

具題庫維護回饋之測驗評量系統設計與實作

何俊輝* 姚家豪** 楊竹星**

*正修科技大學 電子工程系

**中山大學資訊工程系

jhho@cse.nsysu.edu.tw* m923040075@student.nsysu.edu.tw**

摘要

目前一般的測驗評量系統，普遍沒有設計「題庫維護」功能。故本研究旨在設計實作一個具有題庫維護回饋設計的測驗評量系統；透過測驗分析代理人系統，針對選擇題型(例：單選題、複選題)，進行「選項誘答力」(Distraction Analysis)分析，來判斷該試題是否具有鑑別力，再產生「題庫維護回饋」資訊給予命題者(Author)或題庫管理者(ItemBank Manager)，做為參考依據，使「題庫維護」工作更有效率，且題目資源壽命增長，適用性更廣。

關鍵詞：測驗評量系統、選項誘答力、題庫、題庫維護。

Abstract

At present, there is not the general assessment system to design the ItemBank maintainable function. So the goal of this research is to design and implement a test assessment system with ItemBank maintainable feedback mechanism. Through the assessment analysis agent to be directing against the choice type question, proceed with distraction analysis, to judge whether this question has discriminability. Then produce 'ItemBank maintainable feedback' information offers author or ItemBank manager, to be consultation basis, make the work that 'ItemBank maintainable' more efficient, and the life-span of examination question resource increases, suitability is wider.

Keywords: Assessment System, Distraction analysis, ItemBank, ItemBank maintainable.

1. 前言

「測驗評量」，無論在傳統教學或數位學習，皆是十分重要的學習活動，藉由測驗的進行，學習者可以瞭解學習過程中是否已有效掌握學習目標與要點，教師也可深入了解學習者的學習狀況，並評估學習者是否達到標準。

對於數位學習而言，為了能夠有效衡量學習效果，一個優良的測驗評量系統是不可或缺的。然而一個完整的測驗系統，除了測驗評量功能之外，還需要優良的題庫系統作為其基礎。

因為對於一個運作中的題庫系統而言，試題會與日漸增；每一道試題建立完成後，還需經過實際測驗的使用驗證，來決定該試題是否需要進行編修，因為試題可能具有題意不清或是命題不佳等缺點，所以適當地修改試題可以使其適用性更高，更具價值，即使當下之驗證符合需求，但是隨著時間遞移，題意及內涵還是會逐漸模糊或產生變異性。因此，為維護題庫內試題之有效性，命題者必須能夠對其建立之試題進行修改或是刪除的決定；題庫系統必須提供試題之使用資訊，以供命題者或題庫管理者瞭解試題之適用性，所以良好的「題庫維護」功能是不可或缺的。

傳統測驗之題庫在維護上，需要分析、統計與計算每次測驗後試題相關屬性，並記錄到試題檔案卡上，以便未來進行比對，找出適用性較低的題目，這樣的作法需要相當長的時間並耗費相當多的人力，在實行上成本相當高，並不太合乎經濟效益。

目前一般的電腦測驗評量系統，在基本測驗功能方面多半已經開發的相當完善，從教師建立測驗到學生參與測驗等相關功能模組一一齊全；但題庫管理系統部分，還是處於大概具備「新增試題」、「修改試題」、「刪除試題」等基本功能；在「題庫維護」功能的設計上，普遍沒有特別著墨。

有鑑於此，我們所建置的測驗評量系統，特別針對題庫系統部份予以加強，參考原本應用於教學策略修正的「試題選項誘答力」分析理論，將其擴展應用在「題庫維護」功能的設計上，分析受測者的「累計測驗結果」，透過測驗分析代理人子系統，將分析結果回饋給命題者與題庫管理者，做為維護題庫試題的參考，緊密結合測驗功能與題庫系統，讓本測驗評量系統的實用價值更高，適用性更廣。

本文結構說明如下，首先從文獻整理歸納出，題庫應該具備的特性，依據這些特性將傳統題庫與電腦化題庫做比較，探討其優缺點，以「試題選項誘答力分析」理論為基礎，從「判斷試題好壞」的角度切入，來設計題庫維護回饋資訊。在系統實作方面，利用 Use Case Diagram 分析使用者(受測者，命題者…)需求，再依其需求分析，建立系統主要功能架構，並逐一對各功能進行實作。

2. 文獻探討

2.1 題庫定義

題庫(Item Bank、Item Pool 或 Item File)是指根據雙向細目表命題，且經過施測與試題分析，建立起試題特徵資料的一群試題，每道試題均按照行為目標、教學內容、以及近似年級層次等予以編碼，並且可用來組成評量各種教學成果的工具。[1]

所以，如果沒有依據規範的程序進行編製、未經試題特徵訊息分析且未經系統編碼的市售題庫集或題庫本，嚴格說來，只能算是一種試題集，不能被稱為題庫。[2]

題庫具有下列改進測驗品質的潛能優勢，在可預期的未來，它對測驗編製者的重要性將與日俱增，同時節省編制測驗的時間和金錢[11]：

- (1). 可使測驗編製者(也許是學校教師或專業測驗機構的專家)隨心所欲地編制能夠符合各種目標的測驗。
- (2). 可使測驗編製者就題庫的範圍內，編制出每個目標都有適當題數測量之測驗。
- (3). 如果題庫能夠包含內容有效且編題技巧純熟的試題，則測驗品質通常會比個別測驗編製者自己編製的測驗品質還高。

2.2 傳統題庫的建立

建立大規模的題庫是一件耗資龐大的艱鉅工程，絕非個人所能勝任，需要學科領域專家、測驗專家、教師等相互合作來完成。但是，若只要建立一個學校或是單一課程適用的題庫，只要依循下列程序，應當不致有太大的困難。

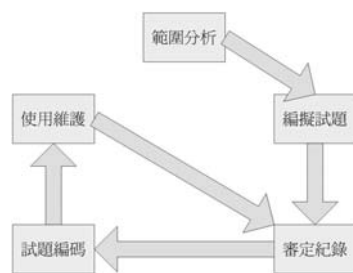


圖 1 傳統題庫建立流程圖

Step 1. 範圍分析：

傳統題庫使用雙向細目表來訂定範圍，雙向細目表是一幅題庫明確規格的藍圖，由教師來擬定設計，主要是有「教材內容」與「教學目標」兩個向度所構成，根據雙向細目表才可以兼顧課程教材與教學目標兩個方向。

Step 2. 編擬試題：

一旦雙向細目表確立之後，教師可以明確

清晰地針對教材內容範圍編擬試題，讓試題的鑑別目標更明確。

Step 3. 審定紀錄：

傳統試題記錄的方式，都是寫在一張 5x7 吋的檔案卡上，由教師(或是其他教師)再次仔細校閱、審定、修正或刪除後，始能定稿。

Step 4. 試題編碼 (Item codes)：

一旦試題通過審定，便要進行編碼(coding)的工作。編碼的功用是為了方便歸檔與管理，可以按照「科目別」、「年級別」、「冊別」、「題型別」、「序號別」等類別，來進行編碼。

Step 5. 使用維護：

每一道試題在編製完成之後，應該至少使用一次以上，並且在使用之後，進行試題分析及測驗分析，以建立起每一道試題的統計特徵指標，並將這些指標記錄在試題檔案卡，作為試題的基本屬性資料，以方便未來檢索使用。

當教師發現試題的特徵資料顯示該試題已經不合時宜，或是因為時代變遷而產生內容不符合需求，或題意及內涵已逐漸模糊，或是產生變異性時，必需再度重新進行審定流程，以做成是否局部修改、大幅修改、甚至整題刪除的決定。[2]

2.3 選項誘答力分析(Distraction analysis)

對選擇型試題進行選項的誘答力分析，可以藉由整體學生的答題情形，來診斷全體學生的學習狀況，以便進一步調整或改變教學策略[4]。

選擇型試題，其主要特色在於提供一個主要的正確答案供判斷選擇外，另提供兩個以上的不正確選項，用來吸引或迷感知識不夠完整或只有片段觀念的學生來選擇它們，以發揮其特有的「誘答」功能，增加試題的鑑別度；意即具有良好誘答能力的不正確選項，能讓程度不夠學生以「隨機猜題」(Randomly Guess)的方式來選擇它，所以選項誘答力的分析從另一角度來看亦可做為判斷試題好壞的參考。

選項誘答力分析在「教學策略」修正的應用上，主要是利用選擇型試題的優良鑑別能力(良好的誘答力)，來鑑別低分組(教學前、未接受教學、非精熟組)與高分組(教學後、已接受教學、精熟組)，兩組學生的能力表現，再依據下列原則來判斷是否需要調整教學策略。

原則：高分組選對正確答案選項的比率會比低分組選對的機率較高。

當與上述原則相反的情形發生時，則表示教師要審思，是否該選項所意涵的教材範圍，有內容編寫錯誤的問題，或是必須針對教學方式來進行改善（當條件設定上，試題不確定具有良好鑑別力的可能時，則要考慮題意不清，命題不當的情形）。

3. 題庫系統的評估研究

傳統題庫與本文所提出具回饋機制之題庫系統的建立都是以試題資源共享性(Sharable)與試題資源再利用性(Reusable)為主要目標，本文僅就以下觀點一一比較討論，並說明本題庫系統的設計。(參考表 1)

表 1 傳統題庫與具回饋機制之電腦題庫之比較

	傳統題庫	具回饋機制之電腦題庫系統
範圍分析	垂直相依性 (雙向細目表)	垂直相依性與平行相依性
儲存方式	試題檔案卡	資料庫
使用管理	效率較低	效率較佳
題庫維護	人工方式	提供題庫維護資訊

- (1).範圍分析：傳統題庫利用「雙向細目表」，在「教材內容」、「教學目標」兩個向度上均能涵括兼顧，而一般電腦化題庫不一定會做此分析；本測驗題庫系統兼具「垂直相依性」與「平行相依性」，在共享性與再利用性的層面較傳統題庫更為廣泛。
- (2).儲存方式：傳統題庫以「試題檔案卡」的方式來儲存題目，本題庫系統則是以資料庫來儲存試題，只要有適當的資料欄位設計、足夠的硬碟空間、良好的資料備份機制，在試題資料越來越龐大的狀況下，電腦化題庫在空間的使用效率、資料整理的便利性、資料完整度上都是較高的。
- (3).題庫使用與管理的效率：傳統題庫雖然按試題屬性編碼來方便搜尋與使用，但是所花費的人力與時間，與本題庫系統比較起來是不成比例的；透過良好的使用者介面，在使用題庫建立試卷、管理更新試題上，電腦化題庫的使用效率會比較高。
- (4).題庫維護：傳統題庫在維護試題上，需要耗費相當可觀的人力與時間，因為計算完試題的特徵資料(統計分析)後，需要一一記錄至試題檔案卡進行比對，耗費人力與時間成本過高；本題庫系統除了提供基本的題庫管理功能，例如：試題修改、試題刪除等功能，尚有支援「題庫維護」功能。

4. 具回饋機制之題庫系統設計

本題庫系統之設計參考傳統題庫的建立流程，亦分成五個主要步驟，分別為：範圍分析、編擬試題、審定紀錄、試題編碼與使用維護，來建構題庫，以下將就這五個步驟一一討論題庫建構流程的設計。

- (1).範圍分析：傳統題庫主要依循「教材內容」與「教學目標」兩大向度來確定題庫的範圍，主要目的是為了與「教材」做「垂直相依關聯」，來確定題庫的適用範圍；而本題庫系統除了與「教材」做「垂直相依關聯性」(題庫資訊功能可檢視)之外，亦兼具「平行相依關聯性」，意即本題庫系統所建構的試題資源，為獨立的教學資源，在別的課程或其他教師開設的課程，也可以分享再利用。例如：網路類的課程有「計算機網路」、「電腦網路」、「網際網路」等相似課程，其內容大同小異，故其試題資源是可以互相參考或移轉的(參考下圖)。^[6]

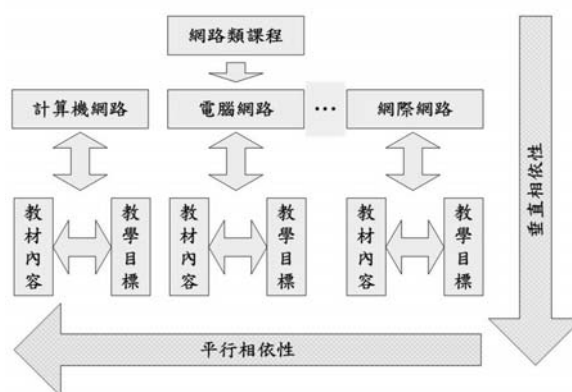


圖 2 「平行相依性」與「垂直相依性」

- (2).編擬試題：提供完整的試題新增功能介面，包含所屬課程類別、試題章節範圍、難易度...等試題屬性，讓教師進行試題編擬。
- (3).審定紀錄：提供齊備的試題管理功能與檢索搜尋介面，讓命題者可以對試題進行校閱、審定、修正或刪除等動作。
- (4).試題編碼：在試題的編碼上，題庫系統會自動為新建立的試題進行編碼，除了系統內部使用的試題識別碼外，尚有試題類型、難易度、專長類別、課程識別碼、章節識別碼...等編碼，以增進試題搜尋與管理維護的效率。
- (5).使用維護：傳統題庫在維護上耗費大量人力、時間；而本題庫系統則建立「題庫維護回饋」機制，來減少維護工作所耗費的成本(詳細請見 5.「題庫維護回饋設計」節)。

5. 題庫維護回饋機制之設計

「選項誘答力分析」理論如果單純應用在「教學策略」上，則是建立在試題皆具有良好鑑別力的條件下，依照低分組(教學前、未接受教學、非精熟組)、高分組(教學後、已接受教學、精熟組)兩組的測驗結果，來找出需要調整的教材範圍或教學方式。

本測驗評量系統應用「選項誘答力分析」理論，從「判斷試題好壞」的角度來切入擴展，修改原先的條件設定，將原本對「單次測驗結果」分析的設定擴大至對「累計測驗結果」分析。意即擴大原先的分組對象範圍，原先對象為單次測驗參加的學生，擴大至只要接受過該試題測驗的學生都會納入分析對象，就各個試題選項，計算低分組(成績為後四分之一的人)與高分組(成績為前四分之一的人)，分別對該試題選項的選擇率，依據以下三個原則予以判斷該試題選項是否具有誘答力，找出適用性低的試題，讓命題者或是題庫管理者能夠有效率的進行試題維護、修正的工作。(參考圖 3.)

- (1). 高分組選對正確答案選項的比率會比低分組選對的比率高。相反的情形發生時，則教師要審思是否該試題選項或是題幹題意不清，甚至是該試題所屬的教材內容編寫錯誤或教學方式不佳，需要調整改善。參考實際案例圖 4，可以看出在正確答案選擇率的比較上，低分組的選擇率是比較高且尚平均，可見其為「隨機猜題」的作答方式；所以該試題不具備鑑別力，甚至會降低使用該試題測驗的鑑別度，所以該試題必須修正。
- (2). 高分組選擇該選項的比率與低分組選擇的比率相等。代表該選項無法鑑別出低分組與高分組，即無鑑別力，需對該選項進行修正。參考實際案例圖 5，可以看到選項二，兩組的選擇率是相等的，所以該選項不具有鑑別力，必須建議修改。
- (3). 某不正確選項，低分組選擇率為零，代表誘答力為零。代表教師出題技巧需要再加強，因為該選項無法對學生的能力鑑別做出貢獻，需對該選項進行修正。參考實際案例圖 6，可以看到選項三與四，其低分組選擇率為零，代表其不具有誘答力，存在與否都不影響該試題鑑別度，必須建議修改。

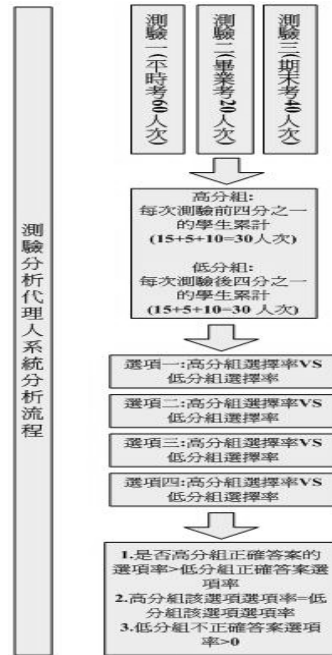


圖 3 題庫管理維護之測驗分析代理人分析流程圖



圖 4 問題試題-試題題意不清



圖 5 問題試題-不具備鑑別度



圖 6 不具備誘答力

6.系統需求分析與設計

按測驗評量系統需求分成兩種主要功能環境，「命題與評量環境」供教師進行題庫管理、測驗評量等相關功能；「受測環境」則是提供學生進行測驗的相關功能，因受篇幅限制之因素，本文在此主要強調「命題與評量環境」的功能，其他環境的功能分析設計請參考[4]。

本研究運用「使用案例圖(Use Case Diagram)」來分析與描述本題庫系統(ItemBank System)所具備的功能案例(Use Case)與所有系統外部參與者(Actor；可以為使用者，例如：命題者、受測者，或是其他外部系統、平台、子系統)之間的相互作用關係(Association)。

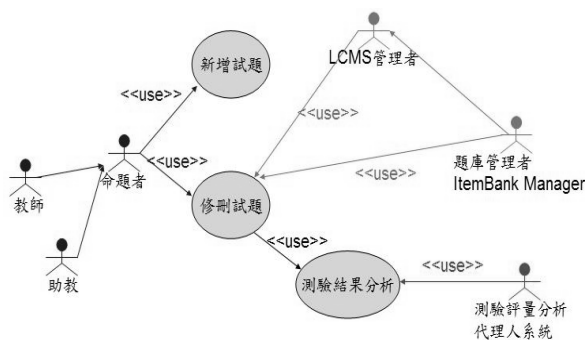


圖 7 命題評量環境下題庫功能使用案例圖

(1).Use Case Name：新增試題。

Description：提供命題者進行試題建立的功能。

Actors：

命題者(Author)為測驗評量系統中執行試題設計與試卷編製等作業之專長領域專業人士(Domain Expert)或教授課程之專長教師；在上圖中教師與助教都是其具體化子類別。

Pre-Conditions：需驗證是否具備命題者之身份。

Events：

- 1.填寫試題屬性，例如：章節範圍、困難度…等。
- 2.填入題幹內容、選項內容。

Post-Conditions：是否繼續建立下一道試題。

(2).Use Case Name：修刪試題。

Description：

修改試題、刪除試題與查詢試題等基本功能外，在修改試題功能介面上，使用「測驗結果分析」案例之功能，以顯示該試題目前的「試題維護回饋」資訊，使「題庫維護」的工作更加有效率。

Actors：

命題者(Author)、LCMS 管理者。LCMS 管理者為學習內容管理系統(Learning Content

Management System)平台之系統管理人員，上圖中題庫管理者(ItemBank Manager)為其具體化子類別，主要是專門管理題庫、試卷等教材資源。

Pre-Conditions：選擇欲管理或維護的試題。

Events：

- 1.修改試題屬性或題目內容。
- 2.刪除試題。

Post-Conditions：修正試題統計資訊。

(3).Use Case Name：測驗結果分析。

Description：

測驗評量分析代理人系統對「累計測驗結果」分析處理，誘答力選項率統計計算，將產生的回饋訊息回饋至「修刪試題」案例。

Actors：

測驗評量分析代理人系統。(針對系統之試題、試卷與測驗結果，個別與群體之使用成效、統計值與交叉影響進行自動分析並自動產生回饋至題庫管理者、受測者、教師等)，在此題庫管理者參考該回饋資訊來改善管理題目試卷資源。

Pre-Conditions：測驗狀態為進行中。

Events：分析測驗結果。

Post-Conditions：測驗狀態為完成。

6.系統功能模組與實作

依據上述之使用案例圖的需求分析，本系統以「命題與評量環境」、「受測環境」兩環境，來設計系統功能架構，呈現各環境下的所有功能模組。(參考圖 8 與圖 9)

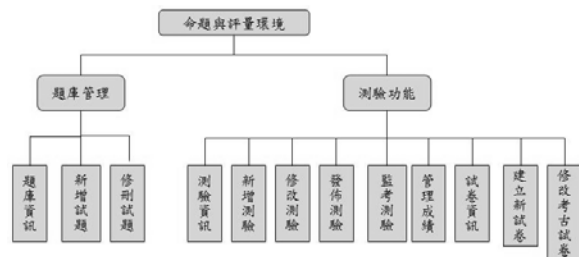


圖 8 命題與評量環境功能架構圖

「題庫管理」功能模組：

- 1.題庫資訊：提供了各章節範圍中各類型題數的數量資訊，讓命題者在新增試題時，能知道哪一範圍的題目數量不足，必須針對該範圍特別新增題目，使題庫具有足夠數量的題目來進行測驗；讓題庫管理者及 LCMS 管理者在管理維護題庫時，能作為題庫當前狀況的參考。

- 2.新增試題：教師填寫試題基本屬性供試題編碼所需，例如試題所屬課程、困難程度、是否開放共享、共享之分類等級，以及填寫題幹與反應項來編擬試題等功能。
- 3.修刪試題：教師審定修改、刪除試題功能，並提供「題庫維護回饋」來進行「題庫維護」的工作。

「測驗功能」功能模組：

- 1.測驗資訊：目前已建立測驗的資訊，例如：測驗開始時間、測驗狀態等。
- 2.新增測驗：新增測驗功能。
- 3.修改測驗：修改、刪除測驗功能。
- 4.發佈測驗：測驗發佈設定功能。
- 5.監考測驗：監考受測者狀態功能。
- 6.管理成績：管理該測驗受測者成績。
- 7.試卷資訊：目前已建立試卷的資訊。
- 8.建立新試卷：新增試卷功能。
- 9.修改考古試卷：修改已建立試卷功能。

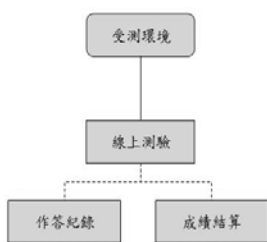


圖 9 受測環境功能架構圖

「線上測驗」功能模組：

當受測者進行線上測驗時，作答紀錄與成績結算為其必經的功能流程，所以用虛線表示其關係。

- 1.作答紀錄：受測者完成測驗後的最後作答紀錄，皆會儲存在資料庫內。
- 2.成績結算：完成測驗時，系統自動結算受測者成績，並給予學習回饋資訊。

7.研究貢獻與未來展望

目前市面上的測驗系統，除了測驗功能與基本題庫管理外，並沒有針對「題庫維護」與「測驗結果」來做分析探討，未能有效利用試題的受測結果，本研究透過分析測驗結果的相關統計資料，來產生管理維護試題的參考資訊，給與題庫管理者、命題者有效率又科學的試題維護參考資訊；妥善利用測驗結果，加強測驗功能與題庫系統之間的相依性與互補性，才能使測驗評量系統實用價值高，適用性廣。

本研究應用「選項誘答力」來設計「題庫維護」回饋，在現代測驗理論上尚有其他試題特徵參數，

例如：鑑別度、困難度等；這些參數在測驗統計的設計上，以單一測驗為樣本母體範圍作試題分析；如果擴展其母體範圍至累計測驗範圍，再做試題分析，應該亦能再設計出其他「題庫維護」回饋參考資訊，使題庫系統在維護上，更加便利，自動化的程度越高。

致謝：本研究係由國科會贊助之研究計畫的部分成果，計畫編號：NSC 92-2213-E-110-044，僅在此向支持者致謝。

參考文獻

- [1] Millman, J. and Arter, J. A. (1984). Issues in item banking. *Journal of Educational Measurement*, 21, 315-330.
- [2] 余民寧，教育測驗與評量(2005 再版)，心理學出版社股份有限公司。
- [3] 余民寧(2004 年 10 月)，心理與教育統計學，三民書局。
- [4] Haladyna, T. (1994) Multiple-Choice Formats. En Haladyna, T. (1994). *Developing and validating multiple-choice test items* (pp. 35-57). New Jersey: Lawrence Erlbaum Associates.
- [5] 何俊輝、楊竹星、姚家豪 (2004 年 12 月)。符合東方文化並可以跨平台的測驗系統。2004 教育資訊國際學術研討會，佛光人文學院。
- [6] 楊竹星、何俊輝、楊宗穎、涂嘉壽 (2003 年 6 月)。具有高度擴充性的網路教學與學習平台之研究。第十屆三軍官校基礎學術研討會，陸軍軍官學校。
- [7] 劉海峰(2003 年 12 月)。科舉傳統與東亞考試文化圈的形成。第七屆中華文明的二十一世紀新意義系列學術研討會，台灣桃園大溪。
- [8] 方鄒昭聰、詹健明(2003 年 12 月)。如何製作符合 IMS QTI 規格的線上學習評量試題。2003 電子商務研討會，私立逢甲大學。
- [9] 陳新豐、林邦傑、余民寧(2002 年 10 月)，線上題庫系統之研發，TANET 2002，國立清華大學。
- [10] C.Smythe, E.Shepherd, L.Brewer and S.Lay(February 2002)，IMS Question & Test Interoperability ASI Information Model Specification, Final Specification Version 1.2，IMS，February 2002.
- [11] Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer.