



心理測驗

一、試述心理測驗標準化（standardization）的意義。

【參考解答】

心理測驗的目的在於利用科學的方法，將個人的心理特質測量出來，以區分出個別差異或群體間的相對差異，因此，測驗工具需以系統性與科學性的方法來建立。所謂心理測驗的標準化，主要包含測驗編製的標準化、實施程序的標準化、計分與解釋的標準化三部份，以下分別敘述之。

(一)測驗編製的標準化：

一個標準化測驗的編制過程，一般來說，包含下步驟：

1. 確定測驗目標
2. 發展測驗題目
3. 選擇題型、編擬試題
4. 預試
5. 試題評鑑與分析
6. 編輯正試題目
7. 正式施測
8. 考驗信、效度，以及常模
9. 編寫指導手冊
10. 修訂測驗

(二)實施程序的標準化：

1. 施測情境的控制：舉凡施測情境的空間擺設、溫度、照明度、測驗擺放方式等等，均需按指導手冊中的規定流程安排。
2. 施測步驟的標準化：如指導語的說明、施測的順序等等，亦應遵守指導手冊中的規定而進行。

(三)計分與解釋的標準化：

計分方法應有一定的方式，而儘量免除主觀評分的影响。且測驗的解釋方式、解釋順序與解釋語言與注意事項，都應有一定的規範。此部份亦可遵照指導手冊的規定，以使整個測驗的使用都符合一致、標準的流程。

二、請說明試題分析的主要方法及內涵。

【參考解答】

一份良好的測驗的編製過程，需要經過試題分析，方能提升試題品質，提高測驗的信度與效度，並建立起優良的試題題庫。一般而言，試題分析可分為質的分析與量的分析兩部分。以下分別進行說明：

(一)質的分析：

1. 針對試題的內容和型式，從取材的適切性與編擬試題技術方面加以評鑑。
2. 包括內容效度的評鑑和有效命題評鑑。

此部份可協請理論專家或測驗專家協助進行。

(二)量的分析：

是基於試題經過預試的結果，根據試題的統計特性逐一分析其難度，鑑別度與選項有效性的評鑑，做為修改試題或選擇試題的依據。

1. 難度分析：

- (1) 試題的難易程度，通常以全體受試者答對或通過該題的百分比表示該題的難度。或是以某一題高分組答對的百分比，加低分組答對的百分比後，再除以2。
- (2) 難度通常以P值表示，P是一種順序尺度，當P值愈大時，則該試題難度愈低；反之，P值愈小，難度愈高。分析题目的難度最主要的目的乃在選取難度適當的题目。

2. 鑑別度分析：

所謂鑑別度，是指測驗試題能區分受試者優劣好壞作用的程度。而鑑別度的分析可分為內部一致性（Internal Consistency）與外在效度（External Validity）兩種方式。

(1)內部一致性的分析：其目的在於檢查各別試題與整個測驗的功能是否符合一致，其主要方法有兩種，第一種是以高分組答對或通過某題的百分比，減去低分組答對或通過的百分比所得的差數，作為鑑別指數（Index of Discrimination）。第二種是以受試者對每一個別試題答對或答錯的情形，與其測驗總分求相關係數，以表示試題內部一致性的高低。

外在效度的分析：目的在於檢驗每一個試題是否具有預定的某一種鑑別作用。其求法與內部一致性之方法相同，不同之處在於高、低分組是以外在效標之分數而決定。

3. 選項有效性分析（誘答選項分析）：

此分析的目的在于檢視測驗中，每道題目選項的反應型態。良好的誘答選項設計為「不會的人，會從錯誤的選項中隨機挑選，所以每個錯誤的選項選擇的人數會一樣。」誘答力過強或過弱的選項，將影響題目的難度，減低測驗的信、效度。

三、請解釋何謂效度？舉例並說明四種不同種類的效度證據？解釋影響效度的相關因素？

【參考解答】

(一)效度之意義：

一個測驗如能測量出其所欲測量的屬性，此測驗即屬有效。效度即是指此測驗的有效性。易言之，效度表示一個測驗能否（或在何種程度內）測出其所欲測量的屬性，或是測驗分數能代表測量到想要測量之潛在特質的程度。

(二)四種不同效度證據：

1. 內容效度：內容效度（content validity）是指測驗內容的代表性或取樣的適切性（adequacy）。如果一位教師給學生做一個數學成就測驗，該測驗的題目涵蓋了教學所欲達成的目標，及教材的重要內容，我們就可說該測驗有內容效度。內容效度可以使用「雙向細目表」加以建立及判定。
2. 校標關聯同時效度：係指測驗分數與當前的效標資料之間的相關而言。這種效度常用的效標資料包括在校學業成績、教師評定的等第、其他同性質測驗的結果等。
3. 校標關聯預測效度：係指測驗分數與將來的效標資料之間的相關而言。其常用的效標資料包括專業訓練的成績與未來實際工作的成果等。
4. 構念效度：所謂構念效度，就是指測驗能夠測量到理論上的構念或特質的程度。易言之，就是指測驗分數能夠依據某種心理學的理论構念加以解釋的程度。考驗構念效度的方法有：相關研究法、團體差異分析、實驗研究、內部一致性分析、因素分析、多項特質—多項方法分析等。

(三)影響效度的因素

1. 測驗的品質：試題是構成測驗的要素，測驗之效度取決於試題的性能。舉凡測驗的題材、測驗的長度、試題的鑑別度、難度及其編排方式等皆與效度有關。若測驗材料經審慎的選擇，測驗的長度恰當，試題具有相當的鑑別力且難易適中並作合理的安排，則效度高；反之則效度低。
2. 測驗的實施：測驗的實施程序是影響效度的重要因素。若主試者能適當控制測驗情境，遵照測驗手冊的各項規定而實施之，則可避免外在因素影響測驗結果的正確性。在測驗實施的過程中，無論是場地的佈置，材料的準備、指導語的說明、時間的限制，均需安指手冊之說明，否則必然使效度降低，失去測驗的意義。
3. 受試者的因素：受試者的興趣、動機、情緒、態度和身心健康狀況等，都足以決定其在測驗情境中的行為反應，而受試者是否充分合作與盡力而為，均能影響測驗結果的可靠性與正確性。無論是能力測驗或人格測驗，惟有藉受試者的真實反應，始能推斷其心理特性如適應狀態。
4. 效標的品質：選擇適當的效標是經驗效度的先決條件，若因所選的效標不當，以致測驗的效度不能顯現出來，則測驗的價值可能被淹沒。一個測驗因其採用的效標不同，其效度係數可能大相逕庭。
5. 團體的性質：效度考驗所依據的樣本，必須能代表某一測驗的所擬應用的全體對象。一個測驗應用於不同的對象，由於他們在性別、年齡、教育程度與經驗背景上的差別，其測驗功能不一致，

效度亦隨之而異。此外樣本的異質性與測驗的效度係數有關，如果其他條件相等，樣本分數的差距愈大，則效度係數也愈高。

四、請就百分等級常模與標準分數常模分別解釋與舉例說明，並比較二者之建立方法。

【參考解答】

常模的建立，是一份良好的測驗所需具備的特徵，其功能有二：(1)指出受試者分數在常模團體中的相對位置；(2)提供比較的量數，使個人在不同測驗上的得分數可以相互比較。以下將分別說明百分等級常模與標準分數常模之意義，並比較二者之建立方法。

(一)百分等級常模與標準分數常模：

1. 百分等級常模：百分等級是表示受試者的測驗分數在其所屬團體中相對位置的統計量數。亦即，把所有參加測驗的同學依照其得分之高低排列，並將人數予以一百等分，則此等分點就是百分點，各百分點以下所包含人數之百分比即為百分等級。例如將某校高二國文科測驗結果轉化為百分等級後，得知甲生的百分等級為95，則表示該生的國文科測驗分數勝過該校百分之九十五的高二同學。
2. 標準分數常模：標準分數係以平均數為參照點，用離均差的數值表示個別測驗分數在所屬團體中的相對位置。亦即，標準分數是依據標準差單位來表示個人的分數落在平均數之上或之下的距離。例如某生智力測驗分數為45分，而該測驗之平均分數為40分，該測驗之標準差為5，則該生之標準分數（Z分數）為： $(45-40)/5 = +1.00$ ，代表該生分數是在平均分數之上一個標準差的位置。

(二)百分等級常模與標準分數常模之建立方法：

1. 百分等級常模：

- (1)未歸類資料的求法：樣本人數不多時，可先將測驗分數依高低順序排列，並給予名次，然後利用下列公式，求各分數之百分等級：

$$PR = 100 - \frac{100}{N} \left(R - \frac{1}{2} \right)$$

其中，N：全體人數；R：名次

- (2)歸類資料的求法：當樣本人數較多時，可先將全部測驗分數分為幾組，並將原始分數歸入適當的組別，而得到測驗分數的次數分配表，然後依據下列公式求每各原始分數的百分等級：

$$P_{PR} = L + \frac{\left(\frac{PR}{100} \cdot N - F \right)}{f} \cdot h$$

其中：PPR：任何原始分數；L：PPR所在組的真正下限；N：全體人數

f：PPR所在組的次數；F：L以下的累積次數；h：組距

2. 標準分數常模：

- (1)直線標準分數：是將原始分數轉換成直線標準分數，分數的分配型態仍然維持和原始分數一樣。所有原始分數的特性，並不因轉換後而有所改變。常用之直線標準分數有：Z分數、T分數、離差智商分數、AGCT分數、CEEB分數等。
- (2)常態標準分數：當兩個分數的分配很明顯的不同時，為了使不同分配型態的分數可以直接比較，就必須將原始分數都轉換為符合常態分配的標準分數，此種分數稱為常態化標準分數。常用之常態標準分數有：T量表分數、標準九、C量表分數、Sten分數等

五、何謂「投射測驗」？常用的投射技術有哪些？

【參考解答】

投射測驗是測量人格的主要方法之一。投射測驗種類繁多，其中又以「羅夏克墨漬測驗」之應用最為普遍，以下將分別進一步說明：

(一) 投射測驗：

投射技術 (Projection Technique) 是指提供受試者一種曖昧的刺激情境，讓受試者在不受限的情形下，自由反應，從而分析探究受試者隱而未顯的人格特質；其基本假設是認為受試者面對曖昧刺激情境時，因為不受明確題意的限制，反而容易自由表達，在不知不覺中將內心隱藏的情感、態度、需要、價值、情緒、動機與人格特質等流露在反應之中。如果所提供的刺激越無結構和模糊不清，受試者越能將其真正的情感投射出來。而根據此種假設與技術所編製的人格測驗，即為投射測驗。

(二) 常用的投射技術：

1. 聯想技術：提供給受試者一些圖畫、墨漬圖、單字或詞語等刺激物，然後請其將所看到的、想到的、或感覺到的說出來。羅夏克墨漬測驗和文字聯想測驗屬於此類技術。
2. 完成技術：提供一些不完全的刺激，由受試者去填補或完成。語句完成測驗即屬此種技術。
3. 編造技術：讓受試者編造或創造一些東西，通常是故事或圖畫。刺激可以是簡單的，例如要求兒童說一個昨天所發生的故事，刺激也可以是複雜的，如根據一些圖片編造一個故事。應用較普遍的編造技術有主題統覺測驗和逆境圖畫測驗。
4. 表現技術：和編造技術極為相似，唯一不同之處是表現技術偏重於形成結果的方式，亦即分析故事或其他結果的內容和型態。主要的表現技術有畫、畫人測驗、玩洋娃娃、玩玩具、塑造黏土和角色扮演等。

六、解釋名詞

【參考解答】

(一) IRT

IRT (item response theory) 即「試題反應理論」，是一種試題分析方法。在試題的選擇、題庫的建立、測驗編製等工作上，皆已廣泛採用此法。它是以「項目特徵曲線」(item characteristic curve, ICC) 來進行試題分析工作。所謂項目特徵曲線代表的是「答對某題的機率」與「所欲量測的心理屬性」的函數關係。每一條曲線均可由三個參數來描述：鑑別力參數、難度參數、猜測參數。

(二) 月暈效應 (halo effect)

是指評分者對被評定者的一般印象會影響其對被評者個別特性的評定結果。如果評定者對某幾位被評定者有較佳印象，則很容易在每項特質上均給較高的等級；反之，如果對某些被評者的印象不佳，則很容易將每項特質均評給較低的等級。由於月暈效應會導致被評者在各項特質上均獲得相類似的評價，因此無法發現被評者的長處和短處，這樣就喪失了評定法應有的價值。

(三) 社會讚許反應傾向 (social desirability response set)

是指以社會所接受的標準回答題目，而非以個人的真正感受或情況作答，例：「在飯前我會洗手」在回答這題目時，受試者可能不是依據自己的情況或感受，而是依據社會所接受或期望的答案。受測者可能因為許多因素而以滿足社會期許的方式作答，例如期待被稱讚、避免被負面評價等等，因此對於人格測驗的結果有極大的影響。

(四) 文化公平智力測驗 (culture-fair intelligence test)

許多文化的差異皆可能影響智力測驗的結果，因此，智力測驗常因文化的偏差而受到嚴厲的批評。是故文化公平智力測驗就是想免除或減少這些差異對測驗結果的影響。最負盛名的文化公平智力測驗之一，是由卡泰爾所編製的文化公平智力測驗，此測驗是非語言測驗，用來測量卡氏智力型態論中的流體智力，測驗中所使用的圖形均為各文化團體所共同熟悉或不熟悉的圖形材料。

(五) 「常模參照測驗」(norm-referenced tests) 和「標準參照測驗」(criterion-referenced tests)

測驗的種類繁多，因著使用的目的、對象的不同，而有各種區分方式。若以「測驗分數的解釋方法」作區分，則測驗主要可分為常模參照測驗與標準參照測驗兩類。以下將分別說明之。

1. 常模參照測驗：是指測驗的結果，根據分數在團體中的相對位置而加以解釋的一種測驗。此測驗的目的，是在區分受試者之間的成就水準。
2. 標準參照測驗：指測驗的結果，根據學習前所訂定的標準而加以解釋的一種測驗。此測驗是要瞭解受試者能做什麼、學會了什麼，而不是和別人比較起來他是怎樣。