

97 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

一、何謂標準化測驗？台灣現行基本學力測驗是不是屬於標準化測驗？為什麼？

【擬答】：

(一)標準化測驗 (Standardized Test)：

測驗在編製程序與實施程序均由專家擔任。測驗題目編製經由試題分析選擇而來，尚須建立常模、信度、效度，而測驗實施標準化方式，如施測步驟、計分與結果解釋均須依指導手冊進行。智力測驗與性向測驗均屬於標準化測驗。標準化測驗具備條件包括：

1. 一般標準化測驗均有由原始分數轉換來的相對地位量數常模，用來解釋測驗分數的依據，如魏氏智力測驗其平均數 100，標準差 15。
2. 標準化測驗的內容必須經過修正的程序，每一題目均需經過詳細檢驗，其次經過統計方法計算每一試題的鑑別度與難度，刪除不佳試題；上述三種測驗均需經過試題分析過程。
3. 標準化測驗實施程序均需標準化，舉凡說明用語、實施過程、測驗時間、答案用紙均需在指導說明手冊說明，施測者遵照實施；魏氏、比西或中華等個別化智力量表均有指導說明手冊提供實施之用。
4. 計分方式要經過標準化，事先準備計分鑰，而且各分測驗及總分計算也應事先規定。

(二) 1. 台灣現行基本學力測驗係由教育部委託臺灣師大心理與教育測驗研究發展中心設計編製標準化測驗。

2. 「國中基本學力測驗」屬於標準化測驗，由測驗研究的專業機構，約請課程、教育心理、測驗統計學者、學科專家、中學資深優秀教師共同參與研究發展工作，按照測驗編製的標準化程序，分析各科教學目標和教材內容，建構基本學力指標，訂定命題藍圖，依據命題原則和要點，審慎撰擬試題，再經試題審查評鑑，進行預試、試題分析等實徵性研究過程，檢驗試題的難度和鑑別力，擇優汰劣，依試題性能存入題庫中。基測的試題都先經過預試和篩選，然後放入題庫，在闈內抽題和組卷。基測題庫中的每道試題都透過「試題反應理論」(Item Response Theory, 簡稱 IRT) 的估計程序，將試題難度連結在同一把尺上，所以每道題目的難度值是已知的。

二、同一群體在不同測驗的表現通常會因試題難度、鑑別度、題數等因素，而呈現出不同的測量結果，為了方便測驗結果的解釋或使用，測驗分數多改以標準分數或量尺分數呈現，使各個測驗的平均數(M)與標準差(SD)均相同。假設某項綜合測驗包含閱讀(M=48, SD=8)、數學(M=33, SD=12)及科學(M=58, SD=10)等三科測驗，各科測驗並轉換成平均數 50、標準差 10 的量尺分數，最後加總三項測的量尺分數代表考生在該項綜合測驗的表現。試問：(一)量尺分數與原始分數之平均數與標準差的變化意涵為何？(二)若某校以此綜合測驗結果遴選資優班學生，原始分數和量尺分數的遴選結果有何不同？

【擬答】：

(一) 1. 量尺分數：所有考生在這一科的表現，也就是「多少考生答對多少題」這樣整體的統計資料所換算出來的。這是國際上大型標準測驗(例如托福)通用的方式，因為具有和群體比較的意義，好像一把用來測「量」個人表現的「尺」，所以稱為「量尺分數」。量尺分數之平均數與標準差不會因為測驗類型不同而改變。

2. 原始分數：在古典測驗理論中客觀測驗的評分方式，以學生在每一試題之答對題數，再乘以每題分數得到總分，答錯者給予 0 分；其計分的本質都沒變，都是以每位受試者在整份測驗之答對題數的總和(稱作原始分數(Observed scores))，即視為其能力估計值。原始分數之平均數與標準差會因為測驗類型不同而改變。

(二) 1. 原始分數的優點即是以受試者作答情形或答對題數直接作計算，簡單易懂，大部分人都明瞭計分方式；缺點即沒有考量測驗試題具有不同難度的事實，因而計分容易受到測驗難度

公職王歷屆試題 (97 地方政府特考)

的影響，而出現不公平、不精確的現象。原始分數測驗結果遴選資優班學生，可能會因為學生在某一科目分數拿取高分，導致測驗原始分數總分較高而被錄取，但未必是資優學生。

2. 標準分數不具有單位，優點即是可以顯示每位受試者在團體中的相對地位量數和排序大小，大多數的受試者尚容易理解該分數的意義；但其缺點即是容易受到極端分數 (outliers) 的影響，而有扭曲次數分配的形狀之嫌，對個別受試者的表現會有過度放大或過度縮小之影響。標準分數測驗結果遴選資優班學生，較能錄取學生在每一科目分數均為相對地位量數較高，較原始分數能正確篩選資優學生。

三、請簡要比較「瑞文氏(Raven's)智力測驗」與「魏氏(Wechsler)智力測驗」在智力評量的差異為何？

【擬答】：

(一)瑞文氏(Raven's)智力測驗

瑞文氏智力測驗係於 1938 年完成，主要是測量 Spearman 之普通能力 (G 因素)，測驗題目是一組矩陣，或是排成行列的圖行元素，但是矩陣中某一部分被移除，受測者的作業必須從選項中挑選被移除遺漏部分。屬於團體智力測驗。

(二)魏氏(Wechsler)智力測驗

美國紐約 Bellevue 醫院精神病科主任魏克斯勒 (D. Wechsler) 於 1939 年編製一套成人智力測驗，以解決傳統智力測驗過於重視速度與學科內容，較不受成人忠誠合作。屬於個人智力測驗。

(三)智力評量的差異

1. 魏氏(Wechsler)智力測驗為個別測驗：

- (1) 意義：同一時間內，只能實施一個人的測驗，採面對面實施。如比西量表與魏氏兒童力量表均屬之。
- (2) 優點為主試者與受試者建立親善關係，深入瞭解受試者反應，在教育上具有診斷功能。
- (3) 缺點為實施不易，費時、費事，難以大量採用。主試者需受專業訓練與涵養才可擔任。

2. 瑞文氏智力測驗為團體測驗：

- (1) 意義：同一時間內，可以同時實施許多人的測驗。如陸軍普通分類測驗 (AGCT)、期中末考、瑞文氏非文字推理測驗。
- (2) 優點包括經濟簡單，花費較少，短時間內蒐集到大量資料。測驗內容較精簡，大都選擇題型，評分較為客觀，加以容易計分與容易解釋測驗結果。適用簡便施測者不需要專業訓練或練習，只要能熟悉測驗題型，依指導說明手冊進行即可擔任主試者。
- (3) 缺點包括受試者行為不易觀察與控制，可能會有疲勞、情緒、焦慮等影響測驗結果因素產生。缺乏個別測驗的親善關係與臨場感，難以維持受試者的合作與興趣。受試者反應侷限於選擇或其他標準項目形式，往往影響測驗的功能，不像個別測驗具有彈性與適應性。

四、七隻老鼠跑迷津抵達終點的秒數如下：13、11、89、24、21、31、17。

(一)請問此研究中的秒數屬於名義變項、次序變項、等距變項或是比率變項？請說明你的理由。

(二)請計算這組數據的全距與中位數。

(三)請問全距、標準差、四分差三者中，何者最宜作為這組資料的離散量數？為什麼？(5分)

(四)為了使研究結果更穩定，研究者又讓另外 8 隻老鼠跑迷津，結果發現 15 隻老鼠的數據所得的全距與中位數與 7 隻時一樣，而平均數變小。請問這組資料的偏態係數較可能為正值或負值？請說明你的理由。

【擬答】：

(一)1. 秒數屬於比率變項。

2. 「秒數」不僅能標示事物與事物間不同特質或類別而指定的數字，亦能指出大小次序 (等

公職王歷屆試題 (97 地方政府特考)

級順序)、多寡、優劣、高低,尚可以數值方式表示二事物差別之大小的變數,還可以表示相對比例大小。比率變數最重要條件是具有絕對零點(Absolute Zero)與有倍數關係,具備四則運算中加、減、乘、除運算。

(二) 1. 全距 $R = M - m = 89 - 11 = 78$ 。

2. 排序 11、13、17、21、24、31、89

$$Md(\text{項次}) = \frac{N}{2} + \frac{1}{2} = \frac{7}{2} + \frac{1}{2} = 4 \therefore \text{中位數} = 21$$

(三) 四分差最宜作為這組資料的離散量數。因為四分位差(Quartile Deviation)克服全距、標準差易受極端值影響,利用將一群數值資料分成四等分,即 Q_1 、 Q_2 、 Q_3 等三個分割點,其中 Q_1 為第一個四分位數占全體人數前 25%。四分位差指群體中,中間 50% 人的全距之半。屬於變異量數定義公式 $QD = \frac{Q_3 - Q_1}{2}$ 。

(四) 偏態係數可能是負值。因為當偏態係數是負值時為負偏分配,其平均數會小於中位數。

五、某研究者欲瞭解女性與男性的工作行為是否不同。他自勞動年齡人口中隨機抽取一個樣本,並依性別與工作類型分類,獲得下表統計人數:

工作類型	女性	男性
熱衷工作者	20	41
工作狂者	32	37
未受僱者	43	52
懈怠工作者	24	27
不滿意工作者	37	30

你會以何種方式檢驗工作類型是否有性別差異?你無需計算考驗統計值,但須清楚說明你的虛無假設與對立假設,如何計算該考驗統計值及如何評估統計顯著性。

【擬答】:

(一) 卡方考驗百分比同質性考驗。

$$(二) H_0: P_1 = P_2$$

$$H_1: P_1 \neq P_2$$

$$(三) \chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

(四) χ^2 值與臨界值 $\chi_{.975(4)}^2$ 比較, χ^2 值 $>$ 臨界值 $\chi_{.975(4)}^2$ 則落入拒絕區,亦即有充分證據支持性別與工作類型有顯著差異; χ^2 值 $<$ 臨界值 $\chi_{.975(4)}^2$ 則落入接受區,亦即無充分證據支持性別與工作類型無顯著差異。