

行政院國家科學委員會補助  
大專學生參與專題研究計畫研究成果報告

\* \*\*\*\*\*  
\* 計畫名稱：線上測驗系統之試卷選題策略研究：以關聯式規則為基礎 \*  
\* \*\*\*\*\*

執行計畫學生：高永昊  
學生計畫編號：NSC 97-2815-C-216-002-S  
研究期間：97年07月01日至98年02月28日止，計8個月  
指導教授：應鳴雄

處理方式：本計畫涉及專利或其他智慧財產權，1年後可公開查詢

執行單位：中華大學資訊管理學系

中華民國

98年03月23日



行政院國家科學委員會補助  
大專學生參與專題研究計畫研究成果報告  
以關聯式規則與基因演算法為基礎之試卷選題策略

計畫編號：NSC97-2815-C-216-002-S

執行期限：97年7月1日至98年2月28日

指導教授：應鳴雄 教授 中華大學 資訊管理學系  
執行計畫學生：高永昊

## 1. 中文摘要

測驗的目的在增進學習的效果，因此電腦基礎測驗及線上測驗已成為重要的趨勢。目前許多線上測驗系統的試卷出題方式，是由事先編制好的題庫中隨機選題，此方式產生之試題組合是否能涵蓋足夠的深度及廣度？是否能評量學生不同認知層次的能力？是否能避免試題間高度相關？都是值得被探討的問題。本計畫將藉由測驗評量理論、資料探勘、基因演算法以及 Bloom 認知分類修正版等概念，提出一個電腦自動編制高品質試卷的選題策略，以使得電腦產生試卷的試題組合能夠具備適當的深度及廣度、涵蓋用以評量不同認知層次的試題，並使試題間具有高度獨立性，以確保選出的試題組合具有評量的意義與試卷品質。

**關鍵字：**出題策略、關聯式規則、基因演算法、Bloom 認知分類

### Abstract

Since the rapid Internet development, the e-learning has been one of important issue of information education. The main purpose of tests is to improve the effect of learning. The computer-based test (CBT) and online test (OLT) have been the important trends.

Nowadays, many online testing systems are still randomly generates test papers from the item bank. A high quality test paper must to consider the some issues as bellow: Is the depth and breadth of test items enough? Can test items examine student ability with different cognitive

levels? Can test items avoid high relationship among test items? Can test paper identify student ability and provide proper learning suggestion. Therefore, it is the important issue to solve above problems by using information technology.

This research provides an item selection strategy by computer arrangement automation base on assessment theory, data mining, genetic algorithm and revision of Bloom's taxonomy, to ensure that the test paper have a high quality.

**Keyword:** Item selection strategy, Association rule, Genetic algorithm, revision of Bloom's taxonomy

## 2. 計畫緣由與目的

測驗能增進學習成果，也能獲得學生學習成效 (Brueckner & Band, 1955; Lipscomb, 2001)。近年來線上測驗 (online test, OLT) 及電腦基礎測驗 (computer-based test, CBT) 具有提高測驗效率、評量客觀性、豐富回饋、生動與互動性的優點 (何榮桂, 1991)，更已成為電子化學習重要的趨勢。目前許多線上測驗系統的試卷出題方式，是由題庫中隨機選題，然而運用這種方式產生之試題組合是否能包含足夠的深度及廣度？是否能評量學生不同認知層次的能力？是否能避免試題間高度相關？都是有必要深入研究的議題。

本計畫將以測驗評量理論、關聯規則、基因演算法以及 Bloom 認知分類修正版等概念，針對試卷試題做一深入的研究，並評量透過 GA 演算法選題的試題組合是否能具備適當的深度及廣度、試題間

的獨立性，以及是否能在評量不同認知層次的試題組合上，獲得較佳的分佈。綜合上述，因此，本計畫的選題策略，將考量多種可能影響試題組合適當性與題目相關度性的因素，例如難度、鑑別度、Bloom 認知分類、試題內涵...等。

而測驗理論主要分為古典測驗理論及當代測驗理論，前者是以真實分數模式為主體 (Lord, 1968; Gulliksen, 1987)，後者則以試題反應理論 (item response theory) 為架構 (Lord, 1980)。因此在具備「自我學習」的線上學習環境中，學習者可藉由線上測驗來評量學習的成效，但透過測驗及評量並非只希望單純獲得一個分數而已，重要的是應重視試題與教育目標的配合，使得測驗成為輔助學習的一部份。若要達到這個目標，線上測驗系統中的試題便需包含每個試題所能評量的層次及試題資訊。

余民寧 (1997) 認為試題分析是針對試題的統計特徵進行量化分析，其特徵包括鑑別度 (discrimination) 以及難度 (difficulty)。經由特徵分析，教師可以得到各個試題的資訊，有助於提高日後編製試卷的經驗，也可以間接提升測驗的信度和效度，並且以此挑選優良試題，建立試題的題庫，因此教師若能在實施測驗的前後，就進行試題分析的動作，並且保留優良的題目以供建立題庫之用，經過一段時間便可以累積許多優良的試題，提供未來出題的參考，不僅省時省力，更可以集合測驗和評量的功能，來達到改進教學評量的目的，並且經由學生實際施測來進行試題的難度及鑑別度檢驗。

關聯規則 (association rule)，是資料探勘中常使用的一種技術，它能在大量資料中，挖掘出資料間隱藏的相關知識。例如：零售商店店長發現顧客買汽水時會和零食一起選購。所以將兩類商品貨架放在附近，並只針對其中一種商品進行折扣促銷，連帶另一項未促銷的商品，也會同樣獲得青睞，達到商場銷售提高的目的。因此本計畫希望藉由計算試題之間的相關係數，將相關度較高的試題避免出現在同一份試卷之上，此種泛相關係數將透過後續

的試題分析來進一步探討。

Holland 的基因演算法是一種利用達爾文進化論概念，找尋最佳解的演算法 (Holland, 1975)。此類演算法是透過操作染色體的方式進行演化，並在期間藉由運算繁衍出子代，透過適應函數 (fitness function) 的設計可選出適合環境的子代繼續繁衍，並獲得較適合環境的物種 (Chen, Y. P., 1999)，此方法的運算過程包含了產生初始族群、染色體基因解碼、計算適應函數值、複製、交配、及突變等六階段。

由於本計畫的試題組合種類數字相當龐大，而運用基因演算法能在合理的時間範圍內求得近似最佳解，採用基因演算法的優點如以下所述：

- (1) 基因演算法是採取「多點搜尋法則」，能夠在同一時間考慮搜尋空間上的多個點而不針對單一個點來進行計算，透過適應函數值的高低以及生物演化的過程，留下優良的基因來逐漸逼近最佳解。
- (2) 基因演算法利用「機率準則」的方式來引導搜尋的方向，因此在處理不同類型的問題最佳化上，都有不錯的效果，這種隨機性和適應性，可以針對整個搜尋空間做更仔細的且徹底的檢查，有效避免落入區域最佳解的機會，這個特性也是使用基因演算法的最大優點。

當前基因演算法已經廣泛的運用在工程及科學上最佳化求解問題，以及商業與財務預測上，諸於解決財務決策和投資組合決策問題上。

近年來 Bloom 等人 (1956) 的教育目標分類在國內外教育界廣為採用，而 Anderson 等人重新修訂了 Bloom 教育目標認知分類系統，並將原先單向度的認知分類區分為「知識向度」與「認知歷程向度」向度，知識向度又區分為事實、概念、程序、後設認知等四類知識 (Krathwhol, 2002)，而認知歷程向度則區分為記憶、了解、應用、分析、評鑑、創造等六種層次能力 (葉連祺、林淑萍, 2003; Krathwhol, 2002)。因此本計畫先將題庫中的每個試題先經由過去測驗記錄產生難度、鑑別度、

Bloom 認知分類等資訊，再透過系統中的 GA 演算法產生出足以涵蓋各種知識層次及認知層次的試題，並同時考量最適當的試題難度與鑑別度，以確保產生的試卷擁有適當的深度及廣度，後續再經過學生的測驗結果來對產生之試卷試題層次比例加以分析，再從教師的認知角度來評量試題的可用性及品質。

過去的研究指出，要建立與維護測驗試題題庫是一項曠日費時的工作，同樣地，教師為了編制一份高品質的試卷所花費的時間與心力也非常多，因此本計畫結合上述優良試卷的四項衡量指標，諸如適當的難度、良好的鑑別度、較低的試題相關以及最廣泛的試題層次，希望透過自動化的試卷選題策略來達成產生高品質試卷這項目標

本計畫計畫共計八個月，目前呈現的成果是植基於選題策略產生高品質試卷為主，後續計畫則針對試題內容資訊剖析及儲存、試卷編製及品質實證進行研究，探討選題策略之有效性。

### 3. 研究架構與系統建置

#### 3.1 研究方法

在銷售領域中，研究者希望藉由關聯規則找出商品間的相關性，並利用相關性高的商品組合來提升營運績效。在測驗命題領域中，教師則希望試卷中試題間的相關性越低越好，更期望能選出具備最大廣度、適當的難度與鑑別度，以及能涵蓋評量多種不同認知層次的試題組合。本計畫基於此觀點，結合資訊技術及試題理論，在線上測驗系統中依據這個試卷選題策略進行實作，最後再針對試卷選題結果，進行試卷品質的適性函數分析。

#### 3.2 研究步驟

##### 3.2.1 題庫建立與資料分析

研究初期，本計畫先於線上測驗系統中建立了「企業資源規劃」課程之 1 至 6 章試題共計 229 題，每一試題均依據內容，設定其所屬之 Bloom 認知分類向度資料，之後實際測試階段將讓學習者於課後

登入至本系統進行課後評量。由於題庫中的試題數量龐大，因此在選題策略效能檢驗中，僅選擇其中兩個章節共產生 20 題試題為題庫研究範圍，但為了計算試題難度與鑑別度，本計畫後續實際測試階段將針對每位學生的測驗成績進行標準分數轉換，以進行高低分群的排序。由於在建立題庫難度、鑑別度等資料時需要依照課程進度安排之隨堂測驗來做實際的數據建檔作業，因此採用每位學生都測驗不相同的隨機考卷，所以每位同學的試卷整體難度並不相同，因此本計畫將利用以下公式  $NS=OS*(1.5-AD)$  來進行標準分數轉換，以利產生成績排序及確認高低分群組。公式中的 OS 為學生測驗後的原始成績，AD 為該學生所填寫試卷的試卷平均難度，NS 為經過轉換調整後的標準成績。若計算後的  $NS>100$ ，則  $NS=100$ 。當成績轉換後，進一步對各試題進行試題難度與鑑別度分析，並利用關聯規則找出試題間的相關係數，再將這些數據結果儲存於資料庫中，以作為期中與期末正式測驗時所要產生統一試卷的分析資料，本計畫試圖利用試題間的相關係數、試題難度與鑑別度、試題所屬之 Bloom 分類之分析，由系統自動選取出最佳試題組合的期中考與期末考試卷。

- (1) 試題難度：依受試者答對的百分比來表示，答對的百分比越高則表示題目越容易，通常試卷平均難度接近 0.5 為佳，其公式為  $DP=R/N$ ，公式中  $DP$  代表試題難度， $N$  為全部受試者人數， $R$  為該題答對之人數。
- (2) 試題鑑別度：依據學生測驗總分排序後，選取成績最高及最低之 25%~33% 學生為高分組及低分組。本計畫設定之分組選取門檻為 33%，各試題之鑑別度公式為  $DI = P_H - P_L$ ， $P_H$  為高分組答對人數百分比， $P_L$  為低分組答對人數百分比， $DI$  為鑑別度，整體試題鑑別度越高，則測驗信度就越高 (Ebel, 1967)。

##### 3.2.2 相關係數分析

此階段針對題庫試題間的關係進行分析，並從資料庫中擷取出任兩試題（試題

編號分別以  $I_a$  及  $I_b$  表示) 同時出現在學生課後評量試卷中的次數  $ST_{a,b}$ ，以及  $I_a$  與  $I_b$  兩試題同時在試卷中被答對的次數  $CT_{a,b}$ 、與同時被答錯的次數  $IT_{a,b}$ 。

$S = \{I_1, I_2, \dots, I_n\}$  表示為試題集合， $n$  為題庫中的試題數， $R_{a,b}$  為  $I_a$  與  $I_b$  試題間之相關係數，其公式為  $R_{a,b} = (CT_{a,b} + IT_{a,b}) / ST_{a,b}$ 。例如，50 位學生參與課後評量測驗，系統會產生 50 份不同試卷組合的試卷，其中的試題 7 與試題 29 同時出現在相同試卷的次數共有 25 次，而試題 7 與試題 29 在這 25 份試卷中同時被答對 12 次，同時都答錯的試卷有 8 次，則試題 7 與試題 29 的相關係數  $R_{7,29} = (12 + 8) / 25 = 0.8$ 。

### 3.2.3 染色體與適性函數設計

本計畫之染色體設計採二進位編碼，1 代表該題將選取於試卷上，0 代表該題將不被選取於試卷上。假設題庫中包含本次測驗範圍的試題共 10 題，並以  $C_1 \sim C_{10}$  表示，若要產生僅包含 4 個試題的試卷，則其中一種試卷組合的染色體如圖 1 所示。

$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$C_6$	$C_7$	$C_8$	$C_9$	$C_{10}$
0	1	0	1	1	0	0	0	1	0

圖 1  $s_3(2)$  的染色體設計

本計畫之適性函數設計是基於測驗品質理論的難度、鑑別度等指標，再搭配試題間的相關係數、Bloom 認知分類之題型比例，以尋找較佳之試題組合。研究中的  $t$  表示為基因演算法目前執行的世代數， $S(t)$  表示為第  $t$  世代的母體， $s_i(t)$  是  $S(t)$  中第  $i$  個試題組合成員， $x$  為  $s_i(t)$  中的基因個數 (題庫中測驗範圍試題總數)， $y$  為每份試卷組合  $s_i(t)$  所需選取的試題總數， $f(s_i(t))$  是  $s_i(t)$  的適性函數值，TOTFIT( $t$ ) 是所有  $s_i(t)$  的  $f(s_i(t))$  總和。試卷品質適性函數  $f(s_i(t)) = W_{dif} * ADif(s_i(t)) + W_{dis} * ADis(s_i(t)) + W_{ac} * AC(s_i(t)) + W_{rb} * RB(s_i(t))$ ，其中  $ADif$  代表第  $s_i(t)$  個試卷組合的試題難度適切值， $ADis$  代表第  $s_i(t)$  個試卷組合的試題鑑別度適切值， $AC$  代表第  $s_i(t)$  個試卷組合的試題間平均相關係數， $RB$  代表第  $s_i(t)$  個試卷組合的 Bloom 認知分類題型比例適切值。而

$W_{dif}$ 、 $W_{dis}$ 、 $W_{ac}$ 、 $W_{rb}$  分別是上述四項變數的加權值，而實務應用時可依據教師對此四項變數的重視程度與偏好來設定加權比例。

(1) 難度適切值  $ADif$  與鑑別度適切值  $ADis$ : 在計算試卷品質適性函數值  $f(s_i(t))$  時，需先針對第  $s_i(t)$  個試卷組合進行試題難度的數值轉換。 $ADif(s_i(t))$  為第  $s_i(t)$  個試卷組合的難度適切值，此數值越高表示此試題組合越佳，其計算公式如下。公式中的  $x$  為此次測驗範圍在題庫中之試題總數，亦即  $s_i(t)$  的染色體個數， $y$  為試卷需求之總試題數， $PD_j(s_i(t))$  為  $s_i(t)$  中第  $j$  個基因對應到題庫中試題的試題難度，其中  $C_j = 0,1$ 。

$$ADif(s_i(t)) = 1 - \frac{\sum_{j=1}^x [C_j * PD_j(s_i(t))] / y - 0.5}{\sqrt{\sum_{i=1}^x (PD_i - \overline{PD})^2 / x}}$$

而鑑別度適切值  $ADis(s_i(t))$  則為第  $s_i(t)$  個試卷組合的鑑別度適切值，該值可由  $s_i(t)$  中所包含的  $j$  個染色體各別對應到的試題鑑別度之平均值來取得，其計算公式為  $ADis(s_i(t)) =$

$$\frac{\sum_{j=1}^x [C_j * DI_j(s_i(t))] / y + \sqrt{\sum_{i=1}^x (DI_i - \overline{DI})^2 / x} - \sum_{i=1}^x DI_i / x}{2 * \sqrt{\sum_{i=1}^x (DI_i - \overline{DI})^2 / x}}$$

(2) 平均相關係數適切值  $AC(s_i(t))$ : 針對  $s_i(t)$  中所有  $c_j=1$  的染色體，將其所對應的所有試題編號，從資料庫中取出這些試題編號間的相關係數。則平均相關係數適切值公式為  $AC(s_i(t)) =$

$$\frac{\sum_{i=1}^{c_2^x} AC_i / c_2^x + \sqrt{\sum_{i=1}^{c_2^x} (AC_i - \overline{AC})^2 / c_2^x} - \sum_{j=1}^{c_2^y} (AC_j / c_2^y)}{2 * \sqrt{\sum_{i=1}^{c_2^x} (AC_i - \overline{AC})^2 / c_2^x}}$$

。以圖 1 為例， $C_2$ 、 $C_4$ 、 $C_5$ 、 $C_9$  為第 2 世代第 3 個試題的被選取組合，而這 4 個為 1 的染色體分別對應到題庫中試題編號 8、12、23、32 的試題，則可獲得  $C_2^4$  種組合的試題相關係數  $R_{a,b}$ ，平均相關係數公

式可得出平均相關係數適切值。

(3) Bloom 認知分類題型比例適切值：由於後設知識通常不列入正式課程評量中，而創造層次的試題也不易透過是非、選擇等題型來評量，因此本計畫僅針對線上測驗系統能自動評分的基本題型進行試卷選題策略研究。所以本計畫僅採用「知識向度」上的事實、概念、程序等三類知識，「認知歷程向度」則採用記憶、了解、應用、分析、評鑑等五個層次。而 Bloom 認知分類各題型比例適切值，是以表 1 中 15 種試題向度分類的試題數量  $n_{qk}$  佔出題範圍總題數  $n$  之比例所計算出的期望值  $E_{qk}$  與實際選出該分類試題數之相似比例，期望值  $E_{qk}$  中的  $q$  與  $k$  分別表示為表 1 中知識向度 A~C 的知識及認知歷程 1~5 的五種認知。

表 1、本計畫採用之 Bloom 分類表範例

認知歷程 知識向度	1.記憶	2.了解	3.應用	4.分析	5.評鑑
A.事實知識	20	10	10	5	0
B.概念知識	10	5	8	0	0
C.程序知識	10	0	2	0	0

表 1 為某份特定測驗範圍的試卷分類分佈狀況表，題庫中包含此測驗範圍的試題共有 80 題。以  $qk='A2'$ （屬於事實知識及了解層次）的試題為例， $n_{qk} = 10$ ，此類試題被命題的機率  $P_{qk} = 10/80 = 0.125$ ，假設要產生一份試題數量  $x$  為 20 的試卷，則此類型的期望值  $E_{qk} = 20 * 0.125 = 2.5$ ，四捨五入後  $E_{qk} = 3$ ，即表示  $qk='A2'$  類型試題

在產生試卷的期望出題數為 3 題。但是若各類試題的  $E_{qk} > 0$  且  $E_{qk} < 0.5$ ，為確保試題中能包含此測驗範圍中的所有類型試題，則強制  $E_{qk} = 1$ 。為了強制各類試題都能被納入試卷中，因此當  $\sum E_{qk} > x$  時，試卷的期望總題數會大於實際規範的試題總數  $x$ 。此時電腦將會將表 1 中的最大期望值  $\text{Max}(E_{qk})$  找出（即  $qk='A1'$  的  $E_{qk}$ ），並讓  $qk='A1'$  的  $E_{qk}$  減 1，本計畫之系統會反覆此動作直到  $\sum E_{qk} = x$  為止。若  $E_{qk}$  相同則比較其未進位前之期望值，優先由較低之期望值之  $E_{qk}$  減去。而 GA 演算法中  $s_i(t)$  所轉換得到的 Bloom 認知分類題型比例適切值  $RB(s_i(t))$  的公式則表示如下，其中的  $T_{qk}$  為在  $s_i(t)$  中實際選取到的各類試題數量。

$$RB(s_i(t)) = \left[ \sum_{q=1}^3 \sum_{k=1}^5 \left( 1 - \left| \frac{E_{qk} - T_{qk}}{E_{qk}} \right| \right) \right] / 15$$

根據上述適性函數指標的轉換，本計畫期望試題組合之適應函數值越高越好，並能藉以提升試卷品質及確保評量意義。

### 3.3 系統建置

#### 3.3.1 出題系統建議架構

此階段將實作發展雛形系統，此系統包含以下幾個模組，各模組間和資料庫的關係如圖 2 所示，以下針對「出題子系統」與「線上測驗與評分子系統」進行詳細的模組元件及運作方式說明

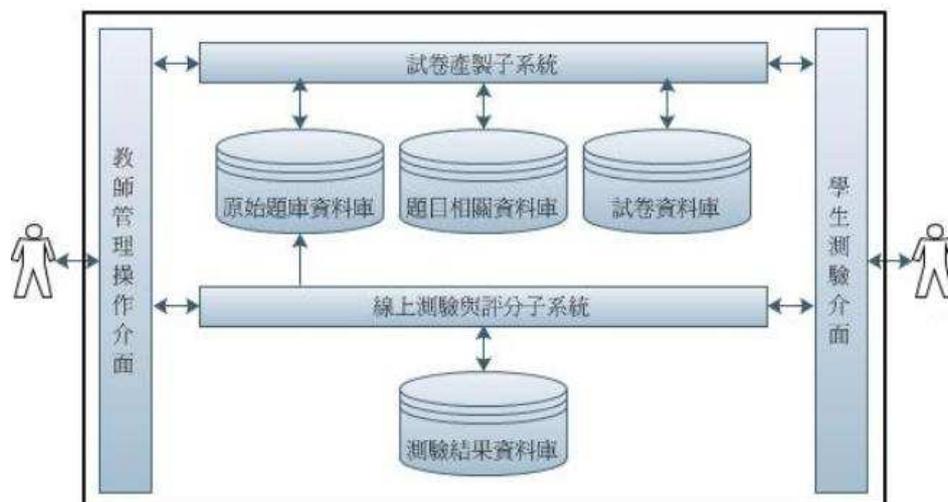


圖 2 出題系統建議架構圖

### 3.3.2 出題子系統之系統架構

出題子系統之系統架構如圖 3 所示，各元件之內容說明如下：

(1)原始題庫管理模組：此模組包含了題目資訊查詢、試題新增、試題刪除與修改等功能。

A. 題庫資訊查詢：顯示目前題庫試題狀況，提供了各個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較少，必須針對這些範圍特別新增題目，使題庫具有足夠數量的題目來進行測驗；讓題庫管理者在管理以及維護題庫時，能作為一個題庫狀況參考。

B. 試題新增：新增試題功能中可編輯許多試題屬性資訊，例如試題編號、試題難易度、試題鑑別度、配分、布魯姆認知分類、試題內容、試題選項以及試題章節等等。

C. 試題修改與刪除：基本修改試題屬性的功能，包含了試題內容、試題選項、配分、試題難易度、試題鑑別度、布魯姆認知分類以及試題章節等等，提供命題者執行試題的重新編輯或是刪除不適當的試題。

(2)資料分析模組：主要目的在於利用基因演算法根據適性函數來挑選出組成一份試卷最佳的試題組合，適性函數的設計為分析題目之間相關係數，搭配難度與鑑別度所得，經由此模組得出之試題相關係數會作為試卷出題篩選之標準。

A. 答題狀況統計：統計出二試題之間，均答對的作答次數與均答錯的作答次數總和來計算相關係數。

B. 出題模組：主要是由題目相關資料庫中篩選出經過基因演算法產生出較佳的試題來產生試卷。

C. 出題方式選取：供命題者選擇產生測驗試題試卷之方式，例如使用隨機出題或是採用策略出題。

(3)測驗模組：

A. 測驗資訊：包含目前已建立的各個測驗的狀況、開始以及結束時間、是否正在測驗中以及成績公告時間等。

B. 新增測驗：出現測驗表單供命題者填寫建立一項新測驗所需的相關資訊，例如該項測驗之名稱、測驗類型、測驗配分以及測驗標準等等；並且進入試卷選擇畫面，利用試卷資料搜尋檢索功能的設計來協助命題者取得合適的試卷，以滿足此次測驗之需求。

C. 修改測驗：主要提供以下功能選項來協助命題者修改測驗基本屬性，修改測驗名稱、總成績配分比例等等。

D. 測驗發佈：對已建立完成的測驗，進行測驗時間的設定，例如測驗開始時間、測驗結束時間以及測驗考試時間等等屬性。

(4)出題模組：教師可選擇使用隨機出題或是策略出題來產生試卷，選擇策略出題時可以設定難度、鑑別度、相關係數以及布魯姆比例分配之權重，並且搭配布魯姆比例分配之試題進位方式選擇希望之試題分佈，提供一個彈性的出題環境。

A. 原始題庫資料庫：存放在學期開始前建立之題庫試題資訊，包含試題內容、試題選項、配分、試題難度、試題鑑別度、布魯姆認知分類以及試題章節等。

B. 題目相關資料庫：存放兩試題之間相關係數值，作為後續計算試卷平均相關係數之參考，由於本計畫所探討的相關是屬於泛相關的範疇，因此在後續試題分析上必須對試題間的相關係數做解釋。

C. 試卷資料庫：存放經過隨機出題或是策略出題後所產生之試卷，提供教師發佈考試所需的試卷支援。

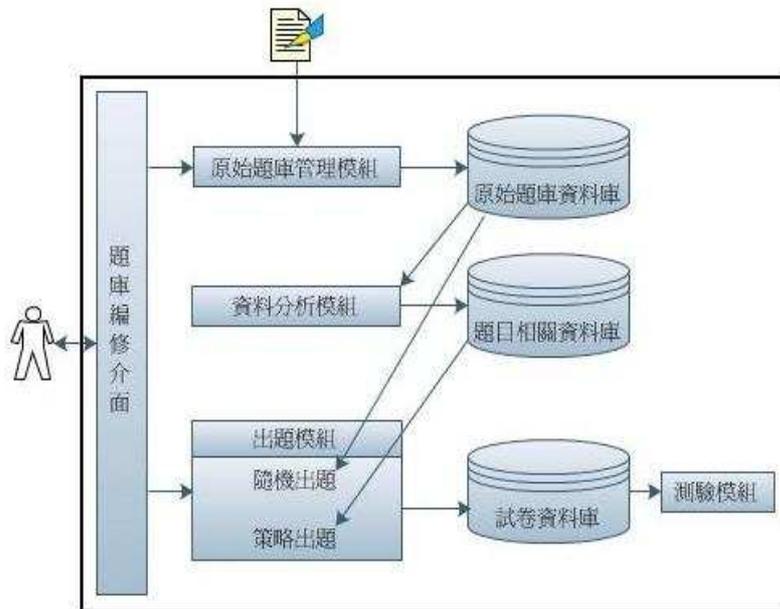


圖 3 出題子系統之系統架構圖

### 3.3.3 線上測驗與評分之系統架構

#### (1) 評鑑模組：

- A. 測驗成績：列出每位學生在各測驗中的成績，並且可以進入該測驗觀看其測驗歷程，作為教師評分的參考。
- B. 學習警示：提供教師或助教瞭解學習及測驗異常的學生資料，作為擬定學習輔導之參考。

- (2) 測驗結果資料庫：透過學生答卷資訊重新計算試題難度、試題鑑別度，將結果存回資料庫中。

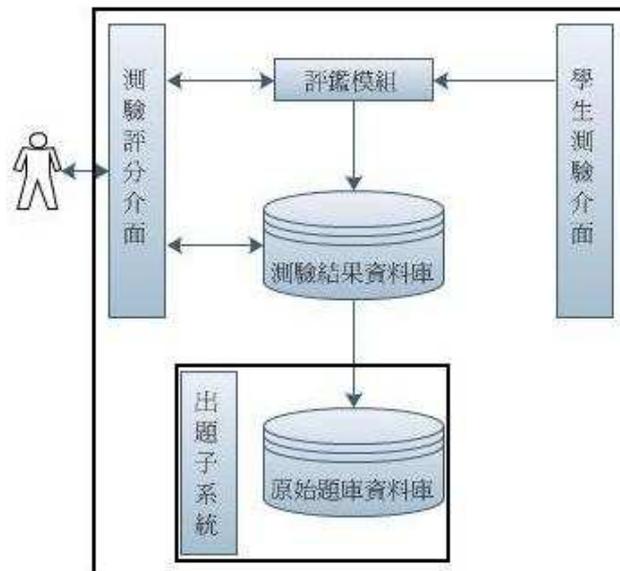


圖 4 線上測驗與評分之系統架構圖

## 4. 系統成效評估

說明：

### 4.1 雛形系統功能說明

以下僅針對主要出題功能之介面進行

- (1) 授課課程出題介面：教師選擇欲出題之課程後，進入系統之試卷產製頁面如圖

5所示，教師可以設定測驗名稱、章節範圍、出題題數、適應函數之參數設定以及Bloom比例分佈之進位方式，而系統將顯示選定章節範圍中題庫試題之

布魯姆分佈題數以及系統計算後最佳的布魯姆題數分佈表，供教師作為出題的參考。而最佳試卷試題組合如圖6所示：

授課課程選擇

應鳴雄老師，您教授2門課：

企業資源規劃

程式設計(一)

---

設定測驗

測驗名稱	GA測驗
章節範圍	3 - 4
出題題數	20
參數設定	適性函數=1 *難度函數+ 1 *鑑別函數+ 1 *相關函數+ 1 *布魯姆函數。
Bm進位方式	<input type="radio"/> 無條件進位 <input checked="" type="radio"/> 四捨五入
<input type="button" value="送出"/>	

題庫布魯姆分布					
知識\認知	記憶	了解	應用	分析	評鑑
事實知識	26	25	0	8	0
概念知識	0	1	0	1	0
程序知識	0	6	0	0	0
總題數為 :67					

布魯姆題數分布					
知識\認知	記憶	了解	應用	分析	評鑑
事實知識	8	7	0	2	0
概念知識	0	1	0	1	0
程序知識	0	2	0	0	0
總題數為 :21					

四捨五入結果					
知識\認知	記憶	了解	應用	分析	評鑑
事實知識	7	7	0	2	0
概念知識	0	1	0	1	0
程序知識	0	2	0	0	0
總題數為 :20					

圖 5 授課課程出題介面圖

第1001代				
組	染色體	適性值	機率	累進機率
1	00000100000010000000000000010001101111000000000110000010111111101	3.51783118502414	0.0446695528246459	0.0446695528246459
2	00000100000010000000000000010001101111000000000110000010111111101	3.51431960083777	0.044426560569495	0.089096113394140
3	000000011001000100000110010100000011100000000011000001000101101	3.14188664883229	0.0186551941157359	0.10775130750987
4	00100000100010000000100000000001001111100000100011000011011001001	2.97347303226698	0.00700142125558458	0.11475272876546
5	00000010000100000000000000010001101111000000000110000010111111101	3.31213004878241	0.0304355831033997	0.14518831186886
6	00000100000010000000000000010001101111000000000110000101011111101	3.39618870888228	0.0362522180745552	0.18144052994341
7	00000100000100000000000000011000101111000000000110000010111111101	3.46347799228529	0.0409084569268152	0.22234898687023
8	00000100000100000000000000010001101111000000000110000010111111101	3.51783118502414	0.0446695528246459	0.26701853969487
9	1000010000100001000000100101000010111100000001011000000100011101	2.87229252283882	0	0.26701853969487
10	000111010100100000000000000000010011101000000000010001000111011101	2.95474808549018	0.00570570490553136	0.27272424460040
11	00000100000100000000000000011000101111000000000110000010111111101	3.46347799228529	0.0409084569268152	0.31363270152722
12	0000010000001000000000011000100010011100000000011000001011101101	3.308959327989	0.0302161776854595	0.34384887921268
13	00000100000010000000000000011000101111000000000110000010111111101	3.46403697674114	0.0409471371595283	0.38479601637221
14	00000100000100000000000000010001101111000000000110000010111111101	3.51783118502414	0.0446695528246459	0.42946556919685
15	000001000000100000000011000100010011100000000011000001011101101	3.308959327989	0.0302161776854595	0.45968174688231
16	00000100000100000000000000010001101111000000000110000010111111101	3.51783118502414	0.0446695528246459	0.50435129970696
17	00000100000100000000000000011000101111000000000110000010111111101	3.46347799228529	0.0409084569268152	0.54525975663377
18	00000100000100000000000000010001101111000000000110000010111111101	3.51783118502414	0.0446695528246459	0.58992930945842
19	00000100000100000000000000010001101111000000000110000010111111101	3.51783118502414	0.0446695528246459	0.63459886228307
20	00000100000010000000000000010001101111000000000110000010111111101	3.51431960083777	0.044426560569495	0.67902542285256

圖 6 第 1000 代最佳試卷試題組合示意圖

## 4.2 電腦隨機出題與電腦策略出題的布魯姆分布比例比較

本計畫持續改善「出題子系統」的出題規則，以提升電腦輔助產製試題的品質，並與電腦隨機出題之布魯姆分佈比例進行比較：

### 4.2.1 電腦隨機出題實驗

本計畫透過系統之隨機出題機制產生一張包含 20 題試題之試卷，其布魯姆分佈比例如表 2 所示：

表 2、電腦隨機出題布魯姆分佈圖

電腦隨機出題之布魯姆分佈					
認知歷程 知識向度	記憶	了解	應用	分析	評鑑
事實知識	7	8	0	2	0
概念知識	0	1	0	1	0
程序知識	0	1	0	0	0
總題數：20					

### 4.2.2 電腦策略出題實驗

本計畫透過系統之策略出題機制產生一張包含 20 題試題之試卷，其布魯姆分佈比例如表 3 所示：

表 3、電腦策略出題布魯姆分佈圖

電腦策略出題之布魯姆分佈					
認知歷程 知識向度	記憶	了解	應用	分析	評鑑
事實知識	8	7	0	2	0
概念知識	0	0	0	1	0
程序知識	0	2	0	0	0
總題數：20					

## 4.3 電腦隨機出題與電腦策略出題的試題品質比較

為了考量教學評量的目的與價值，郭生玉（1988）認為試題品質可透過試題之量化分析如試題難度與鑑別度的評量，因此本計畫將針對電腦隨機出題及電腦策略出題的試題品質進行量化分析檢驗，目的在於探討本計畫的系統出題品質，在探討難度時，本計畫採用 Ebel 與 Frisbie（1991）、余民寧（1997）等人提出之試題評鑑原則，個別試題的難度應介於 0.4 至 0.8 間，整張試卷之試題平均難度應該接近 0.5，而在鑑別度部份，本計畫採用 Ebel

與 Frisbie（1991）、余民寧（1997）等人提出之試題評鑑準則，鑑別度大於 0.4 是屬於非常優良試題；0.3 至 0.39 屬於優良試題；0.2 至 0.29 屬於尚可試題；0.2 以下則屬於劣質試題，而本計畫之電腦隨機出題與電腦策略出題之試卷難度與鑑別度計算結果，彙整於表 4 至表九：

表 4、隨機出題與策略出題之試卷難度

題型	試題產製方式	試題品質-難度			平均難度
		易	適中	難	
選擇題 (20 題)	電腦隨機	2 (10.0%)	18 (90.0%)	0 (0.0%)	0.689 (適中)
	電腦策略	0 (0.0%)	16 (80.0%)	4 (20.0%)	0.502 (適中)

表 5、隨機出題與策略出題之試卷鑑別度

題型	試題產製方式	試題品質-鑑別度				平均鑑別度
		劣	尚可	優	極優	
選擇題 (20 題)	電腦隨機	3 (15.0%)	3 (15.0%)	4 (20.0%)	10 (50.0%)	0.426
	電腦策略	2 (10.0%)	1 (5.0%)	1 (5.0%)	16 (80.0%)	0.646

表 6 以知識向度為基礎來分析隨機出題與策略出題之試題品質數量分佈，表 6 以知識向度為基礎來分析隨機出題與策略出題之試題品質數量分佈，電腦隨機出題或是電腦策略出題在事實知識及概念知識二個向度的平均難度大都介於「適中」的範圍，但電腦策略出題之試題平均難度較接近一般期望平均難度 0.5，因此可以判斷電腦策略出題在這二個向度上優於電腦隨機出題，在程序知識向度上，電腦策略出題因為布魯姆分類比例條件限制下必須出二題，因此平均難度偏難；而在試題平均鑑別度方面，無論電腦隨機出題或是電腦策略出題之試題平均鑑別度均屬於「極優」，但整體而言電腦策略出題在三個向度的試題平均鑑別度表現上還是優於電腦隨機出題。

表 7 以認知歷程向度為基礎來分析隨機出題與策略出題之試題品質數量分佈，電腦隨機出題或是電腦策略出題在應用及評鑑二個認知歷程向度的平均難度因為沒出題而無法進行分析外，電腦策略出題在記憶、了解及分析三個向度的平均難度都較電腦隨機出題之數據更接近一般期望平均難度 0.5，因此可以判斷電腦策略出題在這三個向度上優於電腦隨機出題，而在試

題平均鑑別度方面，電腦隨機出題之試題平均鑑別度在分析向度上屬於「優」，遜於電腦策略出題在此向度獲得「極優」的評價，其餘向度無論電腦隨機出題或是電腦策略出題之試題平均鑑別度均屬於「極

優」，但各項度平均試題鑑別度值均優於電腦隨機出題，因此綜合而言電腦策略出題在各個向度的試題平均鑑別度表現上還是優於電腦隨機出題。

表 6 隨機出題與策略出題之試題品質數量分佈(以知識向度為基礎)

知識向度	試題產製方式	試題品質-難度			平均難度	試題品質-鑑別力				平均鑑別力
		易	適中	難		劣	尚可	優	極優	
事實知識	電腦隨機	1 (5.9%)	16 (94.1%)	0 (0.0%)	0.689 (適中)	2 (11.8%)	3 (17.6%)	4 (23.6%)	8 (47.0%)	0.42 (極優)
	電腦策略	0 (0.0%)	15 (88.2%)	2 (11.8%)	0.522 (適中)	1 (5.9%)	1 (5.9%)	1 (5.9%)	14 (82.3%)	0.666 (極優)
概念知識	電腦隨機	1 (50.0%)	1 (50.0%)	0 (0.0%)	0.75 (適中)	1 (50.0%)	0 (0.0%)	0 (0.0%)	1 (50.0%)	0.432 (極優)
	電腦策略	0 (0.0%)	1 (100.0%)	0 (0.0%)	0.68 (適中)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (100.0%)	0.69 (極優)
程序知識	電腦隨機	0 (0.0%)	1 (100.0%)	0 (0.0%)	0.571 (適中)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (100.0%)	0.405 (極優)
	電腦策略	0 (0.0%)	0 (0.0%)	2 (100.0%)	0.236 (難)	1 (50.0%)	0 (0.0%)	0 (0.0%)	1 (50.0%)	0.461 (極優)

表 7 隨機出題與策略出題之試題品質數量分佈(以認知歷程向度為基礎)

認知領域	試題產製方式	試題品質-難度			平均難度	試題品質-鑑別力				平均鑑別力
		易	適中	難		劣	尚可	優	極優	
記憶	電腦隨機	0 (0.0%)	7 (100.0%)	0 (0.0%)	0.68 (適中)	0 (0.0%)	0 (0.0%)	4 (57.1%)	3 (42.9%)	0.471 (極優)
	電腦策略	0 (0.0%)	6 (75.0%)	2 (25.0%)	0.492 (適中)	1 (12.5%)	0 (0.0%)	1 (12.5%)	6 (75.0%)	0.627 (極優)
了解	電腦隨機	1 (10.0%)	9 (90%)	0 (0.0%)	0.67 (適中)	2 (20.0%)	2 (20.0%)	0 (0.0%)	6 (60%)	0.423 (極優)
	電腦策略	0 (0.0%)	7 (77.7%)	2 (22.3%)	0.494 (適中)	1 (11.1%)	1 (11.1%)	0 (0.0%)	7 (77.8%)	0.653 (極優)
應用	電腦隨機	0 (0.0%)	0 (0.0%)	0 (0.0%)	NA	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	NA
	電腦策略	0 (0.0%)	0 (0.0%)	0 (0.0%)	NA	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	NA
分析	電腦隨機	1 (33.3%)	2 (66.7%)	0 (0.0%)	0.771 (適中)	1 (33.3%)	1 (33.3%)	0 (0.0%)	1 (33.3%)	0.329 (優)
	電腦策略	0 (0.0%)	3 (100.0%)	0 (0.0%)	0.549 (適中)	0 (0.0%)	0 (0.0%)	0 (0.0%)	3 (100.0%)	0.68 (極優)
評鑑	電腦隨機	0 (0.0%)	0 (0.0%)	0 (0.0%)	NA	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	NA
	電腦策略	0 (0.0%)	0 (0.0%)	0 (0.0%)	NA	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	NA

表八、隨機出題個別試題分析表

題號	難度	鑑別度	布魯姆分類
159	0.656	0.406	事實了解
144	0.7	0.666	事實了解
208	0.666	0.666	事實了解
133	0.684	0.264	事實了解
118	0.548	0.083	事實了解
179	0.781	0.2	事實分析
164	0.75	0.333	事實記憶
128	0.69	0.554	事實了解
139	0.651	0.569	事實記憶
134	0.844	0.098	事實分析
655	0.812	0.174	概念了解
116	0.648	0.364	事實記憶
130	0.67	0.313	事實記憶
147	0.666	0.666	事實記憶
169	0.754	0.261	事實了解
190	0.687	0.69	概念分析
200	0.571	0.405	程序了解
140	0.72	0.383	事實記憶
150	0.666	0.666	事實記憶
203	0.625	0.75	事實了解
平均	0.689	0.426	

表九、策略出題個別試題分析表

題號	難度	鑑別度	布魯姆分類
119	0.493	0.625	事實了解
126	0.098	0.066	程序了解
142	0.441	0.133	事實記憶
143	0.593	0.6	事實記憶
147	0.666	0.666	事實記憶
148	0.5	0.666	事實了解
150	0.666	0.666	事實記憶
151	0.36	0.35	事實記憶
153	0.75	1	事實了解
156	0.461	0.6	事實記憶
190	0.687	0.69	概念分析
191	0.5	0.75	事實分析
197	0.444	0.25	事實了解
203	0.625	0.75	事實了解
206	0.666	0.666	事實了解
209	0.375	0.857	程序了解
211	0.5	1	事實了解
233	0.461	0.6	事實分析
654	0.25	1	事實記憶
663	0.5	1	事實記憶
平均	0.502	0.646	

由表八分析此份試卷難度值最高為題庫第134題，達到0.844，因此可以得知此題屬於非常簡單的試題，在電腦隨機出題機制下只有二題是屬於簡單之試題，其餘均為難易度適中之試題。

由表九分析此份試卷鑑別度值最高為題庫第153題、211題、654題、663題，達到1，因此可以得知這些試題屬於非常具有鑑別度的試題，在電腦策略出題機制下只有二題是屬於鑑別度不佳之試題，分別是第126題與142題。另將本系統電腦策略出題初步實驗數據列出如下表10：

表10 電腦策略出題初步實驗數據表

題庫平均難度	0.6505	題庫平均難度標準差	0.161
題庫平均鑑別度	0.4231	題庫平均鑑別度標準差	0.3103
題庫平均相關係數	0.4564	題庫平均相關係數標準差	0.3232

適應函數值為： $Fitness = 3.5013$

## 5. 結論與建議

本計畫針對試題難度、試題鑑別度、試題相關係數以及布魯姆認知分類四個項度，提出一個產製高品質試卷之選題策略，搭配線上測驗平台的技術做整合，進行了更完整的自動化測驗評量系統，此外本計畫初步也針對電腦隨機出題與電腦策略出題所產生之試卷試題的試題品質做檢驗。但是，如果要對於試題品質進行更客觀的評估，則需要另外針對大量的試題進行額外的研究，取得每個試題的難易度與鑑別度結果來加以分析。此外，出題子系統的系統成效評估分析結果顯示，在試題難易度方面，無論是以知識向度或認知向度來進行分析，電腦策略出題與電腦隨機出題的平均試題難度，

在布魯姆認知分類方面，利用電腦策

略出題產製出用以評量「事實知識」、「概念知識」及「程序知識」等各種知識類型的試題，以及產製出用以評量認知歷程的「記憶」、「了解」、「應用」、「分析」及「評鑑」等類型的試題，此結果相較於電腦隨機出題分部較為均勻，顯示出本計畫較能夠進行涵蓋 Bloom 分類概念資訊的試題發展，評量學生不同認知層次的能力，且避免同一試卷中出現過多評量相同知識概念的重複試題，以及達到試卷產製的時間效益，雖然本計畫題庫在較低層次認知的試題數量上較為充足，但教師透過電腦策略出題所節省下來的時間與精神將可以用來進行更高層次的試題發展。

### 5.1 研究貢獻

本計畫藉由網際網路及電腦之運算能力，使線上測驗系統能發揮最大效益，並能對電子化學習系統及教育領域有所貢獻，本計畫之貢獻歸納如下：

- (1) 研究結果顯示，透過本計畫的出題子系統所產製的測驗試卷，其試卷平均難易度相較於電腦隨機出題產製的試卷更能夠接近難易度適當的範圍；而在試卷平均鑑別度方面，電腦策略出題產製之試卷平均而言優於電腦隨機產製之試卷。
- (2) 研究結果針對平均相關係數方面，電腦策略出題產製之試卷平均相關係數將盡量確保試題最大的獨立性，初步可以將試題關聯較高之試題組合排除出現在同一份試卷上，後續將利用試題分析來針對試題之間相關係數進行分析與探討。
- (3) 本計畫的雛形系統在初期研究成果中，已可獲得許多輔助出題的效益，除了考量試卷整體的平均難度、平均鑑別度以及平均相關係數之外，也能涵蓋評量事實、概念及程序等三種知識向度及記憶、了解、應用、分析及評鑑等五種認知向度的試題，做試卷試題的最佳分配。就教育領域而言，本計畫的貢獻主要在於結合了布魯姆教育目標分類概念，確保產生的試卷涵蓋不同認知層次的試題。

### 5.2 研究限制與未來研究建議

由於時間及教學環境上的因素，本計畫之系統仍有一些限制，茲條列如下，並討論其改進的方向。

- (1) 由於本計畫在資料數據的獲得上必須配合學期課程進度安排而定，例如試題之難度與鑑別度之數據必須等到該測驗範圍隨堂測試結束後才能獲得，而在試題相關係數方面也必須配合試卷出題範圍而定，因此若第一次測驗為一、二章，第二次測驗為三、四章，則僅能針對個別範圍進行策略出題。
- (2) 本計畫之試卷產製之後，目前僅能透過教師來判斷試題內容，未來希望增加語意分析及自然語言技術來進行電腦自動化之試題分析，透過人工智慧使題庫試題更臻完善。
- (3) 由於本計畫目前著重系統架構研究及初期成效之評估，因此未來實證階段將以課程領域專業教師為主。未來研究則可邀請教育專家檢視電腦策略出題與電腦隨機出題在試題內容上以及學生測驗上進行比較分析，透過不同出題機制之測驗事件將班級學生對於二份試卷之測驗表現與以分析，針對相關係數做一個廣泛的解釋，驗證電腦策略出題之成效。

### 5.3 計畫成果自評

- (1) 電腦隨機出題相較於電腦策略出題的效益自評

經過電腦策略出題產製試卷方法的設計及雛形實驗系統的建置後，相較於電腦隨機出題，本計畫的電腦策略出題產製試卷的效益說明如下：

- A. 本計畫中電腦策略出題產製的試題涵蓋 Bloom 認知分類的各個向度資訊，使得試題評量更加富有教育意義。而電腦隨機出題通常未考慮試題可評量的向度，並不足夠來衡量學生在各種認知層次上的學習理解成效。
- B. 本計畫藉由提出一個有效的高品質試卷出題策略系統，評量電腦選

題的試題組合是否能具備適當的深度及廣度、試題間最大的獨立性等高等品質測驗試題的衡量指標，期望能夠準確公平地衡量學生的學習成效，並且作為後續教學上之參考。

## (2) 研究發表

本計畫成果已發表於第二屆資訊教育與科技應用研討會，目前將針對研究成果持續修改，並將論文再投稿至期刊發表。

## 6. 參考文獻

- 【1】 余民寧(1991)，試題反應理論的介紹(一)--測驗理論的發展趨勢，研習資訊，8卷第6期，13-18頁。
- 【2】 余民寧(1997)，教育測驗與評量：成就測驗與教學評量，心理出版社。
- 【3】 何榮桂(1991)，電腦教學系統中的測驗設計，中等教育，41卷2期，29-34頁。
- 【4】 郭生玉(2001)，心理與教育測驗(十五版)，台北中和：精華。
- 【5】 葉連祺、林淑萍(2003)，布魯姆認知領域教育部標分類修訂版之探討，教育研究月刊，第105期：94~106。
- 【6】 Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (2001). A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives. New York: Longman.
- 【7】 Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., and Krathwohl, D. R. (1956), A Taxonomy of Educational Objectives: Handbook 1, The Cognitive Domain, N.Y., David McKay Co.
- 【8】 Brueckner, L. J. & Band, G. L. (1955), The diagnosis and treatment of learning difficulties, NY: John-Wiley.
- 【9】 Chen, Y. P. (1999), Extending the Scalability of Linkage Learning Genetic Algorithms, Baker & Tayl.
- 【10】 Ebel, R. L. (1967), The relation of item discrimination to test reliability. *Journal of Educational Measurement*, 4(3), pp.125-128.
- 【11】 Ebel, R. L. & Frisbie, D. A. (1991). *Essentials of educational measurement, 5th ed.*, Englewood Cliffs, NJ: Prentice-Hall.
- 【12】 Gulliksen, H. (1987), *Theory of mental test*, Hillsdale, NJ: Lawrence Erlbaum Associates.
- 【13】 Holland, J. H.(1975), *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, University of Michigan Press.
- 【14】 Krathwohl, D. R. (2002), A revision of Bloom's taxonomy: An overview, *Theory Into Practice*, 41(4), 212-219.
- 【15】 Lord, F. M., & Novick, M. R. (1968), *Statistical theories of mental test scores*, Reading, MA: Addison-Wesley.
- 【16】 Lord, F. M. (1980), *Applications of item response theory to practical testing problems*, Hillsdale, NJ: Lawrence Erlbaum Associates.