

# 心理測驗與評量講義

## 第一回

403110-1



社團法 考友社 出版發行

# 心理測驗與評量講義 第一回

## 目錄

第一講 心理測驗之基本概念.....	1
命題大綱.....	1
重點整理.....	3
一、測量、評量與測驗.....	3
二、心理測驗的發展.....	5
三、心理測驗的結構.....	6
四、心理測驗的類型.....	6
五、心理測驗的應用領域與注意事項.....	12
六、學校測驗計畫.....	13
七、心理測驗之相關議題.....	17
精選試題.....	21

# 第一講 心理測驗之基本概念



## 一、測量、評量與測驗

- (一)測量的意義與特性
- (二)評量的意義與特性
- (三)測驗的意義與特性
- (四)測驗與評量的負作用

## 二、心理測驗的發展

- (一)十九世紀對智遲兒童的訓練
- (二)十九世紀實驗心理學家的研究
- (三) F. Galton 和 James Mckeen Cattell 的早期研究

## 三、心理測驗的結構

- (一)行為樣本
- (二)標準化
- (三)測量的難度和客觀
- (四)信度
- (五)效度

## 四、心理測驗的類型

- (一)依測驗人數區分
- (二)依測驗時間點區分
- (三)依測驗性質區分
- (四)依測驗功能區分
- (五)依測驗材料區分
- (六)依標準化過程區分
- (七)依測驗試題難易區分
- (八)依測驗評分方式區分
- (九)依測驗反應型態區分
- (十)依測量分數解釋區分

## 五、心理測驗的應用領域與注意事項

- (一)應用領域

## 403110-1

(二)注意事項

### 六、學校測驗計畫

(一)心理測驗在教育上的應用

(二)心理測驗之誤用情形

(三)心理測驗在教育上的功能

(四)學生資料解釋之模式

(五)解釋測驗結果的原則

### 七、心理測驗之相關議題

(一)傳統測驗理論

(二)試題反應理論

(三)電腦適性測驗

(四)心理測驗發展新趨勢



## 一、測量、評量與測驗

### (一)測量的意義與特性：

- 1.測量係指獲得個體具某種特質之數量的過程；亦即以數量來描述某個體具有某種特質之程度的過程。
- 2.在本質上，測量是使用數量來描述特質，而非使用文字或品質描述。
- 3.測量最常使用紙筆測驗執行。尚可使用評定量表、晤談、觀察、檢核表、問卷等測量工具。
- 4.例如：利用比西量表（Binet Simon Scale）測量個體，獲得其智力商數（Intelligence Quotient），並用此商數來描述其智力特質的過程，即為測量。
- 5.測量量尺：係將研究搜集到的數量資料，把數字分派於人、事件和物體上的法則。心理與教育研究中，有四種不同測量量尺，即名義量尺、次序量尺、等距量尺、比率量尺。

#### (1)名義量尺：

- ①又稱類別量尺，其係依據特質的同等性而將其分類，並分派數字給每一種類別，此種量尺即名義量尺。
- ②例：性別、職業、種族等，均屬之，其除了說明類別之不同並易於指認外，別無其他意義。
- ③此種量尺之層次最低，不僅不能表示大小、好壞、高低之順序，亦不適合作代數運算。

#### (2)次序量尺：

- ①此種量尺不但可指出類別的不同，亦可說明大小、高低、優劣的順序，但無法表示類別大小之間差異量，其層次較名義量尺為高。
- ②例：3 個學生之作文得甲、乙、丙 3 個分數，除可說明甲優於乙優於丙外，但無法指出甲、乙之間的差異量等於乙、丙之差異量。

#### (3)等距量尺：

- ①此種量尺除具有名義量尺及次序量尺之特性外，尚可說明數量差

異大小的重要特性，其主要特質在於各分數之間的單位相等，但此種量尺的起始點是任意選擇的，無所謂的真正零點。

- ②例：3 天的溫度分別是  $20^{\circ}$ 、 $25^{\circ}$ 、 $30^{\circ}$ ，我們可說  $30^{\circ}$  高於  $25^{\circ}$  高於  $20^{\circ}$ ，且可說  $30^{\circ}$  與  $25^{\circ}$  之差等於  $25^{\circ}$  與  $20^{\circ}$  之差，但無法說  $30^{\circ}$  為  $20^{\circ}$  之 1.5 倍。

(4)比率量尺：

- ①此種量尺除具有上述 3 種量尺之特性外，尚具有絕對的零點，為最高層次的量尺，故除了可指出類別、大小、順序和差量之外，尚可表示出比率之大小。

- ②例：重量 50 公斤重於 25 公斤，但二者相差 25 公斤，另外 50 公斤為 25 公斤之 2 倍。

6.測量的特性：

- (1)個體具有可測量特質：測量時個體具有真實分數，用以算出誤差分數。
- (2)心理與教育的測量是間接的：心理特質屬於抽象，只能間接測量或根據測驗行為結果推論個體。
- (3)測量均有誤差存在：實際測到分數，並非真正分數，總會有誤差存在。測量誤差分系統誤差與非系統誤差。
- (4)測量時常是相對的：測量得到的分數皆為相對分數，而非絕對分數。測量分數必須和他人比較才有意義，如「常模參照測驗」為相對性質。

(二)評量的意義與特性：

- 1.評量係採用科學方法與途徑，多方面蒐集適切並量化的資料，再參照合理的衡量標準，加以比較分析與綜合研判的過程。
- 2.評量所涵蓋內容比測量還大，因此測量只是評量的一部分。測量是依據數量做客觀的描述，而評量除測量數字的依據，尚包括主觀價值判斷。

(三)測驗的意義與特性：

- 1.狹義：測驗是測量的工具與材料。如智力測驗、性向測驗…等。可以是文字或圖形較有利於幼兒、文盲、不同文化及族群表示。
- 2.廣義：測驗是指採用數字量尺或分類系統，來觀察和描述個人特質的一種系統程序；亦即是指採用一套標準的刺激，對個人的特質作客觀測量的有系統程序。根據此定義，測驗具有以下 3 個基本觀念：
  - (1)標準的刺激：標準的刺激是指測驗中讓受試者回答的試題，或讓受試者操作的工作。

- ①試題可以是圖形的，也可以是語文的；工作則如積木、方塊或儀

器。

②無論是試題或工作，它們必須有適當的難度，而適當難度的試題，受試者在測驗上的分數才能呈常態分配。

③試題或工作必須數量適當而具代表性，能概括所欲測量的特質。

(2)個人的特質：測驗所要測量的特質，包括認知能力和情感特質兩方面。

①認知能力：包括性向、智力、學習成就、思考能力與創造能力等。

②情感特質：包括人格特質、動機、興趣、態度、自我觀念與情緒穩定等。

(3)有系統程序：有系統程序就是指測驗的標準化過程；亦即指測驗在實施、記分與解釋程序上的一致性。

①測驗的情境：包括具有適當的光線、溫度、通風、座位與安靜的條件。

②實施的步驟：包括指導說明、時間控制、疑問處理、主試者與受試者的和諧關係等，均須加以控制，以免影響受試者的測驗成績。

③測驗的記分：必須客觀，且依據常模或標準來解釋測驗分數所代表的意義。

(四)測驗與評量的負作用：

1. 測驗可能侵犯學生的隱私權。
2. 測驗可能會造成學生焦慮並干擾學習過程。
3. 測驗可能會將受試者分類，會造成學生心靈傷害。
4. 測驗可能對具有創意性與聰明的學生不公平。
5. 測驗可能對語文能力不佳學生不公平。
6. 測驗的範圍可能只涵蓋部分的學習內容。
7. 測驗可能造成教師自我判斷的效果。

## 二、心理測驗的發展

(一)十九世紀對智遲兒童的訓練：

1. 十九世紀興起一陣研究智遲者和精神病患的風氣，許多機關開始建立分類智遲者和精神病患的標準，另外，使用語文材料來區別智遲者的心智程度。
2. 訓練智遲者，首見於法國一名醫師，採用感覺和肌肉訓練的方法，成為往後智力測驗中實作和非語文測驗的主要內涵，貢獻良多。

(二)十九世紀實驗心理學家的研究：

1. 十九世紀初期的心理學家探討的目標是人類在相同情境下不同的行為表現稱為個體差異。
2. 迄十九世紀末期，實驗心理學家開始使用科學儀器和方法探究人類的行為，研究內涵包括視覺、聽覺等。

(三) F. Galton 和 James Mckeen Cattell 的早期研究：

1. F. Galton的研究貢獻以人類行為和遺傳關係的相關為研究重心，並以統計的方法分析個別差異，量化心理測驗的結果。
2. James Mckeen Cattell於 1890 年首先在其著作中提及「心理測驗」(mental test) 的名詞。
3. Cattell的研究以個別反應時間的差異為主，測驗採個別施測，包含有肌肉力量、痛覺的敏感度、聽覺、重量辨識、反應速度等，在十九世紀末時，其測驗廣泛使用於學校和一些機構，測驗是以感覺為主。

### 三、心理測驗的結構

- (一) 行為樣本：心理測驗是針對個體的行為樣本作客觀和標準化的測驗，透過適合的測驗可以預測行為樣本的未來行為。
- (二) 標準化：欲求測驗結果的可靠性，須有標準化測驗的過程、情境、答案和記分方式。測驗標準化的重要工作是建立常模 (norms)，意謂個人在測驗中的表現和意義。
- (三) 測量的難度和客觀：測驗的實測、記分和解釋結果不因主試者的主觀因素而影響其結果是測量的客觀條件；測驗的難度以通過每個題目人數的百分比為依據；選擇測驗的題目須考慮到題目的難度和客觀性。
- (四) 信度：信度是指測驗結果的一致性，測驗的信度愈高愈佳，但與效度未必成正相關。
- (五) 效度：效度是指測驗是否達到測驗功能的程度，高效度的測驗能預測個人未來行為。

### 四、心理測驗的類型

(一) 依測驗人數區分：

1. 個別測驗：同一時間內，只能實施一個人的測驗，採面對面實施。如比西量表與魏氏兒童智力量表均屬之。
  - (1) 用途：個別測驗是一種臨床工具，其主要用途在於診斷，適合作個案研究之用，例如特殊兒童之甄試。
  - (2) 優點：主試者與受試者建立親善關係，深入了解受試者反應，在教育上具有診斷功能。
  - (3) 缺點：
    - ① 實施不易，費時、費事，難以大量採用。





## 精選試題

一、何謂測驗焦慮？如何減低受試者的測驗焦慮，避免影響成就測驗的結果？

答：(一)測驗焦慮 (test anxiety)：係受試者對測驗產生焦慮不安的情緒。施測前，主試者可試著緩和受測者的情緒，諸如測試結果的保密、測驗的功用等。不過，輕微的焦慮有助於測驗的結果。

(二)減低受試者測驗焦慮的方法：

1. 建立關係：

主試者應了解受測者的生理、心理發展，測驗的目標、性質、年齡、性別等相關的詳細資料，以便能在測驗之前從容不迫地和受測者寒暄一番，了解受測者目前的心理狀態和情緒，緩和彼此的陌生、衝突等，讓受測者在稍放鬆心情的狀態下做測驗。

2. 有效的溝通：

建立彼此的默契，如受試者如何告知主試者需要協助，主試者該用何種方式協助受試者，還有一些受試過程中的特殊狀況的溝通，如上廁所、飲水的問題等，可透過有效的溝通來達到共識。

3. 測驗的相關資源：

如測驗的時間起迄、使用的資源、工具，測驗的目標、常模、信度和效度等，在測驗前告訴受測者，讓其減輕焦慮亦除去對測驗的未知而產生的莫名恐懼感。另外對於測驗的流程也須讓受試者清楚地了解，以便於能順利完成測驗。

4. 主試者對測驗完全地熟悉：

主試者須熟知指導語（即便是團體測驗亦是如此），施測流程和必須使用的工具資源材料等，並且提早和受測者見面並確定所需的資料齊全，建立受測者的信賴感和解除其不確定性的焦慮情緒。

5. 回饋：

在整個測試的過程中，給予受試者安全無顧慮的測試環境，隨時注意受測者的需求，並即刻給予適當的回饋和適時的協助，使測試能順利進行。

## 二、從心理測驗發展歷史來看，教育的普及、人權的重視及科技的發展等，對心理測驗的發展有何重要影響？

答：(一)教育普及：由於受教育的人數和廣度增加了許多，民智開化的程度高於心理測驗剛萌芽、成長的階段，人們正視心理測驗的正面意義和功能，且由於教育普及，測驗的信度、效度和樣本取材更為可靠、正確、廣佈於一般大多數的民衆，使得心理測驗不再只是針對特殊的個體和某些特定的人們，而能更測入且廣泛的發展不同領域、年齡、性質特徵、研究需要等的心理測驗類型，以符合人類發展趨勢的需求。

### (二)人權重視：

- 1.最早的心理測驗者針對特殊狀況者做測驗和研究，諸如心理、精神異常、智能不及同齡發展的受試者等，受試者多受到不平等的待遇，諸如拷問、逼供、嘲弄等，不過，心理測驗由於人權的重視情境下，受測者逐漸能得到公平、人性的對待，由行為學派、認知學派到人本主義心理學的理論發展過程發現其對人性的看重逐漸攀升。
- 2.人格測驗是對人權重視的另一個例子，其測驗諸如人格問卷、興趣、價值等的測驗，是以個體為主角，非以研究、實驗為出發點。

### (三)科技發展：

- 1.測量工具的進步、普遍使用，跨文化的使用評估等，最明顯的例子是電腦的使用幾乎是計分和分析的主要工具，諸如興趣測驗中的測驗結果顯示等。
- 2.由於科技的發達，促進心理測驗的快速發展，省卻諸多麻煩、複雜的過程，縮短其發展的時間，如知識般爆炸性的發展。

## 三、何謂常模參照測驗？何謂標準參照測驗？其性質為何，試比較說明之。

答：(一)常模參照測驗（norm-referenced test）：指測驗結果，根據分數在團體中位置而加以解釋，採用相對性標準比較，一般以平均水準為參照點。如大學聯考、國家高普考試、月考、期末考等教師自編測驗、標準化成就測驗、智力測驗、性向測驗。

(二)標準參照測驗（效標參照測驗，criterion-referenced test）：根據教學前事先所訂定的絕對性標準加以解釋測驗。一般均以學生所學習知識或技能，判定其「及格或不及格」、「精熟或不精熟」、「通過或不通過」，不須參考他人表現來比較。如國小教師自編測驗（專為平時考、小考、隨堂測驗）、國家技師執照考試、Bloom 提倡「精熟學習」、中醫師檢定考試、汽車駕照考試等。

## (三)常模參照測驗與標準參照測驗異同點：

	常模參照測驗	標準參照測驗
1. 相同點	(1)均與學習工作的成就領域相關，且均具有甄別學生目的。 (2)皆需明確陳述教學目標作為編製測驗之原則。 (3)均參考應用同一套常見編擬有效試題原則。 (4)儘量控制影響誤差的各種因素。 (5)均使用各種不同測驗類型。 (6)均會重視有利於測驗結果解釋的各種因素。 (7)二者皆重視測驗結果的信度。	
2. 量尺準點	中間、事後決定的	二端、事前決定的
3. 變異性	分數變異性愈大愈好	分數變異性愈小愈好
4. 計分方式與解釋	百分等級或標準分數	二分類數字
5. 用途	分班編組 (安置性、總結性評量)	補救教學 (形成性、診斷性評量)
6. 試題代表性	學習範圍較廣，每一範圍 試題較少(強調試題的鑑 別力)	學習範圍較窄，每一範圍 試題較多(強調試題在學 生學習工作表現)
7. 測驗計畫性質	使用雙向細目表	使用詳細的教材領域細目 表
8. 評量功能	鑑別	檢定
9. 效度考驗	內容效度、建構效度、效 標關聯效度	內容效度

## 四、測驗使用者拿到測驗指導手冊之後，需根據那些標準，以評估決定該測驗是否適用？

答：測驗使用者藉由測驗的效度、信度、常模和實用性去評估一份測驗是否適用。茲說明如下：

(一)效度 (validity)：效度是測驗中重要的一個指標，意謂著顯示測驗分數的正確度高與否，並且藉由測驗結果得知是否能測試出其測驗所欲呈現的特質。因此，一份測驗的效度愈高愈具有使用的價值存在，其中效度係數 (validity coefficient) 若能在.80 以上，則是一份值得參考與測試的測驗。如一份效度高的自然科學成就測驗能區分不同自然科學程度的學生，並且自然科學程度高的學生得分高於自然科學程度低的學生，但