

第三章 研究設計與實施

第一節 研究設計方法

由於研究標的為紙筆測驗，且測驗性質屬總結性評量中的成就測驗。編製一份物理實驗評量紙筆測驗，當然首重測驗的品質，提供一份穩定且良好的測驗是首要任務。本研究之設計，斟酌研究對象抽樣的可行性，並加強研究標的測驗之內部一致性信度與測驗內容之內部一致性的效度，以求測驗結果的說服力。

一、研究對象

本研究以紙筆測驗為主，測驗對象為學習過現階段『高中物理實驗』的高中學生，由於現階段學習過高中物理實驗課程，以高二、高三的自然組學生為主，考量高二實驗課程內容過少，測驗編制困難，是以測驗對象選定為高三學生。

理論上標的母群體為全台灣所有高中三年級的學生。因為時間、經費的困難關係，考量地緣的因素，選擇可接近的母群體「大台北區的高中三年級的學生」。

斟酌經費、時間與取樣學校配合施測的困難性，樣本的選取採『分層隨機取樣』，依高中入學成績選出分屬高、中、低分群的學校各一所，其目的為考慮抽樣的全面性，配合研究者可掌控之學校即以「方便取樣」之原則，從大台北區的學校中選出分屬高、中、低分群的北一女中（A 班（組）25 人、B 班（組）32 人）、板橋高中（A 班（組）35 人、B 班（組）40 人）、泰山高中（A 班（組）35 人、B 班（組）31 人）三所學校，每所學校依「配額取樣」的方法，由各校選兩班高三學生接受 A、B 卷測試，總共有效測試樣本為測驗 A 卷 95 份、測驗 B 卷 103 份，共 198 份。

二、研究設計

設計一份實驗評量 A、B 兩種測驗卷（附錄四），學生基本資料填答及相關問卷一份（附錄二）。另外亦參考大考中心的研究資料（陳信雄，民 79），改編為一份教師問卷

(附錄三), 提供有興趣的讀者研究使用。

(一) 學生問卷設計：

本分問卷參考大考中心的研究資料(蔡尚芳等, 民 88), 修正編輯而成。主要設計考量為(1) 調查學生基本資料(2) 研究學生贊成物理考科命題改選擇題與其測驗成績的關係(3) 研究認真準備物理考科的學生與其測驗成績的關係(4) 研究學生重視動手實驗學生與其測驗成績的關係(5) 其他問卷資料與學生測驗成績的關係。

(二) 實驗測驗卷設計：

1. 設計原則：考量各校高三的課業因素，測驗時間約 50 分鐘一節課。並分析各校於高三上學期前，因為教科書版本的不同，上學期前所完成的實驗有異，從各版本的實驗教學共通部分，選取高二課程部分 9 個實驗，高三課程部分 1 個實驗。由於估計回答問卷與準備時間約 10 分，實驗評量測驗約為 40 分鐘，基於測驗樣本為高三學生及考試領導教學的考量，教師教學評量測驗必以『大學入學考試』為標竿，是以研究者設計測驗的標準，即以去年『指考』選擇 18 題、非選擇題 4 題考試時間為 80 分鐘為基準，設計一份選擇 6 題(包含單選與多重選擇)、非選擇 4 題的實驗評量測驗卷。

2. 設計原理：實驗評量測驗卷，設計為平行的 A、B 兩種測驗卷，針對每個學校兩個班級學生給予不同試卷同時測驗。A、B 兩種試卷試題測驗概念完全相同(附錄四)，同樣實施十題高中物理實驗屬『知識』、『理解』、『分析』等層次的概念測驗，包含(1) 完全相同的題目共 6 題，其中單一選擇題第 1、2 兩題、多重選擇題第 3、4 兩題、問答與說明題第 1、2 兩題，此六題(以下簡稱前測試題，分別簡稱前測 1、2、3、4、5、6 代號)作為檢測學生的程度。(2) 相同概念但不同題型的測驗共四題(以下簡稱後測試題)，即多重選擇第 5、6 兩題與問答與說明題第 3、4 兩題。其中相同實驗概念(金屬的比熱與靜力平衡實驗)測驗 A 卷的多重選擇第 5、6 題，在 B 卷中研究者將其設計為

問答與說明題第 3、4 題(以下簡稱後測 1 及 2 題);另外的相同實驗概念(量熱器水當量與共鳴空氣柱實驗)測驗 B 卷的多重選擇第 5、6 題,在 A 卷中研究者將其設計為問答與說明題第 3、4 題(以下簡稱後測 3 及 4 題)。

3. 設計目的:(1)藉前測試題檢測全體學生程度,並區分為高、中、低分群。(2)利用 A、B 卷的後測試題,檢測相同概念但不同題型的測驗,高、中、低分群的學生,表現是否有差異。(3)利用上述分析,找出『多重選擇題』與『問答與說明題』不同題型的優缺點。

三、研究方法

(一)問卷的統計分析:

1. 利用學生問卷,進行『問卷選項百分比的差異檢定』(即計算 χ^2 卡方值)以期瞭解問卷所顯示的統計資料是否有顯著差異的情形。
2. 利用學生問卷,進行『問卷選項單因子的前測成績變異數分析』(ONE-WAY ANOVA),若出現顯著差異則立即進行 Scheffe 考驗,以期瞭解各選項群學生的前測成績的差異出現在何處,其中各項檢定的顯著水準(Level of significance) 值取 0.05。

(二)測驗卷的統計分析:

1. 檢驗測驗的信度:

(1) 檢驗交互評分者的信度:檢驗測驗中的問答題在兩位評分者之間的一致性,利用 Pearson 積差相關係數檢驗。

(2) 檢驗內部一致性的信度係數,本研究採用折半信度係數之 Guttman 公式估計。

2. 檢驗測驗的效度:

(1) 對每個試題進行『選項分析』。

(2) 對每個試題進行難度指數、鑑別度分析。

(3) 檢驗內容效度：利用表 2-5.1 雙向細目表對教材內容與教學目標進行分析。

(4) 檢驗內部一致性的效度：分為 (a) 相關係數分析法：研究各題型分測驗與總測驗的相關係數，以檢驗測驗的效度 (b) 團體對照分析法：依學生的測驗總分，將學生分為高低兩群，利用兩群的測驗成績進行『統計平均分數 T-考驗』，若有顯著差異，表示試題具有高的內部一致性。

3. 檢驗前測試題：檢測學生是否均質及試題區分高中低分群的有效性。

(1) 檢驗 A、B 卷全體學生前測成績：利用平均分數統計 T-考驗。

(2) 檢驗前測成績高中低分群中 A、B 卷學生前測成績：利用前測成績區分全體學生為高中低分群，並比較各分群中 A、B 卷學生的前測成績，即平均分數統計 T-考驗。

4. 檢驗後測試題：檢測學生是否均質及 A、B 卷試題是否為有效的平行卷製作。

(1) 檢驗 A、B 卷全體學生後測成績：利用平均分數統計 T-考驗。

(2) 檢驗後測成績高中低分群中 A、B 卷學生後測成績：利用後測成績區分全體學生為高中低分群，並比較各分群中 A、B 卷學生的後測成績，即進行平均分數統計 T-考驗。

(3) 以前測成績為共變量檢驗後測成績：以前測成績為『共變量』(covariate) 對後測成績進行『共變數分析』(analysis of covariance, ANCOVA) 即排除前測成績的影響，對後測成績分析。程序為首先進行『組內迴歸係數同質性』(homogeneity of within-class regression coefficient) 考驗，若具同質性則可進行『共變數分析』，經調整後測成績後，分析學生在後測成績的表現是否有顯著差異。

(4) 檢驗前測成績高中低分群中 A、B 卷學生後測成績：利用前測成績將學生區分

為高中低分群，並比較各分群中 A、B 卷學生的後測成績，即進行平均分數統計 T-考驗。

5. 檢驗後測試題的題型差異：利用相同概念不同題型的測驗，分析題型的影響。

(1) 檢驗後測 1 及 2 題：兩題合併計分，並進行 (a) 全體 A、B 卷學生平均分數統計 T-考驗 (b) 前測高分群中 A、B 卷學生平均分數統計 T-考驗 (c) 前測中分群中 A、B 卷學生平均分數統計 T-考驗 (d) 前測低分群中 A、B 卷學生平均分數統計 T-考驗 (e) 以前測成績為共變量對 A、B 卷全體學生進行『共變數分析』 - 若經排除前測成績的影響因素後，有顯著差異，分析題型所造成的影響 (f) 分析題目在 A、B 卷中的得分率、難度指數、鑑別度，以瞭解題目與題型的優缺點。

(2) 檢驗後測 3 及 4 題：兩題合併計分，檢測步驟同 (1)。

(3) 若上述步驟無法分析出題型所造成的差異，則進行改變策略細部分析，並同時檢驗分析方法的可行性。初步擬定試驗的分析方法：(a) 更改分群策略 - 如改為高低兩群 (b) 更改檢驗的題數 - 如改為檢驗單一題目得分表現。

第二節 研究工具與實施

一、 學生問卷

(一) 問卷內容：此份問卷 (附錄二) 共 19 題，為使問卷與測驗卷選擇題部分填寫在同一張電腦卡上，問卷題號由 21 題編至 39 題。主要內容為學生基本資料第 21 30 及 34 題、學生在物理課業上的努力程度調查第 31、32 題、學生對物理課程的重視與態度調查第 33、35、36 題、命題方式意見調查第 37 38 題、此次測驗卷意見調查第 39 題。

(二)問卷施測：請於作答完測驗卷後配合施測，不限時間。配合教師發放試卷，講解『作答說明』及其他準備，與學生回答問卷時間估計共約需 10 分。

二、 實驗測驗卷

(一)測驗的準備：

測驗試題依據研究者近 3 年來，在高中教學經驗與教學評量中找出命題，並參酌相關文獻及大學聯考命題資料，經過研究者對試題的鑑別度、難易度、以及選項分析後，並考量研究的目的與雙向細目表，加以修改，期使文字更簡潔易懂，圖文能夠更清晰，經過研究者學校老師做初步的審查，同時在沈教授青嵩的指導下，完成整份試卷編輯。

(二)測驗的編製與內容：

本研究測驗的編製，依文獻探討中的步驟進行。

此份測驗卷(附錄四)，依據高中物理實驗課程，選取高二物質科學物理篇九個實驗，高三選修物理一個實驗，合計十個實驗，編成一份十題測驗題的試卷，每個實驗出一題，無論在單選(五個選項)、多選(五個選項)、問答題(五個單項問答)，皆在每個實驗題中選出五個概念考題命題，以確保整份實驗考題總結測驗的完整性，以及概念命題的均勻、配分的公平性。十題測驗題，分為單一選擇題 2 題、多重選擇題 4 題、問答與說明題 4 題。

其中『自由落體』、『二維空間的碰撞』實驗以單一選擇出題，『牛頓第二運動定律』、『波以耳定律』實驗以多重選擇命題，『斜面上的運動』、『水波的反射、折射、繞射與干涉』實驗以問答與說明題命題，以上六個實驗命為 A、B 卷的共同題目。另外『靜力平衡』、『共鳴空氣柱』兩個實驗先以多重選擇題方式出題；而『量熱器水當量』、『金屬的比熱』兩個實驗則以問答題方式命題。最後再將上面四個實驗題重新命

題，將多重選擇題轉換為問答題命題，而問答題則轉換為多重選擇題命題，而無論多選與問答皆配置五個選項或答案項。最後將上述四個實驗題分配為 A、B 卷：(1) 將兩個實驗題中多選題型的放入 A 卷的第 5、6 題中，並將其問答題型的編為 B 卷的問答第 3、4 題。(2) 將另外兩個實驗題中問答題型的放入 A 卷的問答第 3、4 題中，並將其多選題型的編為 B 卷的多選第 5、6 題。完成 A、B 平行測驗卷的設計。

(三) 測驗的實施：

本測驗卷共十題，考試時間為 40 分鐘。依據作答編號，將單一選擇題與多重選擇題共 6 題，依序劃記在電腦答案卡上；問答與說明題共 4 題則作答於非選擇題答案卷。

第三節 資料的收集

一、問卷資料收集

將電腦答案卡資料，利用讀卡機讀出並輸入電腦轉換為檔案以便分析，答案卡中第 21 至 39 題為問卷回答內容。本問卷可藉由第 21、30 及 34 題收集學生基本資料。而問卷中的第 31、33 及 35、39 題的資料，可以統計回答情形，以瞭解學生的概況與分析其與測驗成績的關係。其中第 21、30 題必然全校答案相同，而第 22 題為性別資料，故可藉由此三題瞭解學生填答問卷或測驗卷的態度，並找出無效樣本，予以剔除。

二、測驗卷資料收集與計分

(一) 選擇題資料收集與計分：將電腦答案卡中第 1 至 6 題測驗卷答案資料，利用讀卡機讀出並輸入電腦轉換為檔案，配分後統計各題分數與總分以便分析。

(二) 問答與說明題資料收集與計分：由兩位老師同時進行評分。事先兩位老師各抽樣 15 份評分，並分析受試者答題的各種可能情形，再討論一致性的給分方法，最後再進行重複閱卷評分，兩次評分若發現每大題總分 10 分中差距在 2 分以上再

進行三閱，最後評分取兩位教師評分的平均值。最後統計各題分數與相關總分並儲存為檔案。

(三) 合併上述(一)(二)選擇題與問答題的檔案資料與分數，以便分析。

第四節 試題的信度與效度分析

為確保測驗的可信度，必須對測驗試題進行分析與評鑑。

一、測驗的信度分析

本研究由於時間、人力的限制，未實施預試、重測與複本施測，所以信度的分析限於評分者信度與內部一致性信度，茲分述如下：

(一) 交互評分者信度 (interscorer reliability):

本測驗中含有四大題『問答與說明題』題型，為使測驗的分數，不因評分者不同，而造成不同的結果，所以研究者邀請另一位高中物理教師，進行第二次評分，並分析相關係數如表 3-4.1。

表 3-4.1 問答與說明題兩次評分相關係數

	題號	相關係數
A 卷 (N = 95)	1 (前測 5)	0.96**
	2 (前測 6)	0.99**
	3 (後測 3)	0.94**
	4 (後測 4)	0.95**
	總分	0.98**
B 卷 (N = 103)	1 (前測 5)	0.97**
	2 (前測 6)	0.99**
	3 (後測 3)	0.94**
	4 (後測 4)	0.94**
	總分	0.98**

**P<.01，即達 0.01 的顯著水準

評分方式：事先兩位老師各抽樣 15 份評分，從受試者答題的初步各種可能情形，討論可預期的一致性給分標準，再分別進行評分，兩次評分若發現每大題總分 10 分

中差距在 2 分以上再進行三閱，最後評分取兩位教師評分的平均值。

交互評分者信度可由表 3-4.1 知：每題測驗之相關係數皆在 0.94 以上，且達 0.01 顯著水準，故評分的信度相當高。研究者也由試卷難易度資料表（參考表 3-4.6、3-4.7、3-4.8）中發現，A 卷問答第 4 題及 B 卷問答第 3 題，因題目難度較高所以評分困難度增加，以致相關係數降低。

（二）內部一致性之信度係數：

由於測驗實施的困難，無法實施預試，研究者利用測驗後的樣本，進行信度分析。由於測驗的題數不多，在進行折半信度分析時，發現奇數題與偶數題的變異數並不相同，若使用史布公式（Spearman-Brown prophecy formula），會導致高估測驗的信度，即所估計的信度將會比其他內部一致性所估計的信度還高。是以研究者改採 Guttman(1945)所發明折半信度係數公式考驗，考驗結果如表 3-4.2。

測量標準誤（standard error of measurement）也是用以表示測驗可靠性的一種量數，其最簡便的計算方式是從測驗的信度係數加以估計，其公式如下： $S_m = S_x * (\sqrt{1 - R_x})$ 其中 S_m 為測量標準誤， S_x 為測量分數的標準差， R_x 為測驗的信度係數。

表 3-4.2 測驗卷信度資料

	選擇題答錯倒扣題分後總成績				選擇題答錯不倒扣題分後總成績			
	平均分數	標準差	Guttman 折半信度係數	測量標準誤	平均分數	標準差	Guttman 折半信度係數	測量標準誤
A 卷 (N=95)	37.00	14.10	0.49**	10.07	47.91	12.32	0.50**	8.71
B 卷 (N=103)	35.99	14.78	0.55**	9.91	46.98	13.00	0.54**	8.82

**P<.01，即達 0.01 的顯著水準；總分滿分為 100 分。

由表 3-4.2 知，即使研究者對選擇題計分採用倒扣題分方式，對測驗的信度影響並不大，而本研究測驗針對『入學考試』編製，是以此處的分析，支持研究者將計分方式採用倒扣題分的方法。

本測驗題數為 10 大題，但在測量實驗概念上可說是測量 50 個概念，雖然題數少不利於測驗的信度，但仍可由表 3-4.2 知：測驗之信度係數皆在 0.49 以上，且達 0.01 顯著水準，故仍為可靠之測驗。

二、試題的效度分析

在考驗測驗的效度上，本研究採用下列方法：

(一) 試題選項分析：

表 3-4.3 前測六題選項百分比與問答總分分析（樣本 198 份）

題號		未答	A	B	C	D	E	總分
1 (前測 1)	T	0	9.6	*32.3	12.1	12.6	33.3	
	H	0	7.6	71.2	6.1	3	12.1	
	L	0	12.7	6.3	14.3	20.6	46	
2 (前測 2)	T	1	2	13.6	*58.1	13.3	12	
	H	0	0	3	86.4	6.6	4	
	L	1.6	4.8	20.6	31.7	17.5	23.8	
*3 (前測 3)	T	0	*41.9	*69.2	72	*56.6	23	
	H	0	48.5	72.7	57.6	62.1	13.6	
	L	0	31.7	60.3	88.9	49.2	33.3	
*4 (前測 4)	T	1	*65.2	*71.7	66.2	*78.3	47.0	
	H	1.5	74.2	78.8	54.5	83.3	36.4	
	L	1.6	57.1	73.0	65.1	69.8	68.3	
問答 1 (前測 5)	T		1.03	0.5	1.48			3.01
	H		1.47	0.71	1.52			3.67
	L		0.52	0.27	1.48			2.27
問答 2 (前測 6)	T		1.17	1.85	1.28			4.30
	H		1.42	2.64	1.64			5.70
	L		0.76	0.84	0.70			2.30

註：1. 本表的分組係依上述題目總得分而為之。

2. 題號前有星號者為多重選擇題；前測 1、2 為單選題。

3. T、H、L 分表全部、高（前 33%）、低（後 33%）到考生之答案百分比。

4. 各答案百分比前有星號者為標準答案。

5. 各答案百分比 = 劃記該答案人數除以該組（T、H、L）全部到考人數。

6. 問答題的係以總得分列表，每題滿分為 10 分。

表 3-4.4 A 卷後測四題選項百分比與問答總分分析 (樣本 95 份)

題號		未答	A	B	C	D	E	總分
*5 (後測 1)	T	0	74.7	49.5	*91.6	*73.7	12.6	
	H	0	65.6	50.0	100.0	93.8	9.4	
	L	0	76.7	70.0	86.7	46.7	16.7	
*6 (後測 2)	T	0	8.4	46.3	*60	*93.7	*86.3	
	H	0	3.1	12.5	68.8	100.0	100.0	
	L	0	16.7	80.0	46.7	80.0	76.7	
問答 3 (後測 3)	T		1.65	0.86	0.17	0.22	0.24	3.14
	H		1.83	1.06	0.24	0.35	0.30	3.78
	L		1.30	0.67	0.10	0.11	0.09	2.27
問答 4 (後測 4)	T		0.33	0.49	0.87	0.36	0.56	2.61
	H		0.47	0.66	1.22	1.38	0.72	4.44
	L		0.03	0.17	0.33	0.07	0.27	0.87

註： 1. 本表的分組係依 A 卷十題題目總得分而為之。

2. 其餘設計方法如表 3-4.3。

表 3-4.5 B 卷後測四題選項百分比與問答總分分析 (樣本 95 份)

題號		未答	A	B	C	D	E	總分
*5 (後測 3)	T	0	*87.4	29.1	*80.6	*60.2	54.4	
	H	0	100.0	9.1	87.9	81.8	45.5	
	L	0	79.4	47.1	67.6	44.1	50.0	
*6 (後測 4)	T	0	*60.2	*69.9	*62.1	45.6	*66.0	
	H	0	66.7	87.9	87.9	18.2	75.8	
	L	0	55.9	47.1	41.2	73.5	50.0	
問答 3 (後測 1)	T		0.66	0.72	0.50	0.08	0.42	2.38
	H		0.82	0.82	0.88	0.26	0.70	3.48
	L		0.56	0.50	0.24	0.01	0.20	1.50
問答 4 (後測 2)	T		1.71	0.35	0.37	0.42	0.60	3.45
	H		2.00	0.58	0.64	0.79	0.91	4.91
	L		1.35	0.15	0.12	0.18	0.29	2.09

註： 1. 本表的分組係依 B 卷十題題目總得分而為之。 2. 其餘設計方法如表 3-4.3。

評鑑：(1) 各選擇題的誘答效用性：從表 3-4.3、3-4.4、3-4.5 可知，誘答的答案項皆

能使低分組選答它的人數多於高分組，而正答案皆為高分組選答它的人

數多於低分組，符合期待。

(2)問答題的效用性：由於問答題的評鑑較難，考量鑑別度與難度指數後再分析。

(二) 試題之鑑別度與難度指數分析：

表 3-4.6 前測六題鑑別度與難度指數分析 (樣本 198 份)

題號	答對率 PR(%)	高分組	低分組	難度		鑑別度	
		答對率 PH(%)	答對率 PL (%)	P		r	D
1(前測 1)	32.30	71.20	6.30	0.35	14.60	0.69	0.65
2(前測 2)	58.10	86.40	31.70	0.60	12.00	0.56	0.55
*3(前測 3)	8.90	24.80	-12.40	0.10	18.20	0.57	0.37
*4(前測 4)	21.00	38.50	7.00	0.21	16.20	0.45	0.32
問答 1(前測 5)	30.10	36.70	22.70	0.30	15.10	0.17	0.14
問答 2(前測 6)	43.00	57.00	23.00	0.39	14.10	0.36	0.34
平均	32.23	52.43	13.05	0.37	14.40	0.45	0.40

註： 1. 本表的分組係依上述題目總得分而為之。

2. 題號前有星號者為多重選擇題；前測 1、2 為單選題。

3. PR = 全體到考考生的答對率。

4. PH = 高分組 (前 33%) 考生的答對率。

5. PL = 低分組 (後 33%) 考生的答對率。

6. D = PH-PL 為全體到考考生的鑑別度。

7. 多重選擇題、問答題以得分率代替答對率

(多選) 得分率 = { (該題答對答案數 - 該題答錯答案數) / (該題答案數 * 考生人數) } * 100

(問答) 得分率 = 該題得分 / 該題總分

8. P 為難度指數， 為美國教育測驗服務社 (Educational Testing Service ETS) 訂定的標準難度指數分數表 ($\mu = 13 + 4Z$, Z 為標準 Z 分數值)，本轉換公式的平均難度為 13，而值愈大表難度愈高。

9. P、 μ 、r 可由范氏題目分析表經由 PH、PL 查得。

10. 當 PL 為負時，由於范氏題目分析表無法查得，取 PL 最低值 0.01 處理。

表 3-4.7 A 卷後測四題題鑑別度與難度指數分析 (樣本 95 份)

題號	答對率 PR(%)	高分組	低分組	難度		鑑別度	
		答對率 PH(%)	答對率 PL(%)	P		r	D
*5 (後測 1)	31.4	47.5	8	0.26	15.6	0.5	0.40
*6 (後測 2)	54.1	81.3	22.7	0.52	12.8	0.57	0.59
問答 3 (後測 3)	31.4	37.8	22.7	0.3	15.1	0.18	0.15
問答 4 (後測 4)	26.1	44.4	8.7	0.25	15.7	0.46	0.36
平均	33.8	51.2	15.6	0.35	14.60	0.40	0.36

註： 1. 本表的分組係依 A 卷十題題目總得分而為之。
2. 其餘設計方法如表 3-4.6。

表 3-4.8 B 卷後測四題鑑別度與難度指數分析 (樣本 103 份)

題號	答對率 PR(%)	高分組	低分組	難度		鑑別度	
		答對率 PH(%)	答對率 PL(%)	P		r	D
*5 (後測 3)	37.9	66.1	17.6	0.41	13.9	0.49	0.49
*6 (後測 4)	25	60	-11.8	0.23	15.9	0.76	0.72
問答 3 (後測 1)	23.8	34.8	15	0.24	15.8	0.26	0.20
問答 4 (後測 2)	34.5	49.1	20.9	0.34	14.6	0.31	0.28
平均	31.3	50.8	13.1	0.35	14.64	0.43	0.38

註： 1. 本表的分組係依 B 卷十題題目總得分而為之。
2. 其餘設計方法如表 3-4.6。

表 3-4.9 A、B 卷試題鑑別度與難度指數之分佈範圍分析

測驗	平均 分數	難度範圍				鑑別度範圍			
		P				r		D	
前測六題 (樣本 198 份)	32.09	0.21	0.60	12	16.2	0.17	0.69	0.14	0.65
		(0.37)		(14.4)		(0.45)		(0.40)	
A 卷後測四題 (樣本 95 份)	37.09	0.25	0.52	12.8	15.7	0.18	0.57	0.15	0.59
		(0.35)		(14.6)		(0.40)		(0.36)	
B 卷後測四題 (樣本 103 份)	32.23	0.24	0.41	13.9	15.8	0.26	0.49	0.20	0.72
		(0.35)		(14.6)		(0.43)		(0.38)	

註：括弧內之數字為平均數

評鑑：(1) 單一選擇題的分析：從表 3-4.6 可知，難度指數 P 與 、答對率 PR、鑑別度 D，皆相當良好。

(2) 多重選擇題分析：

(a) 由表 3-4.6 知，前測 6 題中多選第 3、4 題（前測 3、4）得分率分別為 8.9%、21%，唯鑑別度分別為 0.37、0.32 仍為良好試題，得分率偏低是因多重選擇題，研究者採用『大考中心』標準以倒扣方法計分所致。

(b) 由表 3-4.7 知，A 卷後測 4 題中多選第 5 題（後測 1）得分率為 31.4%，鑑別度為 0.40，屬難度不高的理想試題；A 卷後測 4 題中多選第 6 題（後測 2）得分率為 54.1%，鑑別度為 0.58，屬難度適中且鑑別度相當理想的試題。

(c) 由表 3-4.8 知，B 卷後測 4 題中多選第 5 題（後測 3）得分率為 37.9%，鑑別度為 0.49，屬難度不高且鑑別度相當理想的試題；B 卷後測 4 題中多選第 6 題（後測 4）得分率為 25%，鑑別度為 0.72，屬難度稍高但鑑別度相當理想的試題，造成此現象的因素可能為 B 卷低分群學生猜題現象明顯。

(d) 整體多選題，試題鑑別度皆為良好以上，雖然多選第 1 題（前測 3）得分率偏低，唯本試卷定位為『總結性成就測驗』，參照『大學入學考試』的標準，應可採用。

(3) 問答與說明題分析：

(a) 由表 3-4.6 知，前測 6 題中間答第 1 題（前測 5）得分率為 30.1%，鑑別度為 0.14，而難度指數 $P = 0.30$ 、 $n = 15$ ，雖難度不高但鑑別度不良。從表 3-4.3 選項分析中知，其 C 答案項，高分組與低分組得分相當接近且容

易得分，顯現該分項的題目鑑別力較差致降低整大題的鑑別度。但從全體試卷體察，該分項佔 2 分為整體比例的 2% 相當低，而為『總結性成就測驗』中基本試題，尚可接受。

- (b) 由表 3-4.6 知，前測 6 題中問答第 2 題（前測 6）得分率為 43%，鑑別度為 0.34，而難度指數 $P = 0.39$ 、 $\frac{1}{P} = 14.1$ ，屬難度適中且鑑別度良好試題。
- (c) 由表 3-4.7 知，A 卷中問答第 3 題（後測 3）得分率為 31.4%，鑑別度為 0.15，而難度指數 $P = 0.30$ 、 $\frac{1}{P} = 15.1$ ，雖難度不高但鑑別度不良。從表 3-4.4 可知，此題的 C、D、E 答案項得分偏低，可知該部分題目難度高造成整體鑑別度降低，唯 A 答案項得分容易，致總得分平均 3.14 分尚可。因為本題為後測試題，主要為檢定功能，可以從第 4 章分析中知其為『可再修正的題目』，留待第 4 章分析。
- (d) 由表 3-4.7 知，A 卷中問答第 4 題（後測 4）得分率為 26.1%，鑑別度為 0.36，而難度指數 $P = 0.25$ 、 $\frac{1}{P} = 15.7$ ，難度稍高但鑑別度良好。從表 3-4.4 知，此題中各答案項得分相當均勻，可知該部分題目難度一致，但總得分平均 2.61 分（低於全部得分的平均比率）稍低，其中高分組的得分平均為 4.44 分，顯現該題的題目鑑別度良好。因為本題為後測試題，主要為檢定功能，留待第 4 章分析。
- (e) 由表 3-4.8 知，B 卷中問答第 3 題（後測 1）得分率為 23.8%，鑑別度為 0.20，而難度指數 $P = 0.24$ 、 $\frac{1}{P} = 15.8$ ，難度稍高且鑑別度尚可。從表 3-4.5 知，此題的 D 答案項得分偏低，可知該部分題目難度高，致總得分降低，而 A、B 答案項則鑑別度低，使整題鑑別度降低，但並非不良試題。因為本題為後測試題，主要為檢定功能，可以從第 4 章分析中知其為『可再修

正的題目』，留待第 4 章分析。

(f) 由表 3-4.8 知，B 卷中間答第 4 題得分率（後測 2）為 34.5%，鑑別度為 0.28，而難度指數 $P = 0.34 = 14.6$ ，難度適中且鑑別度尚可。從表 3-4.5 知，此題因 B 答案項得分不高且鑑別度稍低所致，其餘大致良好。因為本題為後測試題，主要為檢定功能，留待第 4 章分析。

(4) 整份試題分析：由表 3-4.9 知，整份試卷的 P、 α 、D 值的分佈範圍，若以『大學入學考試』的標準衡量，在平均分數的可接受度上，應屬良好試卷。

(三) 試題的內容效度研究：

本研究考量時間、經費的問題，對試題的內容效度研究，僅由兩位高中教師針對雙向細目表做初步的教材內容與教學目標檢測。

由本測驗的雙向細目表(表 2-5.1)知(1)對教材內容的代表性樣本方面：共測驗 10 個實驗且每個實驗皆佔 10%，涵蓋全部教材且達到內容的平均分配
(2)對教學目標的代表性樣本方面：其測量目標層次分別(a)知識層次佔 26%、理解層次佔 50%、分析層次佔 24%，與『入學考試』相當接近，符合研究目的(b)有 8 個(佔 80%)實驗測量 3 個層次，有 2 個(佔 20%)實驗測量較低的『知識』與『理解』兩個層次，亦符合對實驗內容、研究目的的要求。

(四) 內部一致性的效度分析：

1、相關係數分析法：

表 3-4.10 前測各單項測驗題分與總分相關係數(樣本 198 份)

	單選題 總分	多選題 總分	問答題 總分	前測六題 總分
單選題總分		0.04	0.11	0.76**
多選題總分	0.04		0.23**	0.52**
問答題總分	0.11	0.23**		0.57**
前測六題總分	0.76**	0.52**	0.57**	

** $P < .01$ ，即達 0.01 的顯著水準。

表 3-4.11 A 卷各單項測驗題分與總分間相關係數 (樣本 95 份)

	前測單 選總分	前測多 選總分	後測多 選總分	多選 4 題總分	前測問 答總分	後測問 答總分	問答 4 題總分	前測 6 題總分	後測 4 題總分	A 卷 總分
前測單 選總分		0.02	0.03	0.03	0.06	0.13	0.10	0.74**	0.07	0.52**
前測多 選總分	0.02		0.26**	0.73**	0.33**	0.33**	0.38**	0.58**	0.34**	0.58**
後測多 選總分	0.03	0.26**		0.82**	0.36**	0.39**	0.43**	0.27**	0.92**	0.66**
多選 4 題總分	0.03	0.73**	0.82**		0.42**	0.46**	0.51**	0.51**	0.82**	0.80**
前測問 答總分	0.06	0.33**	0.36**	0.42**		0.49**	0.88**	0.56**	0.48**	0.62**
後測問 答總分	0.13	0.33**	0.39**	0.46**	0.49**		0.85**	0.42**	0.72**	0.66**
問答 4 題總分	0.10**	0.38**	0.43**	0.51**	0.88**	0.85**		0.57**	0.68**	0.74**
前測 6 題總分	0.74**	0.58**	0.27**	0.51**	0.56**	0.42**	0.57**		0.38**	0.87**
後測 4 題總分	0.07	0.34**	0.92**	0.82**	0.48**	0.72**	0.68**	0.38**		0.78**
A 卷 總分	0.52**	0.58**	0.66**	0.80**	0.62**	0.66**	0.74**	0.87**	0.52**	

** $P < .01$ (2-tailed test), 即達 0.01 的顯著水準。

表 3-4.12 B 卷各單項題分與總分間相關係數 (樣本 103 份)

	前測單 選總分	前測多 選總分	後測多 選總分	多選 4 題總分	前測問 答總分	後測問 答總分	問答 4 題總分	前測 6 題總分	後測 4 題總分	B 卷 總分
前測單 選總分		0.03	0.11	0.04	0.16	0.19	0.19*	0.78**	0.16	0.56**
前測多 選總分	0.03		0.26**	0.67**	0.13	0.12	0.14	0.45**	0.24**	0.43**
後測多 選總分	0.11	0.26**		0.85**	0.35**	0.43**	0.43**	0.33**	0.94**	0.73**
多選 4 題總分	0.04	0.67**	0.85**		0.33**	0.37**	0.39**	0.45**	0.80**	0.77**
前測問 答總分	0.16	0.13	0.35**	0.33**		0.57**	0.91**	0.58**	0.49**	0.64**
後測問 答總分	0.19	0.12	0.43**	0.37**	0.57**		0.86**	0.42**	0.72**	0.65**
問答 4 題總分	0.19*	0.14	0.43**	0.39**	0.91**	0.86**		0.58**	0.66**	0.73**
前測 6 題總分	0.78**	0.45**	0.33**	0.45**	0.58**	0.42**	0.58**		0.42**	0.85**
後測 4 題總分	0.16	0.24**	0.94**	0.80**	0.49**	0.72**	0.66**	0.42**		0.81**
B 卷 總分	0.56**	0.43**	0.73**	0.77**	0.64**	0.65**	0.73**	0.85**	0.81**	

** $P < .01$, 即達 0.01 的顯著水準 ; * $P < .05$, 即達 0.05 的顯著水準。(2-tailed test)

評鑑 :

- (1) 由表 3-4.10 知 , 前測試題中各單項題型分數與總分之間之相關係數不但達到顯著水準 , 且係數皆在 0.52 以上。而單選題型與各題型間未達顯著水準 , 應為單選題數不多 , 且研究者以『入學考試』為標準 , 答錯倒扣所造成的干擾 , 以及各題型所測量的實驗概念不相同所致。
- (2) 由表 3-4.11 知 , A 卷的後測試題中各單項題型分數與總分之間之相關係數不但達到顯著水準 , 且多相當高。而單選題型與各題型間未達顯著水準 , 應為單選

題數不多，且研究者以『入學考試』為標準，答錯倒扣所造成的干擾，以及各題型所測量的實驗概念不相同所致。而多選題型與問答題型之間亦達到顯著水準，但係數較低，亦為各題型所測量的實驗概念不相同所致。

- (3) 由表 3-4.12 知，B 卷的後測試題中各單項題型分數與總分之間之相關係數不但達到顯著水準，且多相當高。而單選題型與各題型間未達顯著水準，應為單選題數不多，且研究者以『入學考試』為標準，答錯倒扣所造成的干擾，以及各題型所測量的實驗概念不相同所致。而前測多選兩題總分與問答題間亦未達顯著相關，其道理同單選題。而多選題型與問答題型之間亦達到顯著水準，但係數較低，為各題型所測量的實驗概念不相同所致。
- (4) 由表 3-4.10、3-4.11、3-4.12 知整體而言，試題中各單項題型分數與總分之間之相關係數不但達到顯著水準，且多相當高。其測驗內部一致性相當高，故測驗的內部一致性效度應相當可靠。

2、團體對照法：

這種方法與試題的鑑別度分析法類似，對測驗內容的同質性能夠提供部分證據。

由於本研究對答題後做倒扣題分的修正，所以在此僅提供群組試題的考驗分析。方法為依測驗成績高低，分成高低兩組，然後比較這兩組學生在每個群組試題上得分表現，經過統計 T - 考驗後，若有顯著差異，表示試題具有較高的內部一致性。

由表 3-4.13、3-4.14、3-4.15，無論為前測試題高低分組或 A 卷、B 卷成績高低分組，其在各群組試題中，高分組的學生平均分數皆高於低分組的學生，且經統計 T-考驗後達 ($P < .01$) 顯著水準。可知各群組試題，其測驗的內部一致性效度應相當可靠。

表 3-4.13 前測 A、B 卷高低分組平均分數統計

以前測 A、B 卷平均分數 區分為高(H=100)低(L=98)分組統計					Levene 變異數 同質性考驗			平均分數差異 統計 T-考驗		
題型	分群	平均分數	標準差	平均數 標準誤		F 值	P 值	t-test	df 自由度	P 值 2-tailed
前測 單選 總分	高分組	1.1800	0.7231	0.0723	變異數 相等	131.039	0.000	9.794	196.000	0.000
	低分組	0.3750	0.3769	0.0381	變異數 不相等			9.851	149.763	0.000
前測 多選 總分	高分組	0.6220	0.4263	0.0426	變異數 相等	16.209	0.000	7.994	196.000	0.000
	低分組	0.2000	0.3053	0.0308	變異數 不相等			8.021	179.542	0.000
前測 問答 總分	高分組	0.9170	0.3559	0.0356	變異數 相等	0.573	0.450	7.875	196.000	0.000
	低分組	0.5408	0.3145	0.0318	變異數 不相等			7.885	193.950	0.000
後測 1 及 2 總分	高分組	0.8420	0.4755	0.0475	變異數 相等	0.017	0.895	3.525	196.000	0.001
	低分組	0.6020	0.4823	0.0487	變異數 不相等			3.525	195.768	0.001
後測 3 及 4 總分	高分組	0.8100	0.5338	0.0534	變異數 相等	1.393	0.239	4.332	196.000	0.000
	低分組	0.5051	0.4523	0.0457	變異數 不相等			4.339	192.015	0.000
前測 6 題 總分	高分組	45.317	11.1305	1.1130	變異數 相等	28.916	0.000	20.433	196.000	0.000
	低分組	18.597	6.6772	0.6745	變異數 不相等			20.531	162.672	0.000
後測 4 題 總分	高分組	41.300	19.2192	1.9219	變異數 相等	0.358	0.550	4.997	196.000	0.000
	低分組	27.679	19.1314	1.9326	變異數 不相等			4.998	195.952	0.000
全卷 總分	高分組	42.990	11.5706	1.1571	變異數 相等	1.042	0.309	14.158	196.000	0.000
	低分組	21.148	10.0699	1.0172	變異數 不相等			14.177	193.313	0.000

$P < .01$ ，達顯著差異水準。

本表各單項總分皆以原始分數表示，滿分為 2 分。

而前測 6 題、後測 4 題、全卷總分則換算為 100 分滿分。

表 3-4.14 A 卷成績高低分組平均分數統計

以 A 卷平均分數 區分為高(H=47)低(L=48)分組統計					Levene 變異數 同質性考驗			平均分數差異 統計 T-考驗		
題型	分群	平均 分數	標準差	平均數 標準誤		F 值	P 值	t-test	df 自由度	P 值 2-tailed
前測 單選 總分	高分組	1.0266	0.7961	0.1161	變異數 相等	21.575	0.000	3.977	93.000	0.000
	低分組	0.4896	0.4864	0.0702	變異數 不相等			3.957	75.855	0.000
前測 多選 總分	高分組	0.6723	0.4179	0.0610	變異數 相等	0.262	0.610	4.945	93.000	0.000
	低分組	0.2667	0.3811	0.0550	變異數 不相等			4.941	91.825	0.000
後測 多選 總分	高分組	1.2596	0.4790	0.0699	變異數 相等	1.112	0.294	8.114	93.000	0.000
	低分組	0.5000	0.4327	0.0625	變異數 不相等			8.105	91.627	0.000
多選 4 題 總分	高分組	1.9106	0.6799	0.0992	變異數 相等	0.662	0.418	9.662	93.000	0.000
	低分組	0.6583	0.5805	0.0838	變異數 不相等			9.646	90.149	0.000
前測 問答 總分	高分組	0.8809	0.3055	0.0446	變異數 相等	0.420	0.518	5.615	93.000	0.000
	低分組	0.5125	0.3330	0.0481	變異數 不相等			5.620	92.610	0.000
後測 問答 總分	高分組	0.7979	0.2847	0.0415	變異數 相等	0.004	0.948	6.790	93.000	0.000
	低分組	0.4208	0.2560	0.0369	變異數 不相等			6.783	91.516	0.000
問答 4 題 總分	高分組	1.6787	0.4695	0.0685	變異數 相等	0.790	0.376	7.650	93.000	0.000
	低分組	0.9333	0.4800	0.0693	變異數 不相等			7.652	93.000	0.000
前測 6 題 總分	高分組	42.997	14.043	2.0484	變異數 相等	4.832	0.030	8.565	93.000	0.000
	低分組	21.145	10.624	1.5335	變異數 不相等			8.540	85.672	0.000
後測 4 題 總分	高分組	51.436	13.092	1.9097	變異數 相等	0.045	0.832	10.277	93.000	0.000
	低分組	23.021	13.838	1.9973	變異數 不相等			10.283	92.892	0.000
全卷 總分	高分組	46.160	8.2337	1.2010	變異數 相等	0.990	0.322	14.612	93.000	0.000
	低分組	20.813	8.6625	1.2503	變異數 不相等			14.620	92.919	0.000

$P < .01$ ，達顯著差異水準。

本表各單項總分皆以原始分數表示，滿分為 2 分。

而前測 6 題、後測 4 題、全卷總分則換算為 100 分滿分。

表 3-4.15 B 卷成績高低分組平均分數統計

以 B 卷平均分數 區分為高(H=53)低(L=50)分組統計					Levene 變異數 同質性考驗			平均分數差異 統計 T-考驗		
題型 題號	分群	平均 分數	標準差	平均數 標準誤		F 值	P 值	t-test	df 自由度	P 值 2-tailed
前測 單選 總分	高分組	1.1226	0.7315	0.1005	變異數 相等	21.078	0.000	5.300	101.000	0.000
	低分組	0.4700	0.4861	0.0687	變異數 不相等			5.360	90.932	0.000
前測 多選 總分	高分組	0.4868	0.4394	0.0604	變異數 相等	4.348	0.040	3.368	101.000	0.001
	低分組	0.2320	0.3139	0.0444	變異數 不相等			3.401	94.218	0.001
後測 多選 總分	高分組	1.1019	0.6021	0.0827	變異數 相等	11.082	0.001	8.307	101.000	0.000
	低分組	0.2880	0.3526	0.0499	變異數 不相等			8.428	84.793	0.000
多選 4 題 總分	高分組	1.5132	0.8589	0.1180	變異數 相等	15.618	0.000	8.217	101.000	0.000
	低分組	0.3680	0.4971	0.0703	變異數 不相等			8.339	84.216	0.000
前測 問答 總分	高分組	0.9623	0.3548	0.0487	變異數 相等	0.019	0.890	6.036	101.000	0.000
	低分組	0.5540	0.3303	0.0467	變異數 不相等			6.049	100.984	0.000
後測 問答 總分	高分組	0.7472	0.3010	0.0413	變異數 相等	3.043	0.084	6.301	101.000	0.000
	低分組	0.4080	0.2398	0.0339	變異數 不相等			6.343	98.284	0.000
問答 4 題 總分	高分組	1.7094	0.5443	0.0748	變異數 相等	0.314	0.576	7.329	101.000	0.000
	低分組	0.9620	0.4869	0.0689	變異數 不相等			7.353	100.724	0.000
前測 6 題 總分	高分組	42.862	13.820	1.8983	變異數 相等	11.009	0.001	9.503	101.000	0.000
	低分組	20.933	8.9273	1.2625	變異數 不相等			9.618	89.576	0.000
後測 4 題 總分	高分組	46.226	18.310	2.5150	變異數 相等	13.950	0.000	9.762	101.000	0.000
	低分組	17.400	10.326	1.4603	變異數 不相等			9.912	82.964	0.000
全卷 總分	高分組	43.453	10.361	1.4232	變異數 相等	6.649	0.011	14.615	101.000	0.000
	低分組	18.000	6.8490	0.9686	變異數 不相等			14.785	90.679	0.000

$P < .01$ ，達顯著差異水準

本表各單項總分皆以原始分數表示，滿分為 2 分。

而前測 6 題、後測 4 題、全卷總分則換算為 100 分滿分。