

## 第二章 文獻探討

### 2.1 測驗理論

#### 2.1.1 傳統測驗理論的基本概念

傳統測驗理論 (Classical Test Theory, 簡稱 CTT) 的理論基礎主要是建立在真實分數模式 (True Score Model) 的七個基本假設之下, 分別敘述如下 [17,19,24,30,31]:

1. 在傳統測驗理論中觀察分數與真實分數的函數關係為  $X=T+E$ ,  $X$  代表觀察分數, 是受測者在某個測驗上的一個觀察值;  $T$  代表真實分數, 是對受測者能力取樣無數次, 其結果的平均數或期望值;  $E$  代表誤差分數, 是觀察分數與真實分數之間的差距。因此  $X=T+E$  意指任一次施測, 我們所觀察到的分數, 是受測者真實分數與誤差分數的總和。因為傳統測驗理論的主要工作是去估計測驗中觀察分數與真實分數關聯的強度, 所以它有時又稱作「真實分數理論」(True Score Theory)。

2. 觀察分數的期望值 (亦即母群平均數) 等於真實分數,  $E(X)=T$ 。也就是說, 重複對同一受測者實施同一測驗, 所得測驗分數分配之平均數等於真實分數。

3. 誤差分數與真實分數之間的相關為零,  $\rho_{ET}=0$  也就是說, 真實分數較高的受測者, 其測量誤差沒有比真實分數較低的受測者為高或低的傾向。

4. 兩個不同測驗的誤差分數是無關的， $\rho_{E_1E_2}=0$ 。也就是說，如果受測者在測驗一的誤差分數是正的那麼他在測驗二的誤差分數並不會因此而有正或負的傾向。

5. 某一測驗的誤差分數與另一測驗的真實分數無關，即 $\rho_{E_1T_2}=0$ 。

6. 若兩個測驗的  $T_1=T_2$ ， $\sigma^2_{E_1}=\sigma^2_{E_2}$ ，則這兩個測驗稱為「複本測驗」(Parallel Tests)。

7. 若兩個測驗具有  $T_1=T_2+C_{12}$  ( $C_{12}$  為一常數) 的關係，則這兩個測驗為「本質上  $\tau$  相等測驗」(Essentially  $\tau$  - Equivalent Tests)， $\tau$  表示真實分數  $T$ 。

傳統測驗理論的內涵，便是依據上述七個基本假設而來，傳統測驗理論的主要目的，便是希望從受測者的觀察分數去推估受測者的真實分數。

### 2.1.2 試題反應理論的基本概念

傳統測驗理論至今仍是最實用的測驗理論，許多測驗依舊是根據傳統測驗理論的方法來建立其資料間的實證關係，但是隨著測驗需求量的日益增加及形式的多樣化，傳統測驗理論漸不敷需求，試題反應理論 (Item Response Theory, 簡稱 IRT) 於焉崛起，試題反應理論由於試題架構較精細，考慮層面較廣，不僅可以延續傳統測驗理論的功能，並且藉電腦普遍推廣的幫助，試題反應理論因而得以突破傳統測驗理論在應用上的瓶頸。

## 2.2 測驗編制的流程

測驗的編製必須遵循一定的流程，以確保測驗內容與測驗目的相符，也能夠減低其他因素對測驗分數的影響，致使測驗的結果能夠如實的反映受測者所具備的知識和能力。

### 2.2.1 規劃必須符合測驗目標

測驗的編製計畫會因測驗目的的不同，而有不同的考量，因此在進行測驗的規劃時，必須清楚測驗的目標為何。譬如說：一份安置性的測驗，其試題的困難度較低，但必須包含每一項必要的起點行為，以便將學生安置在適當的教學計劃中；一份診斷性的測驗，其試題難度不必太高，但試題必須包含學生常犯的錯誤，以找出其錯誤的類型和原因；一份總結性測驗，其試題必須反映教育目標，且試題難度的分布範圍較廣，以評量學生的學習結果，評量結果主要是用來確定教育目標是否達成。

### 2.2.2 確定測量的「範圍」和「能力層次」

所謂「範圍」是指測驗所要測量的「內容」，如果是學科的知識測驗，則所指的是知識架構中的哪些部分。而「層次」是指測驗所要測量的「能力」，包括哪一個教育領域中的什麼層次。以一份健康教育的測驗為例，在實際命題之前，就必須對測驗所要測量的「內容」和「能力層次」有一個清楚的規劃。首先，確定在整個健康教育的知識架構中，本次測驗所要測量的重點為何。譬如說：健康教育的教材內容包括健康的身體、健康的心理、意外傷害與疾病的預防、環境與

健康等…而測驗的「內容」鎖定在健康的身體、意外傷害與疾病的預防。其次，必須確定測驗所要測量的能力層次。譬如說：在認知領域教育目標中，包括記憶、理解、應用、分析、綜合、評鑑，而本測驗所要測量的「能力」鎖定在記憶、理解和應用三個層次。決定了測驗所要測量的「範圍」和「能力層次」之後，雙向細目表的架構也就跟著確定了。

### 2.2.3 建立雙向細目表的「細目」，並為命題的依據

傳統的做法是，在雙向細目表的細格中，只填入各類試題的比重（百分比）或題數。而比較精確的做法是，必須以具體的行為描述來界定學生的學習結果。這「具體的行為描述」，也就是雙向細目表所謂的「細目」，或所謂「指標」。以健康教育測驗為例：當測驗所要測量的「內容」是「健康的身體」、而測驗所要測量的「能力層次」是「記憶」時，對學生學習結果的具體描述就可以是「能說出促進健康的基本要則」或「能說出人一生當中生長發育最快的時期」；當測驗所要測量的「內容」是「健康的身體」、而測驗所要測量的「能力層次」是「理解」時，對學生學習結果的具體描述就可以是「能區分貧血、血友病與白血病的不同」或「能解釋影響生長發育的原因」，如表 2。

表 2 雙向細目表

健康教育測驗		認知領域教育目標之能力層次		
		記憶	理解	應用
內容	健康的身體	能說出促進健康的 基本要則 能說出一生當中 生長發育最快的時 期	能區分貧血、血友病與 白血病的不同 能解釋影響生長發育 的原因	略
	意外傷害與 疾病的預防	能說出常見創傷的 種類 能說出 A 型肝炎的 傳染途徑	能以自己的話來描述 不同的急救方法 能了解大量出血時對 人體的危害	略

#### 2.2.4 依據命題原則來編擬試題

在實際進行試題的編製時，除了必須依據雙向細目表的規劃，針對特定的「內容」和「能力層次」來命題之外，還需注意一般的命題原則。以選擇題為例，選擇題係由三個部分所構成，分別是題幹 (Stem)、誘答選項 (Distractors) 與正確答案 (Key) (或統稱為選項)。

##### 1、題幹部份：

- 題幹要清楚表達題意，避免過短或過長，且一次只問一個問題。
- 在各選項重覆出現的文字，應放在題幹內。
- 題幹應盡量用正面的敘述，避免使用否定句，如用否定句，則需特別強調否定字。
- 若測驗某「詞語」的定義，則該「詞語」必須放在題幹內。
- 避免提供正確答案的暗示性線索。
- 試題應能測量到重要的學習結果。

- 題幹的敘述應保持完整，避免被選項分割成兩個部份或段落。

## 2、選項的內容：

- 誘答選項必須具誘答性，混有常見的錯誤以及錯誤訊息，應具有與題幹相關性或合理性，以發揮應有的誘答功能。

- 不要任意或過度使用「以上皆是」或「以上皆非」。
- 選項應相互獨立。
- 變化正確答案的長度以排除提供線索的可能。
- 標準答案必須是正確的答案或是最佳答案。
- 選項的敘述應力求簡短，相同的字詞宜放在題幹中。

## 3、選項的排列：

- 盡可能將選項按邏輯次序（數字）或時序（日期）排列。
- 選項的排列如無法用邏輯次序或時序排列，可按筆劃來決定排列位置，或使用「隨機字母表」來決定選項的排列。
- 如果上述原則均不適用，則以隨機方式排列正確選項的位置。

試題產生的程序，從「確定測驗的目的」、「確定測驗所要測量的範圍和能力層次」、「建立雙向細目表的『細目』，並以雙向細目表為命題的依據」到「依照命題原則來編擬試題」，到這裡為止可以算是告一段落了。但是，從試題的產生到一份標準化測驗的形成，事實上還需要經過相當多的程序。如：審題和修題的工作、試題公平性的審查、預試、試題分析一直到題本的組成與定稿等。任何專業測驗的編製，都應該依照這樣的程序來進行〔10〕。

名詞釋義：

題幹：若以  $x$  代表小明身上的錢數，以  $y$  代表小華身上的錢數，且已知  $x$  與  $y$  的關係式為  $x + 30 = y$ ，則下列哪一個敘述是正確的？

選項：(A) 小明與小華共有 30 元 (誘答選項)

(B) 小明的錢比小華多 30 元 (誘答選項)

(C) 小明與小華的錢差 30 元 (答案)

(D) 從此關係式無法得知誰的錢多 (誘答選項)

(1)「題幹」：指試題中陳述問題的部分。

(2)「選項」：指所有試題所列的備選答案。本測驗中每一試題均提示四個選項，並標示為(A)、(B)、(C)與(D)。

(3)「答案」：指該題的正確解答，本身並須具備充足的理由以說明其正確性。

(4)「誘答選項」：指答案以外的所有選項，且每個都需具有足夠“不對”的理由。

a、一個有誘答力的選項，會吸引概念不正確或不具備該能力的學生選擇，但概念正確的學生，仍能分辨出正確答案與誘答選項。

b、為使誘答選項具有誘答力，設計選項時，每個選項的詞語表達、長度、題幹間的邏輯性都應該儘量相同。

一般命題原則

1. 原創性
2. 公平性
3. 其他注意事項

- (1)題意完整明確、提供足夠解題資訊且切合評量目標。
- (2)每個試題只問一個問題，避免同時包含太多概念。
- (3)試題的文字敘述應簡潔、明白，避免出現跟答案無關的內容。
- (4)取材需為課程中重要觀念且情境自然合理，符合學生的生活經驗。
- (5)試題若含圖表，圖表必須是答題重要資訊的來源且須簡潔清楚。
- (6)答案明確唯一，不會引起爭議。
- (7)正確答案隨機變化位置，且其出現在各選項的機率應大致相等。
- (8)選項宜依某種邏輯次序排列並盡量字數相近。
- (9)標點符號的使用要準確。
- (10)圖示一般以參考為主，但不宜與事實差距太大。
- (11)試題盡量避免問「何者為非/錯誤」。

除了一般命題原則外，在題幹和選項上也有以下幾點需要注意。

#### 【題幹方面】

1. 試題的題幹應僅提出「一個明確的問題」。
2. 不要將四個是非題拼湊成一個選擇題，以致缺少一個明確的中心問題。
3. 題幹本身應為完整的敘述，不能只是一個字、詞、或是一個名詞。
4. 若以「未完成敘述句」作為題幹，題幹應能傳達完整問題。
5. 題幹應只有一個核心概念，否則失去診斷價值。
6. 題幹要簡潔，避免無關陳述(100字以內約為合理範圍)。
7. 題意需明確完整，使受測者不必閱譯選項即了解題意。



8. 儘可能將共同的用字移到題幹上。
9. 題幹避免用否定敘述，如須以否定句陳述問題，則需強調否定用字。
10. 試題必須重新組織，避免抄襲課文。
11. 若題幹要求學生從選項中選出一正確者或錯誤者，請使用以下之固定敘述方式：
  - (1)下列有關...的敘述，哪一個是正確的？(哪一個是錯誤的？)  
或
  - (2)(引言)...，哪一個敘述(或選項)是正確的？(哪一個是錯誤的？)

#### 【選項方面】

1. 盡量以隨機方式變化正確答案的位置，正確答案出現在各選項的機率應大致相等。
2. 選項不要放在題幹中間，以免題幹分裂為二。
3. 標準答案必須是唯一的正確答案或是最佳答案。
4. 誘答選項必須具有同質性與似真性。
5. 選項避免出現「總是、一定、絕對、從來、所有」等副詞。
6. 不要過度使用「以上皆是」或「以上皆非」的選項。
7. 選項宜依邏輯次序(如由小至大、時間先後等)排列。
8. 選項之間應為獨立、互斥，避免意義重疊。

### 2.2.5 試題如何產生

確立評量目的→擬定學科雙向細目表→邀題階段 & 命題階段  
→試題修整與審訂→進行預試→分析預試結果→建立題庫→組合正式題本。

## 命題階段工作流程

確認命題範圍→決定命題的能力指標→發揮教師專業，發展試題  
→勾選答案，填寫試題參數→利用試題檢核表，檢查試題→寄回試題。

表 3 數學科雙向細目表〔16〕

細格為 評量目標及比例		能力層次		
		能力層次一	能力層次二	能力層次三
學 習 內 容	學習內容一			
	學習內容二			
	學習內容三			
	學習內容四			

在適性測驗眾多的模式之中，項目(試題)反應理論(Item Response Theory, IRT)由於具有高效率(可大幅縮短測驗時間與長度)，準確度高(可精確計算出受測者的成就等級)，一直都是最常見被研究者提及的適性模式，使得項目反應理論幾乎成為適性測驗的同義詞〔2〕。但是項目反應理論依據其所使用的模式通常需要 200 至 1000 名受測者參與，以校準其題庫中的項目參數(難度、鑑別度與猜測指數)〔3,4,5,7〕。

### 2.2.6 數學科命題理念

從小到大我們由實物上的數計算、實際測量的去學習數學，一直到學習如何將所學得的數學知識應用到生活層面上，更進而抽象式的去思考數學，運用數學於各個學科領域中。這樣一系列的學習數學、理解數學後，我們知道其實數學在我們的生活當中扮演了相當重要的地位。透過數學學習的程序，使我們擁有辨別事實的能力。透過數學學習的方法，使我們加強了對事物組織的能力。透過數學學習的思考方式，使我們擁有邏輯推理的能力。數學，必須是一種連續性的學習。在這一連貫漸進式的理解過程中，適時地加上生活上各種體驗，才能

更深刻的體會出理解進而去運用數學是一件有趣且重要的事。

實施的「國民中學基本學力測驗」其目標就希望所有的國中生都能在學習的過程中擁有充實又快樂的歷程。在數學科的命題理念上，我們期望的是將數學知識朝向運用於生活化的命題方式。雖然在現行國中課程範圍上，數學所教授的內容有其限制，但如能將這些內容充分運用於生活上，我們相信這學習者本身就具備了基本能力。

在分析這些內容時，我們期望經過國中數學教育後，學生在數學上可能可以達成的目標可區分為下列數點：

(一) 習得基本數學概念與對數學大致輪廓的瞭解。

(二) 諸多數學定義、公理、定理、性質的知識與運用，進而能使用於生活中解決問題。

(三) 透過數學問題思考、理解、計算、欣賞等等過程而對學生產生多方面的訓練效果。

以上所列各點，我們將在以下做進一步的分析，分析的方式則是按「學力」與「基本」等兩個概念在各點的意義來闡述，最終目標是推論出基本學力的精神。

就(一)(習得基本數學概念與對數學大致輪廓的瞭解)的「學力」來說，數學元素與體系的構建是人類思考能力由實體觀察邁向抽象概念的一個重要例子，而此一建構過程對此一時期的學生而言無疑是極為重要的，透過此一建構程序，學生將能獲取認知並接受抽象事物的重要經驗。現今國中數學教材多半採取逐步導入的方式來引進各種數學概念，透過相當時間的學習程序，學生將能有系統地建立對數學物件正確的概念。

在這個過程中，我們的目標有二：(1) 學生將習得若干基本的數學知識並且它們將伴隨學生相當長的時間成為他們未來「常識」的一部份。(2) 伴隨正確數學概念的建立，學生得到了某種基本層次的抽象化能力。關於「學力」另外有一點補充，學力乃是經由系統化教育而獲致的能力，而非生而具稟的天賦，或經由自然成長而獲得的能力。因此基本學力的測驗材料與內容必須充分考慮學生正確回答的基礎是否建立在上述的天賦能力上。舉例來說，一個問題正確回答的重點若是在於細心與否，則我們必須考慮「細心」是否是我們認定的「基本學力」。再舉一個例子，若「快速理解新名詞的定義」是我們要培養學生具備的「基本學力」，則在題中說明一個新數學名詞並立刻要求做應用的試題便是恰當的，否則，這樣的試題便是在測驗其天賦的反應力，那便不是我們所要的，此一試題便是不妥的。

就(二) (諸多數學定義、公理、定理、性質的知識與運用) 而言，伴隨著人類歷史而來的數學發展有著相當驚人的龐大知識，這些定理、性質當中對國中生而言僅有極為稀少的材料是適宜於此階段修習的，其中日後能被記住並加以運用在生活中的更是少數中的少數。當然，在國中階段有系統的數學教育後，記住並運用若干人類偉大數學發現與成就是一件重要的事，並且應該列入基本學力的一份，這些知識應該是核心的、具廣泛運用的。至於其他的數學結果與知識，當然有其在數學世界中應有的地位與重要性，只是其性質上可能是特定條件下的推論、非一般性較為零碎的知識與發現或非國中數學核心的部分，但這並無損每一被證明為真的定理、性質或被廣泛接受之公設、公理之價值。

就(三) (透過數學問題思考、理解、計算、欣賞等等過程而對學

生產生多方面的訓練效果)而論,在數學訓練的過程中,除了概念的理解與建構外,一般也輔以相當數量的練習,透過不同的數學命題、不同的條件等等,來讓學習者多面向地深刻認知數學物件,經由一連串命題給定、推論、回答的數學活動,學者可得到適當的能力訓練,這些能力不僅僅是對命題結論的知曉、記憶與運用,更重要的是在此過程中學習到面對不同的數學問題素材,可具有類似的解決能力與模式,這些凌駕於學問題之上,獨立於教材之外可廣泛應用於諸多領域與生活層面上的能力將是最寶貴的成果與資產。這些能力是透過系統化學習而達成的,它們的取得依賴了教材的存在,但有趣的是,在完成了學習的過程後,我們對這些能力的測驗卻能與教材獨立無關,甚至於與學問無關,它們是獨立存在的能力,即使學生遺忘了課上上瑣碎碎的內容,這些能力也不會失去,因為它們已經成為學生們腦中的一部份,就如同吃飯使用筷子一般理所當然。這就是基本學力測驗追求的測驗目標,我們必須對這些能力先有抽象層面上的瞭解與認知,而後方能察覺它們的存在,再將這些能力落實到學科層面,描述為測驗目標,進一步賦予可觀察的操作型定義(指標),最後透過適當的材料選擇作為測驗的命題,巧妙地表現評價學習者在這方面的達成情形。

「國民中學基本學力測驗」根據上述所提三點目標,形成了數學科能力指標的依據(概念理解能力、基本運算能力、問題解決能力)。在命題方向上,希望能運用命題上的技巧能測出真正理解的學生,而非記憶性的背公式或高精熟度的記憶試題模式的學生。「國民中學基本學力測驗」最終的理想是期許學生對數學產生興趣並瞭解其重要性,進而願意進一步學習更高深的數學知識。雖說學生是否願意深入瞭解更高深的學科知識,往往牽涉到學生性向、天賦與需要性,「國

民中學基本學力測驗」還是期望透過快樂正常的學習下能引發出學生對學習真正的興趣〔16〕。

### 2.2.7 數學科基本學力測驗命題理念

數學，就它的歷史而言，是一部人類思考能力由實物觀察邁向抽象概念的發展史。在這些定理、性質當中，對國中生而言僅有極稀少的材料是適合此階段修習的，而其中日後能被記住並加以應用在生活中的更是少數中的少數。儘管如此，我們分析，經過國中階段有系統的數學教育後，學生還是有一些基本的、核心的、廣泛的知識可以獲得，我們期望他在數學上可能可以達到的目標分為下列數點：

- (1) 基本數學概念與數學大致輪廓的了解。
- (2) 諸多數學定義、定理、性質的認識及運用。
- (3) 在數學訓練的過程中，解決問題能力的獲得。

就「基本」概念而言，數學世界的基本概念何其多，且每一概念都有其目前數學體系中的位階與角色，對於其過於基礎者（基本公設、集合論、語意邏輯）、過於應用而分支過細者都不適合此階段學生。我們所要的「基本」概念是指學生將來可以轉換成「常識」的一部份、或是可以帶著走的能力。關於「學力」另外有一點補充，學力乃是經由系統化教育而獲致的能力，而非生而據稟的天賦或經由自然成長而獲得的能力。因此基本學力的測驗教材與內容必須充分考慮學生正確回答的基礎是否建立在上述的天賦能力上。舉例來說，一個問題正確回答的重點若是在於細心與否，則我們必須考慮「細心」是否是我們認定的「基本學力」：若是，這樣的試題是恰當的；若不是，此一試題便是不妥的。

然而在數學訓練的過程中，除了概念的理解之外，一般也輔以相當數量的練習，透過不同的數學命題、不同的條件等等，來讓學習者多面向地深刻認知數學物件，經由一連串命題給定、推論、回答的數學活動，學習者可得到適當的能力訓練；所以記住並應用若干偉大數學發現與成就就是一件重要的事，並且應該列入基本學力的一部份。有趣的是在完成了學習的過程後，即使學生遺忘了課本上瑣瑣碎碎的內容，在面對不同的題材也具有類似的解決能力與模式。這些凌駕於數學問題之上、獨立於教材之外、可廣泛運用於諸多領域與生活層面上的能力將是最寶貴的成果與資產。這就是基本學力測驗追求的測驗目標〔16〕。

### 2.3 適性測驗

適性測驗 (Adaptive Test) 係應用現代心理計量理論—項目(試題)反應理論(Item Response Theory, IRT)—所發展出來之一種新的實施測驗方式。事實上此種施測方式只根據一個簡單的原則，即：如受測者正確地回答一個題目，接下來呈現給它的題目會較前一個題目難一點；反之，倘若受測者答錯了一個試題，則下一個要回答的題目會比前一個簡單一些〔28〕。在施測的過程中，受測者每做完一個題目（不論對或錯），其回答該題的能力將被重新估計，然後根據被重新估計的能力再呈現適合此階段能力估計 (Ability Estimated) 的題目。每次被重新估計的能力將會更可信。循此原則，直到滿足一個預先設定的信賴水準或終止標準時，測驗即結束。

蓋因此種施測方式係採機動的選題方式去配合受測者的表現，換言之，需要從題庫 (Item Bank) 中根據試題的統計特質 (即試題參

數，Item Parameters) 去選題，且受測者每完成一個反應，期能力水準要再被估計，期間，涉及的計算過程頗為複雜，因此需藉助電腦方能實施，所以此種測驗方式又稱為電腦化適性測驗。

適性測驗的目的即是要提供最符合受試能力水準的題目，以獲得描述該受測者能力之最豐富的訊息。為了達到此種目的，目前雖已發展出許多不同的適性程序，然而大抵根據下列步驟〔25,26〕來實施：

在呈現第一個題目之前先約略估計受測者的起始能力，起始能力的估計可由受測者自己估計自己，或根據其他與受測者有關的資料（如年級、成就分數、或其它測驗分數等的加權）來估計。另一種方式是不對受測者做起始能力的估計，而在施測前假定所有的受測者皆為中等能力。

然後電腦即從題庫中選擇一個最適合或最接近受測者之起始能力的題目加以施測。

受測者對所呈現的題目加以反應。電腦隨即對受測者的反應加以評定「對」或「錯」之後，受測者的能力水準又重新估計。

如果受測者被估計的能力已足夠精確，或是達到一個預先設定的終止標準時，測驗即結束。否則，測驗又從第二步驟重新估計。

### 2.3.1 金字塔型適性測驗

如圖 1 所示，金字塔型或樹狀結構型 (Tree-Structure) 適性測驗是一種題目間分支策略 (Interitem Branching Strategy)〔35,36〕。圖中



每一個節點 (Node) 代表一個題目，「+」與「-」號分別表示受測者之「對」與「錯」的反應。

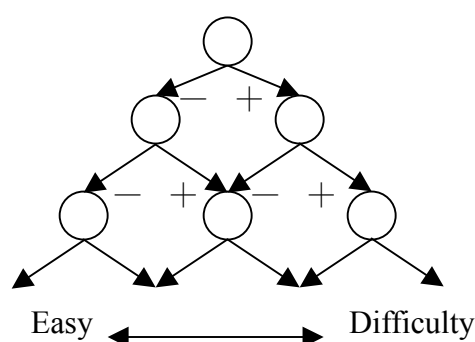


圖 1 金字塔型適性策略圖示

圖亦顯示此種分支策略是一種「階——題一致」之金字塔型 (Constant Step Size Pyramids) 結構；亦即第一階一個題目，第二階兩個題目，餘此類推。測驗的進程則令所有的受測者皆從塔尖的題目 (也即第一題) 開始，若在某一階段的反應正確，則分支至右下一題則接受一題較難的題目；反之，則接受一題較易的題目。每一受測者在任一階段皆只對一個題目做反應。

圖亦顯示，當測驗的階段增加時金字塔式的結構則需要更多的題目。與兩階段式相較，當一個十五題的兩階段式結構則四十五題即足夠，但金字塔式結構則需一百二十八題。另一典型的多階段適性策略即為 Lord 於 1971 年 [29] 所提出之 Robbins-Monro 適性測驗。

### 2.3.2 Robbins-Monro 適性測驗

如圖 2 所示之 Robbins-Monro 適性測驗，其測驗進程與金字塔式類似，但結構上則略有差異。從其結構觀之，此測驗亦可稱為「減階增題式」程序 (Shrinking Step Procedure)，一般言之，第  $n$  階段其需

的題目數為  $2^n - 1$ 。此適性測驗的優點在於能力估計時，能夠迅速的達到收斂。

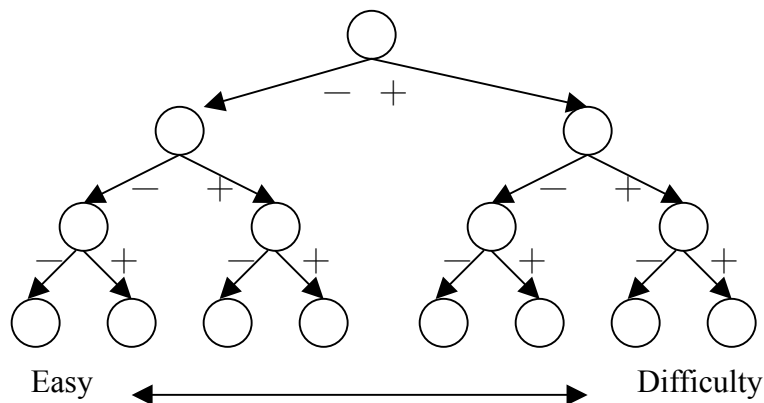


圖 2 Robbins-Monro 適性測驗

### 2.3.3 堆疊式適性測驗

Weiss 於 1973 年 [ 38 ] 提出一種堆疊式適性測驗 (Stratified-Adaptive 或 Stradaptive Test)。此種測驗是依題目的難度分為  $n$  個層次 (Level) 或堆疊 (Strata)。如圖 3 所示，每一堆疊可視為一個難度較叢集於某一點之尖峰型測驗，其難度則由左至右遞增。

堆疊式適性測驗之實施方式是根據受測者之事先能力估計而進入一個適合其能力水準之堆疊。當受測者在同一堆疊中對題目作反應之後，則依其反應之「對」或「錯」而分支至右 (較難) 或左 (較易) 之堆疊，而在任一堆疊中施予受測者之題目則為其未曾做過的題目。如此反覆進行，直到滿足一個預先設定之終止點為止。

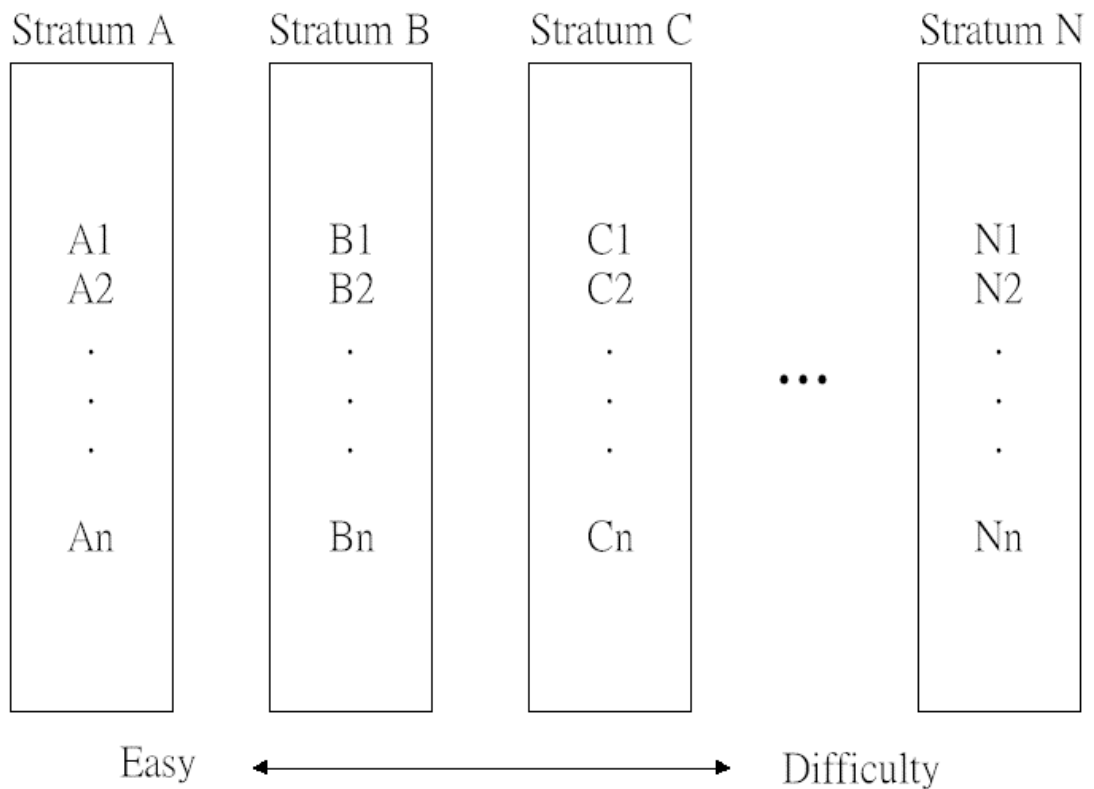


圖 3 堆疊式適性測驗

## 2.4 電腦化適性測驗

電腦化適性測驗(Computerized Adaptive Test;CAT)是測驗理論與實務伴隨電腦科技的發展而形成的產物。至目前為止，CAT 的理論與技術均已十分成熟，一些常見的測驗工具(如 ETS 的托福測驗等)也以 CAT 的方式施測。

近年來電腦測驗(Computer Test)結合項目反應理論及適性測驗之優點，已發展成電腦化適性測驗(Computerized Adaptive Testing)。這種施測方式乃是電腦依據受測者估計的能力水準估計值自動選擇符合受測者能力之題目來加以施測，且可隨時依據受試在新呈現的題目答題的對與錯情形，更新受測者的能力估計並據以選取下一題目，直

到受測者能力之估計已相當地精確為止。此種施測方法，每位受測者被施測之題目不同，施測時間及題數也可以不同。可以在較少的題數達到事先設定的某一精確水準，精確地估算出受測者的能力。

電腦化適性測驗包括幾個要素，不同學者的見解略有差異〔5,1,7,9,12,14,18,23,26,34,39〕，其中 Hulin 等於 1983 年〔26〕的研究指出電腦化適性測驗包括了決定起始點、估計能力與試題選取及結束測量的標準三個要素；而 Weiss 及 Kingsbury 在 1984 年〔39〕則認為有：試題反應模式、題庫、起始點、試題選擇的方式、計分方式、及終止標準等六個要素，其中每一要素都有幾個可能的選擇，端視應用目的的不同而選取最適的組合。

歸納上述學者的見解，茲將電腦化適性測驗的要素分為：(一) 試題反應模式，(二)題庫，(三)起始點的決定，(四)試題選擇的方式，(五)能力估計的方法，(六)終止的標準。茲分述如下：

### 2.4.1 試題反應模式

一般所處理的是二元計分資料，即答對、答錯的資料，受測者答對以 1 表示，答錯以 0 表示。處理二元計分資料的試題反應模式主要有單參數模式、雙參數模式、三參數模式，每種模式各有其適用的範圍及基本假定，需考慮試題的性質適合何種模式，才能選出正確的試題反應模式。

### 2.4.2 題庫

電腦化適性測驗必須含有一個具有試題反應理論參數的題庫。上

述試題反應模式決定之後，即可求出題庫中各試題的參數，而這些試題參數都應在同一量化的單位上。至於題庫的大小，Ree 於 1981 年 [34] 的研究指出，當題庫大於兩百題時，能力的估計並不會比較準確，但他也指出此現象不能推論到所有的情境，必須視題庫的試題性質和試題參數的分配而定。所以目前題庫大小對測驗結果的影響，學者間的看法頗為分歧，實際的研究證據也尚欠充分。比較重要的觀點為試題的難度應均勻地涵括所欲測量群體的所有能力範圍。除此之外，試題的鑑別度愈高則愈能作經濟有效的測量 [27]。

總之，試題的多寡應與試題的品質同時考慮，試題愈多的題庫未必比品質好的試題題庫佳；也就是說，題庫中包括品質好的試題，則題庫的大小就不會構成問題。另外，過多的試題亦會延宕施測時選擇適嘗試題的時間 [1]。

### 2.4.3 起始點的決定

電腦化適性測驗施測時，理論上第一題的難度應與受測者的能力相配合，但並非每位受測者的能力背景我們都能很清楚的瞭解，所以隨機的起始點，或適中難度試題的起始點，都是可以採用的起始點方法。由於電腦化適性測驗試題的難度會隨受測者的作答反應快速的調適，所以偏差的起始點應不至於影響測量結果，但較正確的起始點，的確有助於縮短施測題數，而使能力估計提早完成 [39]。

### 2.4.4 試題選擇的方式

目前研究者或施測者最常採用的兩種試題選擇方法為最大訊息選題法 (The Maximum Information Item Selection Method) [40] 及貝

氏選題法 (Bayesian Item Selection Method) [32,33]。此兩種方法均是題庫中尚未施測的試題中選取一題，其中最大訊息選題法係選取一題能提供受測者目前能力水準最豐富訊息的試題施測於受測者；而貝氏選題法則是選取一題能使受測者能力估計的事後變異數 (Posterior Variance) 為最小的試題施測於受測者。

#### 2.4.5 能力估計的方法

Wright 於 1979 年 [41] 指出，目前研究者最常採用的兩種能力估計法為最大概率估計法 (The Maximum Likelihood Estimation Method) 及貝氏估計法 (Bayesian Estimation Method)。

#### 2.4.6 終止的標準

電腦化適性測驗的特徵之一及受測者接受測驗的長度因人而異，測驗可以一直進行到研究者設定的終止標準為止。研究者可以根據測驗的性質和目的，設定不同的標準，也可以同時設定兩種以上的終止標準 [9]。

電腦化適性測驗終止的標準，主要有三：(一) 當設定的題數達到時即停止；(二) 當能力估計準確度達到設定標準，及估計標準誤差低於預設標準時，即可停止；(三) 當整個測驗的訊息達到某一特定標準時，測驗即可停止 [26,37]。

許擇基及劉長萱於民國 81 年 [12] 指出，終止電腦化適性測驗的方法和前述的試題選擇及能力估計的方法有連帶的關係。以最大訊息選題法選題的話，可設定整個測驗的訊息達到某一特定標準，測驗

即可停止。若以貝氏法選題及估計能力時，則可設定能力估計之變異數小於某一特定值作為終止標準，達到此標準時即可停止測驗。上述兩種方法得到的結果其實是相同的。

為了防止施測時太慢達到上述兩種標準，而讓測驗無止境地施測下去，浪費研究者及受測者的時間，研究者可設定另一終止標準，及設定測驗題數的上限，若此標準先達到，則測驗即可終止。受測者能力值的估計誤差可能很大，若然，則此能力估計值的準確度便值得商榷。

以上所述為電腦化適性測驗的要素，這些要素即為電腦化適性測驗主要的程序茲將電腦化適性測驗的程序簡要，如表。

表 4 電腦化適性測驗的程序簡要

程序	內容
試題反應模式	考驗試題適合何種模式。
題庫	試題反應模式決定後即可求出題庫中各式題的參數（包括：鑑別度參數難易度參數及猜測度參數）。
起始點的決定	選擇題庫中的一題作為受測者第一個施測的題目，通常選取中等難度的試題或採隨機選題的方式。
試題選擇的方式	選擇題庫中其他試題作為下一個施測的題目選題法，通常採用最大訊息選題法或貝氏選題法。
能力估計的方法	估計受測者的能力水準，通常採用最大概率估計法或貝氏估計法。
終止的標準	施測過程終止，終止的標準主要有：設定測驗題數的上限、估計標準誤差低於預設標準、及測驗的訊息達到某一標準。終止的標準可同時設定兩種以上。