

《教育測驗與統計》

一、某教育學者根據文獻評閱心得發現，一份「學科成就測驗」編製的好壞，將影響到能否有效評估學生學習成就的正確性。請問：你會提醒他注意那些因素，以免他所發展出來的測驗工具缺乏效度 (validity)？(25分)

試題評析	本題考的是心理測驗的效度。但是，題意特別針對學科成就測驗，並提及測驗編製的好壞，因此，答題上如果考生針對一般性的測驗效度影響因素進行作答，容易因無法貼近題意而失去焦點，所以答題前對於題目的閱讀與理解非常關鍵。不過，題目難度不高，正常情況下考生應能取得基本得分。
考點命中	1.《高點·高上教育測驗與統計講義》第三回，傅立葉編撰，Ch.16，頁27-29。 2.《高點·高上教育測驗與統計總複習講義》，傅立葉編撰，壹.考情分析，教育測驗重點內容2，頁3。

答：

根據題意，該教育學者根據文獻評閱心得發現，一份學科成就測驗編製的好壞將影響是否能夠有效評估學生學習成就的正確性。基本上，測驗工具如果要具備高效度，應該注意以下幾個因素：

(一)測驗所要評估學習成就的定義與內涵：

就不同科目而言，學科學習成就應有數種不同的面向，不同面向的教學目標應能清楚定義或依理論上之高低層次加以清楚說明並針對各項目標進行命題。

(二)測驗試題組合的來源與題數：

成就測驗一般著重於個別學科的重點章節與內容，在考試範圍內盡可能就教學目標與教材內容將題目出處與測驗之教學目標作對應，也就是使用雙向細目表，除提供試題分布與組合的清楚架構外，並盡可能增加題數，應能達到對於內容效度(content validity)的要求。此時，試題組合的適切性若能尋求並取得同學科領域專家的共同認定，對於內容效度更有保障。正常情況下，題數愈多，測驗長度愈長，試題涵蓋的考試範圍愈齊全，不僅達成測驗的高信度，測驗的效度也會愈高。

(三)測驗的品質：

一個好的測驗工具應經過難度與鑑別度等兩大品質指標的檢視與考驗。難度接近0.5之難易適中的個別試題會有較高的鑑別度，美國的測驗學者Ebel & Frisbie(1991)曾提出一套鑑別度的判斷標準，0.4以上的鑑別度為非常優良，0.3以上但未達0.4的鑑別度則為優良，但仍須小幅修改。當試題都有適當的難度與高的鑑別度，自然就會有高的效度。

(四)測驗編製階段，效標的選擇與品質：

當學科成就測驗的施測目的有可能與未來的教學、輔導或安置有所連結，則解讀測驗效度高低經常與選擇的效標有很大的關係。效標的新近性、客觀性、適切性、經濟性在在影響在編製階段對於測驗的評價，因此效標的選擇非常重要，效標的品質影響測驗的最終題本，也影響測驗的效度。

其他諸如：受試者的心理認知、團體變異性、測驗的實施……，雖然不是測驗編製時應注意之影響效度高低的因素，但是最終仍會對測驗效度有所影響，以上是值得提醒該位教育學者的幾項因素。

二、某教育學者想探究自然科學教育中，教師是如何進行教學評量的。他建議第一線的教師，應該採行實作評量 (performance assessment) 方式來進行，方能評估出學生動手操作的學習成就。請問：第一線的教師可以採用什麼評定方法，來評量學生動手操作的學習成就？(25分)

試題評析	本題考的是多元評量之一的實作評量。實作評量的類型、設計原則、可使用工具，甚至評分規準都是近幾年的考試重點。有準備的同學應該不覺得困難，惟答題得分關鍵為題意鎖定的自然科學教育，考生應能在各種評量方式中善用自然科學領域相關的舉例，以便拿高分！
------	---

考點命中

- 1.《高點·高上教育測驗與統計講義》第三回，傅立葉編撰，Ch.19，頁53-57。
- 2.《高點·高上教育測驗與統計總複習講義》傅立葉編撰，肆.重點摘要，各章重點提示，Ch.19，頁68-70。
- 3.《高點·高上106高普考題神》教育測驗與統計，傅立葉編撰，一，頁1。

答：

「實作評量」一詞翻譯自英文Performance Assessment，簡而言之，就是由教師設計相關的情境，由此情境，針對學生所應達到的學習成果，設計一些問題，讓學生在情境中或實際參與實驗操作或觀察之後，以分組活動或個別思考的形式，進行問題的解決，同時針對學生在過程中的表現，以客觀的標準加以評分的一種評量方式。

第一線教師可以採用的實作評量方式有以下幾種：

(一)紙筆表現

紙筆表現有別於傳統所使用的紙筆測驗，它更強調在模擬情境中應用知識與技能。這種紙筆評量方式的應用可能是最後學習成果的展現，例如：在自然科學教育中，要求學生寫出檢查和啟動儀器設備的安全步驟，也可以作為動手操作前的初步評量。

(二)辨識測驗

是指以實物作為刺激，但所要求的反應仍是語文反應。例如：在某種情境下，可能只要求學生辨認一套工具或一組器具，並指名它們的功用而已；或比較複雜的，教師操作一組實驗設施給學生看，要求他們指出最可能故障或最關鍵的部分，並說出應採取的檢查步驟及所需用到的工具。

(三)結構化表現測驗

此評量要求學生在標準且控制下的情境完成實作作業，測驗情境的結構性甚高，它要求每位學生都能表現出相同的反應動作。這種測驗和標準化測驗一樣，特別強調施測和計分過程的一致性和公平性，所以在儀器設備、材料、時間、要求的實作上都有一致的規定，而在計分上，通常根據多個標準來評定，例如：成品精確性、實作步驟的適當性、工作速度、符合規定的程度。

(四)模擬表現

模擬表現即為配合或替代真實情境中的表現，局部或全部模擬真實情境而設立的一種評量方式。例如：在生物課程中，針對一個假想的生態鏈中之互為天敵的兩種昆蟲，模擬其互動與相互攻擊的原因，進行記錄。

(五)檢核表

把要觀察的行為或作品先分解成幾個要素，列成一清單，然後逐一作判斷，最後再合計總分。而檢核表即是一組列出表現或成果的測量向度，並且提供簡單記錄「是」或「否」判斷的資料表。適用於當作品或動作可以細分為一系列明確且具體的要素，而這些要素又可以作二分法判斷（是或否、有或無、合格或不合格）。

三、在描述統計學中，試比較平均數（mean）、中位數（median）及眾數（mode）三者，在常態分配、正偏態分配、負偏態分配中的大小關係？（25分）

試題評析	本題命題屬於心理統計學的描述統計範圍，有關三個集中量數在不同資料分配型態中的大小關係，應該是今年整份試題難度最低的一個題目。答題時，應能搭配對於三種分配型態的簡單說明才能顯示考生對於此題的了解深度。當然，能夠輔以圖示並標示三個集中量數的大小位置(在此略過)加以說明定能取得高分！
考點命中	1.《高點·高上教育測驗與統計講義》第一回，傅立葉編撰，Ch.3，頁14。 2.《高點·高上教育測驗與統計總複習講義》傅立葉編撰，肆.重點摘要，各章重點提示，Ch.3~5，4.重要應用題型，頁14。

答：

在描述統計中，屬於集中量數的平均數、中位數及眾數不僅提供研究資料或實務數據集中的數據或位置，三者在不同資料分配中的大小關係如下說明：

(一)常態分配

常態分配是對稱分配的一種特例，常態分配下，平均數、中位數及眾數大致相等，三者的數值皆落在常態曲線最高點往數線軸所作垂直線的交接點。

(二)正偏態分配

正偏態分配又稱右偏分配，因為其圖形具備緩長坡在數值較大的右邊之特性，三個集中量數的大小關係，由大到小依序為：平均數、中位數、眾數。

(三)負偏態分配

負偏態分配又稱左偏分配，因為其圖形具備緩長坡在數值較小的左邊之特性，三個集中量數的大小關係，由大到小依序為：眾數、中位數、平均數。

四、某教育學者想知道學習成就上「城鄉差距」的情形。他利用中研院TEPS資料庫中的「綜合分析能力測驗」成績（以T分數表示）為資料分析依據，獲得下列統計報表。

Descriptive statistics

地區分類	N	Mean	Sd	Se	Min	Max
鄉村	927	44.27	9.360	.307	12	71
城鎮	5200	48.58	9.843	.136	15	82
都市	7841	51.62	9.783	.110	16	82
全體	13968	50.00	10.00	.085	12	82

Summary table of analysis of variance

SV	SS	df	MS	F	p
Between	61508.9	2	30754.43	321.67	.000
Within	1335191.1	13965	95.61		
Total	1396700.0	13967			

試問：（每小題5分，共25分）

(一)本研究所列舉的虛無假設為何？

(二)本研究所使用的統計檢定方法為何？

(三)本研究檢定結果所達成的第一類型錯誤率為何？

(四)都市與鄉村之間的測驗成績相差多少分？

(五)本研究檢定結果是否可以支持該學者所認為的「學習成就上確實存在著城鄉差距」的看法？

試題評析	本題主要考的是一因子變異數分析的F檢定。解題不涉及數學計算，卻需要能夠有閱讀統計報表的能力，特別是檢定結論依據的p-值法，以及第一類型錯誤的意義與應用，在本班日常授課已經充分練習，甚至在本班考前的考猜也精準命中，應該在本題的答題與拿分上沒有太大問題。
考點命中	1.《高點·高上教育測驗與統計講義》第二回，傅立葉編撰，Ch.13，頁48-49。 2.《高點·高上教育測驗與統計總複習講義》，傅立葉編撰，壹.考情分析，統計學重點內容4，頁2。 3.《高點·高上106高普考題神》教育測驗與統計，傅立葉編撰，九，頁10-12；十四，頁14。

答：

根據題意，該教育學者想知道學習成就之城鄉差異情形。利用資料庫進行資料分析所得之「描述統計」與「變異數分析摘要表」中的資訊，分別回答各子題如下：

(一)本研究所列舉的虛無假設為，鄉村、城市與都市的平均綜合分析能力沒有顯著差異，以符號表示為：

$$H_0: \mu_1 = \mu_2 = \mu_3$$

(二)由於針對三種城鄉規模所進行之綜合分析能力是否存在城鄉差異的檢定參數為三個獨立樣本的母體平均數，本研究所使用的統計檢定方法為一因子變異數分析的F檢定。

(三)依據「變異數分析摘要表」中最後一個欄位.000的p-值，表示在321.67這個非常大的F檢定統計量以右極端區域的機率為0，因此，在13968個學生樣本下，該檢定應能毫無疑問的正確拒絕虛無假設，所以檢定結果達

成的第一類型錯誤率為0。

- (四)基於前述(三)之F檢定達顯著的結論，表示三個城鄉規模的平均綜合分析能力不全等，由「描述統計」表中資訊可知，平均數最大的都市與平均數最小的鄉村之間測驗成績必然存在差異，其值為：
 $51.66 - 44.27 = 7.39(\text{分})$
- (五)由於p-值小於任何一個顯著水準，檢定達顯著，本研究檢定結果確實足以支持該學者所認為學習成就上存在城鄉差異的看法。

高
點
·
高
上

【版權所有，重製必究！】