



# Introduction to Computerized Adaptive Testing

Speaker : YI CHENG LI





# Outline

- ❑ Definition
- ❑ CAT (Computerized Adaptive Testing)
- ❑ Item Banking





# Definition (1)

- 電腦適性測驗(computerized adaptive testing, CAT)  
電腦依受測者的作答反應，來選取下一階段施測試題，以提高施測的效率及受測者的能力估計精確度。
- 題庫(Item Bank)  
一群試題經過校準、分析、歸類與評鑑後，而集合在一起的組合體。
- 試題反應理論(item response theory, IRT)  
能夠針對每位受測者，提供不同的標準誤，因此能夠評估受測者的能力值範圍，並且提出以試題訊息及測驗訊息的概念，來作為評定某試題或某測驗的施測準確性。





## Definition (2)

- 訊息函數(Information Function)  
在電腦適性測驗中，扮演一個重要的角色，在測驗的進行中，藉由試題訊息函數，挑選適合於受測者的試題，使得試題對受測者為難易適中，並求得最大測驗訊息，使其能力估計的標準誤為最小，達到提高能力估計的精確度，並且能提高測驗的效率。
- 試題訊息函數定義如下：

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

- 其中 $I_j(\theta)$ 為能力值 $\theta$ 在試題 $j$ 的試題訊息函數， $P_j(\theta)$ 為能力值 $\theta$ 在試題 $j$ 的答對機率， $Q_j(\theta)$ 為能力值 $\theta$ 在試題 $j$ 的答錯機率， $P'_j(\theta)$ 為能力值 $\theta$ 的試題反應函數的導數





## Definition (3)

- 試題反應模式(Item Response Model)  
用來描述能力特質與正確反應機率間的關係
- 常用的三種模式：
  - 單參數對數模式
  - 雙參數對數模式
  - 三參數對數模式

$$P_j(\hat{\theta}) = c_j + (1 - c_j) \frac{1}{1 + \exp(-D \cdot a_j(\hat{\theta} - b_j))}$$

- 其中試題 $j$ 的鑑別度為 $a_j$ ，難度為 $b_j$ ，猜測度為 $c_j$ ， $\theta$ 為考生的能力估計值， $D=1.702$





## Definition (4)

在三參數模式下，試題訊息函數可表示成：

$$I_j(\theta) = \frac{d^2 a_j^2 (1 - c_j)}{[c_j + \exp(da_j(\theta - b_j))] \cdot [1 + \exp(-da_j(\theta - b_j))]^2}$$





## Definition (5)

### ■ 測驗訊息函數

測驗訊息為各試題訊息之總和，定義如下：

$$I_T(\theta) = \sum_{j=1}^n I_j(\theta)$$

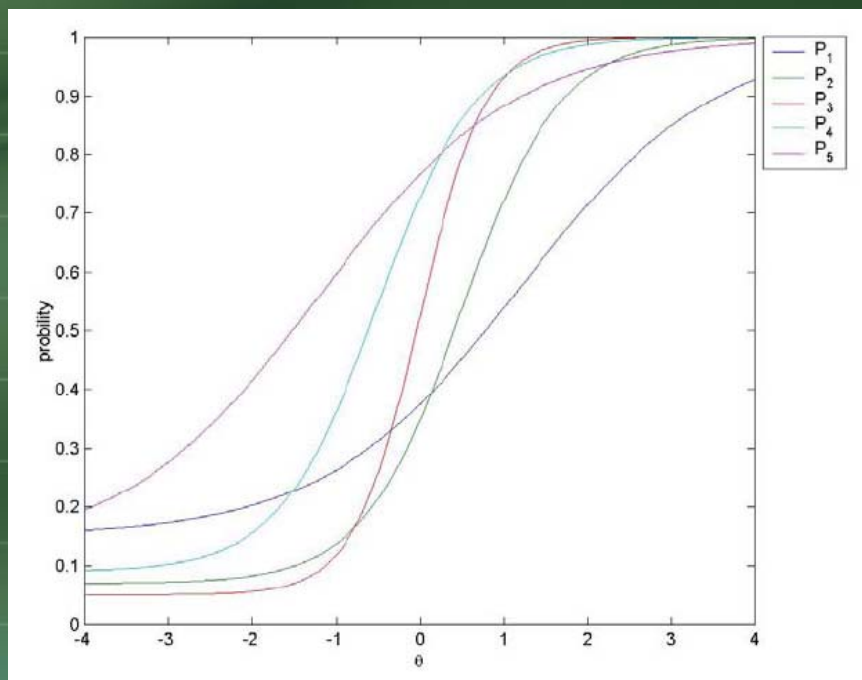
- 若要增加測驗訊息，可以有二種方法，一是測驗題數不變，增加各試題訊息，另一是各試題訊息不變，增加測驗的題數。





## Definition (6)

- 試題特徵函數(item response function)又稱試題特徵曲線(item characteristic curve, ICC)，受測者在某試題上的表現，根據受測者的能力值與受測者的作答機率，經由機率的統計，所得之對應函數。





# CAT (1)

- 測驗的目的：測量考生的能力值
- 測驗的難度適合考生的能力程度時，則測驗所測量到的考生能力最為準確
- 理想的施測狀態：能夠針對每位考生不同的能力程度，來提供適合個別情況的測驗方式
- 傳統的紙筆測驗難度不可能同時適合每位考生的能力需求，因此無法精準地測量出考生能力





# CAT (2)

- 電腦適性測驗是一種為受測者「量身打造」的個別測驗。
- 符合「因材施教」、「經濟有效」、「誤差最小」的原則
- 主要特點：
  - 適性選題施測：使施測試題的難度，符合受測者的能力範圍內，因此試題具有適性。
  - 節省施測時間：每位受測者的施測試題不同，在適性測驗的過程中，若符合測驗終止條件，則施測結束，受測者不用作答偏易、偏難的試題，因此節省施測時間。
  - 精確估計能力：在適性測驗裡，能對估計精確度提供最大貢獻量，會被優先選取為施測試題，因此受測者能力估計的精確度最大。





# CAT (3)

## ☐ 電腦的工作：

- ☐ 根據考生先前的反應表現，預測他在尚未接受測驗的試題上的種種可能反應。
- ☐ 根據上述的理解，有效地挑選試題，接著呈現給考生作答。
- ☐ 最後在測驗完畢時，能夠計分，以分數來表示考生能力的大小。





# CAT (4)

## 電腦適性測驗的流程





# CAT (5)

## ■ 初值設定

■ 適性測驗的開始，對受測者的能力高低一無所知，無法作為任何選題方法的依據。通常有幾種方法：

1. 全域完全隨機法：在題庫中隨機選取一題為起始題。
2. 中等難度隨機法：在設定一個中等難度區域中，隨機選取一題為起始題。
3. 背景經驗調查法：根據受測者的背景資料，比對經驗資料，作為選擇起始題之依據。
4. 設定能力選題法：設定受測者之能力值，再依各種選題方法擇題。





# CAT (6)

## 選題方法

### 最接近難度法

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

### 最接近偏移難度法

選擇試題偏移  $m_j$  最接近受測者能力估計值  $\theta$  的試題，作為下一階段施測的試題。 $m_j$  定義如下：

$$m_j = b_j + \frac{1}{d \cdot a_j} \ln\left(\frac{1 + \sqrt{1 + 8c_j}}{2}\right)$$

### 單點式最大訊息法

選擇  $I_j(\theta)$  最大的試題，作為下一階段施測的試題。





# CAT (7)

## 能力估計

常用的方法有：

- 最大近似值估計法

- 貝氏估計法

最大近似值估計法的估計效能很好，但題數少或估計值無法收斂時，會產生問題；貝氏估計法，如事前分配假設不當時，會產生能力值估計偏差





# CAT (8)

## ☀ 終止條件

### ☀ 常用的終止條件有：

- ☀ 固定長度法：當受測者完成固定測驗題數，則終止條件成立。
- ☀ 估計標準誤：當受測者的能力值估計標準誤小於預設值，則終止條件成立。
- ☀ 測驗訊息法：當受測者的測驗訊息大於預設值，則終止條件成立。
- ☀ 事後變異法：當受測者的事後變異小於預設值，則終止條件成立。
- ☀ 測驗時間法：當受測者的測驗時間大於預設值，則終止條件成立。





## CAT (9)

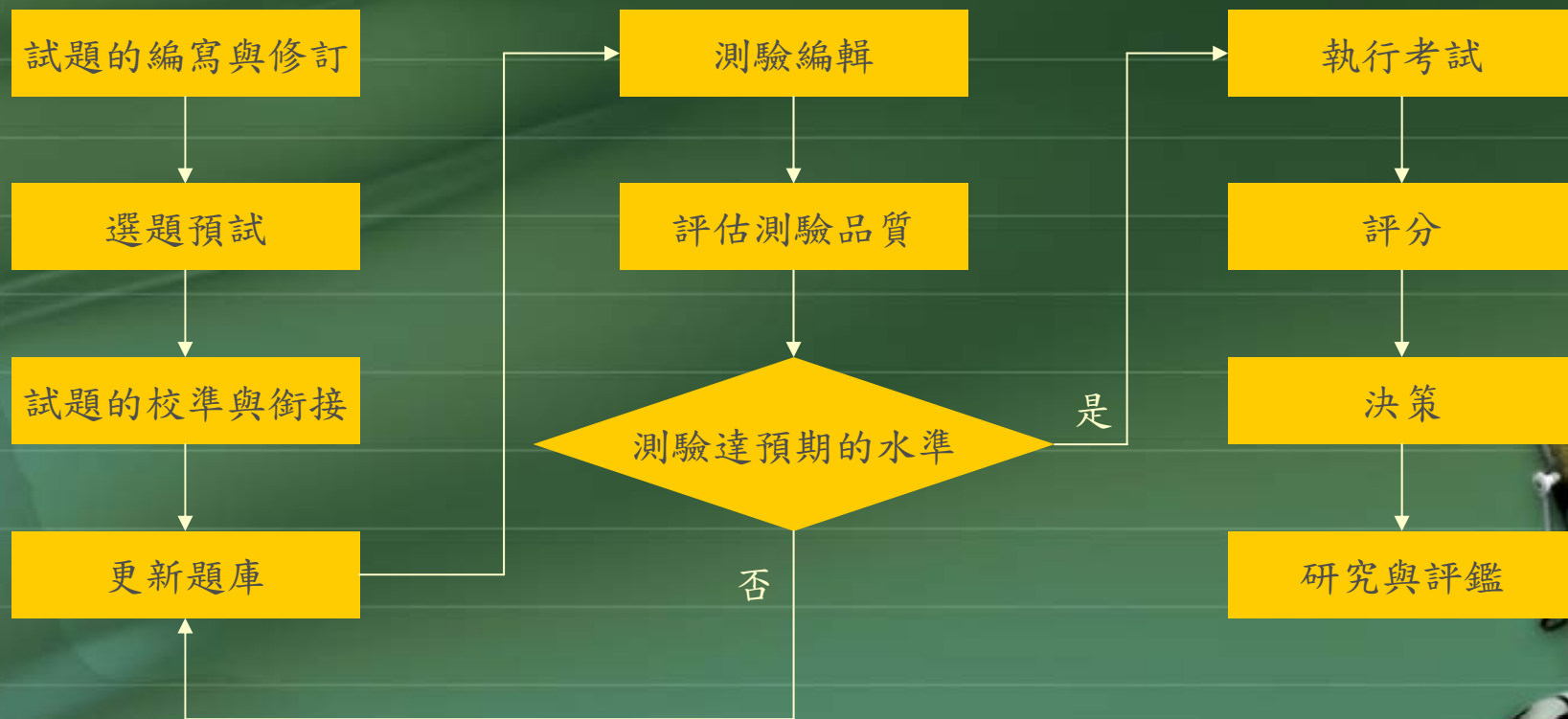
- ☐ 「三參數對數模式」是最適合應用到適性測驗上的試題反應模式
- ☐ 選題方法使用「最大訊息法」，可同時降低測驗長度與提高測驗訊息量





# Item Banking (1)

## 建立題庫的步驟：





# Item Banking (2)

## ■ 建立題庫之重大課題：

### ■ 題庫應包含多少試題？

■ 基本上越多越好

■ 每個概念至少要包含10個試題，每一單元課程內容至少要包含50題

■ 建議一百至二百個難度均勻分佈，且具有合理的鑑別度的試題，便可適用在電腦化適性測驗

### ■ 題庫系統應該如何分類？

■ 依主題或教學目標來檢索試題

■ 依關鍵字來檢索試題

### ■ 題庫是否可以公開？

■ 教學內容窄化？

■ 公開少數的樣本試題

### ■ 題庫是否安全？

■ 考古題的出現？

■ 隨時更新題庫的內容，確保試題的內容效度和統計品質

