

101 年公務人員特種考試原住民族考試試題

等 別：四等考試

類 科：教育行政

科 目：教育測驗與統計概要

一、有一題選擇題經測驗後，各選項的答題反應分布情形如下表所示（此題正確答案為 B，以 * 表示）。

(一)計算這試題的難度值（要列出計算式，且計算至小數點第二位），並判讀其好壞。

(二)計算這試題的鑑別度值（要列出計算式，且計算至小數點第二位），並判讀其好壞。

(三)分析各錯誤選項的誘答力，以建議修題方向。

選項	A	B*	C	D
高分組	0	15	3	2
低分組	3	6	11	0

【擬答】：

(一)難度： $P = \frac{P_H + P_L}{2} = \frac{\frac{15}{20} + \frac{6}{20}}{2} = 0.525$ ， P 值以 0.5 最恰當，此題 $P = .525$ 接近難易適中。

(二)鑑別度： $D = P_H - P_L = 0.75 - 0.3 = 0.45$ 。

鑑別度指標值愈高，表示試題鑑別度愈大；鑑別度指標值愈低，表示試題鑑別度愈小。美國測驗專家 Ebel 在 1979 年提出一套評鑑標準，當 D 大於 0.4 以上顯示題目非常優良，題目具備區辨力。

(三)1. 誘答項分析：客觀測驗進行選項誘答力分析的目的在進一步提供試題分析的指標，作為判斷測驗試題編製優劣與診斷學生反應組型的依據。分析選擇題是否具有誘答功能，只要分析高低分組學生在每個試題選答的次數分配再加以判斷即可。國內郭生玉（民 95）提出判斷原則：①每一錯誤選項至少有一低分組學生選擇；②選擇錯誤選項的低分組學生人數應比高分組學生還多。

2. (1)A 選項誘答力佳，每一錯誤選項至少有一低分組學生選擇，能有效鑑別高分組與低分組。

(2)C 選項誘答力佳，選擇錯誤選項的低分組學生人數應比高分組學生還多。

(3)D 選項誘答力不佳，設計不良，因為高分組答錯人數高於低分組，無法有效鑑別高分組與低分組。

二、有一心理測驗的平均數為 100，標準差為 15，小明在此測驗的得分為 105。但由於測驗分數都可能誤差，解釋小明的測驗結果宜採何種方式？

【擬答】：

信度係數適用於判斷一個測驗對某一團體測量結果的可靠性信度，而測量標準誤 (SE_{meas}) 適用解釋個人測驗分數的可靠性。

(一)測量標準誤意義：

1. 用以估計受試者測驗之真實分數所在範圍的誤差單位與分數穩定性的信度。在標準情境下，使用相同測驗或複本測驗測量一個受試者許多次，其所得分數的平均數即為個人的真實分數。一般來說，會以受試者的真實分數為中心形成常態分配。而每次測驗實得分數與真實分數的差，稱為測量誤差，這許多次測量誤差分配之標準差即為測驗標準誤，又稱為分數的標準誤差。

2. 公式： $SE_{meas} = S_x \sqrt{1 - r_{XX}}$

(二)個人真實分數可信賴範圍

1. 解釋個人的測驗分數：避免在測驗後只給學生數字化測驗分數，而未解釋該數據具有的意義，解釋測驗者最好面對面口述分數代表意義，或附註文字說明；解釋個人的測驗分

公職王歷屆試題 (101 原住民特考)

數宜考量測量標準誤大小，根據測量實得分數推測真實分數的可信範圍與可靠性。

$$2. X - Z \cdot SE_{meas} \leq t \leq X + Z \times SE_{meas}$$

三、請分別說明信度、效度，以及信度與效度兩者間的關係。

【擬答】：

(一)信度意義：

1. 信度是指相同受試者在不同時間，使用相同測驗測量（或複本測驗測量多次）或在不同情境下測量，所得測驗結果穩定性；不同測驗分數一致性、不同評分者間評分一致性、評量結果精確性，藉以反應真實量數程度的一種指標。
2. 測量誤差：信度是指測量或測驗分數，反映出真實分數的程度或無誤差程度。即信度表估計測量的誤差有多少。

(二)效度意義：

1. 效度 (Validity)：測驗分數的正確性，指測驗分數能夠代表測量到想要測量之潛在特質的程度，亦即效度為有多好的程度、使用目的的有效性、測驗能達到某種目的的程度。效度乃根據測驗所做推論的恰當性、意義性及合用性的程度。
2. Messick (1995) 認為效度的新定義應該包含四項：①傳統的構念效度證據；②測驗解釋的價值分析；③測驗在特定用途應用解釋的效用分析；④潛在的及實際的社會後續效應評估。
3. 從教育決定的角度，效度是指一個測驗能夠提供適切的資料以做決定的程度。

(三)信度與效度關係：

1. 信度是效度的必要條件而非充分條件：

- (1) 信度與效度均是優良測驗所必須具備的主要條件，信度包含效度的數學關係。
- (2) 效度需要測驗分數的正確性、一致性，但信度只需一致性即可。信度低，效度必低；信度高，效度未必高；效度低，信度未必低；效度高，信度必定高。

2. 信度與效度的變異關聯性：

- (1) 依據因素分析原理：一個測驗的總變異是由共同變異量、特殊變異量和誤差變異量三部分所構成。

$$\frac{S_X^2}{S_X^2} = \frac{S_{co}^2}{S_X^2} + \frac{S_{sp}^2}{S_X^2} + \frac{S_e^2}{S_X^2}$$

$$Val = \frac{S_{co}^2}{S_X^2} = 1 - \frac{S_{sp}^2}{S_X^2} - \frac{S_e^2}{S_X^2} \quad (\text{信度 } r_{XX} = 1 - \frac{S_e^2}{S_X^2})$$

$$Val = r_{XX} - \frac{S_{sp}^2}{S_X^2} \quad (\text{效度} = \text{信度} - \text{獨特性})$$

- (2) 上述推論，效度係數不會大於信度係數的平方根（信度指標）：

$$Val \leq \sqrt{r_{XX}}$$

四、在推薦甄試時，5 位口試委員針對考生的專業背景與研究潛能兩項加以評審，評分結果顯示這 5 位委員給考生的專業背景評分都很高，但對研究潛能評分都很低。請分析這個評分結果中專業背景與研究潛能兩項分數之間的可能關係（相關）？必要時可舉例（數字）說明。

【擬答】：

如果評分結果會受到評分者主觀判斷和意見的影響時，則須採用評分者信度，以估計數位評分者是否評分一致。

(一) 評分者信度意義：係指從測驗卷中隨機抽取一些樣本，由不同的評分者或同一個評分者在不同的時間，對每一份試卷評分，在測驗過程中觀察、記錄、評分、計分等各方面相互間的一致性而言，然後根據每份測驗卷評分的兩個分數計算其相關係數，即得評分者信度。評分者信度愈高，表示評分者間評分愈一致。

(二) 多項特質—多項方法分析法 (Multitrait-multimethod Approach)

公職王歷屆試題 (101 原住民特考)

學者 Campbell 與 Fiske 在 1959 年提出一份良好的構念效度，須符合下列二條件：

1. 聚斂效度（輻合性效度）（convergent Validity）：一份測驗分數必須和相同構念其他測量的分數間存在高相關。如專業背景分數間、研究潛能間特質相同有高相關，即為聚斂效度。
2. 區別效度（辨別效度）（Discrimination Validity）：一份測驗必須和不同構念或潛在特質其他測量的分數間存在低相關。如專業背景分數與研究潛能間特質不同為低相關甚至零相關，即為區別效度。
3. 採用多項特質—多項方法分析基本條件需有至少二種以上測驗方法（很大差異），去測量二種以上的特質，其所測量特質均相同。

特質	方法 1		方法 2	
	專業背景	研究潛能	專業背景	研究潛能
方法 1	專業背景 研究潛能	(0.75) 0.31	(0.84)	
方法 2	專業背景 研究潛能	0.60 0.21	0.20 0.65	(0.75) 0.32 (0.80)

五、請問下列 A、B、C 三個分配，何者的平均數最小？

分數	A 分配次數	B 分配次數	C 分配次數
210	3	3	1
211	12	5	2
212	14	10	5
213	4	10	10
214	2	5	14
215	1	3	4
合計次數	36	36	36

【擬答】：

分組平均數計算採用 $\bar{X} = \frac{\sum_{i=1}^n x_i f_i}{n}$ ，題目中 A 分配次數低分的較多，同樣合計次數 36 次，A 分配的平均數最小。

【同學亦可將分配帶入公式實際計算 A 分配 $\bar{X} = \frac{\sum_{i=1}^n x_i f_i}{n} = \frac{210 \times 3 + \dots + 215 \times 1}{36}$ 】