

# 初始階段選題對電腦適性測驗試題曝光率之比較

錢永財 劉家惠 郭伯臣  
台中師範學院測驗統計研究所  
cdr2292002@yahoo.com.tw

## 摘要

電腦適性測驗之題庫建立不易，如不能有效控制題庫題目曝光率，當題目過度曝光，受測者則容易施測到相同的題目使得測驗的安全性與公平性產生危機。在測驗開始階段三至五題時，由於能力估計仍未精確，此時若能有效分散試題，提升題庫中低曝光率試題的使用率，並可有效降低題庫中未使用試題題數，對整體試題曝光率達到有效控制，以期達到提升題庫經濟效益。

本研究針對測驗前期三至五題採取隨機選題、McBride & Martin(1983)提出之5-4-3-2-1隨機選題法及b值分層隨機選取等方法，比較其試題曝光率控制效果，試圖找出有效控制題庫中未使用試題題數、降低試題曝光率、改善施測時前後測試題重疊率及試題重疊率的方法。

研究結果：電腦適性測驗測驗前期三至五題採取b值分層隨機選取對試題題庫題目曝光率有明顯改善。以b值分層選題法控制測驗測驗前期五題選題，六至二十題採鄰近法選題在曝光率控制上的表現最佳；若以b值分層選題法控制測驗測驗前期三題選題，四至二十題採最大訊息法選題，則能力估計誤差最小。

**關鍵詞：**電腦適性測驗、題目曝光率、題目反應理論、選題法。

## Abstract

To maintain test security and fairness of computerized adaptive testing (CAT), the exposure rate of item pool should be well controlled. The items administered to different examinees are most frequently the same in the early stage of CAT for the estimation of latent trait of examinees were not accurate enough. It can be expected to balance the exposure rate of items in the item pool and therefore efficiently usage of the item pool.

This study compared the performance of exposure control of three methods in the early stage of CAT: random selection, 5-4-3-2-1 random selection, and b-stratified random selection. The results of percentage of unused items, exposure rate, test overlap rate and item overlap rate of these methods were presented in this study.

It was found that the b-stratified random

selection method improved the exposure control rate of the item pool significantly. To use the b-stratified random selection method combined with Nearest-Neighborhood method performs better than other combination on item exposure control. The bias of latent trait was minimized under the b-stratified random selection method combined with maximum information method.

**Keywords:** computerized adaptive testing, item exposure rate, item response theory, item selection criterion.

## 1. 前言

電腦適性測驗在測驗開始階段三至五題時，由於能力估計仍未精確，此時若能有效分散試題，對題庫中低曝光率試題提升使用率，並可有效降低題庫中未使用試題題數，對整體試題曝光率達到有效控制，以期達到提升題庫經濟效益。若初始值假設受試者為中等能力，在題庫中一選題法挑選的題目作為施測的起始題；易形成每位受試者都使用相同的題目開始，其保密性需要考量。然由電腦隨機選題起始對初期能力估計的精確度不佳情形下易應估計誤差大而出現極大或級小的估計值而使受試者都使用相同的題目，其保密性亦不佳。因此針對測驗前期三至五題採取b值分層隨機選取方法使測驗前期三至五題試題加大b值差異，試圖找出有效控制題庫中未使用試題題數、降低試題曝光率、改善施測時前後測試題重疊率及試題重疊率的方法。

## 2. 文獻探討

### 2.1 電腦適性測驗 (computerized adaptive testing, CAT) 簡介與探討：

初始值設定：

電腦化適性測驗的基本原則是依受試者能力提供適當的試題呈現給受試者施測，但在測驗之始，受試者之能力高低未能得知，因此，必須決定測驗的起始點，以選擇第一個試題供受試者施測。常用於起始題的決定方式，有以下幾種(王寶壙, 1995；陳麗如, 1998)：中等難度題目：即假設受試者為中等能力，

在題庫中挑選難度適中的題目作為施測的起始題；中等難度題目開始，因題目有限，若每位受試者都使用相同的題目開始，其保密性需要考量。由受試者之基本資料（年齡、學習、經驗或其他測驗結果）估算受試者能力初始值，以決定測驗起始點。自由選題：由受試者在接受測驗的時候，自行判定自己的程度，以決定施測的起始題。隨機選題：由電腦隨機選題，但一般限定試題難度參數  $b$  介於 -0.5 至 0.5 間為選取範圍。

題庫：

電腦化適性測驗與傳統紙筆測驗的不同在於必須建立一個含有試題反應理論測驗試題參數的題庫，題庫中之參數必須以共同量尺來表示，才能有一致的單位。適性測驗之效率與效率，與選題題庫大小具有密切關係(李茂能, 2000)。若要使電腦適性測驗與傳統紙筆測驗具有相同的測驗水準，假如電腦適性測驗採固定長度約為傳統紙筆測驗的一半時，選題題庫大小最好為傳統紙筆測驗的 6 至 8 倍長(Stocking, 1994)。當選題題庫長度為 3 倍以上，精確度與作答效率才有顯著差異(Hung, 1988)。

選題策略：

一、接近難度法(Reckase, 1973; Urry, 1970; Weiss, 1974)

選擇試題難度  $b_j$  最接近受測者能力估計值  $\hat{\theta}$  的試題，最為下一階段施測的試題；

$$F_j(\hat{\theta}) = |\hat{\theta} - b_j|$$

二、KL 訊息法(Kullback-Leibler information index)(Chang & Ying, 1996)

定義：KL 訊息(KL information)

令  $\theta_0$  為真實參數。對任意  $\theta$ ，反應為  $X_i$ ，則第  $i$  題的 KL 訊息定義為

$$K_i(\theta \| \theta_0) = E_{\theta_0} \log \left[ \frac{L_i(\theta_0; X_i)}{L_i(\theta; X_i)} \right]$$

其中  $E_{\theta_0}$  為對  $X_i$  的期望值，且

$$L_i(\theta; X_i) = P_i(\theta_0) \log \left[ \frac{P_i(\theta_0)}{P_i(\theta)} \right] + [1 - P_i(\theta_0)] \log \left[ \frac{1 - P_i(\theta_0)}{1 - P_i(\theta)} \right]$$

定義：KL 測驗訊息(KL test information)

定義：平均 KL 訊息指標(average KL information index)

令  $\hat{\theta}_n$  為第  $n$  題的能力估計值，則第  $i$  題的平均 KL 訊息指標定義為

$$K_i(\hat{\theta}_n) = \int_{\hat{\theta}_n - \delta_n}^{\hat{\theta}_n + \delta_n} K_i(\theta \| \hat{\theta}_n) d\theta$$

其中  $\delta_n$  為平均值的計算區間大小。

此指標表示 KL 函數在  $\hat{\theta}_n - \delta_n$  與  $\hat{\theta}_n + \delta_n$  間的區

域面積，若  $\delta_n$  值小，則指標(式\*\*)決定於  $K_i(\theta \| \hat{\theta}_n)$  在  $\hat{\theta}_n$  上的曲度(curvature)；若  $\delta_n$  值大，則指標易受  $K_i(\theta \| \hat{\theta}_n)$  尾端值影響。故  $\delta_n$  應隨  $n$  遞減到 0，並且區間  $(\hat{\theta}_n - \delta_n, \hat{\theta}_n + \delta_n)$  應包含  $\theta_0$ ，又因  $\theta_0$  的最大概似估計  $\hat{\theta}_n$  為平均數為  $\theta_0$ ，變異數為  $1/I^{(n)}(\theta_0)$  的近似常態分佈，故將區間設為

$$\left\{ \hat{\theta}_n - c / [I^{(n)}(\hat{\theta}_n)]^{1/2}, \hat{\theta}_n + c / [I^{(n)}(\hat{\theta}_n)]^{1/2} \right\}$$

其中常數  $c$  依據收斂機率選擇。因  $I^{(n)}$  為  $n$  階，故可設  $\delta_n$  為

$$\delta_n = c / \sqrt{n}$$

即第  $i$  題的平均 KL 訊息指標為

$$K_i(\hat{\theta}_n) = \int_{\hat{\theta}_n - c/\sqrt{n}}^{\hat{\theta}_n + c/\sqrt{n}} K_i(\theta \| \hat{\theta}_n) d\theta$$

KL 訊息法以此平均 KL 訊息指標選出最大訊息者，作為適性測驗的下一階段施測的試題。

三、單點式最大訊息法

(Birnbaum, 1968; Lord, 1980)

$$I_j(\hat{\theta}) = \frac{[P'_j(\hat{\theta})]^2}{P_j(\hat{\theta})Q_j(\hat{\theta})}$$

選擇訊息函數

四、鄰近法(Nearest-Neighbors criterion, NN)  
步驟一：全部題目皆有對能力最大訊息  $M$  與不偏難度  $b'$

$$M = \text{Max}_{\theta} [I(\theta)] = \frac{(Da)^2}{8(1-c)^2} [1 - 20c - 8c^2 + (1+8c)^{2/3}]$$

$$b' = b + \frac{1}{Da} \log \left( \frac{1 + \sqrt{1+8c}}{2} \right)$$

步驟二：設定非遞增整數  $n^{(k)}$   $k=1, 2, \dots, L$

步驟三：初始化能力值估計值  $\hat{\theta}^{(k)}$

步驟四：找新題  $n^{(k)}$  個，其題目不偏難度  $b'$  最接近  $\hat{\theta}^{(k)}$

步驟五：選  $n^{(k)}$  個題目中，有最大  $M$  的當作  $(n+1)$  題，重新估計能力值為  $\hat{\theta}^{(k+1)}$ ，回到步驟三，直到停止條件成立。

終止條件：

電腦化適性測驗施測題目與題數因人而異，而依測驗的目的與性質，測驗終止的標準一般有下列方式(陳麗如, 1998; 陳新豐, 1999)：設定固定的施測題數。當所有受試者答題數達到預設之題數限制時，即終止測驗，一般以二十至三十題之間為原則。此法常用於模擬研究，其優點是易於設計開發，試題使用率可較精確地預測，但可能使受試者能力估

計的精確度具變動性，因此在實際情境的應用較少。

當受試者的能力估計標準差低於預設標準時，測驗即終止。此即能力估計的精確度已達預定標準，使用此種終止標準通常是以貝氏選題法為選題策略。

當題庫中未使用的試題，再也無法獲得更多的測驗訊息時，即終止測驗。換言之，能力的估計已穩定，再做題目已經沒有幫助，採用此終止標準時，通常以最大訊息選題法為選題策略。

## 2.2 MM 演算法(McBride and Martin, 1983)

McBride 與 Martin 在 1983 年提出此法，減少一個題目出現的次數，以增加題目的安全性。MM 演算法是使用隨機策略使初選題目的題目曝光率下降，其實施方法為，對某一受試者，依能力初始值從題庫中選出 5 題最佳的題目，從 5 題中選出一題施測並重新估計能力值，依新估計的能力值從題庫中選出 4 題最佳題目，從 4 題中選一題施測並重估能力值；依新估計的能力值從題庫中選出 3 題最佳題目，從 3 題中選一題施測並重估能力值；依新估計的能力值從題庫中選出 2 題最佳題目，從 2 題中選一題施測並重估能力值；依新估計的能力值，施測題庫中最佳題目。

## 2.3 b 值分層隨機選取法

將題庫依 b 值分層分三層測驗前期三題採取隨機選題、分五層測驗前期五題採取隨機選題使受試者在測驗前期施策難易度相差較大試題。

## 3. 研究設計

### 3.1 本研究探討的自變項有三：

(1) 控制測驗測驗前期三至五題選題方法：

選題方法	控制題數	代號
中間能力起始選題	無	(A1)
中間能力起始 MM 演算法	三題	(A2_3)
	五題	(A2_5)
隨機起始能力 MM 演算法	三題	(A3_3)
	五題	(A3_5)
隨機選題	三題	(A4_3)
	五題	(A4_5)
b 值分層選題	三題	(A5_3)
	五題	(A5_5)

(2) 後期選題方法：最接近難度法、KL 訊息函數、最大訊息法、鄰近法四種方法；

(3) 控制題數：三題、五題二種情形。

### 3.2 在資料產生部分

是以研究者自行撰寫 MATLAB 6.5 的模擬

程式。受試者為 1000 人，能力服從平均數為 0，標準差為 1 的常態分配，亦即  $N(0,1)$ ；試題題庫共 1200 題採用三參數模式，其中 a 參數服從 0.5 到 2.5 之間的均勻分配，b 參數服從  $N(0,1)$ ，c 參數則為 0。

### 3.3 測驗長度：

為顯現能力估計的收斂情形，故本研究將統一設定為 20 題

### 3.4 能力估計部分：

本研究採用常用的最大概似估計法 (MLE)

### 3.5 曝光率控制效能評估指標：

題目曝光率的均勻度 ( $\chi^2$ )

為量化曝光率，需要一個目標分配，因所有題目應有相同的曝光率，故假設曝光率為均勻分配。則表示每題之曝光率的期望值為平均曝光率，故利用 Pearson 的卡方檢定檢定題目曝光率是否呈現均勻分配，其檢定統計量為

$$\chi^2 = \sum_{i=1}^n (er_i - \bar{er})^2 / \bar{er}$$

表示平均曝光率與觀測的題目曝光率的變異程度，並將題庫使用的效率量化，若  $\chi^2$  值小，則題目曝光率為均勻分配，表示題目被充分的使用，故題庫使用有效率。

題目重複率(test overlap rate)

題目重複率(test overlap rate)是另一個評量題目曝光控制的重要指標。題目重複率為區塊或成對的題目在不同測驗同時讓受試者施測的程度，Way(1997)將題目重複率定義為題目被兩位受試者施測，所有成對比較平均的比例。簡單的定義可將題目重複率為兩個隨機選出的受試者施測的重複題目數，除上測驗長度。所以若 N 為受試者人數，題目重複率可由以下所計算：(1)計算  $N(N-1)/2$  對受試者重複題目數；(2)加總此  $N(N-1)/2$  個數；(3)將此總數除以  $N(N-1)/2$ 。理想上，任一對受試者重複題目數應該被最小化。

能力估計誤差評估：(RMSE)

利用真實能力值  $\theta$  與估計能力值  $\hat{\theta}$  的均

方根差(root mean squared error, RMSE)，評估電腦適性測驗系統對受試者能力估計的準確度。

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{\theta}_i - \theta_i)^2}$$

以及試題中未選用題數、最低曝光率、最高曝光率、曝光率大於.3 題數、曝光率大於>.2 題數、前後測試題重疊率。

#### 4. 研究結果

曝光率的均勻度和能力估計的準確度比較：

最接近難度法中(表一)

以 b 值分層選題法控制測驗測驗前期五題選題曝光率的均勻度 46.9、能力估計誤差.23 為九組中最佳。且最高曝光率.224 亦為最低。

KL 訊息函數中(表二)

以 b 值分層選題法控制測驗測驗前期五題選題曝光率的均勻度 128、能力估計誤差.25 為九組中最佳。且最高曝光率.336 亦為最低。

最大訊息法(表三)

以 b 值分層選題法控制測驗測驗前期五題選題曝光率的均勻度 109 為最低但能力估計誤差.19，比中間能力起始選題提高.02。

以 b 值分層選題法控制測驗測驗前期三題選題曝光率的均勻度 134 為最低但能力估計誤差.18，比中間能力起始選題提高.01。

鄰近法(表四)

以 b 值分層選題法控制測驗測驗前期五題選題曝光率的均勻度 57.4、能力估計誤差.21 為九組中最佳。且最高曝光率.278 亦為最低。

b 值分層選題法控制測驗測驗前期三題、五題等二種方法對四種選題法的未選用題數達成明顯改善、最低曝光率達到提升效果、最高曝光率達到下降、曝光率大於.3 題數及曝光率大於>.2 題數均有明顯減少、前後測試題重疊率和試題重疊率亦能達到下降。

就控制曝光率而言以 b 值分層選題法控制測驗測驗前期五題選題，六至二十題採鄰近法選題是最佳選擇；若考慮能力估計誤差可採以 b 值分層選題法控制測驗測驗前期三題選題，四至二十題採最大訊息法選題。

#### 5. 結論與建議

電腦適性測驗測驗前期三至五題採取 b 值分層隨機選取對使用接近難度法、KL 訊息

法、最大訊息法鄰近法等選題策略，試題題庫有以下明顯改善：

(1)有效控制題庫中未使用試題題數提升題庫建制之經濟效益

(2)降低試題曝光率延長題庫使用時限

(3)改善施驗時前後測試題重疊率及試題重疊率

未來可將電腦適性測驗測驗前期三至五題採取 b 值分層隨機選取結合現有整體曝光率控制方法以達到最低曝光率提升、最高曝光率下降控制的組合方法。

#### 參考文獻

- 王寶壙 (1995)，現代測驗理論，心理出版社，臺北。
- 李茂能 (2000)，中文電腦化適性測驗系統之應用與評鑑。文景書局。
- 陳新豐 (1999)，「多媒體線上適性測驗系統發展及其相關研究」，碩士論文，國立臺南師範學院國民教育研究所，臺南。
- 陳麗如 (1998)，「電腦化適性測驗之題庫品質管理策略」，碩士論文，國立臺灣師範大學資訊教育研究所，臺北。
- Birnbaum, A. (1968). *Some latent trait model and their use in inferring an examinee's ability*. In F. M. Lord and M. R. Novick, *Statistical theories of mental test scores*. Reading, Mass: Addison Wesley.
- Chang, H. H. & Ying, Z. (1996). *A global information approach to computerized adaptive testing*. *Applied Psychological Measurement*, 20, 3, 231-229.
- Cheng P. E., & Liou, M. (2003). *Computerized adaptive testing using the nearest neighbors criterion*. *Applied Psychological Measurement*, 24, 257-265.
- Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York: Wiley.
- Davey, T., & Parshall, C. G.(1995, April). *New algorithms for item selection and exposure control with computerized adaptive testing*. Paper presented at the annual meeting of American Educational Research Association, San Francisco.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Hingham, MA: Kluwer Boston, Inc.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, N. J.: Lawrence Erlbaum.
- Stocking, M. L. (1994). *Three practical issues for modern adaptive testing item pools*. Educational Testing Service, Princeton, N. J. (ERIC Document Reproduction Service No. ED 385 551)
- Veerkamp, W. J. J., & Berger, M. P. F. (1997). *Some new item selection criteria for adaptive testing*.

Way, W. D. (1997, March). *Protecting the integrity of computerized testing item pools*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago.

表 1

最接近 難度法	A1	A2_3	A3_3	A4_3	A5_3
$\chi^2$	170	103	72.06	57.46	57.5
RMSE	0.26	0.26	0.27	0.25	0.26
		A2_5	A3_5	A4_5	A5_5
$\chi^2$		80.53	69.9	48.79	46.9
RMSE		0.26	0.27	0.26	0.23

表 2

KL 訊息 函數	A1	A2_3	A3_3	A4_3	A5_3
$\chi^2$	333	280.4	223.7	257.6	153
RMSE	0.28	0.27	0.29	0.28	0.25
		A2_5	A3_5	A4_5	A5_5
$\chi^2$		255.3	215.9	282.3	128
RMSE		0.27	0.27	0.29	0.25

表 3

最大訊 息法	A1	A2_3	A3_3	A4_3	A5_3
$\chi^2$	227	187.3	174.3	193.5	134
RMSE	0.17	0.17	0.18	0.18	0.18
		A2_5	A3_5	A4_5	A5_5
$\chi^2$		169.3	163.6	224.5	109
RMSE		0.17	0.17	0.19	0.19

表 4

鄰近法	A1	A2_3	A3_3	A4_3	A5_3
$\chi^2$	173	168.6	77.06	91.96	73.2
RMSE	0.2	0.25	0.24	0.21	0.21
		A2_5	A3_5	A4_5	A5_5
$\chi^2$		169.2	76.75	92.35	57.4
RMSE		0.25	0.26	0.2	0.21

表 5

最接近 難度法	未選用 題數	最低選最高使		最低 曝光率	最高 曝光率
		用 次數	用 次數		
A1	117	0	1000	0	1
A2_3	93	0	351	0	0.351
A2_5	58	0	242	0	0.242
A3_3	30	0	342	0	0.342
A3_5	18	0	232	0	0.232
A4_3	10	0	316	0	0.316
A4_5	28	0	240	0	0.24
A5_3	0	2	301	0.002	0.301
A5_5	0	4	224	0.004	0.224

表 6

KL 訊息 函數	未選用 題數	最低選 用次數	最高使 用次數	最低曝 光率	最高曝 光率
A1	1047	0	1000	0	1
A2_3	1044	0	568	0	0.568
A2_5	1027	0	484	0	0.484
A3_3	1016	0	435	0	0.435
A3_5	1016	0	355	0	0.355
A4_3	1024	0	508	0	0.508
A4_5	1032	0	580	0	0.58
A5_3	0	2	341	0.002	0.341
A5_5	0	4	336	0.004	0.336

表 7

最大訊 息法	未選用 題數	最低選 用次數	最高使 用次數	最低曝 光率	最高曝 光率
A1	1006	0	1000	0	1
A2_3	1008	0	444	0	0.444
A2_5	1008	0	304	0	0.304
A3_3	1003	0	361	0	0.361
A3_5	1002	0	291	0	0.291
A4_3	1013	0	389	0	0.389
A4_5	1017	0	459	0	0.459
A5_3	0	2	326	0.002	0.326
A5_5	0	4	296	0.004	0.296

表 8

鄰近法	未選用 題數	最低選 用次數	最高使 用次數	最低曝 光率	最高曝 光率
A1	252	0	1000	0	1
A2_3	112	0	1000	0	1
A2_5	108	0	1000	0	1
A3_3	35	0	488	0	0.488
A3_5	34	0	486	0	0.486

A4_3	258	0	322	0	0.322
A4_5	289	0	287	0	0.287
A5_3	0	2	300	0.002	0.3
A5_5	0	4	278	0.004	0.278

表 9

最接近 難度法	前後測試		試題重疊	
	$r \geq 0.3$	$r \geq 0.2$	題重疊率	率
A1	5	10	0.41	0.55
A2_3	5	9	0.33	0.47
A2_5	0	15	0.29	0.42
A3_3	2	6	0.27	0.39
A3_5	0	10	0.28	0.41
A4_3	1	4	0.24	0.36
A4_5	0	3	0.21	0.31
A5_3	1	5	0.25	0.38
A5_5	0	3	0.22	0.33

表 10

KL 訊 息函數	前後測試		試題重疊	
	$r \geq 0.3$	$r \geq 0.2$	題重疊率	率
A1	14	35	0.69	0.78
A2_3	17	37	0.65	0.69
A2_5	14	40	0.61	0.62
A3_3	14	40	0.54	0.49
A3_5	12	42	0.53	0.48
A4_3	18	37	0.54	0.50
A4_5	19	36	0.52	0.49
A5_3	6	23	0.48	0.47
A5_5	3	18	0.42	0.41

表 11

最大訊 息法	前後測試		試題重疊	
	$r \geq 0.3$	$r \geq 0.2$	題重疊率	率
A1	5	17	0.73	0.76
A2_3	7	21	0.68	0.71
A2_5	1	23	0.64	0.68
A3_3	4	27	0.63	0.65
A3_5	0	22	0.64	0.65
A4_3	9	31	0.61	0.64
A4_5	11	33	0.57	0.59
A5_3	2	18	0.57	0.59
A5_5	0	13	0.48	0.51

表 12

鄰近法	前後測試		試題重疊	
	$r \geq 0.3$	$r \geq 0.2$	題重疊率	率

		題重疊率	率	
A1	5	8	0.49	0.60
A2_3	5	9	0.41	0.55
A2_5	5	9	0.41	0.54
A3_3	1	12	0.29	0.44
A3_5	1	8	0.29	0.41
A4_3	1	6	0.35	0.44
A4_5	0	5	0.30	0.38
A5_3	1	6	0.34	0.44
A5_5	0	5	0.29	0.38

圖 1

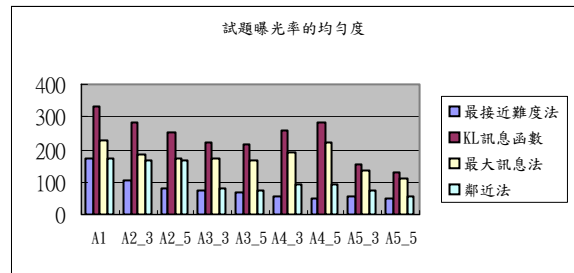


圖 2

