

## 100年公務人員特種考試身心障礙人員考試試題

等別：四等考試

類科：教育行政

科目：教育測驗與統計概要

一、(一)實作評量的評分規準 (scoring rubric)，分為整體式 (holistic) 和分析式 (analytic) 兩種方式，請說明兩者的意義及適用時機。(二)簡述發展評分規準的流程。

【擬答】：

(一)意義

1. 整體式評定法 (Global Rating)：整體式評定法時每一個作品都當作一個整體相互比較，而不對作業細項進行分析判斷，適用於作品數量較多、評定時間有限制、作品本身不易細分評分。可以分成下列三種方式：作品等第量表、等第排列法、心像比較法 (歐滄和，民93)。適用情境包括快速瞭解學生整體表現或相對地位、所測特質單純，適合評斷簡單的成果或表現、計分速度比提供明確回饋訊息重要、計分規準容易瞭解，計分快速，適合總結性評量及大規模施測。(張麗麗，2002)
2. 分析式評定法：分析式評定法 (Analytic Rating) 將觀察行為或作品細分成幾項作業，逐一對各細項作業進行評分，然後加總細項得分，此評分方式較易客觀性與學習診斷價值。可以分成下列三種方式：項目檢核表 (Checklists) 評定量表 (Rating Scales)、評分規程 (Scoring Rubric)。適用情境包括教師擬定教學計畫、與學生溝通重要學習結果，提供學生明確且清楚的預期方向、評斷學生長處與弱點，具有診斷功能，提供明確的回饋訊息、測量複雜特質、適合課堂形成性評量、評量長時間的成長與改變。

(二)發展評分規準的流程：

1. 蒐集學生相關作品。
2. 將學生作品分為不同程度的組別 (如：高、中、低三組)，並分析各組的表現特質。
3. 對表現特質進行層面的分類，並寫下分類的原因。此階段的目的是在建立教師與學生共通的語言，學生透過參與，能對規準及成功的標準有更深一層的瞭解。
4. 以清楚明確的語言界定層面。
5. 列出每一層面不同等級行為表現的描述語，根據各程度的描述語，編製檢核表或等級評量表。
6. 選取各個層面不同等級具有代表性的樣本作品。
7. 修訂與改進，試用訂定的標準，若無法有效描述不同程度的表現，可在過程中不斷修訂與改進。

二、在其他條件相同的情形下，你認為下列作法會提高或是降低測驗結果的信度？請說明理由。

(一)以投射人格測驗取代結構式人格測驗。

(二)以資優班學生取代常態編班的學生，作答一份數學推理測驗。

(三)以低於平均能力的學生，取代中等及中等以上程度的學生，作答一份語文診斷測驗。

(四)以KR-20取代複本信度，估計一般成就測驗 (涵蓋內容廣泛) 的信度。

(五)以奇數題、偶數題方式，估計限時成就測驗之折半信度，取代以延後複本方式估計信度。

【擬答】：

- (一)以投射人格測驗取代結構式人格測驗會降低測驗結果的信度；投射人格測驗在解釋上不具客觀性，實施亦不方便，評鑑者對結果解釋也容易加以自己主觀看法，導致測驗信度與效度較低。
- (二)以資優班學生取代常態編班的學生，作答一份數學推理測驗會降低測驗結果的信度。因為在其他條件相等下，團體變異性愈大，其信度愈高；從相關係數來看，當團體內個別差異愈大，測驗分數分散範圍愈大，其相關係數也愈大，其測驗信度亦較高。
- (三)以低於平均能力的學生，取代中等及中等以上程度的學生，作答一份語文診斷測驗會降低測驗結果的信度；因為大部分診斷錯誤題目都較少，診斷測驗結果信度較低。

## 公職王歷屆試題 (100 身心障礙特考)

(四)以 KR-20 取代複本信度，估計一般成就測驗（涵蓋內容廣泛）的信度會降低測驗結果的信度；因為 KR-20 反應內容取樣誤差與內容取樣異質性誤差，會較複本信度只反應內容取樣誤差產生較多誤差。

(五)以奇數題、偶數題方式，估計限時成就測驗之折半信度，取代以延後複本方式估計信度會提高測驗結果的信度；內部一致性方法如折半信度不適合速度測驗信度的考驗，因為藉這些方法所估計的信度，會有假性的高相關現象。

三、T分數是常態化的標準分數，其平均數是50，標準差是10。若現在欲將一份非常態化的成就測驗分數資料轉換成常態化T分數，應該如何進行？請詳細說明轉換步驟。

【擬答】：

常態化標準分數適用時機為二個分數分配形狀很明顯不同，為了使不同分配型態分數可以直接比較，就必須將原始分數轉換為符合常態化分配的標準分數。詳細說明轉換步驟：

(一)計算已歸類次數之平均數與標準差。

(二)將已歸類每組之組中點減去 $\bar{X}$ 除以 $S$ ，求出各組中點的Z分數。

(三)將各組次數自下向上累積計算，列出累積次數（cf），以便調整各組組中點。

(四)計算至每組組中點累積次數（cf 中點），其方法為 $\frac{f}{2} + F$ （亦即每組次數之一半加該組下限以下的累積次數）。

(五)求出各組累積次數的累積百分比。

(六)將各個值當成百分等級，查標準常態分配表，將各概率化為Z值。

(七)將常態化後Z分數直線轉成各種標準常態分數。

【參閱謝廣全（民82）與林清山（民86）修改】

四、百分等級運用在常模參照測驗的解釋是非常普遍重要，若某群體原始分數分配接近常態分配，則將原始分數化為百分等級後，試回答下列各問題：

(一)中間百分等級的差異（如 PR=41 到 PR=50）及兩極端百分等級的差異（如 PR=1 到 PR=10 或 PR=90 到 PR=99），何者較能精確反應原始分數間之真正差異？請簡單說明原因。

(二)若將每個人的原始分數化為標準分數 $z$ ，試問每個受試所對應的百分等級有何變化？

(三)若將(二)中的 $z$ 乘以10再加上50，試問群體的平均數與標準差變成何值？如此得到的量尺分數其名稱為何？每位受試的相對地位有何改變？

【擬答】：

(一)百分等級量尺單位並不相等，是一種次序量尺，靠近常態分配中央原始分數單位小、二端單位大；相較於中間部位，在分數分配的極端部位之百分等級的差距應給予較大加權，接近中間部位小差距統通常影響較小。

(二)若將每個人的原始分數化為標準分數 $Z$ ，每個受試所對應的百分等級不會變化，因為相對地位量數不變、不會影響在團體中贏過多少百分比。

(三)1. 平均數=50，標準差=10。

2. T分數。

3. 不會改變。直線轉換並未改變原始分數之分配情形；原量表與新量表分配型態相同均為常態分配，將原量表轉換成標準常態分數的方法稱之。

五、(一)說明下列四種狀況，研究者犯了第一或第二類型錯誤（type I, type II error），還是決策正確。

	$H_0$	$H_1$	$\mu$ 的真值	研究者的決策
a.	$\mu = 0$	$\mu \neq 0$	0	拒絕 $H_0$
b.	$\mu = 0$	$\mu \neq 0$	5	拒絕 $H_0$
c.	$\mu = 0$	$\mu \neq 0$	0	不能拒絕 $H_0$

d.  $\mu = 0$        $\mu \neq 0$       -3      不能拒絕 $H_0$

(二)如果將  $\alpha$  由 0.05 變更為 0.01，那麼對第一類型、第二類型錯誤有何影響？

(三)樣本加大，對第一類型、第二類型錯誤以及統計考驗力的影響又是如何？

【擬答】：

(一) a. 犯第一類型錯誤。因為  $\alpha = P(\text{拒絕 } H_0 \mid H_0 \text{ 為真})$ 。

b. 決策正確。因為  $\text{Power} = 1 - \beta = P(\text{拒絕 } H_0 \mid H_0 \text{ 為假})$

c. 決策正確。 $1 - \alpha = P[\text{接受 } H_0 \mid H_0 \text{ 為真}]$

d. 犯第二類型錯誤。 $\beta = P[\text{接受 } H_0 \mid H_0 \text{ 為假}]$

(二)當其他條件不變情況下， $\alpha$ 與 $\beta$ 具有互相消長關係，如 $\alpha$ 值降低 $\beta$ 值增加，相對統計考驗會減少。此題條件會降低第一類型與增加第二類型錯誤。

(三)如果樣本大小  $n$  增加，則第一類型、第二類型錯誤會同時變小、統計考驗力會增加。

公  
職  
王