

101年公務人員特種考試身心障礙人員考試試題

等 別：三等考試

類 科：教育行政

科 目：教育測驗與統計

一、一般測驗介紹的資訊中，常提供信度係數或測量標準誤來說明測驗的可靠程度。就這兩項指標而言，那一項指標在不同樣本呈現較佳的穩定性？那一項指標比較適合作為不同測驗可靠程度的直接比較？為什麼？

【擬答】：

(一)信度：是指相同受試者在不同時間，使用相同測驗測量（或複本測驗測量多次）或在不同情境下測量，所得結果一致性。如重測信度指同一份測驗在不同時間或不同情境下，針對相同群體學生前後重複測量二次，根據二次分數所求得其相關係數，即稱再測信度係數，又稱穩定係數 (Coefficient of Stability)、外在信度 (External reliability)。

(二)測量標準誤：二項測驗分數的原始分數，如果參照相同常模表解釋結果，使可轉換成相同量尺的標準分數互相比較，但其差異需達統計顯著差異，始具有意義。必須使用測量標準誤的差異標準誤 (SEdiff)。

(三)原因

1. 當同一人在同一份測驗前後二次測驗比較，同一人在同一測驗之不同分測驗比較或不同的人在同一測驗比較，用以判斷上述情況測驗分數的差異，是否達統計顯著水準，就需使用差異標準誤 (SE_{diff})。

2. 差異標準誤使用可以檢驗差異分數顯著性，藉以判斷該差異是代表特質上真正差別，或純為測量誤差結果。

二、請說明常模參照測驗的各類信度作法，以及各類作法考量的主要測量誤差來源。

【擬答】：

常模參照測驗信度類型估計信度方法有許多種，皆在說明測量誤差的不同來源。但多有共同解釋意義，信度愈高，即誤差愈小；信度愈低，誤差愈大。最常用信度估計方法有四種：①重測信度；②複本信度；③內部一致性信度；④評分者信度。敘述如下：

(一)重測信度：

1. 指同一份測驗在不同時間針對相同學生前後重複測量二次，根據二次分數所求得其相關係數，即稱重測信度係數，又稱穩定係數。

2. 重測信度基本假設測驗所測驗潛在特質，在短時間內不隨時間消逝而變化。其誤差來源——來自不同時間下測量所造成誤差。適用感覺辨別測驗、動作技能測驗。

(二)複本信度：

1. 複本信度愈高，即表示二份測驗測量同一特質和屬性的程度愈高。複本信度的測量誤差來源，主要是來自試題抽樣產生的誤差。

2. 實施複本測驗方式有二種：

(1) 同一時間連續實施：反映出測驗內容造成誤差有多少（亦即試題抽樣誤差），無法反應出受試者本身狀況在施測期間（即時間誤差），此種複本信度稱等值係數。

(2) 間隔一段時間實施：可反映出題抽樣誤差亦可反應出時間誤差，此種信度稱穩定等值係數。其使用複本方式的再測方式，因此是估計測驗最嚴謹、最佳的方法，因為它已包括測驗分數中所有可能誤差變異。最適用於性向測驗。

(三)內部一致性信度：

上述二種測驗，均需對受試者實施二次測驗，不僅增加編製的負擔，同時受試者合作意願降低、動機減少、疲勞與厭煩增加，均會影響施測的品質。因此學者嘗試一次施測產生正確估計信度結果，會較受歡迎；此方法包括：①折半信度；②庫李方法；③ α 係數方法；④霍意特變異數分析法。說明如下：

1. 折半信度 (Split-half Reliability)：折半方法採用單獨一次測驗結果，利用隨機方

公職王歷屆試題 (101 身心障礙特考)

式將試題分成二半，求二半分數的相關，稱折半相關或內部一致性係數 (Coefficient of Internal Consistency)。反應試題取樣誤差。

2. 庫李方法 (K-R 方法) 意義：係由庫德 (G. F. Kuder) 和李查遜 (M. W. Richardson) 於 1937 年在《心理計量期刊》發表「測驗信度估計理論」，設計一種估計測驗項內部的一致性信度的方法。僅適用答案非對即錯的二分計分測驗。反應試題取樣誤差與試題異質性誤差。
3. α 係數方法：克朗巴賀 (Cronbach) 在 1951 年在《心理統計》學報發表「測驗的 α 和內在結構」，當測驗採用多重計分方式，如態度量表一李克特式量表 (「非常同意」5 分、「同意」4 分、「無意見」3 分、「不同意」2 分、「非常不同意」1 分)、個別智力測驗的加權計分、行為評定量表，皆採用 α 係數方法。簡稱阿爾發係數 (Coefficient Alpha)。反應試題取樣誤差與試題異質性誤差。

$$\text{公式：} \alpha = \frac{n}{n-1} \left(1 - \frac{\sum S_i^2}{S_x^2} \right)$$

4. 霍意特變異數分析法 (郭生玉，民 95)：霍意特 (Hoyt) 在 1941 年提出變異數分析法，亦為估計測驗內部一致性方法：

$$\text{公式：} r_H = 1 - \frac{MS_{err}}{MS_{ind}}$$

(四) 評分者信度：

係指從測驗卷中隨機抽取一些樣本，由不同的評分者或同一個評分者在不同的時間，對每一份試卷評分，在測驗過程中觀察、記錄、評分、計分等各方面相互間的一致性而言，然後根據每份測驗卷評分的兩個分數計算其相關係數，即得評分者信度。評分者信度愈高，表示評分者間評分愈一致，反應評分者間誤差越小。評分者信度又分為評分者間評分者信度 (斯皮曼等級相關、Kendall 和諧係數、積差相關) 與評分者內評分者信度 (同質性信度係數)。

三、請簡要說明「簡單隨機取樣 (simple random sampling)」的執行過程為何？優點為何？缺點為何？

【擬答】：

(一) 簡單隨機抽樣

1. 意義：最典型、最基本的機率抽樣的方式，它是其他抽樣方式的基礎，合乎抽樣概率均等的原則與獨立原則。亦即母體中每一個樣本被抽中的機率均相同，而且被抽到的單位並不會影響其他單位被抽到的機率。簡單隨機抽樣分為放回式與不放回式抽樣兩種，大規模調查研究都採用不放回式抽樣。
2. 執行過程：必須針對母體建立適當抽樣架構，使母體中任何一個體均有同樣被抽取機率，給予母體名冊中每一個體一個號碼，然後使用抽籤式、亂數表方式進行抽取樣本。

(二) 優點：

1. 合乎抽樣均等的原則與獨立原則，抽取樣本較具代表性，不會因調查研究人員主觀偏好而選取特定的樣本。
2. 最簡單易行最常用抽樣方式。
3. 簡單隨機抽樣的誤差很容易計算，且其抽樣誤差相對於其他機率抽樣較小。

(三) 缺點：

1. 抽樣單位愈大，完全隨機抽樣愈不易進行且費時費成本。完整的研究族群名單不易獲得，或取得很費時，費力且費錢。
2. 因隨著機遇而變動，缺少安定性，致同一母群體的樣本所表現的情況可能彼此有差異。
3. 樣本分配較分散，行政作業較不易；樣本代表性恐有不足 (尤其當樣本點差異大時或重要性不同時)。
4. 當其間變異甚大，即分散度很大或具有歪度 (skewness) 時，若採用簡單隨機抽樣，則可能造成分散在兩端的樣本將不被抽中或抽中太多，如此抽出的樣本不具高度代表性，反

而使估計誤差過大。

四、隨著教育改革浪潮與評量方式多元化的需求，實作評量／真實評量 (performance/authentic assessment) 已成為一種重要的評量方式，但有學者認為 (Wiggins, G.)，測驗要有較高的真實性 (authenticity) 並非是新近的論點。請以一個實作評量／真實評量的設計來說明如何達成認知目標中「應用」與「綜合」之評量。

【擬答】：

(一)實作評量編製步驟

Linn 與 Gronlund (1995)、Linn 與 Miller(2005)均提出發展實作評量作業重點的六項建議，包括：

1. 陳述複雜認知技巧與學生表現高學習目標的學習結果；
2. 選擇或發展能代表重要學習結果之內容與技巧作業；
3. 將干擾評量目的與提高實作作業困難度的技能減到最低。
4. 提供學生必要的鷹架以便理解作業與期待，以及注意完成作業所需的先備知識與技巧，方能提供富挑戰性的實作作業；
5. 研擬具體明確的作業指導語，讓學生清晰了解如何實作；
6. 清晰說明實作所期望的表現方式或內涵，並告知評量標準。

(二)實作評量／真實評量的設計

勤洗手保健康--洗手檢核表

姓名：	班級：	總分			
養成勤洗手好習慣，可以預防腹瀉、感冒與腸道寄生蟲等疾病，為了了解你洗手步驟是否正確，請你做一次洗手動作，並由小組長檢核。					
檢核項目	行為特質	是	否	不確定	
濕	1 兩手全弄濕				
搓	2-1 將手擦上肥皂、洗手液				
	2-2 雙手搓揉，包括手心、手背、指縫、虎口、指甲等，應搓洗至有泡泡				
	2-3 至少搓洗 20 秒以上				
沖	3 用清水將雙手加以沖洗				
捧	4 以雙手捧水將水龍頭加以沖淋				
擦	5 用手紙巾或乾淨毛巾將手擦乾，用擦手巾包覆水龍頭予以關閉，且不可以他人共用手紙巾或乾淨毛巾。				
註記： 有該項行為就在「是」打「✓」，無在「否」打「✓」，無法確定在「不確定」打「✓」。					
分享：					
小組長簽名：		教師簽名：		日期：	

【本文摘自：李坤崇 (2012)。實作評量編製步驟及實例。教育研究月刊，214，92-109。】

五、建民老師最近想進行一項行動研究，他自己編製一份國小四年級學生適用的閱讀態度量表，他同時準備以學生最近三篇作文的成績和最近一次國語文段考成績作為關連變項。就這三個變項，建民老師在研究計畫中擬提供最適切而重要的一項信度係數。最適合作為學生學習態度量表、作文、和國語文成就的信度係數分別為何？簡單說明支持你決定的理由。

【擬答】：

(一)學習態度量表

1. 內部一致性信度： α 係數方法。
2. 支持理由：克朗巴賀 (Cronbach) 在 1951 年在《心理統計學報》發表「測驗的 α 和內

公職王歷屆試題 (101 身心障礙特考)

在結構」， α 係數反應受試者在同一份測驗各題目間作答一致性，或題目作答與測驗總分間一致性。當測驗採用多重記分方式，如態度量表—李克特式量表（「非常同意」5分、「同意」4分、「無意見」3分、「不同意」2分、「非常不同意」1分）、個別智力測驗的加權計分、行為評定量表，皆採用 α 係數方法。簡稱阿爾發係數(Coefficient Alpha)。

(二)作文

1. 評分者信度。

2. 支持理由：對於一些無法進行完全客觀記分的測量工具，例如：投射測驗、舞蹈測驗、道德判斷測驗、教學評量（採用觀察法、判斷法、評定量表法）、人格測驗（羅夏克墨漬測驗、主題統覺測驗）、創造性思考測驗及申論題、作文題等必然涉及評分者主觀的判斷，會導致評分者誤差存在，故須對評分者間的變異加以衡量。

(三)國語文成就測驗

1. 複本同時信度。

2. 當同一種測驗有兩種以上的複本測驗（測驗性質類似在內容、試題格式、長度、難度、鑑別度均一致）時，可將此兩種複本分別實施於同一群受試者，再根據此兩份測驗分數計算其相關係數，即得複本同時信度；反映出測驗內容造成誤差有多少（亦即試題抽樣誤差），無法反應出受試者本身狀況在施測期間（即時間誤差），此種複本信度稱等值係數。因此測驗目的在推論個人某一學科具有知識程度，則只需考量是否具備適當的複本信度。

公職王