

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
* : 以基因演算法為基礎之多目標試卷選題策略研究
* 名 稱
* *****

執行計畫學生： 林仁祥
學生計畫編號： NSC 99-2815-C-216-010-S
研究期間： 99年07月01日至100年02月28日止，計8個月
指導教授： 應鳴雄

處理方式： 本計畫可公開查詢

執行單位： 中華大學資訊管理學系

中華民國 100年03月31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 以基因演算法為基礎之多目標試卷選題策略研究

計畫編號：NSC99-2815-C-216-010-S

執行期限：99年8月1日至100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：林仁祥

1. 中文摘要

網際網路的盛行使得網路上的教育學習受到重視，越來越多的學習者藉由遠距教學(Distance Education)的環境來學習知識，提升自我價值。然而為了瞭解遠距教學平台學習者的學習成效，線上測驗平台成為遠距教學系統的重要功能。大多數的線上測驗系統在選題時，往往會以事前編製的題庫為基礎，而且試題的選擇並未以多目標之觀點進行篩選。因此本計劃將採用試題的難度、鑑別度、Bloom 認知分類以及曝光率等四項目標指標，做為衡量試題品質及試卷選題的依據。

本計劃期望藉由 Bloom 認知分類修正版(Revision of Bloom's taxonomy)、測驗理論(Test Theory)、基因演算法(Genetic Algorithms)等概念，提出一個多目標的選題策略。在選題策略中透過難度、鑑別度、Bloom 認知分類以及試題曝光率等參數，使試題組合能涵蓋不同的認知層次且具有適合的難度和鑑別度並有效抑制曝光率，以保證試卷的品質。

關鍵字：測驗理論，基因演算法，Bloom 認知分類

2. 計畫緣由與目的

網路的迅速性與即時性，使人們在經濟、生活上都越來越便利，使用科技來輔助學習也已經成為目前的重要趨勢。

數位學習是由 Jay Cross 在 1999 年所提出來的名詞，美國教育訓練發展協會(ASTD, American Society of Training and

Education)將數位學習定義為「利用數位媒介來學習之過程就是數位學習。」。數位學習即是架構在網路技術上的一種學習方式，它可以讓學習者不論身在何地、任何時間都可以學習[3]。它的設計觀點是以學習者為中心導向，它具有互動性、可重複學習性，可客製化等特性。數位學習具有許多優點，例如：使用者可以不受時間、地點的限制，能在任何時間與場域進行學習，而數位學習系統的教材與學習資訊也能即時更新與傳送，並能改善傳統教育上人力不足方面的問題。因此，數位學習(E-learning)與線上測驗(OnLine Test, OLT)在近幾年來越來越普及，也越來越受重視。

測驗可以瞭解學生的學習成效與增進其學習成果[1]，但是多數的線上測驗系統是事先將已經編製完成的試題放置到測驗系統的題庫中，再採用隨機選題或適性選題的策略來挑選考卷中的試題，在這些挑選題目的過程中，通常都以古典測驗理論所使用的測驗指標[4]，試題難度與鑑別度為主要依據。雖然此兩種測驗指標被認為可以協助教師挑選優良的試題[5]，提升教學評量效果，但仍未加入試題的 bloom 認知層次及試題曝光率。所謂的試題曝光率(item exposure rate)是指某試題出現次數與所有參與測驗受測者人數之間的比率。曝光率過高的問題通常會在測驗題庫過小或是適性測驗上出現[6]。因此，當曝光度過高也可能導致測驗成效或試卷品質受到影響，甚至也有可能影響試題的難度與鑑別度，

Stocking 認為試題當題庫中的試題

數量太少，或是某些試題過度曝光，容易引發受測者記憶試題的情形，進而影響了試卷的品質及試卷的評分效力，因此應將曝光率列入選題時的重要指標[2]。

試題的 Bloom 認知層次則是具有區別分段能力的功用，Anderson 在 2001 年重新把認知領域修改為兩個向度為認知歷程向度與知識歷程向度[7]。認知歷程包括六個層次，分別為記憶、瞭解、應用、分析、評鑑與創造(表 1)，在知識向度則區分為事實、概念、程序、後設認知等四類知識[8,9]，詳見表 2。所以，本計劃認為一份高品質的試卷，應具有以上四個試題指標。

基因演算法 (Genetic Algorithms) 是遵循達爾文所提出的物競天擇，自然進化的法則，並能在有限時間內找到近似最佳解的一種演算法。此類演算法是透過操作染色體的方式進行演化，並在演化期間內經由交配、複製與突變等運算，再透過適應函數(fitness function)的設計，選出適合環境的子代繼續繁衍，以獲得較適合環境的物種。一般而言，基因演算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[10]。因此，本計畫嘗試用測驗評量理論、Bloom 認知分類、基因演算法等概念，產生具備高品質的選題策略，並於選題策略中加入試題曝光度參數，使題目組合在恰當題目曝光率的前提下，能涵蓋各種認知層次試題，且能讓試題能具備適當的難度和鑑別度，使本計劃的選題策略以達到多目標發展的主要目的。

本計畫共計八個月，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。

表 1. 認知歷程向度

認知歷程向度	解釋
Remember	從長期記憶中取得有關知識
Understand	從口述、或圖像等教學資訊中建構有意義的知識
Apply	運用程序解決問題的能力
Analyze	將整體結構拆解為數個部份，並且了解其中關係
Evaluate	根據規則或標準下判斷的能力
Create	重組要素形成新結構或模型

表 2. 知識歷程向度

知識歷程向度	解釋
Factual	學習科目後和解決問題應知的基本元素
Conceptual	要素與較大的結構共同發揮功能的互動關係
Procedural	指如何以一系列或有步驟的流程以完成事情
Meta-cognitive	指一般認知與對自我的認知

3. 研究架構與系統建置

3.1 研究方法

教師會希望題庫中的個別試題曝光度能夠均衡，而不希望某些試題的曝光度太高或太低，以免試題因為過度曝光及未被使用，而導致難度與鑑別度受到影響。因此，在電腦進行試卷選題的過程中，有必要將試題的難度、鑑別度、曝光度，以及試題分佈在 Bloom 認知層次的均衡性都一併列為選取試題的指標。本計劃基於此想法，結合資訊技術及試題理論概念，在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。

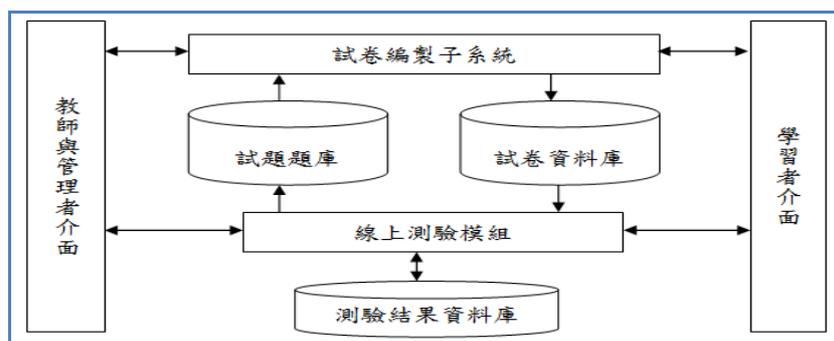


圖 1 系統架構

3.2 系統架構設計

此階段將實作發展雛形系統，主要子系統為「試卷編製子系統」，而操作介面則分為「教師與管理者介面」與「學習者介面」，如圖 1 所示：

a. 教師與管理者介面：

1) 題庫資訊查詢：

顯示目前題庫試題狀況，提供每一章各類型題數的數量資訊，讓命題者新增試題時，能知道哪一章節的題目數量較少，必須針對這些章節特別新增題目，使題庫有足夠的題目來進行測驗。

2) 試題新增：

基本編輯試題屬性的功能，提供了各類型的試題資訊，例如試題編號、試題所屬章節、試題難度、試題鑑別度、Bloom 認知分類、試題內容、試題選項以及試題答案等等。

3) 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 Bloom 認知分類以及試題章節等，提供命題者執行試題的重新編輯或刪除不恰當試題。

b. 學習者介面：

1) 註冊會員：學生透過申請會員功能，輸入個人資料與帳號密碼後，即可登入本計劃線上測驗平台。

2) 線上測驗：學生在測驗發佈公告

後，於測驗開放時間登入系統進行線上測驗，系統會根據答卷結果計算試題難度，試題鑑別度。

3) 測驗結果查詢：學生在測驗結束後可以查閱電子試卷與成績，系統會標記答對與答錯之題目，並統計相關認知層次試題的答題狀況。

c. 試卷編製子系統：

透過基因演算法進行策略選題可以符合教學者所設定之試卷選題需求，而且也大大縮短了編製測驗試卷的時間，圖 2 詳細介紹試卷編製子系統中模組的功能：

1) 試題編修模組：試題編修模組的主要功能為提供一個方便使用的視窗話介面，包含新增與編修兩大功能，新增試題可以設定 Bloom 分類向度、章節範圍、試題難度、試題鑑別度、試題內容與備選答案等，試題編修則可以讓使用者依照類別與關鍵字進行試題查詢，並且對上列試題資料進行編輯與修改。

2) 試卷選題模組：試卷選題模組主要是提供一有效的策略選題演算法（詳見染色體與適性函數設計），利用此演算法根據適應函數多項指標來決定試卷之試題組合，透過快速且自動運算求出最佳化的試卷試題配置，並利用所產生之試卷讓學習者進行測驗，以評估受測者的學習程度，本計劃之試卷選題模組除了策略選題演算法之外，另外也能使用隨機選題模式來編製測驗試卷。

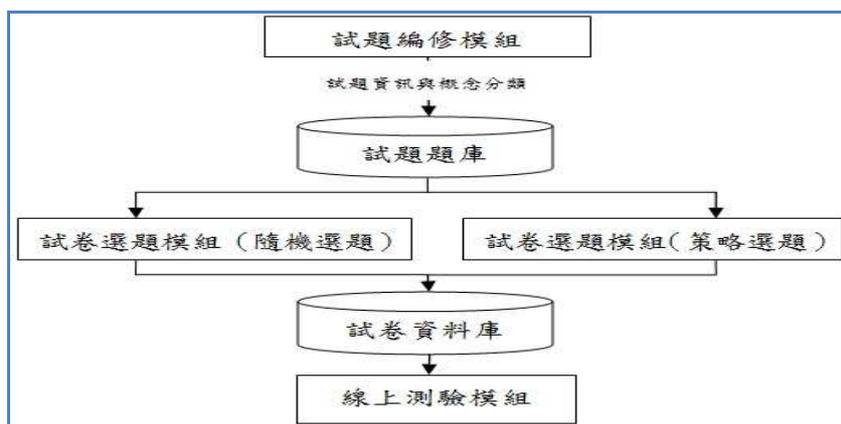


圖 2 試卷編製子系統架構

- 3) 試題題庫：存放建立之題庫試題資訊，包含試題內容、試題選項、試題難度、試題鑑別度、Bloom 認知分類向度以及試題章節等。
- 4) 試卷資料庫：存放經過隨機出題或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。

3.3 研究步驟

3.3.1 題庫建立與資料分析

為了取得試題相關資料，本計劃在進行現階段實驗前一年，先建置了一個線上測驗系統，並在系統內建立「ERP 軟體應用」課程教材與 13 個章節的題庫，題目合計 313 題。題庫中的每個試題均請多位專家依據其題意、內容，將題目進行 Bloom 認知層次的分類，然後讓某大學資管系 54 名學生於課堂考試時進入線上系統進行測驗，試卷題目採隨機選題，因此每位學生之試卷試題均為不同，再經由測驗系統的自動評分與成績計算，取得每個題目的鑑別度、難度及曝光率資料，並加以儲存。

本計劃試圖將試題曝光率、試題難度、試題鑑別度及試題所屬之 Bloom 分類分析後的資料，當作系統選題策略之

主要參數，由系統自動製作一份最佳試題組合的試卷。本計劃主要的試題相關參數公式如下所述。

- (1) 試題難度：依照答對的百分比來表示，數值越高表示越容易，一份試卷的平均難度接近 0.5 為佳，公式如(1)所示。Dif 代表該試題難度，N 代表該試題受測人數。

$$D i f = C / N \quad (1)$$

- (2) 試題鑑別度：學生測驗分數經過排序後，本計劃設定之分組選取高、低分群門檻為前 33% 與後 33%，試題鑑別度公式如(2)所示。PH 為高分群答對人數百分比，PL 為低分群答對人數百分比，DI 表示該題鑑別度，若整體試題鑑別度越高，則測驗信度就越高[11]。

$$D I = P H - P L \quad (2)$$

- (3) 試題曝光率：試題曝光率為測驗中試題所使用的相對次數[6]。曝光度公式如(3)所示。公式中的 i 表第幾題試題，Epr(i) 表示第 i 試題之曝光率，Ti 表示第 i 題被受測次數，N 表示參與該次測驗的人數。

$$E p r (i) = T i / N \quad (3)$$

3.3.2 染色體與適性函數設計

本計劃之染色體設計採二進位編碼的方式，1 代表該試題被選取到，0 代表該題將未被選取。若題庫中的試題數共為 10 題，以 $X_1 \sim X_{10}$ 表示，若要產生 5

個試題的試卷，則其染色體就如圖 3 所示。

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
0	1	1	1	0	1	0	0	0	1

圖 3 染色體設計

本計劃的試卷品質適性函數是由測驗理論中的難度與鑑別度指標，再配合試題曝光率與 Bloom 認知分類的比例共四個目標參數所組成。公式如(4)所示。

$$f(s_i(t)) = W_{dif} * Dif(s_i(t)) + W_{dis} * Dis(s_i(t)) + W_{epr} * Epr(s_i(t)) + W_{bm} * Bm(s_i(t)) \quad (4)$$

t 表示基因演算法的世代數，S(t) 表示第 t 代的母體，si(t) 是 S(t) 中第 i 種試題組合的成員，f(si(t)) 表示第 t 代的第 i 個試題組合的試卷品質適切值，其中 Dif(si(t)) 代表第 si(t) 個試卷組合的試題難度適切值，Dis(si(t)) 代表第 si(t) 個試卷組合的試題鑑別度適切值，Epr(si(t)) 表示第 si(t) 個試卷組合的試題曝光度適切值，Bm(si(t)) 表示第 si(t) 個試卷組合的試題 Bloom 認知分類題型比例適切值。Wdif、Wdis、Wepr 及 Wbm 分別是上述四項適切值的加權值，教師可以在實際應用上依對各適切值的偏好與重視度來設定其加權比例。

a. 難度 Dif :

Dif(s_i(t)) 表示為第 s_i(t) 個試卷組合的難度適切值，當 Dif(s_i(t)) 越高表示此試題組合越佳。在計算難度適切值 Dif(s_i(t)) 時，會經過兩個公式進行試題難度的數值轉換。公式(5)是計算第 s_i(t) 個試卷組合的平均難度 Adif，其中 q 表測驗題數，n_u 表示 s_i(t) 種試卷組合的第 u 個試題是否被選取，n_u=0,1，dif(u) 表示 s_i(t) 種試卷組合的第 u 個試題的難度。

$$ADif(s_i(t)) = \left\{ \sum_{u=1}^q [n_u * dif(u)] \right\} / q \quad (5)$$

計算出平均難度 Adif 後，就可依照公式(6)求出 Dif(s_i(t)) 的難度適切值，EkDif 為教師期望的試卷平均難度，而 Difsdv 為題庫的難度標準差。

$$Dif(s_i(t)) = 1 - |(ADif(s_i(t)) - EkDif) / Difsdv| \quad (6)$$

b. 鑑別度 Dis :

Dis(s_i(t)) 表示為第 s_i(t) 個試卷組合的鑑別度適切值，當 Dis(s_i(t)) 越高表示此試題組合越佳。在計算鑑別度適切值

Dis(s_i(t)) 時，會先透過公式(7)計算第 s_i(t) 種試卷組合的平均鑑別度 Adis，其中 q 表測驗題數，n_u 表示 s_i(t) 種試卷組合的第 u 個試題是否被選取，n_u=0,1，dis(u) 表示 s_i(t) 種試卷組合的第 u 個試題的鑑別度。

$$ADis(s_i(t)) = \left\{ \sum_{u=1}^q [n_u * dis(u)] \right\} / q \quad (7)$$

計算平均鑑別度 Adis 後，再依公式(8)求出 Dis(s_i(t)) 的難度適切值，式子中的 Dissdv 為題庫的鑑別度標準差，而 Disavg 為題庫平均鑑別度。

$$Dis(s_i(t)) = \frac{((ADis(s_i(t)) + Dissdv) - Disavg)}{2 * Dissdv} \quad (8)$$

c. 曝光率 Epr :

Epr(s_i(t)) 表示為第 s_i(t) 個試卷組合的曝光率適切值，當 Epr(s_i(t)) 越高表示此試題組合越佳。在計算曝光率適切值 Epr(s_i(t)) 時，會經過兩個公式進行試題曝光率的數值轉換。公式(9)是計算第 s_i(t) 個試卷組合的平均曝光率 AEpr，其中 q 表測驗題數，n_u 表示 s_i(t) 種試卷組合的第 u 個試題是否被選取，n_u=0,1，epr(u) 表示 s_i(t) 種試卷組合的第 u 個試題的曝光率。

$$AEpr(s_i(t)) = \left\{ \sum_{u=1}^q [n_u * epr(u)] \right\} / q \quad (9)$$

計算平均曝光率 AEpr 後，再照公式(10)求出 Epr(s_i(t)) 的曝光率適切值。

$$Epr(s_i(t)) = 1 - AEpr(s_i(t)) \quad (10)$$

d. Bloom 認知分類題型比例適切值:

本計劃主要以選擇題的方式來做測驗，但 Bloom 認知分類中的後設知識與創造層次試題，不易透過選擇題的方式來評量。所以本計劃在「知識向度」上只採用事實、概念、程序等三類知識，而在「認知歷程向度」則採用記憶、了解、應用、分析、評鑑等五個層次，共計 15 種 Bloom 分類題型。表 3 為本計劃

的選題策略範圍題庫 Bloom 認知分類的題型分布，總題目數為 70 題。

表3. Bloom認知分類的題型分布

認知層次 知識向度	a. 記憶	b. 了解	c. 應用	d. 分析	e. 評鑑
X. 事實知識	4	0	0	0	0
Y. 概念知識	4	29	2	12	0
Z. 程序知識	5	12	0	0	0

以程序知識了解層次(Zb)的題型為例，該題型題庫總題數為 12，故此 Zb 被命題的機率 $PZb = 12/70=0.171$ ，以產生一份題數 x 為 20 的試卷為例，Zb 的期望出題數 $EkZb$ 為 $20 * PZb=3.43$ ，四捨五入後 $EkZb$ 為 3，本計劃為了確保試卷的 Bloom 向度能夠涵蓋測驗範圍中所有 Bloom 類型試題，故當 $Ek > 0$ 且 $Ek < 0.5$ 時，強制期望出題數 Ek 為 1。為了確保試卷的 Bloom 向度夠廣，所以當 Ek 的總合 $> x$ 時，表示試卷總期望出題數大於試卷總出題數限制，此時系統會逐步比照 Ek 大小挑出 $Max(Ek)$ ，將 Ek 減 1，一直持續以上動作直到 Ek 總合與測驗題數 x 相等為止。若 Ek 相同，系統會優先從命題機率較大的 Ek 減 1。Bloom 認知分類題型比例適切值的公式為(11)。

$$Bm(S_i(t)) = \left[\sum_{j=1}^3 \sum_{k=1}^5 \left(1 - \left| \frac{E_{jk} - R_{jk}}{E_{jk}} \right| \right) \right] / 15 \quad (11)$$

$Bm(S_i(t))$ 表示第 $si(t)$ 個試卷組合的 Bloom 比例適切值， j 、 k 表示為 Bloom 認知向度與認知歷程， E_{jk} 表示期望出題數， R_{jk} 代表實際出題數。

上述為本計劃基因演算法各項適性函數指標的轉換，期望藉由這些指標提升試卷品質及確保評量意義。

4. 實作與績效評估

圖二的試卷品質適性函數 $f(si(t))$ 所採用的加權值分別為 $Wdif = 1$ ， $Wdis = 1$ ， $Wepr = 1$ ， $Wbm = 1$ 。在圖 4 中，第 22 組染色體之適性函數值最大，因此被選到的機率最大。機率值為 0.0552 而第 18 組染色體之適應函數值最小，因此被選到的機率值為 0，以符合進化論中的物競天擇、適者生存，不適者淘汰之概念。

接下來藉由電腦產生二個亂數值，範圍為由 0 至 1，來對應圖 4 之累進機率值，選出一對父代染色體，透過基因演化過程的複製、交配與突變流程來對選出的父代染色體進行運算而產生更具競爭力的子代。由於每一群體只有 30 組染色體，因此本計劃將交配率 CR 設為 0.975，產生更多新的染色體結構。而將突變率設定為 $MR = 0.015$ 主要是不希望產生區域最佳解的情況。

圖 5 為經過一千代演化後之染色體，試卷最佳適合度函數(Fitness)為 3.747，圖 6 可以看到該組合的難度適切值為 0.995，鑑別度適切值為 0.869，曝光率適切值為 0.884 及 Bloom 認知分類題型比例適切值，而試卷 bloom 分佈完全符合期望 bloom 分佈，故值為 1。

圖 8 為本計劃基因演算法每代適合函數最大值、最小值及平均值的變化。從圖中可以清楚了解基因演算法在各世代尋找最佳適性函數值之過程，最大值在第 550 代後呈現穩定狀態，平均適合函數值在初期很快的上升，並逐漸接近最大值，表示出每一世代都比先前的世代好，但到後期世代開始出現一些上下震盪，主要是因為群體找到穩定最大函數值後又再試圖突變尋找有無更好的染色體。

本計劃用隨機篩選試題方式產生五組試題組合，再套入本計劃所提出的試卷品質適性函數(4)，計算出適合度函數(圖7)，再與本計劃使用基因演算法所產生的試題組合之最佳適合度函數相互比較(圖6)。

本計劃採用基因演算法所產生的試題組合，難度適切值為 0.995，而在五組隨機產生的試卷組合上，只有一組隨機試卷難度適切值達到 0.945，但其他適切值皆較

差。

初代				
PSIZE	$S_i(t)$	F_t	機率	累進機率
1	11000000010100001011010000110001001010001101001000101010100000101000	2.85727393074235	0.0315876558740085	0.0315876558740085
2	110001010011100100110000010001001110000100100101001100100000100010000	3.01325781170743	0.0438486592690219	0.0754363151430304
3	1010101100110001001010011000100000000100001000101011000110000011001	3.04231497387963	0.0461326770908543	0.121568992233885
	⋮			
17	101000100001010001010101100001101010110011100001001011000110000000000	2.83617161019633	0.0299289227904575	0.554057401311079
18	100110101011100000011010010001011000000111010010000001100110000001000	2.4554173293633	0	0.554057401311079
19	100100000100000100100000011101100010010100101000001101100100010001110	3.03557804376222	0.0456031254312177	0.599660526742297
	⋮			
21	01010000110001010000011100001110100001110011011000010100000000010011	2.92875726611176	0.0372065532384573	0.662773842388573
22	0001000010010101000101000110000001010000111000101111001100100000011001	3.15742027099697	0.0551804481169879	0.717954290505561
23	1000111000101001001011000000000101010111001110100011010000000000100	2.88218776117196	0.0335459900146275	0.751500280520188
	⋮			
30	01011000100000011110010001000100101010000010111010000011011000001	2.88349328908757	0.0336486101193881	1

圖 4 初代染色體適應函數最大值與最小值之機率表

第1000代				
PSIZE	$S_i(t)$	F_t	機率	累進機率
1	0000100010010011011011010001101010000110000100110010000001000100010101	3.74726524786712	0.0372880458451597	0.0372880458451597
2	0000100010010011011011010001101010000100100000110010000001000100010111	3.74609833494434	0.0371879800999875	0.0744760259451472
3	0000100010010011011011010001101010000100100100110010000001000100010101	3.7428081634707	0.0369058395444885	0.111381865489636
4	0000100010010011011011010001011010000110000100110010000001000100010101	3.68866333987192	0.0322627834819186	0.143644648971554
5	1000010010010011011011010001101010000100100000110010000001000100110001	3.5159101969242	0.01744876296474	0.161093411936294
6	000010001000001001011101000100101010110011000110100000000000101010111	3.31243210404501	0	0.161093411936294
7	0000100010010011011011010001101010000110000000110010000001000100010111	3.73945795851949	0.0366185509659007	0.197711962902195
	⋮			
28	000010001000001001011101000100101010110011000110100000000000101010111	3.31243210404501	0	0.92995765098255
29	0000100010010011011011010001101010000100100000110010000001000100010111	3.74609833494434	0.0371879800999875	0.967145631082537
30	0000100010010011011011010001101010000110000100110010000000000101010101	3.69556209117268	0.0328543689174631	1

圖 5 第1000代染色體適應函數最大值與最小值之機率表

在鑑別度上部份，隨機試卷組合最高鑑別度適切值只有0.564，本計劃試卷組合則達到0.869。隨機產生的試卷組合在曝光率適切值部份最高僅拿到0.797，本計劃試卷組合則拿到0.884。最後在bloom適切值上本計劃藉由基因演算法所產生的試題組合完全符合期望的bloom出題比例，得到滿分，而隨機試卷最高bloom適切值卻只有0.855。

經由隨機篩選產生的試題組合與本計劃之最佳化試卷組合比較後，證明本計劃所提出的選題策略與隨機產製試題在同時考量試題難度、鑑別度、試題曝光率及Bloom認知分類題型比例的情形下，獲得較佳的試卷試題組合。

5. 結論與建議

本計劃藉由Bloom認知分類修正版、測驗理論、基因演算法等概念，提出一個高品質的選題策略。在選題策略中透過難度、鑑別度、Bloom認知分類以及試題曝光率等多目標參數，使試題組合能涵蓋不同的認知層次且具有適合的難度和鑑別度

並有效抑制曝光率，以保證試卷的品質。

經由基因演算法所產生的最佳試卷組合可以發現，本計劃選題策略在選取試卷試題上，除了能有效的產生符合教師期望的高品質試題組合，也能保證所選取的試題組合具有實質的評量意義。說明了本計劃提出之多目標發展選題策略，能夠使試卷能具備適當的廣度、深度及鑑別度之外，還能避免選取到試題曝光率高的試題，以維持整份試卷的高品質。

5.1 研究貢獻

本計劃提出一個多目標性的選題策略，除了參考古典測驗理論中所著重的試題難度、鑑別度外，更結合Anderson在2001年所提出的Bloom修正版，Bloom中的知識歷程可以協助教師去區分教些什麼(what to teach)，而認知歷程則可讓學生保留和轉移所學習的知識，同時亦考量到試題若曝光率過高時，所會產生的弊端，如學生會採取背誦的方式去記憶題目。

本計劃所提出的多目標選題策略不僅可以提升試卷之品質，還可以減輕教師

出題之負擔與人力成本，提供給教師做為 出題時的參考策略或方針。



圖6. 基因演算法產製之最佳適合度函數

第1組隨機試卷					
Ri(t)	Ft=Dif + Dis + Epr + Bm	avg_Dif	avg_Dis	avg_epr	
000000010010001000111000010100111011000110010000001110000010010010011	$3.002 = 0.838 + 0.524 + 0.797 + 0.843$	0.458	0.337	0.549	
第2組隨機試卷					
Ri(t)	Ft=Dif + Dis + Epr + Bm	avg_Dif	avg_Dis	avg_epr	
1101000000001011101110000010001100100000101000001010010100011011000010	$2.655 = 0.834 + 0.391 + 0.744 + 0.687$	0.544	0.254	0.574	
第3組隨機試卷					
Ri(t)	Ft=Dif + Dis + Epr + Bm	avg_Dif	avg_Dis	avg_epr	
010000010001111011011010001011010000100001000000001110001001100000011	$2.880 = 0.787 + 0.475 + 0.764 + 0.855$	0.556	0.307	0.565	
第4組隨機試卷					
Ri(t)	Ft=Dif + Dis + Epr + Bm	avg_Dif	avg_Dis	avg_epr	
110001000110000010011100011001001000100010000011011110000001100010	$2.691 = 0.755 + 0.409 + 0.740 + 0.787$	0.564	0.265	0.576	
第5組隨機試卷					
Ri(t)	Ft=Dif + Dis + Epr + Bm	avg_Dif	avg_Dis	avg_epr	
0101110011111100000011001001011000000100000000011010000100001101001000	$2.995 = 0.900 + 0.564 + 0.787 + 0.744$	0.526	0.363	0.554	

圖7. 隨機篩選試題組合與適合度函數

5.2 研究限制與未來研究建議

由於時間及教學環境的因素影響，本計劃雖然達到初期的目標，但尚未對多目標選題策略所產生之試卷組合進行實際的測驗。因此，本計劃下階段之工作便是將針對本計劃所選出之試卷組合進行測驗，再分析本計劃策略所產生之試卷組合在Bloom試題分佈、試題難度、鑑別度及試題曝光率等方面的成效。在未來有下列幾項研究建議

- (1) 參考過去文獻所提出其他影響試卷品質指標，修正適應函數之參數計算方式與調整權重計算的比例，或是加入新的適應函數參數例如章節重要性等。
- (2) 比較其他試卷選題演算法與本計劃提出之策略選題方法之差異，並改良試卷選題演算法之成效。
- (3) 本計劃方法仍著重於歷史資料的輔助，未來期望能發展出新的選題策略，降低對歷史資料的依賴性。
- (4) 與現有線上測驗平台系統結合，例如搭配技職體系技能檢定學科模擬測驗、企業資源規劃證照模擬等。

5.3 計畫成果自評

本計劃所提出的四個試題指標，在難度、鑑別度、曝光度以及Bloom向度各指標的相互比較下，由基因演算法所產生的最佳化試卷組合，皆優於使用亂數所產生的試題組合。證明使用基因演算法再套入本計劃所提出的選題策略將可以有效及快速的找出以符合教師需求的最佳試卷組合，提供給教師做為出題之參考。

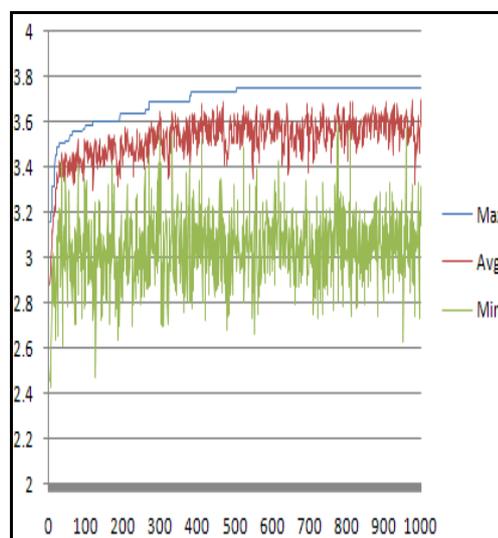


圖8 第1000代之適合函數變化圖

6. 參考文獻

- 【1】 Brueckner, L. J. & Band, G. L., 1955 “The diagnosis and treatment of learning difficulties”, NY: John-Wiley.
- 【2】 Stocking, M. L., & Lewis, C., 1995 “A new method of controlling item exposure in computerized adaptive testing, Princeton”, NJ: Educational Testing Service.
- 【3】 陳年興, 楊錦潭, “數位學習理論與實務“, 臺北縣 汐止市: 博碩文化, 民 95.
- 【4】 余民寧, “試題、分數和測驗理論“, IRT 測驗與教學, 民 94.
- 【5】 余民寧, “教育測驗與評量: 成就與學習評量“, 初版, 臺北市, 心理出版社, 民 86.
- 【6】 謝友詩、劉湘川、郭伯臣, 2006, “電腦適性測驗題目曝光率之模擬研究”。測驗統計年刊, 14 輯上, pp.59~74。台中市: 國立台中教育大學.
- 【7】 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.
- 【8】 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, pp.94-106.
- 【9】 Krathwohl, D.R., A, 2002 “revision of Bloom’s taxonomy: An overview, Theory Into Practice”, 41(4), pp.212-219.
- 【10】 Chen, Y. P., “Extending the Scalability of Linkage Learning Genetic Algorithms: Theory & Practice”, Baker & Tayl, 2004.
- 【11】 Ebel, R. L. “The relation of item discrimination to test reliability”, Journal of Educational Measurement, 4(3), pp.125-128.