

112 年公務人員高等考試三級考試試題

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、某教育學者認為，COVID-19 疫情期間，大多數國民中小學校皆採行線上教學來代替實體教學。但是，線上教學的學習成效深受各種因素的影響與干擾，因此，他認為有必要重新制訂線上教學時，學習成績是否達到「及格」的通過標準。請問：你會建議他該如何設定此通過標準？請至少列舉五項步驟的建議。(25 分)

【解題關鍵】

1. 《考題難易》★★★★★
2. 《破題關鍵》從編制評量過程，探討編制評量標準，所以可以從教學目標需求進行討論
3. 《使用法條》or《使用學說》教育測驗～實務評量
4. 《命中特區》教育測驗 第十章 實務評量

【擬答】

(一)評量觀點轉變

近來新型態的教學評量方式的興起，主要是對傳統評量方法的反思以及引進了新的知識理論。

新評量方法受到重視的原因有四點：

1. 過去標準化測驗評量只測驗最低能力，導致教師為提高學生成績而訓練學生練習測驗題目，而不鼓勵學生參與更有用的學習活動。
2. 對現有的心理評量理論不滿意，因為只強調量化和客觀的評量。此觀點受到社會建構論以及質的評量和後現代思想的論點的反對。反對原因是認為太過強調客觀性會降低對學生反應的解釋品質，並且也會忽略對個別學生的特殊考慮。
3. 認知心理學和建構論的影響認為學生的學習是主動建構，因此影響教師中心的教學轉成學生中心的教學。並且重視或得高層次思考能力的學習取代關於事實、知識和基本技能的學習。因此評量應該用來測量應用知識解決問題的能力。
4. 教育的主要目標之一是為了促進社會公平。而傳統標準化測驗的目的之一是為了區分學生。但是因為其功能為篩選學生，所以一但被認為不成功者，測驗的使用很可能會被認為是不公平的，由其是對來自不同文化的學生。在這種情況下，標準化測驗對於促進社會公平是種阻礙，因此出現反對此種測驗而支持另類評量方法的觀點。

(二)線上評量的精熟標準設定，五項步驟的建議

1. 了解學習目標：確定該教育學者所設定的學習目標，這些目標應該符合課程大綱或課程目標。例如，學習者在特定主題或技能方面應該達到什麼程度的理解或能力。
2. 定義評估方法：選擇適當的評估方法，以確定學生是否達到學習目標。這可能包括測驗、作業、專案或課堂參與等方式。要確保評估方法能夠準確地測量學生的學習成果。
3. 設定成績標準：根據評估方法，像是利用雙向細目表，設定明確的成績標準。這些標準可以基於百分比、等級或其他評分系統來表示。例如，學生需要達到多少分數或獲得什麼等級才能通過該評估。
4. 考慮彈性和公平性：考慮到線上教學的多樣性和學生的不同情況，應該考慮設立彈性和公平的成績標準。這可能包括給予學生額外的時間或資源，提供不同的評估選項，或者考慮到學生的背景和環境因素。

5. 測試和修訂：在實施成績標準之前，進行測試和修訂。可以在教學過程中使用樣本測試，觀察學生的表現，並評估成績標準的有效性。根據反饋和結果，進行必要的調整和修訂，以確保通過標準是公平和可行的。

這些步驟可以幫助該教育學者設定一個合適的線上教學學習成績通過標準，以確保學生能夠達到預期的學習目標。然而，需要注意的是，標準的設定應該根據特定情境和學習環境進行調整，並且持續的評估和改進是必要的。

(三)除前述建議外，也應考慮線上評量特色要包含新的教學評量特點：

1. 重視高層次能力的評量：

重視程序性知識而非陳述性知識。評量的方法採用建構反應的題目，要求學生統整所學知識回答問題。

2. 強調使用真實的問題：

評量所使用的問題和實際生活的問題一樣，比如讓醫學生在臨床情境中診斷病情，而非使用紙筆測驗測量。另外新評量方法也強調直接評量，比如在評量寫作能力時，就直接要求完成一篇文章，而非要求回答文法或是字彙的相關問題。

3. 採用多元評量方法及多向度分數：

新式評量方法有各種不同的評量方式來獲得結果。此特點可提高評量的信度和效度，且可適應學生個別差異的需要，讓不同能力的學生都有表現機會。多向度的評量分數可提供豐富的瞭解學生學習成長與進步的資料，也有益於教師做教學計劃。比如檔案評量中除了學生代表作品的多樣性之外，還有教師以及家長的評語以及學生的自我評量和反省等都是多向度分數的一種。

4. 重視歷程與結果：

傳統評量只強調學習的結果，因此新的評量方法則是反過來重視學習的歷程。通常以觀察或是問問題的方式評量過程。當然，為了配合教學目標的需要，也可以強調學習的結果，或是兩者並重。

(四)根據學習評量目的，把學習評量類型分為下列三種：

1. 促進學習的評量(Assessment For Learning, 簡稱 AFL)是一種形成性評量(課堂作業、課堂測驗、單元作業、單元測驗)，教師使用評量方法與策略，來獲取學生學習回饋，掌握學生學習情況與需求，給予學生適當回饋，以及作為調整教學方法與策略的參考依據，進而幫助學生學習，加速達成學習目標。

2. 學習結果的評量(Assessment Of Learning, 簡稱 AOL)是一種總結性評量，在教學活動告一段落時，對學習結果的總體性評量，例如月考、期中考、期末考、畢業考、能力檢測等，評量內容包括學生學習成果、教學目標達成程度、教學方法適切性，以及課程效能等。

3. 評量即學習(Assessment As Learning, 簡稱 AAL)，也可稱為歷程性評量，是在學習活動中，設計評量任務來引導學生學習，使學生為了完成評量任務，而願意自發地學習，以提升自我能力。此外，評量任務要能引發學生後設認知能力，使學生從評量任務中發現自己不足，自省和反思學習情況，進而激發更好的學習表現。簡言之，評量即學習就是以評量任務為主要學習活動的教學設計。例如，智慧型教室就是一種比較容易實施 AAL 學習即評量的教學情境。

引用來源：<https://www.habook.com/zh-tw/academic.php?act=view&id=59>

公職王歷屆試題 (112 高考三級)

二、某市教育局舉辦一年一度的國小學生才藝競賽，並聘請兩位表演藝術專家針對 10 名進入總決賽的小學生才藝表演進行客觀的評分。這兩位專家針對這 10 名學生的評定名次如下表，請問：這兩位專家的評分間有無關聯性或一致性存在？(25 分)

(查表的 t 值臨界值為： $t_{.975,8} = 2.306$)

| 學生 | A | B | C | D | E | F | G | H | I | J |
|-----|---|---|---|---|---|---|---|----|---|----|
| 甲專家 | 2 | 8 | 5 | 7 | 4 | 6 | 1 | 9 | 3 | 10 |
| 乙專家 | 5 | 4 | 1 | 8 | 3 | 7 | 6 | 10 | 2 | 9 |

【解題關鍵】

1. 《考題難易》★★★★★
2. 《破題關鍵》因為名次不存在相同組，所以直接採用 Pearson 積差相關估計相關結果，再進行 t 檢定考驗相關係數是否顯著存在。
3. 《使用法條》or 《使用學說》相關～Spearman 等級相關在相同名次組較少的情況下，可採用 Pearson 積差相關去估計相關結果
4. 《命中特區》教測統 第五章 相關迴歸～其他相關

【擬答】

(一) Spearman 的等級相關～當評分者對不同受評者作品採評定名次 (屬於次序量尺) 時

$$\text{公式： } \rho = 1 - \frac{6 \sum D^2}{N(N^2 - 1)}$$

但是依題意受評者人數多，加上評定名次結果在評分者間並未出現相同，所以可以採用 Pearson 的積差相關估計相關結果。

(二) 利用 Pearson 的積差相關～依題意計算結果

$$r = \frac{\sum Z_X Z_Y}{N} = 0.5636$$

(三) 並利用考驗積差相關的檢定～t 檢定

假設： $H_0 : \rho = 0$ $H_1 : \rho \neq 0$

假設： $H_0 : \rho = 0$

$$\text{公式： } T.S. = (r) / \sqrt{(1 - r^2) / (N - 2)} = 1.93$$

結果： $T.S. = 1.93 < t_{.975,8} = 2.306$ ；故拒絕 H_1 ，接受 H_0

統計結論：故二位評分者的評分並不存在顯著一致性

志光×保成×學儒 做你的學習靠山



快速考取班

掌握考取節奏
安心學習無負擔

公職輔考第一品牌 只給你最好的

- 學費省很大**
全年課程不間斷，一次繳清學費輔導至考取
- 課程最完整**
完整課程循環，基礎班→正規班→專題課→總複習..等
- 上榜賺獎金**
報名考取班第一年考取同職等考試，頒發獎學金
- 加選最超值**
輔導期間加選其它科目增加考試機會，加選另享專案優惠
- 公約有保障**
考取班簽訂公約，保障您的權利與義務至考取為止

雙料金榜 高考公職社會工作師 普考社會行政

楊○安 考取班 一年考取

因為考試科目大多為第一次接觸，所以希望透過補習班的課程減少自己盲目鑽研的時間，會選擇考取班是因為可以持續學習與複習，銜接比較沒有空檔。

優異考取 普考教育行政

陳○宇 考取班 一年考取

因為家人過去有使用過志光.保成.學儒系列補習班的課程，效用甚佳，其中考取班亦可以輔導直到考取為止，作為努力的後盾再適合不過。

三、某教育學者想知道某校六年級學生的平均身高，他自該校隨機抽取 9 名學生為樣本，並測得其身高分別為 140、142、140、143、139、140、141、142、141 公分（已計算出這 9 名學生身高的 $M=140.89$ ， $SD=1.26$ ），請問：你會推估出該校六年級學生的平均身高之 95%信賴區間（查閱 t 分配表，臨界值為 ± 2.306 ）為何？請寫出計算過程並說明你的答案。（25 分）

【解題關鍵】

- 《考題難易》★★★
- 《破題關鍵》過去考古題常常出現相關命題觀念，所以同學應該可以順利取得好分數
- 《使用法條》or《使用學說》估計～區間估計
- 《命中特區》教測統 第三章 推論統計～區間估計

【擬答】

(一)區間估計的內涵

根據點估計值、抽樣分配及機率原理去推估母群參數在某一區間或範圍的統計方法，即稱為「區間估計」。

(二)區間估計的公式

$$P(t_{\frac{\alpha}{2}} \leq t \leq t_{1-\frac{\alpha}{2}}) = 1-\alpha \dots\dots\dots t \text{ 分配的機率區間}$$

$$P(t_{\frac{\alpha}{2}} \leq \frac{\bar{X}-\mu}{\hat{s}/\sqrt{n}} \leq t_{1-\frac{\alpha}{2}}) = 1-\alpha \dots\dots\dots (1-\alpha) \text{ 信心水準的信賴區間}$$

$$P(\bar{X} + t_{\frac{\alpha}{2}} \cdot \frac{\hat{s}}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{1-\frac{\alpha}{2}} \cdot \frac{\hat{s}}{\sqrt{n}}) = 1-\alpha \dots \text{母群平均數 } \mu \text{ 的信賴區間}$$

(三)計算過程和結果

$$P[140.89 + (-2.306) \cdot \frac{1.26}{\sqrt{9}} \leq \mu \leq 140.89 + (2.306) \cdot \frac{1.26}{\sqrt{9}}] = 95\%$$

四統計結論

在 95%信心水準下，該校六年級學生平均身高的信賴區間位在 139.92~141.86 (公分)。

文教前三巨星

一致推薦 志光 × 保成 × 學儒

全國狀元/榜眼

連過三榜

甘○藜

高考教育行政 榜眼
普考教育行政 榜眼
初等教育行政 狀元

打從下定決心準備考試時起，我就沒想過要自讀，為了節省走彎路的时间，選擇年年都在準備公職考試的補習班肯定比我來得有經驗，能幫助我避免可能遇到的障礙。

全國榜眼

雙料金榜

沈○君

高考教育行政 榜眼
普考教育行政

志光、保成、學儒是輔考的權威，基本上師資篩選與課程安排讓人很放心，而且除了正規班，也有總複習班做考前加強，對於非本科系沒經驗與背景知識的我十分有幫助。

全國榜眼/探花

雙料金榜

張○瑄

高考文化行政 榜眼
普考文化行政 探花

個人推薦志光、保成、學儒系統的文化行政教材，書籍或講義觀念整理清晰，對於建立知識架構頗有助益，也謝謝補習班讓我在考試時擁有如此堅實的輔考資源，著實感謝！

全國榜眼

九個月考取

邱○勛

高考文化行政 榜眼

我覺得補習的好處是能讓老師告訴你重要的考點在哪裡，也會跟你說申論作答的技巧。另外我覺得補習班的講義都是精要地綜合各家的精華，可以省去一些看其他書的時間。

四、請解釋下列專有名詞：(每小題 5 分，共 25 分)

- (一) 參數估計不變性 (invariance of parameter estimation)
- (二) 差異試題功能 (differential item functioning)
- (三) 可能值 (plausible values)
- (四) 效度量尺 (validity scale)
- (五) 試題特徵曲線 (item characteristic curve)

【解題關鍵】

1. 《考題難易》★★★★
2. 《破題關鍵》本題最大的問題應該是不容易處理完答題內容，加上 IRT 相關概念考了三題，並且涉及測驗計分解釋可能需要注意現象，所以難度較高。
3. 《使用法條》or 《使用學說》IRT 的不變性、DIF、ICC 等和實務測驗
4. 《命中特區》教育測驗第一章 測驗緒論、第八章 項目分析

【擬答】

(一) 參數估計不變性

現代測量理論主要是以項目反應理論為架構。項目反應理論處理資料的單位是受試在單個「測量項目」上的反應，與古典真分數理論以整個測驗為處理對象不同。其中 IRT 與 CTT (古典測驗理論) 的主要差異之一在於樣本和試題獨立，故延伸出參數估計不變性和不定性 (Lord, 1980)，不變性可以區分成下列二種，分別是：

1. 項目參數的不變性

從不同族群受試估計而得的試題參數估計值，除了測量誤差外，不受參與受試族群的影響，亦即，它是樣本獨立的 (sample-independence) 的參數估計值。

2. 能力參數不變性

受試的能力估計不因測驗題目不同而改變。亦即，它是試題獨立 (item-independence) 的能力估計值。如有改變則只是選擇量尺的原點及單位不同而已，而非能力的改變。

(二) 差異試題功能

測驗的公平性一直是社會大眾關注的問題，雖然在古典測驗理論也談試題偏差的診斷和補救，但項目反應理論對此問題所提出的理論基礎和考驗架構，卻是相當完整、周延和嚴謹的。來自不同族群，但能力相同的個人，如果答對特定題的機率不同，則該試題便顯現出 DIF 現象。DIF 的操作型定義如下：「若試題特徵曲線對不同族群不同的話，該試題顯現出 DIF 現象。」依據 DIF 的操作型定義，只要比較兩個 (多個) 族群在試題參數曲線上的差異，就可以判定是否有此現象。

常用的方法有下列三種：

1. 比較試題特徵曲線的參數
2. 比較介於試題特徵曲線間的面積
3. 比較反應模式與資料間的適合度

(三) 可能值

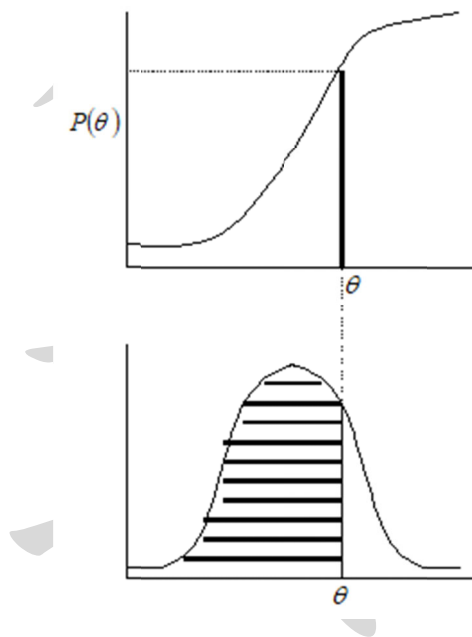
在心理測驗中，「可能值 (plausible values)」是指根據統計模型估計得出的潛在變項或能力的多個可能值。心理測驗通常基於某個理論或模型，該模型假設被評估者的能力無法直接觀察，因此使用可能值來估計該能力的範圍。這些可能值是統計推斷的結果，它們反映了該能力的多個可能水平，以提供更全面和精確的測量。像是 PISA (2009, 2012, 2015) 將閱讀素養界定為實現個人目標、增長知識、發展個人潛能以及參與社會活動，而理解、運用、評鑑、省思與參與文本的能力。在能力估計就是採用可能值進行評估，但是在 PISA 的能力估計常常會發生解釋錯誤的問題。

(四) 效度量尺

有效度量尺 (validity scale) 是一個心理測驗中的指標，用於評估測驗的效度，即測驗是否測量到它所聲稱要測量的概念或目標。有效度量尺通常是測驗中的特殊項目或量表，它們旨在檢測受試者是否在測驗中存在不真實的回答、模擬行為、無關的反應或其他測量失真。有效度量尺可用於檢查測驗結果的可信度和信度，並確保測驗結果的解釋是可靠和有效的。有效度量尺的使用有助於識別潛在的問題，例如回答不真實的受試者、測驗的解讀可能受到的干擾或測驗設計上的缺陷。通過分析有效度量尺的分數或結果，評估者可以確保測驗結果的有效性，並提供有關受試者的信賴程度和測量結果的可靠性的相關信息。像是 MMPI-2 編製於 1989 年採用實徵效標計分法進行，同 MMPI 一樣也是一個在國際上用途廣泛的人格測驗量表。該表共包括 567 個自我報告形式的題目，分基礎量表、內容量表和附加量表三大類，其中基礎量表包括有 10 個臨床量表和 7 個效度量尺。適用於 18 歲~70 歲的受測者，教育程度在國小畢業以上。在進行測驗前，主試人必須熟悉測驗的全部材料 (包括量表的內容、簡介及指導語)，了解受測者的情況 (如受測者的理解力及身體狀況)。進行測驗的房間在亮度與溫度方面要適當，並且盡可能地安靜。MMPI 是第一個採用效度量尺觀念的測驗，用來測試受測者態度對施測結果的影響，此作法一直延續 MMPI II 和 MMPI-A 的版本，效度量尺中以 F 量尺對於裝病的偵測，表現最佳，K 量尺對於淡化問題的偵測效力在單獨使用下，表現不太好。

(五) 試題特徵曲線

項目特徵曲線是表示受測能力與項目答對概率的函數曲線。以 θ 表示考生或受測者的能力或潛在特質， $P_i(\theta)$ 表示能力或潛在特質為 θ 的考生或受測者，其答對第 i 試題或第 i 試題正確反應的機率。如果我們相信人類能力呈現常態分配，即大部分的人能力居中，能力強和能力弱的人數都較少；那麼對於一個試題（項目），我們將不難理解概率分布圖中，不同能力的人答對此項目的機率相當於常態分配的積分，也就是能力極低的人答對的可能近乎 0，而能力極高的人答對機率為 1。這個「常態分配的積分圖」呈現出的肩形曲線（ogive curve），即是理想的項目特徵曲線，而現代測量學者已經發展出不同的項目反應模式來描述 θ 與 $P_i(\theta)$ 的這種關係曲線圖。在現代測量理論中的項目特徵曲線，正是與常態分配積分曲線等同或相近的曲線圖。



王

感謝志光×保成×學儒 讓我們成為文教之星

近八年全國22狀元 19榜眼 23探花 與你相約明年上榜

| | | |
|-----------------------------|-----------------------------|-----------------------------|
| 狀元 111 高考教育行政 陳○兒 | 狀元 110 高考文化行政 何○綺 | 狀元 110 普考教育行政 吳○書 |
| 狀元 110 普考文化行政 韓○群 | 狀元 109 高考教育行政 施○豪 | 狀元 109 普考教育行政 黃○萍 |
| 狀元 109 普考文化行政 張○迎 | 狀元 108 普考文化行政 龔○雯 | 狀元 107 高考教育行政 林○吟 |
| 狀元 107 普考教育行政 林○吟 | 狀元 107 普考文化行政 邱○婷 | 狀元 106 高考教育行政 張○淇 |
| 狀元 106 普考教育行政 徐○瑛 | 狀元 106 普考文化行政 林○雅 | 狀元 105 高考教育行政 陳○坊 |
| 狀元 105 高考文化行政 楊○頤 | 狀元 105 普考教育行政 徐○容 | 狀元 105 普考文化行政 吳○璇 |
| 狀元 104 高考教育行政 羅○婷 | 狀元 104 高考文化行政 李○純 | 狀元 104 普考教育行政 劉○彥 |
| 狀元 104 普考文化行政 謝○齡 | KEEP FOR YOU | |

| | | |
|-------------------|-------------------|-------------------|
| 111 高考教育行政 榜眼 沈○君 | 105 高考文化行政 榜眼 吳○璇 | 107 普考教育行政 探花 陳○霖 |
| 111 普考教育行政 榜眼 甘○藜 | 105 普考文化行政 榜眼 沈 ○ | 107 普考文化行政 探花 吳○文 |
| 110 普考文化行政 榜眼 官○好 | 104 高考文化行政 榜眼 馮○騏 | 106 高考教育行政 探花 張○文 |
| 109 高考教育行政 榜眼 廖○雅 | 104 普考文化行政 榜眼 方○涵 | 106 高考文化行政 探花 白○瑄 |
| 109 高考文化行政 榜眼 張○瑄 | 110 高考教育行政 探花 莊○婷 | 106 普考教育行政 探花 張○文 |
| 109 普考文化行政 榜眼 陳○寧 | 110 高考文化行政 探花 韓○群 | 105 高考文化行政 探花 蔡○軒 |
| 108 高考文化行政 榜眼 彭○筠 | 110 普考教育行政 探花 黃○鈺 | 105 普考教育行政 探花 陳○坊 |
| 108 普考教育行政 榜眼 黃○萍 | 109 高考教育行政 探花 鄭○澤 | 105 普考文化行政 探花 楊○頤 |
| 108 普考文化行政 榜眼 彭○筠 | 109 普考文化行政 探花 張○瑄 | 104 高考教育行政 探花 陳○霖 |
| 107 高考文化行政 榜眼 邱○勛 | 108 高考教育行政 探花 高○吟 | 104 高考文化行政 探花 郭○冶 |
| 107 普考教育行政 榜眼 張○佑 | 108 高考文化行政 探花 許○謙 | 104 普考教育行政 探花 林○容 |
| 107 普考文化行政 榜眼 邱○勛 | 108 普考教育行政 探花 陳○宏 | 104 普考文化行政 探花 張○容 |
| 106 高考文化行政 榜眼 蔡○呈 | 108 普考文化行政 探花 邱○絹 | |
| 106 普考文化行政 榜眼 于○文 | 107 高考教育行政 探花 陳○霖 | |
| 105 高考教育行政 榜眼 徐○容 | 107 高考文化行政 探花 李○儀 | |



上榜席次 為你保留