

110 年公務人員高等考試三級考試試題

等 別：三等考試

類 科：教育行政

科 目：教育測驗與統計

思澤老師

甲、申論題部分：

一、請說明以下措施對測驗信度或測驗效度造成的影響，並說明你的理由？（25 分）

(一)增加相同品質的測驗題數。

(二)加入更多符合測驗目標特質的內容。

(三)加強評分員培訓，確保其都能依評分規準評分。

(四)提高測驗情境的標準化，例如：相同的指導語、光線、溫度……。

(五)使用與測驗目標特質較相似的評量工具或行為作為效標。

【擬答】

【破題關鍵】

1. 《考題難易》：★★★

2. 《解題關鍵》：信效度一直以來是高考教測統的命題重點，近來更是著重對測驗觀念理解程度，所以對準備良好的同學來說，應可以拿到不錯的分數

【使用學說】 由真分數理論出發的信效度觀念，著重在影響信效度因素的論述

(一)增相同品質的題目題數

若我們假定測驗的品質、試題所測量的特質、及受測者的性質都保持一樣，則試題題數與信度的關係可用下列公式來表示：

$$R_{xx} = \frac{nr_{xx}}{1 + (n-1)r_{xx}}$$

R_{xx} 表題數為原測驗 n 倍的測驗的信度係數； r_{xx} 表原測驗的信度係數；

若有一 20 題的測驗，其信度係數為 0.60，若把試題增加成 100 題（為原來的 5 倍），則其信度係數便提高為：

$$R_{xx} = \frac{5(0.60)}{1 + (5-1)(0.60)} = 0.88$$

(二)增加更多符合測驗目標特質的內容

內容效度是指某測驗的試題能否適切地測到此測驗編製目的上所要測的內容或行為領域。當增加更多符合測驗目標特質的內容時，內容效度會提升。

主要是內容效度的確立，採用邏輯的分析方法，判斷每個題目是否符合教材內容及教學目標。如果測驗中的整體題目很能代表教材內容以及所預期的行為改變，而無其他無關因素的影響（比如理解、閱讀能力），則表示測驗有好的內容效度。因為這種分析是依靠邏輯的分析與合理的判斷。因此又稱為邏輯效度或合理效度。

(三)加強評分員培訓，確保都能依評分規準評分

為了大量施測和機械計分而設計的團體測驗通常具有高度標準化的施測及計分程序，所以主試者或計分者所造成的誤差變異少得可以忽略不計。但是某些類型的測驗—例如著名的創造力測驗和採用投射技術的性格測驗—相當依賴評分者的判斷，因缺乏客觀的評分標準，在計分時容易受到評分者個人主觀及價值判斷的影響。所以加強評分員培訓，確保都能依評分規準評分，這樣能有效避開評分者的主觀介入，提升評分者信度。

(四)提高測驗情境的標準化，例如：相同指導語、光線、溫度……

從測量的誤差來看，信度則是指測驗的分數反映出真實分數的程度，或有無誤差的程度。簡言之，信度是在估計測量的誤差多寡。信度也可以說是在測驗的實得分數中，有多少比率是由於所測特質的真實分數差異造成，以及有多少比率是由於誤差產生所造成。如果測驗當中的誤差成份高，信度就低；誤差成份低則信度就高。所以提高測驗情境的標準化，例如：相同指導語、光線、溫度等，將能夠降低誤差對施測結果的影響，測驗信度能夠提升，也因此一個優良的測驗為何需要標準化的條件，正是反應本題的措施能有效提升信度。

(五)使用與測驗目標特質較相似的評量工具或行為作為效標

效標關連效度係指用一測驗在一特定情境下預測某一個體行為的有效程度。為了證明這個有效程度，研究者把這測驗的分數與外在效標（criterion）求其相關程度。因此效標的如何選擇對效標關連效度的影響很重要，所謂效標是指對測驗所欲測的特質的一種直接且獨立的測量，足以用來考驗測驗結果真確性的外在標準。故又稱為實證效度（empirical validity），意指這類效度的驗證是依據證據資料進行的。在教育、臨床、工商企業、政府機構、和軍隊中，測驗常被用作為了解或預測一個人的實際能力或心理狀況的主要參考資料，並據以作成各種重要決策，因此，更需要有良好的、實證的效度研究結果來支持這些決策。

參考來源：

郭生玉著。教育測驗與評量。精華出版。

二、請說明申論題評分方式中的整體式評分（holistic scoring）與分析式評分（analytic scoring）的特徵及適用時機？（10 分）並詳細說明申論題的評分原則？（15 分）

【擬答】

【破題關鍵】

1. 《考題難易》：★★★★

2. 《解題關鍵》：命題原則在公職考試命題比率並不高，但是只要一出題，同學常常會有措手不及的狀況，同學可能再面對本題的回答，會有望題興嘆的感受。

【使用學說】評量命題原則

(一)申論題性質與命題原則

1. 性質

申論題依據允許學生自由反應的程度，可分為申論題與限制反應兩種：申論題類型：毫無任何限制，具有極大自由可陳述自身觀點的特性。限制反應題類型：局部限制作答範圍，要求提出明確而精簡的答案。

2. 命題原則

- (1)專用於測量高層次認知能力的命題方式，採取多題短答的方式取代少題長答的申論題。
- (2)命題應盡量與學習結果相互關聯，且敘述力求明確使學生理解問題要求。
- (3)不允許學生擁有選擇其中幾道試題作答，並給予學生充分作答時間。

(二)整體式與分析式評分

1. 整體式評分

評估學生整體表現，如明列達成優、佳、好、普、劣等級的質性描述，包括對報告主題、演說論點、演說方式、台風的表現等。以學生上台口頭報告為例，評估學生整體表現為劣時，表示學生未能描述報告問題以及重要性，主題不明確，且沒有適宜的結論，內容

難以理解，沒有準備且缺乏組織，僅以最基本的答案回答觀眾問題，或未能解答。整題式評分的好處在於可快速、簡單評分；缺點在於老師或同儕不易評分，如學生清楚說明理由，但無使用視覺輔助工具，究竟要給學生「優」，還是給「普」呢？此外，學生也不易瞭解自己若拿到4分，代表的是哪項表現不好？

2. 分析式評分

設立重點目標，列為若干標準，再依表現指標，分別給予等第評量。例如：以書面報告為例，針對統計分析的作業，從評量標準與表現指標建立列聯表，在評量標準利用統計的教學目標，像是建立假設、公式、計算過程、查表、統計結論為一個軸向；另一個表現目標的軸向，則分列出得分標準像是學生可以拿到A，是因為在建立假設可以描述正確題意及虛無假設撰寫正確詳盡。分析式評分的好處在於學生可清楚知道自己的表現及可努力的方向；缺點在於不易評分，因此不宜列過多標準(4-8項即可)

(三) 申論題評分原則

1. 預先擬定評分標準與依據。
2. 避免干擾公平性之因素。最好在一段時間內評完所有試卷。
3. 一次評閱一道試題，等所有試卷的同一題評完才能繼續下一題的評分。
4. 讓兩位以上評分者獨立評閱每一試題。
5. 使用匿名者作答及匿名評分原則，排除不必要主觀因素與月暈效應影響。

參考來源：

郭生玉著。教育測驗與評量。精華出版。

<http://iweb.ntnu.edu.tw/aa/gpaweb/pic/Rubrics%20QA.pdf>

志光×保成×學儒

快速考取

全方位智慧
服務系統



線上.線下 給您 **最強大的支援**

激推！考生必看

公職王影音頻道



名師精析各科目考試重點、口面試準備技巧、上榜生經驗分享等全方位影音資訊。

線上模擬測驗



運用海量題目，協助考生訓練解題速度，檢視學習成效並及時修正弱點。

歷屆試題下載



收錄各年度國考試題及解題，讓考生練習考古題時更易突破盲點，找到解題關鍵。

國考申論加分



各考試領域專業文章分析解讀趨勢動態，協助考生加深各科目的答題內涵。

三、下表為某測驗針對不同性別、年級、科系類型進行分層隨機抽樣 (stratified random sampling) 的樣本人數，請依此表回答下列問題：

科系類型	低年級 (一、二年級)		高年級 (三、四年級)		小計
	男	女	男	女	
藝術型	392	600	161	343	1,496
實用型	1,000	275	700	160	2,135
研究型	636	565	346	203	1,750
商業型	285	516	272	672	1,745
社會型	317	1,400	159	605	2,481
事務型	66	98	50	179	393
小計	2,696	3,454	1,688	2,162	10,000

(一)何謂分層隨機抽樣 (stratified random sampling)？它與簡單隨機抽樣 (simple random sampling) 有何不同？(10 分)

(二)上表分層隨機抽樣的細格中，人數最多及人數最少的群體在各變項上的類別特徵是什麼？他們分別占總抽樣人數的比例有多少？(10 分)

(三)如果要了解「樣本中各分層群體的樣本人數比例與母群中各分層群體的人數比例是否符合」，應如何進行分析？不必實際計算，只需完整敘述你的分析程序及方法？(10 分)

【擬答】

【破題關鍵】

1.《考題難易》：★★★★

2.《解題關鍵》：本題目針對教育統計的觀念，涉及抽樣觀念與卡方檢定，所以準備好的同學可以從教育統計的抽樣觀念以及卡方考驗順利回答

【使用學說】教育統計的抽樣觀念以及卡方檢定的觀念

(一)分層隨機抽樣，以及與簡單隨機抽樣的差異

1. 分層隨機抽樣是基於母群底冊的訊息，若發現母群體是由不同特徵的群體組成時，為了增加樣本的代表性，可將母群體依據「層內同質、層間異質」的原則進行分層，再將各層中依據比例、變異情況或抽樣成本等狀況考量，進行隨機抽取所需的有效樣本數。例如：研究全台灣使用 5G 的手機使用者比例，由於使用 5G 的費用較高，所以直接針對全台灣進行隨機抽樣的樣本，恐有代表性不足的問題，因此依據地區特性，將全台灣區分成北、中、南、東（含離島），共四個區域，在考量各區使用手機人數與全台灣使用手機總人數，以前述二者比例取得有代表性的樣本進行調查。
2. 簡單隨機抽樣與分層隨機抽樣的差異在於抽樣的對象不同，是針對整體母群進行還是先將母群依據特定特徵進行分層後，再依據研究關心的特定條件，進行各層隨機抽樣。

(二)

1. 人數最多的特徵類別使於社會型中低年級的女生；人數最少為事務型的高年級男性。
2. 社會型中低年級女性佔整體抽樣人數比例為 0.014；事務型高年級男性佔整體抽樣人數比例為 0.0005

(三)卡方檢定

1. 理由

(1)當依變項的資料為人數比例時，可以考慮由 Pearson(1900)首度利用卡方分配進行檢定

的統計技術來回答題意要求。

- (2)在進行卡方考驗時，另外需要注意因為題目的列聯表是採用三個向度進行資料整理，對於向度超過二個以上，卡方檢定會出現問題，因此會建議將各個分層向度一一區分出來，這樣即可以採用多次單一向度的適合度考驗進行考驗題意的要求。

2. 檢定程序

- (1)適合度考驗(test of goodness of fit)

就某單一變項考驗觀察次數（樣本人數）與期望次數（母群體）是否相符合。

- (2)假設建立

$$H_0: f_0 = f_e$$

$$H_1: f_0 \neq f_e$$

- (3) $\alpha=0.05$ ，雙側考驗，透過查表找出 $X^2_{.95(df)}$ 的臨界值

- (4)依據卡方分配的公式計算檢定量， $X^2 = \sum (f_o - f_e)^2 / f_e$

- (5)比較檢定量與臨界值的大小進行統計決策並說明統計結論

參考來源：

郭生玉著。教育測驗與評量。精華出版。

林清山著。心理與教育統計。東華出版

第一名的輔考實力 志光.保成.學儒

文化行政/教育行政 10大全方位課程

從基礎到精通，一系列專業輔導課程
幫助您快速上榜，成功考取公職。

實力養成班

提早準備
提高上榜機會

總複習班

考前觀念統整
法條時事最新補充

正規班

課程最完整
奠定考取實力

成效卓越 讀書會

學員有口皆碑
最具成效的方式

高分作文班

名師指導
拆解高分答題技巧

全國線上 模擬考

藉由測驗了解
各科分數及總排名落點

申論作答課

針對法科、學科
之區別深入探討

能力指標 檢測系統

線上測驗同時診斷
各科目章節強弱

題庫班

教您以最快速度
解出正確答案

30線上 練題批閱

在家也能好好寫申論
線上批閱更彈性

(各班輔導規劃略有不同，部分課程需自費加選，詳情請洽各班服務櫃台)

四、下表是某班學生英文科段考成績、每週補習時數，以及參加托福考試 (TOEFL) 的成績，請依此表回答下列問題：

學生編號	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16
補習時數	10	10	10	10	14	14	14	14	18	18	18	18	22	22	22	22
段考成績	75	70	60	70	90	75	90	80	80	90	90	90	70	80	90	75
托福成績	490	480	475	465	540	520	560	525	580	600	610	605	565	580	590	580

(一)若要比較「這四種補習時數的英文段考成績是否有顯著差異」，應使用何種統計檢定？請敘述檢定程序，並說明這種檢定方法有何基本假設？(10 分)

(二)老師以托福考試成績 (Y) 為依變項，使用兩種迴歸模式進行分析，得到結果如下表。請說明此兩種迴歸模式分析的結果提供了那些訊息？(10 分)

	迴歸模式的公式	預測變項	模式 F 檢定值	R ²
模式一	$Y=243.8+3.8X_1$	段考成績 (X_1)	16.8**	0.54
模式二	$Y=241.5+2.4X_1+7.1X_2$	段考成績 (X_1)、補習成績 (X_2)	69.4**	0.91

**表示 $p<0.01$

【擬答】

破題關鍵】

1. 《考題難易》：★★★★★

2. 《解題關鍵》：第一小題為教育統計的實驗設計觀念，過往比較少出現在公職命題，一般常見的考法是採用 SPSS 報表，要求同學能夠閱讀並解釋統計結果，這一次出單因子受試者間設計的假定與程序，算是比較少見的考法，所以一般同學比較容易忽略，不過第二小題是教育統計的迴歸概念，則是重要的統計概念過往也是公職重要的必考觀念，所以準備好的同學應該可以因應得宜。

【使用學說】教育統計中實驗設計與相關迴歸的概念，結合研究方法觀念切入

(一)單因子變異數分析與假定

1. 理由

依據題意，自變項為「補習時數」涉及四個水準 (10、14、18、22 小時)，依變項為「英文段考成績」，這樣吻合單因子受試者間設計。

2. 程序

(1)假設各組母群平均數相同，即 H_0 和 H_1

(2)計算各變異數來源的量與自由度，如 SS_t , SS_b , SS_w ; df_t , df_b , df_w 。

(3)計算變異數的不偏估計值，即「均方」($MS = \frac{SS}{df}$)

(4)進行 F 檢定， $F=MS_B/MS_W$ 再與臨界 F 值進行比較，看是否落入拒絕區。

(5)列出「變異數分析摘要表」(summary table of analysis of variance)

例如：

ANOVA 摘要表				
SV	SS	df	MS	F
組間(SS_b)	30.8	3	10.267	5.48*
組內(SS_w)	30.0	16	1.875	
總變異(SS_t)	60.8	19		

$$*P < .05$$

$$F_{.95(3,16)} = 3.24$$

(6) 計算關聯強度(綜合考驗顯著後)

關聯強度(strength of association)：若變異數分析後，發現 F 值達顯著，即表示自變項與依變項有關係，可以下列公式來表示其關聯強度：

$$\hat{\omega}^2 = \frac{SS_b - (k-1)MS_w}{SS_t + MS_w}$$

(7) 將統計結果解釋為統計結論

3. 假定

(1) 常態性(normality)

樣本所來自的母群必須都是常態分配。

(2) 觀察值獨立性(independent observation)

因為樣本是經過隨機抽樣獲得，故每一筆從母群中所抽出的樣本觀察值，彼此間不能有相關存在。

(3) 加成性(additivity)

總變異正好可分割為數個變異來源的總和。如 $SS_t = SS_b + SS_w$ 。

(4) 同質性(homogeneity)

進行假設考驗之前，必須以 H_0 為真的前提，進行考驗，所以各組樣本所來自母群的變

異數必須相同，即 $\sigma_a^2 = \sigma_2^2 = \sigma_3^2 = \dots = \sigma_k^2$ 。

(二) 迴歸

依據題意，統計分析採用二個模式進行：

1. 模式一利用段考成績對托福成績進行預測

由報表中呈現 F 值顯著，表示由樣本估算的模式一迴歸式可以推論到母群也存在模式一的迴歸線，但是仍有 1% 下錯結論。報表中的 R^2 表示利用段考成績的變異可以解釋托福成績變異的 54%。

2. 模式二利用段考成績、補習時數對托福成績進行預測由報表中呈現 F 值顯著，表示由樣本估算的模式二迴歸式可以推論到母群也存在模式二的迴歸線，但是仍有 1% 下錯結論。報表中的 R^2 表示利用段考成績的變異和補習時數的變異可以共同解釋托福成績變異的 91%。

3. 模式一與模式二的解釋力差異將模式一、二的 R^2 相減 ($91\% - 54\% = 37\%$)，可以瞭解補習時數多納入迴歸式時，可以增加模式的多少解釋力，藉以說明補習時數的重要性。但是這樣的方式常常受陷於研究方法論的問題無法解決，所以實務研究，已經不太建議使用。

參考來源：

郭生玉著。教育測驗與評量。精華出版。

林清山著。心理與教育統計。東華出版。