

# 《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? (20分)

(二)PISA何以採素養 (literacy) 而非學習成就 (achievement) 的觀點設計題目? (5分)

## 試題評析

本題因PISA在台灣是由教育部與國科會統籌規劃的大型國際計劃。加上台灣高一的15歲學生樣本的部分領域素養曾高居全世界第一, 當然就成為教育部門得意之教改成果的正面績效與政績宣傳重點。雖然, PISA已經有多年執行的成果, 但是為呼應明年啟動的延長國教政策, 今年再度被提出大加論述, 雖然在高考試題出現不免有所爭議, 但是將此歸為時事題也不為過。考生平日接觸教育時事與新聞的付出, 在此刻獲得豐碩成果。命題上屬於意外, 考生不須受單一題目影響後續題目作答心情。

答:

(一)國際學生能力評量計畫(PISA), 為經濟合作暨發展組織(Organisation for Economic Co-operation and Development, OECD)自1997年起所籌劃, 此跨國評量計劃從終身學習的面向來看待教育的真諦; 包含正式與非正式的環境, 諸如正規課程、課外社團、家庭環境、學校氣氛等。這個計畫是由所有OECD會員國與其他非會員國的夥伴國家(地區)共同合作執行。從2000年正式推出後, 參與國逐次遞增, 截至目前為止, 有超過68個OECD會員國與夥伴國(地區)參與計劃, 約略涵蓋了世界87%的經濟體, 有超過一百萬名學生接受評量, 除了紙筆測驗外, 亦有開發線上評量系統。

PISA的施測對象是十五歲學生, 目的是就其生活知能的學習成效提供跨國際的比較, 以及各國教育效能的分析, 並由此界定國民素養的內涵。PISA評量計畫為檢視學校教育結構如何影響教育產出的品質和均等, 而關於教育均等議題的探討, 主要聚焦在社會經濟背景指標所能解釋的變異比率。亦即國家內公私立學校、城鄉間、各族群(種族、移民類別)間、教育品質和受教育的機會是否公平均等。

PISA 評量內容涵蓋閱讀、數學和科學三個領域的素養程度。以下就三個領域的評量內容進行說明:

### 1. 閱讀:

所謂閱讀的素養意旨, 產生閱讀行為的各種情境。情境架構概括為: 個人、教育、職業、公眾。由於題材取自生活中廣泛的文字訊息, 文本的形式相當多元, 如: 散文、敘事、論述、廣告文宣、官方文件或聲明、故事寓言、報告表單等。評量的重點在於對文本訊息的擷取、發展解釋、省思與評鑑文本內容、形式與特色。

### 2. 數學:

主要定義包含四大概念: 數量、空間與形狀、改變與關係、不確定性。所須的技巧如: 數學語言的理解、建模、解題假設執行。評量架構則定義為五種情境: 個人的、教育的、職業的、公眾的、科學的, 題材融入了現在國民都會碰到的全球暖化、溫室效應、人口成長、浮油與海洋、酸雨或運動常識等課題, 生活化的課題都是可涵蓋的範圍。

### 3. 科學:

主要概念包含了物理、化學、生物、科學及地球與太空科學等。應答能力上, 要求受測學生須解釋及預測科學現象、提供假設、使用科學證據並瞭解科學調查, 命題領域擴及: 生命與健康科學、地球與環境科學及科技科學。

(二)PISA試題的設計著重在應用及情境擬態, 並不圈限於課程內容, 受測學生須先把資訊理解並將統整、評鑑、省思能力靈活運用, 自行建構問題情境的答案。而評量的焦點是在於年輕人能否使用習得知識技能, 面對真實世界的挑戰而非僅是學校課程的精熟程度。因此, PISA以素養而不以學習成就之觀點設計題目。

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。(15分)

- (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。(10分)

<b>試題評析</b>	本題提供不同測驗的信效度數據與證據，要求考生給予使用者與施測者適當的採用建議。考生可依平日學習對於各種信效度的定義與適用時機，甚至對應不同測驗類型的強調與探討重點，應可容易對答。
<b>考點命中</b>	《高點教育測驗與統計講義第三回》第14、15章，傅立葉編撰，頁12、23-24。

**答：**

(一)就教育局先後編製的甲乙兩套閱讀成就測驗，依題目敘述各有信度與效度的數據與說明。若某老師擬評量對於學生閱讀教學的效果，則前測宜採甲測驗，後測可採乙測驗。理由如下：

- 1.甲測驗因為高達 0.9 的折半信度，且與某閱讀測驗有同時效度，儘管因某些測量內容未被老師教到，而有效度不足的疑慮，但是折半信度具備不須編製複本且不需考量重測間隔時間的優點，較適合於前測階段使用。
- 2.乙測驗的重測信度為 0.75，且具備教學目標的內容效度，對於閱讀教學的效果而言，在後測階段是一個較佳選擇，因為成就測驗以內容效度為重。

(二)至於，某一測驗機構所發行的兩套閱讀準備度測驗，依題目提供的信度證據，本人會推薦校長採用測驗 A，理由如下：

- 1.因測驗適用於小一學生，其年齡小，施行重測的可能性較低，且影響重測結果分數的因素不少，測驗 B 不予考慮。
- 2.測驗 A 的內部一致性信度高達 0.92，表示該測驗的題目組合同質性高，而且閱讀準備度的測驗目的一致，能夠較準確的測量小一學生閱讀準備度的程度高低。

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。(10分)

表二：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

- (二)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)?(10分)

(三)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。(5分)

<b>試題評析</b>	本題考的分層隨機抽樣是推論統計之抽樣方法的基礎與重點。難度不高，卻有鑑別力，是過去公職考試罕見的命題。本班總複習講義中特別編列為重點考試方向。
<b>考點命中</b>	《高點教育測驗與統計總複習講義》，傅立葉編撰，頁19-20。

**答：**

(一)因研究限制，該體育系教授僅能抽樣10%的385位學生樣本，依分層隨機抽樣方法，抽樣工作須依10%的比例配置，抽樣結果可以完成如下表：

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)由於該研究的重點在於身體形象與運動價值觀，以性別與地區作為抽樣的層，可能是文獻回顧中所獲得先前類似研究並無人著墨，而有相當大的研究創新與貢獻考量外，其他有關學生變項因有測量問題以及本質上非類別變項而有分組問題，較不適合作為抽樣之層。

(三)母群體資料應有以下幾個條件才適合採用分層隨機抽樣：

- 1.分層不能有重疊現象
- 2.母體內樣本單位之差異較大時
- 3.分層後能達到層間變異大、層內變異小

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成就的測驗，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 × 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?(5分)
- (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 $\alpha=.01$ )。(10分)
- (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)。(5分)
- (四)綜合上述,你如何給此教學實驗做結論?(5分)

試題評析	本題考的是變異數分析較複雜之包含交互作用的雙因子設計。但是,題目難度不高,完成表四的五個未知數的簡單計算後,足以進行檢定結論,關鍵影響在於交互作用達顯著後的進一步分析與做法,一般程度考生在本題應可拿到20分左右的分數。本班考生因平日教材與考古題的講述與練習,充分吸收之應用者應能從容應對。
考點命中	《高點教育測驗與統計講義第二回》第13章,傅立葉編撰,頁49-50。

答:

(一)欲檢定此題對於數位化教學成效之探討,因教學實驗採取獨立樣本雙因子設計進行,對於主要效果與交互作用效果的虛無與對立假設可以分別陳述如下:

1.教學方式的主要效果

$H_0$ :不同教學方式的教學成效無顯著差異

$H_1$ :不同教學方式的教學成效不全等

2.教學環境的主要效果

$H_0$ :不同教學環境的教學成效無顯著差異

$H_1$ :不同教學環境的教學成效不全等

3.教學方式與教學環境的交互作用效果

$H_0$ :教學方式與教學環境的不同組合之交互作用對於教學成效無顯著差異

$H_1$ :教學方式與教學環境的不同組合之交互作用其教學成效不全等

(二)表四中之五個未知數由左而右,由上而下依序計算如下:

誤差平方和

$$315.867 - (10.800 + 135.467 + 58.400) = 111.2$$

誤差自由度

$$29 - (1 + 2 + 2) = 24$$

教學環境\*教學方式均方

$$58.4 / 2 = 29.2$$

誤差均方

$$111.2 / 24 = 4.633$$

教學環境\*教學方式F值

$$29.2 / 4.633 = 6.303$$

在0.01的顯著水準下,利用表四最後一個欄位提供的顯著性(p-值),可以進行在(一)中之假設的檢定結論為:

教學方式的主要效果,以及教學方式與教學環境的交互作用效果,因為顯著性皆小於顯著水準,而得拒絕虛無假設的結論,也就是檢定結果達顯著。但是,教學環境的主要效果並未達顯著,意即一般教室與e化教室的不同,對於學習成效並未造成顯著差異。

(三)當教學方式與教學環境的交互作用效果因達0.01的顯著水準,而具有統計上的意義,研究者必須進一步固定教學環境兩個不同水準(level),分別進行教學方式之單因子ANOVA;另固定教學方式之各level,進行教學環境之單因子ANOVA。因此題意,共需再進行五個單因子變異數分析,以進一步釐清單純主要效果背後的詳細資訊。

(四)綜合以上分析,此一教學實驗起碼提供身為校長的研究者,有關傳統講授法、數位化教學以及兩種方式搭

配下的相輔相成等三種不同教學方式，確實對於學生的學習成效造成顯著影響。儘管教學環境的主要效果不顯著，但是因兩因子的交互作用達顯著所進行的後續分析，提供進一步的細緻資訊，此一實驗設計可以稱得上是成功的。

# 高點 · 高上

【版權所有，重製必究！】