

第四章 結果與討論

本研究之目的在於分析文字系列完成測驗中各種預測變數對試題難度與鑑別度的影響，以及研究該測驗對受試者歸納推理能力的評估效果，以下針對各研究結果分節敘述。

第一節 預測變數對試題難度的影響

本研究採用多元迴歸分析 (multiple regression analysis) 研究各種預測變數對試題難度之影響，多元迴歸方程式假設如下：

$$Y = \alpha + \beta_1 X_{CNT} + \beta_2 X_{REL} + \beta_3 X_{ADJ} + \beta_4 X_{EQU} + \beta_5 X_{SI} + \varepsilon$$

各變數說明如下：

Y：為多元迴歸方程式的依變項，代表試題的難度。

α 與 β_i ： α 是常數， β_i 是各自變項的參數，兩者皆為迴歸分析所要進行估計的未知數。

X_{CNT} 、 X_{REL} 、 X_{ADJ} 、 X_{EQU} 、 X_{SI} ：為多元迴歸方程式的自變項，分別代表預測變數 CNT、REL、Adjacency、Equality 與 SI。

ε ：為隨機誤差項，是實際值減去預測值以後的殘差值，代表依變項不能被自變項解釋的部份。

在進行多元迴歸分析之前須滿足多個假設，故本研究先行驗證以下數個假

設：

- 一、自變項是非隨機的 (nonstochastic)
- 二、常態性 (normality)
- 三、無自我相關 (nonautocorrelation)
- 四、變異同質性 (homoscedasticity)
- 五、自變項無線性重合 (no multicollinearity)

本研究每個題目所擁有的預測變數皆為固定值，並非隨機，故符合第一個假設無疑。

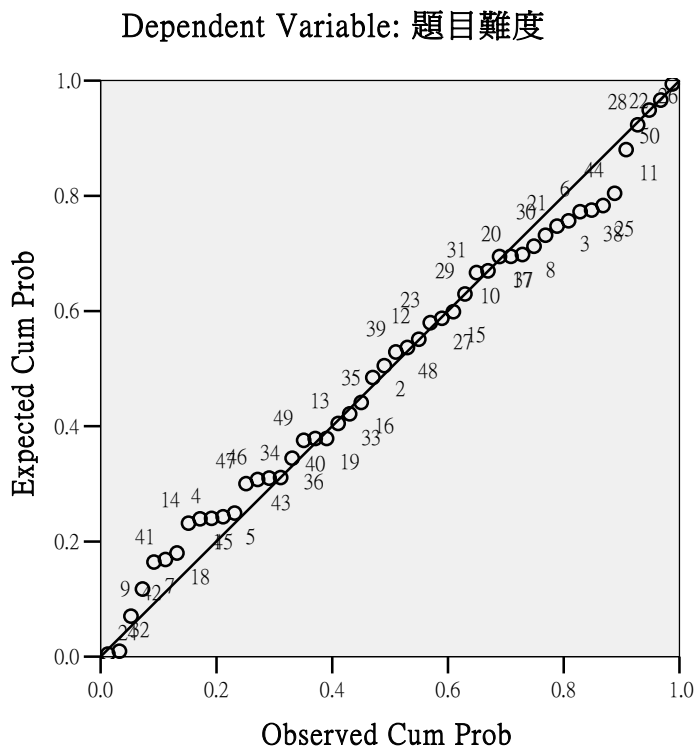


圖 4-1-1 題目難度迴歸分析的樣本殘差值的常態機率分布圖 (normal probability plot)

常態性之假設可由樣本殘差值的常態機率分布圖 (normal probability plot) 來判斷，若該圖中的累積機率分布接近為右上到左下的 45° 線，則表示樣本觀察值符合常態性之假設，由圖 4-1-1 可見，本研究之樣本觀察值大致符合常態性之假設。

無自我相關假設可由 Durbin-Watson 值加以判斷，經由 SPSS 軟體分析，本研究的 Durbin-Watson 值為 1.857，經由查表得知，在自變項為 5 個，而樣本數為 50 時，殘差值有自我相關的臨界值為小於 1.53 或大於 2.47，而 1.857 介於該兩值之間，故未違反無自我相關之基本假設。

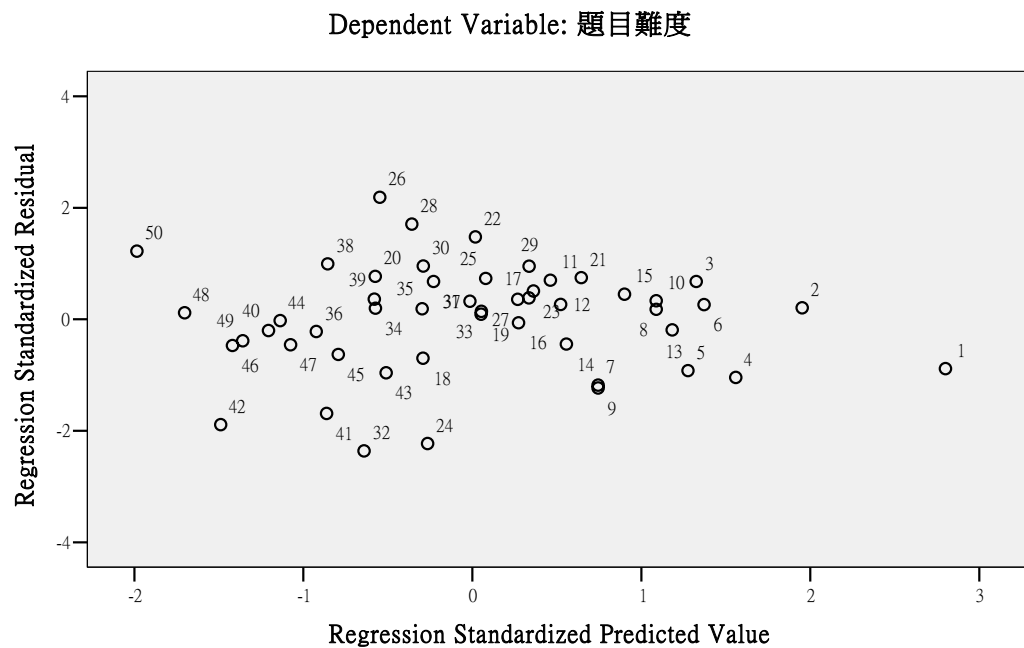


圖 4-1-2 題目難度迴歸分析的殘差值與預測值之交叉散佈圖

變異同質性假設可由殘差值與預測值之交叉散佈圖加以分析，當散佈圖呈水

平的隨機散佈時，表示樣本觀察值符合變異同質性之假設。由圖 4-1-2 可見，本研究之樣本觀察值亦大致符合變異同質性之假設。

表 4-1-1 題目難度迴歸分析的自變項與依變項積差相關係數表

		題目難度	CNT	REL	Adjacency	Equality	SI
Pearson Correlation	題目難度	1.000	-.533	-.485	-.466	-.447	-.362
	CNT	-.533	1.000	.562	.446	.322	.111
	REL	-.485	.562	1.000	.193	.258	.061
	Adjacency	-.466	.446	.193	1.000	.570	.064
	Equality	-.447	.322	.258	.570	1.000	.073
	SI	-.362	.111	.061	.064	.073	1.000
	Sig. (1-tailed)	題目難度	.	.000	.000	.000	.001
	CNT	.000	.	.000	.001	.011	.221
	REL	.000	.000	.	.089	.035	.336
	Adjacency	.000	.001	.089	.	.000	.329
	Equality	.001	.011	.035	.000	.	.307
	SI	.005	.221	.336	.329	.307	.

表 4-1-1 為各自變項與依變項的積差相關係數表，由表中可見任一預測變數與題目難度之相關皆達 0.01 之顯著水準，而由自變項間的相關係數可知，CNT 與 REL 及 Adjacency 之間的相關達到 0.562 與 0.446，Adjacency 與 Equality 之間的相關達到 0.570，可能會出現線性重合的問題，但尚須以其他方式進一步檢定。

表 4-1-2 題目難度迴歸分析的自變項線性重合檢定表

Model Dimension		Condition Index	Variance Proportions					
			(Constant)	CNT	REL	Adjacency	Equality	SI
1	1	1.000	.00	.00	.00	.00	.00	.01
	2	3.237	.00	.01	.01	.00	.01	.96
	3	5.333	.03	.29	.14	.02	.13	.01
	4	7.067	.07	.44	.47	.00	.14	.00
	5	7.866	.15	.10	.26	.05	.53	.01
	6	15.216	.75	.16	.11	.93	.19	.01

a. Dependent Variable: 題目難度

表4-1-2為自變項線性重合的檢定結果，條件指數（Condition Index，簡稱CI值）可以用來判斷線性重合問題，Beleley等（1980）指出，若CI值在10左右，表示線性重合問題並不嚴重，若CI值在30到100之間，表示變項間具有中度至高度的線性重合問題，若CI值在100以上，則線性重合問題相當嚴重，本研究的CI值最大為15.216，表示線性重合問題並不嚴重。變異數比例（Variance Proportions）也是用來檢查線性重合的指標之一，當任兩自變項在同一特徵值上之變異數比例值都非常接近1，表示該兩自變項間可能有線性重合的問題，而觀察表4-1-2之後發現，本研究並無此狀況發生，故以表4-1-2的分析結果而言，本研究應無線性重合問題存在。

表4-1-3 試題難度的描述統計

	N	Minimum	Maximum	Mean	Std. Deviation
題目難度	50	.491289	.959459	.75687381	.109090134
Valid N (listwise)	50				

由表4-1-3可知，整體試題難度平均為0.76左右，略偏簡單，而最難的題目難度為0.49左右，難易適中，最簡單的題目難度為0.96左右，幾乎所有的受試者都能答對該題目。

表 4-1-4 題目難度迴歸分析以 Enter 法進行分析所得的迴歸模式的統計量

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.721(a)	.519	.465	.079819013	1.857

a Predictors: (Constant), SI, REL, Adjacency, Equality, CNT

b Dependent Variable: 題目難度

若採用Enter法進行迴歸分析，可以得到迴歸模式的統計量如表4-1-4，由表中可知所有預測變數共可以解釋題目難度總變異量的51.9%。

表4-1-5 題目難度迴歸分析以Enter法進行分析所得的迴歸模式參數檢定結果

Model		Unstandardized Coefficients		Standardized Coefficients
		B	Std. Error	Beta
1	(Constant)	1.016	.050	
	CNT	-.008	.006	-.201
	REL	-.019	.009	-.269
	Adjacency	-.094	.062	-.207
	Equality	-.033	.025	-.173
	SI	-.064	.023	-.297

a. Dependent Variable: 題目難度

表4-1-5為以Enter法進行分析所得的回歸模式的各參數檢定結果，由該表可得，有關本研究各預測變數與試題難度的未標準化回歸模式為：

$$Y=1.016-0.008X_{CNT}-0.019X_{REL}-0.094X_{ADJ}-0.033X_{EQU}-0.064X_{SI}+\varepsilon$$

標準化回歸模式則為：

$$Y=-0.201X_{CNT}-0.269X_{REL}-0.207X_{ADJ}-0.173X_{EQU}-0.297X_{SI}+\varepsilon$$

第二節 預測變數對試題鑑別度的影響

試題的鑑別度分析可分為內部一致性（internal consistency）與外在效度（external validity）兩方面，本研究將分別對這兩方面進行分析。

一、內部一致性的分析

表4-2-1為鑑別指數（index of discrimination）的描述統計，由該表中可發現，在50題測驗試題中，鑑別度最高的題目其鑑別指數有0.90左右，鑑別力不錯，而鑑別度最低的題目其鑑別指數只有0.07，可以說是毫無鑑別力，整體平均鑑別指數約為0.51。

表 4-2-1 鑑別指數的描述統計

	N	Minimum	Maximum	Mean	Std. Deviation
鑑別指數	50	.072289	.904472	.50579948	.254355510
Valid N (listwise)	50				

表4-2-2為鑑別指數與各預測變數的相關表，由表中可觀察得知，除了預測變數SI以外，其餘的預測變數與鑑別指數的相關皆達顯著水準。

表 4-2-2 鑑別指數與各預測變數的積差相關表

		CNT	REL	Adjacency	Equality	SI
Pearson Correlation	鑑別指數	.864	.601	.430	.359	.188
Sig. (1-tailed)	鑑別指數	.000	.000	.001	.005	.096

進行迴歸分析前所需滿足的五個假設中，由於第一個假設（自變項是非隨機的）與第五個假設（自變項無線性重合）的驗證與上一節相同，故該兩個假設不重新驗證，僅就其餘三個假設進行驗證以利進行迴歸分析。

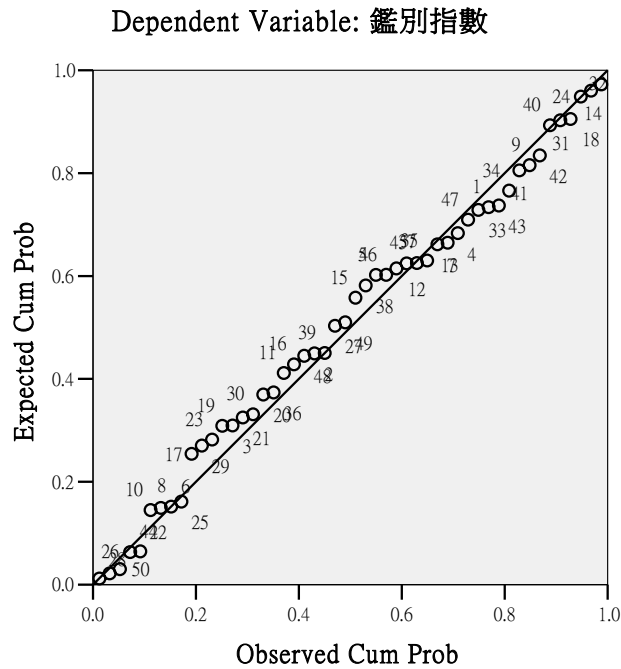


圖 4-2-1 鑑別指數迴歸分析的樣本殘差值的常態機率分布圖

圖 4-2-1 為樣本殘差值的常態機率分布圖，由圖中可以了解，本研究之樣本觀察值大致符合常態性之假設。

經由 SPSS 軟體分析，本研究的 Durbin-Watson 值為 1.717，經由查表得知，在自變項為 5 個，而樣本數為 50 時，殘差值有自我相關的臨界值為小於 1.53 或大於 2.47，而 1.857 恰好介於該兩值之間，故未違反無自我相關之基本假設。

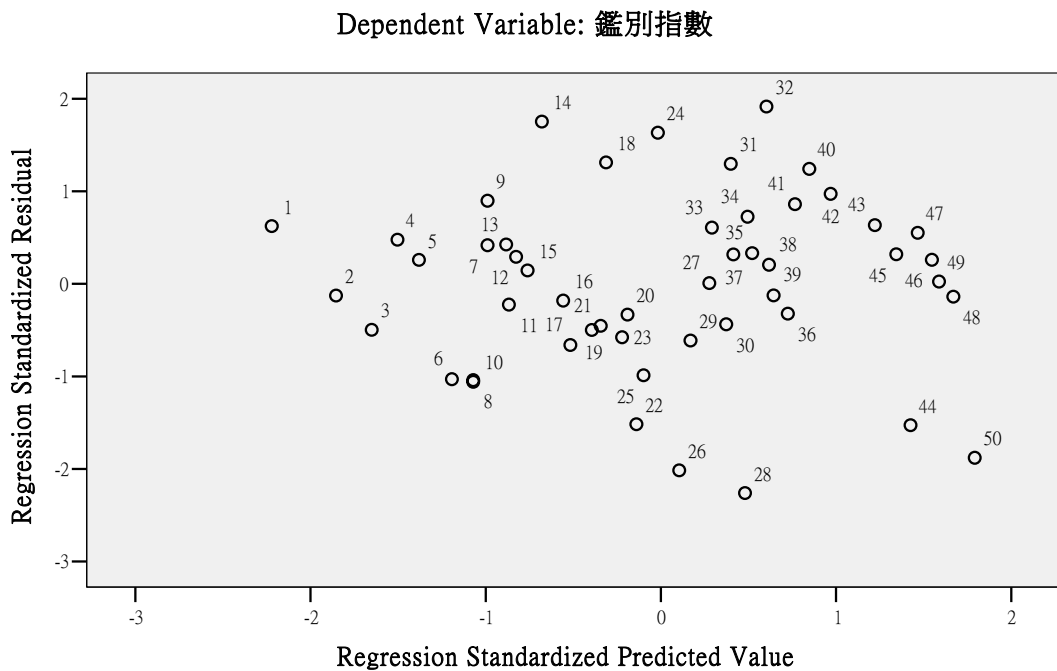


圖 4-2-2 鑑別指數迴歸分析的殘差值與預測值之交叉散佈圖

由圖 4-2-2 殘差值與預測值之交叉散佈圖可見，本研究之樣本觀察值亦大致符合變異同質性之假設。

表 4-2-3 鑑別指數迴歸分析以 Enter 法進行分析所得的回歸模式的統計量

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.883 ^a	.780	.755	.125873139	1.717

a. Predictors: (Constant), SI, REL, Adjacency, Equality, CNT

b. Dependent Variable: 鑑別指數

表 4-2-4 鑑別指數迴歸分析以 Enter 法進行分析所得的迴歸模式參數檢定結果

Model		Unstandardized Coefficients		Standardized Coefficients
		B	Std. Error	Beta
1	(Constant)	.007	.079	
	CNT	.070	.009	.727
	REL	.027	.014	.165
	Adjacency	.040	.098	.038
	Equality	.024	.039	.054
	SI	.046	.036	.091

a. Dependent Variable: 鑑別指數

若採用Enter法進行迴歸分析，可以得到迴歸模式的統計量如表4-2-3，由表中可知所有預測變數共可以解釋題目鑑別指數總變異量的78%。

表4-2-4為以Enter法進行分析所得的迴歸模式的各參數檢定結果，由該表可得，有關本研究各預測變數與試題鑑別指數的未標準化迴歸模式為：

$$Y=0.007+0.07X_{CNT}+0.027X_{REL}+0.04X_{ADJ}+0.024X_{EQU}+0.046X_{SI}+\varepsilon$$

標準化迴歸模式則為：

$$Y=0.727X_{CNT}+0.165X_{REL}+0.038X_{ADJ}+0.054X_{EQU}-0.091X_{SI}+\varepsilon$$

值得注意的是，仔細觀察各預測參數的標準化迴歸係數（Beta），可以發現預測變數 CNT 的標準化迴歸係數顯得特別大，也就是其相對重要性特別高，故研究者採用 Stepwise 法重新進行迴歸分析，其迴歸分析統計量結果如表 4-2-5。

表 4-2-5 鑑別指數迴歸分析以 Stepwise 法進行分析所得的迴歸模式的統計量

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.864 ^a	.747	.742	.129317027	1.626

a. Predictors: (Constant), CNT

採用Stepwise法後，統計軟體僅採用CNT一個預測變數進行分析，其餘預測變數皆被省略，而預測變數CNT可以解釋題目鑑別指數總變異量的74.7%。

表 4-2-6 鑑別指數迴歸分析未進入迴歸模式的自變項相關統計量

Model		Beta In	t	Sig.
1	REL	.168(a)	1.969	.055
	Adjacency	.056(a)	.687	.495
	Equality	.090(a)	1.171	.247
	SI	.093(a)	1.281	.206

a Predictors in the Model: (Constant), CNT

表4-2-6為未進入迴歸模式的自變項相關統計量，觀察被統計軟體捨棄不用的其他預測變數，可以發現這些預測變數的t值皆未達顯著水準，即使強制投入迴歸模式中，其標準化迴歸係數與預測變數CNT的標準化迴歸係數相比也顯得相當小，相對重要性不足。

表 4-2-7 鑑別指數迴歸分析以 Stepwise 法分析所得的迴歸模式參數檢定結果

		Unstandardized Coefficients		Standardized Coefficients
Model		B	Std. Error	Beta
1	(Constant)	.122	.037	
	CNT	.083	.007	.864

參考表4-2-7的檢定結果，有關本研究各預測變數與試題鑑別指數的未標準化迴歸模式可調整為：

$$Y=0.122+0.083X_{CNT}+\varepsilon$$

標準化迴歸模式則調整為：

$$Y=0.864X_{CNT}+\varepsilon$$

二、外在效度的分析

本研究採用多因素性向測驗中的抽象推理分測驗作為外在效標，表4-2-8為效度指數（index of validity）的描述統計，從表中可以發現，效度指數顯得相當低，平均效度指數約只有0.22，最高效度指數約為0.48，最低的效率指數甚至是負值，為-0.18左右，可能原因將於第六節進一步探討。試題的難度、鑑別指數與效度指數對照表請參見附錄C。

表 4-2-8 效度指數的描述統計

	N	Minimum	Maximum	Mean	Std. Deviation
效度指數	50	-.017857	.482143	.21964286	.144589082
Valid N (listwise)	50				

第三節 採用本系統對試題曝光率的影響

本實驗為避免試題曝光問題的產生，採用虛擬題庫系統讓系統自動命題，為證明運用本系統將可以有效避免試題曝光的問題，研究者使用本系統進行模擬命題，命題試卷數共1000份，每份試卷有50題，題目共計50000題，經由下面的試題重複率公式可求得試題重複率：

$$R_t = \frac{T_0 / C_2^N}{(\sum_{i=1}^N L_i) / N} = \frac{2T_0}{(N-1) \sum_{i=1}^N L_i}$$

R_t ：試題重複率

T_0 ：兩份試卷中試題重複的總數

Li：試卷中的題目數量

經由模擬命題比對後求得，在1000份模擬試卷中，試題共重複了2108題，代入上面的試題重複率公式，可求得試題重複率約為0.0085172，在試題重複率相當低的條件下，試題曝光的問題也將隨之下降。

第四節 本系統與外在效標之相關分析

本研究採用多因素性向測驗中的抽象推理分測驗作為外在效標，以該分測驗成績與本研究中「文字系列完成測驗虛擬題庫測驗系統」的測驗成績進行Pearson積差相關分析，得到相關係數如表4-4-1所示。

表4-4-1 文字系列完成測驗成績與抽象推理能力原始分數的相關係數

進行分析的受試者組別	Pearson 積差相關係數
全部受試者（共209人）	0.45**

** . Correlation is significant at the 0.01 level (2-tailed).

由表中可見，全體受試者使用「文字系列完成測驗虛擬題庫測驗系統」進行測驗所得到的成績，與其在多因素性向測驗中抽象推理分測驗成績的相關係數為0.45，相關係數達到顯著水準但並不高。

第五節 不同性別進行測驗的結果比較

表4-5-1為不同性別進行測驗的描述統計結果，表格中前八項為受試者在多因素性向測驗中，各項分測驗成績的描述統計，最後一項則是受試者在接受「文字系列完成測驗虛擬題庫測驗系統」進行測驗後所得到的分數之描述統計。

表 4-5-1 不同性別進行測驗的描述統計結果

測驗項目	性別	N	Mean	Std. Deviation	Std. Error Mean
語文推理	男	83	18.65	4.647	.510
	女	126	19.64	4.179	.372
數學推理	男	83	16.34	3.958	.434
	女	126	14.75	3.523	.314
機械推理	男	83	14.25	3.652	.401
	女	126	12.61	2.814	.251
空間關係	男	83	22.06	4.031	.443
	女	126	18.20	4.209	.375
抽象推理	男	83	20.86	4.234	.465
	女	126	19.33	3.580	.319
錯別字	男	83	42.75	12.028	1.320
	女	126	43.67	10.697	.953
文法修辭	男	83	22.40	5.203	.571
	女	126	23.04	4.129	.368
知覺速度與確度	男	83	75.14	19.478	2.138
	女	126	73.59	16.963	1.511
文字系列完成測驗總得分	男	83	37.42	10.486	1.151
	女	126	36.88	9.232	.822

若欲比較不同性別在各項測驗中的表現是否有差異，尚需進行t檢定，檢定結果如表4-5-2所示。

表 4-5-2 不同性別在各項測驗得分差異的 t 檢定結果

	t-test for Equality of Means		
	t	df	Sig. (2-tailed)
語文推理	-1.606	207	.110
數學推理	3.041	207	.003**
機械推理	3.473	144.218	.001**
空間關係	6.600	207	.000**
抽象推理	2.795	207	.006**
錯別字	-.579	207	.563
文法修辭	-.991	207	.323
知覺速度與確度	.612	207	.541
文字系列完成測驗	.392	207	.695
總得分			

** . Correlation is significant at the 0.01 level (2-tailed).

總和表4-5-1與上表可以發現，在數學推理、機械推理、空間關係與抽象推理等四個分測驗中，男性受試者與女性受試者的成績有顯著差異，而在包含本研究測驗結果在內的其他項目的表現上，兩性受試者並無顯著差異。

第六節 研究結果與綜合討論

經由以上各節的研究結果分析之後，可得到以下數點結論：

一、以本研究所採用的五個預測變數對文字系列完成測驗的難度進行多元迴歸

分析之後，所得到的標準化回歸模式為：

$$Y = -0.201X_{CNT} - 0.269X_{REL} - 0.207X_{ADJ} - 0.173X_{EQU} - 0.297X_{SI} + \varepsilon$$

由各預測變數的標準化回歸係數可以了解，本研究採用的五個預測變數對試題難度皆具有相當影響力，其中以預測變數SI的相對影響力較大，此結果與Butterfield等的研究相符。

二、以本研究所採用的五個預測變數對文字系列完成測驗的鑑別指數進行多元

迴歸分析之後，所得到的標準化迴歸模式為：

$$Y=0.727X_{CNT}+0.165X_{REL}+0.038X_{ADJ}+0.054X_{EQU}-0.091X_{SI}+\varepsilon$$

但在多元迴歸分析中，所有預測變數共可解釋題目鑑別指數總變異量的78

%，而若採用Stepwise法進行多元迴歸分析，則系統僅保留預測變數CNT，

並可以解釋題目鑑別指數總變異量的74.7%，其他預測變數的影響力皆可被

忽略，故若將標準化迴歸模式調整如下或許將更適當：

$$Y=0.864X_{CN}+\varepsilon$$

三、採用「文字系列完成測驗虛擬題庫測驗系統」的試題重複率為0.0085172，充

份表現出該系統自動命題功能在試題曝光率方面的優勢所在。

四、本研究的效度指數偏低，與外在效標的相關係數也只有0.45，表現不如預期，

分析可能原因如下：

1. 題目規則的週期長度皆為3，受試者有可能在作答途中發現所有題目的週期長度皆一致，此發現將可能導致該受試者在測驗中的表現超出其原本能力。
2. 題目變化性太低，據施測老師表示，有許多受試者在測驗剛開始時相當專注，但因題目型太過於單調，題數又多達50題，故受試者在測驗中後段即開始分心，難以專注在測驗作答上，此現象將可能導致測驗結果出現誤差。

3. 題目難度偏低，由題目難度的描述統計可得知，系統中最難的題目也僅僅達到難易適中的程度，而最簡單的題目難度極低，完全喪失鑑別效果，觀察圖4-6-1可以發現，受試者的成績普遍偏高，並未呈常態分布，此狀況將造成統計結果出現誤差。

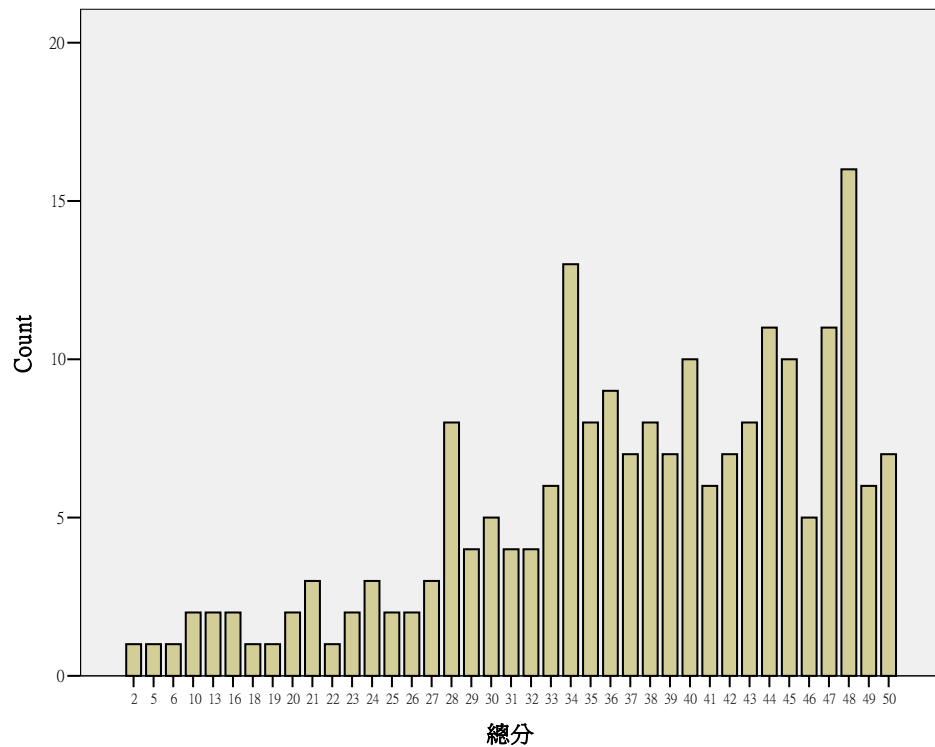


圖 4-6-1 文字系列完成測驗的受試者成績人數統計

4. 受試者作答態度差異相當大，施測老師發現每名受試者的配合度差異相當大，相對於專心進行測驗的受試者，也有許多受試者並不專心進行測驗，這將導致測驗結果與外在效標的相關程度受到影響。
5. 受試者同質性太高，本研究的受試者為台南新營高中的高中一、二年級學生，基於升學制度的影響，同一學校中的學生同質性極高，即能力極為相近，這將導致實驗結果與外在效標的相關係數大幅度降低。

五、不同性別受試者在「文字系列完成測驗虛擬題庫測驗系統」所獲得的成績無顯著差異，表示本系統所測量的推理能力在兩性之間沒有差異，但多因素性向測驗中抽象推理的結果卻出現性別差異的情形，推測可能原因是文字系列完成測驗與抽象推理測驗雖然同為測量非語文量數的推理能力，但實際測量到的能力部分仍有差異存在。