

# 測驗的基本概念

## 一、測驗、測量、評定、評量

	意義	特徵	實施方式
<b>測量</b> measurement	透過測量工具，用數字描述個人特質或所測事物屬性。	不涉及價值判斷	檢核表、問卷實施。如測量身高
<b>測驗</b> test	對於具代表性的樣本進行針對個人特質的系統性程序。	強調客觀性	智力測驗、性向測驗等。
評定	即蒐集、統整和解釋訊息，幫老師做決定的歷程。		教師收集訊息所使用的各種方法。
<b>評量</b> assessment	根據一項標準，對所測量到的數量做價值的判斷。即教師蒐集、整合與詮釋學生所有的學習表現，以擬定出適當的教學決定之歷程	要進行價值判斷	真實評量、變通性評量、實作評量
<b>評鑑</b> evaluation	經由鑑定、判斷的過程，據以檢討得失，力求革新進步。	要進行價值判斷	形成性評量、總結性評量、安置性評量、診斷性評量。

## 二、測驗的發展

### (一) 產生背景：

1. 比奈受法國教育部委託，1905 年發展出比西量表，為智力測驗之父。奠立智力測驗編製的基礎。
2. 實驗心理學之父馮德在 1879 年德國創立了第一個心理實驗室。奠定了測驗程序標準化的基礎。
3. 英國高爾登致力個別差異的研究，提供了測驗資料分析的統計方法。
4. 心理測驗一詞，為美國卡泰爾於 1890 年的文章中首先創用。
5. 桑代克在 1904 年出版第一部教育測量的教科書，開創基本的測量理論與技術。

## (二) 理論背景

	學習理論	評量理念
<b>1</b> 實徵/行為 主義觀	1. 強調知識與經驗的一致性。 2. 知即為聯結，技能部分的組合。 3. 學習即為形成新的聯結級技能的獲取。 4. 強調可觀察的行為的聯結。	1. 有效公平客觀的方式測量。 2. 重視信度、效度。 3. 對知識的抽樣。 4. 量化、標準化的紙筆式評量
<b>2</b> 理性/認知 主義觀	1. 強調概念間的連貫性，真理的形式判準。 2. 知是對各領域概念的理解。 3. 學習即為形成適當的知識結構。	1. 看出學生的心智能力成長。 2. 評量實作，長程表現。 3. 以多元標準衡量學生的進展 4. 實作評量、卷宗評量。
<b>3</b> 社會/歷史 主義觀	1. 知識由人與人、人與物的互動中建構。 2. 情境主義。 3. 學習是一種調和與其所處情境之物理、社會系統的支援與限制的歷程。	1. 評量學生參與學習的品質。 2. 評量學習歷程，尤指社會互動歷程。 3. 學生自省自評。 4. 卷宗評量。

## 三、測驗的功能

(一) **教學評量的功能**：透過教育測驗，教師在做教學決定時，有下列四項功能：

1. 瞭解學生的起點行為：在教學前，教師可以先針對學生實施一份教育測驗，用來評測驗估學生在學習前已具有的背景知識，以作為決定有效教學的起點。
2. 確定教學目標達到的程度：教學後，需確定教學目標是否切實可行？教學目標實現的程度如何？須借助於一些成就所得的結果，以做正確的判斷。
3. 作為改進教學的參考：提供教師明瞭自己在教學上的缺失，作為改進教學的參考。
4. 評定學生的學習成果：針對學生的學習成果進行評定等第，以作為其學習成就的代表，同時又可以提供作為其他教育研究用途的指標。

(二) **落實輔導與諮商效能**：透過測驗的實施，可使個人進一步的了解其自我及各種困擾與不適應行為的緣由。所以獲取學業性向、成就、興趣及人格特質的量化數據資料，就成為輔導諮商歷程中一項很重要的工作。

(三) **行政決定的功能**：

1. **選擇決定**的功能：指一個人是否被錄取或被拒絕。  
例如，大學入學考試或就業申請，常以測驗做為甄選的工具。
2. **安置決定**的功能：指決定分派學生到不同層次的教學班級中。  
例如，根據數學能力測驗，決定那些學生分派到一般課程班級？在安置決定中，沒有任何人被拒絕，所有的人都接受不同層次的教學計畫。
3. **分類決定**的功能：指將個人分派到不同類別的工作、職務、學校或科系中。  
例如，根據性向測驗結果，可以據以分派入伍的軍人到不同工作的崗位。
4. **課程與教育計畫**決定的功能：實驗過程常採用測驗做為收集資料的客觀工具，據以評鑑新的課程或教育計畫的有效性，然後決定新的課程是否需要修改？有無採用的價值？

#### 四、高品質測驗的特徵

- (一) **具有高信度**：信度是指測驗的可靠性，包含測驗分數的穩定性與測驗內容的一致性。穩定性是指同一份測驗經多次施測所得的分數，相當一致，並沒有太大的差異。一致性是指測驗中的各個題目，在功能上或性質上而言，是屬於同一類或趨近於同一類。信度係數的值域，也是介於 0 到 1 之間；並且是數值愈接近 1，即表示信度愈好。
- (二) **具有高效度**：效度是指測驗能夠測量到它所欲測量之特質的程度，或是指測驗能達到其目的的程度。一個測驗的效度愈高，表示它愈能夠測量所欲測量的特質。
- (三) **具有適當常模**：常模指某特定團體在某一測驗上之得分的平均數。
- (四) **高實用性**
1. **經濟**(易於取得所需的測驗)
  2. **容易實施**：容易實施的範圍包括容易安排測驗情境、指導語易說易懂、施測人員易於安排及施測時間不宜太長等。
  3. **容易計分、解釋和應用**

#### 五、測驗者分數的影響因素

影響因素	說明
主試者	主試者的性質、對測驗之瞭解程度、專業程度等，都會影響測驗分數的真實性。
評分程序	評分程序的客觀性與標準化與否，都會影響測驗信度，也會影響測驗分數的真實性。
測驗情境	測驗指導語、測驗環境的光線、噪音、通風性等，都會影響分數真實性
測驗材料	測驗試題多寡會影響測驗信度，也會影響測驗分數的真實性。
受試者	受試者的性質、變異性、代表性等，都會影響測驗分數的真實性。

#### 六、測量與量尺

##### 1. 名義/類別量尺 categorical scales

- (1) 不具備任何的數學屬性，單純將資料或物件用數字來分類、標示其身分
- (2) 如：人、事、物等名稱、種類；男生表示為 1、女生為 2

##### 2. 順序/次序量尺 ordinal scales

- (1) 有大小、規模的方向、次序性，可以判斷是否有  $A > B$ ,  $A = B$ ,  $A < B$  的關係
- (2) 相鄰單位間並無等距關係，如地震震度，百分位數

##### 3. 等距量尺 interval scales

- (1) 可以比較大小/規模，例如： $A > B > C$
- (2) 相等單位：相鄰單位間為等距關係  $A - B = B - C$
- (3) 沒有絕對的零點，如溫度、智商

##### 4. 比率量尺

- (1) 擁有等距所擁有的一切特質
- (2) 具備絕對零點的特質，如個人所得、工作時間、年齡等。

# 信度與效度

## 一、信度 Reliability

### (一) 信度的意義

概念性的定義

1. 一個測驗結果的一致性，信度越高表示該測驗的結果越一致。一般而言，分數越一致，受誤差的影響越小，所以信度主要在說明測驗分數中測量誤差所占成分。
2. 也可說信度高的測驗有良好的穩定性。
3. 受試團體在測驗分數的差異越大，異質性越高，則信度通常也會較高。
4. 在原有條件(學生人數、難度、鑑別度等)都不變的情況下，測驗題目越多，則信度會提高。

計量上的定義

1. 真分數模式  $X = T + E$  (實得分數= 真正分數+誤差分數)
2. 三者變異量的關係  $S_x^2 = S_t^2 + S_e^2$
3. 真實分數的變異數  $S_t^2$  佔總變異量  $S_x^2$  的比率，即定義為「信度」。

$$\frac{S_t^2}{S_x^2} = \frac{S_t^2}{S_x^2} + \frac{S_e^2}{S_x^2}$$

### (二) 信度的性質

1. 信度是指評鑑工具所獲得「結果」的可靠性，而非指工具本身。
2. 每一個信度的估計值，僅指某一特定類型的一致性，而非泛指一般的一致性。
3. 信度是效度的必要條件，但不是充分條件。

### (三) 信度功能

1. 代表測驗分數具有一致性、穩定性和可預測性，可反應測量工具與程序。
2. 測驗要能夠為同一位學生提供前後一致的測驗結果。
3. 為測驗品質是否優良的指標。

### (四) 信度係數

1. 題目越多，信度越高。
2. 難度中等的成就測驗，比難度高的或低的成就測驗有較高的信度。
3. 愈能區分高、低成就者的測驗，題目信度越高。
4. 當分數的全距越低時，信度係數會越低。
5. r 值範圍在 0.00~1.00 之間，沒有負值，r=0 表示無信度；r=1 表示信度最高。

(五) 信度指數 index or reliability：又叫決定係數，是個人測驗分數與真實分數的相關，而信度指數的平方即為信度係數，是真實特質的變異佔測驗分數總變異量的程度。

(六) **測量誤差**可分為兩類：

1. **系統誤差/偏誤**：學生的學習、訓練、遺忘、生長等，是指一種以固定、一致的方式影響測驗分數高低的測量誤差。不同情境中，對不同學生影響都是一致的。
2. **非系統誤差/隨機誤差**：學生的身心狀況(動機、情緒、態度、意願)、施測情境(噪音、溫度、照明)、測驗試題(編擬、抽樣、計分、解釋)等，是以隨機、無規則、不可預測的方式，在不同情境中，隨時影響不同學生的測驗分數。

(七) **信度的影響因素**

1. **測驗的長度**：測驗越長，內容愈具有代表性，分數受到猜測因素的影響越小，所以信度越高。
2. **分數的分布情形**：分數的分布範圍越大,第二次測量時,分數相對位置改變的可能性就愈小,亦即測量誤差對相對位置的影響愈小。
3. **測驗的難度**：測驗難度適中，能使分數變異程度加大,信度變高。
4. **客觀性**：具客觀性的測驗其評分結果較不會受到評分者的判斷與意見的影響，所以相對而言其穩定性高，信度高。

(八) **信度的補救方法**

1. **增加試題或刪除不良試題**：增加試題可以提高測驗的信度，但需考慮學生是否會疲勞。
2. **校正相關係數的萎縮**：對於低信度的測驗最好不要使用，若不慎使用了，則務必進行相關係數萎縮的校正，才能獲取接近真實測量下的真正相關。如斯布校正公式使信度提高

(九) **信度概念在教學評量的應用**

1. 幫助教師評鑑已發行評量工具的信度
  - (1) 教師應熟悉信度的主要概念。
  - (2) 教師應辨別有多種的信度估計方法。
  - (3) 教師應認識某測驗的信度愈高，該測驗越佳。
  - (4) 教師也應進一步認知到即使是相同的測驗，也會因所採用的估計方法不同，所得到的信度係數也會有高低。
2. 幫助教師增進自行編制之評量工具的信度

## 二、信度的類型與考驗方法

### (一) 常模參照

#### 1. 重測信度 test-retest method

- (1) 同一份測驗於不同時間對相同學生前後重複測量兩次，並根據這兩次測驗分數所求得的相關係數。通常用皮爾森積差相關(r) 代表。
- (2) 例：一份測驗在不同的時間點測驗同樣五位學生的表現
- (3) 再測信度的高低和兩次間隔時間的長短有關。
- (4) 間隔大多以一至兩週最為恰當。
- (5) 前測與後測之間，測驗分數常有提高的現象，是因「練習效應」和「累積性的成長」。

#### 2. 複本信度 equivalent form

- (1) 根據相同編制藍本使兩份題數、難度、指導說明、施測時限與例題舉隅等都相當的測驗拿給同一批學生施測，在依據施測後所得的兩份測驗分數來求相關係數。
- (2) 例：五位學生在正本和複本兩種測驗上的表現情形。
- (3) 優點：不受練習和記憶的影響，為最好的信度。
- (4) 複本信度的缺點：
  - ① 編製不易，費時且成本高，因此多數測驗都沒有複本。
  - ② 易受練習影響的測量行為，即使採用複本，影響只能減少，不能避免。

→ 重測與複本信度都必須進行兩次施測或使用兩份測驗，使用不方便。

#### 3. 評分者信度 score reliability

- (1) 採用不同評分者評閱測驗券。

(2) 公式：

$$\rho = 1 - \frac{6 \sum d^2}{N(N^2 - 1)}$$

$\rho$  為評分者信度， $d$  為評定等第之差， $N$  為被評者的人數。

(評分者為兩人的 Spearman 等級相關)

(3) 公式：

$$W = \frac{R_i^2 - \frac{(R_i)^2}{N}}{\frac{1}{12} k^2 (N^3 - N)}$$

$W$  為評分者信度， $R_i$  為被評者之評定等第分數， $k$  為評分者人數， $N$  為被評者人數。

(評分者不只兩位的肯德爾和諧係數)

- (4) 舉例(評分者為兩人)：兩位評分者使用等第方法評定五位學生的作文成績
- (5) 舉例(評分者為兩人以上)：五位國文教師每人評定九篇作文的等第
- (6) 採用時機：對一些無法進行完全客觀記分的測量工具，例如：人格測驗、創造性思考測驗及申論題、作文題等必然涉及評分者主觀的判斷，故須對評分者間的變異加以衡量。

4. 內部一致性信度 internal-consistency methods：只根據一次測驗結果就來估計信度的作法。常用的有三種：

(1) 折半信度 split-half reliability

- ① 利用單獨一次測驗結果，以隨機方式將其分成兩半，再求出這兩半測驗結果間的相關係數。

② 公式：
$$r_{xx'} = \frac{gr_h}{1 + (g-1)r_h}$$

$r_{xx}$  為完整測驗的估計信度， $r_h$  為原測驗的信度， $g$  為測驗加長或減短題數的倍數。

- ③ 例：五位學生在一個四個题目的測驗反映情形

(2) 庫李信度 Kruder-richardson reliability

- ① 依據學生對所有試題的反應，分析試題間的一致性，以確定測驗中的試題是否都測量到相同特質的一種信度估計方法。

- ② 須符合三種基本假設

- A. 試題的計分是使用「對或錯」的二元化計分方式。  
B. 試題不受作答速度的影響。  
C. 試題都是同質的，亦即都測量到一個相同的因素。

③ 公式：
$$KR_{20} = \frac{n}{n-1} \left[ 1 - \frac{\sum pq}{S^2_x} \right]$$

$n$  為測驗的題數， $p$  為答對某一題的人數比率， $q$  為答錯某一題的人數比率，為全部題目答對與答錯的百分比的總乘積， $S^2_x$  為測驗總分的變異數

- ④ 例：「對或錯」的二元化計分測驗-五位學生在是非題測驗上的表現情形

(3)  $\alpha$  係數 coefficient  $\alpha$

- ① 可適用於多元計分的測驗，如學生在五點式量表上的表現。

② 公式：
$$\alpha = \frac{n}{n-1} \left[ 1 - \frac{\sum S^2_i}{S^2_x} \right]$$

$\alpha$  係數為測驗的估計信度， $n$  為題數， $S^2_i$  為每一試題得分的變異數， $S^2_x$  為測驗總分的變異數

- ③ 例：李克特 Likert 五點量尺。

信度類型	信度的涵義	主要的誤差來源
再測信度	同一份測驗的兩次測量結果間的相關係數	時間抽樣
複本信度	兩份複本測驗間的測量結果的相關係數	時間抽樣與內容抽樣
評分者信度	各評分者間或各評分者內之評分結果的相關係數	評分者誤差
內部一致性信度	同一測驗的測量結果內各試題間的相關係數	內容抽樣與內容異質
(1)折半信度	測驗試題分成兩半，這兩半間的相關係數	內容抽樣
(2)KR <sub>20</sub> 公式 KR <sub>21</sub> 公式	試題間的同質性或反應一致性程度的關聯性指標	內容抽樣與內容異質
(3) $\alpha$ 係數	試題間的同質性或反應一致性程度的關聯性指標	內容抽樣與內容異質

## (二) 標準參照

### 1. 百分比一致性指標

(1) 分析前後兩次分類決定結果是否為一致，並以百分比之和來表示。

(2) 公式：
$$P_A = \frac{b}{N} + \frac{c}{N} = \frac{b+c}{N}$$

b 和 c 表示兩次結果都相同的人數, N 表總人數

### 2. K 係數

(1) 由 Cohen(1960)所提出，分析評分者實際評定為一致的次數百分比，與評分者在理論上評定為一致的最大可能次數百分比(經校正誤差後)的比率。

(2) 公式：
$$k = \frac{P_A - P_c}{1 - P_c}$$

$P_A$  表實際的情形,  $P_c$  表預期的情形

## 三、測量標準誤

1. 亦為表示信度的方法，指個人測驗上實得分數和真正分數的差，稱為測量誤差。
2. 信度係數較適合於比較不同測驗的信度；而測量標準誤則較適合於解釋個人的分數。
3. 利用測量標準誤來解釋個人的測驗分數，以推估出個人真實分數的可信範圍。
4. 測量標準誤與信度呈現反比關係。信度愈高，測量標準誤愈低；信度愈低，測量標準誤愈高。  
當信度等於 1 時，測量標準誤為 0；當信度等於 0 時，測量標準誤為 1

5. 公式：
$$SE_{meas.} = S_x \sqrt{1 - r_{xx}}$$

$SE_{meas.}$  為測量標準誤  $S_x$  為測驗的標準差  $r_{xx}$  為測驗的信度係數

➤ 舉例：某生在一智力測驗上得到 IQ 為 130，該測驗的信度為.96，標準差為 15，求測量標準誤之值為何？

$$SE_{meas.} = 15\sqrt{1-.96} = 3 \text{ 以 95\% 為信賴水準，則單邊的標準誤等於 } 3 * 1.96 = 5.88,$$

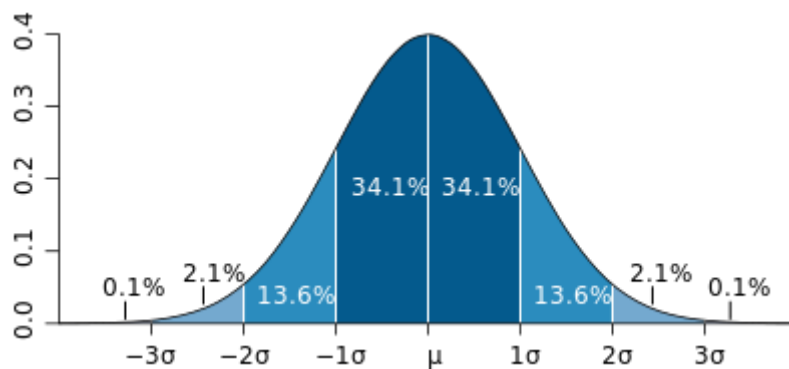
約略等於 6 → 則該生真實分數有 95% 的機會可能落在 124 至 136 之間。

◇ 標準差：統計分布的程度。

如兩組數集合[0,5,9,14]&[5,6,8,9]平均數皆為 7，但後者標準差較小。

◇ 變異數：所有資料到平均數的距離。是「標準差」的平方。





**68%** 的信賴區間：真實分數介於實得分數 正負 **1** SEmeas(標準誤)。

**95%** 的信賴區間：真實分數介於實得分數 正負 **1.96** SEmeas (標準誤)。

**99%** 的信賴區間：真實分數介於實得分數 正負 **2.58** SEmeas (標準誤)。

#### 四、其他標準誤相關

##### (一) 標準誤 Standard Error

由母體中每次抽出若干個體為一組樣本，再由各組樣本求出一個平均數，重複此步驟，反覆進行無數次抽樣，由於進行一次抽樣就會得到一次抽樣誤差，進行無數次就會得到無數次的抽樣誤差，這些抽樣誤差形成的標準差即為抽樣分配標準差，又稱為【標準誤】

##### (二) 估計標準誤 Standard Error of Estimate

根據迴歸線上的已知分數來預測未知分數產生的估計誤差，進行無數次預測就會得到無數次的預測誤差，這些預測誤差形成的標準差即為所謂【估計標準誤】

$$\text{公式：} SE = S_{y.x} = S_y \sqrt{1-r^2}$$

##### (三) 差異標準誤：

1. 同一人在同一份測驗前後兩次測驗比較，同一人在同一測驗之不同分測驗比較或不同的人在同一測驗比較，用以判斷上述情況測驗分數的差異，是否達統計顯著水準，就需使用【差異標準誤】
2. 可利用差異標準誤來解釋個人在兩種測驗上的表現，其優劣是否有差異。

3. 公式：  $SE_{diff} = \sqrt{SE_{meas.x}^2 + SE_{meas.y}^2}$

$SE_{diff}$  表差異的標準誤

$SE_{meas.x}^2$  為 x 測驗的測量標準誤

$SE_{meas.y}^2$  為 y 測驗的測量標準誤

- 舉例：某生參與一項性向測驗，此測驗的分數的  $M=50, S=10$ 。該生的語文測驗得分為 54，此測驗的信度為 .85；其數學測驗的得分為 62，此測驗的信度為 .90，是否真的顯示該生的數學性向高於語文性向？

$$SE_{diff} = 10\sqrt{2 - .85 - .90} = 5$$

→ 若以 95% 為信賴水準，則該生的數學分數須與語文分數相差  $1.96(5)=9.8$  分以上，才可以說其數學性向優於語文性向。

## 五、**效度** Validity

### (一) 定義

概念上的定義：

1. 效度指的是一個測驗能確實測量到它所欲測量的特質或功能的程度，效度越高表示該測驗越能達到它所要測量的目標。
2. 一個心理測驗在評量上最重要的特徵。
3. 若其他條件均相等，樣本的異質性越大，則效度會越大。

計量上的定義

- (1) 一個測驗分數的總變異可以分割成三個部分：

① 共同因素的變異      ② 獨特的變異      ③ 誤差的變異

- (2) 關係為  $S^2_x = S^2_{co} + S^2_{sp} + S^2_e$

- (3) 其中共同因素的變異數  $S^2_{co}$  佔總變異  $S^2_x$  的比率，即定義為「效度」，並以符號  $r_v$  來表示

$$(4) \frac{S^2_{co}}{S^2_x} = \frac{S^2_{co}}{S^2_x} + \frac{S^2_{sp}}{S^2_x} + \frac{S^2_e}{S^2_x}$$

### (二) 效度的性質

1. 效度是指「測驗結果」正確性的程度，而非指工具本身。
2. 效度不是「全有」或「全無」，而是「程度的差別」。
3. 效度是針對某種特殊的用途而言，而不可認為具有普遍性。

- (三) **效度係數 validity coefficient**：指兩種不同測量工具對同一群體施測所得結果間的關係。當係數越接近 1.00 表示工具越有效，通常無切截點存在，來表示測驗有效或無效。當個別測驗之係數小於 0.70；團測小於 0.60 可能應該謹慎的選擇。

### (四) 效度的影響因素

1. **測驗本身的因素**：作答說明不清楚、題目中所用的字彙和句型結構太過困難、試題難度太高或太低、試題編製不佳、題意不清、所編製的試題不適合所要測量的學習結果、測驗題數太少、試題的安排和組織不妥當、試題的正確答案有明顯的組型
2. **教學的因素**：測驗的題目為教師之前所教過則測驗變成記憶測驗而無法有效的測量複雜的學習結果。
3. **測驗實施和記分的因素**
  - (1) 以教師自編測驗而言：沒有足夠的作答時間、學生發問時未對全班說明，僅幫助發問的學生、作弊、申論題沒有客觀的評分標準。
  - (2) 以標準化測驗而言：主試者未按指導手冊的規定來實施測驗、給予不必要的說明、計分錯誤。
4. **受試者的因素**：學生的情緒不佳、害怕考試、動機不強和反應心向等。

## 5. 群體和效標的性質

- (1) 要測量的群體之性質與建立效度資料的群體相近時，這個測驗的效度就較為理想。
- (2) 測量的行為與效標所代表的行為越相似，則效度係數越高。

6. 樣本的異質性：樣本異質性越大，則效度越高。

7. 測驗與效標兩者實施時間之間隔：間隔時間短預測效度係數高，間隔時間長預測效度係數低。

8. 信度的影響：效度係數的最大值與信度高低有直接的關聯。

(1) 從公式來看： $\text{效度最大值} = \sqrt{\text{信度}}$

(2) 從計量的定義來看：增加總變異中「真實變異數」或「減少誤差變異數」的情況下，能有效的提高信度與效度。

9. 預測測驗之題數：題數增加能提高測驗的信度，亦造成效度的提高，不過所新增加的試題必須與原來的試題相當，而且受試者與效標也要保持不變，才能提高測驗的效度。

10. 甄選標準改變的影響：如果甄選標準提高，則群體之同質性增加，就會降低測驗的效度。

## (五) 效度的補救方法

### 1. 放棄不用或重新編擬試題

(1) 就內容效度而言

- ① 藉助雙向細目表，若發現試題不適合，即將其刪除。
- ② 若太多不良題目則應放棄整份測驗。

(2) 就效標關聯效度及構念效度而言：若發現若干試題的效度偏低，應檢討是否

- ① 外在效標選取不當
  - ② 理論構念不周全
  - ③ 試題不佳
- 進行修改或重編。

2. 校正相關係數的萎縮：效度係數有時會被不可靠的測驗分數給萎縮，而造成效度係數降低的情形。所以碰到此情況時應加以校正。

## (六) 效度的複核與效標的混淆

1. 效度的複核：指測驗編製者以新的受試者為對象，重新檢驗原先所編製測驗的鑑別力，其作用是證實原先所選的題目在應用到其他群體時是否同樣有效。
2. 效標的混淆：指評分者事先知道學生的測驗成績後才對學生進行評量，結果使測驗成績與該評分者所評定的成績有假的高相關存在。

## (七) 效度概念在教學評量的應用

1. 利用雙向細目表來編製具內容效度的測驗。
2. 利用具預測效度和同時效度的測驗工具來甄選、分類和安置人員。
  - (1) 甄選—使最適合的人進來。
  - (2) 安置—使每個人被分派去接受最適當的工作或訓練。
  - (3) 分類—使每個人分派至最適合其潛能的位置。
3. 幫助教師建立重要測驗的構念效度：能利用內部一致性分析法/外在效標分析法/因素分析法/多特質多分析法來建立某測驗的效度。

## 六、效度的類型

### (一) 內容效度 content validity

1. 意義：抽樣的測驗試題樣本內容具有代表性，能反映出教學目標或教材內容。
  - (1) 別稱：內容關聯的效度、取樣效度、邏輯效度、合理效度。
  - (2) 測驗試題若能涵蓋所有的教學目標和教材內容，並根據雙向細目表來命題，又具有足夠的代表性試題，即能夠確立該測驗的內容效度的適當性。
2. 測驗目的：確定受試者目前對於測驗中所呈現的代表性樣本之表現程度。
3. 考驗方法：包括邏輯的分析方法和實徵的分析方法。
  - (1) 邏輯的分析方法：請學科或測驗專家針對測驗編製的雙向細目表，仔細判斷每個測驗試題是否與教材內容所涵蓋的範圍及教學目標相符。
  - (2) 實徵的分析方法：依照內容效度係數的計算公式，計算出整份測驗的內容效度係數，再將結果與一個可進行顯著性考驗的機率參照表進行比對。
4. 考量因素有：
  - (1) 測驗項目之適當性：是為適合的年齡水準&內容。
  - (2) 項目樣本之完整性：測驗包括在測量範圍上廣泛之任務樣本。
  - (3) 測驗項目測量內容方法上之適當性：任務形式是採多重選擇或填空、是非題哪種方式。
5. **表面效度 face validity**：一種非技術性的效度形式，包括快速評論測驗已決定其在表現上是否有效之程度，但無法替代完善建立的內容效度。

表面效度係指測驗給人的第一印象「好像」是在測量某種特質的指標，而不是指測驗事實上能測量到什麼特質,但表面效度有時可以激勵學生認真作答。然一份適當的測驗不但要具備表面效度，更重要的是要有內容效度。
6. **現金效度 cash validity**：也非技術性之效度形式，考量重點為商業性測驗之銷售量。但高銷售量並不保證在特定情境上使用的效度。
7. **專家效度**：由專家以雙向細目表之方式來得到效度。
8. 使用的注意事項
  - 避免與表面效度產生混淆

## (二) 效標關聯效度 criterion-related validity

1. 意義：以實證分析方法研究測驗分數與外在效標間**關聯性**的一種指標。預測學生未來的表現或估計學生目前在某些效標表現上的未知狀況。

(1) 別稱：實證效度、統計效度。

(2) 建立效標關聯效度時，不易取得適當的外在效標(測驗所要預測的某些行為或表現標準)。

(3) 當我們再蒐集效標證據以進行效度考驗時，預測力強是我們再選取效標時最重要的考慮。

(4) 效標關聯效度又可分：同時效度、預測效度、區辨效度、聚斂效度。

(5) 效標應具備四項特徵：適切性、可靠性、客觀性、可用性。

2. 效標的種類：學業成就(學校成績、升留級記錄、特殊榮譽、教師的評分等)、特殊化的訓練成績(音樂系和美術系的術科成績、商學院的打字、簿記、會計等)、實際工作表現、對照團體、相關人員的評定、精神病學的診斷、先前的有效測驗。

### 3. 預測效度 predictive validity

(1) 指測驗能預測考試者未來表現的程度。

(2) 例如：某一入學學術性向測驗以國中一年級的數學成績作為外在效標，以計算出該性向測驗的效度。

(3) 對於教師特質的測驗，如果受試者分數越高，後來擔任教師工作表現也越好，則該測驗具有預測效度。

### 4. 同時效度 concurrent validity

(1) 指測驗能評估考試者目前表現的程度。

(2) 例如：一般的 IQ 測驗常以魏氏智力測驗作為外在效標，以計算出該 IQ 測驗的效度。

(3) 最有可能產生「效標混淆/汙染」的現象，即收集效標資料進行效度考驗時，常發現效標資料會受到測驗分數影響。如某教師知道學生的性向測驗分數很低，對此學生效標分數評定也很低。

5. 預期值表 expectancy table：以兩向度之表格來說明效標關聯效度之方式，表的直欄為預測項目 predictor category 或變數，橫欄為效標項目，對於預測項目之每一分項之分數，研究者列出其對應效標項目中分項目的等級(百分比值)。

➤ 建立預期表的方法是：

(1) 先依分類標準將測驗分數與效標分數分成若干層次。例如，在下表中，智力測驗分數依 IQ 高低分為五個層次，效標分數(學業成績)也依高低分成五個層次。

(2) 再依學生在這兩項分數加以分類或劃記。例如，下表中，IQ120 以上而學業成績在 90 分以上者有 8 人，IQ 在 90 到 109 之間而學業成績在 81~90 分者有 12 人。

智力	學業成績					總數
	60以下	61-70	71-80	81-90	90以上	
120以上				// 2	+++ /// 8	10
110-119			+++ / 6	+++ +++ 16	/// 3	25
90-109		/// 4	+++ /// 24	+++ +++ 12		40
80-89	+++ / 6	+++ /// 13	1			20
79以下	+++ 5					5
						100

- (3) 最後，以每一橫行的總人數除每一細格的人數，轉換成百分比。例如，上表中第一橫行第一細格的 2 除以 10，即得下表中的 20%。

IQ	學業成績(%)				
	60 以下	61-70	71-80	81-90	90 以上
120 以上				20	80
110-119			24	64	12
90-109		10	60	30	
80-89	30	65	5		
79 以下	100				

- (4) 預期表建立後，就可根據各層次的測驗分數來預測並得到各項效標分數的機率。例如，從上表中可知，智力在 120 以上者，得到學業成績 90 分以上的機率有 80%，而得到 81~90 分的機率有 20%。同理，智力在 110~119 之間者，得到學業成績 90 分以上的機率有 12%，得到 81~90 分之間的機率有 64%。此種預期表對於學生做各種決定，可以提供很大的幫助，學校可以依據學生各項測驗資料和有關的效標，建立自己學校的預期表，俾便瞭解學生測驗分數。不過會有預測誤差，尤其當建立預期表的人數太少時，誤差更大。
6. **中間效標** intermediate criterion：特殊計畫的訓練成績和專門學校的學習成績均屬中間效標，其僅可以代表工作成功的部分效標，或僅是實際工作表現的預備。
7. **終極效標** ultimate criterion：其可代表個人專業生涯成功的評價，是理想的效標，例如：教師性向測驗的終極效標是實際教學成就。
8. **考驗方法**：效標關聯效度的考驗方法以相關法為主，即求出該測驗與一個外在效標的相關
- (1) 相關的意義：在求兩份測驗的相關時常以皮氏積差相關係數來考驗。其係數數值介於+1 到-1 之間。
  - (2) 相關係數的解釋：相關係數本身並不是一種比率變數，也不是一種等距變數，所以不能說  $r=.80$  是  $r=.40$  的兩倍。
9. **使用的注意事項**
- (1) 留意效標關聯效度產生變化的原因：此種效度是建立在某測驗與外在效標間的相關。因此，當外在因素改變時，如施測情境、施測對象或外在效標的選取發生改變時，我們必須謹慎思考是否還要繼續使用它或做其他的調整。
  - (2) 留意外在效標的涵義：選用的外在效標是「有效又可靠」時，效標關聯效度才具有真正的意義。
  - (3) 留意受試者母群體：當建立測驗效度所使用的樣本不具有推論對象的代表性時，最好放棄使用這種測驗，因為推論可能無效。
  - (4) 留意適當的樣本大小：樣本太小容易造成假象的效度值，間接影響效度推論的正確性。
  - (5) 留意外在效標與預測變項間的不同用法：外在效標是指外在的預期行為標準，與用來預測未來行為變化的變項(即預測變項)是不同的。
  - (6) 留意效度推論的證據：在某個情境下所獲得效標關聯效度的證據，並不一定能夠推論到其他類似的情境上，因此，每次推論時，都需要找尋新證據。
  - (7) 留意不同的預測效果：針對不同的受試者團體，分開來建立不同的效標關聯效度是比較恰當的方法。

### (三) 構念/建構效度 construct validity

1. 意義：測驗能測量理論的概念或特質之程度而言。旨在以**心理學**的理論概念來說明並分析測驗分數的意義，即從心理學的理論觀點，就測驗的結果加以詮釋和探討，亦即根據心理學理論上的構想來編製測驗的內容或選擇試題。
  - (1) 通常用因素分析法、多重特質多項方法矩陣加以驗證。
  - (2) 可以透過內部一致性分析、實驗研究、相關研究、團體差異分析、因素分析等，來提高測驗的構念效度。
2. 考驗構念效度的步驟
  - (1) 針對理論建構進行分析，預測可能的行為表現，並據此設計測驗的題目。
  - (2) 提出可考驗建構與其他變項間關係的預測。
  - (3) 蒐集資料，進行實證分析，以驗證上述的預測或假設是否屬實。
  - (4) 收集其他形態的輔助證據，直到上述的預測或假設得到驗證。
3. 分析受試者解答試題時所需的心理歷程：實施個別測驗，並要求受試者在回答問題時，邊想邊說，以了解受試者在解答問題時，所需經過的心理歷程，是否就是想要測量的部分。
4. 包含三個成分：
  - (1) 信度：量表本身可信的程度
  - (2) **符合效度/聚斂效度 convergent validity**：使用不同方法測量相同特質。一個測驗的分數，應該和測量相同構念或特質的其他測驗分數有高的相關。
  - (3) **區辨效度 discriminant validity**：使用相同方法測量不同特質。即一份測驗分數也要能夠和其他測驗不同理論建構或潛在特質的測驗分數間有低相關。  
→ 如：機械性向的測量分數和在校機械科目的成績算相關，如果是高相關，表示有符合效度。而和語文測驗的成績是低相關，則有區辨效度。
5. **內部一致性分析法**：內部品質是否測量相同的心理特質
  - (1) 相關分析法：分析個別試題與總分間的相關。採二系列或點二系列相關，凡相關係數經考驗達到顯著水準者，即可保留下來。
  - (2) 團體對照法：依據學生測驗總分高低，分為高低分組，然後比較這兩組學生在每個試題上的答對百分比，如果該題的高分組其答對百分比顯著的高於低分組的話，即可保留下來。
6. **外在效標分析法**：以一個有效度的測驗為效標，計算此測驗與新測驗的相關係數。
  - (1) 相關分析法：根據理論來建構出一個測驗，並將此測驗施測後與一個適當的外在效標做相關，經統計考驗後，若達顯著，即表示效標關聯的構念效度獲得驗證。
  - (2) 團體對照法：上過某種教學法與未上過此種教學法的學生、達到精熟的學生與未達到精熟的學生、實驗組與控制組的學生在某個測驗的表現上有明顯的差異，此皆可作為支持該測驗其建構效度的證據。
7. **因素分析法**：從一堆試題中抽出少數幾個共同因素，用以反應或代表這堆試題的共同結構。分析大量變項間的關係，常用在人格量表上。
8. **多特質/多分析法**：以聚斂效度和區辨效度來作為構念效度的證據。

類型	意義	考驗方法
內容效度	測驗的內容能否充分代表其所欲測量的行為領域。	比較測驗的材料和所欲測量的教學目標及材料內容是否一致。
效標關聯效度	測驗成績對目前及未來某一行為的表現(由其他適當的工具測量而得)預測力的高低。	求測驗分數與其他測驗成績之相關。其他測驗成績如在同時測量則為同時效度；如在往後測量為預測效度。
構念效度	測驗的成績能以心理學的屬性來加以解釋的程度。	建立理論架構，以解釋個體在測驗上的表現；跟具理論架構推演出各種假設；收集資料考驗假設是否成立。

#### (四) 內在效度 internal validity ★可信度

1. 意義：指實驗研究者所操縱的實驗變項對依變項所造成的影響真正程度，亦即一個實驗研究能夠有效的實驗出所要實驗的因果關係，若實驗的干擾愈多，正確性愈差，則該實驗的內在效度便愈低，反之干擾變項能完全控制，其正確性愈高，則該實驗的內在效度便愈高。
2. 影響內在效度的因素：
  - (1) 成熟：受試者在實驗期間，不論生理或心理均會產生變化，而可能影響實驗的結果；所以實驗結果的差異，到底是個體變項造成，亦或實驗處理造成，將混淆不清。
  - (2) 臨時事故（同時事件）：在實驗期間，受試者從實驗情境內外，經驗到一些實驗處理以外的特定事件，而影響到實驗的結果。
  - (3) 測驗：實施前後測的實驗設計，測試會影響到受試者，使得個人變得敏感些，或了解研究者所隱藏的目標，或對測驗的表現較為熟練，致使後測成績優於前測，而其結果是因測試的經驗造成或實驗處理造成的不易解釋清楚。
  - (4) 測量工具的變化：後測時所用的測驗器材或評分人員與前測不同，例如測驗卷被寫了答案，評分人員變得較粗心、較不嚴格等，都可能使後測成績發生改變，但卻無法完全歸因於實驗處理。
  - (5) 統計迴歸現象：實驗受試者如為極端選樣，則後測的分數會產生趨近平均數的迴歸現象，而影響實驗的結果。
    - ◇ 統計迴歸效應：受試者在前後測中，第二次測驗成績有向團體平均數迴歸的傾向稱迴歸效應。也就是說，再在第一次測試較差的學生可能在第二次測試時表現好些，而第一次表現好的學生則可能相反，這種情形稱為統計迴歸效應。
  - (6) 選樣誤差：指實驗組與控制組的受試者並不相等，尤以選用自願者或整個班級時，可能會發生此種情形。
  - (7) 受試者亡佚：在實驗期間，受試者因為死亡或退出，會使得實驗組或控制組的樣本變成偏差樣本，而影響實驗結果的內在效度。
  - (8) 選樣與成熟交互作用：實驗組與控制組的樣本不相等再加上成熟因素所造成的影響。
  - (9) 強亨利效應：實驗期間，控制組的受試者為了不甘示弱，力圖與實驗組一較長短而有一般水準以上的表現，此一現象即稱為強亨利效應，其會影響實驗結果的精確性



- (10) **實驗的處理擴散**：若實驗的條件與控制的條件具有很密切的關係或受試者彼此很親近，則控制組的成員可能尋求接觸處理條件的機會，實驗的處理擴散，就很容易會發生。
- (11) **實驗者偏見**：指實驗者精心設計或無意處理，而對受試者產生的影響，如實驗者對於受試者地位的認知，可能影響實驗者作客觀的判斷。
- (12) **統計的結論**：指統計的使用若違背其原理，或誤用統計考驗，則無法獲得有效的結果，而影響實驗結果的內在效度。

#### (五) 外部效度 external validity ★可轉換性

1. 意義：指實驗結果的可推論性之程序的大小，實驗結果的可推論性愈大，亦即其適用性，代表性愈大，實驗的外在效度則愈高。易言之，就是指實驗研究結果是否可推論到研究對象以外的其他受試者，或研究情境以外的其他情境。
2. 獲得研究結果是否能推論/一般化 generalizability
  - (1) 母體效度 population validity：樣本推論至所代表母體的程度。
  - (2) 生態效度 ecological validity：樣本推論至其他母體的程度，外在效度的一種，指情境的可推論性，亦即根據研究情境的研究結果能否推論到不同情境的信心程度。
3. 應避免霍桑效應、受測者的抗拒作用等，以提高研究的可推論性。

#### (六) 交叉驗證

1. 透過不同的測驗樣本對已存在的測驗效度再次進行驗證之過程，以瞭解該測驗效度的正確性。
2. 應確保測驗樣本的代表性與異質性。
3. 效度縮減 validity shrinkage：指測驗對新的受試樣本的相關效標再測值的正確性比原始樣本低的常見現象。

(七) **決策效度**：指該測驗能夠預測未來組織運作結果及個人工作結果，並作為與人事任用有關的決策參考依據，其多採用效標關連效度執行之。

#### (八) 增加效度

1. 意義：指測驗剔除掉若干對依變項有影響的因素後，仍可對此依變項具有若干程度的解釋力。
2. 可使用階層迴歸分析及半淨相關等統計策略，來提高「增加效度」。

#### (九) 後果效度 consequences validity

1. 1989 年由 Samuel Messick 提出
2. 包含價值的隱含意義 value implications、社會的影響 consequences 兩個新概念，其爭議為此構想的實用性。
  - (1) 價值的隱含意義：須評估所建構的標籤、據以解釋測驗結果背後之理論、其背後所隱含意義的價值、該理論本身所包括的意識型態等上述幾點價值意涵。
  - (2) 社會的價值：評估所運用的測驗其潛在的與實際產生出來對社會之影響。

## 七、信度與效度之關係

1. 信度指測驗結果的一致性和穩定性；效度是指測驗能夠測量其所欲測量的特質之程度；兩者皆為良好測驗的特徵。
2. 效度的考驗方法包括邏輯的分析與統計的應用，而信度的考驗則完全採用統計方法。
3. 信度是效度的必要條件而非充分條件，測驗要有效度必須要有信度。但是有信度並不一定有效度。



射箭圖的信度與效度關係

4. 信度高，效度不一定高；信度低，效度一定低；效度高，信度一定高；效度係數一定不會大於信度係數的平方根。

# 常模及標準分數

## 一、常模

### (一) 常模的意義

1. 常模指的是特定參照團體在測驗上所獲得的平均分數，所以亦可視為平均數。
2. 常模參照的解釋便是將個人的表現與團體的他人做比較。  
→ 若只知某學生考 85 分，並無法得知其成績是好或壞，須將此分數參照常模來解釋，才能確定該分數在團體中的相對地位。

### (二) 常模的功用

1. 表明個人分數在常模團體中的相對位置。
2. 提供比較的量數，以比較個人在不同測驗上的分數。

### (三) 常模的類型

1. 全國性常模：依據全國性的代表樣本而建立之常模
2. 地區性常模：依據同地區學生來建立的常模
3. 特殊團體常模：依據具有某些特殊行為或心理特徵之族群來建立的常模
4. 學校平均數常模：依據同學校的學生來建立的常模

### (四) 常模的適切性：在選擇測驗與解釋測驗分數時，依據下列三個標準來判定常模適當性：

1. 新近性：常模需為最新的。
2. 代表性：樣本的代表性、抽樣方法。
3. 適切性：常模的樣本是否所欲施測的對象相類似。

## 二、常模的建立方法

### (一) 發展性常模 developmental norms：根據個人所獲得的發展水準所得之常模。

#### 1. 年齡常模

- (1) 年齡常模是指依不同年齡在測驗上所得的平均數  
在測驗得分為 8 分，閱讀年齡 9 歲 2 個月，得 18 分閱讀年齡為 12 歲 6 個月。
- (2) 年齡常模適用於會隨著年齡增長的特質，如身高、體重、智力和閱讀能力等。
- (3) 缺點：心理與教育特徵的成長，各年齡階段不一致，所以心理年齡沒有一致的單位。

年齡常模表

原始分數	閱讀年齡	原始分數	閱讀年齡
0	7-8	13	10-6
1	7-10	14	10-8
2	7-11	15	11-2
3	8-1	16	11-8
4	8-5	17	12-1
5	8-7	18	12-6
6	8-9	19	12-8
7	9-0	20	12-11
8	9-2	21	13-5
9	9-6	22	13-11
10	9-9	23	14-5
11	10-2	24	15-1
12	10-4		

## 2. 年級常模

- (1) 年級常模是指依不同年齡在測驗上所得的平均數。
- (2) 測驗得分為 22 分，年級等值三年級 7 個月，得 38 分年級等值為五年級 7 個月。
- (3) 年級常模適用於隨年級而逐漸增加的特質，例如小學的各種基本技能的學習。
- (4) 注意：年級常模單位不相等、不同學科的年級常模具有不同的意義。

原始分數	年級等值	原始分數	年級等值
19	3.2	31	4.7
20	3.4	32	4.8
21	3.6	33	5.0
22	3.7	34	5.2
23	3.8	35	5.3
24	3.9	36	5.4
25	4.0	37	5.6
26	4.1	38	5.7
27	4.2	39	5.9
28	4.3	40	6.0
29	4.4	41	6.1
30	4.6	42	6.2
		43	

(取自 Thorndike, 1977)

(二) 團體內/組內 within-group norms：根據個人在特殊團體中的相對位置所建立的常模。

### 1. 百分等級

- (1) 將一個團體分成 100 等分，此等分點稱之為百分點
- (2) 百分等級是指各百分點以下包含有多少百分比的人數，百分等級代表的是等第
- (3) 百分等級的求法

樣本數不多時，可將分數依高低排列給名次，再依公式求百分等級。

$$PR = 100 - \frac{(100R - 50)}{N}$$

PR: 百分等級；N: 全體人數；R: 名次

➤ 例題：五位學生成績依高低排列為：93, 84, 75, 62, 46。則第一、二名的百分等級為？

$$PR_{1st} = 100 - \frac{(100 \times 1 - 50)}{5} = 90$$

$$PR_{2nd} = 100 - \frac{(100 \times 2 - 50)}{5} = 70$$

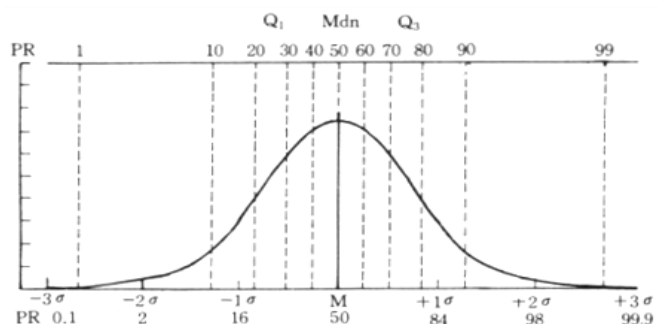
### (4) 百分等級的優點

- ① 易於瞭解
- ② 適用於各種測驗、各種對象（成人及兒童）
- ③ 不論原始分數為何百分等級都可正確解釋分數

### (5) 注意事項

- ① 百分等級與百分比不同。
- ② 百分等級的單位不相等，屬於順序量尺而非等距量尺。
- ③ 分佈於兩端的差異必分佈於中央的差異顯著。順序量尺亦不能進行運算。
- ④ 百分等級分佈：

1 ~ 99



常態分配的百分等級

## 2. 標準分數

直線標準分數	常態化標準分數
Z 分數	T 量表分數
T 分數	標準九

### I. 直線標準分數

#### (1) Z 分數

- ① 意義：最基本簡單的直線標準分數是 Z 分數，它可以表明原始分數在平均數之上或之下幾個標準差。

② Z 分數公式  $Z = \frac{X - M}{S}$

X:原始分數 M:平均數 S:標準差

- ③ 缺點：有負數及小數點，使用不便。

Z分數的計算方法

$Z = \frac{X - M}{S}$	智力測驗：	M = 40	S = 5
甲 $X_1 = 45$	乙 $X_2 = 35$	丙 $X_3 = 32$	丁 $X_4 = 40$
$Z_1 = \frac{45 - 40}{5} = +1.00$	$Z_2 = \frac{35 - 40}{5} = -1.00$	$Z_3 = \frac{32 - 40}{5} = -1.60$	$Z_4 = \frac{40 - 40}{5} = 0$

#### (2) T 分數

- ① 由於 Z 分數有小數點及負數，為了使用簡便，可以強制令其平均數為 50，標準差為 10，使得 T 分數為正整數。

② T 分數公式  $T = 10Z + 50 = 10 \frac{X - M}{S} + 50$

Z 分數	T 分數
甲 +1.00	$10(1.00) + 50 = 60$
乙 -1.00	$10(-1.00) + 50 = 40$
丙 -1.60	$10(-1.60) + 50 = 34$
丁 0	$10(0) + 50 = 50$

## II. 常態化標準分數

### (1) T 量表分數

① 最普遍使用的常態化標準分數是 T 量表分數。與 T 分數一樣是以 50 為平均數，10 為標準差，不同之處在於使用 T 量表分數必為常態分配。

② 計算步驟

- A. 將原始分數轉換為百分等級。
- B. 依常態分配表，查出各百分等級的相對 Z 值。
- C. 利用  $T=10Z+50$  加以轉換，即為 T 量表分數。

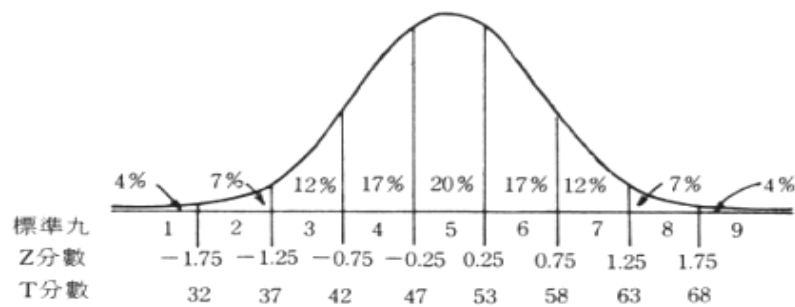
### (2) 標準九

① 將原始分數分為九等分（1-9），這樣標準分數的平均數即為 5，而標準差接近 2。

② 標準九常模的使用逐漸增加，尤其在性向測驗與成就測驗。

③ 特點

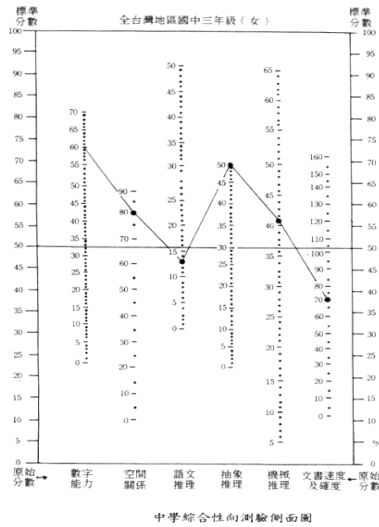
- A. 只用個位數表示分數，資料處理較方便。
- B. 標準九的意義容易被瞭解。
- C. 屬於常態化標準分數，可用於比較不同測驗的分數。



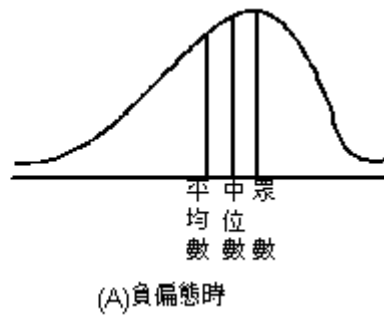
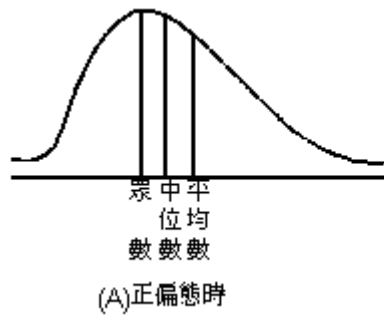
標準九的常態分配機率

### 三、測驗分數的解釋

#### 1. 側面圖分析



#### 2. 正偏態與負偏態



#### 3. 解釋測驗分數應注意事項

- (1) 解釋測驗者應瞭解測驗的性質與功能。
- (2) 測驗分數應為學生保密。
- (3) 解釋分數應參考其他有關資料。
- (4) 解釋分數應避免只給數字。
- (5) 對低分者的解釋應謹慎小心。
- (6) 解釋分數時應設法瞭解學生的感受。
- (7) 解釋分數只做建議，勿做決定。
- (8) 應以一段可信的範圍解釋。

# 各式教育測驗

## 一、測驗分類

分類標準	測驗形式	意義	測驗內容
根據不同教育目標	<b>認知測驗 (能力測驗)</b>	指測量有關個人思考、知識、問題解決等心理能力的測驗。	智力測驗、性向測驗、成就測驗 創造思考測驗
	<b>情意測驗</b>	1. 旨在測量個人的人格態度、動機、情緒、興趣、價值觀、自我觀念等特質。 2. 常見的情意測驗有兩類： (1) 專門用來測量個人對人事物等的看法、動機、興趣、價值觀等的態度測驗，例如民意調查 (2) 專門用來測量個人的人格特質、個性等的人格測驗，如：明尼蘇達多項人格量表 <b>MMPI</b> 。	社會計量法 人格測驗 投射測驗 行為觀察
	<b>動作技能測驗</b>	1. 指測量有關人的手、腳及腦等協調反應的測驗。 2. 多半是使用實作測驗的方式來進行。	檢核表 評定量表
根據不同編製過程的標準化程度	<b>標準化測驗</b>	1. 指由測驗專家根據測驗編製程序而編成的測驗。 2. 通常有一定的編製程序，包括試題適當取樣、明確的施測指導語和施測程序、計分標準、解釋分數的常模、以及信度和效度等指標資料。	智力測驗 性向測驗
	<b>非標準化測驗</b>	1. 由教師以非正式的方式，依自己教學的需要而自編的測驗，故又稱教師自編測驗，或非正式測驗。 2. 測驗的編製過程、實施、記分與解釋，較缺乏標準化，故比較主觀。重要特色是能考慮教學目標和班級個別差異的性質。	教師自編測驗
根據不同試題類型	<b>選擇反應測驗 (客觀測驗)</b>	凡是測驗題目以選擇題、是非題、填充題、配合題及解釋性習題的形式來命題者即屬，這類測驗的作答方式，是將正確或最佳的選項選出來。	大部分心理測驗
	<b>結構反應測驗</b>	1. 以簡答題、限制反應題、申論題或問答題的方式來命題，其答案大都是一段文字或一篇論文 2. 又稱為補充型試題測驗或論文測驗。	主題統覺測驗 羅夏克墨漬測驗 托倫斯創造性思考測驗。
根據測驗的評分方式	<b>客觀測驗</b>	指計分有標準或有正確答案可資遵循的測驗，此種測驗的計分不會因計分者的不同而有差異。	選擇題 是非題
	<b>非客觀測驗 主觀測驗</b>	指計分無固定標準答案可資遵循的測驗，它的評分會因評分者的寬嚴不同而有差異。	口試、論文測驗、投射測驗



根據測量分數的解釋方式	<b>常模參照測驗 NRT</b>	<ol style="list-style-type: none"> <li>1. 指測驗的結果，根據測驗分數在團體中的相對位置而加以解釋的一種測驗。換句話說，凡是將測驗結果，拿來與他人做比較的測驗。</li> <li>2. 目的在區別學生間的不同成就水準，並給予學生的學習成就評定等第；通常也適合在行政上做決策，如分班編組之用。</li> </ol>	專為月考、期末考、學科成就競試而編製的教師自編測驗，多數的標準化成就測驗和性向測驗等
	<b>標準(效標)參照測驗 CRT</b>	<ol style="list-style-type: none"> <li>1. 指參考教師在教學前所訂定的標準，來解釋個別測驗分數是否達成這項既定標準的一種測驗。</li> <li>2. 目的在於瞭解學生的學習精熟水準，以作為教學決定之用。</li> </ol>	專為平時考、隨堂測驗、小考或專為某種診斷用途而編製的教師自編測驗
根據施測人數	<b>個人測驗</b>	<ol style="list-style-type: none"> <li>1. 在同一時間內，只能實施於一個人的測驗。</li> <li>2. 實施個人測驗耗時費力且步驟繁雜，所以僅在有特殊目的，如診斷、個案研究或臨床方面才使用。</li> </ol>	比西量表 魏氏智力量表
	<b>團體測驗</b>	一次可同時施測許多人。	目前所能見到的測驗，大部分都是團體測驗
根據測驗的材料	<b>文字測驗</b>	題目是以文字來呈現，其作答說明與答題反應也都使用文字或語言。	測量於同文化的測驗對象
	<b>非文字測驗</b>	<ol style="list-style-type: none"> <li>1. 題目不用文字來呈現，而是以圖形或實物(如方塊、積木或儀器等)為測驗素材。</li> <li>2. 非文字測驗如屬於操作物體或儀器者，又稱為作業(實作)測驗，例如，拼圖、走迷津、操作儀器、打字、演奏樂器和烹調食物等。</li> </ol>	測量特殊性向(文盲、智能不足者或幼兒) 比較不同文化間人們的心理特質
測驗上的反應形態	<b>最大表現測驗</b>	<ol style="list-style-type: none"> <li>1. 旨在測量個人的最佳反應或最大成就，亦即在確定個人的能力表現。</li> <li>2. 假定所有的受試者有相同而強烈的動機，測驗結果顯示出，當個人盡最大能力發揮時，所表現的程度如何。但是，也會因為個人動機不強，注意力缺乏，或是其他因素的關係，使表現的程度遠在潛力之下。</li> <li>3. 在測驗的分數中，至少有先天能力、實際能力和動機三項決定的因素。分數愈高，能力愈佳。</li> </ol>	智力測驗、性向測驗與成
	<b>典型表現測驗</b>	<ol style="list-style-type: none"> <li>1. 旨在測量個人的典型行為，亦即在正常的情境下，個人通常所表現的行為如何。</li> <li>2. 須假定所有的受試者都很誠實回答問題。但通常會有偽飾作答現象，故典型表現的測量常較困難。</li> <li>3. 並沒有所謂好的分數。例如，對音樂有興趣或沒有興趣，並不代表有興趣就是好，沒有興趣就壞</li> </ol>	人格測驗、興趣測驗、態度測驗與適應測驗

根據測驗的時限	速度測驗	主要在測量速度的快慢，亦即在有限的時間內，工作的反應速度。	測驗難度低，少見於教育測驗中
	難度測驗	旨在評量學生獲得知識與技能的程度，非反應的快慢	成就測驗

## 一、智力測驗

### (一) 智力測驗的發展

1. **生理計量階段**：19 世紀中葉，英國高爾頓 Galton 相信人類的智力可由其感覺器官的敏銳度作鑑定。稱為生理計量法 **biometric method**。因不切實際而未被接受。
2. **心理年齡階段**：1904 年法國教育部請比奈及西蒙設計一種鑑別學童能力的工具—**比西量表 Binet-Simon Scale**，適於 3~13 歲兒童。比西量表有兩大特點：
  - (1) 放棄生理計量法，改採作業法(根據語文、數學、常識等題目)來判定智力高低。
  - (2) 創用**心理年齡(或智力年齡，簡稱 MA)**代表智力高低。心理年齡高於**實足年齡 CA**者智力較高。

✧ 比奈 Binet 被稱為「**智力測驗之父**」。
3. **比率智商階段**：美國斯坦福大學教授推孟修訂比西量表為**斯比量表 Stanford-Binet Intelligence Scale**，將心理年齡作為智力改為以智力商數(智商、IQ)來代表智力。

$$IQ = \frac{MA}{CA} \times 100$$

IQ：代表智商  
 MA：代表心理年齡  
 CA：代表實足年齡

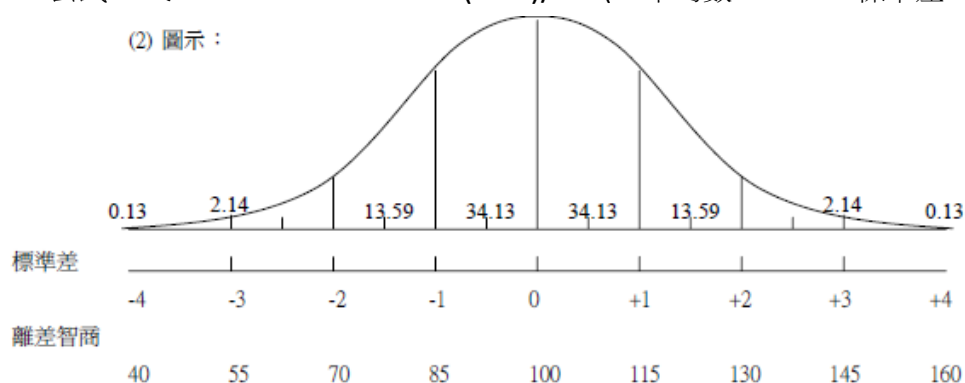
公式：智商的計算公式為：

Ex: 甲生現年 10 歲、心理年齡 150 個月，請問 IQ?  $150 (\text{心齡}) / 120 (\text{現齡}) \times 100 = 125$

4. **離差智商階段**：美國魏克斯勒 David Wechsler 創用，創用離差智商的計量方式。每個年齡階段內全體的智力分佈為常態分配，個人在同年齡組中的位置換算成標準分數，再與平均數的距離有幾個標準差，從而判智力的高低。平均數 100, 標準差 15。(或：每個年齡階段的智力分佈為常態分配，個人分數在其年齡組分佈中離開平均數的距離，用標準差來表示。)

公式： $IQ = 100 + 15Z = 100 + 15(X - M) / S$  (M 平均數=100；S 標準差=15)

(2) 圖示：



## 5. 斯西量表 v.s. 魏氏量表

	斯比量表	魏氏量表
相同	1. 均屬個別智力測驗，不但可用於預測學業成就，且適用於臨床診斷。 2. 花費較大，不夠經濟。 3. 實施程序複雜，需要受過專業訓練者，始能為之。	
相異	1. 第一、二版為年齡量表，第三、四版為點量表。 2. 設計是為測驗兒童智力，較適用於 14 歲以前的兒童。 3. 僅提供單一的智商。 4. 斯比智商是以平均數 100，標準差 16 加以轉換的	1. 點量表。 2. 編製是為測驗成人智力，較適用於 16 歲以後的成人使用。 3. 提供多項的智商，包括語文智商、作業智商、全量表智商。 4. 離差智商是以平均數 100，標準差 15 加以轉換的。

## (二) 智力測驗類別

### 1. 個人智力測驗：

- (1) 每次施測一人，大多為綜合性的成套測驗或測驗組合，可同時測量各種不同的能力，並評鑑受試者各種能力發展上之優劣。用來檢測資賦優異或是心智功能低下的兒童之工具。
- (2) 內容常包括語文、非語文或操作等項目。
- (3) 測驗者可觀察並紀錄受試者測驗情境中的反應，以作為解釋測驗分數的參考。施測者須接受嚴格的專業訓練。

### 2. 團體智力測驗：

- (1) 可團體施測，形式多為紙筆測驗和機器評分，具有經濟、方便之優點。
- (2) 施測者毋須接受臨床訓練。
- (3) 測驗功能主要為提供篩選和分類所需之一般性資料。
- (4) 主要關注的是個體在團體內的智商 PR 值。
- (5) 第一個團體智力測驗：AGCT 美國陸軍普通分類測驗，包含語言理解、算數推理測驗、方塊計算測驗。 $AGCT=20Z+100$

3. 特殊性智力測驗：針對測量特殊能力或測量某一特殊人口之智力所設計之測驗。視其性質和實際狀況需要決定測驗形式。

## 二、性向測驗

1. 性向的定義：一個人在某些活動或領域中（如學科、職業等）所具有的潛在能力，而且此能力本質上含有認知性。

2. 性向測驗的性質：

(1) 為測量特殊能力或能力群而專門設計的工具，可以用來預測一個人在某些學科或職業上成功的可能性。

(2) 不像智力測驗同時測量多種心理功能，並從其中求出智商來代表總體智力；性向測驗的目的在於將各種類別的性向予以客觀而明確地區分。

(3) 大部份的性向測驗包含多個分測驗，每個分測驗測量不同的性向，此類測驗應具有良好的內在效度與外在效度。

① 內在效度：代表每一分測驗在所測的性向因素上有顯著的因素負荷量，以表示它對該性向之測量具有效度，且各分測驗間不具有顯著相關。

② 外在效度：指每一分測驗與其外在效標（例如學習成就或工作表現等）具有高相關，如此才能夠預測受試在效標上的表現。

因此，區分效度和預測效度是性向測驗最應具備的條件。

3. 性向測驗的分類：

(1) 綜合或多元性向測驗：為一種測驗組合，同時包括幾種分測驗，每一分測驗測量一種基本性向，各分測驗的常模以同一標準化樣本建立，因此這類測驗可進行個體內與個體間性向差異之分析。

(2) 特殊性向測驗：測量某些特殊能力，多為單項測驗，只測量一種能力或性向，用來預測在某些學科或行業上成功的可能性。

(3) 教育性向測驗：測量與各學科學習活動有關的各種潛在能力，以用來預測受試者在學習成就上的可能性。

(4) 職業性向測驗：測量從事各行業所需的特殊能力，針對每種職業所需的技能設計

4. 性向測驗的用途

(1) 個人使用：輔導人員可利用性向測驗結果分析受試各項能力的發展，協助受試了解其長處和短處，並參考個人的興趣、志向、動機、以及其它條件，做成有關就學或就業的決策，例如高中學科的選組、大學選系或選擇要從事的職業等。

(2) 團體使用：團體性單位（如學校、公司）利用性向測驗來協助其完成人員揀選、分類和安置。

### 三、成就測驗

1. 成就的定義：成就是指一個人在接受教育或職業訓練過程中，所獲得的成果。
2. 成就測驗的性質：成就測驗為測量個體在各方面學習成就上所達到的結果和程度。廣義而言，它包括學校內的學科測驗、訓練場所評量訓練結果的測驗、大學聯考、高普考等。使用成就測驗可了解個人在某一學科上所吸收的知識有多少、或在一般綜合科目學習上所達到的水準，並且將這些結果與他人比較。
3. 成就測驗的種類
  - (1) 以內容分類：可依學科或課程內容區分，如數學測驗、英文測驗。
  - (2) 以測驗的用途或功能區分：可分篩選性測驗和診斷性測驗。
    - ① 篩選性測驗：除了報告學生的學習成就外，還可區分出成績特優或成績低落的學生。
    - ② 診斷性測驗：多半為單科測驗，藉由不同學習單元的分測驗，例如數學測驗中包括加法（進位或不進位）、減法、乘法（整數或小數）等單元，可藉由分析學生在各單元上作答情形，了解學習的困難所在，並據以設計補救教材。
  - (3) 以編製過程區分：可分為標準化測驗和教師自編測驗。
    - ① 標準化成就測驗：經過標準化程序，有大規模的常模，可適用於各個學校，通常涵蓋面較廣，可比較個人各項學科的差異、個體間差異以及學校、學區的差異。
    - ② 教師自編測驗：多半範圍較小，題目形式較有彈性（如是非題、選擇題、及問答題），在評量一小階段的學習與教學成果較為方便。
  - (4) 以測驗原理區分：分為常模參照測驗與效標參照測驗。
    - ① 常模參照測驗是以受試者的分數與測驗常模比較，以決定其表現與他人相較的高低。
    - ② 效標參照測驗主要目的是決定受試者對所學之教材內容學到了多少，因此，是將受試者的分數與事先設定之精熟標準比較。
4. 成就測驗的用途
  - (1) 考查及報告學習成就
  - (2) 篩選和分類
  - (3) 診斷和矯治
  - (4) 評估教育計劃

	性向測驗	成就測驗
測量的能力	偏向學習新事物的能力，亦即重視未來表現的潛能。	偏向已經習得的能力，強調目前已獲得的知識技能水準。
測量目的	預測未來的表現水準。	確定個人從過去到現在的學習所獲得的成就水準或熟練程度。
測量的內容	取材範圍較廣泛，比較普遍性。	內容僅限於學校教學的特定內容或學習經驗。
測量的時機	教學前測量。	教學後測量。

## 四、人格測驗

(一) 定義：旨在測量一個人各方面的行為特性，其內容相當廣泛而複雜，而人格所包括的態度、興趣、道德、情緒等各方面的特性，不但不易直接衡量，且很難制定完全客觀的評定標準，故屬於主觀的測驗。

### (二) 人格測驗種類

	測驗	例子
依測驗的方法分	1. 自陳測驗 self-report test	吳偉士個人事蹟表格
	2. 情境測驗 situational test	莫里諾的社會關係測驗
	3. 投射測驗 projective test	莫雷的主題統覺測驗
	4. 評定量表 rating test	卜氏兒童社會行為量表
依測驗的內容分	1. 興趣測驗	愛德華個人興趣測驗
	2. 情緒測驗	明尼蘇達多項人格測驗
	3. 性格測驗	塞斯通性格測驗
	4. 社會行為測驗	莫里諾社會關係測驗
	5. 道德意識測驗	

#### 1. 自陳測驗 Self report test：

(1) 定義：依據所要測驗的人格特性，編制許多問題要受試者根據個人的感覺或實際狀況回答，根據答案衡量受試者此種特性所表現的程度。記分客觀、方便，在人格測驗中應用最廣，但容易隱藏事實。

(2) 種類：

- ① 只測量一種人格特質得到一項分數，如：吳偉士個人事蹟表格。
- ② 測量好幾種不同的人格特質，得到好幾種不同的分數，如：少年人格測驗、國民中學人格測驗。
- ③ 根據因素分析研究的結果而編制，如：塞斯通性格測驗。
- ④ 根據人格理論所編製之測驗，如：愛德華個人興趣量表。

(3) 特點：

- ① 測試工具一般是調查表。
- ② 題目編製的數量多。
- ③ 大都包含多個分測量表，以同時測量多維度的個性特徵。
- ④ 易受測試對象和測試形式的影響，通常採用一定的措施控制這種影響。
- ⑤ 多採用紙筆的形式。
- ⑥ 可以測試個體亦可測試團體。

(4) 優點：可團體施測，比其它方法省時省力

(5) 缺點：填答者需要具備閱讀能力，不適用於幼童及視障者&容易受社會期望的影響。

(6) 編製自陳量表的方法：

① 經驗法效標計分法 empirical criterion keying：

- A. 依照題目和效標見的經驗關係編製題目。
- B. 編製的三個步驟：選擇效標組和對照組→編製題目→將題目實施與效標組和對照組。如果題目能夠鑑別兩個不同的效標團體，這些題目就保留下來做為量表的試題。
- C. 局限性：原始效標對題目的效標影響較大；編製的量表缺乏理論依據。
- D. 如：明尼蘇達多項人格量表 MMPI、加州心理量表。

② 因素分析法：

- A. 因素分析法是指同一量表中的題目都應該有較高的相關性，較高的內部統一性。
- B. 採用因素分析法時，應刪除與其他題目沒有相關性的那些題目
- C. 局限性：要具備大量的經驗資料和大量的題目。
- D. 如：石爾頓性格量表。

③ 內容效度：

- A. 題目內容顯示出測驗編製者所要測量的人格特質。
- B. 如：修訂孟氏行為困擾調查表、少年人格特質。

④ 人格理論法：

- A. 測驗的編製是以人格理論為依據，而非以經驗效標為基礎。
- B. 編製者依人格理論詳係界定各項人格特質的意義，然後編擬可以代表這些特質的題目，同時依人格理論加以計分。如在興趣量表中，假定工程人員對閱讀設計圖有興趣，在「我喜歡閱讀設計圖」的題目中，答「是」的，則在工程興趣表上加一分。
- C. 如：愛德華個人興趣量表。

(7) 應注意事項

- ① 題目的文字敘述應就受試者的有利立場或觀點去編擬。
- ② 舉行測驗前，要使受試者瞭解測驗的目的。
- ③ 告知受試者所作的答案和學業成績或操行考核無關。
- ④ 由受試者所信任的人擔任主試，並且保證結果一定守密。
- ⑤ 個別輔導自陳測驗需記名作答；團體統計研究之自陳測驗無須記名，信度較高。
- ⑥ 採強迫選擇式項目，使受試者在兩項同等願望的特性間強擇其一，使受試者無法偽飾。
- ⑦ 隱蔽測驗目的，不要使受試者能從測驗名稱上知道測驗分數要作何解釋。
- ⑧ 設證實分數以檢查受試者的答案是否可靠。
- ⑨ 擬題時所用的字詞需適合受試者的閱讀程度
- ⑩ 宜作為輔導的工具，不適合當作准許入學或選拔升遷之依據。
- ⑪ 只是考查人格特質的方法之一，自陳測驗所獲得的各種資料和結果，應當與其他資料相互校核和補充。
- ⑫ 測驗之後，應依據測驗所得之資料進行輔導工作。

## 2. 情境測驗 situational test：

(1) 受試者在設計的實際情境中，從事各種活動，以觀察受試者在此種情境中所表現的實際行為，再對其人格特性加以評判。如：莫里諾的社會關係測驗。

(2) 種類：

① 日常情境測驗 **everyday life test**：布置的情境與日常生活情境相似，如：哈特宋與梅爾的誠實行為測驗；鮑加都之社會距離量表；莫里諾的社會關係測驗。

② 緊張情境測驗 **situational stress test**：布置的情境有各種不同的壓力(如電擊、困擾、失敗、非難等)，如：美國戰略勤務局甄選人員用的測驗

③ 作業情境測驗 **artificial-task test**：根據受試者的作業情形，可以間接評判某種人格特性，如：包、魏兩氏的藝術測驗、塞斯通的文字流暢測驗

3. 投射測驗：提供受試者一種曖昧的刺激情境，讓受試者在不受限制的情形下，自由反應，因此容易將內心隱藏的問題流露在反應中。種類：

(1) 聯想法 **associative technique**：一連串不相連續的字或墨漬，讓受試者自由聯想  
如：榮格的文字聯想測驗和羅夏默漬測驗。

(2) 構造法 **constructive procedures**：受試者根據看到的圖畫，編造一套含有過去、現在、將來的故事。如：莫雷的主題統覺測驗 **T.A.T**。

(3) 完成法 **compleitive test**：提供受試者不完整的文句、故事，讓受試者自由補充完成。如：羅特之未完成的語句測驗。

(4) 選組法 **choice or ordering devices**：受試者根據特定準則（如美觀、自由），做各式各樣的選擇或組合。如：宋迪測驗 **Szondi test**。

(5) 表達法 **expressive test**：以繪畫、遊戲、心理戲劇等方式，從受試者作畫或玩木偶等行為為探測。如：卡爾柯乞的畫樹測驗。

4. 評定量表：(見情意評量)就欲測量的人格特性，列舉許多有關的項目或問題，由觀察者根據量表中的項目或問題，選擇和受試者日常行為最符合的一項，加以劃記，化為數量予以評定。但易有月暈效應。如：卜氏兒童社會行為量表、猜人測驗。



# 教學評量

## 一、教學評量的概念

1. 意義：欲瞭解學生的行為改變是否如預期的教學結果，就必須在教學後，做客觀而正確的評量。
2. 內涵：
  - (1) 教師教學效率的評量：係以教師為評量的對象，就教師教學方法和技術加以評量，可以包含教學活動設計、教學情境布置、教材的編選、教法的運用、教學進度的掌握、師生參與活動情形、作業的規定和批閱校正、教室管理和常規訓練等項目。
  - (2) 學生學習成就評量：係以學生為評量的對象，旨在評鑑學生的學習行為和學習結果。包括
    - ① 學習前評量：以瞭解學生的個別差異及起點行為，例如：智力、性向、興趣、人格特質、家庭背景。
    - ② 學習中評量：以瞭解學生的學習動機、興趣、態度、方法、習慣及努力情形
    - ③ 學習後評量：即考察學生學習成績，分析其優、缺點，並進而診斷其學習困難所在及原因，以作為實施補救教學及個別輔導的參考依據。
  - (3) 課程設計與實施的評量：係以師生共同參與的課程與教學活動為主，評鑑學校課程計畫與實施的利弊得失，再加以檢討改進，期能有較佳的均衡課程。
3. 功能
  - (1) 透過教學評量可協助教師獲知學生的起點行為。
  - (2) 經由教學評量可診斷學生學習的困難與障礙，作為補救教學或個別輔導的依據。
  - (3) 可瞭解學生的潛能與學習成就，以判斷其努力的程度。
  - (4) 可估量教師教學的效率，以作為教師改進教材、教法的參考。
  - (5) 學生自我評量的結果，可獲悉學習進步的情形，以激發其學習的動機。

### 4. 教學評量計畫 3 步驟

確定教學目標 → 細列具體學習結果 → 設計評量的技術

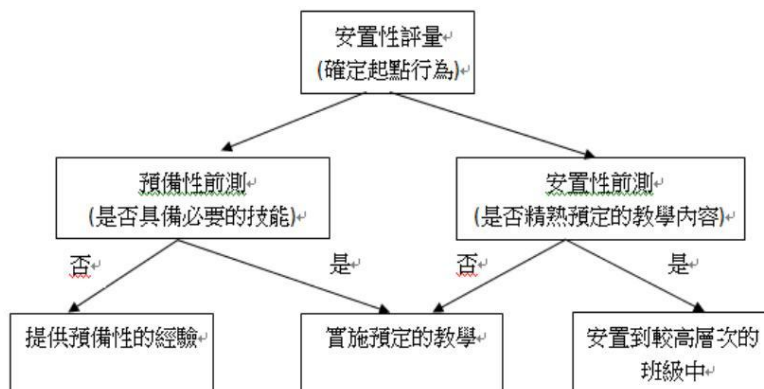
## 二、教學評量的種類：

1. 依據評量的目的可以分成四種：安置性評量、形成性評量、診斷性評量和總結性評量；
2. 依教學歷程可分成：(1)教學前的安置性評量、(2)教學中的形成性和診斷性評量、  
(3)教學後的總結性評量；
3. 依評量結果的解釋可以分成兩種：常模參照評量和效標(標準)參照評量；
4. 依據評量所使用的工具和形式來分，可以分成兩種：紙筆測驗和實作評量。
5. 依評量的時機分：形成性評量、總結性評量
6. 依評量的功能分：安置性評量、診斷性評量

## (一) 教學評量的解釋 (依目的)

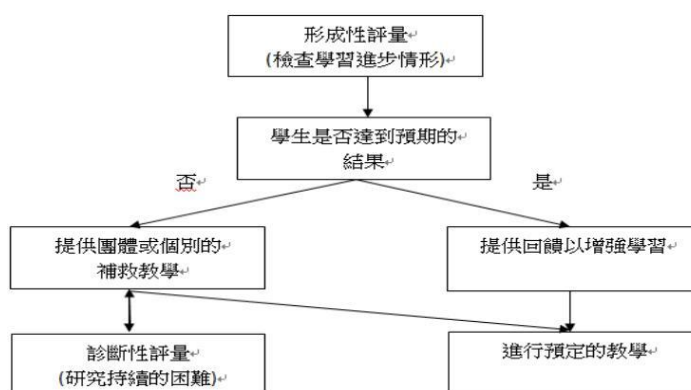
### 1. 安置性評量

- (1) 定義：指教學前對學生所具有的起點行為之評量。它關心的是學生在教學開始前所具有的一些知識、技能、興趣、習慣、人格特質等。
- (2) 目的：在確定學生已經精熟多少預定教學的內容。



### 2. 形成性評量

- (1) 理論基礎：教學歷程要和評量歷程相互結合，以期改進教學，提高學習效果。
- (2) 定義：指在教學活動進行當中，對於教師的教學及學生的學習表現的評量。
- (3) 目的：提供教師和學生連續性的回饋資料，使他們知道學習的成功與失敗。
- (4) 範圍較小，測驗內容限於教學的特定內容，可能是一個概念或原則，或某一單元內容。單元測驗與隨堂考試為此種評量。

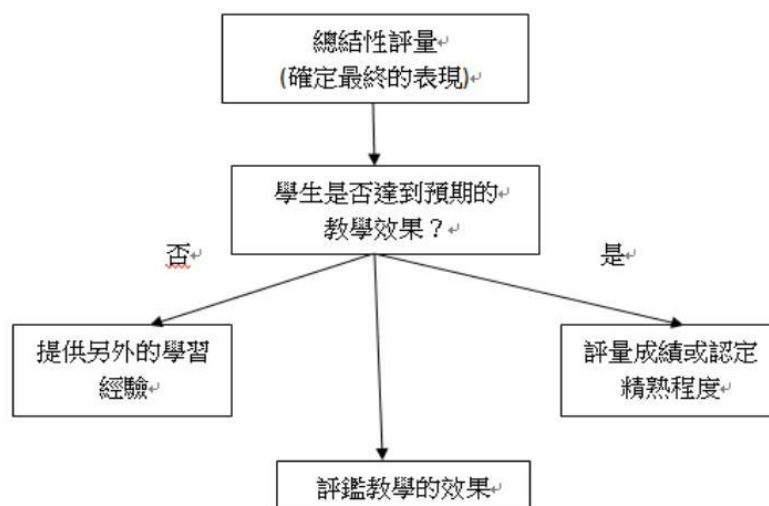


### 3. 診斷性評量

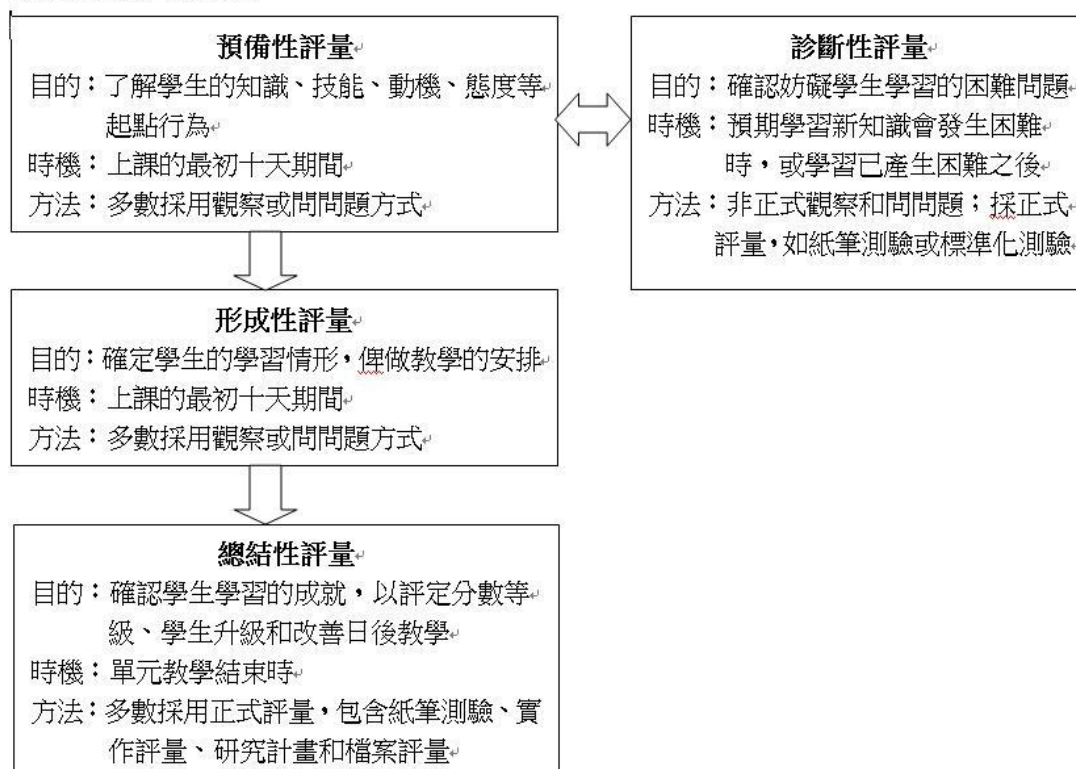
- (1) 定義：在教學活動進行當中，針對學生持續且不斷發生的學習困難之原因，進行診斷和評量。
- (2) 目的：在於鑑定學生的學習困難和形成原因，並提供學生補救的措施，以便設計一種可以排除學習障礙的教學方案。
- (3) 實施診斷性評量，除了採用標準化成就測驗外，為了或得更詳細的學習困難資料，常需使用非正式的教師自編測驗和直接觀察技術，學習困難特別嚴重的，時常需要各種不同的診斷工具，以及專業人員的協助，如心理醫生、學科專家等。

#### 4. 總結性評量

- (1) 定義：指在全部教學課程（或某一單元）結束時，為了確定教學目標達到的程度、學生精熟預期學習結果的程度、以及評定學生成績等級，所進行的評量。學校定期舉辦的週考、月考和期考，均可是為此種評量
- (2) 目的：為了決定預期的教學目標被達到的程度，以及其適切性，也被用於證明學生對於預期學習成果的精熟程度。



#### 教學評量的四種角色



	預備性評量	診斷性評量	形成性評量	總結性評量
評量功能	1. 測量學習前必要的基本技巧 2. 確定已達到課程目標的程度	確定學習困難的原因	提供學習進步的回饋給教師和學生	在教學結束時給予成績等等，或確認精熟程度
評量時間	教學之初	學習發生困難時	教學進行中	在教學之末
評量工具	學前測驗或教師自編測驗	診斷測驗、標準化成就測驗或檢核表	教學須要而特別設計的評定量表、作業或口頭考問、實際演示、問題研討	學期末或教學單元結束時的考試，大多為教師自編的測驗。
取樣考慮	1. 包含每項必要的起點行為 2. 選擇能代表課程目標的樣本	包含共同學習錯誤的樣本	如可能，包含所有單元目標(或最主要的)	選擇能代表課程目標的樣本
難度	大部分試題是基本的、容易的、難度 P 值在 0.65 以上	大部分試題是基本的、容易的、難度 P 值在 0.65 以上	試題配合單元目標的難度	大部分的試題難度 P 值在 0.3~0.7 之間，但較難與較容易的試題也有
評分方式	標準參照	標準參照或常模參照	標準參照、精熟或非精熟	通常是常模參照，也可能是標準參照

#### 常模參照與標準參照評量差異性之比較

	常模參照評量	標準參照評量
評量目的	區分學生成就的個別差異 (分班編組)	確定學生學習的精熟度 (補救教學)
評量內容	涵蓋學習課題的廣大領域，但每一獨特的課題僅以少數題目測量	涵蓋學習課題的有限領域，但每一獨特課題有較多題目測量
題目難度	中等難度，排除極難和極易題目	題目難度和學習內容相配合，不試圖排除極易題目或變更其難度
結果解釋	依據分數在團體中相對地位解釋意義	依據分數達到預定標準解釋
分數種類	百分等級、標準分數、年級常模	內容參照分數、目標參照分數、精熟分數
適用評量	安置性評量、總結性評量	預備性評量、形成性評量、診斷性評量

## (一) 教學評量的種類 (依如何解釋評量結果而分)

### 1. 常模參照評量

- (1) 定義：指評量之後，根據個人的分數在團體中所佔的**相對位置**，加以解釋。易言之，個人分數的意義是和別人比較而獲得的。
- (2) 目的：區分學生彼此之間的成就水準，以做為行政上決策之用。  
例如，分班編級等。
- (3) 多數教師自編測驗與標準化成就測驗都是屬於常模參照測驗。

### 2. 標準參照評量

- (1) 定義：指依據教學前所設定的標準，瞭解學生的學習結果。易言之，教師要求學生的學習要達到預定的標準，並據此判斷學生的成績和優劣。凡是達到標準者，表示學習「精熟」，而未達到標準者，表示學習「非精熟」。
- (2) 目的：確定學生能做什麼，不能做什麼，而不關心學生成就的高低位置。
- (3) 目前我國所舉辦的考試中，包括汽機車駕駛執照、教師檢定等，皆屬於標準參照測驗。
- (4) 圖為標準參照評量的分析

學習內容	甲 生			乙 生		
	答對題數	%	精熟	答對題數	%	精熟
字彙(20)*	18	90	✓	19	95	✓
發音(20)	15	75		18	90	✓
文法(20)	12	60		17	85	✓
翻譯(20)	18	90	✓	14	70	
閱讀(20)	17	85	✓	12	60	

\*表示題數

- (5) 一般而言，標準參照的解釋分數有下列三種類型：
  - ① 內容參照分數：最簡單的形式就是正確答對的百分比，如在 20 題答對 18 題，答對百分比就是 90% 表示精熟了 90% 的測驗內容。
  - ② 目標參照分數：以測驗的目標作為參考架構，以精熟這些目標作為標準，然後以學生精熟目標的百分比或數量作為解釋的依據。
  - ③ 通過/失敗或精熟分數：預先設定可接受的表現水準，稱為預設標準，凡是達到此標準者，表示通過，也就是學習達到精熟程度，而未達到此標準者，表示失敗，亦即學習沒有達到精熟程度。

### 三、變通/替代性評量 alternative assessment

- 指傳統評量以外的另類評量方式，著重在評量與教學的相互配合，以提升問題解決能力與思考能力作為教學與評量的主要目的。
- 重視知識的真實性應用與自我評鑑標準，強調課程本位的能力評量。如多元評量。
- 一般常見的變通性評量有：真實評量、實作評量、學習檔案/卷宗評量

#### (一) 真實評量 authentic assessment

1. 意義：在實際的教學活動中，教學即評量，評量即教學，兩者密切配合；教師在真實教學活動中透過觀察、與學生的談話，以及學生的作品，蒐集各個學生學習情形的資料，直接去測量學生在某一課程的實際操作表現。
2. 目的：強調在使兒童於班級教學的真實活動中，直接去評量學習者在教室活動中直行或操作一項工作，以建構有意義的學習經驗。
3. 特性：強調應用、重視直接評量、使用真實的問題、鼓勵開放思考
4. 三種主要形式：紙筆工作、表演、作品集(歷程集&成果集)

#### (二) 實作評量 performance assessment

1. 意義：在自然或已建構好的情境中，要求學生執行或處理一件指定的工作，並由教師觀察與評鑑學生的建構性反應的過程與結果，看它們是否適當、精確和合理達成教學目標
2. 目的
  - (1) 彌補紙筆測驗的不足。
  - (2) 對學生的學習成就做更正確的推論。
  - (3) 對教師的教學活動產生正向的引導。
3. 實作評量的性質
  - (1) 提供模擬真實情境下施測的方便性
  - (2) 對於年齡幼小和閱讀能力較差的學生，提供有效的簡便評量方法
  - (3) 重點在於真實情境下應用知識和技能的能力
4. 實作評量的特徵
  - (1) 實施費時、耗力
  - (2) 評分以人為判斷為主
  - (3) 問題情境的含糊性和表現的彈性
  - (4) 真實性與直接性
  - (5) 多向度的評分系統
  - (6) 兼重評量的結果與歷程

## 5. 實作評量的優缺點

優點	缺點
(1) 能同時評量認知與技能方面的教學目標 (2) 能提供技能學習方面的診斷資料 (3) 評量情境接近現實生活，有利學生的學習遷移 (4) 提供學生合作的機會 (5) 實作評量較不需要用到語文能力 (語言學習除外) (6) 學生成為評量活動中的積極參與者 (7) 以學生為中心的評量方式 (8) 增進家長扮演觀察者和評量者更積極的角色	(1) 實施上非常耗費人力、時間及金錢 (2) 測驗情境控制困難 (3) 計分不容易客觀 (4) 容易有評量工具上的誤差 (5) 評量次數過少 (6) 合格的評分人員難找 (7) 對容易焦慮的學生不利

## 6. 實作評量的改進建議

- (1) 明確說明預期的表現結果並且描述評量結果的用途
- (2) 擬定清楚、明確且適合學生程度的表現標準
- (3) 將表現標準個數限定在合理的數目範圍內
- (4) 選擇足供學生表現的最真實測驗情境
- (5) 提供清楚且完整的施測指導說明
- (6) 在觀察、判斷和紀錄表現時，儘可能保持客觀、公正
- (7) 在各種條件下觀察表現；儘可能使用多次觀察和判斷
- (8) 在觀察之後，宜盡快做判斷和紀錄
- (9) 使用清楚、適當且方便記錄的評量工具
- (10) 使用評量結果可能會被採用的評分方法
- (11) 告知學生用來評量其表現的方法和標準
- (12) 配合其他方面成就的證據，以補充和驗證實作評量的效度

## 7. 實作評量的類型

(1)	<b>紙筆實作</b> paper-and-pencil performance	① 是一種比較強調在模擬情境中應用知識和技能的評量方式。 ② 可獲得教學所期望達成的學習成果，或作為在更真實的情境中表現的初步評量 ③ 可提供「動手操作」的一個評量初步，有別於傳統式紙筆測驗
(2)	<b>(實物)辨認測驗</b> identification test	① 是指以實物為刺激，但所要求的反應仍是語文反應，而不要求學生實際去操作。 例如：在某些情境下，要求學生辨認一套工具或一組器具，並指明它們的功用，這時又可稱為「物件測驗」object test ② 或讓學生辨認解決某項問題所需要使用的工具、程序等。 ③ 辨認測驗因為能夠以整班(紙筆作答)方式或分站(口頭作答)方式實施，常被放入教學過程中，作為實作之前的準備 ④ 可以顯著地減少材料的浪費、工具的損壞以及意外的發生。

	(實物)辨認測驗	例如： A. 音樂課—發給學生演奏曲樂譜，然後播放該演奏曲的錄音帶，要求學生在聽完後指出演奏時有瑕疵的段落、造成瑕疵的原因及改進方法。 B. 化學課—展示各種化學原料或儀器，要求學生說出它們的正確名稱、功用及使用上應注意的事項。 C. 英文課—播放一段英文對話的錄音，要求學生指出該段對話中發音或文法上的錯誤。										
(3)	結構化實作測驗 structured performance test	① 一份結構化實作測驗，可以作為在標準且有控制的情境下進行評量的工具。測量表現的情境是非常有結構性的，它要求每位學生都能表現出相同的反應動作。 例如：測量天氣溫度，精確度要達到小數點以下第二位。 ② 採用結構化實作測驗必須使用操作型的定義方式，以可被觀察到的、可測量到的、及可被量化的數字形式，加以描述和界定可被接受的最低表現水準，如此的評量結果才會具有客觀性與公正性 <table><tr><th>類型</th><th>例子</th></tr><tr><td>速度</td><td>在 2 分鐘內解決 10 道加法數學</td></tr><tr><td>錯誤</td><td>每頁打字稿不得超過 2 個錯字</td></tr><tr><td>時間</td><td>用西班牙語從 1 數到 20，不能有錯</td></tr><tr><td>精確</td><td>在 5 分鐘內架設好一部實驗儀器 在 3 分鐘內找出一部儀器故障的部位 以 1/8 英吋以內的誤差，衡量一條線的長度 以 0.2 度以內的誤差，閱讀一支溫度計的刻度</td></tr></table>	類型	例子	速度	在 2 分鐘內解決 10 道加法數學	錯誤	每頁打字稿不得超過 2 個錯字	時間	用西班牙語從 1 數到 20，不能有錯	精確	在 5 分鐘內架設好一部實驗儀器 在 3 分鐘內找出一部儀器故障的部位 以 1/8 英吋以內的誤差，衡量一條線的長度 以 0.2 度以內的誤差，閱讀一支溫度計的刻度
類型	例子											
速度	在 2 分鐘內解決 10 道加法數學											
錯誤	每頁打字稿不得超過 2 個錯字											
時間	用西班牙語從 1 數到 20，不能有錯											
精確	在 5 分鐘內架設好一部實驗儀器 在 3 分鐘內找出一部儀器故障的部位 以 1/8 英吋以內的誤差，衡量一條線的長度 以 0.2 度以內的誤差，閱讀一支溫度計的刻度											
(4)	模擬實作 simulated performance	配合或替代真實情境中的表現，在駕駛訓練模擬儀器上學生所顯示出的駕駛技能。 例如：在烹飪課程中，評量學生實際烹飪技巧的表現，在駕駛訓練模擬儀器上學生所顯示出來的駕駛技能。										
(5)	作品樣本 work sample	① 工作範本是真實性程度最高的評量方式 ② 它需要學生在實際作業上表現出所要測量的全部真實技能 例如：學生速記一段口述資料；打一封商業書信；操作電腦分析一份商業資料。										



## 8. 實作評量的實施步驟

### (1) 確立實作評量的目的

- ① 一般而言，有兩個主要目的：給學生評分&診斷學生的學習情形。
- ② 使用目的不同，所決定採用的觀察方法和評分方式就有所不同。
- ③ 在確立好實作評量的目的後，便要詳細說明細節行為的項目，以及教師所期望學生達成的表現標準是什麼。

行為特徵	教學目標舉例
在真實情境中，辨認事物的名稱、功能或瑕疵所在	找出汽車故障所在 說出各種標本的名稱 說出各種工具的用途
依照所給的說明製作出一個成品	設計一套學生制服 畫一張電路線路圖 寫一首合唱曲
表現一組操作性行為或程序	修理故障的收音機 表演華爾滋舞步 演示測量血壓的程序

### (2) 選擇評量的焦點

- ① 實作評量可以把評量重點放在過程或結果上，或兩者兼顧。
- ② 依據所欲評量的技能之性質而決定評量的重點。
  - 通常在學習的開始階段較強調操作過程的正確與熟練。
  - 學習終了階段較注意成果的品質。
- ③ 強調過程
  - A. 沒有成果可以評分（例如：駕駛、體操），或是成果無法取得（例如：急救、消防演習）時
  - B. 操作過程可以按順序直接進行觀察時
  - C. 對於操作過程做分析有助於改進成果時
  - D. 多種不同的操作過程都可以造就好的成果時
  - E. 無法觀察到學生的操作過程（例如：家庭作業）時
  - F. 學生對於操作過程都已經精熟，彼此之間沒有顯著差異時
  - G. 個人的成果之間有很大的差異，可以做明顯的區分時

### (3) 提供適當的表現情境

- ① 表現標準經界定後，教師便需要準備可供進行觀察表現成果的施測情境
- ② 這些情境可以是教室內自然發生的情境，也可是教師特別設計模擬真實的情境
- ③ 決定使用何種施測情境，可以參考下列兩個判斷原則：
  - A. 教室中自然發生表現的頻率
  - B. 做決定的重要性

(4) **選擇計分和評定方法**：依評量的目的、用途、情境和性質而有不同的變化

① **整體式評分法**

- 教師所做的決定屬於一般性質（如分組、評選或評定成績等）
- 只需要教師提供單一的整體分數即可
- 作品等第量表就是一種整體式的計分方法

② **分析式評分法**

- 教師所做的決定具有診斷困難及瞭解學生精熟表現水準的功能
- 依照判斷的形式分成三種：檢核表、評定量表、評分規程

**9. 作品等第量表 product scales**

- (1) 意義：內容包括一系列足以反應出各種不同品質程度的範本作品，將這些範本一一標上號碼，號碼越大代表該範本的品質越優異；號碼越小則代表品質越低劣。
- (2) 將學生的作品對照範本，與範本相同或類似的作品分數即是該範本的號碼。
- (3) 步驟
  - ① 挑選三至七種不同品質程度的作品範本。
  - ② 依序（即依據某種評量標準）排列這些範本，並標上號碼，號碼越大，即代表該作品的品質越高。
  - ③ 將學生的作品依序與該作品量表相對照，找出與範本相似或相同者，並給予該範本所標記的號碼分數，以作為該學生作品的品質得分。
- (4) 使用時機：當某種作品的品質很難以一組特徵來下定義，但使用整體的判斷法還算可行時（如藝術品的鑑賞）。

**(三) 學習檔案/卷宗評量 portfolio assessment**

1. 意義：有系統的蒐集學生作品，透過長時間連續性的蒐集資料，來觀察學生在一個或多個學習領域內的努力、進步與成就。
2. 檔案評量的類型：檔案內容的取舍可依評量的功能分為三方面：
  - (1) 由學生自己決定，呈現學生之自我反省、自我評鑑及自我選擇。
  - (2) 由教師及行政人員掌握決定權，目的在於班級間、學校間或是區域間的評比。
  - (3) 由師生雙方共同決定，教師需教導學生如何蒐集資料來證明自己曾經努力、進步與成長的過程，目的在充分記錄完整之學習圖像。
3. 卷宗評量的實施步驟
  - (1) 決定形成卷宗評量系統架構的標準（課程目標）
  - (2) 將這些標準化成可觀察的行為
  - (3) 使用這些標準，檢查出課程的範圍及順序，並決定出一個收集證據及完成評估的適當時間架構
  - (4) 決定卷宗評量模式中的決策者 stakeholders（老師、學生、家長等）
  - (5) 決定收集資料的形式
  - (6) 決定一個能將證據資料轉換成決策 decision making 的方法
  - (7) 建立一個討論及報告評量資料、以及決定進行方式的系統
  - (8) 建立一系列由不同年齡、年級所完成的卷宗內容及展示品以提供比較之用

4. 卷宗評量的種類：Grosvenor 將學生卷宗評量的形式區分為：

- (1) 展示卷宗 showcase portfolios：讓學生從卷宗選出自認最好的作品，供作展覽
- (2) 描述性卷宗 descriptive portfolios：示範學生能作什麼，清楚的展現出學生的學習情形
- (3) 評估性卷宗 evaluative portfolios：主要供作評估的標準，為了評估出學生作品的品質而定。
- (4) 進步卷宗 progress portfolios：強調必須收集學生在特定學科領域表現出進步的證據而定

5. 檔案評量的優缺點

優點	限制
(1) 是一種動態的、完整的真實性評量 (2) 尊重學生之個別差異 (3) 兼重學習過程及結果 (4) 資料來源廣泛 (5) 學生具有評量的主導權 (6) 可評鑑學生之批判思考能力 (7) 提供學生、教師及家長互相溝通之訊息與途徑	(1) 費時費力 (2) 較高經費與設備的需求 (3) 評分較為主觀，評量結果的信、校度受爭議 (4) 無法提供適時的學習回饋 (5) 學生個別差異大，評分表設計不易

6. 卷宗/檔案評量 vs.傳統評量

	卷宗/檔案評量	傳統測驗
評量的範圍	不受教材範圍限制，可以看到全部學生能力的全貌	侷限於教材範圍之內，看不出特優及落後學生能力的全貌
評量主要目的	培養學生自我評鑑、自我改進的能力，養成學生學習責任感	考核學生學習成果及教師教學效能
評量時學生反應	以應用綜合層次的、建構式的反應為主。可以參考資料、或共同完成活動設計	以知識、理解層次的、選出式的反應為主。不可參考資料，且獨立受測
適用時機	小班級、強調個別教學	大班級、有統一教材進度
個別差異的考慮	活動設計中已經考量到學生個別差異	全部學生使用相同的教材及測驗內容
學生角色	學習者、求助者、自我評量者	被評量者、被獎懲者
家長角色	被邀參與評量及學生學習	被告知學生的學習成就
教師角色	活動設計者、顧問、引導激勵者	考核者、獎懲者、補救教學者
師生關係比喻	像舊式的師徒制	像工廠的製造者與其產品
評量結果說明	學生的努力、進步與成就	只有學生成就

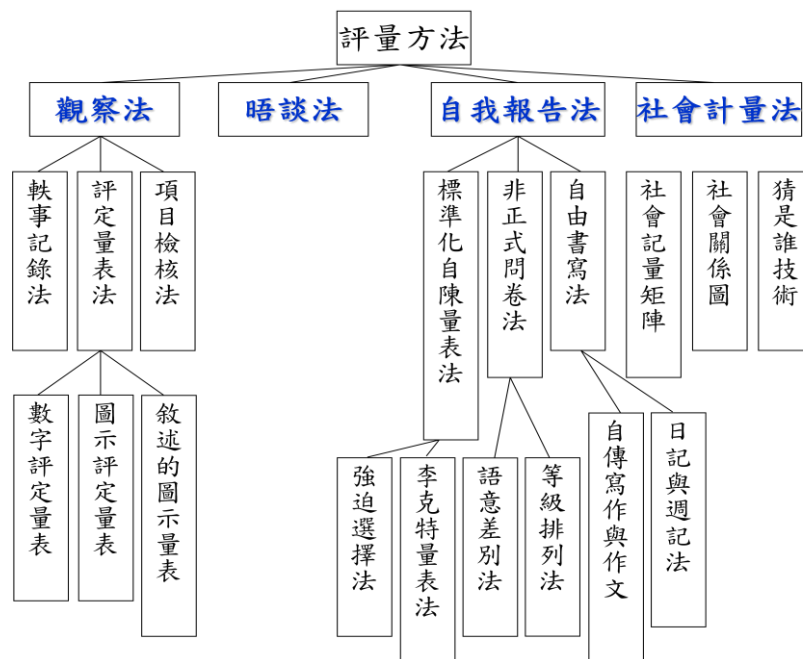
#### (四) 動態評量 dynamic assessment

1. 意義：不在評量過去既有的知識、技巧或經驗，而在於評量學習過程中的改變以及同時在評量程序中包含教學的介入。
2. 目的：在評量兒童的潛能發展水準，以了解兒童在問題情境中能力運作的狀況，為教學方式的決定提供有用的訊息。
3. 動態評量最主要的特質
  - (1) 著重學習歷程或認知改變的評量。
  - (2) 在評量中進行教學。
  - (3) 評量者與被評量者的關係是互動的。
4. 動態評量的理念
  - (1) 時間的連續性：學習評量應考量時間的連續性，亦即定點定時的評量，不足以做為最後價值判斷的唯一依據。
  - (2) 成長的連續性：個體的成長必有其歷程，凡當下的反應自有其生長的次序與因果，動態評量重視此一連貫性。
  - (3) 評量工具的多樣性：不以單一性質的評量工具，作唯一的評量。動態評量以多樣評量工具，做多方評量。
  - (4) 對象的適切性：由於個體特性不一，同一評量工具不可能適合所有個體。動態評量因學生不同，選擇不同的評量方式與工具。
5. 動態評量的特徵
  - (1) 測量過程為「測驗→教學→測驗」或至少是「教學→測驗」的程序
  - (2) 評量的重點是強調知覺、思考、學習和問題解決的過程而非過去學習的成果。
  - (3) 重視可概化認知歷程的教學或學習。
  - (4) 企圖找出影響個體有效學習或行為表現阻礙的因素。
  - (5) 辨識對認知歷程教學有反應者。
  - (6) 掌握積極影響或激勵個體學習和行為反應的情境及條件。
  - (7) 強化未來成功和學習可能的分辨。
6. 心理計量模式：由 Embretson 所提倡。強調認知的可變性，以空間推理測驗為訓練材料，採用前測 – 訓練 – 後測 的方式來評估受試者的能力，在訓練階段是採用標準化介入的方式。在訓練階段受試者是透過具體操作，來增進空間處理的技巧。
  - (1) 評量重點：在訓練後，評估受試者最大的學習能力。
  - (2) 評量目的：發展適當的心理計量模式，來測量認知的改變，及提供受試者特定能力較佳的估計。
  - (3) 評量結果：受試者透過動態測驗介入，空間能力有顯著的進步，而受試者前、後測間的改變分數，可預測文書編輯的表現。

## (五) 生態評量

1. 定義：依生態學的觀點，個體的行為是個體與所處環境交互作用的產物。生態評量便是透過各種方法，對學生在各種生態環境中的能力需求及必備能力進行分析，以利於教師為學生設計功能性的學習目標，並進行教學。
2. 特徵：最終目的在於教導個體適當的社會性行為，協助個體達成社會化。
  - (1) 以學生目前及未來可能接觸的環境為評量範圍:以學生所生活的自然生態系統為主，包括學校、家庭、社區、工作場所和休閒娛樂。
  - (2) 個別化的評量過程:隨學生的生活時間和環境而改變內容和結果。
  - (3) 強調協助學生成功的適應:調整教學結構、提供支援系統，以增進學生的適應行為能力，促成學生的行為與生態環境達成平衡。
  - (4) 發揮學生的潛能:無論學生的優劣程度為何，只要給予學生適當的協助。學生在其生態中均能有適當的行為表現。
3. 步驟：生態評量的步驟分為領域、主要環境、次要環境、活動與技能等五階段。
4. 生態評量的過程
  - (1) 確認學生所處的各项環境
  - (2) 針對每項特定環境設立任務評量表
  - (3) 針對某一特定的環境進行各種可能活動的分析
  - (4) 進行所需技能的工作分析與差異分析
  - (5) 設計教學內容
  - (6) 教學
5. 限制：生態評量固然有其使用上的優點，但相關配合條件仍要十分充分，否則不易執行，包括：
  - (1) 教師須要花費較多的時間評估環境，制定教學目標。
  - (2) 常須戶外教學，須行政及社區資源。
  - (3) 教師要有創造力。
  - (4) 需用較多的輔具。
  - (5) 因為主要透過觀察來獲得，易流於主觀，易產生信效度問題。

## 四、情意評量



### (一) 觀察法

- 評量行為與態度的改變，最好的方法是**觀察行為**。
- 使用自陳量表法測量，收集的資料易產生偏差(自我報告時的社會期許性)。
- 用來記錄觀察結果最常用的方法包括：軼事記錄法、評定量表法、項目檢核表。

#### 1. 軼事記錄法

- (1) 意義：與學生有密切接觸的教師或學校的相關人員，透過直接觀察，簡短地紀錄學生日常生活中出現的特殊行為或經歷的事件，形成累積性紀錄。
- (2) 主要用途：分析與評估學生的社會與情緒適應。
- (3) 記錄內容：著重於事實的描述，而非觀察者對於該行為的解釋或評語。
- (4) 記錄格式：可自己設計，如名片大小的自黏貼紙
- (5) 記錄時最好是每個學生一張小卡片，其中一部份敘述偶發的事件，另一部份作短評
- (6) 軼事記錄法的基本原則
  - ① 事先確定所欲觀察的行為，但對**不尋常**的行為要特別注意。
  - ② 觀察和紀錄各種情境的行為，使資料具有意義。
  - ③ 僅記錄行為的事實、理由、與人的交互關係。
  - ④ 儘可能在行為發生後**馬上**記錄。
  - ⑤ 每項軼事記錄僅限於對**單一特定**事件的簡短敘述。
  - ⑥ 事件的描述和短評宜分開。
  - ⑦ 記錄**積極與消極**的事件。
  - ⑧ 收集相當多的軼事記錄之後，才對學生的行為做推論。
  - ⑨ 從事軼事記錄時應有練習的機會。
  - ⑩ 建立獲得有系統的軼事記錄的計畫。

(7) 軼事記錄法之優點

- ① 能瞭解學生在自然情境下的行為，不被人為情境下的作假反應所**矇騙**。
- ② 能使教師更**勤於**觀察學生，並注意不尋常卻具有高度意義的事件。
- ③ 在使用對象及觀察的變項上具有很大的**彈性**。
- ④ 可作為**日後評量**操行成績、個案討論或學校親職教育座談會時引用或查證之用

(8) 軼事記錄法之缺點

- ① 軼事記錄法的使用上**費時、費力，實用性不高**。
- ② 容易受到教師**主觀**涉入的影響。
- ③ 只能觀察到**有限**的行為樣本。

## 2. 評定量表法 rating scales

(1) 目的：評定所觀察的行為或特質的品質。

(2) 構成的部分：所欲評定行為或特質 & 表示行為或特質程度的量表。

(3) 意義：依據旁觀者的觀察評定一個人的行為，它可以評定個人的能力，也可以評定在情意方面的表現。所評定的表現行為特質通常是屬於等距量尺 **interval scale** 以上的連續性變項資料。

(4) 評定量表法之原則

- ① 所評定的特質需具有**教育的意義** → 評定的特質應和教學目標相符合一致。
- ② 所評定的特質應是可以**直接觀察**到的。。。
- ③ 應清楚界定所評定的特質和量表上的評定點。使用敘述的圖示評定量表可以克服此項困難。
- ④ 量表中的評定點以**三至七個**為宜。評定等級的數量應視評定的目的而定
- ⑤ 無法評定的特質應該省略。
- ⑥ 儘可能綜合幾個人的評定結果。
- ⑦ 寧增加評定的項目而不要增加評定的等級。
- ⑧ 加強對評定者的訓練與溝通。

(5) 評定量表法之優缺點

① 優點：

- A. 適合用於評定工作的過程、結果、及個人的態度與行為。尤其最適合用來評定個人的態度與行為。
- B. 編製容易，**使用簡便**。
- C. 可以**數量化**，便於做統計的分析。

② 缺點：**缺乏客觀性**，因為有不少的評量誤差存在

(6) 使用評定量表法之誤差

- ① **個人的偏差**，包括：寬鬆的誤差、嚴苛的誤差、趨中的誤差
- ② **月暈效應**：評定者因為學生在某一特質上的表現令他印象深刻，而影響了他在其它項目上的評定。
- ③ **邏輯誤差**：評定者沒有根據直接的觀察，將兩項特質視為有直接的關係，而將之評定為比原來實際的關係更接近或相反。

(7) 評定量表的種類：數字評定量表、圖示評定量表、敘述的圖示量表

數字型	<p>根據學生在演講時所表現出的行為程度，圈選適當表示該特質程度的數字；</p> <p>「1」代表「從未」表現出該行為、「2」代表「很少」表現出該行為、「3」代表「偶爾」表現出該行為、「4」代表「一直」表現出該行為。</p> <p>1.肢體表達方面：</p> <p>(1)直直地站著，並且面對聽眾。</p> <p>1 2 3 4</p> <p>(2)隨著說話音調的高低而變化面部的表情。</p> <p>1 2 3 4</p> <p>(3)與聽眾保持目光接觸。</p> <p>1 2 3 4</p>
圖示型	<p>根據學生在演講時所表現出的行為程度，在下列橫線上的適當位置打「ˇ」。</p> <p>1.肢體表達方面：</p> <p>(1)直直地站著，並且面對聽眾。</p> <p>從未 很少 偶爾 一直</p> <p>(2)隨著說話音調的高低而變化面部的表情。</p> <p>從未 很少 偶爾 一直</p> <p>(3)與聽眾保持目光接觸。</p> <p>從未 很少 偶爾 一直</p>
	<p>在下列橫線上的適當位置打「ˇ」，以表示學生在演講時所表現行為的最佳描述。</p> <p>1.肢體表達方面：</p> <p>(1)直直地站著，並且面對聽眾。</p> <p>僵硬、不專心地 不安地來回走動， 站得直直的，並且 移動身體，不敢看著聽眾 目光隨意亂飄 一直面對著聽眾</p> <p>評論：_____</p> <p>(2)隨著說話音調的高低而變化面部的表情。</p> <p>音調和面部表情 面部表情有時恰當， 音調和面部表情能夠 不一致，表情茫然 有時面無表情 配合演說內容和重點</p> <p>評論：_____</p> <p>(3)與聽眾保持目光接觸。</p> <p>不敢正視聽眾， 目光隨意游動， 從頭到尾，一直 目光游離不定 與聽眾有短暫的目光接觸 與聽眾保持目光接觸</p> <p>評論：_____</p>



### 3. 項目檢核表 checklists

(1) 意義：判定行為特質是否出現，並且提供簡單記錄「是」或「否」判斷的資料表。

如：小明上課經常打瞌睡 ☐是 ☐否

因它是作二分的判斷，故不適用於評定某些行為或特質的程度和品質。

(2) 評分者常會加上一些說明來支持他的判斷，或寫一些觀察到的重要事件，所以在每一項目之後留有空白，或在表格右邊留有較大的空間將很有幫助

(3) 項目檢核表之使用原則

- ① 僅在確定學生某項特質是否出現時，才使用項目檢核表。
- ② 清楚界定所要觀察的特質。
- ③ 一次只觀察一位學生，而且僅限於觀察檢核表中的項目。
- ④ 每位學生都有各自的項目檢核表。
- ⑤ 教師應瞭解如何觀察、觀察什麼和如何記錄。

### (二) 晤談法

1. 晤談法比問卷法或觀察法更能正確地瞭解學生的情感反應。因透過師生間的個人溝通，是瞭解學生在學習上的各種情意反應最有效的方法。

- ① 要得到學生的信任
- ② 要有充裕的時間事先安排談話及準備問題。
- ③ 準備大量清晰、簡要的開放性問題。

2. 有助於提高晤談效果之技巧：

- ① 採團體的方式進行。
- ② 由學生來領導團體討論。
- ③ 保持溫暖、積極傾聽的態度。

### (三) 自我報告法

- 由當事人自己在問卷上報告自己的狀況。
- 依其編製原理及形式可分成：

1. 標準化自陳量表法 (見人格測驗)：經由測驗專家主持，擬題、預試、試題分析、因素分析、建立常模和信效度研究等程序的標準化人格測驗。

2. 非正式問卷

- (1) 分為 1.語意差別法 2.等級排列法。
- (2) 語意差別法又稱「兩極形容詞量表」，用來測量個人對某一特定事物的知覺的方法。
- (3) 等級排列法是把代表不同價值系統的敘述句並列在一起，然後要求學生做順序排列。  
該法適用於評量「4-2 價值系統的組織」的教育目標。

### 3. 自由書寫法

- (1) 此法比較像非結構式的晤談。
- (2) 可融入語文科的教學之中，並以團體的方式實施。
- (3) 為使此法充分發揮情意評量的功效，要先讓學生相信：
  - ① 教師所重視的是內容而不是文字與錯別字。
  - ② 教師認真閱讀，並在其回饋中接納任何情感的表達。
  - ③ 它只是個人表達能力的練習或教師瞭解學生生活經驗的方式。
- (4) 分自傳寫作與作文，及日記與週記法。
- (5) 自傳寫作與作文
  - ① 可讓學生自由報告特別值得他記憶的事件、經驗和感覺。
  - ② 使用時，教師可事先朗讀某些具有強烈情感或私人性的自傳性文章給學生聽。
  - ③ 學生完成自傳之後，教師可以從幾個向度來評估其自傳，以瞭解學生在情意方面的發展。
- (6) 日記與週記法
  - ① 要求學生紀錄每天所思所想。
  - ② 此法可配合國小學生提早寫作練習。
  - ③ 教師每週檢閱一次，並給予適當的回饋。
  - ④ 此法到了青少年期就不適用了。
  - ⑤ 把它當作與學生心靈的對話。
  - ⑥ 教師針對其內容作具體、豐富的回應。

### (四) 社會計量法/社交測量法

- 由莫里諾 Moreno 提出
  - 係研究團體結構的一種技術，該法很容易用來探討個人在該團體中的人際關係
1. 社會計量法常使用的問題
    - (1) 你最喜歡和班上的哪位同學坐在一起？ 你最不喜歡和班上的哪位同學坐在一起？
    - (2) 你最喜歡和班上的哪位同學一起作功課？你最不喜歡和班上的哪位同學一起作功課？
    - (3) 你最喜歡班上的哪位同學？ 你最不喜歡班上的哪位同學？
    - (4) 誰是班上最好的領導人物？誰是班上最不好的領導人物？
  2. 社會計量法之限制
    - (1) 只適用於學生之間都相互熟悉的班級。
    - (2) 資料只顯示出所選擇的對象。
    - (3) 結果容易受到問題情境的影響，常隨著問題的不同而有不同的選擇。
    - (4) 當班級人數多時（約超過二十五人以上），社會圖就很難繪製

### 3. 社會計量法的種類

#### (1) 社會計量矩陣

- ① 意義:分析團體人際關係最簡單的一種方法，且是心理與教育研究中使用最普遍的方法。
- ② 範例:以五位同學回答「你最喜歡班上的哪位同學？」和「你最不喜歡班上的哪位同學？」兩題。

#### (2) 社會關係圖

- ① 社會關係圖的繪製比社會計量矩陣複雜，繪製社會關係圖並沒有一致的標準可遵循，圖形也因人而異。
- ② 一般常用的方式是以不同形狀的圖形代表性別，女性為圓形，男性為三角形。
- ③ 缺點：無法做實證性的分析，而且每個社會關係圖僅適用於單一的問題。

#### (3) 猜是誰技術

- ① 又稱提名法，社會計量法的一種，藉此可以獲得友伴評斷的資料。
- ② 此法係用描述各種行為特質的語句，要求學生寫下最適合每一個描述句的友伴姓名。
- ③ 描述句有積極和消極兩類，作答時可自由反應，不需每項描述句都回答。
- ④ 適用於經常同班上課的中小學階段。
- ⑤ 可以找出某一學生被同學喜歡或不喜歡的理由。
- ⑥ 實用性高。
- ⑦ 有助於教師診斷需要輔導的學生。
- ⑧ 五名學生寫下適合六項行為特質的友伴姓名的分析

## 五、課程本位評量 Curriculum-based assessment, CBA

### (一) 何謂課程本位評量

1. 由 Gickling 和 Havertape 提出，原為「課程本位」與「評量」的兩個概念的組合。
2. 「課程本位」強調以實際授課內容與教學材料作為評估學生學習的表現依據；
3. 「評量」則是資料蒐集的過程，所得的資料作為提供教育決策與學生為提所在的确認之依據，因此「以學生在現有課程內容上的持續表現來決定其教學需求的一種程序」就是「課程本位評量」。
4. 即一種整合課程、教學和測驗的一種非標準化評量。是由教師在教學過程中，以實際課程內容為基礎所編制的評量。藉由簡單易行的經常性測驗，教師可及時評估學生的學習困難&教學方向。

(二) 課程本位評量的重要性：自 1980 年代初期起日漸受到重視的一種教學評量方式，目前廣泛應用於輕、中度障礙領域，以及針對學習困難學生所設計的補救教學情境中。原因是它能針對個別障礙學生所要進行的課程，給予老師最適當的線索。

(三) 課程本位評量目的：課程本位評量能解決評量問題、監控學生學習情況、回饋教師教學策略及觀測長程目標的一種有效、簡便、立即的方式

(四) 課程本位的三種模式：

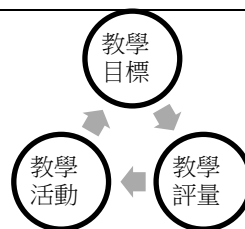
1. 標準參照的課程本位評量模式。
2. 以正確性為基礎的課程本位評量模式。
3. 以流暢性為基礎的課程本位評量模式。

類別	測驗目的	測驗重點	測驗特色
標準參照的 課程本位評量模式	主要在測量學生的表現以直接和教學目標做比較，據以作為教師設計教學的參考。	測量的重點主要以教學目標相對應的行為目標為主，在於了解學生在試題上的表現，以判斷學生是否精熟每一個學習目標	內容效度高。最適合用來監控學生在短期學習目標的進步情形。
以正確性為基礎的 課程本位評量模式	主要在計算學生於課程本位測驗上的正確反映百分比，根據學生正確反應的數據，教師據以修改教學內容的難易度以配合學生目前的技能發展水準。	重點在於教師在教學前將計劃教學的內容轉譯成的不同「教學層次」，在各教學層次的測量中，學生的「獨立學習層次」越高「挫折層次」越低，則代表學生需要挑戰性越高的題目，或應該降低教材的難度。	本模式用於教學的設計，教師需經常性的失策以正確並控制教學項目的難易度，配合學生最大的發展需求。
以流暢性為基礎的 課程本位評量模式	主要在直接測量學生技能發展的流暢程度，據以作為教師長期觀測與修正教學的參考。	重點在於速度的評量，其評量的結果在於顯示個人在單位時間內正確的反應次數。	實施簡單易行，有令人滿意的信度與效度，可靈敏的反應學生在短期的進步情形。

# 教學目標

■ 成功的教學要考慮三要素：

1. 希望學生學習什麼？學習之後預期的結果？
2. 採用何種教材和教法有助於達成預期結果？
3. 採用何種方法評量才能了解學生的行為改變？



## 一、教學目標的性質

(一) 意義：教學目標係指經教學歷程後，預期學習者達致的學習結果；其亦認為教學目標的敘寫應該是明確的，應以能證明學生達致學習目標的行為表現來描述。

(二) 功能：教學目標具有引導、診斷、溝通三種功能。

(三) 選擇適當教學目標的標準

1. 教學目標須力求和國家課程標準、學校一般目標相符合一致。
2. 教學目標須涵蓋課程所有重要結果。
3. 教學目標須力求和有效的學習員則符合一致。
4. 教學目標須切合實際，適合學生的能力、學習時間和可用資源。

(四) 我國各教育階段的教學目標

1. 幼稚教育〈幼稚教育法〉：幼稚教育的目標以健康教育、生活教育及倫理教育為主，並與家庭教育密切配合。
2. 國民教育〈國民教育法〉：以養成德智體群美五育均衡發展之健全國民為宗旨。
3. 高級中學〈高級中學法〉：陶冶青年身心，培養健全公民，奠定研究學術或學習專門知能之預備為宗旨。
4. 職業學校〈職業學校法〉：教授青年職業技能，培養職業道德，養成健全之基層技術人員為宗旨。
5. 專科學校〈專科學校法〉：教授應用科學與技術，養成實用專業人才為宗旨。
6. 高等教育〈大學法〉：以研究學術，培育人才，提升文化，服務社會，促進國家發展為宗旨。

(五) 分類：

1. 認知領域的教學目標—又分為六個層次：

- (1) 1956 年版 Bloom 的分類：知識、理解、應用、分析、綜合、評鑑。
- (2) 2001 年版 Bloom 的分類：記憶、了解、應用、分析、評鑑、創造。

2. 情意領域的教學目標—克拉斯霍爾分為五個層次：

接受或注意、反應、評價、組織、品格形成。

3. 技能領域的教學目標

- (1) 哈樓分六類：反射動作、基本動作、知覺能力、體能、精確動作、有意的溝通(自然化)。
- (2) 辛普森 Simpson 分七類：知覺、心向作用(準備)、模仿、機械練習、複雜反應、技能調適、創作。

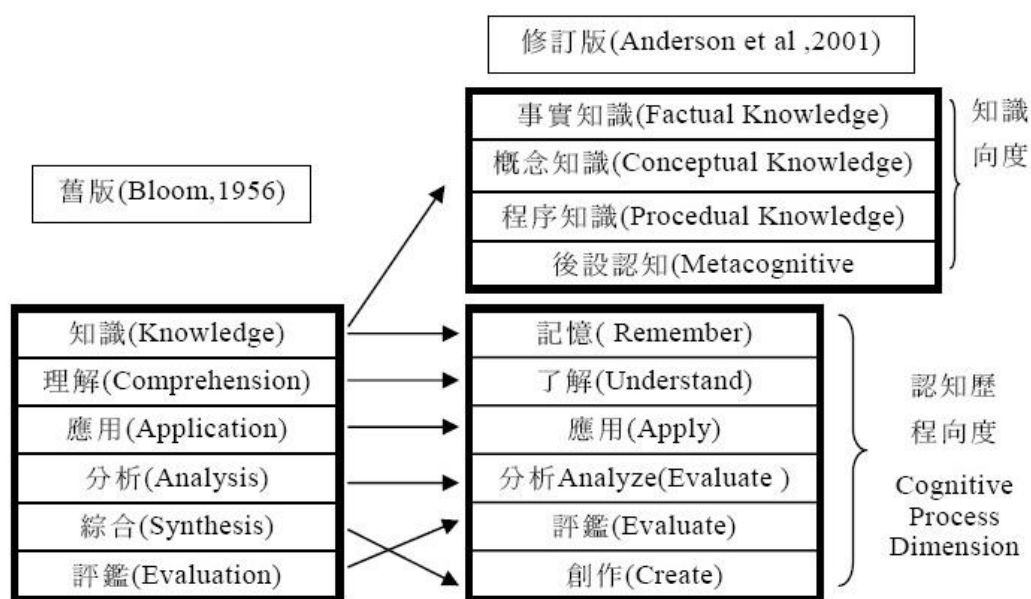


圖 1 Bloom 教育目標分類系統新舊版本對照圖(譯自 Anderson et al, 2001), p268)

## (一) 認知領域

I. 知識向度 Knowledge Dimension：知識向度屬名詞詞態，以學習內容為主，目的在協助教師區分教什麼；並延續跨領域的觀念，將知識區分成四類：

1. **事實知識 Factual Knowledge**：指學生應了解的術語，或學生想進行問題解決時必須知道的基本要素。包括：(1)術語知識。(2)特定細節及元素知識。
2. **概念知識 Conceptual Knowledge**：可連結不同基本概念之相關知識。包括：(1)分類和類別知識。(2)原理和原則的知識。(3)理論/模式/結構的知識。
3. **程序知識 Procedural Knowledge**：指知道如何做某事的知識；通常是一系列或有步驟的流程，以及決定何時運用不同程序的規準。包括：(1)特定學科的技能 and 演算知識。(2)特定學科的技術和方法知識。(3)運用規準的知識。
4. **後設認知知識 Metacognitive Knowledge**：指一般對認知的認知以及對自我知識的認知和覺察；含認知知識、監控、控制、調整認知。包括：(1)策略知識。(2)認知任務知識（包括脈絡和情境的知識）。(3)自我知識 self knowledge。

## II. 認知歷程向度

1. **記憶**：強調對於某一學科領域的內容及特定要素的「回憶」與「再認」。  
例如：能寫出算術平均數的計算公式；能說出古典小說「西遊記」作者是谁？  
(1) 回憶：要求學生自己回想並寫出記得的內容，如填充題、簡答題、默寫。  
(2) 再認：要求學生就所提供的各種答案中判斷哪一個是與記憶的內容符合。
2. **瞭解**：指個人藉著思考歷程把得到的訊息轉換成更具有意義的形式，甚至超出原有的意義之外。並不重視回憶能力，強調對新資料的處理。  
「詮釋」、「舉例」、「分類」、「摘要」、「推論」、「比較」、「解釋」。  
例如：「把一段中文翻譯成英文」、能推論出「溫室效應」的對環境造成的影響。

3. **應用**：學生能把學到的原理和通則運用到新問題或新情境中，即能「舉一反三」。「解決問題」。  
例如：能應用相關的公式計算兩個變項間的關係；成語中的「望梅止渴」，可以用心理學的什麼原理來加以說明？
4. **分析**：把一個溝通材料拆解成幾個構成要素或是幾個部份，以使各概念的相對階層能夠更清楚，或各概念之間的關係能夠更明顯。「區辨」、「組織」、「歸因」。  
例如：提供一段矛盾的陳述，讓學生指出其錯誤或不合理之處；發給每位學生一篇社會新聞報導，要求學生在閱讀之後：(1)以紅色筆畫出可以查證的事實部分(2)以藍色筆畫出撰稿人的個人推論部分
5. **評鑑**：為了某種目的而使用一套準據來對於不同的觀念、作品、方法、材料等的價值做判斷。判斷的結果可以是數量的，也可以是質量的。  
例如：對墮胎提出自己的意見並說出支持與反對理由；試比較兩位不同翻譯者所翻譯的「老人與海」一書，寫一篇評論性文章比較兩者翻譯的優劣。
6. **創作**：集合要素以組成一個具協調性或功能性的個體，重組要素為一個新的模型或結構。「產生」、「計畫」、與「創作」。  
例如：創作歌曲或樂曲。

## (二) 情意領域

1. **接受 receiving**：指學生願意去注意特定的現象或刺激(班級活動、教科書或音樂等)。主要和學生在知覺上的注意力有關。覺知 → 願意接受 → 選擇性的注意  
例如：注意到交通秩序的問題；能夠聽完全首曲子。
2. **反應 responding**：對某些經驗表現出新的行為。指學生主動地參與學習活動，它不只是注意到特定的現象，而且是以某種方式對它做反應。勉強反應 → 願意反應 → 樂於反應  
例如：願意遵守交通規則；主動要求上台朗讀課文。
3. **價值判斷 valuing**：對於接觸到的事物表現出明確肯定的態度傾向。人們對於各種不同的價值觀有著不同程度的認知上的相信與情感上的接納。接納 → 偏好 → 誓約  
例如：熱愛遵守交通秩序的人；相信公開討論有助於衝突的解決。
4. **價值的組織 organization**：將新價值觀與舊有價值觀整合成新的價值系統。指學生把不同的價值觀放在一起，解決它們之間的衝突，並開始建立起一個具有內部一致性的價值系統。強調的是各個價值觀之間的比較、關聯和綜合。價值的界定 → 價值系統的組織  
例如：依照交通的狀況，設計一套改善的計畫；了解並接納自己的優點和限制。
5. **價值的性格化 characterization by a value or a value complex**：指由於個人價值系統的影響，而在某些方面產生主動、長期、一致性的行為，並形成他個人的生活風格(life style)。類化的傾向 → 品格的形成  
例如：養成節儉的生活習慣；培養愛人如己的品格。

### (三) 動作技能領域

1. **知覺 perception**：指用感官注意到物體、性質或關係的過程。  
例如：複誦儀器操作方法。
2. **準備狀態 set**：指對於某一種動作或經驗，在心理上、身體上和情緒上的準備適應或預備狀態。  
例如：正確裝卸儀器之零件。
3. **模仿 imitation**：指在教師的指導下，開始學習動作技能。  
例如：描畫所觀察的標本，表現解剖刀正確的使用方法。
4. **機械 mechanism**：指所學習的動作反應已成為習慣，在表現動作時，學習者已達到某種自信和熟練的程度。例如：正確快速裝置儀器。
5. **複雜的反應 complex overt response**：學習者能夠操作複雜的動作，其操作是順利和有效的。  
例如：完成精確的解剖。
6. **創作 origination**：指依據在動作技能領域所發展出的瞭解、能力和技能，創造新的動作以及處理材料的方法。例如：改良實驗裝置。

## 二、表意目標 expressive objective

- Eisner 提出表意目標，認為學校的教育應該鼓勵學生運用各種感官系統，獲得經驗與感受，除了文字與數字之外，之要發展情意的活動與經驗，以培養廣泛的智慧與素養。
- **表意目標**不在學生從事教育活動後應該展示的行為結果，而在確立學生所經歷的情景。  
如：(1)解釋“失樂園”(Paradise Lost)的意義。  
(2)檢視和評估“老人與海”(The Old Man & The Sea)的重要性。  
(3)使用電線和木材，設計三度空間的形式。  
(4)訪問動物園，討論其中的趣味

## 三、行為目標

### (一) 行為目標的敘寫方法

1. 教學目標必須轉換為行為目標，來進行評量。
2. **行為目標 behavioral objective**：
  - (1) 用具體明確、可觀察、可測量的學生行為所敘寫的目標。
  - (2) 明確的學習結果
  - (3) 實作目標、可測量的目標行為目標的敘寫方法
3. 敘寫的方法(五項基本要素)：
  - (1) **對象**：誰要完成教學所預期的行為，通常指學生
  - (2) **行為**：指達到目標的具體學習行為，如「寫出」「說明」「列出」「預測」等
  - (3) **結果**：行為產生的結果，如「能說出動物名稱」
  - (4) **情境**：表現行為的有關情境或條件，如「能利用字典(情境)查出本課所有生字」
  - (5) **標準**：預期行為可接受的程度教學目標和行為目標，如「能說出校園至少十種以上(標準)的數目名稱



- 教學目標
  - 1. 能了空氣的成分
- 行為目標
  - 1-1 能了解空氣是混合物
  - 1-2 能說出空氣各主要成分
  - 1-3 能推算出各主要成分佔的體積比率

## (二) 良好行為目標的特性

1. 行為目標的敘寫是學生導向，非教師導向
  - 培養學生的閱讀能力 (教師)
  - 能說出本篇文章的大意 (學生)
2. 行為目標的敘寫是學習的結果，非學習的活動
  - 研究蝴蝶的生活過程 (活動)
  - 能標註蝴蝶生活過程中的各個階段 (結果)良好行為目標的特性-2
3. 行為目標是清楚並可觀察的
  - 熟悉面積的求法 (不清楚)
  - 能計算出任何平面幾何圖形的面積 (明確)
4. 行為目標只包含一個學習結果
  - 能說出光合作用的意義 (一個)
  - 能說出光合作用的意義及所需的要素 (二個)

## 一、編製紙筆式評量

### (一) 教學評量的方法可略分三類

1. 選擇式反應測驗題：(1)選擇題、(2)是非題、(3)配合題
2. 建構式反應測驗題：(1)填充題、(2)限制反應題、(3)擴展反應題。
3. 實作評量：(1)口語、(2)書面、(3)展演、(4)視覺、(5)學習檔案

### (二) 教師自編試題步驟

- 1.擬訂計畫 → 2.撰寫試題 → 3.審查試題 → 4.試題分析 → 5.編輯試題



### (三) 雙向細目表設計原則

1. 目的：改進試題的取樣
2. 設計主要依據兩項資料：一是教學目標，另一是教材內容。教學目標不同，測驗的功能和命題的取樣範圍也有所差異。例如，為評量學生的學習進步情形(形成性評量)，其試題的取材範圍將比評量學習是否達到教學目標(總結性評量)為小或獨特
3. 通常以布魯姆 Bloom 的教學目標為依據，分為知識、理解、應用、分析、綜合與評鑑等六種不同能力層次。
4. 依教學目標、教材內容重點、課程內容難易與重要性，決定表中的試題數目與比重
5. 視情況增減表中教學目標與教材內容、及試題數目與比重。
6. 未教過的教材範圍，不應有試題出現。
7. 針對教師自己的命題及評分習慣，選定測驗題型，增減試題數目。

### (四) 是非題

#### 1. 命題原則

- (1) 文字的敘述力求簡要、避免過長複雜
- (2) 避免使用否定的敘述，尤其是雙重否定
- (3) 除非測量因果關係的題目，否則同一題中的敘述避免包含兩個概念
- (4) 避免使用一些含有暗示作用的特殊字詞是非題的命題原則
- (5) 對的題目與錯的題目在敘述的長度上應接近相同
- (6) 在因果關係的題型中，結果的敘述必須是對的，而原因的敘述可對可錯
- (7) 「對」與「錯」的題數宜大致相等，且隨機排列之
- (8) 避免使用含糊、不確定的數量用語

#### 2. 是非題優缺點

優點	限制
(1) 容易命題，且適合多數教材內容 (2) 適用於年幼兒童或閱讀能力有限的學童 (3) 涵蓋教材內容的範圍比其他客觀試題要大 (4) 能夠迅速而客觀的評分 (5) 特別適合於測量有誤解的信念或迷信 (6) 適合測量辨認因果關係的能力	(1) 作答易受猜測因素的影響 (2) 是非題比其他選擇式的題目易引起誤解，故信度較低 (3) 容易作弊 (4) 試題的鑑別作用比選擇題差 (5) 學生容易形成自動「答對」或「答錯」的反應傾向 (6) 在許多情形下，是非題的敘述沒有很清楚對或錯，亦即正確的程度不同。

## (五) 選擇題

### 1. 性質

- (1) 選擇題的組成元素有兩大部分：題幹(stem)與選項(alternatives)
- (2) 在作答方面，可分為正確答案和最佳答案兩種形式。
- (3) 形式，可分為完全的問句、不完全的敘述句、題幹置於指導語中、選項置於題幹中

### 2. 命題原則

- (1) 每題的選目宜求一致，至少三個或不超過五個
- (2) 題幹的敘述須能清楚顯示出題意，但避免過於冗長
- (3) 題幹宜求完整，避免被選目分割成兩部分
- (4) 選目盡量力求簡短，必要的敘述或相同的字詞宜放置於題幹中
- (5) 所有的誘答項目應具有似真性
- (6) 選目的敘述長度力求接近，避免暗示正確答案
- (7) 選目之間宜避免有重疊的現象
- (8) 盡量少用「以上皆非」為選目
- (9) 避免使用「以上皆是」為選目
- (10) 題幹盡量少用否定的敘述
- (11) 如屬最佳答案題型，必須確信只有一個最清楚的最佳答案
- (12) 正確答案出現在各個選目的次數宜接近相同，且須隨機排列

### 3. 選擇題的優缺點

優點	限制
<ol style="list-style-type: none"><li>(1) 比申論題等須專家評量的考題，計分更經濟且客觀</li><li>(2) 考試範圍比申論題更廣泛，內容取樣更具代表性</li><li>(3) 題意比簡答題更清楚，不會造成混淆</li><li>(4) 比是非題，選擇題的誘答選項具有教學診斷的價值</li><li>(5) 比是非題更能偵測學習效果</li><li>(6) 比是非題更能降低反應心向的干擾</li><li>(7) 比是非題更能免於猜測的影響，提高信度</li></ol>	<ol style="list-style-type: none"><li>(1) 相對於申論題等建構反應題型，較不易評量高層次的能力</li><li>(2) 相對於建構反應題，選擇題不易編寫，尤其是誘答選項</li><li>(3) 相對於建構反應題，選擇題無法免於猜測的干擾</li><li>(4) 相對於實作評量，紙筆測驗的選擇題偏重語文能力</li></ol>

### 4. 編寫選擇題的注意事項

- (1) 如果別的題型比選擇題更適合，不要堅持使用選擇題
- (2) 每個問題都與所欲評量的學習成果有著密切的聯結
- (3) 每個題目只針對一個問題
- (4) 問題的描述或複雜度必須切合學生程度
- (5) 用字遣詞應簡明，避免不必要的陳述
- (6) 避免過於弔詭的試題
- (7) 題目不要太多，以免造成猜測
- (8) 難度要適中，勿過份簡單或困難
- (9) 題號用數字，選項用文字
- (10) 題幹本身要有意義

## (六) 配合題-改良式的選擇題

### 1. 性質

- (1) 配合題可以視為改良式的選擇題
- (2) 當選項有多種可能性，且都與題幹中的敘述相關時，教師也可以從選擇題轉向命題「配合題」
- (3) 配合題的命題原則大部分與選擇題相似。

### 2. 配合題的命題原則

- (1) 在同一配合題中，各個項目的性質應力求相同
- (2) 問題項目與反應項目的數量不宜相等
- (3) 適當的問題項目宜在五至八項之間，而反應項目以超過二至三項為宜
- (4) 反應項目宜按邏輯順序編排
- (5) 問題項目避免使用不完全的敘述句
- (6) 問題項目和反應項目盡量力求簡短
- (7) 作答的方法宜清楚說明
- (8) 問題項目與反應項目宜印在同一頁上

### 3. 配合題的優點

- (1) 可在短時間內測量許多問題，因此有較多的內容樣本，可提高信度
- (2) 容易編擬題目
- (3) 如同是非題與選擇題一樣，計分容易且客觀

### 4. 配合題的缺點

- (1) 僅能測量機械記憶的事實知識
- (2) 尋找性質相同的材料不易

## (七) 填充題和簡答題

### 1. 性質

- (1) 填充題是簡答題另一種形式
- (2) 在問題的表示方法上
  - 填充題通常是使用不完全的敘述句來命題
  - 簡答題則是使用直接的問答句來命題

### 2. 填充題和簡答題的命題原則

- (1) 採用直接問答句命題比用不完全敘述句為佳
- (2) 填空的空白必須是重要概念，而非無關的字詞
- (3) 欲填寫地空白避免過於空泛
- (4) 空格不宜太多，以免無法把握題意
- (5) 每題的空白盡可能放置在末端
- (6) 清楚標明所要的精確答案
- (7) 測量定義和專有名詞的理解，以提供名詞給予界定為宜
- (8) 題目的敘述宜使填答者知道是簡短而明確的

### 3. 填充題和簡答題的優點

- (1) 題目容易邊擬
- (2) 可以減少學生的猜測
- (3) 可以用以評量數學問題的解決
- (4) 提供教師了解學生知道和不知道的東西

### 4. 填充題和簡答題的缺點

- (1) 評分比其他客觀測驗題難，書寫字跡和錯別字都會影響計分的客觀性
- (2) 所測量的能力僅是事實之事的記憶，難以測量較高層次的思考力

## (八) 申論題

### 1. 性質

- (1) 申論題一般被認為是用來測量較高層次的認知能力，而客觀測驗則是測量由低層次到高層次的認知能力皆有。
- (2) 申論題可以用來測量學生的組織、統整、歸納、問題解決和表達觀點的能力，這是客觀測驗所無法達成的。

### 2. 優點

- (1) 可以測量較高層次的學習成果（如：分析、綜合和評鑑）
- (2) 可以促進學生思考的統整和應用，以及問題解決的能力。
- (3) 可以影響及改變學生的學習習慣和方法
- (4) 可以增進學生的寫作表達能力
- (5) 命題比較簡便、容易

### 3. 限制

- (1) 效度低：試題取樣不具代表性，取樣範圍較小，無法涵蓋教材的全部內容範圍
- (2) 由於是自由組織和表達觀點，因此很難與所學習的結果牽連上關係
- (3) 信度低：評分容易受到寫作能力和作答技巧的影響；此外，筆跡、錯別字、文法結構的表達與對錯等因素，也會影響評分的公正性。
- (4) 評分主觀且不一致
- (5) 作答及閱卷均相當費時

### 4. 申論題的命題原則

- (1) 較複雜或較高層次學習結果的評量，才需要使用論文題
- (2) 問題儘量與所要測量的學習結果有所相關
- (3) 明確敘述問題，務必使學生都清楚瞭解問題的要求
- (4) 不允許學生可以選擇其中幾道試題作答
- (5) 給與學生充分的作答時間，並提示每一題的作答限制
- (6) 以多題短答的限制反應題取代少題長答的申論題

## 5. 申論題的評分原則

- (1) 妥善預擬評分要點，作為評分的依據
- (2) 評分前後要一致
- (3) 使用最適當的評分方法
- (4) 依據所預期的學習結果來評分，避免受到有關因素影響評分的客觀性
- (5) 一次只評閱一道試題。如果可能的話，由兩位以上評分者獨立評閱每一道試題
- (6) 宜在同一段時間內評完所有試卷，並且應該避免中途停頓或被打斷
- (7) 使用匿名評分

比較項目	選擇型試題	補充型試題
測量的能力	適於測量知識、理解、應用、分析等能力；但是比較不適合於測量綜合與評鑑的能力	比較不適合於測量知識的記憶；適於測量理解、應用、分析等能力；尤其是綜合與評鑑的能力
內容的取樣	使用大量的試題，涵蓋範圍較大，內容的取樣較具有代表性	使用相當少的試題，涵蓋範圍較小，內容取樣較不具代表性
編制的過程	準備優良的試題較難且費時，但評分較易且客觀	準備優良的試題較易，但評分較難且主觀
影響評分的因素	閱讀理解力和盲目猜測	寫作及虛張聲勢的能力
對學習的影響	促進學生記憶、解釋、和分析他人的觀念，可以指出學習錯誤之處	促進學生認識、統整和表達自己的觀念，比較鼓勵創造力的發展測驗的計分

# 試題分析

## 一、試題分析的功能

1. 提供回饋給學生：學生應有權利知道他在每一題上的得分及每一題的正確答案
2. 提供回饋給老師：瞭解每一個題目答對的百分比，可以提供給教師知道學生的學習困難
3. 提供改進課程的依據
4. 增進教師命題的技巧
  - 題目的功能是否如所預期的？
    - 常模參照測驗：區分高、低能力的學生
    - 標準參照測驗：測量教學的結果
  - 題目的難度是否適當？
  - 題目是否沒有缺點？
  - 每一個選目(誘答項目)是否有效？

## 二、難度分析

(一) 常模參照測驗的試題分析方法 → 用來區別學生的程度

1. 試題分析的步驟：

- (1) 依分數排序試卷
- (2) 最高分的 27%學生為高分組，最低分的 27%為低分組
- (3) 分別計算高分組與低分組在每一試題的答對人數與百分比
- (4) 計算「難度指數」  $P = P_H + P_L / 2$
- (5) 計算「鑑別指數」  $D = P_H - P_L$
- (6) 檢查每一題所列選項的回答人數，以確定誘答力

2. 難度分析 (通過百分比分析法)

- (1) 計算全體受試者答對每題的人數百分比，即為「難度指數」

$$P = (R / N) \times 100\%$$

P 試題難度，R 為答對的人數，N 為總人數

- (2) 另一種難度的求法

$$P = P_H + P_L / 2$$

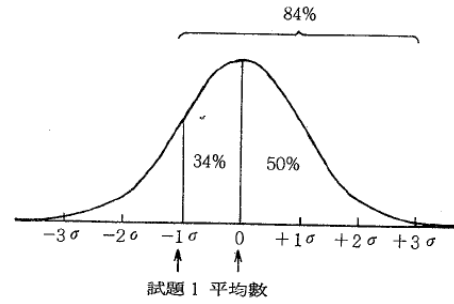
$P_H$  和  $P_L$  分別為高分組和低分組答對的百分比

- (3) 數值愈大，難度愈低
- (4) 難度指數的表示是順序量尺

### 3. 等距量尺分析法

- (1) 假定試題所測量出的特質是常態分配
- (2) 試題的難度可以使用「等距量尺」的方式來表示
- (3) 步驟：

- ① 將每個題目通過的比例，對應至常態分配表上的  $z$  分數
- ② 將  $z$  分數轉換成  $\Delta$ 。  $\Delta = 13 + 4z$
- ③  $\Delta$  界於 1-25 之間，13 為平均， $\Delta$  值愈大題目愈難



$P = .84$	則 $z = -1$	$\Delta = 13 + 4(-1) = 9$
$P = .16$	則 $z = +1$	$\Delta = 13 + 4(+1) = 17$
$P = .50$	則 $z = 0$	$\Delta = 13 + 4(0) = 13$
$P = .9987$	則 $z = -3$	$\Delta = 13 + 4(-3) = 1$
$P = .0013$	則 $z = +3$	$\Delta = 13 + 4(+3) = 25$

### 4. 試題難度的分配

- (1) 試題難度適中，分數的分配型態才會接近常態分配。
- (2) 測驗分數分配與試題難度指數分配的關係。

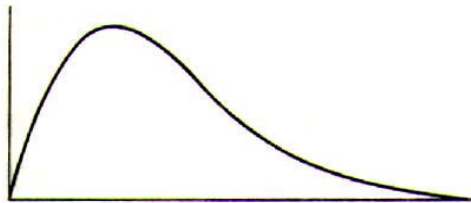


圖 8-1 測驗分數集中在低分一端

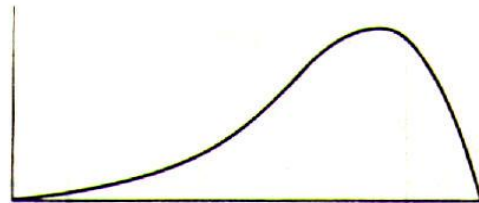
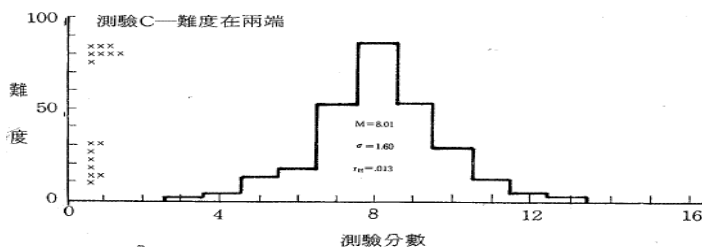
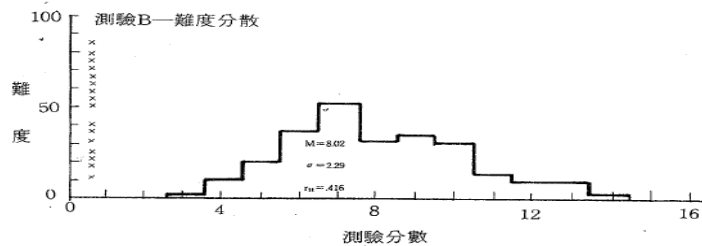
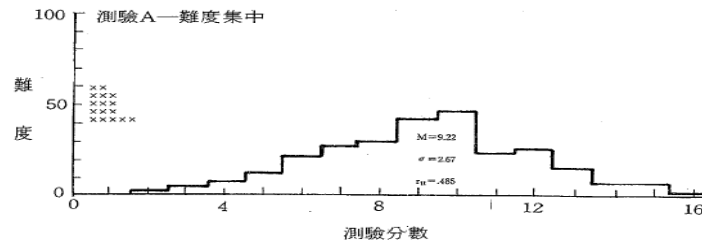


圖 8-2 測驗分數集中在高分一端

- (3) 難度指數與信度的關係：難度適中時，信度最高





### 三、鑑別力分析

(一) 意義：指測驗試題能區分受試者優劣好壞作用的大小。鑑別力是等距量尺。

(二) 目的：確定題目是否有區分能力高下的作用。

(三) 有兩種方法：

#### 1. 內部一致性分析

(1) 目的：了解各個試題的功能是否和整個測驗的功能一致

(2) 假設：

① 整個測驗的分數具有某程度的效度，單一試題的反應若和總分間具有一致性，即表示該題目有某種程度的效度

② 高分組答對的比例應顯著大於低分組的比例，該題的效用和總分的效用是一致的。

(3) 計算公式為： $D = P_H - P_L$

① 鑑別度值介於-1.00 至+1.00 之間

② 指數愈高，鑑別力愈大

③ 若指數若接近 0，代表該題沒有鑑別度

→ 題目太難或太簡單，所有人都錯或都對。

→ 題目不夠清楚，以致猜測。

④ 若指數為負值，此種題目有反向作用，理該淘汰。

#### 2. 外在效度分析

(1) 目的：檢測題目是否具有預定的某種鑑別作用

(2) 方式：

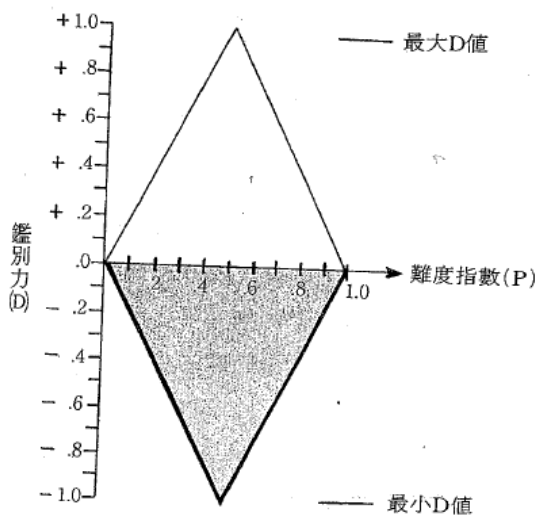
① 尋找一外在效標分數以做為區分高、低分組的依據

② 比較兩組在本測驗題目上通過的比率

➤ 內部一致性法是以預試總分來分組，而外在效度法則是以外在效標總分來分組。

#### 四、難度與鑑別度的關係

- 難度適中，鑑別力最大
- 試題的平均鑑別度愈高，測驗的信度也會愈高。



##### (一) 試題選擇的標準

1. 區分出高分組和低分組，對每一題進行試題分析，得出每題的難度與鑑別度。
2. 先選擇鑑別力較高的試題。
3. 再從中選出難度適中的試題。
  - 鑑別度在.25 以上，愈高愈好
  - **難度最好接近.50**，以.40-.70 為範圍
  - 考量測驗的「雙向細目表」以符合內容效度

##### (二) 選項的有效性分析

- (1) 依高、低分組在各個選項上選答的次數進行判斷
- (2) 每項不正確的選項，至少應有一個低分組的受試者選它；
- (3) 低分組選擇不正確選項的人數應多於高分組
- (4) 修改選項之前，要仔細考量是題目本身的缺陷，或是學生反應上的錯誤。

題 號	高分組答對百分比 (P <sub>H</sub> ) (n=100)	低分組答對百分比 (P <sub>L</sub> ) (n=100)	鑑別指數 $D = P_H - P_L$	難度指數 $P = \frac{P_H + P_L}{2}$
1	.85	.40	.45	.63
2	.62	.21	.41	.42
3	.15	.27	-.12	.21
4	.48	.10	.38	.29
5	.95	.51	.44	.73
6	.83	.32	.51	.58
7	.70	.55	.15	.63
8	.58	.42	.16	.50
9	.96	.26	.70	.61
10	.22	.13	.09	.18

表 7-3 試題的選目分析

組 別	選 目					D	P
	A	B*	C	D	未答		
高 分 組	2	16	1	1	0	.40	.60
低 分 組	5	8	4	3	0		

每個不正確的項目均有低分組受試者選答，甚至高分組亦有之，而且低分組有更多人選擇不正確答案。正確答案 B，高分組選的人多於低分組，由此可見，此題目的選目均屬合理有效。

表 7-4 試題的選目分析

組 別	選 目					D	P
	A*	B	C	D	未答		
高 分 組	14	0	0	6	0	.10	.65
低 分 組	12	0	8	0	0		

選目 C 是有效的，因為它吸引更多低分組受試選答，但是 B 是沒有誘答力的選目，因為兩組均無人選他，故須要修改。D 也是不佳選目，因為它吸引更多高分組受試選答，是否意義不清楚？或有兩個正確答案？或受試者作答不小心？如果 B 與 D 選目能加以修改，則本題鑑別力可能會因之提高

表 7-5 試題的選目分析

組 別	選 目					D	P
	A	B	C*	D	未答		
高 分 組	9	0	10	1	0	.20	.40
低 分 組	5	4	6	5	0		

高分組選答 A 與 C 的人數接近，顯見這兩個選目在正確性上不夠清楚，應修改使之只有一個答案或最佳答案。低分組對每個選答都非常接近，這表示每個選目對他們均有誘答力，可能是由於他們缺乏知識所致。

表 7-6 試題的選目分析

組 別	選 目					D	P
	A	B*	C	D	未答		
高 分 組	5	6	4	5	0	.05	.28
低 分 組	4	5	5	6	0		

高分組選答每個選目的次數都很接近，這顯示他們可能盲目的猜測，由於盲目的猜測會增加測量的誤差，故像此種題目需加以檢討或修改

表 7-7 試題的選目分析

組 別	選 目					D	P
	A	B	C	D*	未答		
高 分 組	3	6	3	8	0	- .20	.50
低 分 組	3	1	4	12	0		

高分組選答正確答案的次數反而比低分組低，表示此是反向作用題目，須檢討原因加以修正，或淘汰不用。

### (三) 試題分析的解釋

1. 鑑別指數不等於題目效度
2. 鑑別指數低未必表示試題有缺點
3. 試題分析資料是假設性的
4. 避免只根據統計的特性來選題

# 統計相關

1. **迴歸分析**：探討數據間是否有一種特定關係→預測。用來分析一個或一個以上自變項與依變項的數量關係，以了解當自變項為某一水準或數量時，依變項的反應數量或水準。
  - (1) 簡單迴歸：只有一個自變項預測一個依變項，例如：社經地位→學業成就；
  - (2) 複迴歸：利用兩個或兩個以上的自變項去預測一個依變項，例如：動機、社經地位、智力、性向→學業成就。但迴歸只能用來做預測，無法做因果推論。
2. **共變數分析**：是一種統計控制的方法，旨在利用統計的方法把影響實驗正確性的誤差(=共變項所帶來的誤差)加以排除，目的在增進統計考驗力和降低實驗誤差。如把教學前可能的因素排除，就要用共變數分析，將共變項。
3. **路徑分析**：又稱「因果分析」或「結構方程式分析」或「同時方程式考驗模式」，用以討論多重變項間因果結構模式統計技術，同時將所有預測變項(自變項)納入迴歸模式中，路徑分析由一系列迴歸分析組成，使用迴歸分析原理，透過假設性架構，將不同方程式加以組合，形成結構化模式。
4. **t 考驗**：指二組平均數的差異比較，例如：教學後的成績是否比教學前的成績高？
5. **變異數分析**：是使用 F 考驗的一種統計方法，檢驗兩個以上獨立樣本平均數差異情形是否達到顯著水準，例如：三種不同教學法，學生數學成績的不同。
6. **多變項分析**：同時分析兩個以上變數計量，如三種不同教學法，學生國英數成績的不同。
7. **三角檢證**：是指使用多種方法來研究同一現象，是質化研究中不可缺少的工具，Denzine 將其分為資料三角檢證、研究者三角檢證、理論三角檢證、及方法論三角檢證等四種。
8. **因素分析**：主要目的是對資料找出其結構，以少數幾個因素來解釋一群相互有關係存在的變數，而又能到保有原來最多的資訊，再對找出因素的進行其命名，如此方可達到因素分析的兩大目標：資料簡化和摘要。
9. **卡方考驗  $\chi^2$** ：用來表示類別資料之間的關係的統計分析方法。
  - (1) 一般多用於調查研究中，用來處理人數、次數等資料，如「男女生在選擇旅遊地點時，是否會有明顯的不同？」、「國中三年級學生在打電動玩具的經驗上是否有所不同？」、「國小高年學生上過兩性平等課後，對兩性平權觀念態度上是否有所改變？」
  - (2) 卡方考驗的統計原理：卡方檢驗是將樣本觀察次數(或百分比)與理論或母群體次數(或百分比)(期望值)相比較，來看二者是否有差異性，卡方值越小，表差異越小，越不易達顯著，卡方值越大，表差異越大，越容易達顯著。
  - (3) 卡方檢定使用時機
    - ① T 考驗：考驗兩個小樣本的平均數是否有很大的差異，而可能來自於不相同的母體
    - ② Z 考驗：當常態分配的母體，其平均數與標準差為已知，考驗樣本是否可能來自於母體或與此母體相近。
    - ③ F 考驗：考驗兩個變異數間有無顯著差異
10. **相關分析**：對總體中具有因果關係標誌的分析。

Pearson 積差相關	特性：數值穩定、標準誤小。 例：工作時數與收入的關係。
Spearman 等級相關	特性：適用於二個評分者評 N 件作品， 或同一位評分者，先後二次評 N 件作品。 例：兩位評審對 N 件學生作品之評定。
Kendall 等級相關	特性：相當簡便 例：兩位評審對 N 件學生作品之評定。
Kendall 和諧係數	特性：特別適用於評分者間信度； 考驗多位評審者對 N 件作品評定等第之一致性。 例：多位評審對 N 件學生作品之評定。

11. **描述統計**：就是計算樣本的各種統計數，沒有對母群進行推論。
12. **多變量統計**：主要用於分析擁有多個變數的資料，探討資料彼此之間的關聯性或是釐清資料的結構。例如：老師的性別、年齡、學歷等變數是否影響對常態編班意見的不同。
13. **自/獨立變項 Independent variable**：(因) 是指在實驗控制情境下，實驗者有系統地處理的變項。
14. **依變項 Dependent variable**：(果) 指由自變項的變化所導致的變化事項。
15. **使用大樣本的時機**
  - (1) 研究的母群體相當異質。
  - (2) 研究的干擾變項難以控制(採大樣本可使干擾變項的影響相互抵消)
  - (3) 測量工作的信度較低。
  - (4) 統計水準定得較小。
  - (5) 較大的統計考驗。
  - (6) 要進行子群的分析。
  - (7) 預期相關研究的相關性較低。
16. **簡單隨機抽樣 simple random sampling**：必須合乎「均等的概率」和「獨立的概率」。如果在抽樣時，母群內的每一個體均有相同的概率可以被抽取為樣本，就合乎均等概率；母群內的某一個體被抽為樣本時，並不影響到其他個體被抽為樣本的概率時，就合乎「獨立概率」。
17. **系統性抽樣 systematic sampling**：抽樣時所用的名單或稱架構 frame 是按照字母順序或其他系統性順序排列，而研究者按此名單順序每隔 k 個元素抽取一個元素為樣本。
18. **分層隨機抽樣 stratified random sampling**：組間(各組之間)異質，組內(受試者之間)同質。抽樣之前研究者知道母群內各「層次」strata 成員的比例，而按照這比例自每一層次中隨機抽取樣本，就是分層隨機抽樣，層間的變異儘量大，層內的變異要儘量小。最能平均分配樣本，所以**最具全國代表性**
19. **叢集抽樣 cluster sampling**：組間同質，組內異質。抽樣單位是一群元素所構成的集合，而不是各元素，並不是母群內所有元素都有均等概率被選為樣本。將母體分成幾個群集(或部落、區域)，而群集間的變異小，群集內的變異大。例如：從十個班級中抽三個班級，或從十個縣中抽兩個。
20. **立意/專家/判斷抽樣**：指研究人員根據某種目的，刻意尋找具備某種特質的個體來組成研究樣本

21. 適度性指標：(1)邊緣機率、(2)近似比值、(3)估計的能力變異數

22. 戈德門 Goldman 根據所期待解答的問題性質，將資料的解釋分成四種類型

1.	描述型 descriptive	根據資料描述個體身心的特性與現況。 如：兒童的人格特質如何？其認知的能力如何？其語文的能力與數理能力比較起來如何？他的興趣為何？
2.	追溯型 genitive	主要目的在追溯兒童發展現狀的形成原因，及其演變過程。 如：該名兒童的學習困擾是因為情緒困擾？或是基本技巧不夠？或是缺乏興趣？或是來自父母過度期待的壓力？或是過去的失敗經驗？
3.	預測型 predictive	根據過去和現有的資料，預測或推斷個體未來可能的發展。 如：此兒童進入小學的數學表現將如何？適應狀況如何？
4.	評鑑型 evaluative	根據既定的標準，對學生的表現作價值判斷，並就其抉擇提供建議。 如：該兒童應該選擇音樂班或語文資優班？

上述四種解釋方式，就層次而言，自描述型至評鑑型，其層次愈來愈深，描述型僅是說明個體的現況，追溯型與預測型則以現況為基礎，分別深入個體的過去以及探究其未來，最後評鑑型的解釋則可以提供價值判斷和決策建議。

實驗法	<p>是一個簡潔有力的工，幫助我們排除或控制兩個因素以外的影響。</p> <ol style="list-style-type: none"> <li>1. 研究者可以介入他所要觀察的情境</li> <li>2. 研究的重點是在找出獨立變項與依變的關係</li> <li>3. 兩個基本特徵為 (1)控制變項 與 (2)隨機化</li> </ol>	
	優點	缺點
	<ol style="list-style-type: none"> <li>1. 因果關係：實驗有助於控制變項的時間順序，幫助建構因果關係。</li> <li>2. 控制：透過對環境、變數與實驗對象的控制，研究者可以設計出獨立而不受常態活動影響的測驗情境。經由環境的控制，減少干擾。</li> <li>3. 重複性高。</li> </ol>	<ol style="list-style-type: none"> <li>1. 人為性：不自然，與真實的世界不符合</li> <li>2. 實驗者的偏見：如為符合顧主的期望，私人的利益等。偏見會使觀察、資料記錄、數學運算和解釋等一環節產生錯誤</li> <li>3. 受試者的偏差</li> <li>4. 倫理：為個人的目的</li> </ol>
調查法	<ol style="list-style-type: none"> <li>1. 目的：希望藉由訪問少數具有代表性的人或團體，而能了解社會大眾的生活狀況、日常活動、以及他們對各種人物事件的意見與態度，不同於實驗法，而是去發現自然產生的變化。亦即，調查法是一種收集、組織和分析資料的方法。</li> <li>2. 適用範圍：人口特質資料、社會與經濟狀況、各類活動、意見與態度</li> </ol>	
	優點	缺點
	<ol style="list-style-type: none"> <li>1. 可用來調查真實環境，而非實驗室的人為條件</li> <li>2. 可依預算及需要選擇不同的調查方式。</li> <li>3. 易從各種人士中採集大量資料，大部份調查不受到地理的限制。</li> <li>4. 有助於調查研究的資料業已存在。如：檔案資料、政府文件可作為第一手的資訊來源（主要資料來源）或二手資訊來源（論據資料）</li> <li>5. 可同時處理多變項和假設</li> <li>6. 可使用大量樣本</li> </ol>	<ol style="list-style-type: none"> <li>1. 無法像實驗法控制變項，因果難証。</li> <li>2. 問卷詞句的表達會影響結果。</li> <li>3. 有找錯對象的可能。</li> <li>4. 某些研究越來越難展開，如拒訪</li> <li>5. 訪員的程度</li> <li>6. 受訪者能力／偏向</li> </ol> <p>(1) 輕重不分、焦點不清</p>
參與觀察法	<p>是一種實地或直接的觀察，研究者為了對某一團體有所有謂科學的了解，而參與此團體，建立並保有多面向的長期關係，以利其研究。</p> <p>特別是企圖回答文化情境脈絡有關的問題都可以參與觀察探索，如：希望了解一個生活環境的活動和互動，如何對某種行為和信仰賦予意義。</p>	
	優點	缺點
	<ol style="list-style-type: none"> <li>1. 第一手觀察到的資料</li> <li>2. 突破傳統研究的限制(時間、空間)</li> <li>3. 在金錢與設備上的花費較經濟</li> <li>4. 在自然的狀態下進行，因此，所得到的結，可能比其它方法有更精確結果</li> <li>5. 可追蹤觀察</li> <li>6. 從觀察者的表現得到多元的訊息</li> </ol>	<ol style="list-style-type: none"> <li>1. 時間較長</li> <li>2. 環境的不適應/被影響</li> <li>3. 道德壓力</li> <li>4. 情感易投入，喪失客觀性（如弱勢團體）</li> <li>5. 非系統性誤差（研究本身的立場、價值與選擇）</li> <li>6. 跨文化的限制</li> <li>7. 推廣（論）性低</li> <li>8. 易出現倫理問題：欺騙、衝突、涉入</li> </ol>



內容分析法	<ul style="list-style-type: none"> <li>透過量化技巧與質的分析,以客觀和系統態度,對文件內容進行研究與分析,藉以產生該項文件內容的環境背景,及其意義的一種研究方法。</li> <li>其程序包括：1.選擇研究題材→2.熟悉研究題材 →3.提出研究問題或假設 →4.抽樣 →5.資料收集 →6.資料分析與解釋</li> </ul>	
	<ul style="list-style-type: none"> <li>內容分析法:使用大樣本時機：               <ol style="list-style-type: none"> <li>當許多未控制的因素會產生影響時</li> <li>當預期變項間的差異或相關很小時</li> <li>研究的組別必須分成次要組別時</li> <li>當母群體被研究的變項甚為異質時</li> <li>當研究的依變項在測驗的信度較低時</li> </ol> </li> </ul>	
	優點	缺點
	<ol style="list-style-type: none"> <li>以量化、系統的方法來探究傳播內容</li> <li>只要傳播內容的蒐集沒問題，研究者便較能掌握整個研究</li> </ol>	<ol style="list-style-type: none"> <li>內容分析法常需配合對受眾的調查</li> <li>內容分析成果受限於研究者的分類與定義</li> <li>須檢視相當大的資料，需大量的人力與時間內 容分析著重的是內容呈現的差異性，但對於內容所隱藏的意識型態，以及影響內容呈現的結構性問題，較難依此方法評估</li> </ol>
民族誌	<p>應用參與觀察法和密集的實地工作</p> <p>使這個研究取向與眾不同的，則是該取向係從文化觀點(culture perspective)來闡釋和應用其發現結果。</p>	