

教育測驗及統計講義

第一回

405580-1



社團人
法考

考友社

出版
發行

第一講 緒論

◎ 命題重點 ◎

一、測驗之意義：

是對行為樣本所做的一種客觀的和標準化的測量

二、測驗之功能：

- (一)評估。
- (二)診斷。
- (三)預測。

三、測驗之種類：

(一)依測量行為特質分：

- 1. 能力測驗。
- 2. 人格測驗。

(二)依測驗材料分：

- 1. 語文測驗。
- 2. 非語文測驗。

(三)依測驗方式分：

- 1. 個別測驗。
- 2. 團體測驗。

(四)依測驗結果解釋的參照標準分：

- 1. 常模參照測驗。
- 2. 標準參照測驗。

四、測量之層次：

- (一)名義量尺：僅作分類沒有大小的意義。
- (二)次序量尺：具大小高低的順序。
- (三)等距量尺：有相等的單位。
- (四)比率量尺：有真正的零點。

五、心理與教育測驗之特徵：

- (一)心理與教育的測量是間接的。
- (二)個體具有可測量的特質。
- (三)測量都有誤差存在。
- (四)測量時常是相對的。

六、我國心理與教育測驗的發展之影響因素：

- (一)文官考試制度的影響。
- (二)西方測驗技能的引進。
- (三)文教與軍事機關對於測驗之提倡。

(四)學校輔導工作的推展。

七、測驗的發展趨勢：

(一)注重穩定的測驗單位。

(二)採最有意義的常模。

(三)採簡短測驗。

(四)增加投射測驗的實用與研究。

● 精選試題 ●

一、測驗之意義為何？

答：一般而言，測驗是指在控制的情境下，使用一群標準化的刺激，引起受試者的反應，藉以對其能力或人格特質作定量的或定性的評量的一種科學工具或方法。

二、試述測驗在學校教育與輔導上之功能。

答：測驗在學校推行輔導工作和增進教育設施的效率方面，都扮演了重要的角色，它的主要功能有三：

(一)評估：

測驗的目的在衡鑑學生的心理特性，幫助教師在較短的時間內客觀地認識學生，也幫助學生自我了解。諸如辨別學生的智愚、性向，測量人格特質，藉以了解學生的個別差異，包括學生個別之間的差異和學生本身的所長與所短，以便因材施教，輔導學生充分自我發展。

(二)診斷：

教師應用測驗，猶如醫師應用體溫計和血壓器診察病人的體溫和血壓一樣，目的在了解情況，幫助診斷(diagnosis)，以便對症處方，藥到病除。兒童在學習和成長時期，在生活和學習適應上都難免遭遇困難，測驗用作診斷的工具，旨在發掘學生的優點和缺點，尤其著重困難原因的發現。

(三)預測：

測驗是否有用，就看它能否準確的預測受試者將來的行為。測驗具有預測的功能，我們才可以了解學生未來發展的可能性，藉此對學生未來的生長作適當的安排，引導他們對升學、選課或就業作明智的抉擇，幫助他們向最有利於自己和社會的方向去發展，以求適才適所，才盡其用。

三、試說明測驗何以是教育上重要的工具？

答：測驗應用到教育方面，使教育行政與教育方法研究走上科學化途徑，使教育行政人員知道學生的能力與學業成就，方可對教育目標、政策、課程、教法等問題作客觀的估量，從而了解學生的性向、志趣，然後因材施教，藉以提高教育效率，改進教育方法，基此，測驗在教育上相當重要的工具之一。

四、何謂測量？何謂評量？試比較二者之異同。

答：(一)所謂測量，就是指根據量尺，用數字描述個人特質的歷程。例如，依據比西量表(Binet-Simon Sale)測量的結果，可以得到智力商數(IQ)，用這個商數來描述個人的智力特質，這個歷程就稱為測量。

(二)所謂評量，就是指根據一項標準，對所測量到的數量做價值的判斷。例如，一位學生在算術成就測驗上獲得80分，這個分數代表他的成績優異呢？中等呢？或低劣呢？單獨從分數去看，無法獲知，但依據其能力而評斷其成績為優秀

的，這就是屬於評量。

(三)由上所述，可知評量所涵蓋的範圍較測量為大，測量僅是評量的一部分而已。評量可以依據測量所提供的資料，也可不必依據它，而做價值判斷。因此，凡是可利用的資料，不論是數量的或品質的，均可做為評量的依據。在測量方面，儘可能力求客觀，減低判斷的影響，但在評量方面，須根據各種有關的資料做價值判斷，故評量通常比測量更為主觀。

五測驗為何必須在控制的情況下實施？

答：通常測驗必須在控制的情境下施行，所測量的結果才能供作比較評定的依據，因為主試者施行測驗是否按照規定的或標準化的手續，以及受試者的外在刺激或其內發的需要、驅力、焦慮……等等因素的情況下受試，往往有不公平的比較，因此這些因素必須儘量加以控制，必須依照規定的手續去施行，使影響測驗的結果的各種因素，減低至最小限度，這樣，測驗的結果才能作公平的比較和解釋。

六試述標準化之意義。

答：所謂標準化(standardization)，含有二方面的意義：其一是編製測驗的題目要經過標準化的程序；其二是測驗的使用要按照標準化的手續。編造一個測驗，舉凡測驗的設計、擬題、預試、選題、編排和建立常模等，都須經過嚴密的考驗程序。另一方面，測驗實施的步驟，包括測驗前的準備、測驗做法的說明、時間的控制、初閱、複核和換算指數等的記分、對測驗結果的解釋和應用、試場環境的布置以及主試者、襄試者和受試者之間和諧關係的建立等等，都有詳細的規定，使用測驗的人不可變更；而且使用測驗的人必須受過專門的訓練，這樣不論誰來實施測驗，都不會影響受試者的作業，受試者的分數也不因主試者的情感或偏見而轉移。測驗的編造和實施具有這樣精密的手續，才能成為一個客觀的科學工具。這樣的手續和過程，就叫「標準化」。

七試比較標準化測驗與非標準化測驗之不同。

答：心理測驗的編製有其根本的心理計量學基礎，若從純粹測量(measurement)的觀點來看，心理測驗可分區分為標準化(standardized)和非標準化(nonstandardized)測驗兩大類。標準化測驗無論在實施和評分上皆有統一的規定，各受試者受到完全相同的待遇，測驗並備有常模，作為測驗結果計分和解釋的依據，施測者在測驗的使用上應嚴格遵守這些規定。非標準化測驗在編製過程中並不一定牽涉到心理計量學或常態分配的測量原理，它們在實施時雖也有一定的程序，但較具彈性，同時，這類測驗上並無常模之建立供作測驗評分和結果解釋之用。

八人類的心理或心理特質（如智力、品格或態度）有無測量的可能？結果能否量化？

答：人類的智力、性向、知識、技能或人格等，都是一些抽象的特質，不能直接摸

● 精選試題 ●

一、一個好的測驗應具備那些重要屬性？

答：行為科學家所使用的測量工具大都是測驗 (test)，而一個好的測驗具有三項必要的屬性：「標準化」(standardization)「效度」(validity)、及「信度」(reliability)。

(一)標準化：

標準化意指已用某種測驗施測過一群特定的個人，並且已將他們的成績或分數予以仔細記錄與保存。

(二)效度：

一個測驗如能測量出其所欲測量的屬性，此測驗即屬有效 (valid)。重要的是測驗的用途，一個測驗在某一種用途上可能有效，但卻並不保證它在其他用途上也有效。

(三)信度：

信度指的是測量的精確程度，它是效度的一個必要成份。如果一個測驗的信度很低，那麼，它的效度便不可能很高。然而一個毫無效度的測驗卻可能有相當的信度。專業性的測驗編製者喜以數學的方式來表示信度與效度。概言之，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，而信度則是一個測驗在測量其所測屬性時所得結果的穩定性與可靠性。

二、何謂信度？

答：所謂信度，就是一種測量工具或一套測驗可靠的程度，通常是指測驗結果的一致性 (consistency) 或穩定性 (stability) 而言。一個測驗的信度表現於兩方面，茲分述於后：

(一)測驗內部試題之間是否相互符合，前後連貫？

(二)重複測驗的結果是否保持不變？無論試題是否符合或幾次測驗結果是否一致，均屬相對程度上的差異，而非全有或全無的分別。任何一種測量，總有或多或少的誤差，而誤差受機遇因素所支配。誤差愈小，信度愈高；誤差愈大，信度愈低。因此，信度亦可視為測驗結果受機遇影響的程度。

總合說來，信度的意義為考驗測驗的可靠性，信度太低，則測驗結果無法令人信服。

三、試述測驗信度計算的方法。

答：由於測驗上誤差的來源有多種，所以用來計算測驗信度的方法也可分為不同的幾類。第一類是專為檢驗同一測驗中不同複本 (equivalent forms) 上分數的一致性，稱為複本信度。第二類是專為檢驗同一測驗中內部試題或刺激之內部一致性 (internal consistency)，包括折半信度、庫李信度、和 α 係數等，第三類是為檢驗同一測驗在兩個（或以上）不同時間重覆實施時其分數的穩定性，稱為再測信

度。第四類是為檢驗不同評分者之間評分結果之一致性，稱為評分者信度。

(一)複本信度：

當同一測驗上具有兩種或以上的複本時，吾人可以將此兩種複本分別實施於同一群受試樣本，然後以 Pearsonr 法計算兩複本上分數之相關係數，此一數值即為複本信度 (alternate form reliability) 係數，或稱相等係數 (coefficient of equivalence)。

(二)折半信度：

折半信度，顧名思義，是將一測驗分「折」成相等的兩半，每一半各自形成一項個別的單元或測驗，並求其信度。其程序為先將測驗以一次實施予同一群受試者，然後將測驗以特定的方法劃分為相等的兩半，受試者在每一半上可得到一個分數然後再以 Pearsonr 來計算兩半分數間之相關係數。

(三)庫李信度和 α 係數：

庫李信度 (Kuder-Richardson reliability) 和 α 係數 (Coefficient α) 也是為檢驗測驗內部一致性而設。它們與折半信度不同的是在計算測驗信度時無須將之「折」為兩半，而直接測定整個測驗中所有試題或刺激之一致性或相關程度。庫李信度為 F. Kuder 和 M.W. Richardson (1937) 二氏所創用，故以二人姓氏命名。庫李信度的求法是建立在變異量數分析法之上，它共有兩個不同的公式，一為 KR20，另一為 KR21。較常用的為 KR20，現在先看此一公式：

$$KR20 = \frac{k}{k-1} \left(1 - \frac{\sum pq}{SD^2} \right)$$

庫李信度的缺點是它們僅可應用於以二分法為計分標準的測驗上。但心理學上許多的測驗常採用多重評分標準，為因應這些多重評分制測驗的需要，Cronbach (1951) 將 KR20 加以修改而設計了下列公式，用來計算這類測驗之內部一致信度。

$$\text{Cronbach's } \alpha = \frac{k}{k-1} \left[1 - \frac{\sum (SD_i)^2}{SD^2} \right]$$

(四)再測信度：

所謂再測信度 (test-retest reliability) 乃是指將同一個測驗以同樣方法但卻在不同時間重覆實施於同一樣本，並求各次測驗間分數的穩定性。是以再測信度又稱為穩定係數。

(五)評分者信度：

評分者信度 (scorer reliability) 乃是指不同的評分者在測驗過程中觀察、記錄、評分、計分各方面相互間的一致性。兩個或以上的評分者在這幾方面所得出之結果愈相吻合，則評分者信度便愈高。