

## 第二章 文獻與理論探討

### 第一節 題庫的概念

#### 一、題庫的意義

題庫 (item banking) 的意義有廣有狹，較廣義的說法指的是以「任何方式」建立的題庫，亦既事先將考試或測驗的題目，妥為擬定，保存在一個適當的處所；等到需要時，再依照測驗的目標去選用（黃堅厚，民63）。就像社會上的其他「庫」，例如，銀行有「金庫」、醫院有「血庫」一樣，功能是類似的。然而，一個堪用的題庫不應只是用來收集試題而已，題庫若要能夠發揮其功能，那麼，題庫中的試題就必須經過適當的分類，它的特性也應經過合理的勘測。

吳幼吾（民74）認為：所謂題庫事實上就是將一羣題目加以編號儲存起來以利日後的尋索，若題目本身未加以歸類只是將其集合在一起並不能稱為題庫。所以，更嚴謹且結構化一點來說，所謂題庫指的應就是一大羣使用起來很方便的題目之匯集（Millman & Arter, 1984）。這「一大羣」指的是題庫內的題目數量應該為任何應用場合中所需題目數的幾倍以上，而使用方便指的是題目經過編碼及資料結構等的處理，使蒐尋很容易（洪碧霞，民81）。

較為狹義的題庫則係指經由「電腦作業」所建立，而題目總數為任何應用場合所需數量之幾倍以上，且經過題目特徵編碼規劃及資料結構化處理等程序，必須利用電腦來作各種處理之題目所組成的集合。Lippey（1974）就堅定的認為，一定要經過電腦化（computerized）後才能稱得上是題庫；Booth（1983）則提及題庫中的所有題目都應該要有一定的品質保證，方得進入題庫；Mead（1981）更嚴格的界定題庫中各個题目的統計（難度）參數都應量尺化。若此，較狹義的

題庫不管從理論或是實際的使用層面上來看，應該可以將其稱之為「電腦化題庫」(computerized item banking)。

電腦化題庫是指經由電腦作業所建立，並經題目特徵編碼及資料結構處理等程序之一大羣题目的匯集。這樣的題庫除了具備廣義題庫應有的良好特性之外，並可利用電腦來作保存、蒐尋、刪除、更新、擴充、修改及維護題庫內試題等題目管理的工作，以作為支援「傳統紙筆測驗」、「電腦化輔助測驗測驗」及「電腦化適性測驗」等相關之其他用途，以充份發揮題庫評鑑與教學的功能。其中，「維護題庫內試題」的基本概念乃是希望能校準(calibrate)題庫中所有的題目成為共同的變數(common variable) (Mead, 1981)。電腦化題庫的另一個泛稱為「精緻化題庫」，若僅依一個特定格式，挑選試題鍵入電腦中，並不對題目進行試題分析，或其他的題目特徵編碼者，他只是一羣利用電腦來儲存的题目的組合，常通稱為「教師參考用題庫」或「基本題庫」，此一類型題庫並不能稱之為「電腦化題庫」。

另有一「標準化測驗題庫」的名詞，所謂「標準化」應該具備如下之三要素：(張榮祖，民81)

1. 屬性：評量教學目的實質內涵之要素。
2. 難度：評量教學目的達成程度之要素。
3. 題型：評量教學目的工具性之要素。

換句話說，至少具備上述題目特徵編碼與題庫規劃之題庫可稱之為「標準化題庫」。當然，一套功能規劃完備的題庫需要考量的地方還很多，例如對試題反應理論IRT的參數編碼規劃等。發展標準化測驗題庫主要目的是提供教育行政當局作為重要教育決策(例如跳級、保送、資優生甄試等)之鑑定工具。

所以說題庫可以簡單如一疊題目卡片，也可以複雜得非以電腦處理不能運作，就如有的系統之題目特徵編碼高達57個變項(MEDSICH, Hazlett, 1973)，針對題目所作的界定越嚴格，那麼所要付出的代價就會越高，但是當題庫大小一定時，則它所將具備的應用潛力與價值

就會越高，但是當題庫大小一定時，則它所將具備的應用潛力與價值就自然越高。有些題庫是專為教師個人而設計，有些則只能由專業機構控制（洪碧霞，民81）。

本研究往後所提之「題庫」，指的乃是前述所提之「電腦化題庫」而言。

## 二、題庫的功能

Hsu & Nitko (1983)在「教師用微電腦測驗軟體」一文中曾經指出，微電腦能夠協助教師們處理十七種測驗發展方面的相關重要工作，而這十七種工作又可歸納為四大類：

1. 試題編製與組合。
2. 測驗之執行。
3. 成績之登錄。
4. 計分與解釋測驗結果。

然而，不管是以上任何一個工作或任何一種類型，題庫的建立是一項最早、同時也是最基本的工作。

題庫依使用層次而言，有的可能用在大學，有些可能用在中、小學，但這些題庫的功能卻都十分接近。通常題庫中的題目數量要比實際選用的題數為多，而在一般更為精緻化的題庫當中，每一個題目都按科別、課目、目標、主題、單元、題碼、難度...等題目特徵編碼加以分類規劃。

題庫的建立是達到學習評量正常化的最好方法。題庫的存在並不百分之百的保證其有用性，如果要能達到真正有用的地步，則必須符合資料的結構性與良好的題庫編碼規劃，並具備良好品質的試題，具備這樣特性的題庫才能發揮其的真正價值與功能。一個精緻化的題庫將可提供評量和教學的功能，現將其分述如下：

## (一). 評量的功能

### 1. 做爲安置性評量的參考

教師在教學前要先考慮是否對受試者進行下述兩種測驗：

#### (1). 準備度測驗 (readiness pretests)。

此種測驗在某一單元或某一課程開始之前實施，測驗內容涵蓋有效學習該課程所需具備的知識及技能。受試者若缺乏必備的知識或技能，則可進行補救教學或者將其編入特殊班級，施予特殊訓練。利用題庫中具有良好品質的題目來對學生做施測，可有效地分析學生所具備的基本知能。

#### (2). 安置測驗 (placement tests)。

利用題庫中的題目測試學生，並就測試結果與現有常模做比較，則將可研判出學生的現有內在環境（如智力、學習成就. . .）之程度與外在環境（如校外的學習、家庭的學習）之現況，並依此分析出一個妥善的位置，以便將受試者依個別差異安排在不同的起始點，並作爲分組、分班，設計不同的教學內容、不同教學方法以及不同的輔導方法。

### 2. 做爲形成性評量的參考

在教學進行過程中，若要測知學生的進步情形，則需要進行「形成性測驗」(formative tests)。一般而言，形成性測驗常用來測量學生對某些章節學習結果的精熟程度，此種測驗類似傳統上教師常用的隨堂考試(quiz)或單元測驗(unit tests)。

形成性測驗以某一「通過標準」（教學目標）為參照的依據，也就是「標準參照的」（criterion-referenced），它只重視學生是否達到預定的標準，或者達成學習目標，而不重視與別人比較之相對地位。而題庫有常模可用來作為學生學習成熟度的測量標準。若發現未達學習成熟的單元，可考慮重新教學或者再次加強輔導。

### 3. 做為診斷性評量的參考

在教學進行過程中，當學生的學習困難一直持續下去，已無法以形成性測驗所提供的方法來解決時，就需要更為深入地來研究學生之學習困難所在，那麼這個時候「診斷測驗」（diagnostic tests）便可派上用場。診斷測驗需在每一特定的教材領域均有大量的測驗題，題目與題目之間的差距甚小，所以能夠鑑別特殊的學習盲點所在。因此，在學習過程中，可使用題庫測出學生學習困難的地方、類別，再從題庫中選出各種適當的題目來進行補救教學。

### 4. 做為總結性評量的參考

在某一課程或教學單元結束時，我們需要了解那些學生已熟悉本課程或本單元教材，可進行下一單元或課程的教學，而每一個學生的成績等第又是如何？

因此，在教學結束時所實施的成就測驗，其主要目的在於決定那些學生以達到學習目標，以及評定學生的成績者，稱之為「總結性測驗」（summative tests）或期終測驗。

用題庫中的題目來做為總結性的評量是最為妥善與公正的，它有廣泛評量所獲致的常模來做為客觀的評量標準，並可提供評定學習成就的依據，以為學生應該重讀或可予升級的指標。

## (二). 教學的功能

題庫可提供精熟學習、進階學習指示及補救教學的教材。

題庫中的試題都是經過雙向細目表分析各學習單元題目分配等先置作業後，並經命題、題目收集與分類、試題考驗及資料分析與修正等反覆處理的編製過程。除此之外，每一個試題都是經過試題分析來附諸予題目本身的難易度，以及題目所屬的等級層次。

經過題庫的編碼以及題目特徵編碼作業之後，當教學者需要某種層次的試題時，便可依照所需的難易度、效度、鑑別度、認知層次、需求題型... 等等題目特徵，透過電腦化的利益來快速挑選出適當的題目作為學生自習及課堂練習的材料。

人們對評量內涵建構的意義有了越來越嚴格的期許，並且希望教學評量能提供豐富而且具體的教學診斷訊息，以為日後成為發展適性教學 (adaptive instruction) 的參考依據。

邇來，評量變項量化的方式由靜態變為動態，測驗的內容（試題）也由嚴格的標準化而漸趨強調適性化 (adaptive)，更進一步希望將測驗融入於教學歷程中，教師（或電腦）只是在學生學習的歷程中搜集記錄學生對一連串設計過之作業的反應，以便掌握學生情況，了解學生對該學習領域的潛能及可能適合的學習方式或途徑，以提供學生的具體學習資源和建議。

雖然評量變項量化的方式已逐漸朝向各學習單元電腦化動態評量系統的發展。然而，在相關電腦輔助教育 (CAE) 系統及專家系統尚未發展成熟之前，它的功能仍然非常有限，但是至少它可以達成特定領域學習過程記錄的工作，事實上，教學評量本來就是學生學習歷程的動態記錄。

### 三、題庫的分類編碼

當前是資訊一日千里的時代。資訊時代，就必須應用科學的方法來做管理的工作，充份運用電腦的利益，將管理工作變為便捷而有效。

題庫的建立，必須先經科學的規劃作業，並對試題之性及其質編碼(coding)預作分析與歸類。從「科際整合」的觀點，題庫建立的前置規劃，應能考量日後有關不同題庫之間的互通性、相容性以及統整性，以期達到資料流通、資源共享的目的。

題庫的分類編碼作業就是要建立易記、易找、易尋的科學管理方法，當一看見資料的名稱，就能夠很快的讀出它的代號來。也就是說，資料或檔案的代號應該變成電腦的符號，當我們有條理、有系統，條理分明、系統分明、層次分明的來進行編碼分類的作業之後，題庫資料的規則性特徵就變成電腦的符號了。若此，題庫建立的內涵就變成一個圖形化、數字化、標準化、電腦化的科學程序了，

張榮祖(民82)認為題庫的組織方式可分為兩種：一種是採用關係資料庫結構，另一種是按層次分類方式。

#### (一).關係型資料庫結構

##### 1. 實體元素

採用「關係型資料庫結構」時，涉及到的實體有下列諸端：

- (1).冊次
- (2).章次
- (3).節次
- (4).屬性
- (5).難易度

- (6). 試題
- (7). 答案
- (8). 解題過程

## 2. 邏輯關係

其箇中又有幾種不同的邏輯關係：

- (1). 一對多關係：
  - a. 冊次和章次
  - b. 章次和節次
  - c. 冊次和試題
  - d. 章次和試題
- (2). 多對多關係：
  - a. 試題和屬性
  - b. 試題和難易度
- (3). 一對一關係：
  - a. 試題和答案
  - b. 試題和解題過程

## 3. 特色與限制

- (1). 特色：
  - a. 採越高級的邏輯範式，將可減少數據的冗餘。
  - b. 大大增加查詢途徑，方便使用者應用。
- (2). 限制：
  - a. 題庫系統的設計與實現較困難。
  - b. 試題與答案的記錄長度可改變。
  - c. 僅能提供文字儲存訊息，較難提供圖形儲存的訊息。

### (二). 層次分類方式



## 1. 層次元素

採用「層次分類方式」時，分類的依據有下列諸端：

- (1). 科目
- (2). 冊次
- (3). 章次
- (4). 節次
- (5). 屬性
- (6). 難易度

## 2. 邏輯關係

循序單向規劃，並據之分類試題和答案。

## 3. 特色與限制

- (1). 特色：
  - a. 題庫結構簡單。
  - b. 題庫系統的設計較簡單。
  - c. 滿足使用者對題庫的一般查詢要求。
- (2). 限制：
  - a. 綜合性試題的隸屬劃分較麻煩。

洪碧霞（民80）也認為分類系統是開啓題庫之鑰，最常見的是依內容分類，兩種可行的方式，一種是依主題或教學目標分，這種分法有時可以有層次結構，像數學底下有整數的運算一支，而整數的運算下又可細分為其他目錄。另一種方式是採用關鍵字，即每個題目都以一些字或碼來描述其內容，也可將目標以關鍵字來編碼。這兩種分類方式並不彼此排斥，有些系統就結合了兩種編碼蒐索方式。

一般而言，關鍵字比較富彈性，可以同時適用於目標、內容、年級及思考歷程等。同時當課程修訂時，關鍵字系統的修訂比較容易。但如果分類系統可以明晰的界定（如生物學的種、屬、科、目等），或者當

電腦軟體無法處理多重關鍵字時，固定的分類方法就比較適當。

蔡松齡（民81）認為題庫命題作業的規劃工作，有下幾項應該加以注意：

### （一）. 題目代碼付與

代碼在電腦作業中應具有唯一性，且各碼應具有標示作用，因此他建議以九個碼來區分：

〔實例〕 0 1 0 1 0 1 0 0 1

〔說明〕 1~2碼——科別區分。如國小、國中、機械科...  
，英數並用，依實際需要區分。

3~4碼——課目區分。如國文、英文...，數字表示，每科別可分99課目。

5~6碼——單元區分。如單元一、二...，數字表示，每課目可分99單元。

7~9碼——題碼區分。如001、002...，數字表示，每單元可分999題。

### （二）. 科別課目代碼編排

就各級學校課程標準所列科目、課目內容加以規劃，且避免產生重複現象。此項工作應由教育主管單位或學校校務單位統籌辦理。

### （三）. 單元編排與題型設計

就單一課目課程內容加以編排，使題庫取材能均勻分佈於全課目中。另就課程特性設計「主觀測驗題」題型，例如問答、計算...。此項工作應由負責命題之教師或教學研究會負責完成，其方式如表2-1所示：

表 2 - 1 主觀題型命題示例

索引	科別	課目	單元名稱	客觀題題數	主觀題型	主觀題題數
010101	共同科	國文	師說	20	問答	5
010102	共同科	國文	廉恥	23	問答	4
010102	共同科	國文	桃花源記	22	問答	3

(四). 教材內容與教學目標相結合

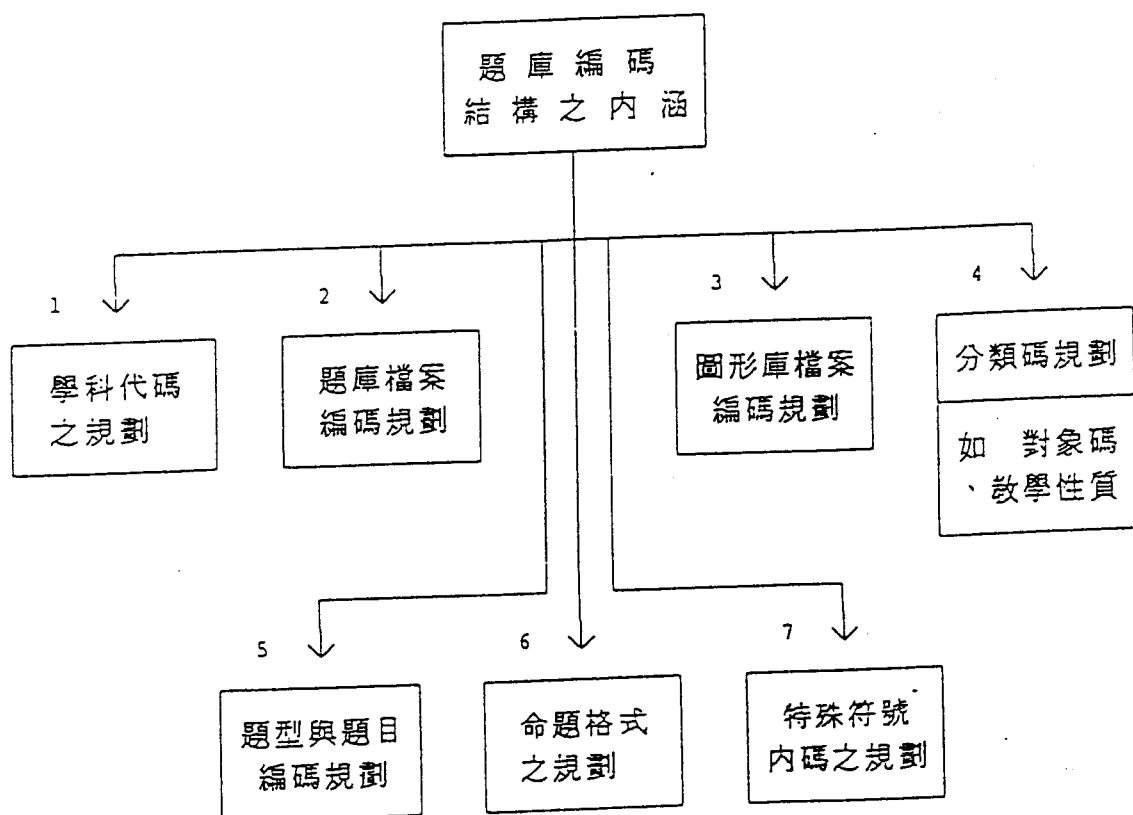
在命題前並應先分析課程材料內容及其行為目標，進而使兩者結合成「雙向細目表」。如此，試題的內容始能具有適當之代表性。如表2-2所示：

表 2 - 2 雙向細目表示例

教材內容	教 學 目 標						合計題數
	知識	理解	應用	分析	綜合	評鑑	
師說	9	5	3	2	5	1	25
廉恥	10	8	3	2	1	3	27
桃花源記	10	3	3	5	2	2	25
.....							

戴建耘（民82）認為：基於未來題庫之整合及維護考量，題庫應該採用開放式的架構以及標準文字檔，而且要妥善規劃檔案編碼結構。他認為題庫編碼規劃工作決定整合發展之可能性，同時也將題庫編碼規劃工作之內涵（如圖1-2）歸納為下列幾項：

- (一)、學科代碼
- (二)、題目檔名編碼
- (三)、圖形檔名編碼
- (四)、分類碼：教學性質碼、認知層次碼、對象碼與年度碼
- (五)、題型與題目
- (六)、命題格式
- (七)、特殊符號內碼



## 圖 2 - 1 題庫編碼結構之內涵

經由上述的探討，吾等可以將電腦化題庫之分類規劃工作分為四大類：

### (一). 題庫總體性規劃

題庫的建立技術是我們要探索的，但是建立題庫的主要目的及其所賦予擬達成的目標才是我們所最希望達到的。未經由妥善、合理、科學並且富有正面意義之總體規劃所建立的題庫，不但將無法達成既定的目的與目標，恐會有形成負面效果。

因此，對於題庫建立的宗旨以及具體目標應先予以釐清與界定，有了建立題庫的明晰宗旨與目標之後，再分析其建立用途與使用範圍，有了這些總體性認知，據之以進行各分項規劃才有意義，也才不致於投入了大批的人力物力財力，卻建立了只有反效果，而且也沒有什麼意義及價值的題庫。

### (二). 檔案分類規劃：

針對試題文字檔（題目檔）及試題圖形檔檔案加以有條理、有系統，條理分明、系統分明、層次分明的來做最科學的編碼規劃與分類，務期建立檔案易記、易找、易尋的檔案分類規劃特性。

### (三). 題目特徵編碼規劃：

對於題目特徵的編碼使得單調的題目變成具有生命力。賦予题目的特徵編碼越多，他所能具備的應用潛力也就越強，當然建立時所必須付出的代價也就越高。常見的題庫特特徵編碼項目列舉如下諸端：

## ● 題目分類屬性部份 ●

1. 學校層級代碼規劃
2. 科目選擇代碼規劃
3. 冊別範圍代碼規劃
4. 章節範圍代碼規劃
5. 題目建立年度代碼規劃

## ● 題目用途部份 ●

1. 教學用途代碼規劃
2. 對象(學習成就)代碼規劃
3. 認知層次代碼規劃

## ● 題目格式部份 ●

1. 題型類別代碼規劃
2. 選項數代碼規劃
3. 答案及解題說明規劃

## ● 題目統計特質部份 ●

1. 傳統測驗－難度 P
2. 傳統測驗－難度指數  $\Delta$

3. 傳統測驗－鑑別度  $D$
4. 傳統測驗－點二系列相關  $r_{pb}$
5. IRT (試題反應理論) 題目參數－鑑別度  $a$
6. IRT (試題反應理論) 題目參數－難度  $b$
7. IRT (試題反應理論) 題目參數－猜測度  $c$
8. IRT (試題反應理論) 題目參數－適合度考驗  $\chi^2$

(四). 建題格式規劃：

進行電腦化題庫題目鍵入工作之前，應先對各種不同題型之建立與儲存方式加以探討規劃，使得建立的格式有統一性的規劃，規劃的過程應考量各種題目特徵的編碼規則。

(五). 特殊字碼分類規劃：

題庫規劃與建立時，不可避免的會使用中文系統所沒有的特殊字元及符號，如英文音標、英文及數理的上下標、罕見的中文字等。以「倚天中文系統」為例，則需借助「造字檔」來加以解決。若各軟體間各有各的設計而互不相通，則勢必將影響到題庫日後的推廣與整合，因此，特殊字碼分類與規劃處理是需事先加以考量的。

#### 四、題庫的建立

題庫是題庫系統的核心，題庫內容的好壞決定題庫系統的成敗。這就是要求選取能覆蓋全部教學範圍、嚴謹地編審選取各種不同題型屬性和難易度的試題，並且數量上必須滿足需求（張榮祖，民81）。

建立題庫最重要的一件工作就是如何來選擇高水準的題目。有了高水準的題目，有關教師命題時的困難就可迎刃而解，Hambleton(1982)及Popham(1981)曾分別提及判斷法(judgement methods)及經驗法(empirical methods)來選擇題目。所謂判斷法乃是聘請學科專家們舉行專家會議(expert conference)，藉以評定所欲選擇的題目品質，並且判定是否與所定的目標之間和諧。經驗法是採用檢查受試者對題目反應的情形，以做為修正題目的參考。

題庫的建立是經常性的，因此無所謂完成，就電腦系統而言，他必須配合科技的進步使運作更快速、精緻、友善，就題目而言，他必須配合課程、考生或目標的更易，不斷的擴充和更新（洪碧霞，民82）。

##### (一) 題庫建立的目標與注意事項

###### ◇ 目 標 ◇

1. 題庫的命題要符合教材的基本精神及學習目標。
2. 題庫的命題需符合評量原則。
3. 題庫試題能引導學生走上正確學習方式和正常學習方向。
4. 能協助教師們編製更適切的教學評量工具。
5. 能協助教師們分析教學決定的形成。

###### ◇ 注意事項 ◇

1. 題目的內容需完整、題意清晰明確。
2. 題目付予之條件須完備，文字敘述簡明扼要。



3. 題目需注重邏輯的推理性與診斷性，內容要涵蓋全部教材。

## (二). 題庫建立的適當時機

不管何種性質的題庫，其建立過程均需投入相當之人力、物力與財力，因此需審慎評估其需求後，始可加以進行，一個題庫至少需符合下述的任何一項需求，題庫才有建立的價值 (Millman & Arter, 1984；何榮桂，民80；洪碧霞，民81)：

1. 現有測驗不被接納，而有編製屬於自己測驗的需求時。
2. 經常需要用到測驗時。
3. 實施多式平行測驗是必要時。
4. 需要實施個別化適性測驗時。
5. 測驗使用者與發行者願意充份合作建構一個滿足自己需求的題庫時。
6. 已有題庫系統(item banking system)可資使用時。

## (三). 題庫建立的步驟

題庫因測驗方式之不同而有不同程序之建立方法，現將其分述如下：

### 1. 以傳統測驗理論為基礎之題庫

以傳統測驗理論為基礎所編製的傳統式測驗，一般以團體為基礎的試題統計量數 (group-based item statistics) (如 P 值) 來建立結構化的試題組合 (亦即題庫)。

蔡松齡 (民81) 認為不論是廣義的一經由任何方式，或是狹義的一經由電腦作業所建立之題庫，均不外乎如圖 2-2 之作業流程：

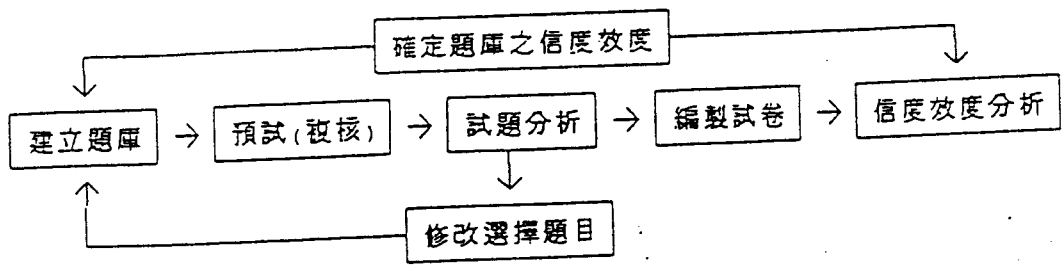


圖 2-2 廣義題庫建立流程圖之一

陳繼逸（民80）認為題庫的建立依其程序而言可為如下所示：

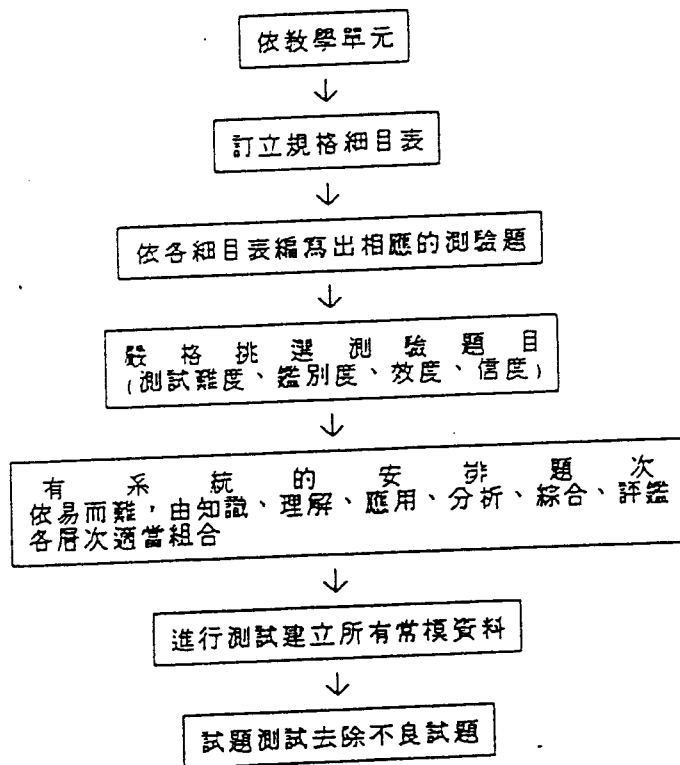


圖 2-3 廣義題庫建立流程圖之二

江文雄等（民82）曾在國中數學科題庫之發展計畫中提及計畫中之題庫建立步驟如下：

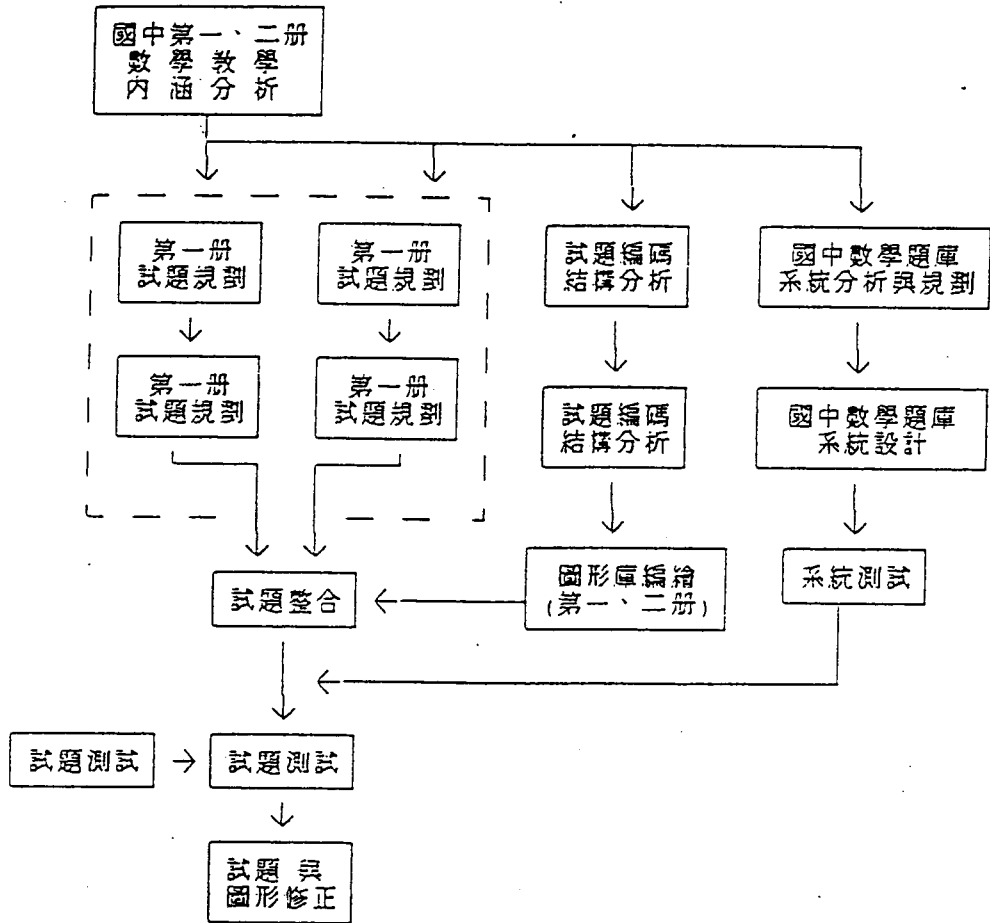


圖 2-4 教育部國教司國中數學題庫建立步驟

張輝政（民82）在高級中學題庫建立之現況報告中提及題庫建立之階段過程如下：

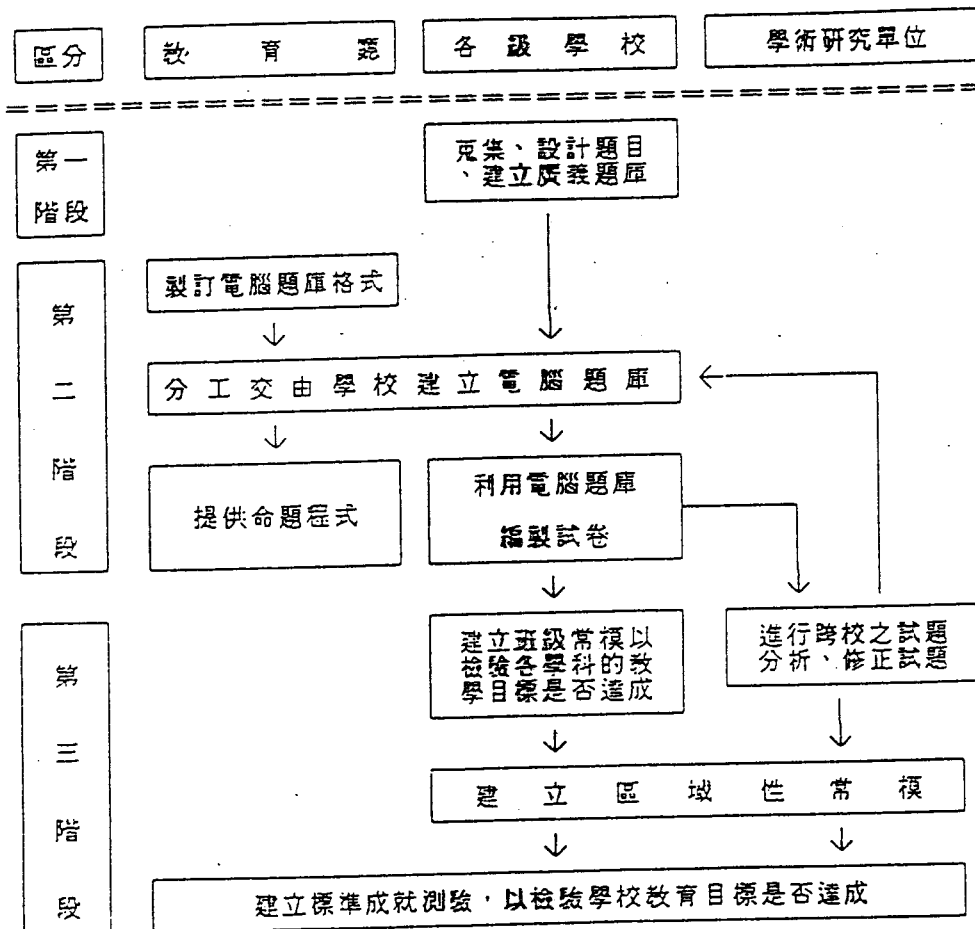


圖 2- 5 教育廳高中題庫建立之階段流程圖

以傳統測驗理論為基礎所建立之題庫有其難以克服的缺點，亦即當題庫試題進行更新（update）時題庫將產生很大的限制。因為以團體為基礎的統計量數，會因不同的測驗或不同的預試樣本而使新舊題目間產生非線性關係，而導致新

舊題目間之量尺的特質 (scale property) 不一致；也就是說，當題庫要進行更新時，不易將不同試題的試題組合之 P 值或 D 值重新量尺化 (rescale)，因此，每次建立的題庫只是用於單一的測驗 (何榮桂，民80)。

## 2. 以試題反應理論為基礎之題庫

當我們實施電腦化適性測驗 (computerized adaptive testing) 時，是以試題反應理論 (item response theory; IRT) 為基礎來建立題庫：亦即在建立題庫之前，須選擇適當之試題反應模式 (item response model) 將試題加以校準 (calibration)，然後將校準過 (calibrated) 的試題參數 (item parameters) 挑選品質較佳之題目，以建立標準化的題庫 (calibrated item bank) (何榮桂，民80)。

由臺灣省教育廳委託臺南師院測驗發展中心所進行之「改進中小學評量專案」，其有關建立國民中小各學科成就測驗題庫建立之研究 (一) 中，由題庫試題組成測驗之編製過程如圖 2-6。

許擇基 (民78) 認為題庫不只是用來收集試題，題庫要發揮功用，試題必須經過適當的分類，其特性也應經過合理勘測。他並具體將建立以試題反應理論為基礎的題庫，其建立與應用歸納為十一個步驟，圖 2-7 便是描述每一步驟的職務：

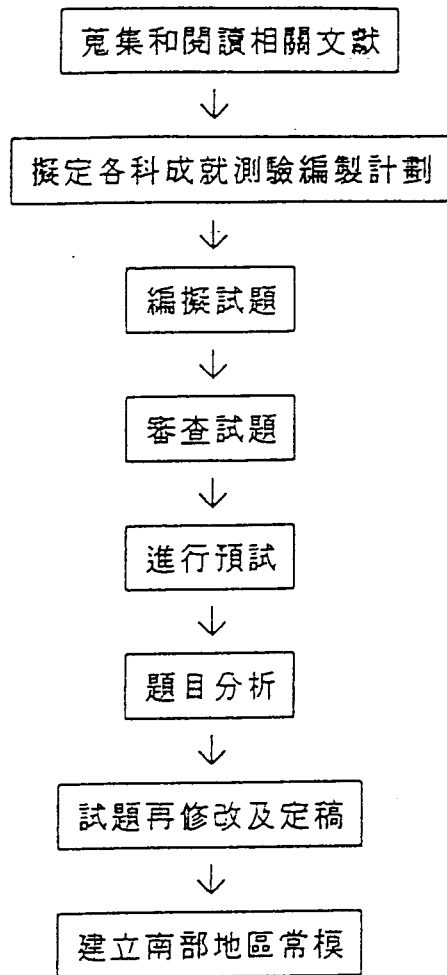


圖 2- 6 教育廳國民中小學題庫試題組成測驗之編製流程圖

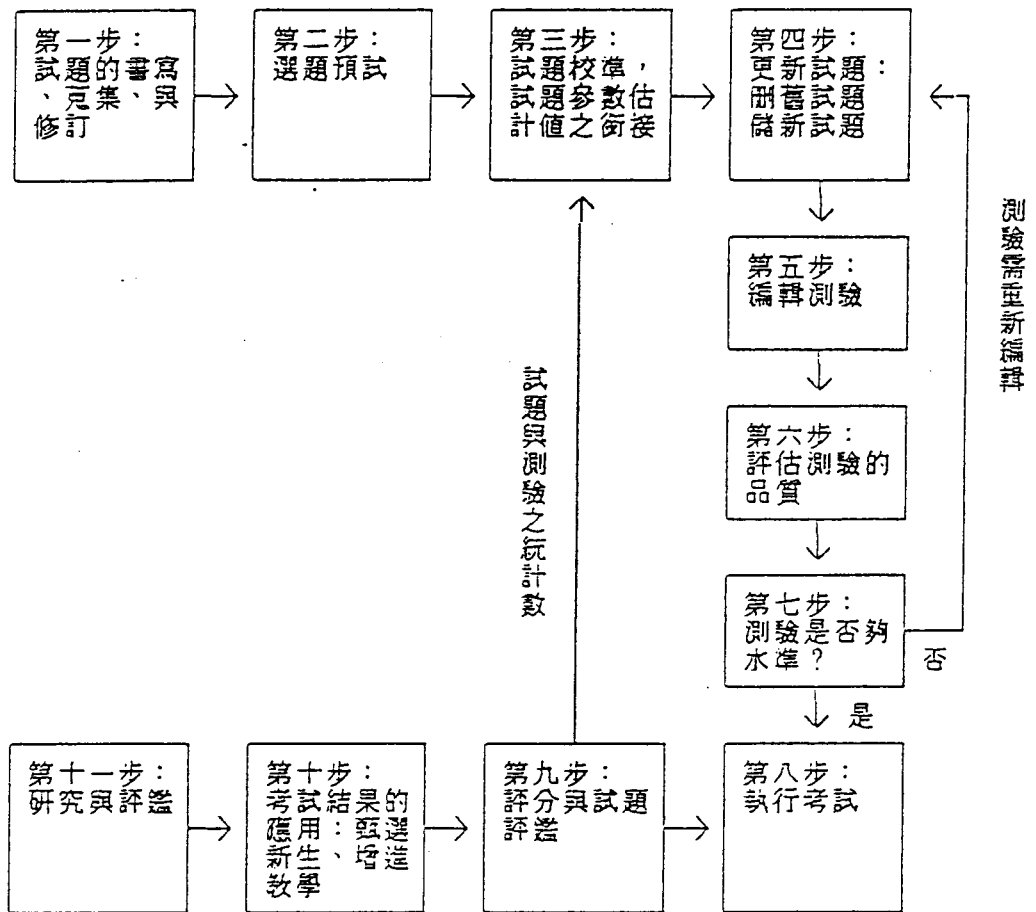


圖 2-7 以試題反應理論為基礎建立題庫之步驟

以試題反應理論的方法來校準，係因樣本獨立 (sample independent)，因此更新題庫中的題目時，仍可用試題反應理論的方法使新舊題目之間等值化 (equating)；亦即使新舊題目之參數在同樣的量尺上具有同樣的統計特質 (Vale, 1986；何榮桂，民80)。

## 第二節 電腦化題庫系統

題庫乃是包含許多為某些目標、技能或工作 (task) 而編製的題目，以為測驗編製者為某些需求 (目的) 編製測驗時的基礎。換句話說它是許多題目的匯集。當一個題庫包含有具備“內容效度”而且技術品質優良的試題時，測驗的編製就容易多了，其所產生的測驗通常比沒有題庫的強況下要高 (洪碧霞，民82)。題庫是題庫系統的核心，題庫的好壞決定題庫系統的成敗，而功能強大的題庫系統才能使良好品質的題庫發揮其應有的性能於極限。

### 一、電腦化題庫系統之發展原則

戴建耘 (民80) 認為：為使題庫之發展能獲致實用並具有彈性化，以下幾點題庫系統的發展原則是必須參照的：

#### 1. 相容性

可使用於所有的PC相容的單、彩色 (MDA、EGA、VGA) 系統。

#### 2. 經濟性

(1). 完全以軟體方式作試題管理與測驗。

(2). PC硬體設備無任何限制，即使PC/XT之機種亦可使用，以利研究後之推廣。



### 3. 操作便利性

- (1). 以圖素符號方式 (Icon)、視窗及下拉式清單作為操作指示與選項介面。
- (2). 系統管理者 (教師) 可使用鍵盤或滑鼠控制。
- (3). 學生用之題庫測驗系統磁片，可在系統下直接開機，不須任何中文系統，減少受試者操作之困難度。

### 4. 功能性

- (1). 提供「試題分析系統」，可直接讀取學生輸入之資料檔作為資料分析。
- (2). 系統可根據教師設定之參數或定義作為試題選題或測試者之環境畫面。
- (3). 提供畫面測試與排版試卷印刷兩種作業模式。
- (4). 具有「圖文」式處理之題庫系統。

### 5. 擴充性與整合性

- (1). 採用開放式架構與模組化之系統整合方式。例如，可讀取任何文書編輯器之標準文字檔，以便作為試題來源，亦可產生一定格式之文字檔，以便供外部之教育統計軟體工具使用。
- (2). 圖形檔可以直接由PC上之任何向量式或位元式圖形來供應。

張榮祖 (民82) 亦從使用者介面的觀點提出說明，為他認為介面設計的適切與否，影響測驗可信度甚鉅，以下幾像是為設計時的考慮重點：

1. 簡單化：受測者對電腦操作的熟悉度差別頗大必須盡量降低排斥心理。下拉是功能表及Icon可避免記憶指令的困擾，提高親和力。對話框可以增加選項的方面性，亦可導引受試者操作路徑獲得到立即式的回饋。滑鼠可以減少答題時間，提高受測效率。選目的Key可為中文：甲乙丙丁，英文：A B C D，數字：1 2 3 4，或他種符號。系統必須能辨認選目型態而顯出來。
2. 一致化：介面規格，功能表文字，Icon顯示地點及大小，功能鍵，Hot-Key的設定，以及題目、選目、繪圖區論文字作區的顯示區域及方式等等，均宜為前後一致。
3. 單題顯示：據行為科學家統計結果，認為一題作答完後再進入下一題比螢幕上同時顯示多題且用ScrollBar來調整較能提高受測者的注意力。
4. 導引及輔助：在線上測驗消除受測者對電腦的恐懼及操作不便帶來的厭煩心理，是很重要的一點。而系統功能強大，與操作複雜度通常是成正相關的。所以應在功能強及不流於繁瑣下，取得平衡。

## 二、電腦化題庫系統之功能需求與架構

一個電腦化題庫最基本的功能至少必須具備編輯題目的能力；進一步來說，也即其功能至少包括編製（create）、儲存（store）、取存（retrieve）、以及修改（modify）試題等功能。具備題目編輯的能力是電腦化題庫系統的必要條件，而不是充份條件。

從電腦輔助測驗的觀點來看，認為系統至少應具備下列四個條

件（何榮桂，民80）：

1. 具有試題編輯的功能，以作為編輯試題之用。且此編輯系統亦能處理特殊字元（如數學符號等）、圖形，並可連接影像掃描機（Scanner），以直接輸入外在的圖形或影像，避免試題中需要圖形時編輯上的困擾。
2. 具有儲存試題（即題庫）的功能。若將題庫與編輯系統配合，則可利用編輯系統更新題庫中的試題。
3. 具有分析個別試題（如難度、鑑別度及誘答情形等），以及估計整個測驗之信度及效度的功能。
4. 具有線上（on-line）施測（個別或團體）的功能。
5. 具有儲存資料（既測驗結果）的功能。所儲存之資料可作為改進試題品質、診斷、或做為進一步研究之用。

張榮祖（民82）也認為線上測驗系統之功能一般包含下述諸端：

1. 測驗生成：包含試卷選定，試卷聯集，測試規則設定，答案規格設定。
2. 試題管理：負責試題擷取，題型判斷，測驗規則的監督執行。
3. 物件顯示：負責控制物件之展現。
4. 答案輸入：按照不同題型提供各式（選擇、填充、問答、作圖、聲音）答題工具。

5. 答案儲存：負責受試者答案的彙整，按照所訂定的答案編排規格存入檔案中。

此外，從使用者介面來考量，下述各項應予妥善規劃：

1. 視窗 (windows)
2. 對話盒 (dialogue box)
3. 圖像符號 (icon)
4. 影像處理
5. 文字—圖形合併式作業系統
6. 滑鼠 (mouse)
7. 觸摸式螢幕
8. 圖形複製
9. 多工能力
10. 多媒體
11. 雙向細目表
12. 數理化符號排版系統

Hambleton (1984) 曾提出數項未來發展微電腦在測驗上功能之幾點建議：

1. 發展新的題目格式及型態，並充份利用圖形、顏色、聲音和動作等電腦的優點。
2. 調察、研究是否許多紙筆型態的題目能適當的在電腦螢幕上呈現，特別是利用在交談式的紙筆測驗。
3. 檢查受試者與機器間的交互作用，例如測驗的實施，可能受到顯示器 (monitor) 的顏色、字體大小... 等因素的影響。同樣地受試者也可能受到個人對電腦過去經驗的多寡而有所變異 (如打字的速度)，因此應提供不同的輸入方式。
4. 發展一些比較複雜的測驗練習，研究者應發展出一套較複雜之套裝程式以配合實際的需要。
5. 探討如何做到試題、軟體套裝程式以及考生受試反應之保密工作。
6. 訂定不同電腦套裝程式之效度以及如何應用在測驗之標準評定程式是否正確，這些程式是否能提供一些個別化的試題等，均為選用時所應注意事項。

綜合上述，題庫系統之各項功能需求可將其統合如下：

1. 整體功能：
  - (1). 系統自動評估的能力
  - (2). 提供標準答案的能力
  - (3). 試題屬性處理能力
  - (4). 試題難易度處理能力
  - (5). 系統文件 (操作手冊等) 完整性
  - (6). 具有網路版本，多人同時使用的能力

- (7). 題庫保密的功能
- (8). 資料壓縮，節省儲存空間的功能
- (9). 資料備份的功能

## 2. 操作環境

- (1). 視窗 (WINDOWS) 操作環境
- (2). 良好的中文環境能力
- (3). 良好的特殊符號字元處理能力
- (4). 線上施測功能
- (5). 題庫儲存功能
- (6). 系統擴充功能

## 3. 操作方式

- (1). 操作容易學習容易
- (2). 以圖像符號 (icon) 來驅動
- (3). 支援滑鼠功能

## 4. 整合方案

- (1). 系統規劃
- (2). 項目定義功能，建立試卷標準化
- (3). 題庫格式與傳統試卷格式接近
- (4). 與外界軟體 (如文書處理PE、桌上排版DTP等) 整合能力

## 5. 試題編輯

- (1). 自行定義科目、冊、章、節、題型、題級
- (2). 試題庫的編輯
- (3). 試卷庫的編輯
- (4). 包含各種題型 (是非、選擇、填充、問答、計算、證明、解釋. . .)

## 6. 命題方式與試卷生成

- (1). 依屬性配合施測目的生成試卷
- (2). 試卷自動生成功能
- (3). 自動命題題目分佈平均
- (4). 試卷人工生成功能
- (5). 多重命題功能

## 7. 文字圖形符號處理

- (1). 在螢幕繪製簡單幾何圖形
- (2). 以掃描器 (Scanner) 錄入圖形
- (3). 能處理科學性特殊符號，不需外接DTP
- (4). 提供數理化特殊圖形
- (5). 良好的圖形／影像處理能力

## 8. 試卷閱覽

- (1). 閱覽試題庫 (圖文符號同時顯示)
- (2). 閱覽試卷庫 (圖文符號同時顯示)

## 9. 排版列印

- (1). 自動列印
- (2). 列印前預試
- (3). 圖形文字符號並印

## 10. 統計分析

- (1). 試題校準的功能
- (2). 試題分析的功能
- (3). 具有解釋測驗結果的功能

表 2-3 題庫系統功能需求內涵分析

類 別	文 獻 探 討	小 組 創 見
整題功能	<ol style="list-style-type: none"> <li>2. 提供標準答案的能力</li> <li>3. 具備屬性處理能力</li> <li>4. 具備難易度處理能力</li> </ol>	<ol style="list-style-type: none"> <li>1. 系統自動評估的能力</li> <li>5. 系統文件 (操作手冊等) 完整性</li> <li>6. 具有網路版本, 多人同時使用的功能</li> <li>7. 題庫保密的功能</li> <li>8. 資料壓縮, 節省儲存空間的功能</li> <li>9. 資料備份的功能</li> </ol>
操作環境	<ol style="list-style-type: none"> <li>1. 視窗 (WINDOWS) 操作環境</li> <li>2. 良好的中文環境能力</li> <li>3. 良好的特殊符號字元處理能力</li> <li>4. 線上施測功能</li> <li>5. 題庫儲存功能</li> </ol>	<ol style="list-style-type: none"> <li>6. 系統擴充功能</li> </ol>
操作方式	<ol style="list-style-type: none"> <li>1. 操作容易學習容易</li> <li>2. 以圖像符號 (icon) 來驅動</li> <li>3. 支援滑鼠功能</li> </ol>	
整合方案	<ol style="list-style-type: none"> <li>1. 系統規劃</li> <li>2. 項目定義功能, 建立試卷標準化</li> </ol>	<ol style="list-style-type: none"> <li>3. 題庫格式與傳統試卷格式接近</li> <li>4. 與外界軟體 (如文書處理PE、桌上排版DTP等) 整合能力</li> </ol>
試題編輯	<ol style="list-style-type: none"> <li>1. 自行定義科目、冊、章、節、題型、題級</li> <li>2. 試題圖的編輯</li> <li>3. 試卷圖的編輯</li> <li>4. 包含各種題型 (是非、選擇、填充、問答、計算、證明、解釋...)</li> </ol>	
命題方式與試卷生成	<ol style="list-style-type: none"> <li>1. 依屬性配合施測目的生成試卷</li> <li>2. 試卷自動產生功能</li> <li>3. 自動命題題目分佈平均</li> <li>4. 試卷人工產生功能</li> </ol>	<ol style="list-style-type: none"> <li>5. 多重命題功能</li> </ol>
文字圖形符號處理	<ol style="list-style-type: none"> <li>1. 在螢幕繪製簡單幾何圖形</li> <li>2. 以掃描器 (scanner) 錄入圖形</li> <li>3. 能處理科學性特殊符號, 不需外接DTP</li> <li>4. 提供數理化特殊圖形</li> </ol>	<ol style="list-style-type: none"> <li>5. 良好的圖形/影像處理能力</li> </ol>
試卷閱覽		<ol style="list-style-type: none"> <li>1. 閱覽其題庫 (圖文符號同時顯示)</li> <li>2. 閱覽試卷庫 (圖文符號同時顯示)</li> </ol>
排版列印	<ol style="list-style-type: none"> <li>1. 自動列印</li> <li>2. 列印前預試</li> <li>3. 圖形文字符號並印</li> </ol>	
統計分析	<ol style="list-style-type: none"> <li>2. 試題分析的功能</li> </ol>	<ol style="list-style-type: none"> <li>1. 試題校準的功能</li> <li>3. 具有解釋測驗結果的功能</li> </ol>



何榮桂（民80）認為一個較完整的電腦化題庫系統架構，應該如下圖所示：

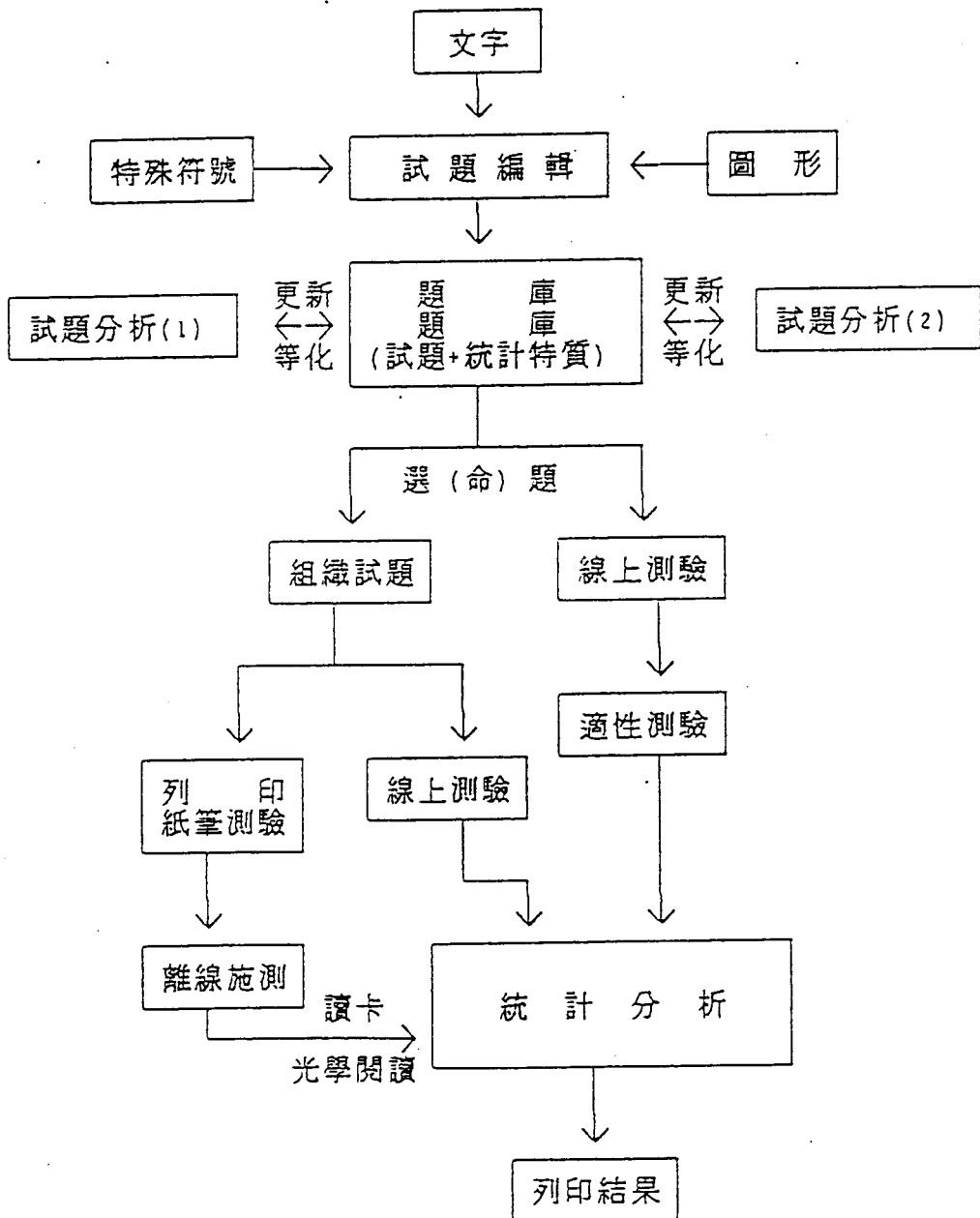


圖 2-8 電腦化題庫系統架構

針對圖2- 8，何教授並有如下的陳述：

1. 電腦化題庫系統之最基本功能即具有編輯試題的能力，以作輸入試題之用. . . 。
2. 顧名思義，題庫是儲存試題之處，但它應該是一種「動態的」(dynamic) 之設計；換句話說，它應該有試題分析的輔助設計. . . 。
3. 在題庫中有了可資使用的試題，接著就可依測驗的目的組織試題成爲一個完整的測驗，因此，在系統裡也應該考慮選(命)題的設計。
4. 因爲傳播測驗的結果之需要，一個可依簡單的統計分析的輔助設計也不可忽略. . . 。

### (三). 電腦化題庫系統之評估

張榮祖(民82)認爲：題庫系統功能的優劣，須經由測試及評量，以及檢驗其功能的好壞，始能對系統本身作全盤性的了解。而測試評量前，則須依據系統運作的範圍及目的，擬出一套完善的評量模式。其中包含評量的項目，每種項目的特性，等級的區分以及評量的方式。若能以數據方式來表達系統功能的完備與否，更能提供更科學及客觀的評鑑。

此外，他並發展如下的題庫系統評分表，以作爲題庫系統發展時或評估一套題庫系統的參考依據：

1. 「基本功能」評分表

表 2- 4 電腦化題庫系統「基本功能」評分表

特性	功 能	說 明	基本 分數	加權 分數
專業性	資料結構	A、具備「試題屬性」處理能力	10	
		B、具備「試題難易度」處理能力	10	
前瞻性	操作環境	具備MS-WINDOWS 3.0 以上彩色操作環境		10
親和性	操作方式	以圖像表示各重操作狀態，並以mouse操作		10
整合性	整合方案	以題庫系統為中心，完成所有相關之規劃以及系統規格，供使用單位抉擇		5
技 術	擴充性 試題修改	A、使用者可自行定義： 科目、冊次、章次、節次、題型、題級	10	
		B、具備對試題庫 增/插/刪/修的功能 C、具備對試卷庫 增/插/刪/修的功能	5 5	
指 標	命題方式 試卷生成	A、可依試題的屬性，配合施測目的，智慧性生成試卷		10
		B、具備隨機選題，生成試卷功能 C、具備人工選題，生成試卷功能	5 5	
文字 圖形 符號 處理	簡單圖形	具備在螢幕繪製： 直線、正圓、橢圓、弧線、方形、雙曲線 等幾何圖形功能		
	複雜圖形	具備以256灰階 Handy scanner 錄入圖形的功能		5
	資料庫	系統自備資料庫，不須外接DBP即能處理科學特殊符號		10
	圖形庫	其中儲存數理化特殊圖形，可供繪製或剪貼此庫中圖形，應用於試題中		10
	數理化符號之處理	A、描述語言 B、造字	5 5	
題庫編輯	題庫編輯	A、編輯時可同時輸入：圖形、文字、公式、符號	10	
		B、編輯時可看到已編輯結果	5	

## 2. 「完備功能」評分表

表 2- 5 電腦化題庫系統「完備功能」評分表

特性	功 能	說 明	基本 分數	加權 分數
專業性 前瞻性	閱覽 試卷閱覽	A、具備閱覽試題庫的功能（圖文符號同時顯示） B、具備閱覽試卷庫的功能（圖文符號同時顯示）	5 5	
	排版 列印	A、具備閱覽試題庫的功能，並可作修改 B、自動排版 C、圖形、文字、符號並行列印 D、印表機不限機種、廠牌、型號	10	10 10 10
標準化 題庫	具備： 屬性、 難易度、 答案、 解題過程	單科 試題量 A、 $\geq 10,000$ B、 $\geq 20,000$		5 10
效應 指標	擴充性 試題修改	價 格 A、\$10,000.00 以下 B、\$20,000.00 以下 C、\$30,000.00 以下 D、\$50,000.00 以上		1 2 3 5
評 分 值	評分值： 得分/ 效應檢損指標			

### (四). 建立電腦化題庫系統之問題與未來發展方向

洪碧霞與吳鐵雄(民78)亦認為電腦化測驗及電腦題庫的建立與實施將為提昇評量的效力提供另一展新的里程碑。然而，如何將教學原理和命題技巧，融合資訊科學技術中的電腦輔助教學、資料庫管理技術、中文系統處理技術、計算機圖學與人工智慧的專家系統，以提昇教學題庫的命題與考驗效率，幫助教師分析診斷與補救教學，則是一項極待教育界努力的方向。

江文雄等(民82)認為電腦化題庫建立過程中經常遭遇下述三

方面的問題，江教授等也相對的提出了妥善的對策：

1. 題庫需大量特殊符號

【對策】

- (1). 盡量使用標準中文內碼的特殊符號
- (2). 使用圖形表示
- (3). 採用中文系統之造字檔

2. 未來題庫之整合與維護問題

【對策】

- (1). 採用開放式架構及標準文字檔
- (2). 規劃檔案編碼架構

3. 題庫之圖形來源問題

【對策】

- (1). 採用單圖形檔與多圖形檔並用並進方式
- (2). 使用外部繪圖工具最為支援

經由技術與經驗的成長，未來除了在測驗方式應朝向電腦化適性測驗的理論與技術發展之外，在電腦技術的成熟與進步配合之下，電腦化題庫的發展將沿下述四個階段漸趨發展：

1. 傳統題庫之發展

2. 電腦化測驗題庫

3. 多媒體題庫

4. 智慧型題庫

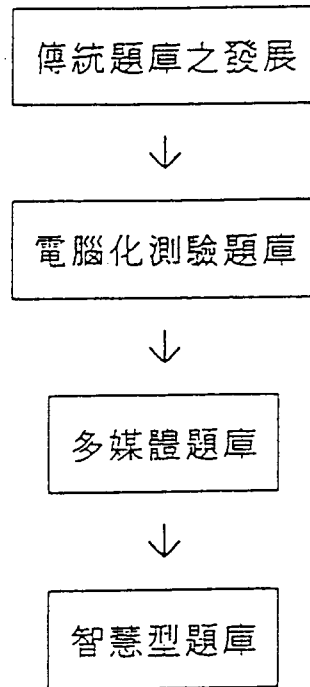


圖 2-9 電腦化題庫技術發展策略圖

### 第三節 題庫的試題分析

#### 一、題庫試題分析概說

題庫乃是包含許多為某些目標、技能或工作 (task) 而編製的題目，以為測驗編製者為某些需求 (目的) 編製測驗時的基礎。換句話說它是許多題目的匯集。當一個題庫包含有具備「內容效度」而且技術品質優良的試題時，測驗的編製就容易多了，其所產生的測驗通常比沒有題庫的情況下的品質要高 (洪碧霞，民82)。

建立題庫最重要的一件工作就是如何來選擇高水準的題目。有了高水準的題目，有關教師命題時的困難就可迎刃而解，Hambleton(1982) 及 Popham(1981) 曾分別提及判斷法 (judgement methods) 及經驗法 (empirical methods) 來進行題目特徵有關「量」的統計分析，並藉以選擇題目。所謂判斷法乃是聘請學科專家們舉行專家會議 (expert conference)，藉以評定所欲選擇的題目品質，並且判定是否與所定的目標之間和諧。經驗法是採用檢查受試者對題目反應的情形，以做為修正題目的參考。

測驗編製有一個共同目標那就是希望以最少的題數 (即最經濟)，達到最大的功能 (即最有效)。吳裕益 (民81) 認為要達到以上的目標，通常是對一個大題庫 (large pool of items) 進行實地測試，然後從題庫中選取對信度和效度貢獻最大的一組題目來組成一份測驗。在編製一個新測驗 (或縮短原有的測驗)，最後所要採用的一組題目通常要經過「試題分析」 (item analysis) 過程來選取。

對測驗工具的評鑑 (evaluation of instrument)，通常也稱為試題評鑑。王昭明 (民75) 認為試題評鑑的目的不外乎下述五點：

1. 改進測量工具 (improve the instrument) 。
2. 增強學生學習 (reinforce learning) 。
3. 作為補救教學 (remedial instruction) 的基礎。
4. 改進教學 (improve instruction) 。
5. 增加試題編製的技巧 (increase skill in test construction) 。

Erickson 和 Wentling (1976) 也認為試題評鑑應包含下述兩個：

1. 對工具個別部份的評鑑。也就是對測驗工具進行試題分析 (item analysis) 。
2. 對工具做整體性的評鑑。也就是考慮工具的效度與信度。

優良的測驗必須具備相當水準的成的信度 (reliability) 與效度 (validity)，而測驗的信度與效度又取決於全部試題 (test items) 的性能。對於試題性能的檢驗，則有賴於邏輯的與統計的分析 (logical and statistical analysis) (簡茂發，民76) 。

因此，對於擬於收集進題庫中各個試題，事先予以進行試題分析的工作是相當重要的。

所謂試題分析就是對每一個測驗試題再做一次檢驗 (Reexamining) 發現試題的優點和缺點 (Strengths and Flaws) 即稱為試題分析。所以，試題分析不但可以使我確定那些題目是不好的 (Poor Items)，同時也讓我們知道這些題目之所以不能達到預期目標的原因。是故，試題分析使我們變成更好的試題編製者 (Chase, 1978) 。

Martuza (1977) 認為試題分析的目的有下述兩點：

1. 評估測驗試題的品質。
2. 發覺學生對教學內容不了解與誤解之處。



Erickson & Wentling (1976) 則認為試題分析的目的有下述四點：

1. 選出已被認為適切的題目。
2. 檢查試題之結構有何缺點。
3. 協助教學者設計個別化之補助教學。
4. 增進教學的效果。

王昭明 (民75) 也提出試題分析的目的有如下諸端：

1. 鑑定試題的結構。
2. 改進測驗試題的品質。
3. 發現學生學習的障礙。
4. 作為設計補救教學的依據。
5. 作為改進教師教學的依據。

試題分析可以分為質的分析 (qualitative analysis) 與量的分析 (quantitative analysis) 兩部份 (簡茂發, 民76)。前者就試題的內容和形式, 從取材的適切性 (relevance) 與編製試題的技術加以評鑑, 就是對試題質的方面做邏輯的分析; 後者則基於試題經過選題預試 (try-out) 的結果, 對試題特徵做量的統計分析, 這可從心理測驗理論中兩個不同學派對試題分析個別作法來加以說明。

心理測驗理論 (或簡稱為「測驗理論 (test theory)」) 是一種解釋測驗資料間實證關係 (empirical relationships) 的系統化理論學說。測驗理論一般劃分為兩大學派, 亦即：

1. 傳統測驗理論 (classical test theory, 簡稱CTT)

傳統測驗理論主要內涵是「真實分數模式」 (true score model) (Gullikson, 1987; Lord & Norick, 1968), 亦即, 觀察分數  $X$  等於真實分數  $T$  與誤差分數  $E$  之和, 其數學表示式為

：  $X = T + E$  。這個簡單而淺顯的弱勢假設 (weak assumption) 便是傳統測驗理論的理論基礎模式和發展骨幹。

## 2. 當代測驗理論 (modern test theory, 簡稱MTT)

當代測驗理論主要是以「試題反應理論」 (item response theory, 簡稱IRT) (Lord, 1980; Hulin, Drasgow, & arsons, 1983; Hambleton, Swaminathan, & Rogers, 1991) 為理論架構。事實上, IRT乃依據強勢假設 (strong assumption) 而來, 也就是可使用一個「數學模式」, 來表示可觀察到的受試者行為 (test performance) 與其不可觀察到的潛在特質間之關係。

## 二、傳統測驗理論試題分析

### (一) 傳統測驗理論試題分析概述

傳統測驗理之試題分析除了前述所說有關試題「質」方面邏輯性的分析之外, 而所謂「量」方面的分析乃是對試題逐一分析其難度 (item difficulty)、鑑別度與受試者對各項配列答案 (options) 的反應情形, 以作為修改試題或選擇試題的依據。

而有關個別試題之題目特徵參數 (item parameters) 的檢驗通常分為下述三類 (Crocker & Algina, 1986; 吳裕益, 民81) :

1. 描述受試於單一題目之反應的分佈情形之指數 (如, 題目反應的平均數和變異數)。
2. 描述受試於單一題目之反應和某些感興趣的效標之關聯程度的指數 (如, 題目與測驗總分之相關)。

數（如，題目與測驗總分之相關）。

3. 與「題目變異數」以及「題目與效標之關係」二者有函數關係的指數（如，題目信度和效度指數）。

以傳統測驗理論為基礎所建立之題庫有其難以克服的缺點，亦即當題庫試題進行更新（update）時題庫將產生很大的限制。因為以團體為基礎的統計量數，會因不同的測驗或不同的預試樣本而使新舊題目間產生非線性關係，而導致新舊題目間之量尺的特質（scale property）不一致；也就是說，當題庫要進行更新時，不易將不同試題的試題組合之 P 值或 D 值重新量尺化（rescale），因此，每次建立的題庫只是用於單一的測驗（何榮桂，民80）。

洪碧霞（民81）也認為傳統的題目參數像通過百分比（P）及題目與總分相關（ $r_{p.b.}$ ）對題庫中試題的描述並不適切，因為這些參數是樣本依賴，換句話說，依考生特質的不同，參數就會跟著浮動。因此，對甲校很難的題目，對乙校可能並不難，如此，題庫中題庫參數的參考價值就會大打折扣，而題庫的功能也就會跟著劇降。

## (二). 試題分析的方法

常模參照測驗與標準參照測驗在目的上有很大的不同，題目分析可「質」的分析與「量」的分析，質的分析乃是分析試題的內容與形式，包括“內容效度”（content validity）的分析以及編擬試題技術之評鑑。量的分析是採用統計方法來分析題目的品質，其項目包括難度（difficulty）與鑑別度（discrimination）分析。一份測驗的信度與效度高低與否，則完全取決於個別試題之品質，所以經由題目分析的程序將可提高測驗的信度與效度。

### 1. 難度分析

(1).以通過百分比表示難度

試題的難度與測驗的效率 (effectiveness) 有關，難度適當的試題是構成優良測驗的必要條件 (簡茂發，民75)。試題的難易程度通常以全體受試者答對或通過該題百分比 (percentage passing) 來表示，此種方法用於當一個題目是採用二元計分法時。其公式如下：

$$P = \frac{R}{N} \times 100\%$$

P：試題難度

R：答對該題的人數

N：全體受試者人數

P值介於 .00~1.00 之間

另一種求難度的方法是以高分組 (測驗總分最高之27%) 答對百分比加低分組 (測驗總分最低之27%) 答對百分比再除以二。當我們欲求得鑑別度時也需計算出高、低分組答對的百分比，所以很多測驗編製者較偏此種表示方法：

$$P = \frac{P_H + P_L}{2}$$

P : 試題難度

P<sub>H</sub> : 高分組答對該題的百分比

P<sub>L</sub> : 低分組答對該題的百分比

P 值介於 .00~1.00 之間

當試題難度 P 值越接近 .50 時，題目所具備的區別能力越高，越是接近 1.00 或 .00，就越無法區分受試者之間能力的差異。

在二元計分法當中，題目難度 (P) 與題目變異數之關係如下

:

$$S_i \text{ (或 } \sigma_i) = \sqrt{p_i \times q_i}$$

$\sigma_i$

S<sub>i</sub> : 某個題目之變異數

p<sub>i</sub> : 答對某題之百分比

q<sub>i</sub> : 答錯某題之百分比

當題目彼此之間之相關係數是一個常數，當  $p_i = .50$  時，總測驗分數變異數為最大。

## (2). 等距量表的難度指數

以通過百分比來表示難度，其試題難度  $P$  值是一種順序量尺 (ordinal scale) 差距單位並不相等，它只能表示試題難易度的相對位置，並無法指出各試題之間難易程度的大小。

針對此一缺點，美國教育測驗服務社 (Education Testing Service) 另創一種具有等距尺度 (interval scale) 特性的難度指數，以  $\Delta$  (delta) 表示。它是一種以 13 為平均數、4 為標準差、下限為 1、上限為 25 的標準分數。 $\Delta$  值愈小，難度愈低； $\Delta$  值愈大，難度愈高。

此外它還有幾樣特點：

- ①. 可表示試題難易的相對位置。
- ②. 可指出不同難易之間的差異數值。
- ③. 假設所測驗而得的題目難易數值成常態分配。
- ④. 試題的難度可在常態分配曲線的橫軸上某一點以離差分數 (deviation score) 表示之。

### § 求法 §

根據答對某一試題的人數百分比與答錯該題的人數 (包括未作答者) 百分比，使前者在右，後者在左，找出兩者在常態分配曲線橫軸上的分界點，此點的相對位置以標準差為單位表示之，既為  $x$ ，其求法如下式：

$$\Delta = 13 + 4x$$

至於在實際的應用上，試題的 $\Delta$ 值可由范氏項目分析表（Fan, 1952）查得。當 $P > .50$ 時， $\Delta$ 值小於13；當 $P < .50$ 時， $\Delta$ 值大於13，所有試題的 $\Delta$ 值介於1~25之間。

## 2. 鑑別度分析

許多測驗的目的是在提供有關受試者在該測驗所欲測量之構念或是該測驗分數所欲預測之外在效標的個別差異之訊息但有時候所測量的構念沒有辦法找到比該測驗總分更恰當的量數。（如課堂上的成就測驗，通常不容易找到適當的外在效標。）吳裕益（民81）認為在此種情況下，就以測驗總分作為受試者在所測量之構念相對地位之操作性定義（operational definition）。使用此種內在效標主要目的在於選擇那些得高分者其答對之機率也較高，得低分者其答對之機率也較低之試題。試題的鑑別力（discriminating power）之大小與測驗的信度與效度都有密切的關係。所以，如想增進測驗之預測與診斷的功能，都必須著重試題鑑別度的分析。試題的鑑別度可分為題目效度（item validity）與內部一致性（internal consistency）分析兩方面。前者在分析受試者在題目上的反應與效標上之表現的情形，後者是在分析個別試題與整個測驗總分的一致性。

### (1). 題目效度分析

#### ①. 題目特徵曲線法

題目特徵曲線 (item-characteristic curves) 可用來表示每個題目之效度。將受試者實際作答資料，繪製於一個以效標分數為橫座標、以通過百分比為縱座標的座標系上的試題特徵曲線。我們可直接由座標系上看出特別突出之題目表示效度越高，越平坦者效度越低。

## ②. 相關係數法

以題目特徵曲線表示效度可以看出效度的高低，但要應用於電腦化題庫上，則最好採用單一數字指數來表示題目的效度。事實上用來求取題目效度指數的方法至少有五十種以上（陳英豪、吳裕益，民71）。以下便是幾種較常表示的方法：

### a. 點二系列相關 (point-biserial correlation)

以點二系列相關來表示題目之效度，其最適用於效標為連續分數，而題目為二分變數。其計算式如下：



$$r_{pb} = \frac{\bar{X}_p - \bar{X}_q}{S_t} \sqrt{pq}$$

$r_{pb}$  : 某個題目之變異數

$\bar{X}_p$  : 答對者之受試者在效標上的平均得分

$\bar{X}_q$  : 答錯者之受試者在效標上的平均得分

$S_t$  : 全部受試者在效標得分的標準差

$p$  : 答對某題之百分比

$q$  : 答錯某題之百分比

b. 二系列相關 (biserial correlation)

此種效度的表示方法假定受試者在試題上的反應為常態分配，而且題目只有答對及答錯兩種情況。其計算式如下：

$$r_{pbis} = \left( \frac{\bar{X}_p - \bar{X}_q}{S_t} \right) \left( \frac{p - q}{y} \right)$$

$r_{pb}$  : 某個題目之變異數

$\bar{X}_p$  : 答對者之受試者在效標上的平均得分

$\bar{X}_q$  : 答錯者之受試者在效標上的平均得分

$S_t$  : 全部受試者在效標得分的標準差

$p$  : 答對某題之百分比

$q$  : 答錯某題之百分比

$y$  : 常態分配下答對百分比 ( $P$ ) 所在位置之曲線的高度

c.  $\phi$  係數 (phi coefficient)

以  $\phi$  係數表示相關用於題目與效標均為二分變數的情況。計算  $\phi$  係數前先將資料化為如下自由度為 1,  $2 \times 2$  的列聯表。

	答對	答錯	
合格	A	B	(A + B)
不合格	C	D	(C + D)
	(A + C)	(B + D)	

其次將上述列聯表之資料帶入下式計算  $\phi$  係數：

$$\phi = \frac{BC - AD}{\sqrt{(A+B)(C+D)(A+C)(B+D)}}$$

## ②. 鑑別度指數

此種鑑別度指數是目前採用最普遍的題目分析法。一般以 D 來表示鑑別度指數 D 的值在 -1.00 ~ +1.00 之間。其計算式如下：

$$D = P_H - P_L$$

D : 鑑別度指數

$P_H$  : 高分組通過人數百分比

$P_L$  : 低分組通過人數百分比

D值越高，表示該題的鑑別度越高。D的平均指數越高，表示測驗的信度越高。

表2-6 兩變項之性質與適用之相關係數

Y	X	二分類別變項	基於常態分配的二分變項	次序變項	等距或比例變項
二分類別變項	$\phi$				
基於常態分配的二分變項	$\phi$		$r_{tet}$		
次序變項	$r_{rb}$		$r_{rb}$	$r_s, r_w$	
等距或比例變項	$r_{pb}$		$r_{bis}$	$r_s, r_w$	$r_{xy}$

(資料來源：簡茂發，民76，頁194)

- $\phi$  :  $\phi$  係數 (phi coefficient)
- $r_{tet}$  : 四分相關係數 (tetrachoric correlation coefficient)
- $r_{rb}$  : 等級二列相關係數 (rank-biserial correlation coefficient)
- $r_{pb}$  : 點二相關係數 (point-biserial correlation coefficient)
- $r_{bis}$  : 二系列相關係數 (biserial correlation coefficient)
- $r_s$  : 斯氏等級相關係列 (Spearman rank-biserial correlation coefficient)
- $\tau$  : 斯肯  $\tau$  係數 (Kendall's Tau)
- $w$  : 斯肯和諧度係數 (Kendall's coefficient of concordance)
- $r_{xy}$  : 皮氏積差相關係數 (Pearson product-moment correlation coefficient)

## (2). 內部一致性分析

內部一致性分析旨在瞭解各個試題的功能是否和整個測驗的功能相符一致。一般在進行題目分析時，經常以測驗的總分作為分析的依據。此種分析可能改善內容效度和建構效度，但無法增進效標關連效度。

內部一致性分析的方法，與題目效度分析的方法很類似，主要是以測驗總分來取代外在效標。其主要方法有下列兩種：

### ①. 求試題反應與測驗總分之關聯性

每一為受試者之測驗總分屬於連續變數，而受試者的作答情形可分為答對與答錯兩種，為二元變數，故可用前述的點二系列相關 ( $r_{pb}$ ) 或二系列相關法 ( $r_{bis}$ ) 來求得其高低，而二系列相關法可由范氏題目分析表查得。

### ②. 比較高低分組通過每一試題的百分比

$$D = P_H - P_L$$

D值越大，表示作答反應與測驗總分的一致性越高。

### 三、試題反應理論試題分析

#### (一)、試題反應理論試題分析概述

當我們擬予實施電腦化適性測驗 (computerized adaptive testing) 時，是以試題反應理論 (item response theory; IRT) 為基礎來建立題庫：亦即在建立題庫之前，須選擇適當之試題反應模式 (item response model) 將試題加以校準 (calibration)，然後將校準過 (calibrated) 的試題參數 (item parameters) 挑選品質較佳之題目，以建立校準化的題庫 (calibrated item bank) (何榮桂，民80)。

試題反應理論 IRT 發跡於 40 年代中葉，Ledyard Tucker (1946) 就已經已提及「題目特徵曲線」 (item characteristic curve, 簡稱 ICC) 這個名詞與觀念，其實早在 1916 年 Binet 及 Simon 編製智力量表時，就已經使用了 IRT 的核心觀念—題目特徵曲線，但一直到 60 年代末，測驗領域仍以古典測驗理論中真實分數模式為主 (亦即，觀察分數等於真實分數與誤差分數之和，數學公式為： $X = A + E$ )。

表2-7 對IRT發展有實際貢獻的代表學者與研究成果(余民寧, 民81)

作者(年代)	代表作及其貢獻	作者(年代)	代表作及其貢獻
Tucker (1946)	第一位提出試題特徵曲線概念的人。	Wright & Masters (1982)	簡述 Rasch 模式的各種模式成員, 證明與部份計分模式相通, 對 Likert 式評定量表與次序反應資料的計分方式改進不少。
Lord (1952)	第一位尋出兩個參數常態肩形模式的參數估計公式, 並考慮試題反應理論應用性的人。	Mislevy & Beek (1982)	發表另一有名的電腦程式: BILOG。
Rasch (1960)	試題反應理論中 Rasch 模式的創始者, 脫俗深遠。	1982年	Applied Psychological Measurement 第四季出版一冊專刊試題反應理論及其應用的專題。
Lord & Novick (1968)	第一本介紹古典與當代測驗理論模式的經典作品, 引發學者對「潛在特質」概念的重視與研究。	Wainer & Messick (1983)	編組成的論文集, 以期表揚 Lord 一生對試題反應理論的貢獻, 並兼論該理論的應用與未來。
Wright & Paschopoulos (1969)	美國地區第一本介紹 Rasch 模式的參數估計法, 並發展有名的 BICAL 電腦程式代表作品。	Weiss (1983)	編撰一本專談試題反應理論的應用與未來, 並介紹它在電腦適性化測驗上應用的論文集。
Samejima (1969)	她的一系列作品描述新的試題反應模式及其應用, 其中還包含處理多分法知能屬性資料的模式, 甚至擴展到多向度的試題反應模式, 為一規模龐大的重要著作。	Hulin, Orasgov, & Parsons (1983)	為一本試題反應理論的教科書, 增加對「適合度測量」概念的說明與應用。
Beck (1972)	提供許多估計模式參數的新概念。	Hambleton (1983)	編撰一本試題反應理論模式與應用的論文。
Anderson (1973)	歐洲地區談論量型模式的重要著作。	Embretson (1985)	編撰一本試題反應理論的未來發展的論文集。
1976年	Lord 等創作第一級有名的電腦程式: LOGIST。	Baker (1985)	為一本專論性的試題反應理論教科書, 專為沒有教學訓練基礎的讀者而作, 並附有 CAS 電腦檔教學影片。
1977年	Journal of Education Measurement 第四季出版一冊專刊試題反應理論的專題。	Hambleton & Swaminathan (1985)	為一本即時的試題反應理論教科書。
Baker (1977)	第一篇評論試題反應模式參數估計法的文獻探討。	Crocker & Algina (1986)	比較古典與當代測驗理論的專論性教科書。
Wright & Stone (1979)	第一本簡述各種 Rasch 模式理論及其應用的專著。	Wainer & Braun (1988)	專談有賦力度方面的論文集, 也討論試題反應理論在效度上的應用。
Lord (1980)	第一本以試題反應理論為命名的專著, 是當代測驗理論發展的里程碑。	Linn (1989)	負責主編第三版「教育評量」(Educational Measurement) 其中一章專門介紹並評論試題反應理論。
Weiss (1980)	第一本討論應用的論文集, 專談試題反應理論的實際應用或題一一電腦適性化測驗。	Freedle (1990)	人工智慧及其在當代測驗理論上應用論文集。
Anderson (1980)	對測量模式參數估計法有貢獻的方法學專論。	Suen (1990)	介紹各種測驗理論方面的教科書。
Beck & Aitchison (1981)	提出連續的最大似然估計法——EM 估計程序, 對參數估計法有貢獻很大。	Wainer 等人 (1990)	專談電腦化適性測驗方面的人門書, 也談試題反應理論在電腦化適性測驗上的應用。
Masters (1982)	第一位發表部份知能計分模式, 對改良 Likert 式評定量表的計分與次序反應資料的計分貢獻	Hambleton, Swaminathan, & Rogers (1991)	試題反應理論方面的人門書, 適於非教學主修的初學者閱讀。

## 1. 試題反應理論的意義

所謂試題反應理論 I R T 就是以一數學模式，來表示可觀察到的受試者行爲 (test performance) 與其不可觀察到的潛在特質間之關係，也就是一種數學函數的形式 (Hambleton & Cook, 1977)。Hambleton (1985) 也認為：試題反應理論是用來解釋測驗題目、受試者反應及個人特質等三者之間相互關係的一種理論架構 (請參閱下圖)。倘依據上述所言，我們可以說：所謂試題反應理論就是以一個預先設定的數理統計學機率模式，將受試者看不見的潛力與他自己作答時的實際得分情形聯結在一起，當得分累積至一定程度時，受試者看不見的潛力便可藉由統計的方法推算出來。所以，試題反應理論的基礎實際上是奠定在預先設定的數學函數上，而函數的表示方法隨著實際應用情形而有調整，我們無法決定函數表示法是否會正確，但可依核對函數決定的得分情形和受試者實際的得分情形來看結果是否相符。

試題反應理論可以分為廣義和狹義的方式 (許擇基、劉長宣，民80)；廣義的 I R T 中，受試者的測驗成績是由一些看不見的能力特質來決定。考生的能力特質有高低之分，所以我們可以用數值來表達不同受試者在能力特質上的相對程度。通常受試者在第  $i$  個特質上的程度以  $\theta_i$  來表示，所以一個測驗若測量  $k$  種能力特質，受試者在  $k$  度空間中的位置是  $(\theta_1, \theta_2, \theta_3, \dots, \theta_k)$ ，函數和對應的假設加起來則組成爲一個 I R T 中的模式 (model)。改變函數表示法或假設，則產生新的反應模式，所以在廣義的 I R T 中，有無數的試題反應模式。但在



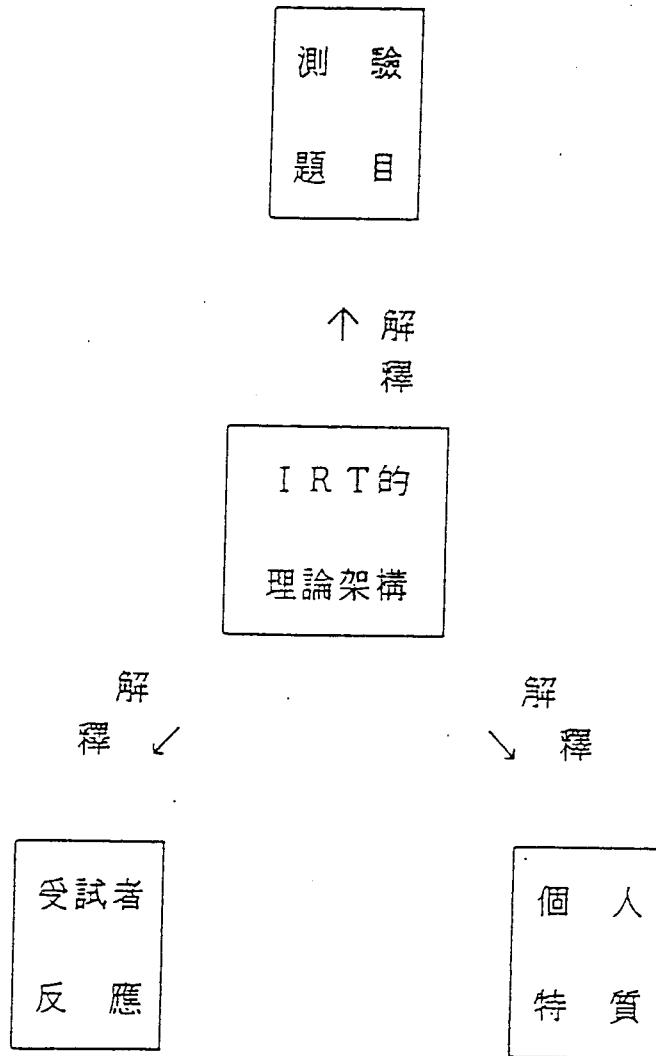


圖2-10 I R T可用來解釋測驗題目、受試者反應與個人特質的理論架構

---

實際應用時，I R T通常指的是狹義的定義；在狹義的I R T中

試題測量的能力空間被縮小到一度。換句話說，將測量範圍，限制在一種能力或特質上，同時函數的表示法，採用的是logisticfamily（數理邏輯家族）中的成員。基本上這些成員有一大家子人，但在狹義的IRT中，經常被用來討論的函數表示法則是：(1).單參數模式（one-parameter model）；(2).雙參數模式（two-parameter model）；(3).三參數模式（three-parameter model）。

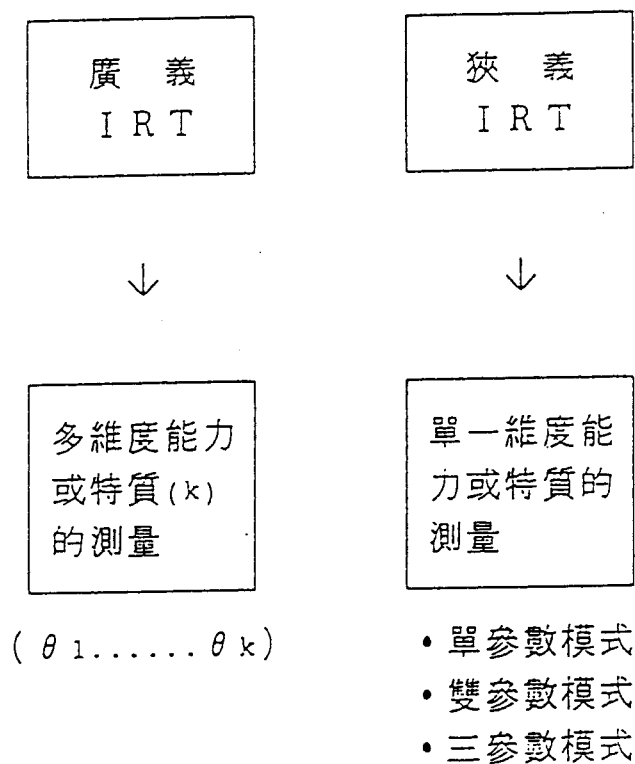


圖2-11 廣義與狹義 I R T

---

## 2. 試題反應理論的基本假設

前面我們曾經提及 I R T 乃一個依據強勢假設 (strong assumptions) 而來，立論與假設均合理且嚴謹的學說。I R T 是利用函數來模擬受試者答題的心路歷程，而模擬的函數之實際應用是否能屬有效，則要視理論的假設與受試者實際的答題情形是否相符而定。有關 I R T 的理論假設有下述諸端：

- (1) 「維度」 (dimensionality) 必須集中。
- (2) 各別試題「局部獨立」 (local independence) 。
- (3) 某個「題目特徵曲線」就是該題答對或然率對受試者能力的迴歸線。
- (4) 「答題速度」 (speedness) 沒有時間限制。

## 3. 試題反應理論的重要概念

基於 I R T 應用的特質，Hulin (1983) 曾提出了五項重要概念：(黃國彥、王以仁，民76)

- (1). 在 I R T 中，許多個人之間不同的特性 (characteristics) 均被加以考慮。在早期如 Lord (1952) 建立 I R T 模式時，所重視的個人特性為能力 (ability)，如語文能力、數學能力等；Parsons (1979) 則曾經將 I R T 應用於職業滿意度之測量的研究。
- (2). 在 I R T 的研究中有一共同的現象：假如所觀察到的行為或測驗反應，係映射出個人的某一潛在特性 (latent

characteristic)。例如，某人在學業性向測驗語文部份 (scholastic aptitude test-verbal section, SAT-v) 中所獲得的分數，就代表了他的某種潛在特性的語文能力。由於潛在特質是無法直接觀察的，故利用 I R T 將個人反應與潛在變項兩者加以連結；亦即以可觀察到的行為反應（通常是對許多測驗項目的回答），來預測其個人的潛在特性。

- (3). 在 I R T 中，是以題目 (item) 來代表觀察的單位，以測驗 (test) 或量表 (scale) 來表示許多題目的集合，並以特質或能力來表示個人的潛在特性。同時採用希臘字母  $\theta$  (theta) 來表示「潛在特質」的數值。
- (4). 潛在特質中  $\theta$  被視為呈一連續數值的分配，但並不需要假設它呈常態分配。事實上，某些母羣體特性的真實分配中，也並非完全是常態的。
- (5). 不同測量理論，對於界定特質與反應行為及變項間關係的內容上，自是有所差異。因此，I R T 僅適用於對個人潛在特質能力  $\theta$  之推估，而不需要對此特質提出建構效度。

余民寧 (民82) 則認為 I R T 建立在兩個基本概念上：

- (1). 受試者 (examinee) 在某一個試題上的表現情形，可由一組因素來加以預測或解釋，這組因素叫作「潛在特質 (latent traits) 或能力。

(2). 受試者的表現情形與這組潛在特質之間的關係，可透過一條連續性遞增的函數來加以詮釋，這個函數便叫作「題目特徵曲線」(item characteristic curve, 簡寫為 ICC)。

## (二). 三參數題目參數估計

試題反應理論包含著一組數學模式 (a family of mathematical models) (Hambleton & Cook, 1977; Weiss, 1983)。Hambleton & Cook (1977) 指出：試題反應理論就是以一數學模式，來表示可觀察到的受試者測驗行為 (test performance) 與其不可觀察的潛在特質間之關係，亦即是一種數學函數的模式。所以試題反應理論也常為試題反應模式 (item response models, 簡稱 I R M) 所代替。

所以在廣義的 I R T 中，有無數的試題反應模式。但在實際應用時，I R T 通常指的是狹義的定義；在狹義的 I R T 中試題測量的能力空間被縮小到一度。換句話說，將測量範圍，限制在一種能力或特質上，同時函數的表示法，採用的是 logistic family 中的成員。基本上這些成員有很多，但在狹義的 I R T 中，經常被用來討論的函數表示法則是：(1). 單參數模式 (one-parameter model)；(2). 雙參數模式 (two-parameter model)；(3). 三參數模式 (three-parameter model)。

三參數模式之計算式如下：

$$P_i(\theta) = C_i + \frac{1 - C_i}{1 + \exp\{-D a_i(\theta - b_i)\}}$$

$P_i(\theta)$  : 能力值為  $\theta$  之考生答對試題的或然率

$\theta$  : 考生之能力值

$D a_i$  : 為一常數，三參數模式一般設為 1.7。

$a_i$  : 鑑別力

$b_i$  : 難度

$c_i$  : 猜測度

### (三). 測驗訊息

試題反應理論以試題及測驗訊息 (information) 取代傳統測驗理論的內部一致性信度 (如  $\alpha$  係數)。單一試題訊息是能力值 ( $\theta$ ) 的函數，它會隨著能力之不同而變化。而測驗訊息則是所有試題訊息之總和，它當然也隨著能力而變

化。表示訊息量時以能力值（-3~+3）為橫軸，以訊息量為縱軸。當能力值為 0 時之訊息量最高，而在+2及-2以外較低。

### 1. 試題訊息的計算方法

$$I_i(\theta) = \frac{D^2 \times a_i^2 \times Q_i(\theta) (P_i(\theta) - c_i)^2}{P_i(\theta) \times (1 - c_i)^2}$$

$I_i(\theta)$  : 對  $i$  個試題的訊息

$Q_i(\theta)$  : 能力值為  $\theta$  之考生答錯試題的或然率

$P_i(\theta)$  : 能力值為  $\theta$  之考生答對試題的或然率

$\theta$  : 考生之能力值

$D_{ai}$  : 為一常數，三參數模式一般設為 1.7。

$a_i$  : 鑑別力

$b_i$  : 難度

$c_i$  : 猜測度

## 2. 測驗訊息的計算方法

測驗訊息等於該測驗所包括之所有各別試題訊息之總和。

$$\begin{aligned} I(\theta) &= I_1(\theta) + I_2(\theta) + \dots + I_n(\theta) \\ &= \sum_{i=1}^n \frac{P_i(\theta)}{P_i(\theta) \times Q_i(\theta)} \end{aligned}$$

### (四). 單元化分析

由題庫組成之測驗應該符合試題反應理論有關單元化 (unidimensionality) 的假設，亦即每次只測量受試者的一種能力。

有關單元化假設的檢視有下述三種：



### 1. 因素分析法

常用的方法為四分相關矩陣而不是 $\phi$ 係數，它和二系列相關係數  $r_{bis}$  有類似的意義。

### 2. 試題相關矩陣特徵法

試題之間相關矩陣的特徵值 (eigenvalue) 也可用來檢視單元化的假設，如果矩陣的特徵值中有一個特別大，則試題多只測一種能力。

### 3. 雙難度比照法

將試題分成幾個部份，並將每一部份試題的難度估計出來，最後再將幾個部份的測驗混合起來，混合起來的試題在作一次難度估計，因此每個測驗有兩個難度估計值。將兩組難度值相互比較其差異，我們就可決定各分測驗是否符合單元化的原則。

## (五). 適合度考驗

三參數模式的試題分析常採用  $\chi^2$  指數來考驗模式的適合度，並藉以考驗測驗中之試題受否有嚴重偏離三參數 logistic 模式。

#### 四、傳統測驗理論與試題反應理論之比較

洪碧霞(民81)也認為傳統的題目參數像通過百分比(P)及題目與總分相關( $r_{p.b.}$ )對題庫中試題的描述並不適切，因為這些參數是樣本依賴，換句話說，依考生特質的不同，參數就會跟著浮動。因此，對甲校很難的題目，對乙校可能並不難，如此，題庫中題庫參數的參考價值就會大打折扣，而題庫的功能也就會跟著劇降。

當我們擬予實施電腦化適性測驗(computerized adaptive testing)時，是以試題反應理論(item response theory; IRT)為基礎來建立題庫：亦即在建立題庫之前，須選擇適當之試題反應模式(item response model)將試題加以校準(calibration)，然後將校準過(calibrated)的試題參數(item parameters)挑選品質較佳之題目，以建立準化的題庫(calibrated item bank)(何榮桂，民80)。

##### (一). 傳統測驗理論與試題反應理論之異同

綜合而言；傳統測驗理論 C T T 與試題反應理論 I R T 有下述幾種同：

##### 1. 受試者能力之表示法不同

- (1). C T T—以觀察值  $X$  表示學生能力高低。
- (2). I R T—以能力值  $\theta$  (+3 ~ -3) 表示學生能力高低。

## 2. 理論基礎不同

- (1). C T T—建立在簡單的函數假設 ( $X = A + E$ ) 上。
- (2). I R T—奠立在一預先設定的數學函數模式上，將受試者看不見的潛力和他作答得實際情形聯結起來，當得分夠多時，看不見的潛力就可用統計方法推算出來。

## 3. 參數值穩定性不同：

- (1). C T T—樣本依賴。試題參數會依考生特質的不同而不同。
- (2). I R T—樣本獨立。樣本不同的參數也可經由線性轉化而成爲樣本獨立的參數值。

## 4. 主要功能不同：

- (1). C T T—
  - a. 可提供測驗編製與評鑑的客觀標準及分析方法
  - b. 協助正確的解釋和運用考試獲得的分數
  - c. 協助發掘及解決現有的測驗應用問題
- (2). I R T—
  - a. 測驗的編製
  - b. 建立題庫
  - c. 適性測驗
  - d. 選項的不同加權
  - e. 試題偏向
  - f. 分數的等化

(二). 試題反應理論 I R T 尚未普遍使用的原因

國內外研究 I R T 的學者已逐漸增加，但由於某些因素，因而大都停留在理論的探討階段，能將 I R T 實際應用於真實測驗情境者目前非常少見。其箇中原因有下列五點（Hambleton & Cook, 1977）：

1. 試題反應理論係一複雜的測驗理論。同時研究者需要具備較佳的數學能力，以便閱讀許多與此理論有關之數學方面的報告與專書，而甚感困難。
2. 大多數研究 I R T 的學者，在理論探討與介紹方面遠較實際應用者為多。
3. 並無迅速而容易使用的有關試題反應模式的電腦套裝軟體程式，使得試題反應各參數模式在應用上相當不便（此一問題，邇來已經大為改善）。
4. 部份學者對於 I R T 的研究與發展，所能獲得的成效深表懷疑。
5. I R T 的各種測驗模式，往往基於被限制在較強的假設上，當實際應用於心理測驗資料時，卻不能完全符合這些假設，而產生相當大的問題。

當代測驗理論相較於傳統測驗理論有其優勢之處，但它被用來解決實際真實資料者，則屬鳳毛麟角，余民寧（民82）也認為主要有下列五項原因：

1. 當代測驗理論係建立在理論假設嚴謹的數理統計學機率模式上，是一種複雜深奧、艱澀難懂的測驗理論，這對於在數學方面訓練有限的教育與心理學介學者而言，莫非是一大挑戰。閱讀有關此理論之數學方面的研究報告與專書，已感頗為困難，實在更難以深入將之發揚光大。
2. 多數當代測驗理論學者都出自於數學界或曾是學數學主修者，或至少在數理統計方面訓練有數者，他們偏愛對理論模式的探討，遠勝於對實際應用的推廣工作。
3. 過去，電腦科技的進步有限，沒有電腦套裝軟體的即時配合，當代測驗理論中對模式參數的估計，難以用手算或小型計算機順利進行。因此，在應用上更受限制。
4. 有些古典測驗理論的擁護者，對當代測驗理論的研究與發展，所能獲致的成效與應用性深表懷疑。為了證明與解釋疑惑，當代測驗理論學派的支持者，便朝向理論模式的量化技術方面探討，致使當代測驗理論的發展愈趨數學化、數量化與電腦化。
5. 礙於嚴苛的基本假設，當代測驗理論所能適用的教育與心理測驗資料有限，並且需要大樣本的配合，因此使得它的應用性大打折扣，未獲一般測驗使用者的全力擁戴。

由上述之說明可知，古典測驗理論在立論上雖不夠嚴謹，但因其理論簡明易懂，比較有利於實際測驗情境，特別是小規模資料的實施，如班級、實習課的分組等。而當代測驗理論之立論雖然既明確且嚴謹，但因其包含一「數學模式」

在內，較為艱澀難懂，僅適用於大樣本測驗資料的分析，如校際、地區、聯招或全國性技術士技能檢定學科筆試等。

## 第四節 我國中等學校題庫發展現況

### 一、我國中等學校數學題庫之發展現況

近年來教育主管機關為推行各級學校「教學正常化」，大力推展多元化的教學評量，從進修研習、溝通觀念，到改進命題技術，舉辦命題競賽，建立學科試題庫等。

八十年十一月，臺灣省教育行政會議結論，以「建立題庫」方式改進教學評量，包括建立校內的常模參照測驗及校際的標準參照測驗，以客觀檢驗各級學校的教學成效及教學品質，是否達到預期目標，確保受教者權益。據此，省教育廳進行一連串有關改進中小學教學評量的專案，除委託臺南師院進行有關國中教師參考用題庫及國中各學科標準化題庫的發展之外，亦成立高中部份的題庫建立中心學校（數學科由彰化高中負責），以團隊合作的方式將題庫的建立列為學校經常性的工作。

此外，教育部中教司、國教司亦先後進行有關國中及高中數學題庫及題庫管理系統的相關研究。

#### (一) 高中數學題庫

臺灣省政府教育廳成立各題庫建立中心學校，數學題庫由省立彰化高中負責，並分為下述三階段進行：

第一階段
------

蒐集、設計題目、建立廣義題庫  
=====

1. 主辦單位—各級學校
2. 預定完成期限—八十一學年度

3. 工作重點

- (1). 有計劃地蒐集過去的試題，並做選擇分類。
- (2). 對於不足的部份，由學校教師們分工，做地毯式的命題。
- (3). 上述兩種來源的題目，以卡片、清冊等書面方式建立題目檔案。
- (4). 由於題目數量較少，教師在設計測驗時須有固定百分比之題目來自題庫，其他不足之部份由教師自行補充。

第二階段

電腦題庫

=====

1. 主辦單位—教育廳業務科、各級學校
2. 預定完成期限—八十三學年度
3. 工作重點
  - (1). 訂定電腦提庫格式。
  - (2). 分工交由學校建立電腦題庫。
  - (3). 提供命題程式。

第三階段

狹義的題庫

=====

1. 主辦單位—教育廳業務科、各級學校、學術研究單位
2. 預定完成期限—八十五學年度
3. 工作重點
  - (1). 進行跨校之試題分析、修正試題。
  - (2). 對電腦題庫所產生之測驗進行信度、效度分析，並進一步對試題做修正，以增加由題庫製成測驗之信度與



效度。

(3). 建立區域性常模，以提供各科目測驗解釋之用。

## (二). 國中數學題庫

臺灣省政府教育廳除委託臺南師院進行國中教師參考用題庫之建立，此外，亦同時進行國中國文、理化及數學標準化題庫的發展。目前由臺南師院發展之國中數學科標準化題庫，共涵蓋國中數學一、三、五冊，各均有70題，分成甲、乙兩式測驗，其中有4題為定錨題，其預試樣本為臺南市、高雄縣市國中各年級學生4441人。目前之發展特色如下：

1. 分析出各試題之難度值  $P$  及全部試題之平均值。
2. 分析出各試題之題目與總分二系列相關及全部試題之平均值。
3. 全部測驗內部一致性信度 ( $\alpha$  係數) 均很理想。
4. 各成就測驗試題除了有傳統題目參數 ( $P$ 、 $r_{bis}$ 、 $r_{pb}$ ) 外，也均有三參數 logistic 模式試題參數 ( $a$ 、 $b$ 、 $c$ )。
5. 各冊數學科成就測驗均以建立南部地區  $T$  分數及百分等級常模。

此外，教育部國教司亦委託中華民國資訊教育學會將題庫發展之理論與技術結合，於八十二年六月完成國民中學數學科電腦化題庫之規劃與建立專案研究。該研究實際整合了測驗評量學者、國中數學教師、程式設計專家為研究主要成員，並且實際規劃及建立國中數學科第一冊電腦化題庫，同時每個試題亦進行傳統測驗參數分析。該題庫發展計劃之主要特色如下：

1. 分析出各試題之傳統測驗參數： $P$  值、 $r$  值、 $\Delta$  值及  $D$  值。

2. 實際分析數學科教學性質內涵並與以編碼規劃。
3. 提出建立題庫之文字、符號、圖形等問題的解決方案。
4. 發展題庫管理工具—題庫管理系統 I B M S 。
5. 結合人工智慧型之題庫編碼結構觀念。

## 二、我國數學題庫發展所面臨之問題

衆人之事難爲！題庫的建立是一個正確的開端，但具備內容效度及完善的題目特徵編碼賦與的題庫是需要大量人力、物力及經費來持續不斷的投入的。此外，一套功能完備的電腦化題庫管理系統是需要不斷的技术開發及經驗累積才能達成。

我國各級學校之相關數學科題庫的建立與開發已經具有相當的成效，但距理想仍有段相當的距離，以下諸端爲目前所遭遇之困難：

### (一). 測驗理論方面：

1. 學校教師之測驗評量素養有待加強。
2. 題庫之定義及實際功能有待進行普遍說明及溝通。
3. 學校教師對試題反應理論 I R T 之意義與用途不瞭解。
4. 適性測驗之原理、原則與重要性認知尚未普遍化。
5. 測驗理論素養的缺乏導致題庫良莠不齊。

(二).題庫建立方面：

1. 題庫建立需要大量特殊符號
2. 未來題庫之維護與整合困難
3. 題庫建立所需之大量圖形
4. 缺乏熱心推動之教師參與
5. 教師普遍不會操作電腦及中文輸入

(三).資訊技術方面：

1. 缺乏功能完備之題庫管理系統。
2. 多媒體型題庫之開發需缺乏人力與經費
3. 電腦化適性測驗之開發缺乏人力與經費
4. 智慧性整合性題庫之開發缺乏人力與經費
5. 中小學之硬體設備尚未充足及完整

(四).整合方案方面：

1. 缺乏統一之題庫檔各類型題目建題格式
2. 缺乏統一之題庫檔各類型題目特徵編碼規劃
3. 試題文字檔建立格式不統一
4. 試題圖形檔之檔案格式與大小缺乏統一規劃

5. 目前已存之題庫及題庫系統資源整合困難