

112 年特種考試地方政府公務人員考試試題

等 別：三等考試

類 科：教育行政

科 目：教育測驗與統計

教育測驗-梁昕老師解題

教育統計-王瑋老師解題

一、雖然目前國內有許多升學及就業考試，仍以古典測驗理論(classical test theory, CTT)作為編製、發展與分析試題的理論依據，然而卻因為該理論具有何種缺失而逐漸式微，進而被新興的試題反應理論(item response theory, IRT)取而代之?請說明之。(25 分)

【解題關鍵】

1. 《考題難易》★★★★☆
2. 《破題關鍵》試題反應理論的題目在最近這兩年的考試中頻繁出現，無論是申論題或名詞解釋均有其蹤跡，各位夥伴務必要勤加準備。而本題作答時，主要考點即在於試題反應理論對古典測驗理論的批判，以及針對這些批判，試題反應理論作何修正。
3. 《使用學說》試題反應理論基本概念

【擬答】

鑒於古典測量理論的諸多限制，洛德(F. Lord)於 1950 年以潛在特質理論(Latent Trait Theory)為基礎，發展出了試題反應理論(Item Response Theory)，以改進傳統測驗理論的缺失，以下將依其意涵、發展脈絡及其與古典測驗理論的差異等分別說明之：

(一)試題反應理論之意涵

1. 試題反應理論是現代測驗理論的重心，其特點在於以「機率」的概念來解釋受試者能力與題目反應間的關係，亦即將測驗結果透過數學模式的運算，推估受試者的基本能力。目前許多大型標準化測驗，例如：國家教育進展評量(NAEP)、國際學生評量(PISA)、國際數學與科學教育成就趨勢調查(TIMSS)等測驗均是以 IRT 理論為基礎發展而成。
2. 考生(examinee)在某一測驗試題上的表現情形，可由一組因素來加以預測或解釋，這組因素叫作潛在特質(latent traits)或能力(abilities)；而考生的表現情形與這組潛在特質間的關係，可透過一條連續性遞增的函數來加以詮釋，這個函數便叫作試題特徵曲線(item characteristic curve, ICC)。

(二)對古典測驗理論的批判

1. 參數估計具有受試樣本依賴性：古典測驗理論對於難度、鑑別度等參數指標的估計，會依受試者性質、能力的不同而變動，易造成估計時的不穩定。
2. 假定所有考生的測驗誤差均相等，不符現實：古典測驗理論假定同一份測驗具有相同的測量標準誤，以此估算測驗者得分信賴區間或進行分數的比較，此種假定是與現實不符，因為忽略了個別學生的作答反應及組型。
3. 對於考生能力的估計會因測驗改變而不同：古典測驗理論以單點估算個人能力，因此，當試題過難或過簡單時，受試者的測驗得分會出現變動，此時會造成對於學生真實能力的推估有所不同，會造成對學生能力的錯誤評估。由於受試能力參數會因受試題特徵之影響，因此，受試者必須接受相同或平行(parallel)的測驗題目，才能比較所測量能力之差異，所以較難進行能力間的相互比較。
4. 缺乏預測力：沒有提供一個受試者可能答對個別試題的概率，當測驗設計者希望預測一個或多個受試群體測驗分數的特徵，或是要為某些受試群體設計某種特性的測驗時，此

種訊息是必要的。

5. 過分依賴複本測驗：對於非複本但功能相近的測驗所測得之分數，無法獲得有意義的比較；此外，對於信度是建立於複本測量概念之基礎，而每個測驗不見得都能建立複本測驗。對於每位受試者多次接受同一測驗，而每次測驗又假設無相關，這樣的假設又與真實情境不符。

(三) 試題反應理論與古典測驗理論的差異

1. 試題反應理論所採用的試題參數 (item parameters) (如：難度、鑑別度及猜測度)，是一種不受樣本影響的指標。
2. 試題反應理論能夠針對每位受試者，提供個別差異的測量誤差指標，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. 試題反應理論可透過測驗等化的過程，對於受試者個人的能力估計，可排除測驗題項的影響，並且對於不同受試者間的分數，或同性質但不同測驗間的分數，亦可進行有意義的比較。
4. 試題反應理論提出以「試題訊息量」或「訊息函數」的概念，來作為評定某個試題或整份試卷的測量準確性，可對試題品質提供更多、更精確的判斷訊息。
5. 試題反應理論同時考慮受試者的反應組型與試題參數等特性，因此在估計個人能力時，除了能夠提供一個較精確的估計值外，對於原始得分相同的受試者，也往往給予不同的能力估計值，例如同樣是得 80 分的受測者，但若是能答對較難題目者，應該有更高的能力值。

志光×學儒×保成
為你絕佳助攻

5大衝刺課程

帶你直攻
高普考

NEW 階段複習課 圖像精要複習鍛鍊 把重要考點烙印腦中 心智圖圖解運用 透過圖解複習 破解考題陷阱 針對考點分析	測驗易點通 埋頭苦練 不如讓老師點通學習之路 常考題型知識強化 易錯題型觀念釐清	總複習 考點update! 時事修法update! 關鍵考點 最新考情 考前複習 短期密集
題庫班 各科名師專業訓練 審題神速、答題神準 讀書精熟+答題精準=快速上榜 題庫演練 精準教學 解題技巧	作文實戰班 作文學得好, 同時提升寫作能力與論述邏輯 高分寫作指引 強化論述深度 架構分層演練 新式作文教戰	

公職王歷屆試題 (112 地方特考)

二、某教育學者隨機抽樣 2485 名男女高中生，測得其對學校整體滿意度的數據，在進行某種統計分析方法後，獲得如下表的結果。請問：

(每小題 5 分，共 25 分)

- (一)該學者所使用的統計分析方法稱作什麼？請說明之。
- (二)該統計方法背後的基本假設是否滿足或成立？請說明之。
- (三)該統計方法的檢定值為何？請說明之。
- (四)若從 95% 的信賴區間角度來判斷，男女同學對學校整體滿意度之間的差異是否達到顯著程度？請說明之。
- (五)若從男女同學的平均數（一為 3.1830，另一為 3.3463）來看，其差異似乎很微小。那麼，是什麼因素影響你對上述(四)的結論？請說明之。

	性別	個數	平均數	標準差	平均數的標準誤
學校整體滿意度	男	887	3.1830	.74982	.02518
	女	1598	3.3463	.65226	.01632

	變異數相等的 Levene 檢定		平均數相等的 t 檢定							
	F 檢定	顯著性	t	自由度	顯著性 (雙尾)	平均差異	標準誤差異	差異的 95% 信賴區間		
								下界	上界	
學校整體滿意度	假設變異數相等	6.759	.009	-5.663	2483	.000	-.16330	.02883	-.21985	-.10676
	不假設變異數相等			-5.443	1627.312	.000	-.16330	.03000	-.22215	-.10446

【解題關鍵】

1. 《考題難易》★★☆☆☆
2. 《破題關鍵》

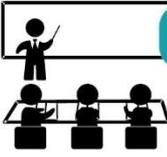
獨立樣本 t 檢定的報表解讀是一再強調的重點，並且去年原住民特考三等亦有非常類似的考題可以參考，是同學必須要拿下的分數。

【擬答】

- (一)比較男女高中生對學校整體滿意度平均數差異，此處採用獨立樣本 t 檢定。
- (二)獨立樣本 t 檢定的基本假設為兩組資料服從常態分配，但本題無此資訊；而變異數同質的假設此處不滿足，因為變異數相等的 Levene 檢定顯著($p\text{-value}=0.009 < 0.05$)，代表變異數異質。
- (三)在變異數異質下，t 檢定統計量為 -5.443，對應的自由度為 1627.312。並由 $p\text{-value} < 0.001$ ，可知男女高中生對學校整體滿意度平均數有顯著差異。
- (四)變異數異質下，兩組平均數差的信賴區間為 $[-0.22215, -0.10446]$ ，區間不包含 0，所以男女同學對學校整體滿意有顯著差異。
- (五)雖然兩組平均數的差僅有 0.1633，但因為兩組樣本數都非常大，所以在此平均數差異與標準差的條件下，兩組平均差異達到統計上顯著。

志光×學儒×保成

五大學習方式 上課超便利



現場面授

名師現場面對面
即時互動解答疑惑



直播教學

即時登入直播跟課
掌握進度免等待



視訊課程

手機APP預約上課
輔導期間 無限重覆看課



WIFI看課

專屬WIFI教室
讓你學習時間更彈性



在家學習

使用在家補課點數
即可在家複習上課
(以老師授權科目為主)

三、在常模參照測驗下，學者專家們採用的信度估計方法可分成四類，分別是：(1)再測方法(test-retest method)；(2)複本方法(parallel-forms method)；(3)內部一致性方法(internal-consistency method)；(4)評分者方法(scorer method)。請問：這四類信度估計方法的測量誤差主要來源各為何？請列表說明之。(25分)

【解題關鍵】

1. 《考題難易》 ★★☆☆☆
2. 《破題關鍵》本題為基本考古題，主要就是從不同信度的估計方式去說明各自可反應的誤差來源，相信有充分準備的小夥伴看到這一題應該會放下心中沉重的大石。而本題作答時，雖然要求列表呈現，但仍可以透過文字論述的方式搭配表格進行補充說明，這樣作答篇幅與內容會更加完整。
3. 《使用學說》信度的誤差來源

【擬答】

從測量誤差的角度，信度是指測驗分數中，「真實分數」與「測量誤差」的程度，因此，信度係數即為「真實分數變異量在實得分數總變異中所占的比例」，也就是測驗分數反映出真實分數的程度或沒有誤差的程度。而不同類型的信度，所反應的誤差來源也不盡相同，以下將分別說明之。

(一)誤差來源

1. 時間取樣誤差：因測量時間不同所造成的誤差。
2. 內容取樣誤差：指不同題目組(題庫)間的內容差異，也就是依照不同標準所選取的測驗題目，對受評者的測驗得分有很大的影響。
3. 內容異質誤差：指各題內涵的相異程度，若差異甚大，則為內容異質；反之，則為內容同質。內容異質或同質會影響受試者對題目反應的一致性，例如：有兩份測驗，其中一測驗僅包含教育心理的題目，但另一個測驗卻同時有教育行政、教育心理、教學原理等題目，對受試者而言，前者題目的反應一致性可能會比後者高。

4. 評分者誤差：不同評分者評閱測驗時，因「個人主觀判斷上的差異而影響到受評者的測驗得分」所導致的誤差。

(二) 不同信度的意涵及其所反應的誤差

1. 再測信度

計測量信度最直接的方式，將同一測量題目針對同一對象前後進行兩次，根據兩次測驗總分求其相關係數，此相關係數即為再測信度，也可稱為穩定係數；若相關係數高，就表示再測信度高，即表示題目具有穩定性，經得起時間考驗。其誤差變異量主要來自「時間取樣」，也就是時間間隔不同所致。此外，兩次測驗之間各種非系統的因素，也會影響測驗的穩定性，例如：測驗情境、受測者的身心狀況、主試人員狀況等。

2. 複本信度

(1) 同時複本信度：若正本與複本在同一時間實施，此種複本信度可說明「內容取樣」所造成的誤差，稱為「等值係數」，或稱為「複本立即信度」，可用以檢驗測量歷經指標變動所呈現的等值性，當複本信度越高，表示所測量的題目，越能代表該範圍的測量內容。

(2) 延宕複本信度：若先測量正本，延後一段時間後再實施複本測量，除了可說明內容取樣的誤差外，亦可顯示出間隔一段時間後，受測者在不同測驗情境所造成的誤差量(即時間取樣誤差)，此種複本信度又稱為「穩定與等值」係數，或稱為「複本延宕信度」，可以同時檢驗測驗經歷時間變化與指標變動下的穩定性。

3. 內部一致性信度

(1) 折半信度：主要是在測量題項間的內部一致性。根據一次測量結果，將測量題目分為兩半計算分數，通常會採用「奇偶切分法」，一半為「奇數題」，另一半為「偶數題」，並計算兩部分總分之相關，若有高度相關即代表此兩部分相似性高，具有信度。折半信度的測量誤差，主要是來自因隨機抽樣試題以決定測驗內容時所造成的「內容取樣誤差」；因此，折半信度越高，表示兩半測驗的內容越一致或相等，也就是測驗內容的抽樣越適當。

(2) 庫李信度：由庫德(Kuder)和李察遜(Richardson)所發展出來的公式方法，該方法是用在一個測驗僅實施一次，而且不必將測驗分數分為兩半，是分析受測者對每個題目的作答反應，分析試題間的一致性，以確定測驗中的試題是否測量到相同特質的一種估計方法。其反應的誤差主要來自於「內容取樣誤差」以及「內容異質誤差」。一般而言，當測驗中的每個試題都測量到相同能力或潛在特質時，即表示試題的測量功能間一致性越高，信度越大；反之，當測驗中的所有試題都測量到兩種以上的能力或潛在特質時，則表示試題的測量功能不一致，信度將越小。

(3) Cronbach α ：Cronbach α 係依據全體受試者對測驗項目的反應，計算每一個項目以及總分的「變異數」，然後估計「項目變異數」占「總變異數」的比例，以此估計測驗分數的信度。其所反應的測量誤差和 K-R 法一樣，主要來自「內容取樣的誤差」以及「內容異質性的誤差」。

4. 評分者信度

所謂評分者信度是指不同評分者或同一評分者在不同時間內，觀察、記錄、評分等各方面相互間的一致性。若信度高，表示不同評分者對所評閱測驗卷或對作品的看法越一致；反之，若信度低，表示不同評分者對給分標準，彼此不一致，反應出此份測量工具不標準、不客觀，且容易受到評分者因素的影響。欲減少此種影響的方法是刪除極端評分，或以多人評分計算平均值。此種信度可反應出評分者誤差的影響程度。

		意涵	誤差來源
再測信度		同一份測驗的兩次測量結果間的相關係數	時間抽樣
複本信度		兩份複本測量間的測量結果的相關係數	時間取樣/內容取樣
內部一致性	折半信度	同一測驗試題分成兩半，這兩半間的相關係數	內容取樣
	庫李信度	同一測驗試題間的同質性或反應一致性程度的關聯性指標。	內容取樣內容異質
	α	同一測驗試題間的同質性或反應一致性程度的關聯性指標。	內容取樣內容異質
評分者信度		各評分者間或各評分者內之評分結果的相關係數	評分者誤差

志光.保成.學儒

教育行政 近三年 **17**狀元**8**榜眼**7**探花

狀元 112 高考教育行政 林○娟	狀元 112 普考教育行政 許○沂	狀元 111 高考教育行政 陳○兒	狀元 111 地特四等教育行政 盧○(花東區)	狀元 111 地特三等教育行政 劉○勇(竹苗區)
狀元 111 地特三等教育行政 邱○菁	狀元 111 地特三等教育行政 吳○婕(花東區)	狀元 111 地特三等教育行政 陳○俞(新北市)	狀元 111 地特三等教育行政 陳○杰(台中市)	狀元 111 地特四等教育行政 王○威(新北市)
狀元 110 普考教育行政 吳○書	狀元 110 地特三等教育行政 吳○書(台中市)	狀元 110 地特三等教育行政 劉○箴(彰投區)	狀元 110 地特四等教育行政 陳○君(新北市)	狀元 110 地特三等教育行政 白○欣(台北市)
狀元 110 地特三等教育行政 洪○婷(台南市)	狀元 110 地特三等教育行政 謝○芸(基宜區)	榜眼 111 地特三等教育行政 李○廷(竹苗區)	榜眼 111 高考教育行政 沈○君	榜眼 111 普考教育行政 甘○燾
榜眼 111 地特三等教育行政 蔡○儒(新北市)	榜眼 111 地特三等教育行政 黃○寧(台北市)	榜眼 110 地特四等教育行政 黃○蓉(花東區)	榜眼 110 地特三等教育行政 蕭○嶺(台中市)	榜眼 110 地特三等教育行政 李○懸(台北市)
探花 112 普考教育行政 王○萱	探花 112 高考教育行政 陳○媛	探花 111 地特三等教育行政 簡○雯(台中市)	探花 111 地特三等教育行政 黃○諶(竹苗區)	探花 111 地特四等教育行政 黃○蓉(花東區)
	探花 110 普考教育行政 黃○鈺		探花 110 高考教育行政 莊○婷	

系統式高效學習 112年 高考教育行政 楊○翰

老師上課模式是先在黑板列出重要考點架構表，並配合課本內容進行講解，這樣的方式讓學生得以在每個單元形成系統性的知識，遇到題目馬上就能將腦海中對應的知識運用於解題。

公職王歷屆試題 (112 地方特考)

四、某教育學者擬根據五個預測變數來對學校整體滿意度進行多元迴歸分析，他獲得如下的局部報表結果。請問：(每小題 5 分，共 25 分)

- (一)此完整的原始分數迴歸方程式應該為何？請說明之。
- (二)該迴歸方程式的最佳預測力為何？請說明之。
- (三)該迴歸方程式的預測模型是否達顯著程度？請說明之。
- (四)那幾個預測變數可能不具有預測作用？請說明之。
- (五)那一個預測變數最具有預測作用？請說明之。

模式摘要^b

R	R 平方	調過後的 R 平方	估計的標準誤	變更統計量				
				R 平方改變量	F 改變	df1	df2	顯著性 F 改變
.742 ^a	.550	.549	.46540	.550	605.573	5	2479	.000

- a. 預測變數：(常數)，學習活力，學習負荷，關係圓滿，學習樂趣，學習效能。
- b. 依變數：學校整體滿意度。

Anova^a

模式	平方和	df	平均平方和	F	顯著性
迴歸	655.832	5	131.166	605.573	.000 ^b
殘差	536.949	2479	.217		
總數	1192.781	2484			

- a. 依變數：學校整體滿意度
- b. 預測變數：(常數)，學習活力，學習負荷，關係圓滿，學習樂趣，學習效能。

係數

	未標準化係數		標準化係數		t	顯著性	B 的 95% 信賴區間		相關			共線性統計量	
	B 之估計值	標準誤差	Beta	分配			下界	上界	零階	偏	部分	允差	VIF
(常數)	.611	.068			8.952	.000	.477	.745					
關係圓滿	.015	.017	.015		.891	.373	-.018	.048	.432	.018	.012	.626	1.598
學習效能	.362	.022	.373		16.139	.000	.318	.405	.692	.308	.217	.339	2.947
學習樂趣	.207	.021	.219		9.886	.000	.166	.248	.659	.195	.133	.368	2.714
學習負荷	.205	.016	.231		12.432	.000	.173	.237	.614	.242	.168	.526	1.901
學習活力	-.010	.014	-.010		-.748	.455	-.038	.017	.151	-.015	-.010	.940	1.063

- a. 依變數：學校整體滿意度

【解題關鍵】

- 1. 《考題難易》★★☆☆☆
- 2. 《破題關鍵》

這幾年慢慢開始加強報表的解讀題，不需要繁瑣的公式與計算機的操作，僅需針對題意來回答報表中內容，尤其多元迴歸分析的問題，更適合這樣的出題方式，在這幾年也陸續有不少命題，110年原特三等與107地特四等與104年薦任考題皆是相當經典的考題可以參考。

【擬答】

- (一)假設Y：學校整體滿意度，X₁：關係圓滿，X₂：學習效能，X₃：學習樂趣，X₄：學習負荷，X₅：學習活力

公職王歷屆試題 (112 地方特考)

原始分數迴歸方程式：

$$\hat{Y} = 0.611 + 0.015X_1 + 0.362X_2 + 0.207X_3 + 0.205X_4 - 0.010X_5$$

- (二)此迴歸方程式的決定係數為 0.550，代表以關係圓滿、學習效能、學習樂趣、學習負荷與學習活力共同解釋學校整體滿意度時，整體預測力為 55%。
- (三)由 ANOVA 表中可知，整體 F 檢定統計量為 605.573，對應之 p-value < 0.001，代表該迴歸方程式的預測模型達統計顯著。
- (四)從係數估計表可知，關係圓滿與學習活力可能不具有預測作用，因為兩者的 p-value 分別為 0.373 與 0.455，未達 0.05 的顯著水準。
- (五)根據係數估計表，學習效能的標準化迴歸係數 0.373 是所有變異中最大的，所以學習效能對解釋學校整體滿意度最具有預測作用。

志光×學儒×保成
十大貼心服務

學習無後顧之憂

- 線上課業諮詢
- 老師申論批閱
- 上榜生經驗親授
- 時事專題講座
- LINE@班導服務
- 班導師制度
- 雙師資雙循環
- 多元補課方式
- 歷屆試題練習
- 線上平時測驗

詳細規劃請洽全國各班門市