108 年特種考試地方政府公務人員考試試題

等別：四等考試

類科：教育行政

科目：教育測驗與統計概要

一、請分別說明：

(一)教學評量四大類：安置性評量、形成性評量、診斷性評量、總結性評量，其實施的時機與主要目的為何？(20分)

(二)形成性評量與總結性評量在試題設計上的比較。(5分)

【擬答】

【解題關鍵】

1.《考題難易》：(最難5顆★)★★

2.《解題關鍵》：傳統評量過去考古題很常見，對考生來說，並不難處理

3.《命中特區》：思澤 教育測驗與統計 P399

(一)四大評量時機與目的

1.安置性評量：

安置性評量主要是為了評估學生在學習前的狀態而使用，可分為預備性評量以及安置性評量。前者目的是為了評估學生是否具備學習新單元的必要技能，如學生缺乏進入新單元的能力則需要補救教學；後者則是為了評估學生是否熟練預定教學的內容，如果精熟了，則可以調整教學計劃或是將學生安置在較高層次的學習班級當中。但是安置性評量不一定在每次的教學前都需要實施。如果教師對所教班級的程度已有一段時間的認識，或是教學新單元所需的基本能力不明確時，都可以不必進行此種評量。

2.形成性評量：

形成性評量的功能在改進教學及提高學習效果。其目的在於提供回饋給學生和教師並及時發現學習困難處並改進教學。其評量的目標著重在運用結果以改進教學以及檢查學生學習進步情況。所以評量的結果通常不會給予等級或是評分，只告訴學生其學習狀況是精熟或非精熟。

3.診斷性評量：

診斷性評量的使用時機可在教學前或是教學進行中對學生進行學習困難與否的診斷。其功能在於可發現學生在某一項特定學習內容或是技能上的困難，並以此採取補救的措施。

4.總結性評量：

總結性評量目的在確定教學目標達到的程度、學生精熟預期學習結束的程度以及評定學生成績等級。通常在學校會以段考或是週期性小考來實施。

(二)形成性與總結性評量的比較

評量	總結性評量	形成性評量
目的	常考慮區分學生個別差異	常考慮確定學生精熟程度
內容	包含學習的廣大領域，但是每一獨特課題只以少數題目測量	涵概學習課題的有限領域，但每一獨特課題以較多題目測量
題目難度	中等難度，排除極難、極易的題目	難度和學習內容配合
結果解釋	依分數在團體中相對地位解釋意義	依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、目標參照、精熟分數

二、相關分析是一種測量變數間關係強弱的方法，請回答下列相關問題：（每小題 5 分，共 25 分）

- (一)散佈圖用於觀察兩個變項之間的關係，水平軸（X 軸）與垂直軸（Y 軸）各代表一個變項，請問各為何種變項？
- (二)某散佈圖呈現「左上右下」的線性關係，請問是為何種相關？
- (三)請說明相關係數（coefficient of correlation）的意涵與功能。
- (四)某學校進行數學能力檢測，發現「身高與分數之間呈現正相關」，這樣的結果顯然與常理不合。試解釋之。
- (五)某相同研究條件下，兩組相關係數之比較研究：在電腦打字能力研究中，女同學組（ $N_1=60$ ）之 $r_1=0.70$ ，男同學組（ $N_2=40$ ）之 $r_2=0.50$ 。因為 $r_1 > r_2$ ，因此定論女同學之電腦使用能力其相關性高於男同學之電腦使用能力。試評論此結論。

【擬答】

【解題關鍵】

- 1.《考題難易》：(最難 5 顆★)★★
- 2.《解題關鍵》：相關分析的考題屬於測驗的基礎觀念，所以拿分並不困難
- 3.《命中特區》：思澤 教育測驗與統計 P151

(一)積差相關係數為最常用的相關係數，是英國統計學家皮爾遜(K. Pearson)所發展出來的統計方法，乃「XY 雙變項所有 Z 分數交乘積和的平均」，適用於 XY 雙變項皆是等距或比率變項資料。

假若考慮變項的量尺特性，例如：點二系列相關，在雙變項中，有一個變項是(真正)二分名義變項，而另一個變項是等距或比率變項(連續變項)時適用。例如：歷史成績和性別(男、女)的相關。

(二)散佈圖為左上右下時，為負相關，像是隨著年紀增加，洗髮精用量越來越少。

(三)指雙變項資料(bivariate data)之間的相關程度。通常以相關係數(correlation coefficient)r 表示，r 的範圍在 $r=-1.00$ 至 $r=+1.00$ 之間。係數的大小代表「關聯的強度」，係數的正負號代表「關聯的方向」。

公式：

$$r_{xy} = \frac{\sum Z_x Z_y}{N}$$
$$r_{xy} = \text{積差相關係數}$$
$$Z_x = \frac{X - \bar{X}}{S_x}$$
$$Z_y = \frac{Y - \bar{Y}}{S_y}$$

$N = \text{總人數}$

(四)r 值意義和樣本大小有密切相關，當樣本 N 很小時，雖然得到高的 r 值，也許是機遇造成。反之，雖然 r 值較小，但樣本 N 非常大，則此 r 值便不可忽視。因此要加以檢定，以確定相關係數是否為 0。

例如：21 名學生國語和數學的相關，是 $r=.45$ ，我們想知道這 $r=.45$ 是不是機率造成？查相關係數檢定的附表，發現 $df=N-2=19$ ， $\alpha=.5$ 的臨界值為 .433，因為計算 r 值 .45 大於臨界 r 值 .433，故可宣稱 21 名學生國語、數學的相關的確是正相關，而非機遇造成。

(五)依據題意，會建議應該從相關係數差異的考驗進行說明，因為相關係數是由樣本的資料統計結果，所以單純透過樣本數值差異來說明，會忽略機率造成的影響，所以可以透過 Fisher 的 Z 轉換法，可以透過 t 檢定對二個相關係數差異進行考驗，這樣才能說明差異真實存在。

三、一般在進行推論統計檢驗時，總會存在一定數量的不可控制的誤差，這些誤差會影響著檢定結果之準確性與效果，試回答下列相關問題：

(一)請分別說明型 I 錯誤 (Type one error)、型 II 錯誤 (Type two error)，與統計考驗力 (Power)。(15 分)

(二)請指出下列五種因素中，何者「增加」統計考驗力？何者「減少」統計考驗力？(10 分)

1. 在預估與已知母群的平均數間，更大的預估差異值。
2. 更大的母群標準差。
3. 更大的樣本數。
4. 使用較嚴格的顯著水準 (例如使用 .01 而不使用 .05)。
5. 使用雙側考驗而不使用單側考驗。

【擬答】

【解題關鍵】

- 1.《考題難易》：(最難 5 顆★)★★
- 2.《解題關鍵》：統計決策理論的觀念並不難處理，一般準備好的同學可以全拿
- 3.《命中特區》：思澤 教育測驗與統計 P81

(一)

1. H_0 為真，卻錯誤拒絕 H_0 的機率：稱為「第一類型錯誤」(type I error)，其機率為 α 。
2. H_0 為假，卻錯誤接受 H_0 ：稱為「第二類型錯誤」(type II error)，其機率為 β 。
3. H_0 為假，正確拒絕 H_0 ：決策正確，其機率為 $1-\beta$ ，稱為「檢定力」或「統計考驗力」(power of test)

(二)

增加 Power 的常見因素，有：

1. 增加 α 。
2. 母群的變異數變小。
3. 增加樣本數 n 。
4. H_0 和 H_1 的平均值差異變大。

所以依據題意所羅列的五個因素，可以發現增加 Power 的有第一個、第三個、第五個因素，減少 Power 的有第二個因素、第四個因素。

四、請回答下列教育測驗評量中「信度」相關的問題：

(一)請分析「試題題數」、「團體分數的變異程度」、「不良試題的刪除」、「試題難易程度」，以及「測驗評分的客觀性」這些因素，對信度產生的影響。(15 分)

(二)某測驗觀察分數及誤差分數的變異數分別是 150 及 30，請算出該測驗的信度。(5 分)

(三)某教師自編英文測驗 25 題，信度係數 0.50，若該測驗再增加 50 題，請算出新的信度係數。(5 分)

【擬答】

【解題關鍵】

- 1.《考題難易》：(最難 5 顆★)★★
- 2.《解題關鍵》：影響信度的因素常考的內，難度不高，但是折半信度的斯布校正公式可能比較容易出錯
- 3.《命中特區》：思澤 教育測驗與統計 P245

(一)

1. 團體分數的變異程度

團體分數的變異程度與信度的關係，可由信度的定義公式去推論

$$r_{xx} = 1 - \frac{S_e^2}{S_x^2} \quad S_e^2 \text{ 為誤差分數變異量；} S_x^2 \text{ 為實得分數變異量。}$$

公職王歷屆試題 (108 地特三等)

由公式中可知，若誤差分數變異量保持不變，實得分數的變異量愈大，則測驗信度愈高。實得分數變異量愈大，表示團體成員間的個別差異也愈大，亦即成員間的異質愈大，則測驗結果的信度也就愈高。

2. 試題難易程度

有些測驗使用在能力較差之團體時，信度可能相當低，因為他們的分數很容易受到猜測因素所影響。若試題較難，則盲目猜測現象必然增加，對能力較低的受測者缺乏鑑別作用，信度係數會降低，若試題太簡單，受測者都能答對，則相當於測驗題數被縮減，所得分數全距也隨之縮減，信度係數也會下降。

3. 測驗題數

在討論折半信度係數時，我們曾說過測驗信度會隨試題題數增加而提高。若我們假定測驗的品質、試題所測量的特質、及受測者的性質都保持一樣，則試題題數與信度的關係可用下列公式來表示：

$$R_{XX} = \frac{nr_{xx}}{1+(n-1)r_{xx}}$$

R_{XX} 表題數為原測驗 n 倍的測驗的係數； r_{xx} 表原測驗的信度係數；

若有一 20 題的測驗，其信度係數為 0.60，若把試題增加成 100 題（為原來的 5 倍），則其信度係數便提高為：

$$R_{XX} = \frac{5(0.60)}{1+(5-1)(0.60)} = 0.88$$

倒過來，若有一個 100 題的測驗，信度係數為 0.90，若想隨機抽取 60 題構成一較短的測驗，則其信度將降為：

$$R_{XX} = \frac{0.6(0.90)}{1+(0.6-1)(0.90)} = 0.84$$

雖然增加題數可以提高信度係數，但隨著題數的增加，所需施測時間也增加，受測者疲勞和厭煩的程度也會增加，所以過多的試題可能會得不償失。此外，我們也許無法擬出那麼多同樣高品質的試題。

4. 不良試題刪減

題目的整體品質會增加，則信度估計結果會更加可靠，信度會提升。但是在刪減不良試題時，也應該注意整體題數過度減少，是否會造成信度估計結果反而更不好。

5. 評分客觀性

評分上若能更客觀，則信度估計會更加可靠，信度會比主觀評分方式的信度來得更好，所以在選擇題的評分與申論題的評分，可以想像選擇題評分會更加客觀，所以信度會更好。

(二)

信度的定義公式為

$$r_{xx} = 1 - \frac{S_e^2}{S_x^2} \quad S_e^2 \text{ 為誤差分數變異量； } S_x^2 \text{ 為實得分數變異量。}$$

依題意，信度為 0.8

(三)

原本有 25 題的測驗，其信度係數為 0.50，若把試題再增加 50 題（為原來的 3 倍），則其信度係數便提高為：

$$R_{xx} = \frac{3(0.50)}{1+(3-1)(0.50)} = 0.75$$