

## 102 年公務人員高等考試三級考試試題

類 科：教育行政

科 目：教育測驗與統計

一、PISA( Programme for International Student Assessment, 中譯為國際學生能力評量計畫)係一種國際性評量計畫,我國各界對於此評量也頗為重視,並於 2006 年首次參與國際評比。試就有關 PISA 回答下列問題:

(一)PISA 係那一個組織所籌劃?其目的為何?包含那些內容?施測的對象為何?

(二)PISA 何以採素養(literacy)而非學習成就(achievement)的觀點設計題目?

【擬答】:

- (一)1. 國際經濟合作與發展組織 (Organization for Economic Co-operation and Development, OECD) 提出。
2. 自公元 2000 年每三年評量一次的國際性學生評量計畫 (PISA), 主要功能評量個人參與社會活動能力、適應現代生活的基本能力、可以進行國際間比較與相互觀摩、評量具有時間先後順序比較、評量嚴謹以專書方式出版、評量結果兼重平均數與離散量數。
3. 評量內涵包括①閱讀能力 (Reading Literacy): 一套知識、技能與策略進步,個人從生活不同脈絡與透過同儕互動所建立之能力;②數學能力 (Mathematic Literacy): 培養學生面對未來挑戰的能力,包括分析、推理及有效溝通概念,透過公式、形式解決不同領域與情境之數學問題;③科學能力 (Science Literacy) 使用科學知識界定問題與「證明基礎」的結論,瞭解、幫助與探究人類活動自然世界。(張鈿富、王世英、吳慧子、周文菁, 民 95)
4. 施測對象: 各國受完義務教育後的 15 歲學生。
- (二)素養 (Literacy) 指的是運用知識和技能處理日常問題的能力,強調教育系統與運作成效評量、評量內涵重視學生生活化與終生學習能力養成。學習成就為一般學校的學科「成就」測驗,較著重學科概念的了解與知識的獲得,較無法在現實生活應用。

二、試回答下列有關測驗選用的問題:

- (一)某教育局先後編製兩套(甲和乙)閱讀成就測驗。甲的折半信度為.90,乙的重測信度為.75。乙有其教學目標的內容效度。甲與某閱讀測驗有同時效度,但甲有一些測量的內容並未被老師教到。若老師擬評量其閱讀教學的效果,其前、後測宜用那一套測驗較適合?請說明你的理由。
- (二)有一測驗機構發行兩套(A 和 B)閱讀準備度測驗。此兩測驗國小一年級新生適用。測驗 A 的內部一致性信度為.92,測驗 B 的重測信度為.76。兩者皆未提供有關效度的資料。如果校長向你請教測驗的選用,你會推薦那一套?請說明你推薦的理由。

【擬答】:

- (一)1. 建議用甲閱讀成就測驗。
2. (1)折半方法採用單獨一次測驗結果,利用隨機方式或奇偶數題將試題分成二半,求二半測驗分數的相關,稱折半相關或內部一致性係數 (Coefficient of Internal Consistency),心理測驗人員大都將折半信度視為重測信度延伸。且具有同時效度 指測驗分數與外在效標取得在同一時間內連續完成,計算此二種資料的相關係數即為同時效度,目的在使用閱讀成就測驗分數估計個人在效標方面的實際表現,亦即在估計或診斷目前的實際情況。
- (2)閱讀成就測驗題目性質會因重測而改變,像推理與思考題目極易因第一次施測時因思考而貫通原則或瞭解解決方法,在第二次時便只需記憶即可。且甲套閱讀成就測驗只有.75,代表成就測驗結果易受時間誤差影響。且乙套閱讀成就測驗其內容效度缺乏部分教學內容,會使得測驗試題樣本內容缺乏「教學目標」或「教材內容」的代表性或適切性。
- (二)1. 建議用 A 套閱讀準備度測驗。

公職王歷屆試題 (102 高普考)

2. 閱讀準備度測驗便是幼兒就學準備度的前身，用來測驗幼兒進入小學一年級時的學業知識準備度；目的在衡量受試者對試題反應的一致性，可以使用內部一致性信度 Cronbach  $\alpha$  係數方法，因此法適用於二分化或多元化計分的測驗資料，依據受試者對所有試題的反應，分析試題間的一致性。而在信度考驗上，一份優秀的教育測驗至少應該具有 0.80 以上的信度值，才具有使用的價值（余民寧，1997）。

三、有一位體育系教授擬對某體育專業學校的學生調查其對「身體形象與運動價值觀」的研究，該校學生人數為 3,850 人，學生來自臺灣北中南東各區，學校提供的人數資料如表一：

表一：體育專業學校學生人數分配

性別	地區				和
	北區	中區	南區	東區	
男生	570	581	555	313	2019
女生	518	523	501	289	1831
和	1088	1104	1056	602	3850

研究者因經費、時間及人力的關係，僅能抽取十分之一(也即 385 人)的學生為受試者。同時為了考慮樣本代表性，於是分層隨機抽樣(stratified random sampling)的方式抽取樣本。試根據以上敘述及表一中的資料，回答下列問題：

(一)試以分層隨機抽樣的方法，計算出表二之「？」，以完成此研究者的抽樣工作。

(二)表一：體育專業學校學生人數抽樣結果分配

性別	地區				和
	北區	中區	南區	東區	
男生	?	?	?	?	?
女生	?	?	?	?	?
和	?	?	?	?	385

(三)有關學生的變項很多(如社經地位、體型、智商等)，何以研究者要以性別及地區為其抽樣的「層」(strata)?

(四)分層隨機抽樣方法雖較能確保樣本的代表性，但要採用此方法時母群體要顯示出何種基本資料?此方法才可進行。

【擬答】：

(一)

性別	地區				和
	北區	中區	南區	東區	
男生	57	58	56	31	202
女生	52	52	50	29	183
和	109	110	106	60	385

(二)1. 分層隨機抽樣 (Stratified Random Sampling)：屬於隨機抽樣方法之一，抽樣前先依某種標準分類成若干層，然後分別以簡單隨機抽樣，自每一層抽取若干個體作成樣本，母體有多少層則樣本亦有多少層，此抽樣方法稱為分層隨機抽樣。提供分層基準如收入、職業、學校規模大小、性別、年齡等。

2. 按性別分為二個副母體，各副母體分別採「分層比例隨機抽樣法」，依地區分層（分為北部地區、中部地區、南部地區、東部地區），各層依學校提供人數資料比率決定應抽樣本數。如採用社經地位、體型與智商較不代表性，抽取樣本無法代表全國體育專業學校的學生，且社經地位並非本文關心研究變項，且社經地位會影響智商，並不適宜分層隨機抽樣。

(三)適用於母體層內同質層間異質、結構複雜，且母體人數比例不一情況。

公職王歷屆試題 (102 高普考)

四、數位化教學目前甚受重視，但其成效說法不一，於是校長邀集專家規劃一項實驗，選定一個學科進行一學期的教學實驗。其教學分為傳統講授、純數位化教學及講授主數位輔三種方式。教學環境分為一般教室（有必要之 e 化設備）及 e 化教室。教學實驗即以獨立樣本雙因子設計進行，實驗完成後再施以該學科就的測驗，經統計分析之部分結果如表三及表四。

表三：後測分數之描述統計

教學環境	教學方式	平均數	標準差	人數
一般教室	傳統講授法	77.40	2.408	5
	純數位化教學	72.60	1.817	5
	講授主數位輔	76.40	2.608	5
e 化教室	傳統講授法	75.00	2.121	5
	純數位化教學	74.20	1.924	5
	講授主數位輔	80.80	1.924	5
總數	傳統講授法	76.20	2.486	10
	純數位化教學	73.40	1.955	10
	講授主數位輔	78.60	3.169	10
	總數	76.07	3.300	30

表四：後測分數之變異數分析摘要表

變異來源	平方和	自由度	均方	F 值	顯著性
教學環境	10.800	1	10.800	2.331	.140
教學方式	135.467	2	67.733	14.619	.000
教學環境 x 教學方式	58.400	2	?	?	.006
誤差	?	?	?		
校正後之總和	315.867	29			

(一)檢定此問題的主要效果及交互作用效果之（虛無 VS. 對立）假設如何陳述？

(二)請先計算表四中「？」之數值後，再解釋你在(一)所陳述的假設（設顯著水準  $\alpha = .01$ ）。

(三)如果交互作用效果有統計上的意義或達到 .01 顯著水準，則下一步驟該如何進行？你可以用文字敘述說明你的做法或根據表三繪製交互作用效果之剖面圖（你可以教學環境或教學方式為圖之水平軸繪製交互作用效果之剖面圖）。

(四)綜合上述，你如何給此教學實驗做結論？

【擬答】：

(一)統計假設

1. 教學環境統計假設

$$H_0 : \mu_{A_1} = \mu_{A_2}$$

$$H_1 : \mu_{A_1} \neq \mu_{A_2}$$

2. 教學方式統計假設

$$H'_0 : \mu_{B_1} = \mu_{B_2} = \mu_{B_3}$$

$$H'_1 : \mu_{B_i} \text{ 不全等}$$

3. 交互作用統計假設

$$H''_0 : \alpha\beta_{ij} = 0$$

$$H''_1 : \alpha\beta_{ij} \neq 0$$

(二)

1. 變異數分析摘要表：

變異來源	平方和	自由度 (df)	均方(MS)	F	顯著性
教學環境	10.800	1	10.8	2.331	.140
教學方式	135.467	2	67.733	14.619	.000
教學環境x教學方式	58.400	2	29.2	6.303	.006
誤差	111.200	24	4.633		
總和	315.867	29			

2. 結論：

- (1)  $F_A = 2.331$ ，顯著性.140，落入接受區，接受虛無假設，亦即無充分證據支持不同教學環境對學科測驗成績有顯著差異，無主要效果。
- (2)  $F_B = 14.619$ ，顯著性.000，落入拒絕區，拒絕虛無假設，亦即有充分證據支持表示不同教學方式對學科測驗成績有顯著差異，有主要效果，惟犯型 I 錯誤 1%。
- (3)  $F_{AB} = 6.303$ ，顯著性.006，落入拒絕區，拒絕虛無假設，亦即有充分證據支持不同教學環境與教學方式對學科測驗成績有顯著交互作用，亦即不同教學環境是否影響學科測驗成績，必須視採用教學方式而定，惟犯型 I 錯誤 1%。

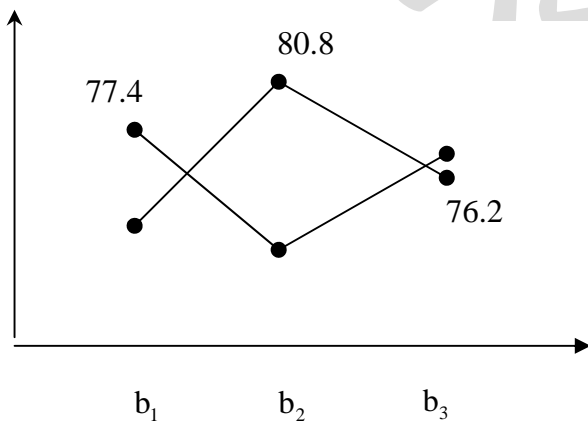
(三) 1. 實驗者必須進行單純效果考驗。

2.

AB 摘要表平均數

	傳統教授法 b <sub>1</sub>	純數位化教學 b <sub>2</sub>	講授主數位輔 b <sub>3</sub>
一般教室 a <sub>1</sub>	77.4	72.6	76.4
E 化教室 a <sub>2</sub>	74.2	80.8	76.2

本題為無次序性交互作用 (Disordinal Interaction)，獨立樣本二因子變異數分析，存在有交互作用時，畫出來的二條線，相交之交互作用稱之，如下圖。



(四) 此教學實驗必須進行單純效果。考驗 A 因子中某一個「水準」在 B 因子中每一個「水準」之間效果差異，或考驗 B 因子中某一個「水準」在 A 因子中每一個「水準」之間效果差異即為單純效果 (Simple Effects) 的考驗。例如：考驗教學環境中一般教室在三種教學方式 (傳統教授法、純數位化教學、講授主數位輔) 對學科測驗成績的影響？