

## 110 年特種考試地方政府公務人員考試試題

等 別：三等考試

類 科：教育行政

科 目：教育測驗與統計

吳迪老師、思澤老師

一、十二年國教新課綱強調核心素養的教育，但是有關核心素養的多元評量方面，實作技能與情意態度的評量一直是課綱中較被忽略的。請問要如何評量學生是否具備實作技能層面或情意態度層面的素養，詳細說明這兩種評量方法的執行方式、優點與缺點。(20 分)

1. 《考題難易》：★★★普通

2. 《破題關鍵》：實作評量以及情意態度評量皆曾出現在考古題中，所以對準備良好的同學來說，不難發揮，應可以拿到不錯的分數

### 【擬答】

(一)評量：

所謂評量 (assessment) 是指「收集、統整和解釋訊息。」(余民寧, 2002) 用來幫助使用者做決定的歷程。比如說教師會使用評量瞭解學生的學習狀況，並用以決定是否安排教學計劃或是補救計劃等等。以教師使用為例，評量的目的在於幫助教師做出下列決定：1. 診斷學生的問題、2. 評量學生表現、3. 提供回饋、4. 安置學生、5. 計畫及執行教學、6. 建立並維持教室的社會平衡 (余民寧, 2002)。因為訊息必需加以解釋才能用於決策，因此通常必須做出價值判斷。由於要對使用評量的對象做出價值判斷，則在資料的收集上則應盡可能的包含到所有的面向。因此在資料的收集上應包含計量 (quantitative) 以及非計量 (qualitative)。

(二)實作評量的執行方法以及優缺點：

1. 實作評量簡介：

評量的意義不但要知道學生知道什麼，也要評量學生會什麼。傳統紙筆測驗只能測出前者，實作評量則能兩者兼顧。所謂的實作評量通常是呈現工作給學生，並要求以口頭、寫作或是完成作品的方式表現。

2. 執行方法：

實作評量目的在測量較複雜的學習目標及能力，因此在設計、實施及評分方面需要花費較多時間。整個實作評量的設計過程主要包含下列六個階段：

- (1) 確定實作評量的目的或要評量的能力
- (2) 確定要採用之實作工作
- (3) 確定在何種情況下實施實作工作
- (4) 確定判斷實作表現的評分標準及系統
- (5) 準備指導語給學生及評分者
- (6) 修正實作的工作或評分系統 (scoring rubric)

在這當中評分的標準的設計是發展有效實作評量的最重要工作。因為評分標準可以使學生知道學習目標和期望，並引導努力的方向和優先順序。對教師而言可以引導公平客觀的評分，並提高信、效度。

以擴大反應的實作評量為例，來說明執行方法，擴大反應的實作評量係指除了工作本身提供的資料之外，需要從各種管道尋找資料。比如做實驗、研究時要使用到網路資源、電腦或是其它設備等。實作過程和結果都是此種評量的重要部份。結果的完成需要好幾

天的時間，並提供修改的機會，使學生能充份展現其應用知識和技能解決問題的高層次思考能力。

3. 優缺點：

(1) 優點：

- ① 提供評量歷程與結果的方法：實作評量可評量歷程，也可評量結果或兩者同時兼顧。觀察實作過程也具有診斷作用。
- ② 可以評量實作和統整能力：實作評量可以評量到實際操作的能力以及比較高層次的解決問題能力，這是其它評量法無法做到的。尤其是採用研究報告的評量方式更具有評量應用及統整知識的能力。
- ③ 符合近代學習理論：建構論強調學習者是知識的主動建構者，因此實作評量本質上要考慮學生的先備知識，並使學生主動建構自己的意義。
- ④ 有助改進教學：傳統紙筆測驗或客觀測驗只能偏重於一些知識的獲得和記憶，其他能力比如獨立思考以及問題解決能力無法得到發展。實作評量因為兼顧學習結果與歷程，因此若配合教學目要將有助於學生學習能力的提升。

(2) 缺點：

- ① 費時費力。
- ② 無法測量到所有學習目標：實作評量應放在測量客觀測驗無法測量的學習結果。如一些客觀測量可測得的能力如概念、事實、規則等記憶或推理就不必使用實作評量。
- ③ 評量信度較低：實作評量容易犯評分者誤差，尤其是缺乏明確的評分系統時，會使學生的實作過程或是作品產生不同的評分結果。這種不一致會導致信度的降低。
- ④ 評量結果推論性較低：在某工作上的實作表現通常難以推論到其它的工作上，比如能發表有說服力演說的學生是否也能發表具有娛樂性演說的能力？為解決這樣的缺點，可採用多元的實作工作，從不同工作表現累積資料。

(三) 情意態度評量的執行方法以及優缺點：

1. 情意態度評量的簡介：

前述的評量方法主要目的在教學領域當中的認知以及技能方面的評量，而教育所要達成的目標除了前述兩者之外，還應包含健全人格的培養。情意評量的目標除了培養健全人格之外，也是為了幫助學生將來能夠適應社會生活。另外，情意與認知之間會互相影響，比如一些天賦不高，但是動機強烈、態度積極的學生透過努力的行為仍有可能在學業上取得成功。因此，教師更應注意學生在情意方面的情況，給予適當的幫助。而情意評量的範圍包含了態度、學習興趣與動機、自我觀念以及適應社會生活的能力還有價值觀 (Popham,2002)。

2. 執行方法：以社會計量法 (sociometric technique) 為例

猜是誰技術 (guess who technique)：此方法是呈現一系列簡短行為的敘述句，再請學生寫出最適合每一項敘述句的友伴的名字。每個句子不限人數。為使學生放心回答，會告知其作答有匿名性。此法的好處是實用性高，並且敘述句可分積極或消極的，因此教師可同時評量數種行為特質。並有助於教師診斷需要輔導的學生。

社會計量技術是一種研究團體交互關係型態的一種方法，可應用在下列幾個方面：(1)組織班級團體；(2)改善個別學生的適應；(3)改善團體的社會結構；(4)評估學校措施對學生社會關係的影響。此方法先請學生為一個活動選擇一個或若干個友伴，並運用不同的分析方法呈現結果。選擇的方向可以是選擇喜歡的友伴，或是選擇不喜歡做友伴的同學；

後者可有助於發現被拒絕的學生，進而協助克服困難。社會計量技術的分析方式，例如：社會計量矩陣：一旦學生完成選擇作答，將學生得到的選擇數量簡單計算後形成的表。從此表可看出學生的選擇型態以及哪些學生受歡迎，而哪些學生不受欢迎。

3. 優缺點：

(1) 優點：

紙筆測驗是傳統情意評量中最常使用的方式。但是情意評量所強調的是學生的行為、習慣、態度、情緒、動機和適應發展等改變。提出價值澄清法的 Rath 曾認為在校表現不好的同學可能是尚未找到價值觀所造成，所以過去傳統評量忽略情意，側重認知評量，卻不一定可以預測同學的學習成就表現，因此情意評量可以提供我們對學生更全面的瞭解。

(2) 缺點：

大多數情意評量容易有不夠客觀的質疑，像是觀察法容易涉入主觀偏誤

- ① 個人的偏差：包含寬大誤差（傾向高分的評定）、嚴格誤差（偏向低分的評定）、趨中誤差（傾向居中評分）
- ② 月暈效應，也就是評定者對受評定者的印象影響。
- ③ 邏輯誤差，也就是說評定者將兩特質視為直接相關而將之評定為比實際的關係更接近。比如評定者認為智力跟學業成就有關，而將學業成績較好的學生的智力評為較高。

參考來源：郭生玉著。教育測驗與評量。精華出版。  
命中：思澤老師，講義第十章實務上之評量

## 考取文化行政 教育行政 就選 志光.保成.學儒

### 上榜學長姐見證

#### 一年考取 雙料金榜

**張○瑄** 109 高考-文化行政榜眼  
109 普考-文化行政探花

感謝補習班，讓我學習不孤單

如何提升弱科和強科增分，讓考試平均分數大致落在60分上下是上榜關鍵。只要能多加練習考古題和熟讀教材，答題上把握旁徵博引和縱、橫向思考原則即可。謝謝工作人員、文化行政學員夥伴的陪伴和共學，讓自己在國考路上不致孤單。

#### 一年考取 全國狀元

**施○豪** 109 高考-教育行政狀元  
109 普考-教育行政

視訊學習，不受限場地和時間影響

選擇視訊補習是因為內容和面授班一樣，但不用受限於場地和時間，還可以調速度和重複播放，防止不小心恍神發呆結果沒聽到重點，如果有問題時，可以記錄下來找時間或線上詢問。報名年度班是因為答應自己一年就要考上，所以沒有留太多後路。



二、下表是甲縣市( $X_1$ )與乙縣市( $X_2$ )每年中途輟學的人數統計表，請依此回答下列問題：

年份	甲縣市( $X_1$ )	乙縣市( $X_2$ )	人數差異( $d$ )= $X_1 - X_2$
2000	63	119	-56
2001	122	161	-39
2002	106	147	-41
2003	66	116	-50
2004	37	71	-34
2005	27	55	-28
2006	16	26	-10
2007	16	12	4
2008	8	6	2
平均數	51.2	79.2	-28.0
標準差	41.1	58.7	21.9

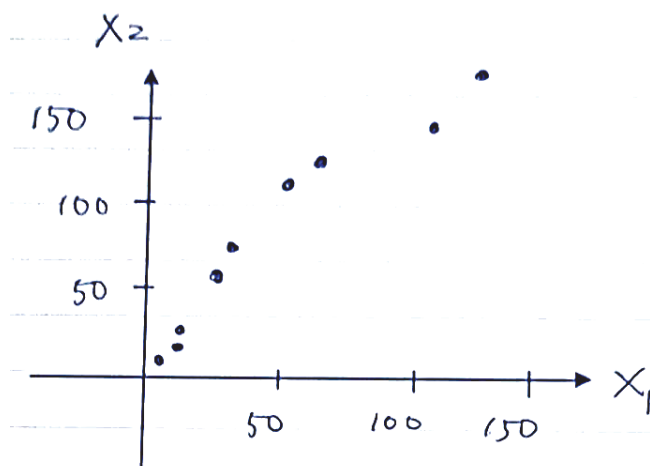
(一)請以甲縣市的中輟學生人數( $X_1$ )為橫軸，乙縣市的中輟學生人數( $X_2$ )為縱軸，繪出兩縣市的中輟學生人數散布圖(scatter plot)，並說明兩者之間有何種關聯？你的證據是什麼？如果要進行假設檢定來檢驗「甲縣市( $X_1$ )與乙縣市( $X_2$ )中途輟學的人數是否有顯著差異」，應採用何種假設檢定方法與統計指標？為什麼？(15分)

(二)請依假設檢定的步驟來檢驗「甲縣市( $X_1$ )與乙縣市( $X_2$ )中輟學生人數是否有顯著差異」，並說明你的結果？(以  $\alpha=0.05$  進行假設檢定)[ $t_{(975,df=8)}=2.306$ ](10分)

1. 《考題難易》：★★簡單
2. 《破題關鍵》：由散布圖判斷二個變項是否相關及檢定二個相依樣本的平均數，近幾年常考的題型。

【擬答】

(一)



1. 由散布圖得知  $X_1$ 、 $X_2$  具有正相關。
2. 因為當  $X_1$  愈大時， $X_2$  也愈大。
3. 由 1. 得知  $X_1$ 、 $X_2$  有相關，所以應採用相依樣本  $t$  檢定來檢定甲乙兩縣市中輟學生平均人數是否有差異。

(二)

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$$

設母體為常態分配，利用相依樣本  $t$  檢定， $\alpha=0.05$ ，令  $d = X_1 - X_2$

公職王歷屆試題 (110 地方政府特考)

拒絕域  $C = \{t \mid t > 2.306 \text{ 或 } t < -2.306\}$

檢定統計量

$$t = \frac{\bar{d} - \mu_d}{\frac{S_D}{\sqrt{n}}} = \frac{-28 - 0}{\frac{21.9}{\sqrt{9}}} = -3.836$$

落入拒絕域  $\rightarrow \text{ReH}$ 。

結論：有證據顯示甲縣市與乙縣市中輟生人數有顯著差異，但仍有 5% 之機率犯型 I 錯誤。

三、某縣(市)新編國一版本智力測驗，抽樣 2500 名國一學生來建置常模，計算出學生們在該測驗的得分平均數為 60 分，標準差為 10 分，並且計算出該測驗的內部一致性信度(KR-20)為 0.91。請回答下列問題：

- (一)根據本測驗的信度，該生的測量標準誤(standard error of measurement, SEM)應為多少分？此測量標準誤是什麼意思？(10 分)
- (二)若將該測驗的成績轉換為平均數 100，標準差 15 的智商分數(IQ)，則在該測驗上得到 80 分的學生，其智商分數為多少？(10 分)
- (三)若要了解該測驗是否能有效測量出國一學生的智商，可以使用何種方法來證明？請舉出兩種方法，並詳細說明你的程序。(10 分)

1. 《考題難易》：★★★普通

2. 《破題關鍵》：測量標準誤、常模以及效度證據建立概念一直以來是教育測驗命題重點，此次命題並未刻意為難同學，所以準備好的同學應可以拿到不錯分數。

【擬答】

(一)測量標準誤：

1. 簡介：

實際上我們無法得知真實分數，因此只能反過來用一次得到的測驗分數來推測其真實分數的所在範圍。因此，以測量標準誤解釋個人分數是以一段分數範圍表示，較為客觀。比如說，A 生 IQ 得分為 150，B 生為 147。從表面上看 A 的 IQ 比 B 還要高，但是如果考量到測量標準誤，則兩者三分之差可能是在測量誤差之內。因此實際上這樣的分數差異並無太大意義。因此，我們不能說 A 生 IQ 必定大於 B 生。

2. 題意計算過程：

$$SE_{meas} = S_X \sqrt{1 - R_{XX}} = 10 \sqrt{1 - .91} = 3$$

(二)測驗分數解釋：

$$Z = \frac{X - \mu}{\sigma} = \frac{80 - 60}{10} = 2$$

$$IQ = 15Z + 100 = 2 \times 15 + 100 = 130$$

(三)建構效度證據建立：

1. 建構效度的簡介

所謂建構 (construct) 是指一種理論性的架構，用以說明某有持續性的心理特質或屬性，例如：智力、性向、人格、精神疾病等。如果一個測驗具有良好的建構效度，它所測得的分數便可真實地反映出該測驗所指稱的心理特質，而測驗者也可以依據該心理建構的理論來詮釋測驗結果。

2. 建構效度的證據建立方式

建構效度的檢驗應由不同的觀點以不同的方法進行，因此無法以單一的統計數據來代表整個測驗的建構效度，相反地，我們要以累積的實證資料對測驗各層面的建構效度作整體的研判。以下列二種為例進行說明：

(1)發展上的改變：

心理學上有些心理特質的發展與生理年齡有密切關係，譬如：智力、學習成就、學前兒童的動作發展等。因此檢驗這些測驗效度時，便可比較不同年齡（年級）層的受試者的平均表現是否隨著年齡（年級）遞增而增加，果若如此，則該測驗效度即獲得肯定。

(2)對實驗處理效果的敏感性：

有些實驗處理是針對所要測的心理特質進行操縱，因此若前測、後測的分數有顯著差異，便可說明此測驗能有效地測到實驗處理所針對的變項。則下列二圖顯示甲測驗比乙測驗有較高的效度。

另外，如果有一測量「考試焦慮」的人格量表，將它於大學聯考之前對一群高三學生施測，然後在大學聯考後進行後測，則前後測之間必然會有顯著差異，否則該測驗便難以證明它所測的是考試焦慮。

參考資料：郭生玉著。教育測驗與評量。精華出版。

命中：思澤老師，講義第三章 信度、第四章 效度、第五章 常模

志光.保成.學儒

快速考取

全方位智慧  
服務系統



線上.線下 給您

最強大的支援

激推！考生必看

公職王影音頻道



名師精析各科目考試重點、口面試準備技巧、上榜生經驗分享等全方位影音資訊。

線上模擬測驗



運用海量題目，協助考生訓練解題速度，檢視學習成效並及時修正弱點。

歷屆試題下載



收錄各年度國考試題及解題，讓考生練習考古題時更易突破盲點，找到解題關鍵。

國考申論加分



各考試領域專業文章分析解讀趨勢動態，協助考生加深各科目的答題內涵。

公職王歷屆試題 (110 地方政府特考)

四、下表是針對不同居住地區隨機抽選某高中，再針對該高中隨機抽選四個班級的所有學生，進行擁有智慧型手機的人數研究，請依此回答下列問題：

	有智慧型手機	無智慧型手機	合計
北部	120	30	150
中部	90	10	100
南部	60	40	100
東部	30	20	50
合計	300	100	400

(一)此研究的抽樣方式稱為什麼？它與分層隨機取樣(stratified random sampling)有何不同？(10分)

(二)若想了解居住地區與是否擁有智慧型手機兩者之間有無關聯，要使用何種統計檢定方法來進行分析？請進行假設檢定，說明你的統計分析程序與結果？(以可接受的犯錯率為5%( $\alpha=0.05$ ))[ $\chi^2_{(.95,df=3)}=7.8$ ] (15分)

1. 《考題難易》：★★簡單  
2. 《破題關鍵》：抽樣方法及卡方獨立性檢定，常考題。

【擬答】

(一)此抽樣方法稱為二階段群集(部落)抽樣。其做法將母體依某種標準分成若干個性質相同的群集，從中抽幾個群集調查。此方法適合用在群集間差異小而群集內差異大時最為有效。而分層隨機抽樣為將資料依某種標準分成若干層，每層抽出一些樣本調查。此方法適合用在層間差異大而層內差異小時最為有效。

(二)利用卡方獨立性檢定：

	有智慧型手機	無智慧型手機	合計
北部	120 (112.5)	30 (37.5)	150
中部	90 (75)	10 (25)	100
南部	60 (75)	40 (25)	100
東部	30 (37.5)	20 (12.5)	50
合計	300	100	400

$H_0$ ：居住地區與是否擁有智慧型手機無關聯

$H_1$ ：居住地區與是否擁有智慧型手機有關聯

$\alpha=0.05$

拒絕域  $C = \{x^2 | x^2 > 7.8\}$

檢定統計量

$$x^2 = \sum \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(120 - 112.5)^2}{112.5} + \dots + \frac{(20 - 12.5)^2}{12.5} = 32 > 7.8 \text{ 落入拒絕域} \rightarrow \text{Re}H_0$$

結論：有證據顯示居住地區與是否擁有智慧型手機兩者之間有關聯，但仍有5%之機率犯型I錯誤。