

第一章 緒論

第一節 研究背景與動機

大學入學考試為各大學取才的管道之一，近十餘年來，教改聲浪風起雲湧，統一的考試招生制度「聯招」也面臨改變，逐漸脫離「考招合一」的傳統模式，學生能否進入大學深造，不再全由每年一次、各校統一的「聯招」考試成績決定。主張「考招分離」的方式，即考試由「大學入學考試中心」統籌辦理，而招生則由「大學招生策進會」負責。「大考中心」並於民國 81 年提出建議：將大學入學方式多元化，兼採「改良式聯招」、「推薦甄選」、「預修甄試」等管道並行，將考試分為「學科能力測驗」與「指定科目考試」兩部分。

由於考試領導教學，是以入學考試的方式及命題須兼顧其對教育的影響。為了引導正常教學，鼓勵學生腳踏實地的學習，而不是一味的練習解題，評量基本能力（知識、概念、理解、分析與應用）是必須的。物理學科可說是一門實驗科學，如何在教學中加強學生的實驗能力、實驗技能，並培養學生正確的科學精神與態度，是全體教育工作者的任務。而如何能了解學生這一方面的能力，評量是一種必要的手段，如何藉由評量的方法與改進，來達到教學的目的，就是教育工作者研究的重點，亦是「大學入學考試中心」成立的宗旨。

近年來常見，「高中聯招」或「國中基本學力測驗」在自然學科的命題上，實驗題的命題數常佔百分之五十以上的份量，題目設計的目的：「希望能引導學校重視實驗教學。」而看到近年來坊間補習班也陸續開始設置實驗室，讓補習的學生親自做實驗，實地學習自然科學，這些發展都值得欣慰。當然「高中物理教學」與「大學入學考試」層次上與「國中基本學力測驗」有差異，要達到「物理實驗命題份量超出全部命題百分之五十以上」有相當困難，唯大多數高中教師仍希望「加重實驗部分的評量」以確實反映

物理科學的本質。

長期來考試領導教學的結果，教師的教學必須重視『大學入學考試』，而大型的入學考試，又多採用『紙筆測驗』。而物理學科為一門思考性的學科，長期來在大學入學考試中命題常偏難，題目難度高，被定位為「投資報酬率低」的學科，例如 91 年指定科目考試，高標為 30 分，均標為 17 分，低標為 5 分，除扼殺學生的學習興趣外，也間接造成物理教師在教學上的困饒。如何提高學生學習興趣與學習成就，就是教師的重點任務。是以大學入學考試的測驗與教師的教學評量如何改進，為當前物理科學教育相當重要的一環。

目前『大學入學考試』仍是以紙筆測驗為主要方式，是以學校的教學評量也以『紙筆測驗』為主。近年來『入學考試』的紙筆測驗可分為選擇式測驗題的單選題、多重選擇題，以及非選擇式測驗題的填充題、問答與說明題、計算題等。在物理實驗測驗命題方面大多為單選、多重選以及問答與說明題。

然而『問答與說明題』的測驗，必須進行人工閱卷，其所需的人力與物力，試務龐雜且規模龐大，評分主觀且困難，容易產生給分不一致之主觀偏差，以致測驗目標的可靠性常招人質疑，是促成研究者對『以選擇題取代問答題測驗的可行性研究』產生動機。

物理科的入學考試，是否維持往年的『選擇與非選擇的方式』或『選擇與非選擇的配分各為何』，經常在教育學者與物理科學專家間廣泛的討論，最明顯的變化即 91 年的大學入學指定科目考試將『選擇與非選擇配分』由各為 50 分調整為 80 分及 20 分，而 92 年則在全球傳染病 SARS（非典型肺炎）疫情的影響下進一步改為全部選擇題，此亦引發各界與學生的議論。然而將非選擇題改為選擇題測驗，對未來高中物理教學的所造成的影響為何？而測驗是否達成學生能力評量測驗的原始目的等，皆是研究者對相關問題有興趣之處，也值得高中教師一同研究與討論。

第二節 研究目的

物理科學由於係實驗科學，在教學後的評量：實驗技能、科學精神與態度的評量不可忽略。而從往年的大學聯考中，發現問答題式的實驗題，往往鑑別度相當低，有違評量的目的；而選擇式測驗題的實驗題，又不容易出題且常招詬病，無法測出學生的實驗技能、科學精神與態度。物理本是一門實驗科學，是以如何落實物理實驗考題，以測驗出學生的實驗能力當然是必須的且是重要的。

實驗科學能力的評量，當前最熱門也被大多數專家學者認定較有效的為『全真評量』或『實作評量』，這類評量在學校教學中採用有其可行性及好的教學效果，但大型的『入學考試』欲採用有其困難度不易克服。當『紙筆測驗』無可避免，且問答與說明題對評量學生的物理實驗各種技能亦非相當有效時，對『紙筆測驗』中的各種題型優缺點分析及相關的命題技術的發展，就顯得相當重要。

國內有關物理教學評量的相關研究相當少，針對物理實驗評量的研究更少，然大考的物理考科今年已變更為全部選擇題型，如何評鑑此類測驗的優缺做為高中物理教師的指引不可或缺，也是研究者試圖嚐試的目標。

研究者試圖努力設計一份物理實驗評量題目（包括選擇式測驗題與問答題），藉以分析出選擇式測驗題與問答題，在各種學生程度中的答題狀況，並分析各種題型在測驗上的優缺點，找出試題測驗目標的關鍵因素，及試題鑑別學生能力高低的重要因子，進而找出選擇式測驗題與問答題互相取代的可行性，當然研究者感興趣的是「用選擇式測驗題取代問答題的可行性」。

本研究希望藉由相關文獻的分析與蒐集，探討高中物理實驗評量的方式，並分析紙筆測驗題型的不同，研究測驗對『評量學生的學習成就與能力指標的檢定』的優缺點，藉由測驗的結果與檢定方法的討論，提供高中教師處理教學評量時的參考。

第三節 研究的假定與限制

一、 研究的假定：

- (一) 一個有效的客觀式測驗試題，應能夠成功的做題型轉換。
- (二) 一個有效的客觀式測驗試題，應能夠測量認知領域中的『知識』、『理解』、『應用』和『分析』層次的學習結果。
- (三) 學生在本研究的『問答與說明題』測驗的行為表現，能完全代表其學習成就，而與語文能力無關。

二、 待答問題：

- (一) 經過『前測』後，檢定為程度相同的學生，其在『後測』時測驗相同概念不同題型的試題時，其在『多重選擇題』與『問答與說明題』的成績表現是否顯著差異？
- (二) 經過『前測』後，檢定為程度相同的學生，其在『後測』時測驗相同概念不同題型的試題時，可否發現試題的鑑別度是否產生顯著變化？
- (三) 不同測驗成績的學生，贊同『物理考科全部利用選擇題方式命題』的比例為何？不同意見的學生其成績的表現有否差異？
- (四) 不同測驗成績的學生，其『在校外加強課業』與『準備物理考科時間』的比例為何？不同情況的學生其成績的表現有否差異？
- (五) 不同測驗成績的學生，認同『確實動手做過實驗』與『實驗是物理科學最重要的一環』的比例為何？不同看法的學生其成績的表現有否差異？

三、 研究的範圍與限制：

- (一) 本研究以大台北區公立高中高三自然組學生為對象，實施物理實驗客觀式紙筆測驗，研究推論應與本研究設計類似的學校及學生與高中物理實驗課

程測驗為範圍。

- (二) 本研究以實施高中物理實驗客觀式紙筆測驗，分析『多重選擇題』與『問答與說明題』的取代性，研究推論應限制在此範圍，不宜過度推論到其他的題型測驗與其他不同的學科。

第四節 名詞解釋

1. 總結性評量：首要目標是給學生評定成績，或為學生做證明、或者是評定教學方案的有效性、或課程的有效性（邱淵譯；教學評量，頁 111 至 112）。在某一教學單元、課程或學期之末，就學生們的學習成就進行評量，決定其成績的等第、及格與否（簡茂發，民 85）。
2. 形成性評量：主要的目的不是給學習者評定成績或做證明，而是既幫助學習者也幫助教師把注意力集中在達到精熟程度所必須具備的特定知識上。其功能為提供回饋訊息，以期增強學習效果，即評量結果可指出「有哪些單元知識還沒有達精熟程度」。而這種評量並不包括評等第。（邱淵譯；教學評量，頁 112）
3. 診斷性評量：為了使教學適合學習者的需要和背景所設計的一種評量，評量的功能為鑑定造成學習困難的因素，目的為設計一種可以排除學習障礙的教學方案。（邱淵譯；教學評量，頁 189 至 190）
4. 安置性評量（或預備性評量）：為教師在進行單元教學之前，先給學生實施的教學評量，為一種以「了解學生之基本能力與起點行為」所設計的評量，以做為教師設計教材內容及決定教學方法的參考。（張景媛；教學評量與教學輔導）
5. 操作評量：適用於儀器使用、標本辨識學習的一種評量，為一種確定目標，非語言文字所能區辨的能力測試。（簡茂發、李虎雄等著，民 88）

6. 專題報告：針對一主題提出說明包括動機、問題闡述、策略、執行結果、提出解釋、發現新問題等所作的報告，可以了解撰寫者規劃、執行、研究等各方面的能力。(簡茂發、李虎雄等著，民 88)
7. 卷宗評量 (portfolio assesment)(或檔案評量)：針對某主題，蒐集相關資料，整理分析、批判後展示其成果。為一種「對問題的認識程度，對變因的重要性認識、能作有層次有組織的表達能力」的評量。常用在社會科學方面的學習，或調查報告、實驗報告方面的評鑑。(簡茂發、李虎雄等著，民 88)
8. 進階式評量 (Hierarchy assessment)：為利用電腦作滋生式的評測，以思考的可能路徑，沿路徑設計問題，達到評測思考的歷程、行為的動機等評量的目的。(簡茂發、李虎雄等著，民 88)
9. 實作評量 (performance assessment)：將察覺問題、形成問題、提出解決問題策略、進行規劃、執行以及資料的整理、統計分析、歸納研判、提出論點解釋資料、應用推廣等整個探索問題的過程列入評量的範圍，這種評量稱為實作評量。為一種針對數學、自然科學的學習方面所作的評量，以解決一個小問題為題，一邊處理問題一邊記錄，以評測「知識、知識的應用、規劃執行、建構概念、推理運用」的能力。(簡茂發、李虎雄等著，民 88)
10. 常模參照評量：以常態分配曲線為理論，將學生的學習評量結果，以在團體中的相對位置描述其學習成就。
11. 標準參照評量：以「標準分數」為理論，將學生的學習評量結果，與某一參照點比較，而求得其在團體中佔有的地位，以描述其學習成就。
12. 教育統計學之名詞解釋(余民寧，民 86；陳英豪、吳裕益，民 87；簡茂發，民 82；張紹勳，民 89；林清山，民 81；全士豪，民 65；王文科，民 91)：

(1) 虛無假設 (null hypotheses): 為以虛無形式呈現的假設, 與其敘述變項間『沒有差異』、『沒有影響』、或『沒有關係』。當以推論統計考驗統計假設時, 即為考驗虛無假設。虛無假設所敘述的, 與實驗者預測的或期待的, 恰好相反。

(2) 顯著性考驗 (test of significance): 根據某項預定水準來拒絕或接受虛無假設, 此預定的水準即為顯著水準 (level of significanc)(通常以符號 α 表示)。一般選定的顯著水準, 即在決定第一類型 (type I error) 錯誤的機率。

在 $\alpha=0.05$ 的顯著水準下拒絕了虛無假設, 即表示該項差異有 95% 的機會是由實驗造成的, 其由抽樣誤差所造成的機率低於 5%; 亦即拒絕以抽樣誤差做解釋的危險性為 5%。

研究者在獲致結果之前所訂的為 α 顯著水準, 根據計算而得的結果決定的為 P 值 (probability value P) 水準。

(3) 自由度 (degree of freedom, 縮寫 df): 在一項分配中, 自由度的數目係指彼此能獨立的觀察數值的數目。

(4) 積差相關係數 (product moment correlation coefficient) 通稱為 Pearson r : 由英國統計學家 Karl Pearson 所倡。(公式略, 請參考林清山, 民 81, 頁 130)。理論上 r 值介於 +1.00 與 -1.00 之間, 當 (a) $r > 0$ 稱正相關 (b) $r < 0$ 稱負相關 (c) $r = 0$ 稱零相關 (d) $r = 1$ 稱完全正相關 (e) $r = -1$ 稱完全負相關。

當求出的 r 值大於查表所得的 r_{α} 顯著性臨界值時, 即表示兩樣本在顯著水準下具有相關性。

(5) Guttman(1945) 所發明折半信度係數: 用以計算測驗的內部一致性的係數。Guttman 的公式不必假定兩半測驗分數的變異量必須相等。

其計算公式如下:
$$r_{tt} = 2\{S_x^2 - (S_o^2 + S_e^2)\} / S_x^2$$

其中 S_x^2 : 測驗總分之變異數, S_o^2 : 奇數題分數之變異數, S_e^2 : 偶數題分數之變異數。

Guttman 折半信度係數的顯著水準判斷方法, 可參考 (5) 的 Pearson 相關係數顯著性判斷法。

(6) 兩個獨立樣本之變異數 (variance) 的差異顯著性考驗 :

() F-考驗 (F-test) : 決定樣本是否符合變異數相等的規準。

$$F \text{ 值計算公式 : } F = \frac{S_{x1}^2}{S_{x2}^2} \text{ (即兩個變異數的比值) }。$$

其中 S_{x1}^2 和 S_{x2}^2 表示兩個所欲考驗樣本的變異數。

() 當決定考驗的顯著水準 時, 並根據自由度查出 F 分配表 (F-distribution) 的左、右兩端 F 臨界值, 若 F 計算值介於 F 分配表中之左、右兩端臨界值間時, 虛無假設成立。亦即, 當由計算公式所求得的 F 值, 在 F 分配中出現的機率 (即 P 值) 在信賴水準 (confidence level, 可用 $1 - \alpha$ 計算) 內 (即 $P > \alpha$ 值), 即兩個變異數是同質 (homogeneity of variance); 若 F 計算值, 落於 F 分配表中的信賴區間 (confidence interval) 外, 即此時 P 值小於顯著水準, 則兩個變異數不同質。

(7) 兩個獨立樣本平均數的差異顯著性考驗 :

() 當兩個變異數同質時, 此兩平均數差異檢定可採用 t-test

(students' t-test) 其計算公式如下 :
$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p^2 \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}}$$

() 當兩個變異數不同質時, 此兩平均數差異檢定, 須用無母數考驗, 以柯克蘭和柯克斯 (Cochran & Cox, 1957) 所發展的 t 考驗法。

其計算公式如下：
$$t' = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left(\frac{S_{X1}^2}{N_1} + \frac{S_{X2}^2}{N_2}\right)}}$$

其中 \bar{X}_1 及 \bar{X}_2 表示兩個所欲考驗樣本的平均數

S_P^2 表示樣本的變異數不偏估計值，其計算公式為

$$S_P^2 = \frac{\sum(X_1 - \bar{X}_1)^2 + \sum(X_2 - \bar{X}_2)^2}{N_1 + N_2 - 2}$$

N_1 和 N_2 表示兩個所欲考驗樣本的個數

S_{X1}^2 和 S_{X2}^2 表示兩個所欲考驗樣本的變異數

() 由上述 () 及 () 計算公式所求得的 t 值，在 t 分配中出現的機率小於顯著水準 (即 $P <$ 值)，表示在 的顯著水準下拒絕了虛無假設。亦即兩個平均數之間有顯著差異，研究者如此推論，犯第一類型錯誤的機率小於 。

(8) 兩個獨立樣本的百分比的差異顯著性考驗：

() $I \times J$ 交叉表或列聯表 (contingency table)：進行 χ^2 (卡方) 考驗時，常將所收集到的資料安排為 I 個橫列、 J 個縱行的方格形狀的表，稱為 $I \times J$ 列聯表 (contingency table)。

() χ^2 (卡方) 考驗：

(a) 定義公式為： $\chi^2 = \sum \left\{ \frac{(f_o - f_e)^2}{f_e} \right\}$ ；其中 f_o 是觀察次數 (observed frequency)， f_e 是期望次數 (expected frequency)。

(b) 運算公式： $\chi^2 = N \left\{ \sum \left[\frac{f_{ij}^2}{(f_i \cdot f_j)} \right] - 1 \right\}$ ；其中 f_{ij} 表 $I \times J$ 列

聯表中的第 J 組受試者的第 I 個反應的數值， f_i 為樣本中所有的第 I 個

反應的總數， f_j 為樣本中的 J 組受試者總數，N 為樣本總數。

- () 百分比同質性考驗 (test of homogeneity of proportions): 使用 χ^2 考驗進行百分比的同質性考驗。其主要目的在於考驗被調查的 J 組受試者在 I 個反應 (或問卷選項) 中選擇某一選項的百分比是否有顯著差異。

當應用運算公式所計算的 χ^2 (卡方) 值較利用 χ^2 (卡方) 分配表所查到的值為大時，拒絕虛無假設，即 I*J 列聯表中的百分比並不同質，亦即表中各選項的百分比達顯著差異水準。

(9) 獨立樣本單因子變異數分析 (ONE-WAY ANOVA):

- () 離均差平方和 (sum of square of deviations from the mean, 簡寫 SS):
公式: $SS = \sum (X - X_a)^2$ 。其中 X 表一般樣本數值， X_a 表樣本平均數。
- () SS_t (總離均差平方和): 表示全部樣本的個別數值與樣本總平均值的差的平方和。即 () 中的 X 取全部的樣本數值， X_a 取全部樣本平均數。
- () SS_w (組內離均差平方和): 代表各個組內 (within groups) 的離均差平方和。即 () 中的 X 取組內的樣本數值， X_a 取組內的樣本平均數值，並將所有組的 SS 加總。
- () SS_b (組間離均差平方和): 代表各組組間 (between groups) 平均數離開總平均數之變異情形的量數，即組間的離均差平方和。而 $SS_t = SS_w + SS_b$ 。
- () 均方 (Mean Square, 簡寫 MS): 等同於變異數的不偏估計值 S^2 (variance estimate), 即為離均差平方和除以自由度。

變異數分析中，有組間變異數不偏估計值 (MS_b) 和組內變異數不偏估

計值 (MS_w), 而 $MS_b = SS_b / df_b$, $MS_w = SS_w / df_w$, 其中 df_b 為組間自由度, df_w 為組內自由度。

() F 考驗值：表示組間的 MS 除以組內的 MS；即 $F = MS_b / MS_w$ 。

當 F 值大於 1 且落入臨界區，則應拒絕虛無假設，即平均數間有顯著差異存在。若計算的 F 值大於查 F 分配表的臨界值，則應拒絕虛無假設，即平均數間有顯著差異存在。

當 F 值達顯著水準時，就表示各組平均數之間變異很大，且至少有一對平均數之間有顯著差異存在。此時必須進行事後比較，以找出哪些平均數之間有顯著差異。

() Scheffe 考驗 (簡寫 S 法)：當變異數分析結果其 F 值達顯著水準時，則要進行事後的比較，採用 Scheffe 考驗，首先計算 F 值 (公式略，請參考林清山，民 81，頁 361)。

當所計算的 F 值大於查表的 F 分配值與組間自由度的乘績時，表示兩個平均數達到顯著差異水準。

當由計算公式所求得的 F 值，在 F 分配中出現的機率小於顯著水準 (即 P 值 <)，表示在 的顯著水準下拒絕了虛無假設，即表示兩個平均數之間有顯著差異，表示研究者如此推論，犯第一類型錯誤的機率小於 。

(10) 獨立樣本單因子共變數分析 (ONE-WAY ANCOVA)：

先利用直線迴歸分析將共變量 (covariate) 的影響排除後，再利用變異數分析去考驗各組平均數之間是否仍然有顯著差異存在。共變數分析的主要步驟有下列三個：

- () 組內迴歸係數同質性考驗 (homogeneity of within-class regression coefficient):

當計算的 F 值在 F 分配表中之左、右兩端臨界值間時，虛無假設成立。亦即，當由計算公式所求得的 F 值，在 F 分配中出現的機率在信賴水準內 (即 $P > \alpha$ 值)，即組內迴歸係數是同質；若 F 計算值，落於 F 分配表中的信賴區間外，即 P 值小於顯著水準 α ，則組內迴歸係數不同質。

當組內迴歸係數不同質時，不宜進行共變數分析。

- () 共變數分析：將同質的迴歸係數合併成一個共同適用的組內迴歸係數，然後在這種假定下進行共變數分析，看排除掉共變量的影響之後，各組平均數之間是否仍然有差異存在。

當所計算的 F 值大於查表的 F 分配值時，即表示即使排除共變量的解釋量部分後，各組平均數 (亦即調整後的平均數) 之間仍然有顯著差異存在。可進行事後考驗，以瞭解各組的平均數的真正差異在何處。

當由計算公式所求得的 F 值，在 F 分配中出現的機率小於顯著水準 (即 $P < \alpha$)，表示在 α 的顯著水準下拒絕了虛無假設，即各組平均數之間有顯著差異，研究者如此推論，犯第一類型錯誤的機率小於 α 。

- () 求調整後的平均數：如果共變數分析的結果，F 值達到顯著水準，表示即使排除共變量的解釋量部分後，各組平均數 (亦即調整後的平均數) 之間仍然有顯著差異存在。