)＝$[T(q), T(q),\ldots., T(q)]$，其中 T(θ)＝

$\frac{1}{n}\sum_{j=1}^{n} P_j(q)$。

P(θ)與 T(θ)的殘差向量(residual vector)為[P(θ)－T(θ)]，且[P(θ)－T(θ)]＝[P_1(θ)－T(θ),P_2(θ)－T(θ),…,P_n(θ)－T(θ)]

P(θ)與試題反應組型 x 的殘差向量為[P(θ)－x]，且[P(θ)－x]＝[P_1(θ)－x_1,P_2(θ)－x_2,…,P_n(θ)－x_n]

將[P(θ)－x]與[P(θ)－T(θ)]兩個殘差向量以內積(inner product)的形式表示成函數，則該函數為公式(2)。

$$f(x)=[P(\theta)-T(\theta)]'\,[P(\theta)-x] \tag{2}$$

利用乘法的分配率(distributive law)，則公式(2)可改寫為公式(3)。

$$f(x)=[P(\theta)-T(\theta)]'\,P(\theta)-[P(\theta)-T(\theta)]'\,x\text{，亦可寫為}$$

$$f(x)=\sum_{j=1}^{n}(P_j(q)-T(q))P_j(q)-\sum_{j=1}^{n}(P_j(q)-T(q))x_j \tag{3}$$

當θ值固定時，則 f(x)函數中的第一項([P(θ)－T(θ)]' P(θ))即為常數，因此，f(x)值的大小將隨著 x 而改變。

假定 P(θ)值的大小是有遞減關係，則

[P_1(θ)＞P_2(θ) ＞・・・ ＞P_{n-1}(θ) ＞P_n(θ)]

因為 P_1(θ)＞P_2(θ) ＞・・・ ＞P_{n-1}(θ) ＞P_n(θ)，顯示前面的試題答對率較高，即題目的難度較低，相對答對的人數會較多。因此，典型的反應組型應該是答對前面的試題較多，亦即較多 1 的反應型態，而答錯後面的試題較多，亦即較多 0 的反應型態。如此，公式(3)的 f(x)值將會產生負值。Guttman 量尺就是屬於典型反應組型的一個好例子。相反地，若是非典型反應組型，則公式(3)的 f(x)值將會產生正值。由此，可藉由 f(x)的大小來判斷受試者是否有不尋常的反應組型，f(x)值越大，代表受試者的反應組型越不尋常，越需要進行診斷教學。

要判斷受試者是否具有不尋常的反應組型時，首先必須先將 f(x)標準化，Tatsuoka(1985)推演出第 i 個受試者的反應組型函數 f(xi)的期望值與變異數分別為公式(4)與公式(5)。

$$E(f(x_i))=0 \tag{4}$$

$$\mathrm{Var}(f(x_i))=\sum_{j=1}^{n}P_j(q_i)(1-P_j(q_i))(P_j(q_i)-T(q_i))^2 \tag{5}$$

標準化的 f(x)值，我們以ζ(θ,x)表示，則ζ(θ,x)的大小為公式(6)。

$$\zeta(\theta, \mathrm{x}) = \frac{f(\mathrm{x})}{\sqrt{\mathrm{Var}(f(\mathrm{x}))}} \tag{6}$$

所謂的規則空間是以ζ(θ,x)為縱座標，以θ為橫座標所形成的笛卡兒座標，規則空間中的每個座標點(θ,ζ(θ,x))，代表一種反應組型，也就是一種知識狀態。θ與ζ(θ,x)假定為以(θ,ζ(θ,x))為形心(centroids)的雙變項常態分配(bivariate normal distribution)，且其共變數矩陣(covariance matrix)為公式(7)。

$$\begin{bmatrix} 1/I(\theta) & 0 \\ 0 & 1 \end{bmatrix} \tag{7}$$

5.對受試者的反應進行分類

當所有可能的典型反應組型都映射到規則空間的笛卡兒座標後，即可根據受試者的座標值(θ,ζ(θ,x))大小，來決定受試者可能具有的知識狀態。分類的方法是將受試者的座標值(θ,ζ(θ,x))和最接近它的兩個知識狀態的座標值相比較，依據馬氏距離(Mahalanobis distance)大小來決定，受試者的座標值(θ ,ζ(θ,x))是比較類似哪一種知識狀態。

決定出受試者比較類似的知識狀態後，即能據此瞭解受試者的學習狀況，進而針對受試者的學習盲點，進行補救教學。

(二)、應用實例

Tatsuoka(1995)針對 2000 名參加學術性向測驗(SAT)的學生為研究對象，以規則空間的研究方法，探討這些學生在數學測驗方面的得分，其可能代表的知識狀態為何，如圖 4 所示。

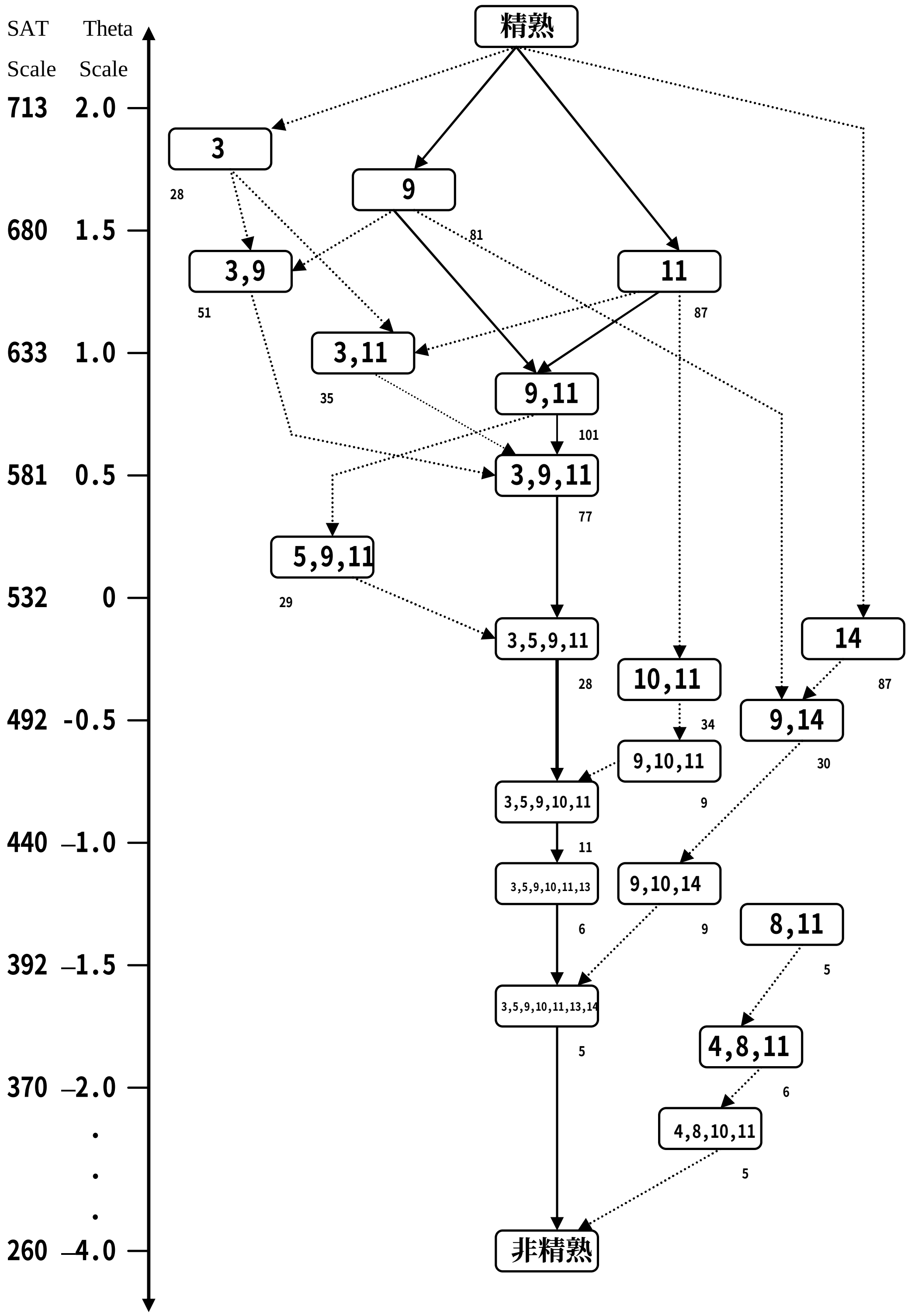

圖 4　　知識狀態的網路圖(Tatsuoka, 1995)

圖 4 的最左邊的一行代表 SAT 的分數，第二行代表相對的能力值(θ)，例如 SAT 為 680 分時，能力值大約為 1.5。此圖中總共有 14 個認知屬性，橢圓形中的數字，代表尚未學會的認知屬性，而其餘的認知屬性則已經學會。圖 4 最上面橢圓形是精熟，代表 14 個屬性全部學會，而最底下的橢圓形是非精熟，代表 14 個認知屬性中，沒有學會任何一個認知屬性。橢圓形外的數字，代表被分類為這種知識狀態的學生人數。

例如 SAT 分數約為 660 分時，有兩種可能的知識狀態，左邊的橢圓形包含 3 和 9，即顯示該知識狀態尚未學會認知屬性 3 和認知屬性 9，但學會了其它的 12 個認知屬性，這一類的受試者共有 51 人；而右邊的橢圓形包含 11，即顯示該知識狀態尚未學會認知屬性 11，但學會了其它的 13 個認知屬性，這一類的受試者共有 87 人。由此可知，SAT 同樣是 660 分時，確有兩種不同的知識狀態，而未具備認知屬性 11 的受試者，只要學會認知屬性 11 後，就能達到精熟的程度；至於未具備認知屬性 3 和認知屬性 9 的受試者，有兩種方式可以達到精熟的程度，一種是先學會認知屬性 3 後，再學會認知屬性 9 即可達到精熟的程度，另外一種即是相反的歷程，亦即先學會認知屬性 9，再學會認知屬性 3。

(三)、對規則空間模式的評價

綜合上述對於規則空間模式的探究，茲分析該模式的優缺點如下：

1.規則空間模式最大的優點在於可以提供豐富且準確的診斷訊息，供教學者判斷學習者的學習歷程與學習成果。由圖 4 所舉 SAT 約為 660 分的例子，可以清楚比較出規則空間的評量模式，比傳統評量方法，能提供教學者較準確的診斷訊息。傳統評量方法對於同分的受試者，並無法提供有效的訊息，供教學者判斷他們是否具備相同的知識結構。但採用規則空間的評量模式，對於同分的受試者，則能提供有效的診斷訊息。例如，在 SAT 的成績表現同樣是 660 分，有些受試者是未具備認知屬性 11，而有些受試者則是未具備認知屬性 3 和認知屬性 9。

2.雖然規則空間模式具有能提供學習者學習歷程的診斷訊息的優點，然而，目前可能還無法被廣泛應用到教室的教學評量。規則空間模式無法廣泛被應用的主要的原因有兩個：首先，規則空間模式的估計方法，是以試題反應理論(IRT)為理論基礎，具有較複雜的數理公式，評量者可能會因不熟悉規則空間模式的複雜公式，而不知如何採用此種評量方法。其次，規則空間模式所編製的評量試題，必須根據心理學的研究成果，如此，才能透過工作分析，讓每道試題包含某些特定的認知屬性。然而，目前教學心理學對於各學科領域的學習教材，尚未建立豐富的研究成果，因此，教學者並不易編製出符合規則空間模式要求的評量試題。

二、線性邏輯測驗模式

線性邏輯測驗模式(linear logistic test model，簡稱 LLTM)是 Rasch 模式的一種延伸，它於緣起 Scheiblechner(1972)提出對試題難度(b_i)進行更精細的分解。

(一)、理論模式

線性邏輯測驗模式是由 Fischer(1973, 1974)所發展出來的，Fischer 以 Scheiblechner(1972)所提的試題難度(b_i)理論為基礎，將 Rasch 模式中的試題難度(b_i)，分解成許多認知操作(cognitive operations)的線性組合，認知操作也就是解題的規則，其反應的機率如公式(8)與公式(9)。

$$p=\frac{\exp(q_j-b_i)}{1+\exp(q_j-b_i)} \tag{8}$$

$$b_i=\sum_{l=1}^{p} w_{il}a_l+c \tag{9}$$

a_l (l=1,2,.....,P)是所謂的基本參數(basic parameters)，w_{il}是a_l的權重(weights)，c 通常是常態係數(normalization constant)。a_l是指答對該題，所需要的認知操作，因此，每道試題的難度不同，是因為每道試題所包含的認知操作不同的緣故。例如，任意兩道試題 I_i,I_k的難度差異($\beta_i - \beta_k$)可表示成公式(10)。

$$b_i - b_k = \sum_{l=1}^{p}(w_{il} - w_{kl})a_l \tag{10}$$

運用線性邏輯測驗模式的評量方法，即可透過受試者的試題反應組型，推估出受試者可能因沒有具備某種認知操作的知識或技能，而導致無法答對包含該種認知操作的題目。同時，也可推估出全部試題的所有認知操作，哪些認知操作是受試者比較容易學習獲得的，而哪些認知操作對受試者是比較艱深的。

(二)、應用實例

Spada & Kluwe(1980)以 949 名中學生為研究對象，採用線性邏輯測驗模式，探究中學生解決 24 道槓桿平衡的試題。Spada & Kluwe 歸納出解決這些槓桿平衡問題時，所可能需要具備的 8 項認知操作，如表 4 所示。

表 4　解決槓桿問題所需的認知操作

認知操作	
1	注意並且從法碼重量的差異上進行推論
2	注意並且從力臂長度的差異上進行推論
3	當槓桿的某一邊法碼(或力臂長度)改變時，懂得改變槓桿另一邊的法碼(或力臂長度)
4	當槓桿的某一邊法碼(或力臂長度)改變時，懂得改變槓桿另一邊的力臂長度(或法碼)
5	當槓桿的某一邊法碼(或力臂長度)改變時，懂得調整槓桿同一邊的力臂長度(或法碼)
6	能考慮影響槓桿平衡的其他因素
7	能利用槓桿原理，推論出槓桿平衡時，兩邊法碼與力臂長度的乘積要相等
8	當某一邊的法碼(或力臂長度)不等時，懂得將同一邊的力臂長度(或法碼)的大小，調整為原法碼(或力臂長度)大小的倒數

(Spada & Kluwe, 1980)

表 4 中的 8 項認知操作，分別顯示出受試者在解題時，所使用的認知技能。例如第 1 項認知操作，受試者對於槓桿平衡的問題，只關注到槓桿兩邊的法碼重量是否相等，顯示受試者對於槓桿平衡的問題，只具備最初淺的認識。第 7 項認知操作，受試者能推論出槓桿兩邊法碼與力臂長度的乘積要相等外，顯示受試者能充分瞭解槓桿原理。

Spada & Kluwe 將 24 道有關槓桿平衡的試題，全部經過特別的設計，學生在解決每道試題時，都需要使用到表 4 中 8 項認知操作中的某些項目。每道試題(i)都有一個相對應的向量ω_i，向量ω_i中的元素包含 1 或 0，1 代表解答該道試題時需要使用到某個認知操作，0 代表解決該道試題時，不需要使用到該項認知操作。例如，解第 1 題時，需要具備第 1 項認知操作才有可能答對，則第 1

題難度中的基本參數α，所相對應的權重其向量為ω_i =(10000000)。解第 5 題時，需要具備第 2 項與第 8 項認知操作才有可能答對，則則第 2 題難度中的基本參數α，所相對應的權重的向量為ω_2 =(01000001)。全部 24 道試題與 8 項認知操作，就構成一個以 1 或 0 所形成的矩陣，如表 5 所示。

表 5　24 道試題 ×8 項認知操作的矩陣

試題	認知操作							
	1	2	3	4	5	6	7	8
1	1	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0
3	0	1	0	0	0	0	0	0
4	1	0	0	0	0	0	0	0
5	0	1	0	0	0	0	0	1
6	1	0	0	0	0	0	0	1
7	1	0	0	0	0	0	0	1
8	0	1	0	0	0	0	0	1
9	1	0	1	0	0	0	0	0
10	1	0	0	1	0	0	0	0
11	0	1	0	1	0	0	0	0
12	0	1	1	0	0	0	0	0
13	1	0	0	0	1	0	0	0
14	0	1	0	0	1	0	0	0
15	0	1	0	0	1	1	0	0
16	1	0	0	0	1	1	0	0
17	0	1	1	0	0	1	0	0
18	0	1	0	1	0	1	0	0
19	1	0	0	1	0	1	0	0
20	1	0	1	0	0	1	0	0
21	1	0	0	0	0	0	1	1
22	0	1	0	0	0	0	1	1
23	1	0	0	0	0	0	1	1
24	0	1	0	0	0	0	1	1

(Spada & Kluwe，1980)

Spada & Kluwe 採用線性邏輯測驗模式，結果實際估計出來的基本參數a，分別是$\hat{a}_2 - \hat{a}_1 = 0.4$，$\hat{a}_3 = 0.6$，$\hat{a}_4 = 1.4$，$\hat{a}_5 = 1.2$，$\hat{a}_6 = 1.6$，$\hat{a}_7 = 1.4$，$\hat{a}_8 = 0.6$。$a$ 的數值越大，顯示其對應的認知操作越困難。而由實際估計出來的數值，顯示第 6 項認知操作的數值較大，顯示其較為困難。而第 1 項與第 2 項認知操作，卻無法估出實際數值，只能知道第 2 項認知操作比第 1 項認知操作較困難。

(二)、對線性邏輯測驗模式的評價

綜合上述對於線性邏輯測驗模式的探究，茲分析該模式的優缺點如下：

1.試題反應理論(IRT)中的 Rasch 模式，對於試題難度(b_i)，並未加以細分構成的要素。而線

性邏輯測驗模式則是關注到試題包含不同認知操作的線性組合，會導致不同的試題難度。因此，線性邏輯測驗模式比 Rasch 模式更能診斷出受試者因缺乏某種認知操作的知識或技能，而無法答對包含該種認知操作的試題。另外，線性邏輯測驗模式也可以將全部試題包含的所有認知操作，一起進行認知操作的基本參數 α 之估計，即可得知哪些認知操作是比較容易，哪些認知操作是比較困難。據此，教學者可由線性邏輯測驗模式所提供的診斷訊息，更清楚掌握受試者的學習歷程。

2.線性邏輯測驗模式是以 Rasch 模式為基礎，再加以擴增對試題難度(b_i)的探討。而 Rasch 模式常被批評的缺點是並未考慮鑑別力參數與猜測參數，而這個缺點同樣也出現在線性邏輯測驗模式。另外，若想採用線性邏輯測驗模式的評量方法，則必須使用 Fischer 所開發的 LpcM 套裝軟體，才能進行各種參數值的估算。礙於 LpcM 套裝軟體的取得與 LpcM 套裝軟體的使用能力等因素，目前線性邏輯測驗模式，並未廣泛的被應用在教室的教學評量。

三、混合策略模式

Mislevy & Verhelst(1990)認為 Fischer(1973)的線性邏輯測驗模式對於試題的測驗內容，已經關注到不同的認知操作。但這些不同的認知操作，通常被假定為使用於同一種解題策略，亦即所有受試者會採用同一種策略，解決所有測驗的試題，而這些試題則包含了各種不同的認知操作。然而在實際的測驗情境中，受試者常會使用不同的解題策略。因此，Mislevy & Verhelst 針對受試者可能採用不同的解題策略，發展出新的診斷評量模式，即是混合策略模式(mixed strategies model)。

(一)、理論模式

根據試題反應理論(IRT)，受試者 i 對試題 j 的答題結果 x_{ij}，其機率為公式(11)。

$$p(x_{ij}\,|\,q_i,b_j)=[\,f(q_i,b_j)]^{x_{ij}}[1-f(q_i,b_j)]^{1-x_{ij}} \tag{11}$$

q_i 為能力參數，b_j 為難度參數，則 $f(q_i,b_j)$ 為已知函數。在 IRT 局部獨立(local independence)的假定下，受試者 I 對 n 道試題的答題反應組型($\mathbf{x}_{ij}$)的機率為公式(12)。

$$p(\mathbf{x}_i\,|\,q_i,\boldsymbol{\beta})=\prod_{j=1}^{n}p(x_{ij}\,|\,q_i,b_j) \tag{12}$$

若以 LLTM 的形式表示，則試題反應函數為公式(13)，亦即前面的公式(8)。

$$p(x_{ij}\,|\,q_i,b_j(\boldsymbol{\alpha}))=\frac{\exp[x_{ij}(q_i-b_j)]}{1+\exp[(q_i-b_j)]} \tag{13}$$

其中，試題難度(b_j)可以分解成認知操作($\boldsymbol{\alpha}$)的線性組合，即為公式(14)。

$$b_j(\boldsymbol{\alpha})=\sum_{m=1}^{M}Q_{jm}a_m=Q_j'\boldsymbol{\alpha} \tag{14}$$

其中，$\boldsymbol{\alpha}$ 表示解題時所需要的認知操作，Q_{jm} 是指解決試題 j 必須用到第 m 個試題特徵(item features)的次數。

如果受試者可以採用 K 種不同的解題策略解題，則此反應模式有如下的假定：

1. 每個受試者對所有試題，皆採用同一種解題策略，不同的受試者可採用不同的解題策略。
2. 只能直接觀察到受試者的試題反應組型，而無法直接觀察到受試者的解題策略，受試者的解題策略必須藉由試題反應組型加以推估。
3. 受試者採用第 k 個策略所產生的試題反應組型，會符合某個已知的試題反應模式。
4. 實質理論聯繫了試題的觀察特徵(observable features)與受試者採用某個策略的答對機率之間的關係。

假定受試者參數(subject parameter)ψ是 K 維度的向量空間，代表所採用的 K 種解題策略，則

$$\begin{cases} \psi_{ik}=1 & \text{當受試者 i 採用第 k 個解題策略時，則}\psi_{ik}=1 \\ \psi_{ik}=0 & \text{否則，}\psi_{ik}=0 \end{cases}$$

若以試題反應組型(X_{ij})與受試者參數(ψ_i 與 θ_i)表示時，則上面公式(11)與(12)式可改寫成公式(15)。

$$p(\mathbf{x}_i \mid f_i, q_i, a) = \prod_k \left\{ \prod_j [\, f_k(q_{ik}, b_{jk})]^{x_{ij}} [1 - f_k(q_{ik}, b_{jk})]^{1-x_{ij}} \right\}^{f_{ik}} \qquad (15)$$

同時，公式(14)的試題難度，則改寫成公式(16)。

$$b_{jk} = \sum_m Q_{jkm} a_{km} = Q_{jkm}' \boldsymbol{\alpha}_k \qquad (16)$$

(二)、應用實例

Mislevy & Verhelst(1990)以測量空間視覺能力的心理旋轉(mental rotation)的三道試題為例子，如表 6 所示。根據研究結果顯示受試者對於心理旋轉可以採用兩種不同的解題策略，一種為透過實際的旋轉，以找出相符合的圖形，而不同的旋轉角度會影響試題的難度；另一種為採用分析推理的方法，以判斷出符合的特徵，而特徵數量的多寡，也會影響試題的難度(Lohman, 1979)。

表 6　　試題特徵

試題	旋轉角度的改變	明顯的特徵
1	60 度	3
2	120 度	2
3	180 度	1

(Mislevy & Verhelst,1990)

每個受試者 i 有兩個向量：一為$\psi_i \equiv (\psi_{i1}, \psi_{i2})$，其中，$\psi_{ik}$代表受試者 i 採用第 k 個解題策略的參數。當受試者 i 採用第 k 個解題策略時，則$\psi_{ik}=1$，否則$\psi_{ik}=0$；另一向量為$\theta_i \equiv (\theta_{i1}, \theta_{i2})$，其中，$\theta_{ik}$代表受試者 i 採用第 k 個解題策略時的能力值。當受試者 i 採用第 k 個解題策略時，則受試者使用第 k 個策略的能力值為θ_{ik}。

假定受試者 i 採用第 1 種以實際旋轉的解題策略，解決心理旋轉的問題，則受試者 i 採用第 1 種解題策略的能力值為θ$_{i1}$，採用第 1 種解題策略的試題難度為β$_{i1}$，且解題策略參數ψ$_{i1}$為 1，根據公式(13)，其答對該題的機率為：

$$p(x_{ij} \mid q_{i1}, b_{j1}, f_{i1} = 1) = \frac{\exp[x_{ij}(q_{i1-} b_{j1})]}{1 + \exp[(q_{i1-} b_{j1})]}$$

根據 Cooper & Shepard(1973)研究結果，Mislevy & Verhelst 將使用實際旋轉與分析推理兩種不同解題策略的試題難度界定如下：

1.採用第 1 種解題策略，即採用實際旋轉的策略其試題難度，根據公式(14)，得到

$$b_{j1} = Q_{j11}a_{11} + Q_{j12}a_{12}$$

採用第 1 種策略，Q_{j11}是隨著旋轉角度的增加，而試題難度也跟著增加。因此，可設第一道旋轉 60 度的試題，其Q_{111}為 1，第二道旋轉 120 度的試題，其Q_{211}為 2，第三道旋轉 180 度的試題，其Q_{311}為 3，同時，因為採用第 1 種實際旋轉的解題策略，試題特徵的多寡，並不會影響不同試題的難度，因此，可設Q_{112}＝1，Q_{212}＝1，Q_{312}＝1。最後，若設a_{11}＝1，a_{12}＝-2。則採用第 1 種解題策略，第一道旋轉 60 度的試題，其試題難度為

$$b_{11} = Q_{111}a_{11} + Q_{112}a_{12} = 1\times1 + 1\times(-2) = -1$$

同理，採用第 1 種解題策略，第二道旋轉 120 度的試題，其試題難度為

$$b_{21} = Q_{211}a_{11} + Q_{212}a_{12} = 2\times1 + 1\times(-2) = 0$$

同理，採用第 1 種解題策略，第三道旋轉 180 度的試題，其試題難度為

$$b_{31} = Q_{311}a_{11} + Q_{312}a_{12} = 3\times1 + 1\times(-2) = 1$$

採用第 1 種解題策略，三道試題的難度，分別如表 7 所示。

2.採用第 2 種解題策略，即採用分析試題特徵的策略其試題難度為，根據公式(14)， 得到

$$b_{j2} = Q_{j21}a_{21} + Q_{j22}a_{22}$$

因為，採用第 2 種策略，Q_{j21}是隨著特徵的增加，而試題難度也跟著增加。由表 6 可知，第 1 題(旋轉 60 度)有 3 個明顯的特徵，則可設Q_{121}＝3，第 2 題(旋轉 120 度)有 2 個明顯的特徵，則可設，Q_{221}＝2，第 3 題(旋轉 180 度)有 1 個明顯的特徵，則可設，Q_{321}＝1。同時，因為採用第 2 種分析特徵的解題策略，旋轉角度的多寡，並不會影響不同試題的難度，因此，可設Q_{122}＝1，Q_{222}＝1，Q_{322}＝1。最後，若設a_{21}＝1.5，a_{22}＝-2.5。則採用第 2 種解題策略，第一道有 3 個明顯特徵的試題，其試題難度為

$$b_{12} = Q_{121}a_{21} + Q_{122}a_{22} = 3\times1.5 + 1\times(-2.5) = 2$$

同理，採用第 2 種解題策略，第二道有 2 個明顯特徵的試題，其試題難度為

$$b_{22} = Q_{221}a_{21} + Q_{222}a_{22} = 2\times1.5 + 1\times(-2.5) = 0.5$$

同理，採用第 2 種解題策略，第三道有 1 個明顯特徵的試題，其試題難度為

$$b_{32} = Q_{321}a_{21} + Q_{322}a_{22} = 1\times1.5 + 1\times(-2.5) = -1$$

採用第 2 種解題策略，三道試題的難度，分別如表 7 所示。

表 7　　試題難度參數

試題	解題策略 1 的試題難度	解題策略 2 的試題難度
1（旋轉 60 度）	-1.0	2.0
2（旋轉 120 度）	0.0	0.5
3（旋轉 180 度）	1.0	-1.0

(Mislevy & Verhelst,1990)

使用混合策略模式時，是將受試者的試題反應組型($\mathbf{x}_{ij}$)，透過貝氏定理(Bayes theorem)，即使用公式(17)的聯合事前機率(joint prior probabilities)，與公式(18)的聯合事後機率(joint posterior probabilities)，估計出不同能力值的受試者，可能採用不同策略的概率。

$$p(q_k = q, f_k = 1 \mid \boldsymbol{\pi}) = p_k g_k(q) \quad (17)$$

其中，$p_k = p(f_k = 1)$，$g_k(q_k) = p(q_k \mid f_k = 1)$。

$$p(q_k = q, f_k = 1 \mid \mathbf{x}, \boldsymbol{\pi}, \boldsymbol{\alpha}) \propto p[\mathbf{x} \mid f_k = 1, q_k = q, \boldsymbol{\beta}_k(\boldsymbol{\alpha})] p_k g_k(q) \quad (18)$$

其中，$p[\mathbf{x} \mid f_k = 1, q_k = q, \boldsymbol{\beta}_k(\boldsymbol{\alpha})] = \prod_j \frac{\exp\{x_{ij}[q - \boldsymbol{\beta}_k(\boldsymbol{\alpha})]\}}{\mathbf{1} + [q - \boldsymbol{\beta}_k(\boldsymbol{\alpha})]}$，且

$$p(\mathbf{x} \mid \boldsymbol{\pi}, \boldsymbol{\alpha}) = \sum_k p_k \int p[\mathbf{x} \mid f_k = 1, q_k = q, \boldsymbol{\beta}_k(\boldsymbol{\alpha})] p_k g_k(q)\,\mathbf{d}q$$

透過上述公式(17)與公式(18)，即可推論受試者較可能採用何種策略。以表 7 的三道試題為例，若某位受試者的反應組型為 x =(011)，亦即第 1 道題目答錯，第 2 與第 3 道題目答對，則根據上面模式估計的結果顯示：當受試者能力值為-1 時，則受試者採用第一種解題策略(實際旋轉策略)的概率函數值(likelihood fuction)為 0.021，受試者採用第二種解題策略(分析特徵策略)的概率函數值為 0.091。當

受試者能力值為 0 時，則受試者採用第一種解題策略的概率函數值為 0.033，受試者採用第二種解題策略的概率函數值為 0.256。當受試者能力值為 1 時，則受試者採用第一種解題策略的概率函數值為 0.041，受試者採用第二種解題策略的概率函數值為 0.406。由上述估計的結果可知，若試題反應組型為 x =(011)的受試者，其採用第 2 種解題策略的機率顯著高於採用第 1 種解題策略的機率。

(三)、對混合策略模式的評價

綜合上述對於混合策略模式的探究，茲分析該模式的優缺點如下：

1.有些評量模式假定受試者在解題的歷程中，對於同性質的所有試題，會採取同一種解題策略，然而混合策略模式則是假定受試者對同性質的所有試題，可能會採取不同的解題策略。混合策略模式的這一點假定，是比較符合受試者的真實解題情境，亦即對同性質的試題，受試者可能會採用不同的解題策略。其次，將受試者的試題反應組型，透過混合策略模式的估算，可以估算出不同能力值的受試者，選擇不同策略的概率函數值的高低，據此，可推估不同能力值的受試者，較可能採用何種解題策略。

2.使用混合策略模式的前提是，評量者必須事先確定受試者可能會採用哪幾種不同的解題策略，而這個部分則有賴於心理學提供豐富的研究成果，但以目前教學心理學對於各學科領域的探究成果，可能無法有效滿足評量者的需求。另外，對於某種解題策略，必須透過工作分析的方式，讓每道試題包含不同難易程度的認知操作。而這種將認知操作適當的分配到不同試題的艱難任務，也阻礙了混合策略模式的推行。

四、知識空間模式

Doignon & Falmagne(1985)提出的知識空間(knowledge space)評量方法，基本的假定為任何學習領域的評量，可以將該領域分解成一些問題或試題的集合，這些試題組合成知識表徵的基本元素。想要評量學生該領域的知識，即是在評定學生是否精熟某個次集合(subset)的試題。這個次集合稱為學生的知識狀態(knowledge state)，所有可能知識狀態的組合稱為知識結構(knowledge structure)，知識結構能表現出該領域的認知組織(Koppen, 1994)。而符合某些假設的特殊類型的知識結構，則稱為知識空間(knowledge space)。

知識空間的評量模式，首先要建構可能的知識狀態，再根據學生的反應組型去推估學生學習路徑(learning path)的可能機率，由此達到診斷學生學習狀況的目標。建構知識狀態的方法，主要有兩種：一種是請有經驗的學科專家或教師，以他們的經驗，根據試題的推測關係，推導出可能的知識狀態，這種方法類似於演繹推理的方法；另一種方法是從許多學生的試題反應組型，去歸納出可能的知識狀態，此種方法類似於歸納推理的方法(Dowling, 1994)。在實際應用上，兩種方法可以同時考慮使用。而由學科專家或教師所推導出的知識狀態，可能會因人而異，因此，應該由許多不同的專家或教師共同討論出一個可以被接受的知識狀態。

(一)、理論模式

知識空間的理論模式，主要是由受試者所精熟的問題，來推測其可能具有的知識狀態，再由知識狀態去估測受試者可能的學習路徑。知識空間模式在估測的歷程中，包含了一些基本假定，在這些基本假定中，較關鍵的部分，主要包含四項定義與兩項性質，茲以 Falmagne, Koppen, Villano, Doignon, & Johannesen(1990)所舉的五題基礎算術題，來說明四項定義與兩項性質的實質內涵。

下列有五道基礎的算術問題：

1 378 ×605＝？

2．58.7 ×0.94＝？

3．(1/2) ×(5/6)＝？

4．34 的 30%是多少？

5．冠朵玲(Gwendolyn)的年紀是麗蓓嘉(Rebecca)的 3/4 倍，麗蓓嘉(Rebecca)的年紀是愛德溫(Edwin)的 2/5 倍，愛德溫是 20 歲，冠朵玲是幾歲？

根據上述的 5 道數學題目，我們可以推論任何會解第 5 題應用題的學生，應該也會解第 1、2、3 題計算題。同樣地，不會解第 1 題的學生，大致也不會解第 2、4、5 題。我們可以用圖 5 來表示這樣的關係。

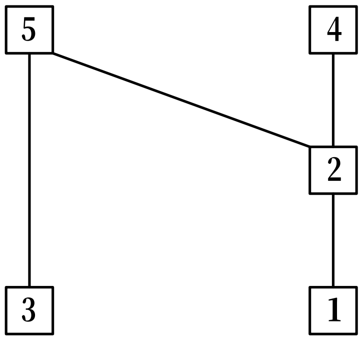

圖 5　5 道基本算術題的 Hasse 圖(Falmagne et al.,1990)

圖 5 所呈現的是一種階層的關係，代表能解決上面的題目代表一定能解決下面的題目，例如會解第 5 題的學生，也就會解第 2、1 和 3 題。而會第 2 題的學生，代表也會解第 1 題，但並無法保證就會解第 4 或第 5 題。

知識空間的理論模式，是以集合論為理論基礎，而下列所要介紹的四項定義與兩項性質，其主要內涵是關於集合理論的部分，茲以圖 5 的 5 道數學題目，說明如下：

定義 1：有限集合 Q，若其某些子集所形成的知識結構為 K，且 K 具有下列兩點性質的話，則 K 稱為在 Q 上的「知識空間」。

(1).空集合(Ø)與 Q 皆為 K 的元素。

(2).K 具備封閉性聯集(closed under union)的性質。

若有限集合 f 代表由圖 5 的 5 道數學題目所組成的集合，亦即 f＝{1,2,3,4,5}，則可以從有限集合 f 的 32 個子集中，根據定義 2 的原則，挑選出Ø,{1},{3},{1,3},{1,2},{1,2,3},{1,2,4},{1,2,3,5},{1,2,3,4},{1,2,3,4,5}等 10 個子集，來當成集合 F 的元素，亦即 F＝{Ø,{1},{3},{1,3},{1,2},{1,2,3},{1,2,4},{1,2,3,5},{1,2,3,4},{1,2,3,4,5}}

因為空集合(Ø)與 f 皆為 F 的元素，並且 F 具備封閉性聯集的性質，故 F 稱為在 f 上的知識空間。

定義 2：假如答對某題 t，也就會答對 q 題，亦即 q 題可從 t 題推測，若以符號「? 」表示「推測的關係」(surmise relation)，則可表示成

q? t ? (t? K ? q? K)

由上述定義 2 可知，圖 5 的 5 道算術問題，具有下列幾種的推測關係：

1? 5; 2? 5; 3? 5; 1? 4; 2? 4; 1? 2

由上述的幾種推測關係，可以找出 Ø,{1},{3},{1,3},{1,2},{1,2,3},{1,2,4},{1,2,3,5},{1,2,3,4},{1,2,3,4,5}等十種知識狀態，每種知識狀態各表示學生會作答的題目，例如{1,3}表示會解第 1 與第 3 題。若有位學生會做第 4 題，則該生可能的知識狀態為{1,2,4},{1,2,3,5},{1,2,3,4},{1,2,3,4,5}等四種知識狀態的其中一種。

令 F＝{Ø,{1},{3},{1,3},{1,2},{1,2,3},{1,2,4},{1,2,3,5},{1,2,3,4}，{1,2,3,4,5}}則 F 稱為知識空間，而 F 集合中的元素稱為知識狀態。

定義 3：若知識結構 K 中的元素，可藉由某些元素的聯集而獲得的話，則這些元素所形成的集合稱為「基礎 B」(basis B)。

如上例，若集合 C＝{{1},{3},{1,2},{1,2,4},{1,2,3,5}}，則集合 C 稱為知識空間 F 的「基礎」，因為知識空間 F 中的各種知識狀態，皆可藉由基礎 C 的聯集而獲得，因此，知識空間的基礎，是建構知識空間本身的最小子集。

定義 4：有限集合 Q 所形成的知識空間 K，若 K 中的最小元素包含某些試題 q，且 q? Q，則這些最小的元素，稱為 q 在 K 的最小狀態(minimal state)。

如上例，{1}是 1 的最小狀態，{3}是 3 的最小狀態，{1,2}是 2 的最小狀態，{1,2,4}是 4 的最小狀態，{1,2,3,5}是 5 的最小狀態。

性質 1：若 k 和 k´都是屬於知識狀態，則 k? k´也是屬於知識狀態。

如上例，{1,3}與{1,2,4}都是 F 的知識狀態，而{1,3}∪{1,2,4}＝{1,2,3,4}，也是 F 的知識狀態。

性質 2：若 k 和 k´都是屬於知識狀態，則 k∩k´也是屬於知識狀態。

如上例，{1,3}與{1,2,4}都是 F 的知識狀態，而{1,3}∩{1,2,4}＝{1}，也是 F 的知識狀態。

上述的四項定義與兩項性質，是知識空間的重要內涵，而學習路徑則是另一個重要的概念。Falmagne(1994)認為任何知識狀態都是屬於學習路徑的一個部分，學習路徑是由開始不具備任何知識技能的空集合狀態(Ø)，逐漸學習到全集合的知識狀態(Q)，亦即學會有限集合(Q)中的所有問題。此種學習路徑呈現一種鍊結的關係，是由沒有學到任何知識的空集合狀態開始學習，每次逐一學會一個問題，最後到達學會全部問題的全集合狀態，這種學習路徑稱為「順序」(gradation)。

為了讓讀者對學習路徑有更清楚的瞭解，茲再引用 Falmagne et al.(1990)所舉的另一個 5 道題目的例子，來說明學習路徑的概念。

設有限集合 Q，包含 5 道試題，則可表示 Q＝{1,2,3,4,5}，若知識空間 K，是由集合 Q 的 15 個子集所組成，而 K＝∅,{1},{3},{1,2},{1,3},{1,4},{2,3},{3,4},{1,2,3},{1,2,4},{1,3,4},{2,3,4},{1,2,3,4},{1,2,3,5},{1,2,3,4,5}}

則知識空間 K 的學習路徑共有 16 種「順序」，如圖 6 所示。

1 2 3 4 5	1 2 3 5 4	1 2 4 3 5	1 3 2 4 5
1 3 2 5 4	1 3 4 2 5	1 4 2 3 5	1 4 3 2 5
3 1 2 4 5	3 1 2 5 4	3 1 4 2 5	3 2 1 4 5
3 2 1 5 4	3 2 4 1 5	3 4 1 2 5	3 4 2 1 5

圖 6　知識空間 K 學習路徑的 16 種順序(Falmagne et al.,1990)

若以第一個「順序」{1,2,3,4,5}為例，即表示先學會問題 1，再依序學會問題 2、問題 3、問題 4，最後學會問題 5。則由圖 7 可推算出，其鍊結的關係為：

∅? {1}? {1,2}? {1,2,3}? {1,2,3,4}? {1,2,3,4,5}

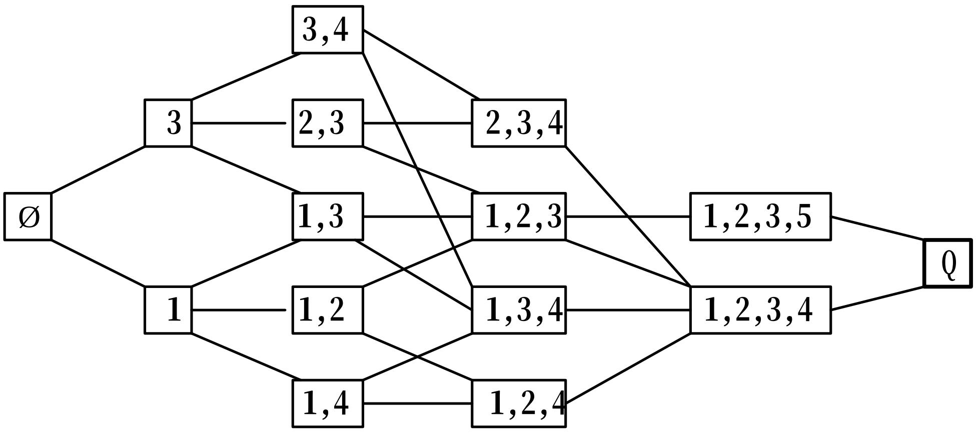

圖 7　知識空間 K 的學習路徑圖(Falmagne et al.,1990)

至於受試者的學習歷程，究竟是圖 6 的 16 個學習路徑中的哪一個順序，則是透過下面的公式(19)、公式(20)與公式(21)來推估。

在進行受試者學習歷程的推估時，假設具 m 個問題的有限集合 Q，其形成的知識空間 K，受試者在測試時間(τ)下，產生的試題反應組型 $R_\tau=(r_1,r_2\cdots,r_m)$機率，如公式(19)、公式(20)與公式(21)。

$$P\{R\}=\sum_{K\in \mathrm{K}} P\{R|K\}P\{K\} \tag{19}$$

$$P\{R|K\}=r_1(r_1,K)\times r_2(r_2,K)\times\cdots\times r_m(r_m,K) \tag{20}$$

$$P\{K\}=\sum_{p\in G(K)}\int_0^\infty P\{K_t=K|L=?,G=p\}p_p\,f(l)dl \tag{21}$$

其中，公式(20)與公式(21)中所包含的符號意義，說明如下：

$$r_i(r_i,K)=\begin{cases} 1-\beta_i & \text{假如 } i? K \text{ 並且 } r_i=1; \\ \beta_i & \text{假如 } i? K \text{ 並且 } r_i=0; \\ \eta_i & \text{假如 } i? K \text{ 並且 } r_i=1; \\ 1-\eta_i & \text{假如 } i? K \text{ 並且 } r_i=0; \end{cases}$$

β_i表示因粗心而答錯 i 題的機率，η_i表示因猜測而答對 i 題的機率，$r_i=1$ 表示答對 i 題，$r_i=0$ 表示答錯 i 題。例如，知識狀態為{2,5}的受試者，其試題反應組型 R=(0,1,1,0,1)的機率為 $P(R=(0,1,1,0,1)|k=\{2,5\})=(1-\eta_1)\ (1-\beta_2)\ \eta_3(1-\eta_4)(1-\beta_5)$。而知識狀態為{1,2,3}的受試者，其試題反應組型 R=(1,0,1,0,1)的機率為 $P(R=(1,0,1,0,1)|k=\{1,2,3\})=(1-\beta_1)\ \beta_2(1-\beta_3)(1-\eta_4)\ \eta_5$。

π代表學習路徑中的一種「順序」，$K\tau$ 表示受試者在時間τ的知識狀態，G(K)表示包含知識狀態 K 的所有順序(π)所形成的集合，λ表示受試者的學習速率(learning rate)，L 表示評量受試者學習速率的隨機變項，$P\pi$表示順序(π)的機率。

一旦施測者獲得受試者的試題反應組型，即可透過上述公式(19)、公式(20)與公式(21)，推估出受試者較有可能產生學習路徑的順序。據此，施測者即能清楚瞭解受試者的整個學習歷程。

(二)、應用實例

Falmagne et al.(1990).以 397 位十年級與十一年級的中學生為研究對象，讓這些受試者在 40 分鐘內作答 24 道數學試題。Falmagne et al.以其中的 5 道試題，進行學習路徑的分析。在可能的 16 種「順序」中，假設{1,4,2,3,5}、{1,4,3,2,5}、{3,4,2,1,5}這三種「順序」的機率為 0，並且假設猜測機率(η_i)為 0，每題因粗心答錯的機率(β_i)都相同。經估計的結果，得到 13 個「順序」的機率如表 8 所示。

表 8　13 種「順序」的估測機率

π	P_π
13245	.11
13254	.06
13425	.11
12345	.00
12354	.09
12435	.00
14235	設為 0
14325	設為 0
32415	.00
32145	.11
32154	.01
34125	.03
34215	設為 0
31245	.45
31254	.03
31425	.00

(Falmagne et al.,1990)

由表 8 可知，受試者在 16 個學習路徑的順序中，以「31245」的機率最高為.45，亦即受試者學會 5 道試題的順序，較可能是 31245，則受試者學習路徑的鍊結關係為：

Ø? {3}? {1,3}? {1,2,3}? {1,2,3,4}? {1,2,3,4,5}

(三)、對知識空間模式的評價

綜合上述對於知識空間模式的探究，茲分析該模式的優缺點如下：

1.傳統評量方法因無法評量出學習者的知識結構，故評量結果只能呈現出學習者零碎且片段的知識，而知識空間模式最大的優點則在於可以評量出學習者的知識結構，透過知識空間模式對學習者知識結構的分析，可以診斷出學習者對各概念之間的異同情形，是否產生適當的連結關係。同時，學習者對各概念的學習順序，也可藉由知識空間模式的推估，瞭解學習者的學習歷程。

2.知識空間模式在推估學習者的知識結構時，需要使用到特定的統計軟體(PRAXIS)，且因運用到許多複雜心理計量公式，因此，知識空間模式尚未普遍的被應用到教學評量領域。另外，進行知識空間評量時，評量者必須事先確定各概念之間的階層關係，這有賴豐富的心理學研究成果來支持，然而以目前心理學對教學領域的研究成果，似乎仍然無法提供充足的實質理論基礎。

上述四種認知診斷評量方式，在認知診斷測驗的編製時，都結合認知心理學的相關研究，透過對試題的特別設計，提供教學者診斷學習者認知發展的有效訊息，且測驗編製的歷程大致都符合 Nichols(1994)所主張的認知診斷測驗的五個編製歷程。然而上述四種認知診斷評量，都尚未能普遍的被應用到教室的教學評量領域，其主要的原因是目前心理學對於教學領域的研究成果，尚未能提供足夠的實質理論基礎，以及上述四種認知診斷評量，都包含相當複雜的心理計量公式。倘若目前的心理學對於各學科領域能提供豐富的實質理論基礎，則認知診斷評量模式是適用在各

學科領域的教學評量

其中，Tatsuoka(1983)的規則空間模式、Fischer(1973)的線性邏輯測驗模式與 Mislevy & Verhelst(1990)的混合策略模式，都是採用 IRT 模式為理論基礎，而 Doignon & Falmagne(1985)的知識空間模式，則是以集合論(set theory)為理論基礎。

因此，雖然上述四種認知診斷評量模式，具有能提供教學者相當豐富的診斷訊息，以及適用在各學科領域的教學評量的優點，可惜目前仍未能有效的被應用在教學評量中。相信不久的未來，若心理學能對教學領域提供更豐富的研究成果，以及認知診斷評量模式的發展者，能讓施測者更熟悉複雜的認知診斷模式，則認知診斷評量定能被廣泛的應用到教室情境的教學評量中。

肆、結語

好的評量方法，除了要能測量出學習者的學習現況外，同時也應該提供學習者學習缺失的診斷訊息，以利教學者進行有效的補救教學。如此，評量方法與教學歷程的結合，才能讓教學活動更完善。

國外的教育與心理評量界，已有許多學者一再呼籲，評量方法應該與認知心理學所獲得的實質理論相結合，也就是從事所謂的認知診斷評量，如此，才能讓評量真正發揮其最大的效益。同時，有多位國外學者，依循著認知診斷評量的評量歷程，發展出多種不同的認知診斷評量模式。而這些認知診斷評量模式，由於建構在複雜的數理公式上，並且需要搭配使用特殊的電腦程式，所以在評量的實務上，並未真正普遍的被使用。然而，可預見的是，認知診斷評量定是教育與心理評量界，大家共同努力追尋的重要評量方法。

國內目前對於學習者學習成效的評量，已從單一的紙筆測驗，轉變為多樣化的評量方式。雖然多元評量的實施，提供了較廣域的評量內容，較多樣的評量方法，並且創造了較真實與公平的評量情境。然而國內當前所推行的多元評量，並未提供教學者充足的診斷訊息，因此，如何在既有多元評量的基礎上，再配合認知診斷評量的評量觀點，是國內未來教育與心理評量界所當共同努力的方向。

參考書目

一、中文部分

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

二、英文部分

Anastasi, A. (1967). Psychology, psychologists, and psychological testing. *American Psychologist, 22,* 297-306.

Bart, W. M., & Williams-Morris, R. (1990). A refined item digraph analysis of a proportional reasoning test. *Applied Measurement in Education, 3*, 143-165.

Bloom B. S., et al.(1956). *Taxonomy of educational objectives: Cognitive domain.* New York: David McKay.

Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase(Ed.), *Visual information processing*(pp. 76-176). Orlando, FL: Academic Press.

Doignon, J. -P. & Falmagne, J. CI.(1985). Spaces for the assessment of knowledge. *International Journal of*

Man-Machine Studies, 23, 175-196.

Dowling, C. E. (1994). Integrating different knowledge spaces. In G. H. Fischer, & D. Laming.(eds.), *Contributions to mathematical psychology, psychometrics, and methodology.* New York: Speinger-Verlag.

Falmagne, J.-C. (1994). Finite Markov learning models for knowledge structures. In G. H. Fischer, & D. Laming.(eds.), *Contributions to mathematical psychology, psychometrics, and methodology.* New York: Speinger-Verlag. Falmagne, J.-C., Koppen, M., Villano, M., Doignon, J. – P., & Johannesen, L. (1990). Introduction to knowledge spaces: How to build, test and search them. *Psychological Review, 97,* 201-224.

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica, 37,* 59-374.

Fischer, G. H. (1974).*Einfü hrung in die Theorie psychologischer Tests.*[Introduction to mental test theory. In German.] Bern: Huber.

Frederiksen, N. (1986). Toward a broader conception of human intelligence. *American Psychologist, 41,* 445-452.

Fuson, K. C. (1992). Research on whole number addition and substraction. In D. A. Grouws(Ed.), *Handbook of research on mathematics teaching and learning.* New York: Macmilian.

Gitomer, D. H., & Van Slyke, D. A. (1988). Error analysis and tutor design. *International Journal of Machine Mediated Learning, 2,* 333-350.

Gitomer, D. H., & Yamamoto, K. (1991). Performance modeling that integrates latent trait and class theory. *Journal of Educational Measurement, 28,* 173-189.

Goldsmith, T. E., Johnson, P. J., & Acton, W. H. (1991). Assessing structural knowledge. *Journal of Educational Psychology, 83,* 88-96.

Katz, I. R., Martinez, M. E., Sheehan, K. M., & Tatsuoka, K. K. (1998). Extending the rule space methodology to a semantically-rich domain: diagnositic assessment in Architecture. *Journal of Educational and Behavioral Statistics, 24,* 254-278.

Koppen, M. (1994). The construction of knowledge spaces by querying experts. In G. H. Fischer, & D. Laming.(eds.), *Contributions to mathematical psychology, psychometrics, and methodology.* New York: Speinger-Verlag.

Koppen, M., & Doingon, J. – P. (1990).How to build a knowledge space by querying an expert. *Journal of Mathematical Psychology, 34,* 311-331.

Lohman, D. F. (1979). *Spatial ability: A review and reanalysis of the correlation literature*(technical Report No.8). Stanford, CA: Stanford University, Department fo Education, Aptitude Research Project.

Lohman, D. F., & Ippel, M. J. (1993). Cognitive diagnosis: From statistically-based assessment toward theory-based assessment. In N. Frederiksen, R. J. Mislevy, I. I. Bejar.(eds.), *Test theory for a new generation of tests.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Master, G. N. & Mislevy, R. J.(1993). New views of student learning: Implications for educational measurement. In N. Frederiksen, R. J. Mislevy, I. I. Bejar.(eds.), *Test theory for a new generation of tests.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Mislevy, R. J. (1993). Foundations of a new test theory. In N. Frederiksen, R. J. Mislevy, I. I. Bejar.(eds.), *Test theory for a new generation of tests.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Mislevy, R. J. & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika, 55,* 195-215.

Nesher, P. (1986). Learning mathematics: A cognitive perspective. *American Psychologist, 41,* 1114-1122. Nichols, P. D. (1994). A framework for developing cognitively diagnostic assessment. *Review of Educational Research, 64,* 575-603.

Resnick, L. B., & Ford, W. W. (1984). *The psychology of mathematics for instruction.* Hillsdale, NJ: Lawrence

Erlbaum Associates.

Riley, M. S., Greeno, J. G., & Heller, J. I. (1983). Development of children ' s problem-solving ability in arithmetic. In H. P. Ginsburg(Ed.), *The development of mathematical thinking.* New York: Academic Press.

Scheiblechner, H. (1972). Das Lernen und Lösen komplexer Denkaufgaben.[The learning and solving of complex reasoning items. In German.] *Zeitschrift für Experimentelle und Angewandte Psychologie, 3,* 456-506.

Sheehan, K. M. (1997). A tree-based approach to proficiency scaling and diagnostic assessment. *Journal of Educational Measurement, 34,* 333-352.

Snow, R. E., & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.)(pp. 263-331). Hillsdale, NJ: Lawrence Erlbaum Associates.

Spada, H., & Kluwe, R. (1980). Two moedls of intellectual development and their reference to the theory of Piaget. In R. Kluwe & H. Spada (Eds.), *Developmental model of thinking*(pp.1-32). New York: Academic Press.

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

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

Tatsuoka, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M. G. Shafto(Eds.), *Diagnostic monitoring of skill and knowledge acqusition*(pp.453-488). Hillsdale, NJ: Lawrence Erlbaum Associates Tatsuoka, K. K. (1995). Architecture of knoweledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach. In P. D. Nichols, S. F. Chipman, & R. L. Brennan.(eds.),*Cognitively diagnostic assessment*(pp.327-359). Hillsdale, NJ: Lawrence Erlbaum Associates.

Verschaffel, L., & Corte, E. D. (1993). A decade of research on word problem solving in Leuven: Theoretical methodological, and practical outcomes. *Educational Psychology Review, 5,* 239-255.

Yamamoto, K., & Gitomer, D. H. (1993). Application of a Hybrid model to a test of cognitive skill representation. In N. Frederiksen, R. J. Mislevy, I. I. Bejar.(eds.), *Test theory for a new generation of tests.* Hillsdale, NJ: Lawrence Erlbaum Associates.

收稿日期：民國 92 年 3 月 26 日
修正日期：民國 92 年 8 月 15 日
接受日期：民國 92 年 8 月 26 日

The Research of Cognitively Diagnostic Assessment

Chin-Tang Tu

Center for Teacher Education
National Kaohsiung Normal University

Abstract

There are two purposes of educational assessment. One is to test the student. The other is to help the teacher through the testing result to observe that if the student really understands the class. If they do not, then the teacher has to give some remedial instruction. In other words, the teacher can realize the learning problems of the student and give them a hand in learning. At abroad, many educational scholars aver that psychology and assessment have a close relationship. That is the reason why they tend to focus on the interaction between the theory of cognitive psychology and learning. In this sense, they do believe that when compiling worksheets, let the interaction be our guide. This kind of testing is the so-called "cognitively diagnostic assessment."

However, in Taiwan, few scholars are concerned with the significance of cognitively diagnostic assessment. In order to highlight the importance, this paper attempts, first, to elaborate the theoretical study of cognitively diagnostic assessment and, second, to examine how to compile the worksheets with some examples.

Keyword: cognitively diagnostic assessment, response pattern, knowledge space