

107 年公務人員特種考試司法人員考試試題

考試別：司法人員

等 別：三等考試

類科組：監獄官

科 目：心理測驗

一、心理測驗中，通常在進行項目分析（item analysis）時，有那些指標（index）可以用來幫助研究者或使用者檢驗題目的好壞？（20 分）

【擬答】

(一)項目分析的意涵

項目分析主要是針對測驗題目是否能反應測驗所關心的目的，一部份可以增加測驗本身的信效度，另一方面，可以提升題目的鑑別能力，尋找出優良題目，以形成題庫，方便之後使用。

項目分析可以初步分成量的分析與質的分析，量的分析可以透過難度、區辨性（鑑別力）、誘答分析等概念，主要會採用一些統計的方法去分析資料，並說明其意涵；質的分析則是透過題目的內容以及形式，採用雙向細目表的方式去瞭解或是藉由編制題目的專家進行審題。

(二)項目分析的指標

1. 質的分析

(1)內容主題專家

主要是有三大類：

一類是針對測驗主題有相當程度熟悉的專家。

另一類是受測對象也可能是熟悉該主題的專家。

最後一類是心理計量的專家，他們可以提供專業的見解，預先找出有疑問的題目，進行篩選。

(2)題目描述的方式

在題目語意表達上，應注意一些特性，如避免否定的描述、避免雙重否定的敘述、注意頻率副詞的使用、避免負向情緒字眼等，舉例來說，頻率副詞的使用指的是在面對 Likert 五點量表進行作答時，因為作答方式是從非常不同意到非常同意，假若在題目描述還加上頻率，會造成作答困難，例如：我不常參加社團活動，假若作答者回答有點同意，那是代表他經常參加還是很少參加，假若將題目更改成我參加社團活動，則受測者作答理解較為簡單。

2. 量的分析

在量的分析指標上，常見的有難度、鑑別力、誘答、社會期許值和因素分析等，例如：在使用 Likert 量表時，受測者的回答沒有對錯，因此常會受到社會期許值影響，造成在單一題目上，大多數受測者會有傾向同一側回答的表現，例如：你喜歡唸書嗎？大多數人回答同意或是非常同意時，這樣在分析受測者的答題反應，會出現同意一邊的高選擇人數。

(三)例證

在編制問卷的過程，必須先透過預試以及正式施測的結果來篩選題目，像是一般的成就測驗，在初步編擬問題完之後，會先嘗試找少量的受測者來回答問卷，透過受測者回答結果，初步瞭解測驗編制狀況，若無太大問題，經過正式施測之後，將受測者結果，由小排到大來看前百分之二十七的高分組通過率，與後百分之二十七的低分組通過率之平均，來

反應題目難度。難度的標準，一般偏好 0.5 左右的題目居多，這樣施測的結果，鑑別力通常是最大的，而且信度相對是較高。

二、「項目反應理論」(Item Response Theory, IRT) 中不變性 (invariance) 具有那兩個特性？請加以說明。(20 分)

【擬答】

(一) IRT 的不變性意涵

試題反應模式不像古典的真實分數模式，它是可能作假的模式(falsifiable models)；換句話說，任何一種試題反應模式都有可能適用或不適用於某份特殊的測驗資料，亦即模式可能會不當地預測或解釋資料。因此，在應用試題反應理論時，我們必須先估計出模式與考生的參數值外，還需要考驗模式與資料間的適合度(model-data fit)。當某一種試題反應模式適用於某種測驗資料時，一些試題反應理論的基本特性也會跟著產生，例如：不變性。

不變性，係指不受測驗種類和不同族群考生的影響，造成參數估計結果發生變化。

(二) 不變性的兩種特性

首先，從不同組的試題估計而得的考生能力估計值，除了測量誤差外，不會受所使用的測驗種類的影響，亦即，它是試題獨立(item-independent)的能力估計值；其次，從不同族群的考生估計而得的試題參數估計值，除了測量誤差外，亦不受參與測驗的考生族群的影響，亦即，它是樣本獨立(sample-independent)的試題參數估計值。上述兩種特性，在試題反應理論中叫作「不變性」(invariant)，這些不變性是從把試題的訊息(information)考慮在能力估計的過程中，把考生能力的訊息考慮在試題參數估計的過程中而得。

(三) 結論

為克服古典測驗理論的缺失，才有現代測驗理論的誕生。現代測驗理論主要是以試題反應理論為理論架構，依據強勢假設而來，其理論發展較晚，所用公式複雜難懂。雖然目前應用的測驗資料範圍有限，但由於其立論與假設嚴謹，仍深受學者喜愛。現代測驗理論具有下列特點，這些特點正是古典測驗理論所缺乏。

1. 所採用的試題參數為不受樣本影響(sample-free)的指標；亦即，參數的獲得，不因樣本的不同而不同。
2. 能針對每位受試者，提供不同的測量誤差指標，而非單一相同的測量標準誤，因此能夠精確推估受試者的能力估計值。
3. 可以藉由同質性試題組成的分測驗，測量並估計出受試者個人的能力，不受測驗的影響(test-free)。此外，對於不同受試者間的分數亦可進行有意義的比較。
4. 提出試題訊息量的概念及測驗訊息量的概念，作為評定某個試題或整份測驗的測量準確性。古典測驗理論中，用以評定內部一致性的信度，有漸被測驗訊息量取代之趨勢。
5. 現代測驗理論同時考慮受試者的反應組型與試題參數等特性，故對於原始的得分相同的受試者，往往給予不同的能力估計值。
6. 現代測驗理論所採用的適合度考驗值，可以提供考驗模式與資料間之適合度、受試者的反應是否為非尋常等參考指標。

參考資料來源：<http://140.119.115.26/bitstream/140.119/15124/1/259.pdf>

三、柯典獄長轉介一位十三歲的受刑人陳小龍給羅心理師做智力評估，羅心理師採用 WISC-IV 做智力評估。智力評估結果，陳小龍的語言智商 (Verbal IQ, VIQ) 是 85，操作智商 (Performance IQ, PIQ) 是 94。柯典獄長想知道陳小龍的 VIQ 與 PIQ 兩者之間是否達到顯著

差異，請問羅心理師如何回覆（註：VIQ 的信度係數是.97；PIQ 的信度係數是.93）？另請問，陳小龍的整體智力應屬於那一個水準？（30分）

【擬答】

(一)心理師的回覆依據

1. 定義

測量標準誤常用於彙整多種測驗結果，來確認特定特質測量後的意義，近代常用不同測驗分數表示測驗結果，但欲以多種測驗同時表示特定特質，則可由多個測驗分數的測量標準誤合併後來表達，並且進行解釋。例如：性向測驗從過去的特定性向測驗發展至現在的多元性向測驗，不同性向測驗的分數，假若有需要合併解釋，就會使用到差異分數的標準誤。

2. 公式

$$SEM_{diff} = \sqrt{(SEM_1)^2 + (SEM_2)^2} = SD\sqrt{2 - r_{11} - r_{22}}$$

(二)結果計算過程和說明

$$SEM_{diff} = 15 * 0.32 = 4.89$$

二分測驗的實得分數相差為 9，透過 SEM_{diff} 的估算，可以知到實得分數必須超過 $1.96 * 4.8 = 9.41$ 以上，二分測驗的得分才有顯著差異存在。

(三)陳小龍的整體智力

各分測驗的原始分數可按年齡組別換算成量表分數。各分測驗的量表分數相加，可得語文、作業及全量表量表分數，再查表換算得三項離差智商（平均數為 100、標準差為 15 的標準分數），以測知並解釋受試整體認知能力的水準。

故 VIQ 的量表分數約為 2.84，PIQ 的量表分數約為 4.24，故 FSIQ 的量表分數為 7.08，依據量表分數與百分等級對照表，可以查出整體智力大約為 85 左右，依據 WISC 的分數解釋為中下的智力表現。

參 考 資 料 來 源 :

<http://zh.wikiteamwork.wikia.com/wiki/WISC:%E9%87%8F%E8%A1%A8%E5%88%86%E6%95%B8%E8%88%87%E7%99%BE%E5%88%86%E7%AD%89%E7%B4%9A%E5%B0%8D%E7%85%A7%E8%A1%A8>

四、心理測驗中有關效度之概念，會出現一個現象稱為「效度縮減」（validity shrinkage），請加說明，並敘述在那些情況會出現「效度縮減」之現象？（30分）

【擬答】

(一)效度複核的觀念

所謂的效度複核是指對已建立的測驗效度進行覆核的過程，主要是採用與原來用以建立測驗的樣本之不同樣本來再行測驗，然後比較兩次測驗結果，以檢查效度的準確性。

(二)效度縮減的觀念及成因

而效度縮減在效度複核研究中是常見現象，測驗對相關效標的預測度，新受試樣本的正确性比原始測試樣本低，即因樣本不同，而產生的效度變小的現象。會發生效度縮減的原因，主要是測驗的內容與效標有很高的相關時以及機率性產生誤差，特別是當選取項目是隨機選取的。

(三)發生的情況常見有三種

1. sample size：由於最初的樣本所獲得虛假的高效度可能是取樣誤差的累積結果，因此樣本越小（取樣的誤差也越大），效度數值也越大。
2. item pool 多而後來保留的項目少則易產生隨機誤差。

3. 在選 item 時，是否有經過事前設定的假設，若沒有的話，則 validity shrinkage 的值會增大。

參考資料來源：

http://www.epc.ntnu.edu.tw/download.php?filename=188_20c3d2d6.pdf&dir=archive&title=%E6%85%8B%E5%BA%A6%E9%87%8F%E8%A1%A8%E7%B7%A8%E8%A3%BD