

國立臺中教育大學 96 學年度研究所碩士在職專班招生考試

測驗與評量 試題

教育測驗統計研究所 用

一、閱讀選擇題（四選一，每題 2%，共 40%）

《第一大題》

在試題反應理論(item response theory, IRT)的多元計分模式中，依據資料結果的不同，較常見的模式有下列三種，其皆由二元計分模式延伸而來。下列為一些 IRT 模式：

二參數模式（two-parameter logistic model, 2PL）

$$P_i(\theta) = \frac{\exp[a_i(\theta - b_i)]}{1 + \exp[a_i(\theta - b_i)]}$$

其中，以 $P_i(\theta)$ 表示潛在能力為 θ 的受試者，答對第 i 題的機率； a_i 和 b_i 分別為第 i 題的鑑別度參數及難度參數。當 $a_i = 1$ 時，即為一參數模式（one-parameter logistic model, 1PL）。

等級反應模式（Graded Response Model，簡稱 GRM）

$$P_{ik}^*(\theta) = \frac{\exp[a_i(\theta - b_{iv})]}{1 + \exp[a_i(\theta - b_{iv})]}$$

部分給分模式（Partial Credit Model，簡稱 PCM）

$$P_{ik}(\theta) = \frac{\exp\left[\sum_{v=1}^k (\theta - b_{iv})\right]}{\sum_{c=1}^{m_i} \left[\exp\sum_{v=1}^c (\theta - b_{iv})\right]}$$

一般化部分給分模式（Generalized Partial Credit Model，簡稱 GPCM）

$$P_{ik}(\theta) = \frac{\exp\left[\sum_{v=1}^k a_i(\theta - b_{iv})\right]}{\sum_{c=1}^{m_i} \left[\exp\sum_{v=1}^c a_i(\theta - b_{iv})\right]}$$

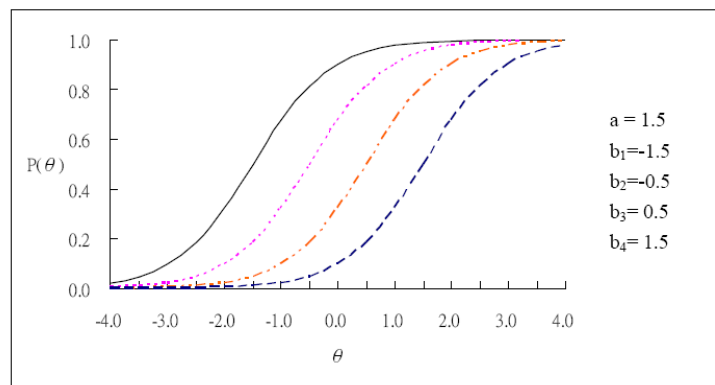
其中，以 $P_k(\theta)$ 表示潛在能力為 θ 的受試者，其答案被評為第 k 類別的機率； $P_{ik}(\theta)$ 表示潛在能力為 θ 的受試者，其第 i 題答案被評為第 k 類別的機率。

a 、 a_i 為鑑別度參數； b_k 、 b_i 及 b_{iv} 為類別閾參數； m_i 為第 i 題的總類別數。

請根據以上敘述回答下列問題：

1. 李克特氏評定量表最適合採用下列何種模式進行分析？
(A) 1PL
(B) 2PL
(C) GRM
(D) GPCM

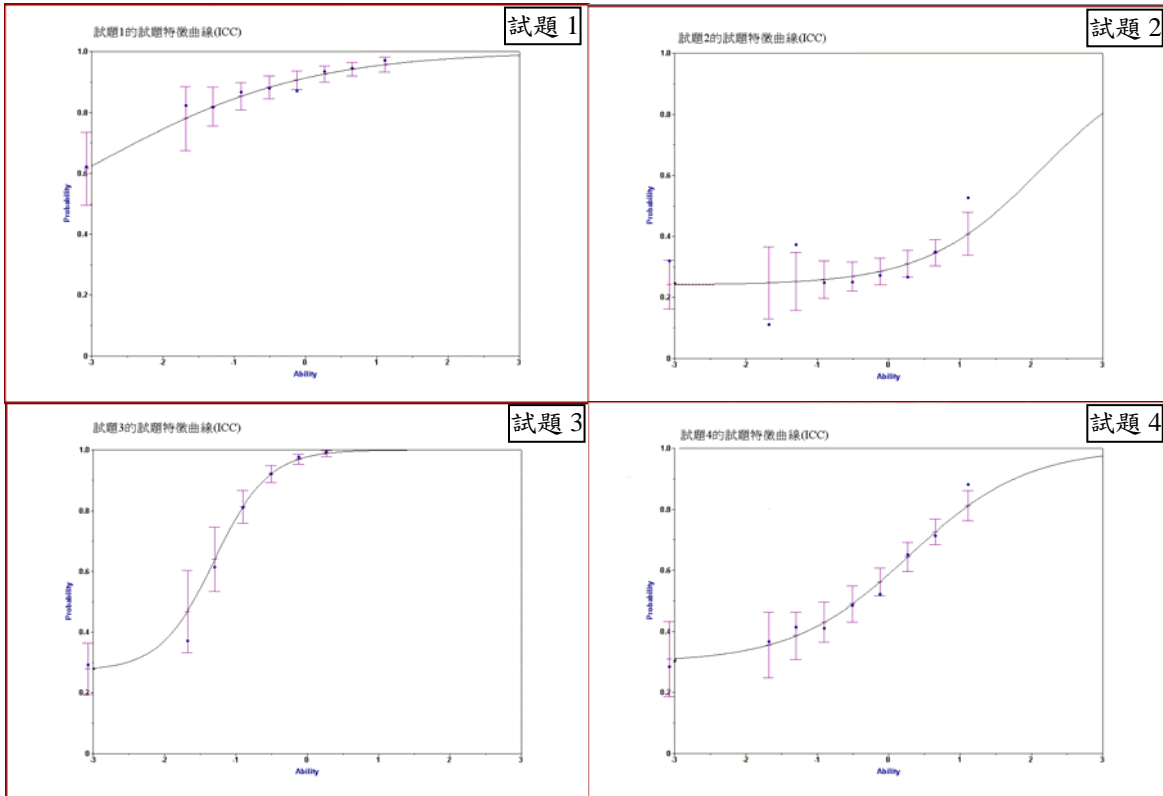
2. 數學領域的四階段給分應用題最適合採用下列何種模式進行分析？
- (A) 1PL
(B) 2PL
(C) GRM
(D) GPCM
3. 有關 GPCM 模式與 PCM 模式之描述，下列何者為真？
- (A) 二種模式下各題之鑑別度相同且各題共享一套類別閾參數
(B) GPCM 模式下各題之鑑別度不同，且各題類別閾參數可不同
(C) PCM 模式下各題之鑑別度不同，但各題共享一套類別閾參數
(D) 二種模式下各題之鑑別度均不同且各題共享一套類別閾參數
4. 某試題具有五個類別閾參數，則我們可推論其為幾點計分的試題？
- (A) 4
(B) 5
(C) 6
(D) 7
5. 下圖為某試題的特徵曲線圖，圖中題目共有四個類別，具有共同的鑑別度 ($a=1.5$)，四個類別難度的難度值分別為-1.5、-0.5、0.5、1.5。請問此圖應屬於何模式？



- (A) 1PL
(B) 2PL
(C) GRM
(D) PCM

《第二大題》

以下是使用試題反應理論軟體 Bilog-MG 分析某一項成就測驗所得之四條試題特徵曲線，請依此回答下列問題。



6. 請問本測驗的試題特徵曲線是屬於試題反應理論中何種模式？
 - (A) 單參數模式
 - (B) 二參數模式
 - (C) 三參數模式
 - (D) 四參數模式
7. 四個試題中，哪一題最簡單？
 - (A) 試題 1
 - (B) 試題 2
 - (C) 試題 3
 - (D) 試題 4
8. 四個試題中，哪一題鑑別度參數最大？
 - (A) 試題 1
 - (B) 試題 2
 - (C) 試題 3
 - (D) 試題 4
9. 四個試題中，哪一題猜測度最高？
 - (A) 試題 1
 - (B) 試題 2
 - (C) 試題 3
 - (D) 試題 4
10. 四個試題中，哪一題試題特徵曲線與實際資料之適配度最好？
 - (A) 試題 1
 - (B) 試題 2
 - (C) 試題 3
 - (D) 試題 4

《第三大題》

下表是取自某研究者在某測驗（共 5 題）所分析的統計資料，以及此研究者以 SPSS 統計軟體分析後的部分輸出結果。試根據以下表格，回答第 11 題至第 15 題：

Task-takers	Items					Total Score
	1	2	3	4	5	
A	3	4	3	4	4	18
B	4	5	4	3	5	21
C	2	3	1	2	2	10
D	1	1	2	1	2	7
E	5	5	4	5	5	24
F	1	2	2	1	1	7
G	5	5	5	4	5	24
H	3	3	4	5	3	18
\bar{X}_i	3	3.5	3.125	3.125	3.375	$\bar{X} = 16.125$
S_i	1.5	1.414214	1.268611	1.536026	1.494783	$S_x = 6.697714$
S_i^2	2.25	2	1.609375	2.359375	2.234375	$S_x^2 = 44.85938$

其中 $\sum_{i=1}^5 S_i^2 = 10.45313$

RELIABILITY ANALYSIS - SCALE (ALPHA)				
Item-total Statistics				
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Alpha if Item Deleted
Item1	13.1250	31.2679	.9718	.9336
Item2	12.6250	33.4107	.8909	.9478
Item3	13.0000	35.7143	.8461	.9560
Item4	13.0000	33.4286	.7975	.9644
Item5	12.7500	31.9286	.9295	.9411

11. 此測驗的 Cronbach α 信度大約為何？

- (A) .858725 (B) .898725 (C) .958725 (D) .978725

12. 下列何種處理最有可能提高此測驗信度？

- (A) 增加受試者人數 (B) 增加題數
(C) 採用最佳的信度估計公式 (D) 減少每題的計分點數

13. 根據測驗學的原理，信度是一個估計值，而真正的信度並不可知。因此，Cronbach 公式計算所得的信度值，會比真正的信度高或低？

- (A) 比真正的信度高 (B) 比真正的信度低
(C) 在小樣本的情形下，比真正的信度高 (D) 在大樣本的情形下，比真正的信度高

14.根據此表格中的數據，下列何者正確的敘述？

- (A) 可據以計算內容效度
- (B) 可計算庫李信度 (Kuder-Richardson reliability)
- (C) 每一題平均數相當高 (大於 3)，所以信度也必定相當高
- (D) 可以計算 Hoyt 信度

15.如果要刪除不良題目以提高信度，則哪一題應該首先應刪除？

- (A) 第 1 題
- (B) 第 2 題
- (C) 第 4 題
- (D) 第 5 題

《第四大題》

下表是取自某研究生在編製量表過程時，就預試資料進行因素分析所得的因素分析摘要表(已轉軸)，空白處 (X1 至 X16) 表示此研究生尚未填寫進去。試根據以下表格，回答第 16 題至第 20 題：

試題	共同因素			共同性
	Factor 1	Factor 2	Factor 3	
1	.02	.73	.21	X1
2	.89	.01	.13	X2
3	.77	.05	.15	X3
4	.86	.13	.09	X4
5	.17	.84	.18	X5
6	.20	.07	.82	X6
7	.77	.16	.05	X7
8	.16	.14	.84	X8
9	.21	.55	.36	X9
10	.18	.17	.88	X10
11	.09	.79	.16	X11
12	.91	.04	.18	X12
13	.86	.23	.17	X13
14	.89	.16	.14	X14
15	.17	.03	X15	.6074
特徵值	5.2857	X16	3.0926	10.7229

16.此研究生在這個量表上，其所要建立的效度是哪一種效度？

- (A) 效標關聯效度
- (B) 內容效度
- (C) 專家效度
- (D) 建構效度

17.在因素分析摘要表中，X10 應該大約等於何值(取小數點至第四位)？

- (A) .8357
- (B) .8800
- (C) .8373
- (D) .4762

18.在因素分析摘要表中，X15 應大約等於何值(取小數點至第二位)？

- (A) .41
- (B) .56
- (C) .72
- (D) .76

19.在因素分析摘要表中，X16 應該等於何值(取小數點至第四位)？

- (A) 4.1005
- (B) 3.9817
- (C) 3.4215
- (D) 2.3446

20.該研究生的指導教授認為有一題不符合簡單結構的原則，亦即該題必須要刪除。則該題可能是哪一題？

- (A) 第 2 題
- (B) 第 7 題
- (C) 第 9 題
- (D) 第 11 題

二、選擇題（四選一，每題 2%，共 30%）

21. 某研究者欲利用徑路搜尋 (pathfinder) 探討學童的加減法知識結構，下列何者是他正確的作法？
- (A) 以統計檢定瞭解節點 (node) 之間的聯節強度是否達統計顯著
 - (B) 以統計學的徑路分析 (path analysis) 尋求最佳模式
 - (C) 以配對比較 (paired comparisons) 獲得接近性矩陣
 - (D) 可事先決定節點的階層性 (hierarchy)
22. 某研究者進行數學科試題鑑別度分析中，欲採用外在效標 (exterior criteria)，下列何者是他的正確作法？
- (A) 此外在效標不能是研究者本身設計的測驗工具
 - (B) 數學科測驗所測的樣本須與外在效標所測樣本相同
 - (C) 外在效標是否測量相同的潛在特質，不是主要考量，因為可利用統計方法排除
 - (D) 外在效標的計分類型應與此數學科測驗相同
23. 某研究者欲根據 J. D. Novak 所提出的概念構圖 (concept mapping) 來探究學童的幾何圖形概念，下列何者是他正確的作法？
- (A) 不宜有交叉聯結 (cross linking)，以免結構圖過於複雜難懂
 - (B) 在概念的連節關係之間加上聯結語
 - (C) 結構圖中概念數目愈多，表示其概念愈完備
 - (D) 只適合以概念圖形呈現，無法以數值進行結構圖之間的比較
24. 關於 J. P. Doignon 和 J. C. Falgout 所提出的知識空間 (knowledge space) 評量模式之敘述，何者最為正確？
- (A) 利用徑路分析 (path analysis) 推估受試者的學習路徑 (learning path)
 - (B) 以 Hasse 圖描述試題的階層關係
 - (C) 以相似性評定描述試題的近似程度
 - (D) 必須先確定受試者之間的階層關係
25. 關於題組反應模式 (testlet response model) 的敘述，下列何者最為正確？
- (A) 適用於多元計分 (polytomous) 的資料
 - (B) 適用於簡答題之題目類型
 - (C) 不適用於速度測驗 (speed test)
 - (D) 針對違反局部獨立性 (local independence) 的試題進行分析
26. 下列何者不是常模樣本的要件？
- (A) 廣泛性
 - (B) 新近性
 - (C) 代表性
 - (D) 適切性
27. 就「效度」的定義而言，以下選項何者正確？
- (A) 效度是指測驗之測量結果的穩定性
 - (B) 效度是指測驗分數的有效程度
 - (C) 效度是指測驗題目的有效程度
 - (D) 效度是指測量工具的穩定程度
28. 打字測驗、速算測驗等類型的測驗適合採用下列何種估計信度的方法？
- (A) 重測信度
 - (B) 折半信度
 - (C) Alpha 係數
 - (D) 庫李信度

29. 某研究者在進行測驗等化設計時，採用將欲銜接的測驗給予兩組相似但不完全相同的考生施測，每組考生另接受一部份共同試題的測驗，且共同測驗的試題內容必須與原二份測驗十分類似，且測驗長度相當於一個分測驗的方式，此種銜接法稱為什麼？
- (A) 分測驗預先銜接設計
 - (B) 試題預先銜接設計
 - (C) 定錨測驗隨機分組設計
 - (D) 單一組共同考生設計
30. 下列何者不是檔案評量實施過程中之限制？
- (A) 學生反思能力不足
 - (B) 學生缺乏學習動機
 - (C) 評分規準建立困難
 - (D) 經費與空間的不足
31. 欲測量學生是否具備應用工具測量圓周長度的能力，最宜採用哪下列何種評量方式？
- (A) 實作評量
 - (B) 動態評量
 - (C) 歷程評量
 - (D) 變通評量
32. 某研究生在製作問卷時，採用以下四個準則來排序問題，這些準則中何者是正確的？
- (A) 將問題長度相同或類似者放在一起
 - (B) 將敏感性問題隱藏在問卷中間的位置
 - (C) 將要用來查核信度的配對問題分開擺置
 - (D) 將事實性問題放在態度性問題之前
33. 學習態度量表（Likert-type 五點式量表）最適合以下列哪一公式來求信度？
- (A) K-R20 號公式
 - (B) K-R21 號公式
 - (C) Alpha 係數公式
 - (D) 折半法公式
34. 在進行觀察評量時，教師對學生先入為主的印象，容易導致評定的結果產生偏差，此稱為何種效應？
- (A) 比馬龍
 - (B) 強亨利
 - (C) 霍桑
 - (D) 月暈
35. 一標準化測驗在一段時間後重複施測，下列何者為其重測信度之主要誤差來源？
- (A) 受測者的學習與成長
 - (B) 考試地點的座位安排
 - (C) 測驗的計分誤差
 - (D) 測驗試題的品質不良

三、選擇題（五選一，每題 2%，共 30%）

以下題目皆為五選一的單選題，但不代表有唯一的標準答案，請詳細比較五個選項之後，選出一個相對的最佳答案。

36. 古典測驗理論若以數學式 $X=T+E$ 表示，其中 T 是代表
- (A) 觀察分數
 - (B) 潛藏分數
 - (C) 真實分數
 - (D) 誤差分數
 - (E) 量尺分數
37. 以下的效度指標中，何者是用來檢驗測驗對於相關能力的預測程度？
- (A) Discriminant validity
 - (B) Concurrent validity
 - (C) Trait validity
 - (D) Face validity
 - (E) Construct validity
38. 公式： $S_m = S\sqrt{1-r}$ 中， S 是測驗分數的標準差， r 是該測驗的信度係數，所以 S_m 稱為？
- (A) Correction for attenuation
 - (B) Cronbach alpha
 - (C) KR-20 coefficient
 - (D) Standard error of measurement
 - (E) Spearman-Brown Reliability
39. 公式： $r = \left(\frac{N}{N-1}\right)\left(\frac{S^2 - \sum S_i^2}{S^2}\right)$ 中， N 是測驗的題數， S^2 是全測驗的變異數， S_i^2 是測驗內逐題的變異數， r 稱為？
- (A) Correction for attenuation
 - (B) Cronbach alpha
 - (C) KR-20 coefficient
 - (D) Standard error of measurement
 - (E) Spearman-Brown Reliability
40. 以下哪一個測驗執行時，「個別施測」的執行方法，並不是唯一的使用方式？
- (A) WISC-III
 - (B) WPPSI
 - (C) Stanford-Binet Scale
 - (D) Raven Progressive Matrices
 - (E) Peabody Picture Vocabulary Test

41. 請問「在家考試」(take-home) 是屬於下列哪一種評量方式？
- (A) 申論題考試
 - (B) 選擇題測驗
 - (C) 紙筆測驗
 - (D) 人格測驗
 - (E) 學習檔案
42. 請問「檔案評量」(portfolios) 是屬於下列哪一種評量方式？
- (A) 真實評量
 - (B) 非紙筆測驗
 - (C) 性向測驗
 - (D) 動態評量
 - (E) 學習檔案
43. 請問「開書考試」(open-book) 是屬於下列哪一種評量方式？
- (A) 檔案評量
 - (B) 實作評量
 - (C) 紙筆測驗
 - (D) 成就測驗
 - (E) 學習日誌
44. 下列何者不是傳統試題分析 (或項目分析) 所提供的資料？
- (A) 難度
 - (B) 效度
 - (C) 鑑別度
 - (D) 選項誘答力
 - (E) 答對率
45. 請問下列哪一項最符合「檔案評量」(portfolios) 的特徵？
- (A) 作品
 - (B) 正確
 - (C) 速度
 - (D) 反省
 - (E) 過程
46. 採用雙向細目表，下列何者可以改進？
- (A) 同時效度
 - (B) 效標關連效度
 - (C) 預測效度
 - (D) 內容效度
 - (E) 建構效度

47. 請問有關「信度」的估計方法，下列何者是無法一次施測完成的？
- (A) 折半法
 - (B) 重測複本法
 - (C) alpha 係數
 - (D) 庫李法
 - (E) 複本法
48. 請問採用「百分等級」為計分基礎的測驗，最接近下列何者？
- (A) 標準參照測驗
 - (B) 常模參照測驗
 - (C) 形成性測驗
 - (D) 診斷性測驗
 - (E) 速度測驗
49. 請問「診斷評量」最類似下列哪一類評量或測驗？
- (A) 標準參照測驗
 - (B) 常模參照測驗
 - (C) 安置測驗
 - (D) 實作評量
 - (E) 檔案評量
50. 請問下列哪一項的記分方式最接近目前臺灣的「基測」或「學測」的記分方式？
- (A) 總結性測驗
 - (B) 形成性測驗
 - (C) 常模參照測驗
 - (D) 標準參照測驗
 - (E) 診斷測驗