

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 I R T 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

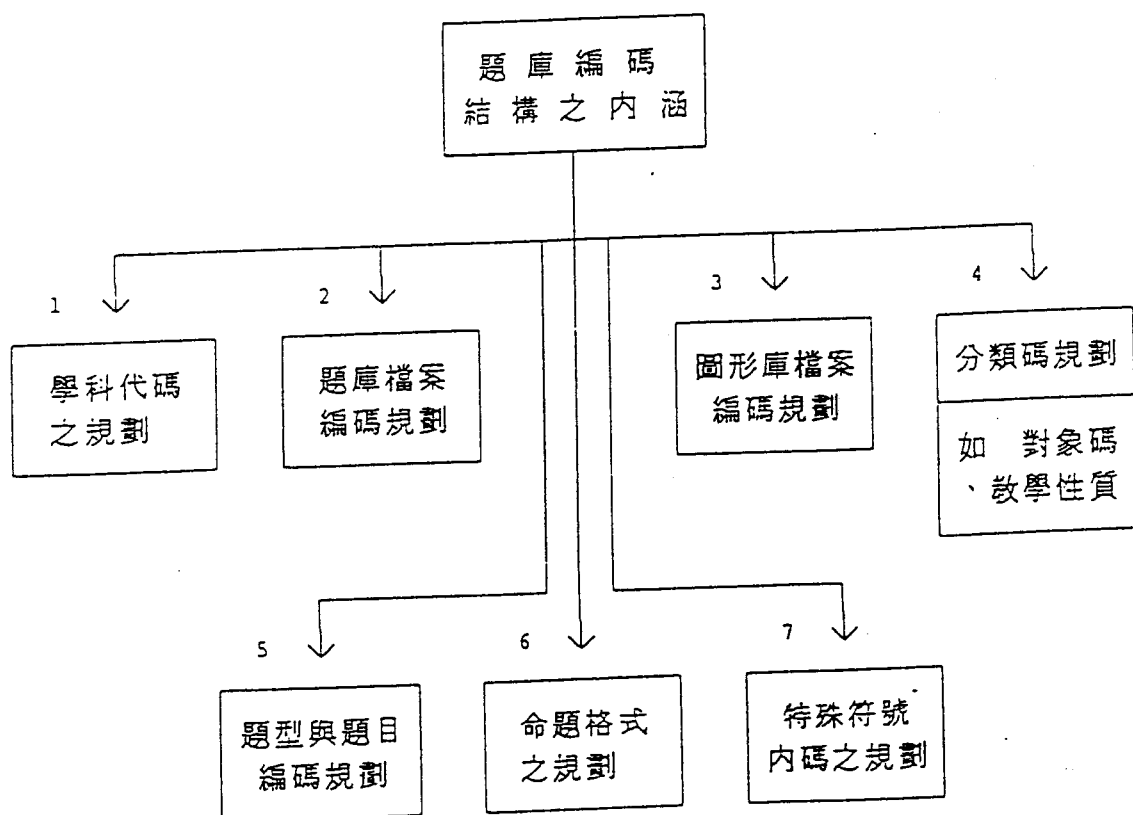


圖 2 - 1 題庫編碼結構之內涵

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 Δ

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 χ^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：

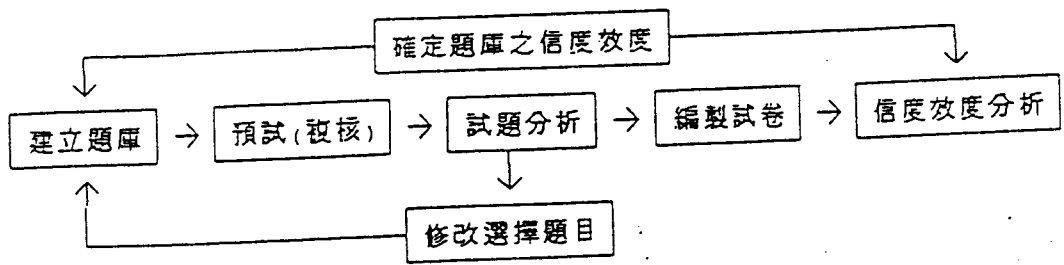


圖 2-2 廣義題庫建立流程圖之一

陳繼逸（民80）認為題庫的建立依其程序而言可為如下所示：

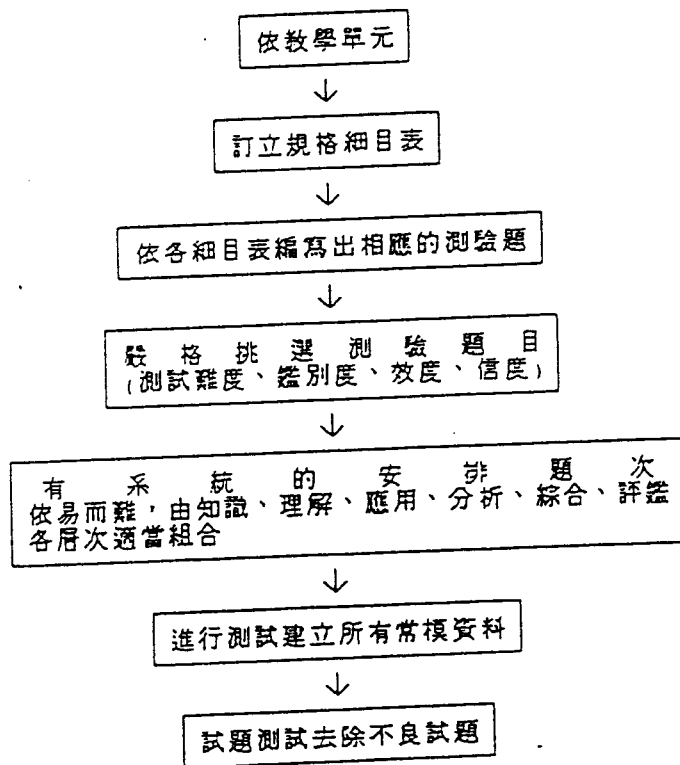


圖 2-3 廣義題庫建立流程圖之二

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：

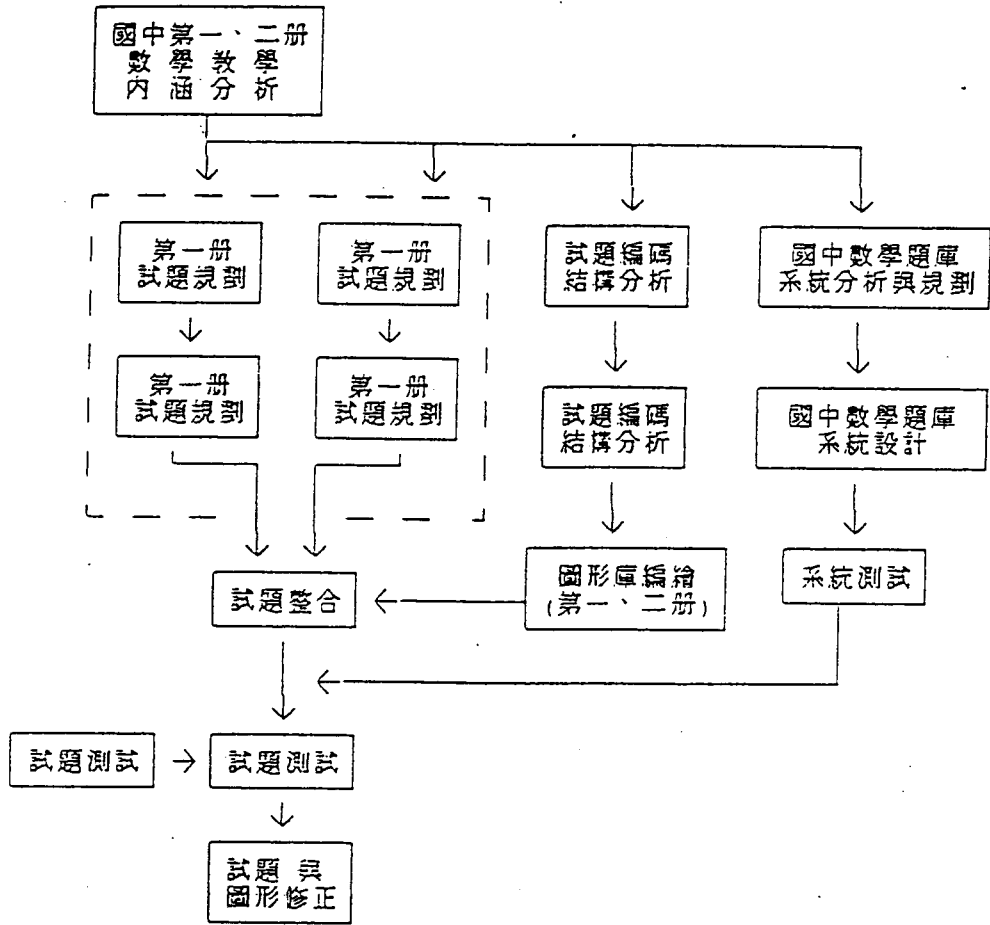


圖 2-4 教育部國教司國中數學題庫建立步驟

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：

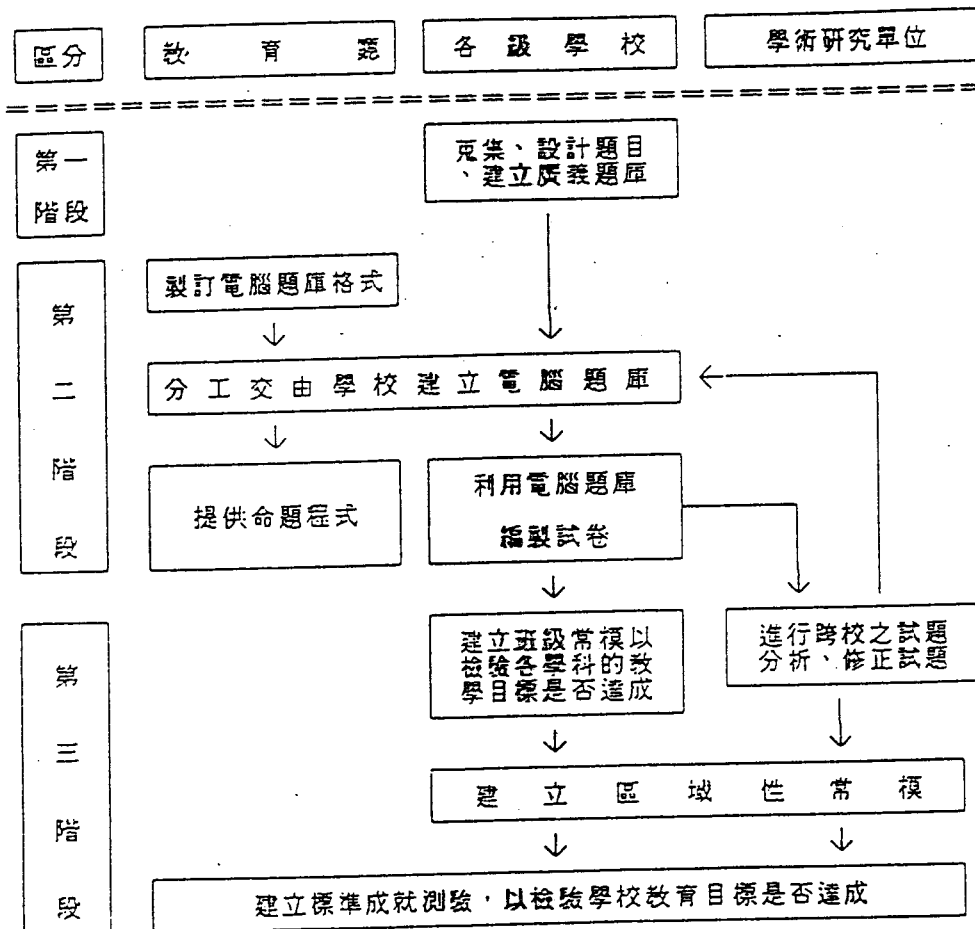


圖 2- 5 教育廳高中題庫建立之階段流程圖

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：

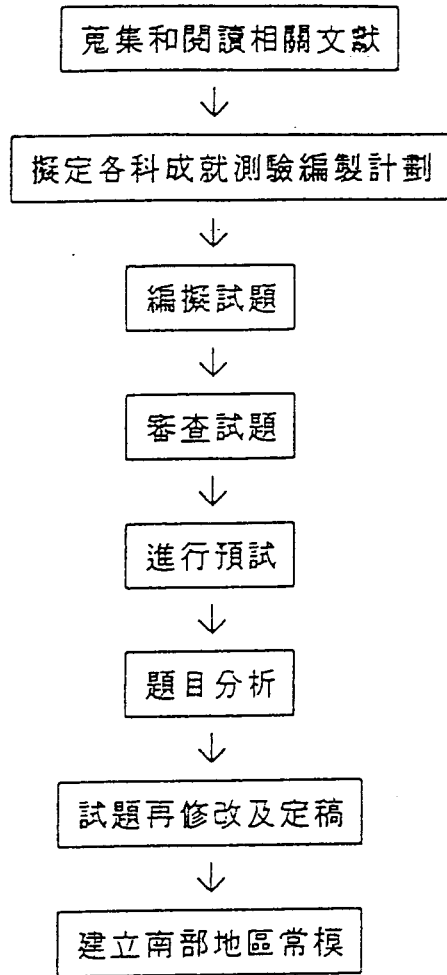


圖 2- 6 教育廳國民中小學題庫試題組成測驗之編製流程圖

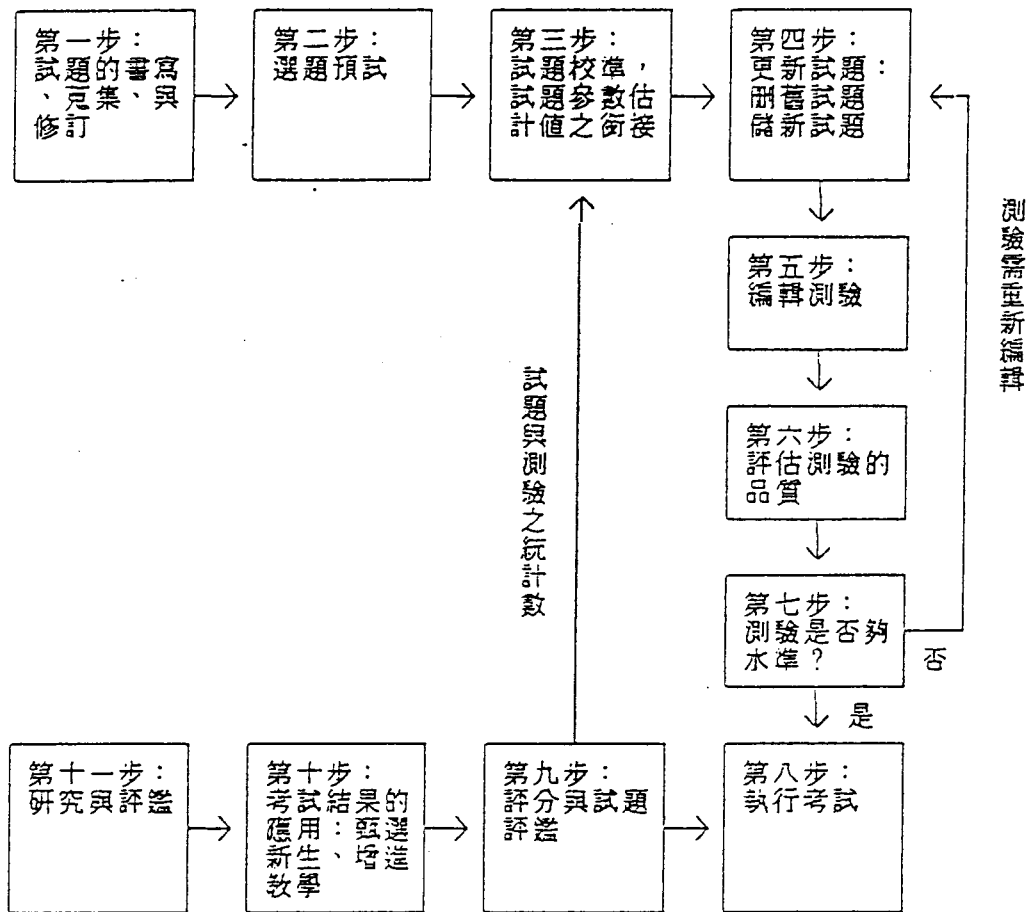


圖 2-7 以試題反應理論為基礎建立題庫之步驟

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)。