

第五章 創造思考的歷史教學評量

本章旨在針對前一章的教學實驗結果進行分析評量，其步驟是首先透過多元化評量來檢測學生在創造思考的歷史教學方面的成效，包括應用布魯姆認知教學目標自編的「歷史學科能力測驗」及實作評量等；接著利用教學評鑑問卷，分析學生在經過此一創造思考教學實驗後的學習心得與建議，作為進行另一次教學的參考依據。

第一節 創造思考能力的多元評量

在整個教學過程中，教學評量有著不可或缺的重要性，可以說是擔任教學過程中承接轉合的關鍵部分。透過教學評量，可以分析之前教學目標的達成與否、衡量教師教學過程及方法的得失、評斷學生的學習困難、提供之後實施補救教學和個別輔導的依據等。

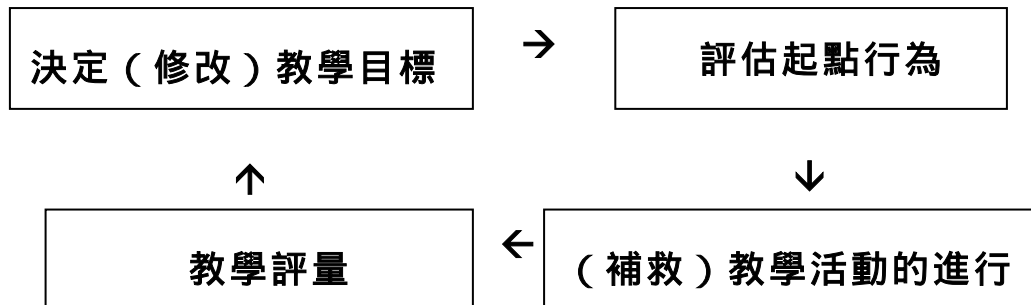


圖 5-1 教學基本模式 (The General Model of Instruction, GMI)

雖然，以往教育工作者未曾輕忽教學評量在教學過程中的重要性，但長久以來，評量卻常被誤用而扮演了錯誤的角色。如以教學模式的順序來說（圖 5-1 所示的），是先決定教學目標，才接續起點行為的評估，再來教學活動的進行，最後才是教學評量，所以應該是教學目標引導且決定教學評量的實施，

即先有目標，評量是根據目標來決定²¹⁴。故教學評量的目的是在分析「應該怎樣表現」(知識)、「真正表現行為」(實作)之間的差異，方能檢討教學成效與實施補救教學²¹⁵。但是家長、學生，甚至教師們常常讓教學評量嚴重影響教學內容、教學目標，產生所謂「考試領導教學」的本末倒置現象。所以近年來教育改革的重點之一就是「評量的改革」，希望以多元化評量帶動大家對於多元能力、全人教育目標的重視。至於何謂「多元評量」，簡單的說就是運用一種以上的評量方式檢驗學生的學習狀況，也包含傳統的紙筆測驗在內。

創造思考教學的目標是在培養學生多元化思考及創造力，透過各種教學策略以求達到上述目的，當然需要經由教學評量以檢視教學成效。本節即在透過多元評量的方式，檢驗創造思考教學的成效；同時也針對各種不同評量方法進行評估，俾能掌握本教學實驗的具體成果。

一、創造思考教學的多元評量

(一) 傳統評量與多元評量

評量的方式有許多種，其中最普遍的方式應屬紙筆測驗。舉凡學校的隨堂考及定期考查，到高中、大學學力基本測驗，大多採用傳統的紙筆測驗，其所以成為最常使用的評量方式，這是因為它亦有以下的優點²¹⁶：

1. 行之多年，大眾（包括教師、家長、學生）對它的實施方式皆

²¹⁴ 張素珍，〈國小社會科之真實性評量：以「學校的自治活動」單元為例〉，〈人文及社會學科教學通訊〉，第十二卷六期，第76頁，民國91年4月。

²¹⁵ Wolf, D. P., Bixby, J., Glen, J., & Gardner, H. new forms of student assessment. In G. Grant (Ed), Review of Research in Education, 17 (pp.31-74). Washington, DC: American Educational Research Association. 1991.

²¹⁶ 多元評量與基本學力測驗之推動 - 簡茂發，教育研究月刊，第98期，第5-6頁，民國91年6月。

不陌生，評量的結果一般人較熟悉，容易了解其相對的意義，實施起來亦較簡便。

2. 紙筆測驗能測量認知領域的學習結果。
3. 學生團體能夠在同一時間內一同參與測驗，符合經濟效益的原則，實施效率佳。
4. 評分標準有客觀的標準答案，較符合社會對評量公平的要求等。

但根據許多學者及站在教學第一線的基層教師觀察，傳統教學評量亦有以下的缺點²¹⁷：

1. 大致上是指「紙筆測驗的評量」，即「一元化評量」，導致教學目的常常只為求出單一標準化的答案，壓縮、壓抑教師及學生教學過程中「思考」及「創意」的空間。
2. 既偏重紙筆的評量，且受限於教師本身的命題能力，以致考試題目多偏重在記憶、理解層次，屬於較低層次能力的評量。而較高層次的能力評量，包括教學目標所指出的應用、分析、綜合、評鑑等高層次能力的評量，一般教師大多難以設計出適當的試題，久而久之，學生們的思考能力自然趨於僵化。
3. 窄化了學生們的能力及學習興趣，致使許多學生沒有機會表現其他評量方式可以察覺的能力，這種現象，與現今多元能力的理論相違背，結果學生們只能往背書、唸書、應付考試的方向學習。至於其他可以用手操作、表演、觀察、實驗……等等評量的能力，學生們幾乎沒有發揮的機會與空間。

²¹⁷ 郭生玉，〈如何落實多元評量〉，《教育研究月刊》，第98期，第11頁，民國91年6月；郭慧龍，〈「多元評量」與「檔案評量」之我見及我思〉，《竹縣文教》，第24期，第20頁，民國90年12月；李坤崇，〈多元化教學評量〉，第3頁，台北心理出版社，民國91年4月。

4. 傳統教學評量的功能，大都僅重視如何分辨學生成績的高下，命題觀念與技術過於依賴量化評量，且大多只著重在「總結性評量」以作為獎懲學生之用，並沒有真正發揮評量應該具備的效果 - 如利用「形成性評量」、「診斷性評量」檢討及補救教學，以致分數帶來的挫折感成了大多數學生學習興趣的殺手，而未考量個別學生努力的過程及程度。
5. 傳統教學評量在技能、情意領域有其先天限制，如文章、圖畫、實驗報告、自然科學展覽設計、演講、小組合作等學習能力，皆難以運用紙筆測驗評量，

綜合以上對於傳統評量的看法，可知紙筆測驗是眾多教學評量方式之一，自有其擔負的角色與侷限，我們不須全盤否定其應有的作用，也不該放大了它的重要性。但是也勿陷入多元評量的迷思中²¹⁸，而應分了解：多元評量並非廢除紙筆測驗、並非「一種評量」(即實作評量、檔案評量、動態評量其中一種) 所以不能為多元而多元。在實施多元評量時應考慮下列向度²¹⁹：

1. 評量人員的多元化、互動化：教師的角色由考核者轉為協助者，納入自我評量、同儕評量、家長評量，彼此之間充分溝通與討論，才能更了解學生學習歷程與結果。
2. 評量內容的多元化、生活化：包含認知(記憶、理解、應用、分析、綜合、評鑑) 情意、技能三個領域，並重視學習方法、

²¹⁸ 李坤崇，〈多元評量〉，《教育研究月刊》，第 25 頁，第 98 期，第 25 頁民國 91 年 6 月。

²¹⁹ 郭慧龍，〈「多元評量」與「檔案評量」之我見及我思〉，《竹縣文教》，第 24 期，民國 90 年 12 月，20 - 21 頁；侯雅齡，〈如何有效運用多元評量結果〉，《測驗與輔導》，第 159 期，第 3339-3341 頁，民國 89 年 4 月；邱上貞，〈評量的適配：多元的評量與變通的考試〉，《教學創新及多元評量》，第 19 - 20 頁，高雄市國民中學九年一貫課程校際策略聯盟之社會學習領域教師研習，高雄市大義國民中學製作，民國 91 年七月；李坤崇，〈多元化教學評量〉，第 120-129 頁，心理出版社，民國 91 年 4 月。

習慣、求知歷程及建立學生「核心的生活能力」。

- 3.評量方法的多元化、彈性化：並非廢除紙筆測驗，而是降低它在現今教師評量的比率，加重其他評量方式的比率，彈性運用各種評量方式來適切評量學生的學習成果，如觀察、討論、面談、研究報告、自傳、標準化測驗、問卷、圖表、繪畫、查核表、活動紀錄、工作樣品展示、日記、軼事紀錄等。
- 4.評量標準的多元化、適性化：評量及格標準應依據學生的個別差異來訂定，以適性測驗的精神，設計具有鑑別度的題目，測量學生可能學會的部分，藉此讓每個學生都體會到自己的價值而適才適性的發展。教師能欣賞創造力的答案，莫以標準抹煞，當學生偶而出現頗具創造力的答案，若教師拘泥於既定標準評分而不給分，壓抑學生創造力，教學評分若發現具創造力或爭議性之答案，宜審慎思維或請學生詳細說明在決定給分多少，方能激發學生創造力。

總結以上有關多元評量的注意事項，可知多元評量並非像大拜拜般將所有評量方法都呈現出來，而是在教學過程中依照學生個別學習狀況，在適當時機運用不同的教學評量方法，才能達成多元化評量的真正目的。

（二）創造思考的歷史教學與多元評量

事實上，多元評量的概念在國民中學課程標準總綱的實施通則中即有提及²²⁰：教學評量內容應涵蓋認知、技能、情意三大領域，評量方法宜採多元化方式實施，包括觀察、實作、表演、口試、筆試、作業練習、研究報告等，教師應按學科性質，相機酌用並隨時記錄。

²²⁰ 教育部中教司，〈國民中學課程標準〉，第12頁，台北中正書局，民國85年7月。

同時，國民中學歷史課程標準也提到教學評量應力求多元化，包括：口頭問答、定期考查、課外作業，如製作圖表、撰寫心得報告、蒐集相關資料及動態活動，如辯論比賽、話劇表演、口頭報告、分組討論、教學觀摩、史蹟考察、參觀訪問等²²¹。

現在實施中的九年一貫課程亦強調：學習評量應掌握多元評量之精神與做法，進行多元評量時應注重學習困難之診斷，而非以取得某一資格或選拔條件為目的；學習評量之內涵，應涵蓋認知、技能、情意等三種層次，不應偏重認知層次的評量，而忽略技能或情意目標之評量；應設計多元評量方式，可採用紙筆測驗、實作評量與檔案評量等方式進行，重視形成性和總結性評量結果²²²。

從以上可知多元評量是符合現今全人教育的目標，希望扭轉以往一元化評量所導致的種種教育亂象。若將多元評量運用於創造思考的歷史教學過程中，必也能貫徹創造思考的歷史教學目標：培養歷史思維、判斷力及創造力，將所學實際運用、落實在「現實生活」中，使之更能適應多元、複雜的社會生活。

在實際教學中最常運用的評量方式不外乎紙筆測驗，如何採用傳統的紙筆測驗以促進學生的創造力發展呢？就紙筆測驗的性質而言，可以加以改進的就是：命題的內容不能只限於是課本知識的記憶或理解，而應盡可能提升到應用、分析、綜合、評鑑等層次²²³；就歷史學科而言，試題設計應能測知學生解釋與分析的思維能力、多角度的思考方向；同時命題資料不限於文字，圖表、照片、統計數值均為採用²²⁴。

²²¹ 同上注，第 228-229 頁。

²²² 教育部，〈國民中小學暫行課程綱要實施要點（草案）〉，民國 89 年。

²²³ 同上注，第 392-393 頁。

²²⁴ 李月桂，〈淺談大學入學考試歷史科試題的命題趨勢〉，〈台灣師大歷史教育〉第三期，第 203 頁，民國 87 年 6 月。

基於上述多元評量理論的認識，本教學實驗針對創造思考的歷史能力所做的紙筆測驗，是根據布魯姆認知教學目標的原則加以新設計，試題的認知層次大致分為三類：知識類（記憶、回憶所學的教材，如時地事物等名詞）；理解類（了解所學過知識的意義，包含圖文史料的轉換、解釋及推論能力）；高層次思考（包含應用、分析、綜合、評鑑等層次的試題），除了評量一定的記憶力外，更加重高層次思考的分析能力的評量，目的即在檢測創造思考的歷史教學目標是否有效達成。

除紙筆測驗外，本教學實驗還採用實作評量等方式。為了能增加此一評量的效度與信度，認知方面，採取「量化」的評量指標如下²²⁵：

- 1.流暢力：指產生多少觀念而言，是所有有關反應的總合，如能列舉十個反應得十分。如以「蔣中正與北伐統一」一節中的圖案聯想活動為例，有一組在一定時間內（三分鐘）能想出5個有關與北伐有關的圖像，就得5分的流暢力。
- 2.變通力：指不同分類或不同方式的思考，是指所有類別的總合。如同上例，那一組想出五種圖像，其中有兩種皆是屬於日本軍閥的聯想，有三種則都是屬於北洋軍閥的雙關語圖案，故該組的流暢力雖然是5分，變通力則是2分。
- 3.獨創力：指反應的獨特性，想出別人所想不到的事物，在評分時是指統計上的稀少次數，依常模給分，反應超過5%則不計分。如在以上的教學活動中，有一組提出的以鮮紅色在山東省的地圖位置，用槍排出「五、三」圖案，是其他組別所未有人想到的，自然就得到獨創力的最高5分。

²²⁵ 陳龍安，〈創造思考教學的理論與實際〉，第 388 - 389 頁，台北心理出版社有限公司出版，民國 77 年 4 月。

4.精進力：是指增加細節的部分，除了第一個觀念的基本結構外，每增加一個概念則得一分。如有一組除了想出以鮮紅色在山東省的地圖位置，用槍排出「五、三」圖案，還讓血跡繼續向東北地區蔓延，並加以說明這個圖案中包含了「五三慘案」及「日軍意圖染指東北」兩個概念，在精進力方面可得2分。

至於在情意方面採用以下的項目建立檢核表評量之²²⁶：

- 1.冒險心：是指有猜測、嘗試、實驗或面對批評的勇氣。例如：
在教師未指定作答學生時，若有同學自願回答並接受其他同學對其答案的批評，這是一種堅持己見的勇氣。
- 2.挑戰心：從複雜混亂中理出情緒，主要在尋求可行的辦法。以第十四章第二節「中國的劇變」(戡亂時期到政府遷台)為例，當教師詢問學生：「如果遇到類似當時嚴重通貨膨脹的情形將如何因應？」，藉此評量學生是否能洞察出變動的原因以找出解決的方案。
- 3.好奇心：對某一種事物感到懷疑產生問題，然後試著去調查、探詢、追問，雖然感到困惑，卻仍然繼續思索、沉思，以求明白事物的真相。如在第十三章第二節「中共的成立」，製作中外歷史年表的教學活動中，從實作中讓學生提出歷史事件之間是否有因果關係的疑問，並試著證實其想法。

二、創造思考的歷史教學評量實施情形

本教學實驗所採用的評量方式，若依評量時機來區分，可分為：課堂觀察評量(評定量表、軼事記錄)、單元形成性評量(評定量表、口語評量)、單元總結性(紙筆)評量、段考總結性(紙筆)評量等四項；若以多元評量的種類來說，可分成以下兩類：

²²⁶ 同上注，第 389 頁。

一是紙筆測驗，包括單元總結性評量的自編「歷史學科能力測驗」及全年級的段考總結性（紙筆）評量；二是實作評量，包含小組合作的評定量表、軼事記錄及學習單的評定量表、口語評量。茲分項敘述分析如下：

（一）自編的「歷史學科能力測驗」

1. 編題

筆者根據上述創造思考的歷史教學評量目標，將本測驗題目分為三個向度：基礎認知性、理解性、高層次性，依此設計雙向細目表以決定各章節、各向度之題數，再自各方題庫中選出合乎實驗教材範圍及評量目標的題目共五十題，題數分配如表 5-1 所示：

表 5-1 「歷史學科能力測驗」初選題數分配

向度 \ 章節	第十二章第二節到第三節	第十三章第二節到第三節	第十四章第二節到第十五章第二節	小計
基礎認知性	5	5	5	15
理解性	7	7	6	20
高層次性	5	5	5	15
合計	17	17	16	50

2. 預試與試題分析

編題後，以高雄市正興國中三年級六個班共 220 人，進行第一次預試，根據預測結果算出難度與鑑別度後，淘汰其中鑑別力較低的題目，剩下 20 題，分配情形如表分配情形如表 5-2 所示：

表 5-2 「歷史學科能力測驗」複選題數分配

基礎認知性	1	1	2	4
問題解決性	3	3	3	9
高層次性	2	3	2	7

合計	6	8	7	20
----	---	---	---	----

再就二十題之試卷，以彰化縣埔心國中及高雄市和平國中三年級各兩個班，共 154 人為對象，進行第二次預試及試題分析，其人數分配及試題分析結果如表 5-3 及表 5-4 所示：

表 5-3 「歷史學科能力測驗」預試取樣人數

學校 \ 性別	男生	女生	合計
彰化埔心國中	37	42	79
高雄和平國中	28	47	75
合計	65	89	154

表 5-4 「歷史學科能力測驗」分析結果

題號	向度	難度 P	鑑別度 D	題號	向度	難度 P	鑑別度 D
1	③	0.62	0.56	11	③	0.82	0.37
2	①	0.63	0.39	12	②	0.51	0.63
3	②	0.78	0.44	13	③	0.76	0.44
4	②	0.71	0.54	14	①	0.71	0.59
5	②	0.67	0.56	15	①	0.62	0.51
6	②	0.66	0.44	16	②	0.62	0.76
7	②	0.74	0.51	17	②	0.61	0.54
8	①	0.60	0.66	18	③	0.56	0.63
9	②	0.72	0.41	19	③	0.63	0.73
10	③	0.80	0.39	20	③	0.62	0.41

說明：表中的向度①代表基礎認知性題目；②是指理解性題目；③則屬於高層次性題目

由表 5-4 顯示，27 % 高分群與 27 % 低分群試題分析結果，發現雖然有些題目之難度 (P) 不夠適中，或鑑別度 (D) 偏低之

情形，然難度及鑑別度的中數各在 0.65 及 0.53，難度的範圍在 0.51-0.82 之間，鑑別度在 0.37-0.76 之間，若根據中華民國大學入學考試中心統計分析的標準，即以 $0.80 > P \geq 0.50$ 為中難度試題， $D \geq 0.40$ 鑑別度為優、 $0.40 \geq D \geq 0.30$ 則為良的分類方法²²⁷，本試題符合了選題的標準，應可作為正式測驗之工具，見附錄二。

3. 信度與效度研究

(1) 信度²²⁸研究：

就第二次預試樣本中和平國中四十人的測驗結果（見表 5-5），以庫李法（Kuder-Richardson Method）來看受試者對所有題目反應的一致性。經過代入庫李 21 號公式後得出信度為 0.775²²⁹，就一般教師自編之測驗大多在 0.60 到 0.80 之間的標準來看²³⁰，本試題信度尚稱適中，即測驗結果的一致性與穩定性是在可接受的範圍。

表 5-5 「歷史學科能力測驗」第二次預試的總分結果

學生	總分	學生	總分	學生	總分	學生	總分	學生	總分	\bar{X} （總分平均） 是 13 標準差為 4.5 S^2 為測驗總分之變異數（標準差的平方）為 4.5^2
1	18	9	13	17	9	25	7	33	15	
2	14	10	16	18	14	26	17	34	7	
3	18	11	5	19	17	27	8	35	10	
4	5	12	19	20	12	28	16	36	20	
5	9	13	8	21	15	29	18	37	14	

²²⁷ 陳豐祥，〈歷史教學評量的理論與實際〉，第 202 頁，台北：台灣學生書局出版，民 83 年 4 月。

²²⁸ 信度是指測量的一致性，一般而言，兩次測驗的結果越是一致，則誤差越小，所求得的信度也就越高。

²²⁹ 庫李 21 號的公式為： $1 - \frac{\sum(X - \bar{X})^2}{K(S^2)}$ ，其中 K 表示整個測驗之題數，故 $K = 20$ ； \bar{X} 為測驗總分的平均數，故 $\bar{X} = 13$ ； S^2 為測驗總分之變異數（即標準差的平方，標準差的計算方法是各分數與平均數之差數的「平方和」除以總人數，再開平方），標準差為 4.5 故 $S^2 = 4.5^2$ ，以此法求得 $r_{KR21} = 1 - \frac{13(20-13)}{20(4.5)^2} = 1 - \frac{13 \times 7}{(20 \times 20.25)} = 1 - \frac{91}{405} = 1 - 0.225 = 0.775$

²³⁰ 陳英豪、吳裕豪，〈測驗的編製與應用〉，第 459 頁，高雄：偉文圖書出版有限公司出版，民 90 年 3 月。

6	15	14	17	22	14	30	4	38	15
7	14	15	16	23	13	31	17	39	6
8	6	16	15	24	19	32	18	40	15

(2) 效度研究：

本實驗試題的內容效度由專家及現任國中歷史教師共20人，協助對內容作質的分析。即筆者在每一個題目的後面列上自身所認定的題型性質，如表5-1所示各有基礎認知性4題、理解性9題和高層次思考性7題，請回答問卷者根據所描述的題型特質，加以評析是否同意該題符合筆者所認定的題型，所使用的問卷表格見附錄三，結果如表5-5所示，各題同意題型分類之百分比為55 % -85 % 之間，可見本測驗的題目大多能達到它所要測量的教學目標。

表 5-6 「歷史學科能力測驗」內容一致性分析表

題號	同意人數	百分比	題號	同意人數	百分比
1	12	60 %	11	13	65 %
2	17	85 %	12	15	75 %
3	13	65 %	13	11	55 %
4	15	75 %	14	16	80 %
5	12	60 %	15	17	85 %
6	13	65 %	16	14	70 %
7	15	75 %	17	15	75 %
8	16	80 %	18	15	75 %
9	16	80 %	19	12	60 %
10	14	70 %	20	14	70 %

4.各組前後測之平均數及標準差²³¹

在實驗教學前(約一個月之前),實驗組及對照組先進行「歷史學科能力測驗」的前測,測驗後並不公佈答案,等到教學實驗結束後再實施後測。表5-7是全體受試在「歷史學科能力測驗」前後測之平均數及標準差,可以知道實驗組無論在前測或後測時的平均成績皆不如對照組,但是實驗組在前測時與對照組差距3.4分,至後測時差距縮小至2.7分;另外在標準差方面,實驗組在前測時大於對照組,至後測時就小於對照組,可見實驗組在此一測驗的成績雖未超越對照組,但可見其有些許的進步,且分數的差異度縮小也代表大多數人都有進步。

表 5-7 各組「歷史學科能力測驗」前後測之平均數及標準差

項目	組別	人數	平均數	標準差
前測	實驗組	26	24.4	14.8
	對照組	25	27.8	13.3
後測	實驗組	26	60.7	21.6
	對照組	25	63.4	25.8

(二) 全年級的段考總結性評量

此一評量的目的,是期望能分析創造思考的歷史教學與歷史學業成就的關係如何。由期末考評量的結果得知,實驗組的平均數為67.3,對照組為70.8,兩者差距三分;標準差卻較對照組小,與「歷史學科能力測驗」的結果大同小異,可見創造思考的

²³¹ 標準差是一種最重要且最有用的差異量數,我們可以藉由標準差了解一個群體各分數之集中與分布的情形,即標準差愈大代表該班學生的分數高低差異愈大。

歷史教學對於有標準答案的測驗助益不大。

表 5-8 各組在段考總結性評量之平均數及標準差

組別	人數	平均數	標準差
實驗組	26	67.3	20.6
對照組	25	70.8	20.9

(三) 實作評量

「實作評量」就是由教師設計相關的情境，由此情境，針對學生所應達到的學習成果，設計一些問題，讓學生在情境中或實際參與操作或觀察之後，以分組活動或個別思考的形式，進行問題的解決，同時針對學生在過程中的表現，以客觀的標準加以評分的一種評量方式²³²。根據 Stiggins 的看法，觀察和專業判斷來評量學生學習成就的評量方式都可以稱作為「實作評量」²³³，其方式非常的多元化，例如書面報告、作文、演說、實驗、資料蒐集、作品展示等，形式則如軼事記錄、評定表、口語評量等，其實施方式可以是日常生活的形成性評量、總結性評量或特殊教學計劃，其評分是以人的專業判斷為主。實作評量可以評量學習的過程與結果，可以將評量潛藏於課程之中，也可以評量情感性的表現以及社會技巧²³⁴。

²³² 何俊青，〈另類評量在社會領域概念評量的應用〉，〈人文及社會學科教學通訊〉，第十二卷六期，第 62-63 頁，民國 91 年 4 月。

²³³ Stiggins, R. Design and development of performance assessment. *Educational Measurement: Issues and Practice*, 6, 33-42, 1987.

²³⁴ Garcia, G. E., & Pearson, P. D. Assessment and diversity. In L. Darling Hammand (Ed.), *Review of Research in Education*, 20, 337-391, 1994.

因為實作評量在測高層次認知、歷程、情意、技能與後設認知等複雜且統整的能力，且由於強調「帶得走」的能力（即習得能力可以遷徙至真實生活中），故作業的真實程度較高²³⁵，相當適合作為創造思考教學的評量方式。

以下即是幾種本教學實驗所採用的實作評量方式：

1. 小組合作與學習單的評定量表

本實驗的教學活動大多是以學生分組的方式進行，尤其是在運用腦力激盪法設計的教學活動時，小組成員是否通力合作對整個組別表現影響特別大，故在教室觀察部分設計了「小組合作評定量表」，以觀察小組活動的情形如表 5-9²³⁶所示，評分方式有小組成員互評與教師評量，並包含評量學生學習過程中努力的程度，給予正增強。

如以第十三章第二節中「西安事變」為例，教學過程中有一『西安事變真相追追追』的教學活動，希望藉由故事接龍的型態完成完整版的西安事變。在情意方面的教學目標是激發學生的冒險心及團隊合作的凝聚力，至於教學方法則是各組的每一個同學，都必須依序說出一句有關事件中的時間、地點、人物、事件發展的字句，才能架構成一個完整的事件。先由組長帶領大家討論出事變的輪廓，事件完整與否端賴全體組員是否能通力合作，小組成員自然最清楚誰在討論中是否充分發揮積極參與又尊重他人的精神，而此一評分標準亦評量同學的努力程度，即觀察是否更用心以達成目標，以鼓勵能力不足但努力達成目標的同學。

²³⁵ 張麗麗，〈從分數的意義談實作評量效度的建立〉，《教育研究月刊》，第 98 期，第 40 頁，民國 91 年 6 月。

²³⁶ 參考及修改自張素珍，〈國小社會科之真實性評量：以「學校的自治活動」單元為例〉，《人文及社會學科教學通訊》，第十二卷六期，第 85、90 頁，民國 91 年 4 月。

表 5-9 「小組合作評定量表」

+ -		評分	評分標準說明
姓名			
努力程度			「+」號越多代表越用心、越進步；代表比以前用心或進步； 「-」號越多代表越不用心、越退步；代表比以前不用心或退步。
	1		只參與少數小組活動，未自己完成工作，又比較不尊重他人的想法，也不能清楚的表達自己的想法。
	2		只參與少數小組活動，未自己完成工作，又比較不尊重他人的想法，但能清楚的表達自己的想法。
	3		參與大部分小組活動，能獨自完成自己的工作，但比較不尊重他人的想法，不能清楚的表達自己的想法，無法與他人合作一起完成工作。
	4		參與全部的小組活動，能尊重他人的想法，也能和組員分工合作，但是比較不能提出自己的看法。
	5		每次的小組活動，都積極參與，能尊重他人的想法，也能和組員分工合作，清楚表達自己的想法。

另外還有各學習單的評定量表，如以第十四章第二節「中國的劇變」(戡亂時期到政府遷台)的學習單為例，教師先將這段時期的內容依這六大主體加以處理，說明並設計成相關的六大問題，請學生思考，分組搶答，教師並再配合相關的相聲(表演工作坊的『今夜我們來說相聲』)讓學生更具體的了解遷台的情形，最後再請學生就「如果遇到類似的情形將如何因應？」，作感想及意見的分享。教師則亦依據「六W」的歸納方法，編出「類心理測驗」的學習單。請各組從「六W」中各個角度去找尋線

索，在一定時間內完成學習單，完成度最高的組別獲勝
設計如表 5-10 的評定量表以進行自評及老師評量²³⁷。

表 5-10 學習單的評定量表（自評及老師評量）

評分題目	勾選	評分標準
1.搶答比賽	<input type="checkbox"/>	0分：未作答、作圖
	<input type="checkbox"/>	1分：作答、作圖不完整
	<input type="checkbox"/>	2分：作答、作圖詳盡但速度較慢
	<input type="checkbox"/>	3分：作答、作圖詳實且速度較快
2.聽相聲、抖包袱	<input type="checkbox"/>	0分：未作答
	<input type="checkbox"/>	1分：所抒發的感想不夠充實
	<input type="checkbox"/>	2分：所抒發的感想十分充實，但文字敘述不夠流暢通順或不夠精簡
	<input type="checkbox"/>	3分：所抒發的感想十分充實，文字敘述流暢通順且精簡
3.發揮題	<input type="checkbox"/>	0分：未作答
	<input type="checkbox"/>	1分：所發表的想法不夠有創意
	<input type="checkbox"/>	2分：所發表的想法十分有創意，但文字敘述不夠流暢通順或不夠精簡
	<input type="checkbox"/>	3分：所發表的想法十分有創意，文字敘述流暢通順且精簡
4.歷史類心理測驗	<input type="checkbox"/>	0分：未作答或作答錯誤
	<input type="checkbox"/>	1分：僅列出 1-2 項之正確答案
	<input type="checkbox"/>	2分：僅列出 3-4 項之正確答案
	<input type="checkbox"/>	3分：能完整答對六個答案
總得分：		分

又以第十五章第二節「中華民國在台灣的發展」為例，請學生藉由訪問家中長輩及蒐集資料了解台灣自 30 年代到現今的轉

²³⁷ 參考及修改自王淑慧，〈美麗的剪貼簿 - 社會科檔案評量〉，〈測驗與輔導〉，第 160 期，第 3363 頁，民國 89 年 6 月。

變與不同之處，除了做成書面報告以外並上台報告，針對此一學習單所設計的教師評定量表如表 5-11²³⁸。

組別	報告主題	小組長	蒐集資料人員	資料整理人員	打字人員	報告人
評分項目	內容切合重點		書面資料整理完善	報告者的表現合宜		團體表現
勾選	1分：報告內容偏離主題		1分：書面資料沒有整理	1分：報告者的神態、說話速度、語調、清晰度、肢體動作皆不佳		1分：輪該組報告人員上台時，全組漠不關心，認為只是報告者的事
	2分：報告內容並非重點		2分：書面資料雖經整理，但仍凌亂	2分：報告者的神態、說話速度、語調、清晰度、肢體動作皆尚可		2分：輪該組報告人員上台時，只有一部份人幫忙，其他人看熱鬧
	3分：報告內容切合主題		3分：書面資料整理完善	3分：報告者的神態、說話速度、語調、清晰度、肢體動作皆十分良好		3分：輪該組報告人員上台時，全組人員互助合作
總得分						
努力程度	+ + + + + + + + + + + +			- - - - - - - - - - - -		
教師評語						

表 5-11 教師評定量表

2. 軼事記錄

本教學實驗是採取行動研究法，即須以日誌紀錄觀察及訪

²³⁸ 同上注。

談學生的情況，藉此真正了解學生的學習意願等情形，作為教師反省思考的有效方法，這可以和實作評量中「軼事記錄」的評量方式作一結合，以作為情意方面評量的重要依據。現以表 5-12 為例，說明在第十二章到第三節「蔣中正與北伐統一」的教學過程中的「軼事記錄」。

姓名：戴 、陳xx、李	時間：91 年 4 月 30 日
地點：二年八班教室	觀察者：陳老師
上課進度：第十二章到第三節「蔣中正與北伐統一」	
事件：	
<p>此次上課主要是以一些與北伐有關的圖案讓學生作聯想。我先示範畫出一些圖案讓學生猜想，學生們反應踴躍。接著反過來，我要同學根據主題想出聯想圖案，戴 很有自信的說：一個軍人拿著塗成黃色的槍指向北方，打倒了十個人、陳xx搶著說：地圖上山東的地方有一個日本武士（衣服上有日本國旗）亂開槍，地上倒了一些人並排成「五三」字樣（五三慘案）、李 跟著說：校門（門上寫著親愛親誠）內有一排軍人，有一人高喊著向北衝，目標是北方的十門大砲。我嘉許他們的表現，而這三人所提出的答案各讓該組加分，三人還意猶未盡的要再提出新的聯想，因時間不夠而露出遺憾的表情。</p>	
解釋：	
<p>這三位同學在紙筆測驗時常常是屬於低成就的一群，平時愛在書上塗鴉，學習興趣是較低落的他們對這樣圖文聯想的課程卻表現出極好的反應，開始在課堂上會與老師有互動、參與活動的進行，不怕回答錯誤而勇於發言，讓我感受到他們在學習興趣上的進步。</p>	

表 5-12 「軼事記錄」卡

3. 口語評量

發問的技巧是創造思考教學極重要的一環，甚至可說運用若得當，就是創造思考教學是否成功的關鍵。且教師於教學過程中以問題問學生乃是常見的師生互動模式，只是較少教師注意提問的時機與技巧，甚至不認為它是教學評量的一部分²³⁹。事實上，「問問題」適用於形成性評量，以口頭問答來評估學生的學習成效，並適時要同學做分組口頭報告，亦是運用口語評量做為總結性評量。如本教學實驗要各組上台作自編歷史故事接龍、不同立場的辯論、感想分享、搶答比賽等。至於評量的標準有台風、內容的創意、內容是否切題、表達的清晰度等方面。

4. 評估實施成效

在實施多元評量過程中，學生對於評量者的多元化往往感到陌生而新奇，以往學生都是被評量的一方，一旦讓他們可以自評、互評，他們會感到惶恐，不知如何評分，也擔心同學互相評分會有不公平的現象。教師需要花時間向學生說明、討論多元評量的用意及評分的標準，教師也要設計一些防止同學會互相串通的評量方式，一段時間後同學才會慢慢適應多元評量的評分方式。

小組分工評量的落實與否是組員是否確實參與活動的關鍵，尤其一些原本學習成就較高的學生常常擔負較大的責任、表現也最多，教師要隨時依據小組分工評量的結果調整各組成員或座位，以便讓每個組員都參與課程活動。

在各學習單的評定量表部份加入個別努力程度的評量，這對於有些在紙筆測驗方面表現不佳，因而沒有信心的學生也可以從

²³⁹ 李坤崇，〈多元化教學評量〉，第 192 頁，台北心理出版社，民國 91 年 4 月。

中獲得成就感及感受到關心，若再加上評語或當著全班加以口頭的鼓勵，學生在接下來的教學活動中會更積極的參與。

在與學生的個別訪談中，有些原本學習成就高的學生，尤其是女同學會抱怨多元評量讓他們的作業份量變重了、對講求標準答案的紙筆測驗幫助不大、不如多做與統編本教材有關的練習、有些組員甚少為該組盡心力等，這些都是需要再多花時間與學生溝通才能引導學生瞭解多元評量與創造思考教學的目的，漸漸矯正學生唯有紙筆測驗才是評量的觀念。

在實施教學實驗開始之初時，教師應先告知同學評量的項目、標準、評量過程中會根據各評分表累積各組的得分，最後在學期末根據各小組的表現頒獎，整個教學實驗過程中，大多數小組成員會因此而發揮極大凝聚力，教師的角色則是在旁提醒引導，各小組為了團體的榮譽會各自約束要求組員，連以往對紙筆測驗不是很在乎的同學，都對多元評量的評分結果極為在意，彼此督促組員同心合作、發表意見、完成作業等。

第二節 教學實驗問卷的設計與結果

一、問卷的設計

本教學實驗參考前人研究所設計的學生問卷，其中所有題目都和本研究目的相符合，即以研究目的的有關變項為根據，每一個變項分別擬定幾個問題加以測量，問卷經過專家修正後，在研究的母體中選取五至十人填答以為預試，大約經二週後，改以訪問方式進行，然後比較填答和訪問結果的一致性而定信度的高低；至於效度，則請教專家分析每個題目的「表面效度」，完成的問卷見附錄四。

本問卷調查的對象包含實驗組與對照組的兩班學生，問卷內容有三部分：基本資料、封閉性及半封閉性題目、開放性題目。其中兩班所實施的問卷內容在前兩部分是相同的，藉以對照在實施不同教學法下，學生的感受與改變。至於實驗組問卷在開放性題目方面，針對創造思考的歷史教學設計有關教學法的建議與感受、同儕互助合作及潛力發揮情形等題目，以便更深入、詳細的了解學生對此教學法的反應，進而作為改進教學的參考。

本問卷發出51份，有效問卷為全部51份，達100%。

二、問卷的結果

(一) 量化的探討

如上所述，實驗組及對照組在基本資料、封閉性及半封

閉性題目的問卷內容是一樣的，其中第二部分的前三題是調查學生之

前的歷史學習經驗，自第五題起是針對教學法所設計的題目，如：實施教學後同學討論情形如何？有沒有學習到思考方法及喜歡思考？有沒有引發創意？對於學習歷史學科知識有沒有幫助？對於培養分析判斷力有沒有幫助？學生是否成上課主角？有沒有更喜歡上歷史課？對歷史是否更有興趣？下學期希不希望再上類似課程活動？以下即針對有關學習效果的問卷四至十六題，將實驗組及對照組所得到的平均數與標準差摘要如表5-13：

組別	人數	平均數	標準差
實驗組	26	36.8	7.03
對照組	25	28.9	4.16

表5-13 實驗組與對照組在問卷統計方面的平均數與標準差

由表5-12可以明顯的看出，實驗組（實施創造思考的歷史教學）在學習效果方面是高於對照組（實施傳統教學），再計算t檢定值之後的結果， $t = 4.1823$ ，達到0.5的顯著水準，表示實驗組學生認為此次實驗教學有達到創造思考的歷史教學目標，讓他們在學習動機、學習成效的認知和情意方面皆有所增強。

（二）質性的分析

首先是調查學生之前有過不同於傳統教學的上課經驗比例有多高。結果在二班共51個學生中有36個學生有類似創造思考教學的上課經驗，比例高達71%，其中有24個人（67%）肯定這樣

的上課方式是比較新鮮有趣的，14個人（39%）認為可以培養合群的態度，6個人（17%）則承認它可以發揮個人長處。相對於對它負面的印象如會讓同學間意見不合（11%）等，可知大多數同學以往的上課經驗是相當不錯，對於創造思考教學的成效採取正面肯定的態度。

再者，是調查關於教學實驗前歷史科學習及接觸的狀況。首先，是兩班學生平時對歷史相關書籍報導、展覽的接觸程度，如圖5-2所示，實驗班有74%、對照班有40%的同學平時有自發性接觸歷史相關知識與資訊的習慣。至於在對歷史學科本質的了解方面，是否認為歷史是一門可以能培養創意、思考能力的科目，如表5-14所示，兩班同學有64%及85%的同學大致認可歷史能培養創意、思考能力。筆著接著詢問對於歷史課程的喜歡程度，如表5-15所示，兩班有64%及77%認為自己是喜歡上歷史課的。至於原因的調查結果如圖5-2所示，勾選「個性興趣」答案的人數最多，占40%的比例，其次是教師的人格特質²⁴⁰有27%的比例；另有15%及13%分別選擇課程內容安排及學習的成就感。由這3個題目的選擇答案看來，可以看出大多數學生對歷史學科的喜愛程度是蠻高的，因為認可這門學科的目標，甚至會自發性的接觸歷史相關的知識與資訊，內化成為生活中的一種習慣與興趣。若是配合適當的教學方法與教材內容，相信歷史這門知識會更受學生們喜歡的。而在各個題目的作答結果方面，實驗班採取同意的比例比對照組高的多，因為此一問卷是在教學實驗後所做的，也許可以看作是實驗班同學對於創造思考教學法運用於歷史教學的肯定。

²⁴⁰ 包含教師個人的言行舉止、上課風格和教學方法的運用等會增強學生學習動機的特質。

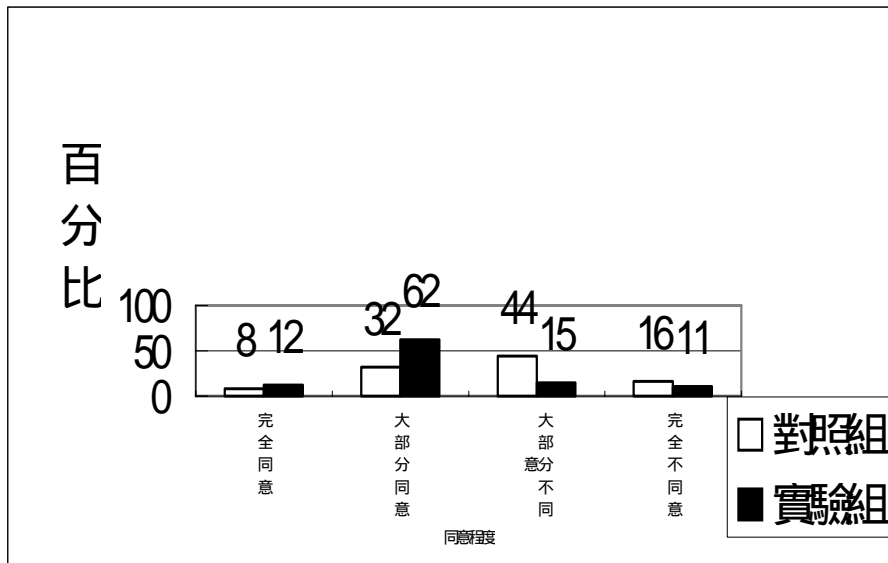


圖5-2 兩組平時閱讀歷史相關報導級書籍比較圖

表5-14 實驗教學前對歷史學科本質的了解

	完全同意	大部分同意	大部分不同意	完全不同意	合計	
					A + B	C + D
對照組	12 % (A)	52 % (B)	28 % (C)	8 % (D)	64 %	36 %
實驗組	58 % (a)	28 % (b)	24 % (c)	0 (d)	a + b	c + d
					86 %	24 %

表5-15 實驗教學前對歷史課程的喜歡程度

	完全同意	大部分同意	大部分不同意	完全不同意	合計	
					A + B	C + D
對照組	32 % (A)	32 % (B)	22 % (C)	14 % (D)	64 %	36 %
實驗組	54 % (a)	23 % (b)	23 % (c)	0 (d)	a + b	c + d
					77 %	23 %

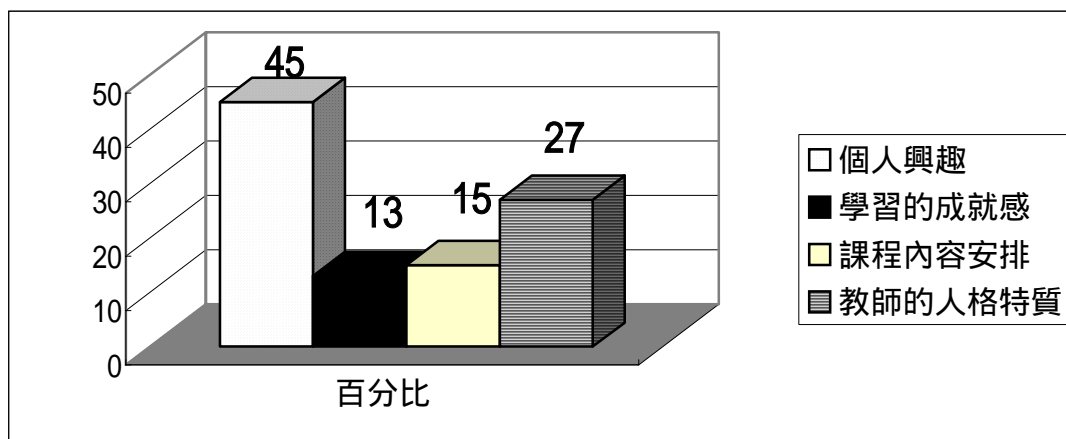


圖5-3 學生喜歡歷史課程的原因

至於兩種教學法是否達到歷史學科之認知目標，如表5-16所示，在增進思考能力的方面，實驗班及對照班分別有73%、68%的比例認同各自教學法可學到思考方法，可見創造思考教學法在增進思考力的方面略勝一籌，但因為筆者在傳統的講演式教學時亦會介紹多方觀點，以引發學生思考，只是未如實施創造思考教學法時讓學生發表意見，故對照組的比例也不低。若再進一步由其引申的下一個題目 - 「增進思考的有趣度」來看，如圖5-4所示，實驗組的教學方法讓高達81%（完全同意的50%、不完全同意的31%）的學生感覺到思考是一件有趣的事，對照組只有60%的比例，且喜歡的程度上亦不如實驗組，可見創造思考的歷史教學讓學生成為學習的主角，這使學生不再是被動式的接受老師的思考模式，讓學生感受到思考的海闊天空及樂趣。接著在增進創造力的調查方面，如表5-17所示，實驗組的同學有高達88%都肯定創造思考的歷史教學可以引發許多有創意的想法，對照組只有52%的比例，這在在證實創造思考的歷史教學成效是相當受到同學認同的。再詢問是否能增進對歷史學科知識的學習方面，如圖5-5所示，實驗組及對照組同意的比例不相上下，各占77%、76%的比例，若再對照之前兩班在「歷史學科能力測驗」及段考成

績的表現，可印證學生的認知與實際紙筆評量結果相差不大；最後在問到增進分析判斷能力與否，如表5-18所示，實驗班有高達84%的比例，高於對照組的68%比例。

表5-16 增進思考能力

	完全同意	大部分同意	大部分不同意	完全不同意	合計	
					A + B	C + D
對照組	12 % (A)	56 % (B)	24 % (C)	8 % (D)	68 %	32 %
實驗組	38 % (a)	35 % (b)	12 % (c)	15 % (d)	a + b	c + d
					73 %	27 %

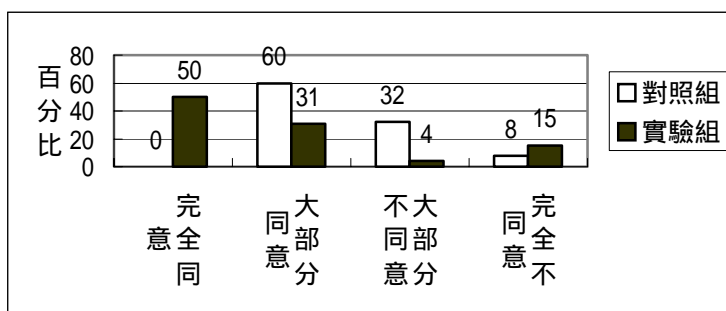


圖5-4 增進思考的有趣度

表5-17 增進創造力

	完全同意	大部分同意	大部分不同意	完全不同意	合計	
					A + B	C + D
對照組	14 % (A)	48 % (B)	36 % (C)	12 % (D)	52 %	48 %
實驗組	46 % (a)	42 % (b)	8 % (c)	14 % (d)	a + b	c + d
					84 %	16 %

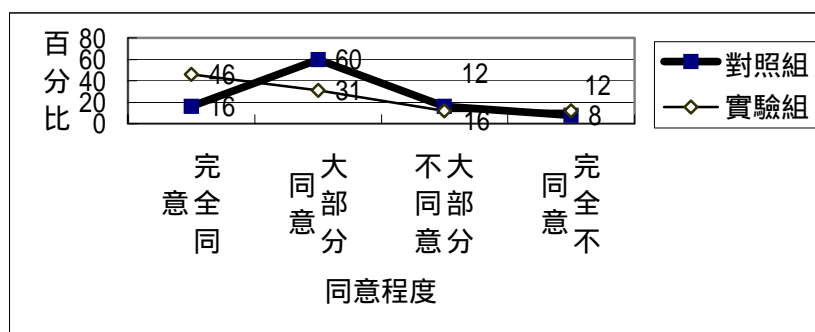


圖5-5 促進學科知識的學習

表5-18 增進分析判斷能力

	完全同意	大部分同意	大部分不同意	完全不同意	合計	
					A + B	C + D
對照組	4 % (A)	64 % (B)	24 % (C)	8 % (D)	68 %	32 %
實驗組	38 % (a)	46 % (b)	8 % (c)	8 % (d)	a + b	c + d
					84 %	16 %

總之，經由以上的問卷調查結果，可知創造思考的歷史教學在認知方面的目標，如促進思考力、創造力及分析判斷力等皆深受實驗班學生的肯定，唯獨在歷史學科知識的學習方面與對照組相差不大，這是下一次實施創造思考教學時可以努力的方向。

再來是針對兩種教學法是否達到歷史學科之情意目標，如表5-19所示，在上課熱烈討論的程度上，實驗組有77 %的比例認為班上同學相當投入在分組或全班的討論上，遠遠高於對照組的32 %比例；而經由下一題的調查結果，如圖5-6所示，實驗組的學生亦有81 %因為創造思考的歷史教學而喜歡大家上課時熱烈討論的氣氛，相對的對照組只有64 %的比例，這在在顯示創造思考教學營造了一個讓學生勇於、樂於發言的教學情境，讓學生不再只是單向式的接受老師的講授教導，師生間是有對話交流的。故在同學上課參與程度的調查結果方面，如表5-20所示，77 %比例的實驗組學生認為他們不再是教學過程中的配角，其中50 %的學生是積極參與整個教學過程，對照組則是有60 %的學生有參與感，但程度亦不如實驗組（完全同意的人數是0）。接下來詢問上課氣氛的問題，如圖5-7所示，54 %的實驗組同學感受到上課氣氛十分活潑有趣，對照組只有4 %的人有相同的感受，累計統計起來，實驗班更有高達92 %的人認同上課氣氛是趣味盎然，對照組則只有44 %的人有如此的感受。故如表5-21、圖5-8所示，自然實驗組喜歡上歷史課、喜歡歷史這門學科的程度就以92 %、88

% 的比例高於對照組72 %、60 % 比例。但無論實驗組或對照組都有高達85 %、82 % 的比例希望下學期能採取創造思考教學的課程模式，可見若能帶動上課氣氛，學生的學習興趣也會隨之高昂，上課就不再感到痛苦而能有學習收穫的喜悅了！

表5-19 上課熱烈討論的程度

	完全同意	大部分同意	大部分不同意	完全不同意	合計	
					A + B	C + D
對照組	0 (A)	32 % (B)	52 % (C)	16 % (D)	32 %	68 %
實驗組	42 % (a)	35 % (b)	11 % (c)	12 % (d)	a + b	c + d
					77 %	23 %

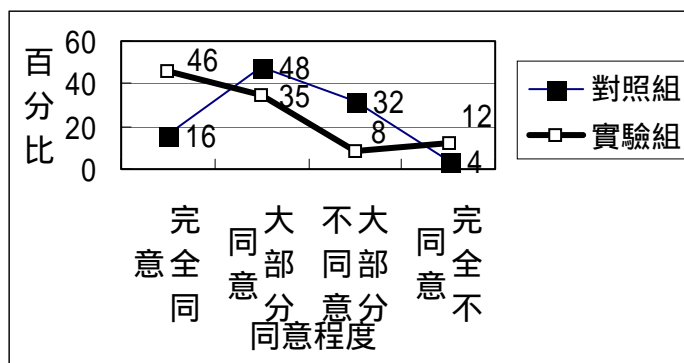


圖5-6 喜歡上課討論的程度

表5-20 同學上課的參與程度

	完全同意	大部分同意	大部分不同意	完全不同意	合計	
					A + B	C + D
對照組	0 (A)	60 % (B)	36 % (C)	4 % (D)	60 %	40 %
實驗組	50 % (a)	27 % (b)	19 % (c)	4 % (d)	a + b	c + d
					77 %	23 %

圖5-7 上課氣氛活潑有趣的程度

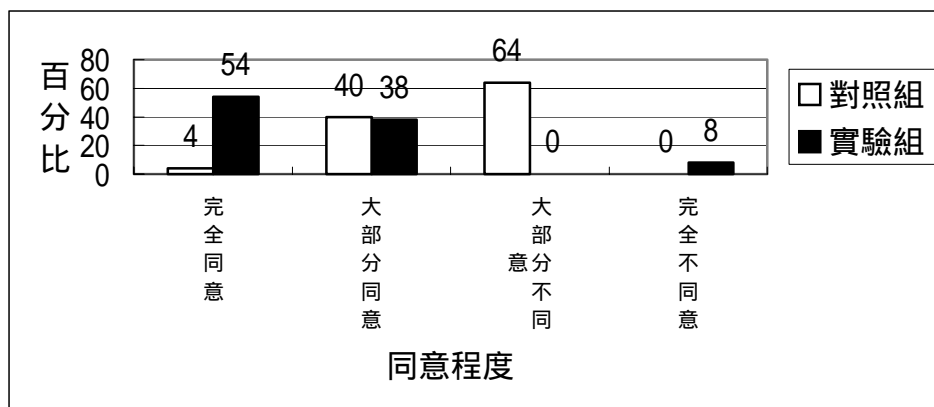


表5-21 喜歡上歷史課程度

	完全同意	大部分同意	大部分不同意	完全不同意	合計	
					A + B	C + D
對照組	12 % (A)	60 % (B)	20 % (C)	8 % (D)	72 %	22 %
實驗組	54 % (a)	38 % (b)	0 (c)	8 % (d)	a + b	c + d
					92 %	8 %

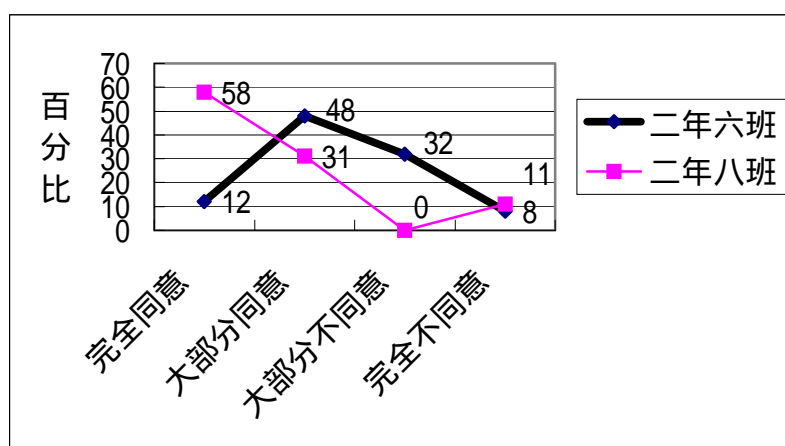


圖5-8 喜歡『歷史』這門學科的程度

表5-22 希望下學期採取討論、活動等上課方式的意願

	完全同意	大部分同意	大部分不同意	完全不同意	合計	
					A + B	C + D
對照組	20 % (A)	64 % (B)	4 % (C)	12 % (D)	84 %	16 %
實驗組	43 % (a)	42 % (b)	8 % (c)	7 % (d)	a + b	c + d
					85 %	15 %

總之，從以上情意目標的探討可知：採取創造思考教學法明顯的提升了學生上課的參與感，無形中藉由熱烈的討論、發言、活動塑造了一個讓他們的學習意願增加，更能深入歷史情境的教學環境，心胸自然開闊寬廣，這正是創造思考的歷史教學在情意層面所欲達成的目標。

最後在開放式題目方面，實驗組及對照組的部分回答可見附錄六、七。從其中更加清楚的知曉實驗組同學對教學實驗進行細節的感想，如同學們都肯定許多活動設計很能帶動同學共同參與，並能發揮合作精神及創意。但也有一些女同學基於這樣的上課方式，以致對課文內容未多加講述、減少單一標準答案的評量練習，以及上課秩序較難控制等因素，顯然對分組討論的效果打了相當大的折扣。意外的是男同學皆是一致相當認同這樣的上課方式，這也可以顯示男女同學對上課的期許大不相同。但也讓筆者思索實施創造思考教學除了要事先仔細規劃課程設計，上課分組時如何讓同學發揮榮譽感以減少混水摸魚的情形發生，如何讓職務分工執行徹底等都需要詳加考量。並且要在課前多與同學溝通創造思考歷史教學目的之所在，矯正同學認為唯有講授單一標準答案的課程內容才是學習的錯誤觀念，進而認同在教學過程中培養思考方法及創造力才是學習的真正目的，以釐清同學對分組討論教學耗時又無益於紙筆評量的成見。

至於實驗組在開放性問卷中有兩題是經過創造思考的歷史教學過程後，同學之間的對話篇：「你覺得在這學期的歷史教學活動中，最能和同學們分工合作的人有誰？你有什麼話想對他

說？」、「你覺得在這學期的歷史教學活動中，發言時表現的最好的人有誰？你有什麼話想對他說？」在這兩題的回答中可以看出有的特長是在溝通協調組員間分工合作、有人在口才台風方面表現突出、有人則是創意層出不窮，同學說出對同組組員的感謝之意、也寫出對班上同學努力付出的鼓舞及嘉許，從中感受到經過創造思考的歷史教學後，同學間無論在認知能力及團隊精神上都增進許多。

在對照組的開放性題目方面，就如之前封閉性題目的回答結果一般，雖然也肯定傳統的講演式教學在講述課文內容相當詳盡，但大多數同學希望課程能更活潑有趣，多補充不同的歷史觀點以及可以採取分組討論等，更加顯示傳統的講演式教學在這些方面的不足，需要兼採創造思考教學法加以彌補。