

國立臺灣大學工學院工程科學及海洋工程學系



碩士論文

Department of Engineering Science and Ocean Engineering

College of Engineering

National Taiwan University

Master Thesis

結合線上測驗診斷的分層教學決策輔助機制

A Support Mechanism for Decision-Making of Stratified Teaching

Using Online Examination Diagnosis

林威言

Wei-Yen Lin

指導教授：丁肇隆 博士

張瑞益 博士

Advisor : Chao-Lung Ting, Ph.D.

Ray-I Chang, Ph.D.

中華民國 104 年 7 月

July, 2015

誌謝



能夠順利完成本篇論文，感謝指導教授丁肇隆教授與張瑞益教授的悉心指教，口試委員王家輝教授、吳文中教授與張恆華教授惠賜寶貴意見。此外，特別感謝烏日國中數學科全體教師費心參與教學實驗，誠致基金會均一教育平台方董事長以及逸文、倚恩在試題方面的協助，沒有兩個合作單位的慷慨協助，本論文概念便無法實現。

回顧研究所生涯，感謝 125A 實驗室的學長姐信義、忠原、育賢、億鑫、宇翔、勤昇、哲論、玟君與學弟妹忠毅、佳儒、建晴、宣儀、乃藩、郁琪、加容、玉琪、怡萱在各方面的協助、中央研究院資訊科學所學長佑鈞與宏敏在研究上的指導、工科所、資工所與醫工所授課教授與助教專業知識的提點。更可貴的回憶是與眾多優秀同屆同學韋成、雯方、竹君、瑞華、宏達、千雅、佳柔以及蔚茵、衍綺、岳銘、詩涵、陳熙患難與共的革命情誼，這一切已是榮幸。而來自各地的持續祝福，包含新竹台北的堂哥與堂弟、紐約的堂妹，以及和英國的 HC 與 MH、日本的 I、韓國的 C 與新加坡的 Z 之間的遙相鼓勵，也已成爲浪漫。

也在此衷心感謝我的家人在求學過程中無條件的支持，將所有的圓滿獻予她們。

今日寫下誌謝，就是一種向椰林大道告辭的手勢。r02 級同學們，我們將站在台大研究生涯建立的基礎上，大步向前。愛國愛人，成爲更好的人。

威言

2015 年 8 月

中文摘要



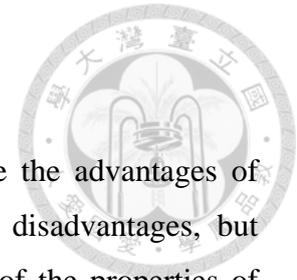
傳統教學過度忽視學生學習能力的差異，個人化教學又常因為教師人力與時間資源不足而失敗，在這兩個極端之間，分層教學成為一個折衷兩方優點的選項。而線上測驗易於重新組合、資料分析的特性，正足以解決分層教學造成教師額外資訊處理負擔的困境。

本研究提出一套結合線上測驗診斷的分層教學決策輔助機制，為有意施行這種教學方式的教師減少處理資訊的負擔。本機制主要由三個模組構成：選題模組、學習成就分析模組、成效衡量模組。選題模組採用線性規劃單體法組合試題；學習成就分析模組參考專家意見，建立學生學習成就追蹤指標；成效衡量模組利用二因子變異數分析分層教學成效。

經教學實驗驗證，結合本機制的國中數學分層教學成效，確實比傳統教學優異。除此之外，實驗發現：分層教學成效與課綱單元難易度有顯著相關的交互作用，特別難的單元，分層教學的成效不彰，然而除此之外，分層教學都有很出色的成效。另一方面，分層教學成效與學習能力分群這兩個因子交互作用沒有顯著性，表示分層教學的成效並不特別明顯侷限在特定的學習能力分群。

關鍵字：分層教學、線上測驗、線性規劃、二因子變異數分析

ABSTRACT



Stratified teaching is a new option which enable to integrate the advantages of traditional teaching and personalized instruction and avoid their disadvantages, but teacher needs to spend more time on data management. Because of the properties of easier recombination and data analysis, online examination can be used to reduce the data management burden on teachers.

Using online examination diagnosis, this thesis provides a supporting mechanism for decision-making for stratified teaching to help the teachers who intend to adopt stratified teaching. The mechanism consists of three modules: item combination module, learning achievement module and effect evaluation module. The item combination module uses linear programming simplex method to combine items. The learning achievement module set up students' achievement tracking criteria by interviewing with experts. The effect evaluation module use two-way ANOVA to evaluate the teaching effect.

Experiments show that the teaching effect with the proposed mechanism is better than the traditional teaching. Furthermore, there exists significant interaction correlation effect between stratified teaching effect and difficulty of the course unit. Stratified teaching has better effect on most course units except of especially difficult ones. Besides, there is no significant interaction correlation effect between stratified teaching effect and learning ability group, which shows that stratified teaching effect doesn't depend on specific learning ability group.

Keywords : *Stratified teaching, Online examination, Linear programming, Two-way ANOVA*

目錄



誌謝	i
中文摘要	ii
ABSTRACT	iii
目錄	iv
圖目錄	vi
表目錄	vii
Chapter 1 緒論	1
1.1 研究動機與目的	1
1.2 論文架構	2
Chapter 2 文獻回顧	3
2.1 差異化教學與分層教學	3
2.2 選題策略	6
Chapter 3 研究方法與實驗設計	11
3.1 系統概觀	12
3.2 選題模組實作方式與實驗設計	13
3.2.1 試題難度與鑑別度分析	13
3.2.2 選題策略演算法：線性規劃單體法	16
3.2.3 選題策略演算法實驗設計	18
3.3 學習成就分析模組實作方式	22
3.4 成效衡量模組實作方式與教學實驗設計	22
3.4.1 成效衡量模組實作方式：二因子變異數分析	23

3.4.2	分層教學實驗方法：準實驗設計法.....	28
3.4.3	分層教學實驗流程設計.....	30
Chapter 4	實驗與結果討論.....	32
4.1	分層教學實驗操作.....	32
4.1.1	線上測驗試題來源.....	32
4.1.2	前後測實驗操作.....	33
4.2	分層教學實驗分析.....	36
4.2.1	教學方法與課綱單元二因子變異數分析.....	37
4.2.2	教學方法與學習能力分群二因子變異數分析.....	40
4.2.3	加試題實驗結果分析.....	41
4.3	選題策略實驗操作與分析.....	43
4.3.1	K=10 前測數據來源實驗結果.....	45
4.3.2	K=10 後測數據來源實驗結果.....	47
4.3.3	K=5 前測數據來源實驗結果.....	49
4.3.4	K=5 後測數據來源實驗結果.....	50
4.3.5	選題策略實驗結果綜合討論.....	51
Chapter 5	結論與未來展望.....	53
	參考文獻.....	55
	附錄一.....	57



圖目錄



圖 2-1	網路流量選題策略演算法架構圖	8
圖 3-1	系統建構流程	12
圖 3-2	“Anscombe's Quartet”示意圖	20
圖 3-3	選題策略演算法實驗設計步驟	21
圖 3-4	學習成就分析模組輸出格式範例（一個課綱單元）	22
圖 3-5	分層教學實驗流程圖	31
圖 4-1	均一教育平台使用者來源比例	33
圖 4-2	實驗組與控制組對照箱型圖	36
圖 4-3	各課綱單元進步幅度點狀圖	37
圖 4-4	各課綱單元實驗組與對照組進步幅度折線圖（一）	38
圖 4-5	各課綱單元實驗組與對照組進步幅度折線圖（二）	38
圖 4-6	高中低分群進步幅度箱型圖	40
圖 4-7	K=10 前測數據來源的難度鑑別度相關係數折線圖	45
圖 4-8	K=10 後測數據來源的難度鑑別度相關係數折線圖	47
圖 4-9	K=5 前測數據來源的難度鑑別度相關係數折線圖	49
圖 4-10	K=5 後測數據來源的難度鑑別度相關係數折線圖	50

表目錄



表 2-1	古典測驗理論與試題反應理論優缺點比較	10
表 3-1	本研究系統模組說明	13
表 3-2	線性規劃常見解法比較表	17
表 3-3	二因子獨立測量實驗設計架構圖	23
表 3-4	二因子 ANOVA 數據表	25
表 3-5	二因子 ANOVA 摘要表	26
表 3-6	二因子 ANOVA 摘要表範例	28
表 4-1	前後測試題來源與數據對照表	34
表 4-2	實驗組與控制組教學輔導方式	35
表 4-3	課綱單元與教學方法二因子 ANOVA 摘要表	39
表 4-4	學習能力分群與教學方法二因子 ANOVA 摘要表	40
表 4-5	後測與加試題學習成就比較表	41
表 4-6	各百分制區間學生的加試題表現	42
表 4-7	實驗難度限制條件設定表	44
表 4-8	K=10 前測數據來源的各項實驗數據結果	45
表 4-9	K=10 後測數據來源的各項實驗數據結果	47
表 4-10	K=5 前測數據來源的各項實驗數據結果	49
表 4-11	K=5 後測數據來源的各項實驗數據結果	50
表 5-1	烏日國中實驗組學生各次數學科測驗表現	53

Chapter 1 緒論



1.1 研究動機與目的

「因材施教」是自孔子以來流傳數千年的教育理念，台灣的十二年國民基本教育更將「因材施教」列為核心理念，與「有教無類」、「適性揚才」、「多元進路」、「優質銜接」等理念並列，擬定於 103 學年度開始實施。因應這樣的趨勢，教師就需要具備差異化教學 (differentiated instruction) 概念，針對不同學習能力程度的學生，提供多層次的學習支援 (learning support)，使得各程度的學生都能夠獲得適性與適才的學習輔導。

雖然「因材施教」已是有數千年歷史的教育理念，然而在教學實務上，仍有許多困難。最大癥結便是如何在既有編班體制與多層次教學輔導之間取得平衡，一方面傳統教學過度忽視學生學習能力的差異，另一方面全然施行差異化教學，又往往因為教師人力與時間資源不足，難以為繼，最後半途而廢，成為實驗性質的教育理想。

在這兩個極端之間，分層教學成為一個折衷兩方優點的選項。分層教學的教學輔導單位以「群」(group) 為單位，常見的分群為高、中、低三群，強調教師只需要對各分群提供對應的輔導與追蹤機制，就能達到改善教學成效的目的。

即便如此，相較於傳統教學，分層教學對於教師仍然是較大的負擔。分層教學過程中，分群、輔導、追蹤、再分群如此週而復始的循環之中，教師需要更快速、更細緻處理資訊的能力，包含設計不同難度的試卷、追蹤學生學習成就的起伏等等，才能做出對應的分層教學決策。簡單來說，分層教學對教師最大的挑戰就在於設計不同難度的試卷與動態追蹤學生學習成就的發展狀況。



本研究擬定提出一套分層教學決策輔助機制的實作方式，在教學施測之前，將線上測驗的試題作為題庫 (item bank)，利用選題策略演算法組合出不同難度的試卷；施測之後，這些線上測驗的作答紀錄，一方面作為學生自動化學習能力分群、學習成就追蹤的依據，一方面作為教師了解分層教學成效的參考。

線上測驗的試題易於重新組合、資料分析的特性，正足以解決分層教學造成教師額外資訊處理負擔的困境。如此一來，教師可以更快速產生不同難度的試題，也能更細緻、更動態地追蹤學生學習成就的情況。

1.2 論文架構

本論文架構，第一章敘述本研究的動機與目的，說明分層教學的價值與如何透過本機制輔助教師施行分層教學，達到分層教學目標；第二章敘述分層教學與選題策略的相關文獻，一方面說明分層教學注重的內涵，另一方面說明與比較各選題策略，以便選擇最適合本研究的演算法；第三章分為四個部份，第一部份介紹系統建構流程與系統三個組成模組，第二部份介紹選題模組的實作方式，第三部份介紹成效衡量模組的實作方式以及驗證分層教學成效的實驗設計，第四部份則是說明驗證選題模組能力的選題策略演算法之實驗設計。第四章敘述分層教學與選題策略兩個實驗的操作方式與結果分析，第五章則是歸納實驗結果，作出結論，並說明未來可再延伸的研究方向。

Chapter 2 文獻回顧



2.1 差異化教學與分層教學

分層教學 (hierarchical teaching 或 stratified teaching) 的前身是差異化教學 (differentiated Instruction)，差異化教學是教師依據學生的準備度 (readiness)、興趣 (interest) 及學習概況 (learning profile) 設計不同的教材教法與評量[1]，以確保學生的學習處於持續進步的狀態。

從起源來說，差異化教學的概念最早起源於美國，20 世紀初之前的美國基礎教育是不分齡教學，之後隨著工業化與都市化發展，才開始出現分齡教育、分級學校。接著，隨著種族融合，學校內種族組成多元化；以及 1975 年「身心障礙人士法案」(IDEA) 法案通過，身心障礙學生也有回流普通班級的權利與趨勢，因此自 1980 年代以來美國基礎教育已經進入融合教育的階段，一個學校、班級的學生可能會有非常歧異的背景[2]。

從定義上來說，差異化教學可泛指為不同學生調整教學環境與教學實況，以開創不同學習需求的適性學習經驗。差異化教學與美國 1970 年代提倡的「個人化教學」(personalized instruction) 並不相同。個人化教學強調的是針對每位學生設計個人化教學計劃，使得學生可以自行掌握學習速度。個人化教學最大的困境還是在於以一般教室內的學生人數，要達到個人化的目標太過困難[3]。再者，強調個別化的同時，也會造成忽略同儕合作與班級教學品質的困境。相對的，差異化教學則強調以小組、次團體為教學策略的目標單位。

許多學者曾以多種不同方式研究差異化的教學，在學習成就方面的成效。以中學階段來說，在 Tieso 的研究[4]中，教師們分別使用 3 種分組方式與 2 種課程模



組進行教學實驗，發現高學習能力的學生在普通班級中，透過彈性小組（flexible small group）分組學習和差異化的課程內容，能完成挑戰性更高的學習任務，而且學習成效顯著高及其他學生。Baumgartner 等人[5]則研究閱讀能力及態度與差異化教學之間的關係，發現在一般課堂教學現場，透過差異化的教學設計，如彈性分組、提供多種閱讀媒材等，能有效提昇不同程度中學生的閱讀能力和態度。此外，Strong 等人[6]則在中學五年級的數學課，針對 4 種不同的學習方式的學生設計差異化教學活動，發現可有效提昇每位學生的數學學習成就。

分層教學是更簡化、也是更著重學科學習成就的差異化教學。它擷取差異化教學當中最簡單的作法，強調以層級為單位，主要以學習成就作為區分層級的方式與改善的目標，容許維持學校體系下的班級編制，強調持續追蹤學習成就與保持層級流動的作法。也因為實行上比差異化教學來得簡便，在一些教學進度壓力較大的教學體制如中國，特別受到注目[7]。

分層教學強調的是在教學中對於各學習能力分群設定不同層次目標，施以不同方式教學和輔導，訂定不同方式的測量基準，使得各個層級的學生都能有所進步。常見具體作法是：一、了解差異之後，將班級中的學生劃分為不同層級；二、針對差異區分目標；三、因材施教；四、針對不同分群考核成果[8]。分層有兩個最重要的原則：第一、不是固定的，學生在層級之間的變化主要是根據學習情況，進步顯著時上調，退步時下移。第二、作法上必須顧及學生心理，避免造成自暴自棄的反效果。

分層教學可分成班內分層與分層走班（class-selection）方式，差別在於班內分層保留學校行政編制下的班級，而分層走班則是將散佈在各班同一層級的學生聚集在一起，例如高分群在一班、中分群與低分群各在一班等等[7]。此外，分層教



學在教學實務上有四大環節：學生分層、分層備課、分層授課、分類指導。一、學生分層：是分層教學的基礎，分層的標準有很多種，視教師教學改善的目標而定。二、分層備課：教師必須確立不同層次的教學目標，把握基本標準與較高標準的差別，再設計分層教學的步驟。三、分層授課：把握授課的學習起點，處理好知識之間的銜接過程，減少各部份的學習坡度。四、分類指導：除了「因材施教」的輔導，教師還可以考慮手上的資源加入作業、課外小組合作等等，才能促成學生由較低層次的分群轉向較高層次的分群[9]。

此外，施行分層教學時，也應注意三個常造成效果不彰的原因：一、分層標準不科學：學生分層標準是否客觀、科學，影響著整個分層教學的效果。涇渭分明公開化的分層，會使得未能進入理想層次的同學產生被歧視的感受，喪失學習的積極性。二、目標定位不合理：有些教師儘管施行分層教學，但在教學目標上仍過於統一，只是為了短期間拉高成績，忽視某些落後學生的學習困境，最後容易造成效果不彰。三、評價方式不恰當：應採取發展性與綜合性評價，要著重「以人為本」原則，促進人才是分層教學的本意[10]。

目前在各種教學領域，已累積一些分層教學的研究實例。溫州的中學曾在一年級施行新生的英語分層教學，結果發現，高分群與中分群學生積極度與自信心都有提昇，但是對於低分群的學生則沒有產生正面影響。瀋陽市更有職業學校連續 5 年採用分層教學，對於高中低分群學生皆進行分層輔導，結果顯示成績越好的學生，在這個過程中進步最快。不過，北京的中學曾在一個年級的英語教學，採用分層走班的教學實驗，打破班級編制，將同一層次的同學歸入同一班，結果顯示低分群同學自尊受影響，使得學習成就不升反降[11]。這些文獻顯示分層走班成效不見得比傳統教學好，而班內分層的成效則大部分都比傳統教學的成效來得

優異。

此外，分層教學不僅能影響學習成就，還包括學習態度、興趣等層面。Yu [12] 針對大學英語課程施行分層教學實驗，分層方式是教師參照前測成績和實驗前的問卷調查結果，根據學生的英語水準與學習態度等方面的情況，將學生大致分為優、中、差三個層次，分層結果不向班級公佈。分層輔導方式除了教師設定不同分層目標之外，也幫助學生建立 3 到 4 人的合作小組。經實驗結果發現，能夠顯著提高學生的英語水準、學生英語學習的自信心，還能緩解學習焦慮情緒。

在台灣方面，李達勝[13]曾將分層教學應用在桌球課程，作法是：教學之初先將學生依球技分為一般生、進階生、優等生三個層次，一般生過去沒有接觸過桌球，進階生對桌球過去曾打過桌球，但未受過專業訓練，偏好自由發揮，而優等生過去曾經受過訓練，球技有一定的水準。不同的層次採取不同的學習目標，一般生著重基本技術訓練，進階生著重精進技巧，優等生著重戰術設計。讓層次相近的學生經過一段時間的練習後，再依據比賽評鑑的結果，重新調整層次。

相對地，分層教學就如許多新式的教學法一樣，並非沒有缺點。從先前的描述可得知，對於教師來說，施行分層教學主要需要建立資訊有這階段學生學習成就、這階段高中低分群的學生名單、學生各階段學習成就比較、高中低分群學生的移動情況，也因此需要花費更多心力去處理與消化資訊，這使得分層教學儘管已經簡化了差異化教學的難度，但是對於有日常教學進度壓力的教師仍然是一大負擔，而這也正是本研究希望介入與協助改善的部份。

2.2 選題策略

針對選題或是編制試題的問題，已累積不少相關研究，大致可分為三大面向：

一、試題參數的理論基礎



二、採用的演算法

三、試題的呈現方式

其中，第一點與第二點與本研究相關，下文分別描述其大致內涵。

以第一點來說，試題參數引用的測驗理論可分成「古典測驗理論」(Classical Test Theory, CTT) 以及「試題反應理論」(Item Response Theory, IRT) 兩大類別[14]。古典測驗理論以心理學的計量理論為架構，主要以真實分數為核心。而試題反應理論則假設受測者的答題情況可用潛在特質或能力，加以預測或解釋；另一方面，受測者能力的程度與作答情形，也可以用數學函數描述[15]。

以第二點來說，主要有網路流量 (network flow)、基因演算法 (genetic algorithm)、動態規劃法 (dynamic programming)、貪婪演算法 (greedy) 與線性規劃 (linear programming) 等演算法[16]，以下介紹各演算法的代表性研究。

採用網路流量演算法的相關研究，可以 Armstrong 等人的研究為代表[17]。將測驗題庫轉換成網路流量問題的架構，將題庫試題設定為供應節點 (supply)，不同限制條件的題數設定為需求節點，接著以網路流量相關演算法解題。其架構可表示如圖 2-1，其中 S_i 代表題庫編號第 i 題， D_j 代表不同限制條件的試題集合， x_{ij} 代表是否選入此試題 ($x_{ij}=1$ 表示選入， $x_{ij}=0$ 則否。)， c_{ij} 代表該試題的訊息量。這個方法的缺點是當限制條件為量化屬性時，轉換成網路流量的架構就會有困難，再者也並非所有限制條件與題庫之間的關係都能用網路流量的架構表示。

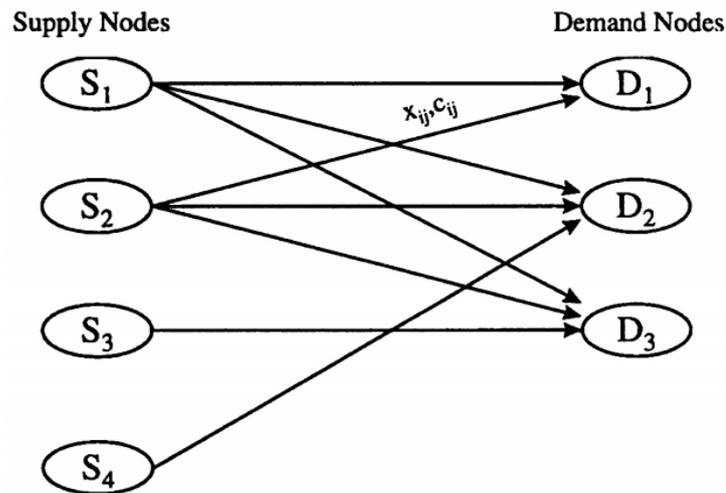


圖 2-1 網路流量選題策略演算法架構圖

採用基因演算法的相關研究，如 Sun 等人[18]使用二進位編碼的基因演算法來處理題目組合問題，設定染色體的基因數為題庫中的數量。亦即將染色體表示為 $[x_{k1}, x_{k2}, \dots, x_{kN}]$ 的序列，而每個基因 x_{kj} 為二進制型態，以 1 代表被挑進試卷中，0 則代表未中選。雖然不需要限制演化操作，但是這個方法為了符合試卷題數限制，使得搜尋空間有過多不合法的解且浪費搜尋時間。Hwang 等人[19]則以實數型基因演算法處理選題的問題，以實數編碼的染色體基因數表示試卷的題數，舉例來說，從 N 種題型的題庫中挑選 M 個試題組成試卷，則其染色體就是 $[x_{k1}, x_{k2}, \dots, x_{kM}]$ ，其中每個基因 x_{kj} 代表入選的試題編號，為介於 1 到 N 的不重複實數。這個實數型基因演算法的優點是可以短時間組合一份試卷，然而為了基因重覆與超界的問題，因此必須對交配與突變過程中，某些操作加以限制，如此一來，使得此基因演算法的演化效能也相應降低。

採用動態規劃法的相關研究，如 Hwang[20]結合了分群 (clustering) 與動態規劃法，每個試題的屬性包含難度、鑑別度、作答時間和課程概念等。首先利用每



位受測者的學習狀態與個人特徵的資料，結合 fuzzy logic theory，計算出每道試題的難度，接著以分群技術將相似概念的題目合為題組，算出試題題組-概念的矩陣，最後以兩個階段的動態規劃法完成選題的目標。根據約 20 個案例的實驗結果，在一定限制條件可以取得不錯的效果。然而這個方法也有兩個缺點，第一是在題庫規模龐大的情況下，會造成計算時間很長；第二是某些限制條件的考量也不夠理想。

採用貪婪演算法的相關研究，如：Swanson 等人[21]結合目標搜尋法（goal seeking）與貪婪（greedy）演算法，以兩階段程序設計選題策略：第一階段是挑選階段，先建構一份符合目標題數的試卷，第二階段則是取代階段，從題庫中挑選一新試題，觀察試卷中的每一題與某一個新選的題目互換之後，試題訊息量與目標訊息量的差距會減少，若確實能減少，則取代互換後能減少最大差距值的題目，接著再從題庫挑出下一個新試題，重複以上步驟，直到取代步驟無法再縮小誤差為止。這個方法的缺點是同一個目標訊息量，卻不一定能求得最佳解（optimal solution）。原因是在某些情況下，題數條件未滿足之前，誤差值就趨近於 0，此時為了符合題數條件繼續選題，導致誤差值反而變大。

採用線性規劃的相關研究，如 Linden 等人[22]提出將試題化為決策變數（decision variables），選題方法化為二元線性規劃（binary linear programming）的架構，如此便能以線性規劃相關的演算法解決。線性規劃最大的優點是利於加入限制條件，也能求得最佳解。因此相較於其他的演算法，具有較佳的彈性，也是在一般情況求取最佳解的理想選擇。

表 2-1 為古典測驗理論與試題反應理論的優缺點比較[14]，這兩種方法各有不同長處，因此在應用的情境亦有差異。以本研究來說，因為分層教學的單位是班

級，使用者是教師，因此古典測驗理論相較之下是比較適用的選項。

表 2-1 古典測驗理論與試題反應理論優缺點比較

	古典測驗理論	試題反應理論
優點	<ol style="list-style-type: none"> 1. 應用最廣。 2. 小規模樣本亦適用。 3. 計算方式教師易於操作與分析。 	<ol style="list-style-type: none"> 1. 試題參數不會相依於受試者能力。 2. 更容易對於不同受試者間的分數進行比較。 3. 對於原始得分相同者，能給予不同的能力值估計。
缺點	<ol style="list-style-type: none"> 1. 各項試題參數如難度、鑑別度等，會相依於受試者的能力。 2. 未考慮受試者的個別差異。 3. 不同試題內容的測驗難以比較。 	<ol style="list-style-type: none"> 1. 統計學公式複雜艱澀。 2. 須配合電腦的套裝軟體程式估計公式參數。 3. 過度趨於數學話導致缺乏質性評量應建立的基準。 4. 必須大樣本的配合，應用性不大，無法適用於一般測驗。

Chapter 3 研究方法與實驗設計



本章將介紹本研究分層教學決策輔助機制的實作方式與實驗設計，探討如何建置出這樣的機制以及如何透過實證方式驗證效果。本章總共分為四個部份：第一部份介紹系統建構流程與系統三個組成模組，第二部份介紹選題模組的實作方式與驗證選題成效的實驗設計，第三部份介紹學習成就分析模組的實作方式，第四部份介紹成效衡量模組的實作方式以及驗證分層教學成效的實驗設計

3.1 系統概觀



圖 3-1 系統建構流程

本系統的建構流程如圖 3-1 所示。首先，以動機為研究起點，藉由閱讀文獻了解相關理論與作法之後，開始尋找實驗合作對象。本系統設計的流程可分成兩部份，基本的系統模組規劃主要由研究者獨立設計（如表 3-1 所示），而各模組功能所呈現數據之種類，則整合專家意見之後，再行設計與實作。

表 3-1 本研究系統模組說明

模組名稱	功能
選題模組	使用者可設定難度區間與重要性順序，系統會自動提供符合條件的試題。
學習成就分析模組	提供高、中、低分群學習成就與學生名單。
成效衡量模組	提供歷次學生學習成就的差異，作為評估教學成效的標準。

系統主要由三個模組構成：選題模組、學習成就分析模組、成效衡量模組。選題模組採用選題策略演算法組合試題；學習成就分析模組參考專家意見，建立學生學習成就追蹤指標；成效衡量模組利用資料分析技術分層教學成效。其中選題模組、成效衡量模組兩個模組的實作方式，將在以下兩節分別闡述之。

3.2 選題模組實作方式與實驗設計

本節將介紹選題模組所採用的選題策略演算法以及驗證選題模組功能的實驗設計。首先介紹演算法所需要的試題相關數值：難度與鑑別度，及其計算方式；接著介紹本研究採用的線性規劃（linear programming）單體法（simplex method），最後說明實驗評估標準與實驗設計。

3.2.1 試題難度與鑑別度分析

試題分析可分成主觀性試題分析與客觀性試題分析。主觀性試題分析通常在試題施測之前，由試題編製者或學者專家判斷每道試題能否達到所欲測量的教學目標、題目語意是否清楚、題目是否有錯別字等部份。客觀性試題分析則在試題



施測之後，藉由受試者作答反應，根據難度 (difficulty) 與鑑別度 (discrimination) 兩項數值指標判斷試題的優劣。另一方面，試題分析又可分為「質化分析」(qualitative analysis) 與「量化分析」(quantitative analysis)，質化分析是分析試題內容與形式，比如內容效度 (content validity)，而量化分析是用統計方法分析題目品質，包括難度和鑑別度分析。

由上述文獻可發現，難度與鑑別度是客觀性與量化試題分析都採用的分析指標，以下分別介紹這兩大指標的內涵。

試題的難度是指題目對於受測者的作答正確容易與否的程度。難度分析可分為採用全體受試者或者採用高分群與低分群受試者兩種計算方式。若採用全體受試者的計算方式，則計分方式是答對某道題目人數與作答總人數的比例，也就是說將答對某道題目的人數除以作答總人數，以 P 表示，如式(3.1)所示。其中 R 代表該道試題答對人數， N 代表該道試題所有作答總人數。

$$P = \frac{R}{N}, 0 \leq P \leq 1 \quad (3.1)$$

另一方面，採用高分群與低分群受試者的計算方式是將受測者的測驗總分，依照高至低的次序排列，最後分別各取一定比例的受測者為高分群與低分群。而難度計算的方式，就如式(3.2)所示，是高分群與低分群答對百分比(分別為 R_H 和 R_L) 之和除以兩群受測者人數(分別為 N_H 和 N_L) 之和。若高分組與低分組受試者人數相等時，則有更簡便的計算方式，如式(3.3)所示。由於鑑別度分析也需要高低分群答對比例 P_H 、 P_L 的數據，因此採用高分組與低分組受試者的計算方式，成為現今主流的難度計算公式。

$$P = \frac{R_H + R_L}{N_H + N_L}, 0 \leq P \leq 1 \quad (3.2)$$

$$P = \frac{P_H + P_L}{2}, P_H = \frac{R_H}{N_H}, P_L = \frac{R_L}{N_L}, 0 \leq P \leq 1 \quad (3.3)$$



上述難度分析的計算結果，可用來判斷試題難易的程度，其值在 0.4 至 0.8 之間，屬於難易度適中的範圍，作為挑選試題之標準最適當。小於 0.4 為偏難，大於 0.8 則偏易，越接近 0.5，其難易度越適中[14]。

另一方面，試題的鑑別度則是觀測高分群答對比例與低分群答對比例的差異情形，計算方式可表示為式(3.4)，其中 P_H 代表高分群答對比例， P_L 代表低分群答對比例。所謂鑑別度好的試題，應當會讓班上高分群答對比例高於班上低分群答對比例，這樣才合理。其計算方式如式(3.4)：

$$D = P_H - P_L \quad (3.4)$$

式(3.4)中， P_H 代表高分群答對比例， P_L 代表低分群答對比例。上述鑑別度分析的計算結果，可用來判斷試題鑑別度的好壞，鑑別度大於等於 0.4 代表非常優良，鑑別度介於 0.4 至 0.3 代表優良，鑑別度介於 0.3 至 0.1 代表尚可，但可能需要修改，而鑑別度介於 0.1 到 0 之間代表鑑別度不佳，需要淘汰。若為負值，則表示答案可能有誤或者不明確[14]。

從以上的介紹，可看出不論試題難度或鑑別度分析，都需要找出高分群與低分群比例的參數。試題分析中所選的高、低分群越極端，則表示鑑別度越大。但是太極端的話（例如高分群與低分群在 20% 以內），則會因選取的受試者太少，影響分析結果的可靠性。決定高低分群常見的比例有三種，分別是前後各佔 25%、27% 或 33%。在常態分配下，根據學者 Truman Kelley 的看法[23]，最適當的比率

是高低分群各佔 27%，這也成為現今分群主要的參考基準。也有學者認為在一般的測驗與評量中，只要介於 25% 至 33% 都屬於合理範圍。但一般最常見的標準化測驗的話，習慣上仍採用 27% 作為高、低組別分群的標準[15]。在高低分群之外的受測者，則歸為中分群。

3.2.2 選題策略演算法：線性規劃單體法

線性規劃最早在 1939 年由 L. V. Kantorovich 所提出[24]，目的在於探討有關組織和規劃問題。線性規劃是用於最佳化（極大化或極小化）的線性目標函數，同時又必須符合一組包含等式及不等式的線性限制條件。建立線性規劃模式的過程，可依循下列三步驟：

1. 依問題的特性定義決策變數 (decision variables)。決策變數是用來表示所欲解決的問題中，決策所需執行的事項，例如：題庫中決定某試題是否選出作為測驗的試題，決策變數即可用來表示這些試題。
2. 分析問題找出限制條件，利用決策變數建立成數學等式或不等式之限制式 (constraints)。
3. 依據問題所欲達到的目標準則，利用決策變數建立目標函數 (the objective function)，並將其極大化或極小化。

線性規劃模式的架構包含目標函數、結構限制式和非負限制式，其中結構限制式與非負限制式又可統稱為「條件不等式」。

1. 目標函數：
$$\text{Max } Z = \sum_{j=1}^n c_j x_j \quad (\text{或 } \text{Min } Z = \sum_{j=1}^n c_j x_j)。$$

2. 結構限制式：Subject to
$$\sum_{i=1}^m \sum_{j=1}^n a_{ij} x_j \leq b_i \quad (\text{不等式可由 } \leq \text{ 或者 } \geq \text{ 組成})。$$

3. 非負限制式： $x_j \geq 0$ ， $j = 1, 2, 3, \dots, n$ 。

線性規劃常見的解法可分為圖解法 (graphical method)、代數法 (algebra method) 與單體法三種，其優缺點可條列如表 3-2 所示[16]。本研究由於應用於選題，決策變數的數目大於兩個，因此採用單體法作為線性規劃問題的解法。



表 3-2 線性規劃常見解法比較表

	圖解法	代數法	單體法
優點	當線性規劃問題只有兩個決策變數，可以節省求解的時間。	當決策變數多於二個，無法畫出可行解區域時，可以利用代數法。	當決策變數及限制式很多時(大型線性規劃問題)，利用本法可以快速求出所需要的解，是一種非常有效率的方法。
缺點	當決策變數為二個以上，圖解法就不適用。	當決策變數及限制式很多時，代數法的計算方式可能很龐大，基本可行解的個數可能會多到無法處理的地步。	雖然在數百限制式以下的問題，本法是最有效率的演算法。

進行單體法之前，首先要將線性規劃標準式中，限制式的部份由不等式改為等式。式(3.5)是一個常見的線性規劃標準式：



$$\begin{aligned}
 &MaxZ = C_1X_1 + C_2X_2 + \cdots + C_nX_n \\
 &Subject\ to: \\
 &a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n \leq b_1 \\
 &a_{21}X_1 + a_{22}X_2 + \cdots + a_{2n}X_n \leq b_2 \\
 &a_{31}X_1 + a_{32}X_2 + \cdots + a_{3n}X_n \leq b_3 \\
 &\vdots \\
 &\vdots \\
 &a_{m1}X_1 + a_{m2}X_2 + \cdots + a_{mn}X_n \leq b_m \\
 &And\ X_1, X_2, X_3, \dots, X_n \geq 0
 \end{aligned}
 \tag{3.5}$$

在單體法之中，式(3.5)會改寫成以下式(3.6)的形式。

$$\begin{aligned}
 &MaxZ = C_1X_1 + C_2X_2 + \cdots + C_nX_n \\
 &Subject\ to: \\
 &a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n = b_1 \\
 &a_{21}X_1 + a_{22}X_2 + \cdots + a_{2n}X_n + X_{n+1} = b_2 \\
 &a_{31}X_1 + a_{32}X_2 + \cdots + a_{3n}X_n + X_{n+2} = b_3 \\
 &\vdots \\
 &\vdots \\
 &a_{m1}X_1 + a_{m2}X_2 + \cdots + a_{mn}X_n + X_{n+m} = b_m \\
 &And\ X_1, X_2, X_3, \dots, X_n, X_{n+1}, X_{n+2}, \dots, X_{n+m} \geq 0
 \end{aligned}
 \tag{3.6}$$

根據單體法，式(3.6)的線性規劃標準式可以用以下的原則求解：

1. 對於具有 m 個限制式、 $n+m$ 個變數的標準式，藉由讓 n 個變數為 0，然後聯立求解剩餘的 m 個等式及 m 個變數問題，可得到的解稱為基本解 (basic solution)。
2. 若得到的所有變數均滿足非負限制式，則此基本解稱為可行基本解 (basic feasible solution)。

3.2.3 選題策略演算法實驗設計

本節首先介紹評估選題策略演算法實驗數據的標準，接著再闡述實驗步驟與內涵。

常見的試題品質評估標準之一為信度。信度是一種評估測量方法品質的標準，指的是測量結果的穩定性（stability）與一致性（consistency）。以試題信度來說，簡單來說，就是指評量結果一致性的程度。試題信度越高，代表評量結果越一致；反之，則代表評量結果越不一致。

再測信度（test-retest reliability）與複本信度（parallel forms reliability）是試題信度估計常見的兩種方式，再測信度指的是讓同一群受測者在前後不同時間點，接受同一份測驗；複本信度指的是讓同一群受測者，在相近時間點接受程度相同的試題測驗，程度相同包含題型、題數、難度等等都相同[25]。這兩種信度估計方式，通常以相關係數表示。兩次量度的相關係數越高，表示信度越高。

在各種相關係數計算方式之中，最常受到引用的就是 Pearson 相關係數（correlation coefficient），通常以 r 表示。Pearson 相關係數描述的是兩個變量線性相關的強弱程度，計算公式主要是兩個樣本的共變異數（covariance）除以標準差（ s ）的乘積，如式(3.7)，其中 n 代表樣本數， S 代表樣本標準差， \bar{x} 和 \bar{y} 分別代表樣本 x 和 y 的平均數。

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3.7)$$

當 r 的絕對值越大，表示兩個變量線性相關程度越高， $r > 0$ 時表示兩個變量正相關， $r < 0$ 時則表示兩個變量負相關。相關係數值區間與對應的相關程度分級如下：等於 1.0 表示完全相關，介於 0.99 ~ 0.70 表示高度相關，0.69 ~ 0.40 表示中度相關，0.39 ~ 0.10 表示低度相關，0.10 以下表示微弱或無相關[26]。

光用 Pearson 相關係數，還不足以完全代表兩個變量的實際分佈情況。統計學

家 Francis Anscombe 就曾提出一個著名的例子“Anscombe's quartet”（如圖 3-2）[27]，說明離群值對於相關係數的影響。這四組數據的相關係數皆為 0.81，但是很明顯可看出右上圖的數據並非高度相關的線性關係；再者，左下圖的數據因為一個離群值，使得相關係數大幅下降；反之，右下圖的數據則因為一個數值，使得相關係數大幅上升。

因此，本研究衡量選題策略產生試題的信度時，除了引用 Pearson 相關係數之外，也會觀察其他統計數值作為輔助。

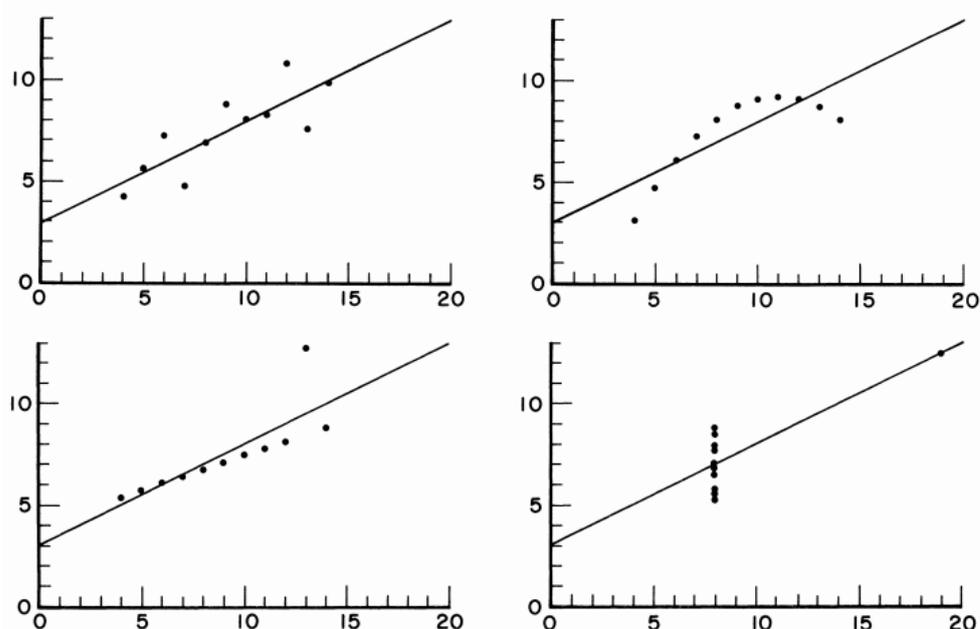


圖 3-2 “Anscombe's Quartet”示意图

本實驗目的在於利用分層教學實驗蒐集的學生作答紀錄數據，模擬再測信度的作法，驗證選題策略演算法的信度。

分別利用前測、後測的學生作答紀錄，以 K-Fold Cross Validation 方式驗證試題信度。所謂的 K-Fold Cross Validation，是建立 K 格的數據集合，以前面 K-1 格的數據集合作為訓練集（training set），第 K 格的資料作為測試集（testing set），應

用在本實驗，可避免取樣偏誤所造成的問題。

實驗流程共有 K 回合，每回合包含訓練與測試兩個階段。整個詳細實驗流程可參考圖 3-3。首先，在訓練階段，步驟為：取不重複的 K-1 格的數據集合作為訓練集，接著計算訓練集的試題難度與鑑別度數據、設定難度區間，然後以選題策略演算法計算出預期難度與鑑別度。接著，在測試階段，步驟為：計算測試集的難度與鑑別度表現，然後紀錄訓練集與測試集的在難度與鑑別度的差距，完成此回合的驗證。接著進入下回合，重複以上訓練與測試階段步驟，直到完成所有 K 回合為止。

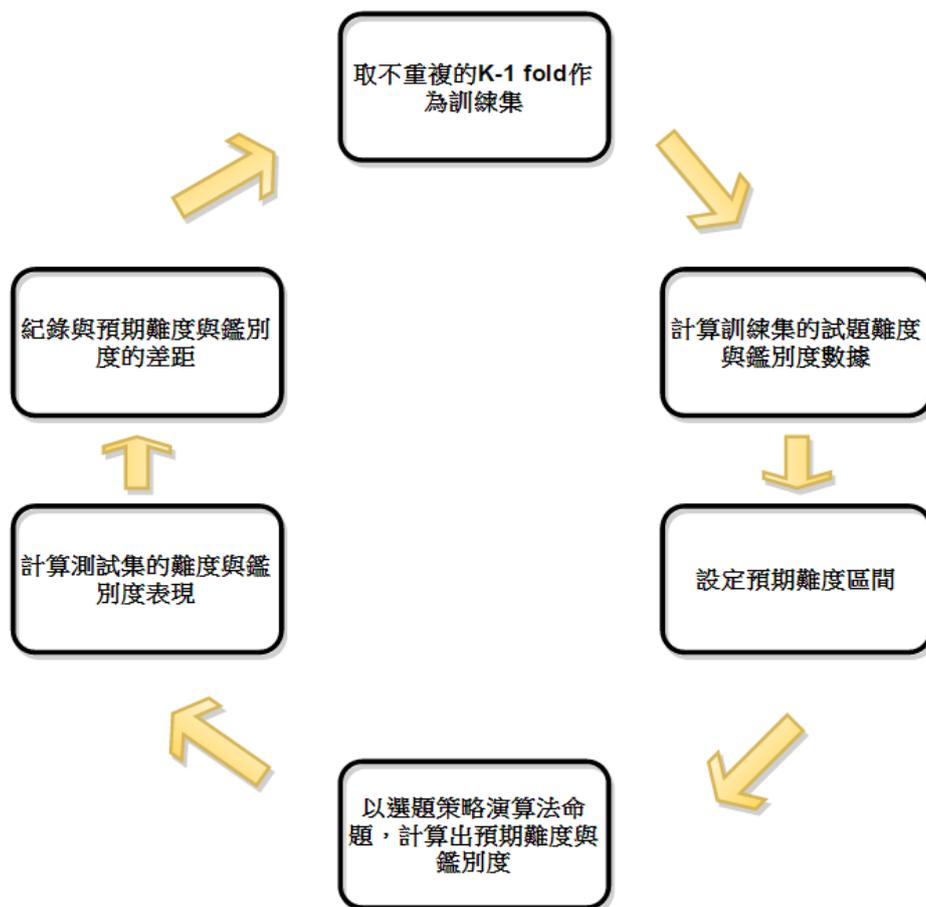


圖 3-3 選題策略演算法實驗設計步驟



3.3 學習成就分析模組實作方式

學習成就分析模組採用一般訪談法 (interview)，參考使用者的意見，建立使用者所關心的學生學習成就追蹤指標。

本研究訪談研究合作對象為烏日國中的 4 位數學任課老師，最後採用教育領域常見的平均數、中位數與標準差作為追蹤指標，並與資料庫中對應的統計量進行比較。圖 3-4 為一個課綱單元的學習成就分析模組輸出格式之範例。

	試題指標	難度(平均答對率)	鑑別度
試題數目			

統計結果	平均數	中位數	內部標準差	與資料庫平均數差距	與資料庫中位數差距	與資料庫內部標準差差距
群組						
高分群						
中分群						
低分群						

高分群名單：

中分群名單：

低分群名單：

圖 3-4 學習成就分析模組輸出格式範例 (一個課綱單元)

3.4 成效衡量模組實作方式與教學實驗設計

本節將介紹成效衡量模組所採用的實作方式以及驗證成效的教學實驗設計。首先介紹二因子變異數分析的內涵，接著介紹教學實驗採用的設計原理，最後闡述本教學實驗進行的流程。



3.4.1 成效衡量模組實作方式：二因子變異數分析

多因子實驗 (factorial experiment) 指的是一個實驗中同時評估兩個或多個因子效果的實驗。因此，二因子變異數分析就是探討一個實驗中兩個自變項的各自效果以及兩者之間的交互作用。若兩因子各自獨立，又可稱為二因子獨立測量變異數分析[26]。

舉例來說，一間學校計畫引進一種新型教學方法，在此之前，管理階層想要比較學生學習能力程度與教學方法對於學習態度的影響，以便了解這個新型教學方法值不值得引進。負責這項實驗任務的人員，首先便可透過以下表 3-3 所敘述之架構圖的方式來設計實驗。

在實驗中，通常會進行三種分析，第一種是不考慮 B 因子時，A 因子是否出現顯著效果；第二種是不考慮 A 因子時，B 因子是否出現顯著效果；第三種分析是觀察 A 因子和 B 因子之間是否有顯著的交互作用 (interaction) [25]。

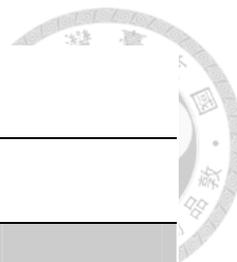
表 3-3 二因子獨立測量實驗設計架構圖

		因子 B 教學方法	
		B ₁ 舊型	B ₂ 新型
因子 A 學生學習能力程度	A ₁ 高分群學生	高分群學生對於舊型教學方法的滿意度	高分群學生對於新型教學方法的滿意度
	A ₂ 低分群學生	低分群學生對於舊型教學方法的滿意度	低分群學生對於新型教學方法的滿意度

二因子變異數分析典型的步驟可歸納為：一、將實驗資料化為表 3-4 的數據紀錄格式；二、逐項計算表 3-5 各欄位數值；三、根據機率欄位判斷結果。其中，

表 3-4 的每個 X 代表一個蒐集到的實驗數據，如果實驗目的是要了解學生學習能力程度與教學方法對於學習態度的影響，那麼 X 就代表學生對某個教學方法的態度數值（如滿意度等）。此外， T_{AB_j} 代表 AB_j 組的數據總和， $\sum X$ 則代表所有組的數據總和。

表 3-4 二因子 ANOVA 數據表



因子 B					
因子 A	B ₁ 組	B ₂ 組		B _b 組	
A₁ 組	X ₁	X		X	
	X ₂	X		X	
	:	:		:	
	X _n	X		X	
	$T_{A_1B_1}$	$T_{A_1B_2}$		$T_{A_1B_b}$	T_{A_1}
A₂ 組	X	X		X	
	X	X		X	
	:	:		:	
	X	X		X	
	$T_{A_2B_1}$	$T_{A_2B_2}$		$T_{A_2B_b}$	T_{A_2}
:	:	:	:	:	
A_a 組	X	X		X	
	X	X		X	
	:	:		:	
	X	X		X	
	$T_{A_aB_1}$	$T_{A_aB_2}$		$T_{A_aB_b}$	T_{A_a}
	T_{B_1}	T_{B_2}		T_{B_b}	$\sum X$

表 3-5 二因子 ANOVA 摘要表

變異來源	自由度	平方和	平均平方數	變異比(F 值)	機率 (p 值)
因子 A	df_A	SS_A	MS_A	F_A	P_A
因子 B	df_B	SS_B	MS_B	F_B	P_B
交互作用 A×B	$df_{A \times B}$	$SS_{A \times B}$	$MS_{A \times B}$	$F_{A \times B}$	$P_{A \times B}$
誤差	df_{Error}	SS_{Error}	MS_{Error}		
總和	df_{Total}	SS_{Total}			

表 3-5 的計算方式從最左欄自由度開始，接著一路由平方和、平均平方數導出變異比，最後透過查表得出機率值。以下逐項介紹各欄的計算方式：

- 一、自由度：各欄位的計算方式如式(3.8)。其中 a 代表 A 因子的組數， b 代表 B 因子的組數， n 代表一個 AB 組中的人數， N 代表總人數。

$$\begin{aligned}
 df_A &= a - 1 \\
 df_B &= b - 1 \\
 df_{A \times B} &= (a - 1)(b - 1) \\
 df_{Error} &= ab(n - 1) \\
 df_{Total} &= N - 1
 \end{aligned} \tag{3.8}$$

- 二、平方和：各欄位的計算方式如式(3.9)。



$$\begin{aligned}
 SS_A &= \frac{\sum T_A^2}{nb} - \frac{(\sum X)^2}{N}, \quad \sum T_A^2 = T_{A1}^2 + T_{A2}^2 + \dots + T_{Aa}^2 \\
 SS_B &= \frac{\sum T_B^2}{na} - \frac{(\sum X)^2}{N}, \quad \sum T_B^2 = T_{B1}^2 + T_{B2}^2 + \dots + T_{Bb}^2 \\
 SS_{A \times B} &= \frac{\sum T_{AB}^2}{n} - \frac{(\sum X)^2}{N} - SS_A - SS_B, \quad \sum T_{AB}^2 = T_{A1B1}^2 + \dots + T_{AaBb}^2 \\
 SS_{Error} &= SS_{Total} - SS_A - SS_B - SS_{A \times B} \\
 SS_{Total} &= \sum X^2 - \frac{(\sum X)^2}{N}
 \end{aligned} \tag{3.9}$$

三、平均平方數：各欄位的計算方式如式(3.10)。

$$\begin{aligned}
 MS_A &= \frac{SS_A}{df_A} \\
 MS_B &= \frac{SS_B}{df_B} \\
 MS_{A \times B} &= \frac{SS_{A \times B}}{df_{A \times B}} \\
 MS_{Error} &= \frac{SS_{Error}}{df_{Error}}
 \end{aligned} \tag{3.10}$$

四、變異比 (F 值)：各欄位的計算方式如式(3.11)。

$$\begin{aligned}
 F_A(df_A, df_{Error}) &= \frac{MS_A}{MS_{Error}} \\
 F_B(df_B, df_{Error}) &= \frac{MS_B}{MS_{Error}} \\
 F_{A \times B}(df_{A \times B}, df_{Error}) &= \frac{MS_{A \times B}}{MS_{Error}}
 \end{aligned} \tag{3.11}$$

求得表 3-5 自由度、平方和、平均平方數、變異比各項欄位的數值之後，最後搭配自由度與變異比，可求得機率欄位的 p 值。藉由 p 值，可分別觀察 A、B 因子的主要效果 (main effect) 與兩者的交互作用。接下來透過表 3-6 這則範例，可以分析這兩個因子的主要效果與交互作用是否達到顯著水準。

表 3-6 二因子 ANOVA 摘要表範例

變異來源	自由度	平方和	平均平方 數	變異比 (F 值)	機率 (p 值)
因子 A	1	1.50	1.50	1.67	0.0715
因子 B	1	48.17	48.17	53.52	0.0038
交互作用 AxB	1	60.16	60.16	66.84	0.0005
誤差	20	18.00	0.90		
總和	23	127.83			

一般來說，p 值以 0.05 作為顯著水準的標準， $p < 0.05$ 代表顯著，而 $p < 0.01$ 則達高度顯著。以表 3-6 此則範例來說，可看出 A 因子未達顯著水準，另一方面，B 因子和 A、B 因子交互作用皆達高度顯著水準。

假設 A 因子對應表 3-3 的學生學習能力程度，B 因子對應教學方法，則表 3-6 的結果表示學生學習能力程度，對於滿意度沒有顯著影響，而教學方法和學生學習能力程度與教學方法的交互作用，則對於滿意度有顯著影響。

3.4.2 分層教學實驗方法：準實驗設計法

準實驗設計 (quasi-experimental design) 是指當研究者無法隨機分配受試者到實驗組或控制組，並控制實驗情境時，此時運用原始群體在自然情況進行實驗的研究方法[28]。

舉例來說，假設現在要研究的議題是國民小學實施新編數學科教材是否優於之前的教材，他與校方接洽後，學校只能提供兩班學生接受實驗，卻不同意將班級學生任意拆散。因應現實情境的限制，此時研究者就必須採用準實驗設計。

準實驗設計最常見的幾種類型，包括平衡對抗設計 (counterbalanced design)，單組時間序列設計 (one-group time series design)，以及不相等前後測設計 (nonequivalent pretest-posttest control group design) [29]。其中，平衡對抗設計指的是所有分組個別輪流接受一種不同的實驗處理，分組數必須與實驗處理數相等，比如三種實驗處理即須有三個分組，順序則隨機決定。單組時間序列設計指的則是對一組實驗樣本進行一系列週期性前測，經實驗處理後再進行一系列週期性後測，觀察實驗結果。而不相等前後測設計，則是教育研究領域相當流行與有用的設計，也是最適用於本研究問題的實驗設計方法，以下針對此方法詳述其內容。

不相等前後測設計亦稱為未隨機化控制組前後測設計 (nonrandomized control group, pretest-posttest design)，這個實驗設計與真實驗 (true experimental design) 的前後加控制組設計 (pretest-posttest control group design) 很相似，差別只在於沒有使用隨機分派實驗組與控制組。研究者採用完整的、已形成的受試者群體，先施予前測，之後為這兩組安排不同處理條件，接著實驗後測，比較其結果[30]。施行方式可用式(3.12)表示。

$$\begin{aligned} O_1 \rightarrow X \rightarrow O_2 & \quad O_1, O_3: \text{pretest score, X:treatment X} \\ O_3 \rightarrow C \rightarrow O_4 & \quad O_2, O_4: \text{posttest score, C:treatment C} \end{aligned} \quad (3.12)$$

整體來說，相較於真實驗研究法則較能掌握內在效度的因素，準實驗研究法則能掌握外在效度的因素。也就是說，就實驗效度來說，不相等前後測設計因與現實環境條件相近，因而有較好的外在效度；然而在內在效度層面，因未採取完全隨機方式分派受試者，所以可能會干擾到實驗的效果[28]。



3.4.3 分層教學實驗流程設計

採用準實驗設計法不相等前後測設計，分為採取分層教學實驗組與實施一般教學的控制組。實驗操作流程主要可分為前測、輔導與後測三大階段。

- 一、前測階段：首先利用資料庫的歷史作答紀錄，挑選難度與鑑別度一致的前測與後測試題。接著對實驗組與控制組學生施行前測。
- 二、輔導階段：實驗組部份，首先會收到學習成就分析模組的結果。輔導方式是會將各班學生依照前測結果分為高中低三群（分群方式同 3.2.1 試題難度與鑑別度分析所闡述的分群理論根據，採前 27% 為高分群，後 27% 為低分群，其餘為中分群），請任課教師施行針對該班各群學生的分層教學。控制組部份則不會收到來自研究者的任何數據，維持一般複習方式。
- 三、後測階段：實驗組與控制組都進行後測，實驗組在後測之後加考加試題，加試題乃利用前測實驗組受測學生作答紀錄，利用本研究選題策略演算法，訂出與前後測一致難度與鑑別度的試題。

後測與加試題結束之後，實驗分析可分為兩個方向：一、本機制利用前測與後測數據，結合二因子變異數分析的方式，探討實驗組與控制組在不同課綱單元、不同學習能力分群的表現，供參與分層教學的教師了解教學成效。二、本機制利用加試題數據，分析不同作答紀錄數據來源的命題方式，在高中低分群的表現是否有所差異。整體實驗流程如圖 3-5 所示。

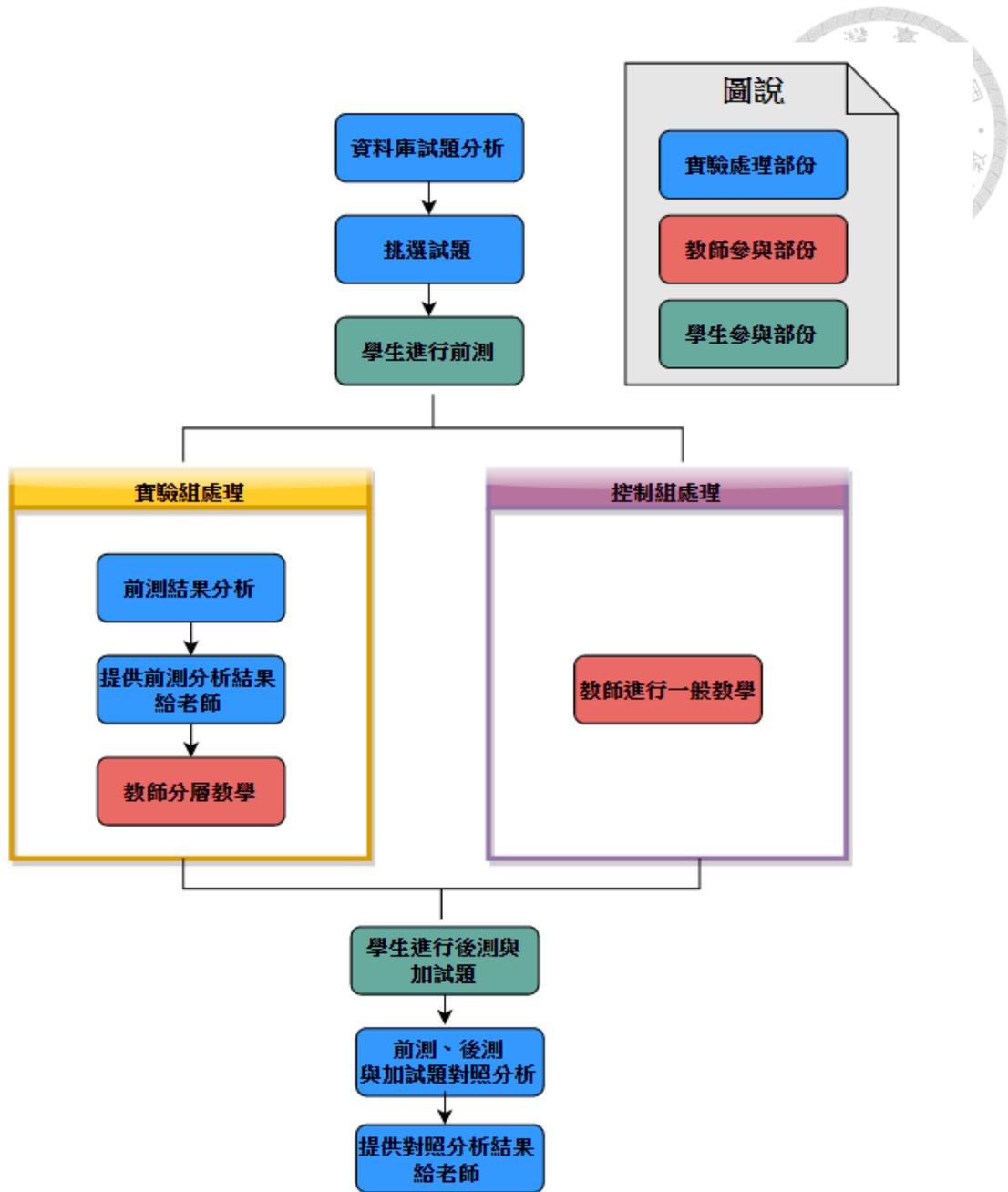


圖 3-5 分層教學實驗流程圖

Chapter 4 實驗與結果討論



本章分為三個部份。第一部份闡述分層教學的實驗操作方式，介紹本研究採用的線上測驗試題來源及分層教學實驗操作之方式。第二部份闡述分層教學實驗分析之結果，針對分層教學在課綱單元、各學習能力分群有何教學成效之差異。第三部份闡述選題策略實驗的操作方式與結果分析。

4.1 分層教學實驗操作

4.1.1 線上測驗試題來源

本實驗的線上測驗試題來源取自均一教育平台，於 2012 年上線啟用，是台灣唯一 K12 使用的大規模網路免費公開課程(MOOC)網站。截至 2015 年 3 月，已經累積 4300 支教學短片，超過 700 組互動式試題，17 萬個註冊者，每週 4 萬 5 千個上線使用者。

均一教育平台的使用者來源，可以從兩個方向來看。首先，各縣市使用者占均一教育平台整體使用者的比例，可以從圖 4-1 的長條圖看出，台北市占 30%、新北市 15%，亦即所謂的「雙北」或「大台北地區」就占了 45%的比例。此外，台北市、新北市、台中市、台南市、高雄市就占了 67%，顯示北部地區、都會地區的使用者仍然是主要來源。

不過，另一方面，以均一教育平台使用者與該縣市 K12 學生的比例來說，從圖 4-1 左側的統計數據可看出，台東縣最為突出，達 10.67%，是排名第二台北市的兩倍以上，第三則是宜蘭縣 4.01%，這顯示均一教育平台在某些非都會區的目標對象觸達率反而較都會區來得高。

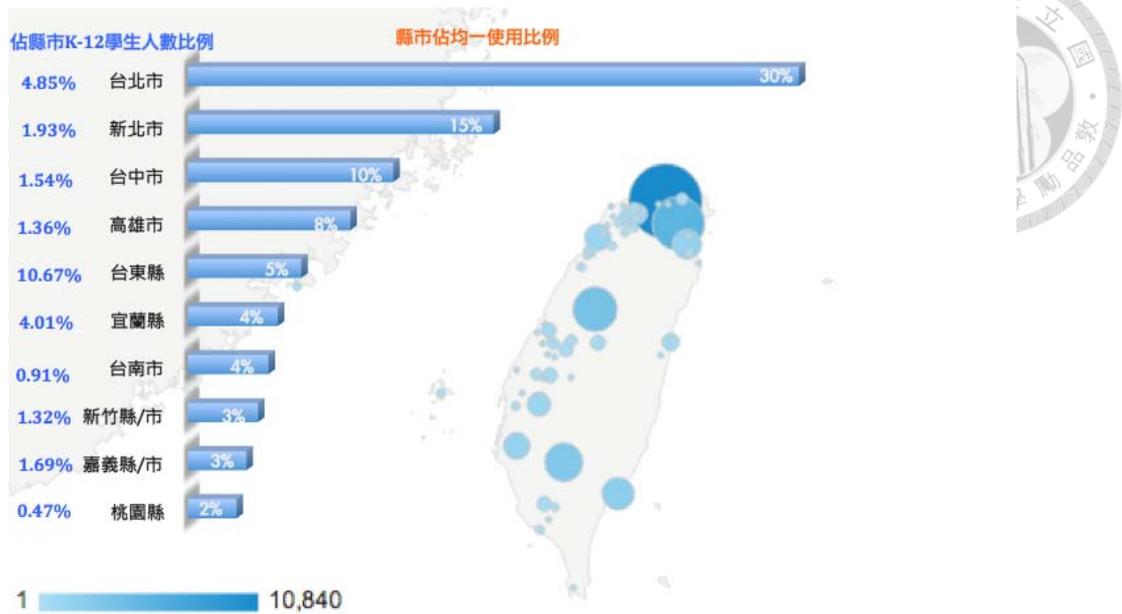


圖 4-1 均一教育平台使用者來源比例

4.1.2 前後測實驗操作

本實驗對象為烏日國中三年級共 4 個班級的學生，設定 1 個班級為控制組，另外 3 個班級為實驗組。共回收 94 個有效樣本，控制組為 23 個樣本，實驗組為 71 個樣本。前測施行時間為：2014 年 11 月 17 日至 11 月 21 日，前測施行一週後將學習成就分析模組的結果提供給任課老師，如附錄一所示（以其中 A 班為例），從 12 月開始進入為期兩週的輔導階段。後測施行時間為：2014 年 12 月 15 日至 12 月 19 日。

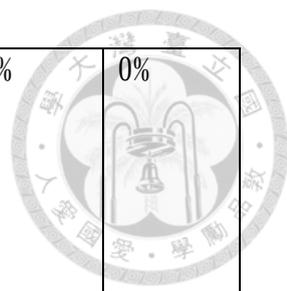
在前測與後測的出題工作方面，經由與實驗合作國中教師們的討論，設定為平時學生習慣的題目數目與作答時間，題數是 30 題，作答時間為 70 分鐘。針對授課進度所對應的均一教育平台題組，計算均一資料庫的作答紀錄求出各題組的難度與鑑別度之後，選擇難度介於難易適中範圍、鑑別度達優良門檻值以上的題組，最後前測與後測整份試題難度皆為：0.73，鑑別度為：0.49。課綱單元、題組與出題比例的分佈情況可參考表 4-1。

此外，在加試題方面，因校務進度限制，使得題數最多只能達 10 題。在此，利用了本研究的選題策略演算法，以烏日國中實驗組前測為數據來源，產生一份難度為 0.73，鑑別度為 0.49 的試題，對應均一資料庫的難度為 0.72，鑑別度為 0.55。課綱單元、題組與出題比例的分佈情況同樣可參考表 4-1。

其中，輔導階段的兩週內，實驗組與控制組皆花費 8 小時針對前測題目複習，細節可參考表 4-2。

表 4-1 前後測試題來源與數據對照表

課綱單元 名稱	均一試題對應題組名稱	難度	鑑別度	前後測 出題比 例	加試出 題比例
1-1 相似形 與比例線 段	平面圖形放大、縮小對 長度與角度與面積的影 響	0.75	0.50	6.67%	30%
	放大圖與縮小圖的面積 關係 2	0.67	0.44	6.67%	0%
	相似形的基本概念	0.69	0.57	6.67%	0%
1-2 相似三 角形	三角形相似性質	0.71	0.58	3.33%	10%
	相似三角形 1	0.64	0.61	3.33%	0%
	相似三角形 2	0.70	0.56	6.67%	20%
	解出相似三角形	0.80	0.39	3.33%	0%
	解出相似三角形 2	0.69	0.62	3.33%	0%



1-3 相似三角形的應用	運用三角形相似性質來計算	0.82	0.82	20%	0%
2-1 點、直線與圓之間的位置關係	圓與弧	0.67	0.65	10%	10%
	點與圓的位置關係	0.67	0.65	10%	10%
2-2 圓心角、圓周角及弦切角	弦切角	0.85	0.31	3.33%	0%
	扇形與圓心角 1	0.72	0.56	6.67%	0%
	扇形與圓心角 2	0.79	0.43	3.33%	0%
	圓心角	0.89	0.24	3.33%	0%
	圓周角與圓心角換算 1	0.75	0.50	3.33%	20%

表 4-2 實驗組與控制組教學輔導方式

班級	實驗組 A 班	實驗組 B 班	實驗組 C 班	控制組 D 班
教師	A	B	C	D
時數	8	8	8	8
方式	同班教學，但教師會在課堂說明與區別複習之題目的程度。各群複習之題目為：高分群取答題正確率最低的 10 題，中低各分群取與答題正確率 50% 絕對值差距最小的 10 題。			維持一般傳統複習方式

4.2 分層教學實驗分析

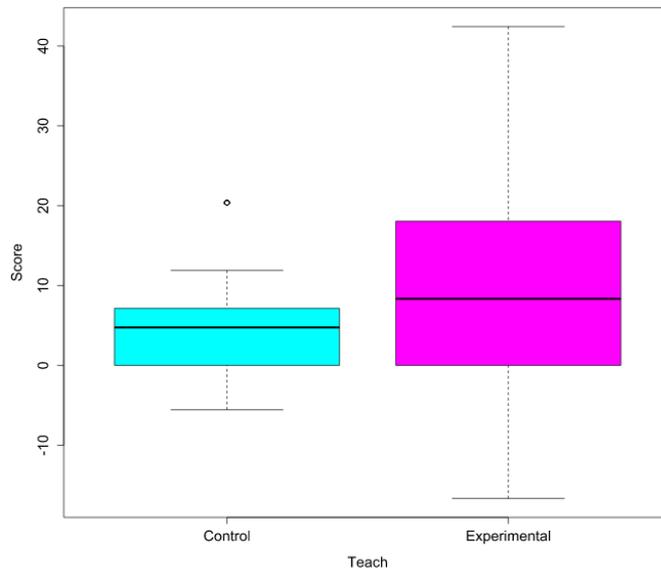


圖 4-2 實驗組與控制組對照箱型圖

圖 4-2 橫軸代表教學方式，縱軸代表換算成百分制之後，後測與前測的分數差距。就整體來說，實驗組的全距比控制組來得大，表示實驗組裡同時存在進步幅度比較大與退步幅度比較大的學生。不過就數據分佈情況來看，上四分位數(Q1)來說，實驗組和控制組的表現不相上下，但是在中位數(Q2)、上四分位數(Q3)，都是實驗組大幅優於控制組。

接下來，在 4.2.1，設定教學方法(下列圖表中以”Teach”表示)與課綱單元(下列圖表中以”Lecture”表示)為因子，以二因子變異數分析分層教學成效與課綱單元之間的關係。另一方面，在 4.2.2，則設定教學方法(下列圖表中以”Teach”表示)與學習能力分群(下列圖表中以”Level”表示)為因子，以二因子變異數分析分層教學成效與學習能力分群之間的關係。

4.2.1 教學方法與課綱單元二因子變異數分析

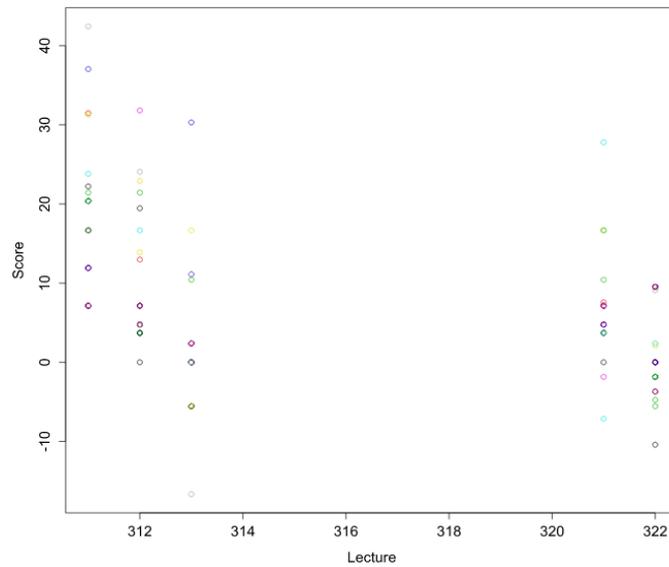


圖 4-3 各課綱單元進步幅度點狀圖

圖 4-3 橫軸代表課綱單元，縱軸代表換算成百分制之後，後測與前測的分數差距。就各課綱單元的難度來說，從圖 4-3 可看出各課綱單元中學生的進步幅度，從中可看出對受測學生來說，大致上課綱單元由易到難的順序大致是： $311 < 312 < 321 < 313 < 322$ 。

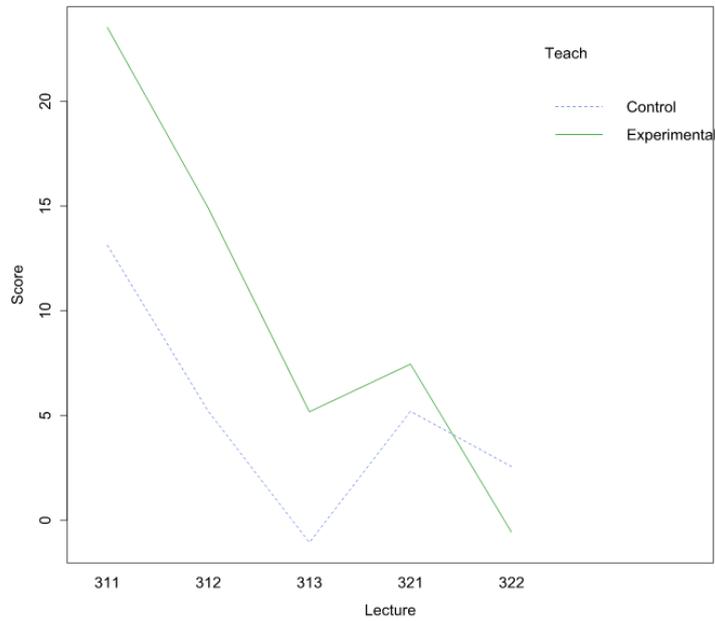


圖 4-4 各課綱單元實驗組與對照組進步幅度折線圖（一）

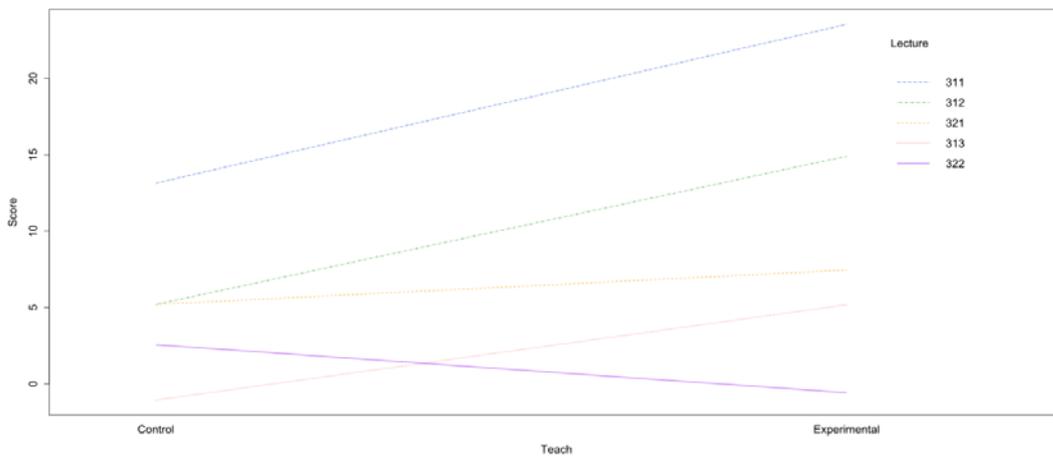


圖 4-5 各課綱單元實驗組與對照組進步幅度折線圖（二）

就實驗組與控制組在各課綱單元的進步幅度來說，從圖 4-4 可看出在 311、312、313 這三個單元，實驗組穩定地大幅優於控制組；在 321 單元兩者差距開始縮小，到了 322 單元出現交叉，控制組進步幅度反而較好。此外，從圖 4-5 可看出，控制組進步幅度由小到大為 $313 < 322 < 321 = 312 < 311$ ，而實驗組進步幅度由小到大為 $322 < 313 < 321 < 312 < 311$ 。由上述數據可歸納，除了明顯較難的課綱單元 322，實驗組表現均優於控制組。

表 4-3 課綱單元與教學方法二因子 ANOVA 摘要表

變異來源	自由度	平方和	平均平方數	變異比(F 值)	機率 (p 值)
課綱單元 (Lecture)	1	1870.4	1870.37	24.4290	2.622e-06
教學方法 (Teach)	1	777.3	777.34	10.1530	0.001851
交互作用	1	678.6	678.60	8.8632	0.003543
誤差	116	8881.3	76.56		

就二因子變異數分析結果，根據表 4-3 中機率欄位，課綱單元、教學方法與交互作用都達到高度顯著（小於 0.01 門檻值）的水準。證實施行分層教學，確實使得學生進步幅度有高度顯著差異。

4.2.2 教學方法與學習能力分群二因子變異數分析

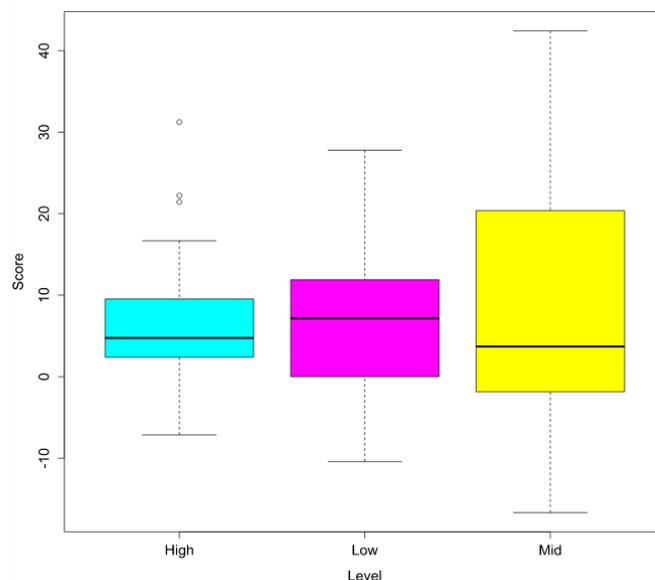


圖 4-6 高中低分群進步幅度箱型圖

圖 4-6 橫軸代表學習能力分群，縱軸代表換算成百分制之後，後測與前測的分數差距。就各學習能力分群的進步幅度來說，可從圖 4-6 看出，中分群全距最大、高分群下四分位數 (Q1) 最好、低分群中位數 (Q2) 最好、中分群上四分位數 (Q3) 最好。就整體看來，3 個分群的 Q1, Q2 各有高低，但差距都在 5% 以內。圖 4-6 最明顯的現象是中分群的 Q3 遠高於高、低分群 10% 以上，可以說中分群前 50% 是進步幅度最大的群體。

表 4-4 學習能力分群與教學方法二因子 ANOVA 摘要表

變異來源	自由度	平方和	平均平方數	變異比(F 值)	機率 (p 值)
學習能力分群 (Level)	2	20.7	10.34	0.1045	0.900854

教學方法 (Teach)	1	777.3	777.34	7.8603	0.005941
交互作用	2	135.5	67.77	0.6853	0.506005
誤差	114	11274.1	98.90		

就二因子變異數分析結果，根據表 4-4 的機率欄位，只有教學方法的因子達到高度顯著的差異，學習能力分群因子跟交互作用則並沒有顯著性差異。這表示學生進步幅度主要取決於教學方法，而非其原本學習能力被劃分在哪一層分群裡。學習能力分群與教學方法這兩個因子交互作用沒有顯著性，表示分層教學的成效並不特別相依於特定的學習能力分群。

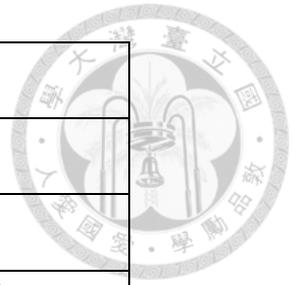
4.2.3 加試題實驗結果分析

為了方便了解受測學生在加試題作答表現的特徵，在此將加試題與同時期施行後測並列比較，計算在各個百分制區間的比例，以了解加試題的表現特徵。

表 4-5 後測與加試題學習成就比較表

百分制區間	後測		加試	
	百分比	累積百分比	百分比	累積百分比
0 ~ 10	3	3	2	2
10 ~ 20	7	10	2	4
20 ~ 30	8	18	0	4
30 ~ 40	8	26	2	6
40 ~ 50	12	38	12	18
50 ~ 60	8	46	13	31

60 ~ 70	10	56	10	41
70 ~ 80	11	67	17	58
80 ~ 90	20	87	20	78
90 ~ 100	13	100	22	100



首先，從表 4-5 可以看出後測與加試題整體受測者作答表現的分佈狀況。最顯著的特徵是在加試題中，低分群的比例大幅減少。相較於在後測當中，40 分以下佔了 26%，在加試題當中 40 分以下僅佔了 6%。雖然在 0~10 分的區間，加試題只比後測少了 1 個百分點，不過在 10~40 分的 3 個區間，加試題都比後測少了 5 個百分點以上。

表 4-6 各百分制區間學生的加試題表現

後測成績百分制區間	加試題表現
0 ~ 10	+ 40%
10 ~ 20	+ 8%
20 ~ 30	+ 27%
30 ~ 40	+ 21%
40 ~ 50	+ 23%
50 ~ 60	+ 23%
60 ~ 70	+ 17%
70 ~ 80	+ 8%
80 ~ 90	+ 6%
90 ~ 100	- 5%

從表 4-6 來看，更可以看到原屬低分群學生的進步幅度。0 ~ 60 分的各百分制區間當中，除了 10~20 分的區間之外，其餘的區間都達到 20% 以上的進步幅度。

從上述兩個表格的結果歸納顯示，在加試題當中，低分群的進步幅度最為顯著。這表示：一、低分群未必是永遠的落後群。適度調整試題難度，低分群便會有所反應；二、也因如此，在計算試題難度與鑑別度等數據時，不可輕易排除低分群，避免造成數據效度不足。

4.3 選題策略實驗操作與分析

選題策略實驗利用教學實驗中收集到的前測與後測的受測者作答紀錄，作為實驗數據來源，以 K-Fold Cross Validation 方式驗證選題策略信度結果。

K-Fold Cross Validation 常見的 K 值為 10，不過以總有效受測學生數目 94 來說，K 值設定為 10 時，將使得訓練組與測試組的比例為 84 : 10 (或者 85 : 9)，是規模相當小的測試組。因此，本研究增加另一種 K 值為 5 的實驗設計，避免過於單一依賴小規模測試組所可能造成的判斷偏誤狀況。設定為 5 的原因是此時訓練組與測試組的比例為 75 : 19 (或者 76 : 18)，測試組的規模恰好等於本研究教學實驗合作國中的單一班級規模，貼近於實際應用情境。

經過 4 位教學實驗合作教師的建議，本選題策略演算法實驗設定目標函數為鑑別度，而結構限制式則包含題目數目、題目難度範圍與各課綱單元選題最小比例，可表示為式(4.1)。其中， m 為選題範圍內的題庫總數目， X_1, X_2, \dots, X_m 代表題目在題庫中的編號， C_1, C_2, \dots, C_m 代表各題的鑑別度數值， a_1, a_2, \dots, a_m 代表各題的難度數值， b_{upper} 代表難度上界， b_{low} 代表難度下界， $ratio_{min}$ 代表最低選題比例， i_1, i_2, \dots, i_5 代表第 1, 2, ..., 5 個課綱單元， n 代表該課綱單元在題庫裡的總數目。在



參數設定方面，前測、後測一致設定總題數為 30， $ratio_{\min}$ 則設定為 10%。用以設定難度區間的參數 b_{upper} 和 b_{low} ，根據學理上「難易適中」設定為 0.4~0.85 的區間，每組之間相距 0.05（如表 4-7 所示），共進行 10 組不同難度的實驗。

$$\begin{aligned}
 &Max\ Z = C_1X_1 + C_2X_2 + \dots + C_mX_m \\
 &subject\ to: \\
 &X_1 + X_2 + \dots + X_m = 30 \\
 &a_1X_1 + a_2X_2 + \dots + a_mX_m \leq b_{upper} \\
 &a_1X_1 + a_2X_2 + \dots + a_mX_m \geq b_{low} \\
 &X_{i11} + X_{i12} + \dots + X_{i1n} \geq 30 \times ratio_{\min} \\
 &X_{i21} + X_{i22} + \dots + X_{i2n} \geq 30 \times ratio_{\min} \\
 &X_{i31} + X_{i32} + \dots + X_{i3n} \geq 30 \times ratio_{\min} \\
 &X_{i41} + X_{i42} + \dots + X_{i4n} \geq 30 \times ratio_{\min} \\
 &X_{i51} + X_{i52} + \dots + X_{i5n} \geq 30 \times ratio_{\min}
 \end{aligned}
 \tag{4.1}$$

表 4-7 實驗難度限制條件設定表

實驗編號	設定難度下界	設定難度上界
1	0.40	0.50
2	0.45	0.55
3	0.50	0.60
4	0.55	0.65
5	0.60	0.70
6	0.65	0.75
7	0.70	0.80
8	0.75	0.85
9	0.80	0.90
10	0.85	0.95

接下來，分別討論各項實驗之結果。



4.3.1 K=10 前測數據來源實驗結果

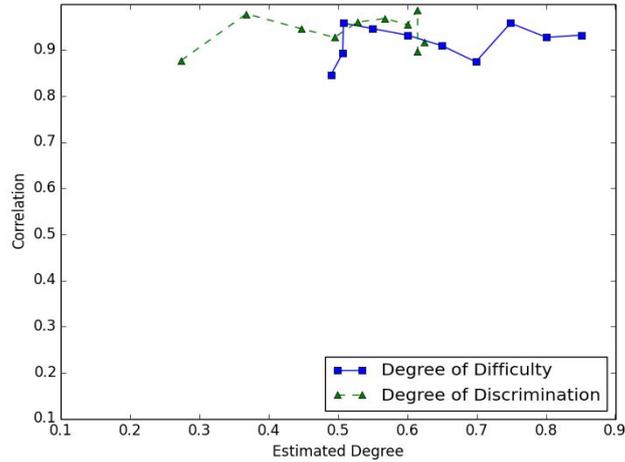


圖 4-7 K=10 前測數據來源的難度鑑別度相關係數折線圖

表 4-8 K=10 前測數據來源的各項實驗數據結果

實驗編號	難度結果	鑑別度結果	難度 R 值	鑑別度 R 值	難度差距中位數	鑑別度差距中位數
PreK10-1	0.490	0.615	0.846	0.986	0.093	0.138
PreK10-2	0.507	0.624	0.893	0.916	0.130	0.155
PreK10-3	0.508	0.615	0.958	0.896	0.150	0.232
PreK10-4	0.550	0.600	0.946	0.955	0.096	0.148
PreK10-5	0.600	0.567	0.932	0.968	0.142	0.181
PreK10-6	0.650	0.528	0.909	0.960	0.160	0.242
PreK10-7	0.699	0.495	0.874	0.927	0.127	0.159
PreK10-8	0.749	0.447	0.958	0.946	0.144	0.197

PreK10-9	0.801	0.368	0.927	0.978	0.151	0.178
PreK10-10	0.851	0.274	0.932	0.877	0.086	0.136

就選題品質來說，在難度結果方面，根據表 4-8 的難度結果欄位，數值分佈平均，顯示本選題策略可以成功產生各種難度的試題組合。在接近 0.5 的選題難度結果尤其理想，最明顯的例子例如：PreK10-1 設定的預期難度區間為 0.40 ~ 0.50，最後選題結果的難度為 0.490；PreK10-2 設定的預期難度區間為 0.45 ~ 0.55，最後選題結果的難度為 0.507；PreK10-3 設定的預期難度區間為 0.50 ~ 0.60，最後選題結果的難度為 0.508。在鑑別度結果方面，在難度結果 0.490 到 0.749 區間，都可達到 0.4 以上的「極佳鑑別度」門檻。預期鑑別度與預期難度的走向成正比，最佳的鑑別度出現在難度結果為 0.507 的情況。

就選題信度來說，從圖 4-7 的落點來看，難度與鑑別度 R 值均穩定超過 0.8 的「高度相關」門檻，兩者的 R 值表現也差不多。以表 4-8 的差距中位數欄位來看，難度的信度又比鑑別度略好，因為 10 組實驗當中，共有 8 組的差距中位數在 0.15 以下，相對地，鑑別度則只有 3 組。

4.3.2 K=10 後測數據來源實驗結果

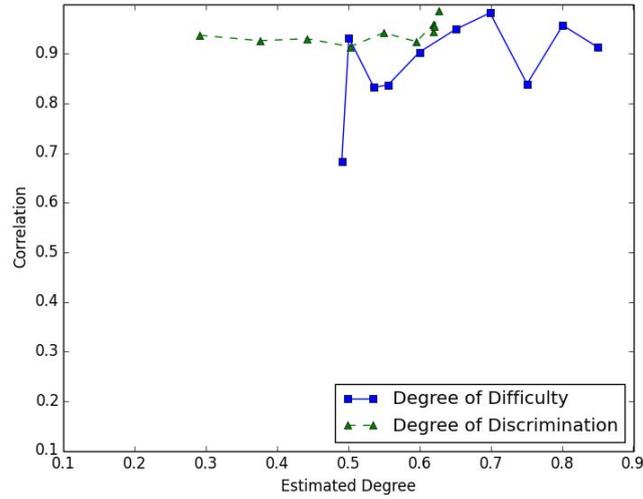


圖 4-8 K=10 後測數據來源的難度鑑別度相關係數折線圖

表 4-9 K=10 後測數據來源的各項實驗數據結果

實驗編號	預期難度	預期鑑別度	難度 R 值	鑑別度 R 值	難度差距中位數	鑑別度差距中位數
PostK10-1	0.491	0.621	0.683	0.955	0.080	0.159
PostK10-2	0.501	0.619	0.932	0.959	0.064	0.170
PostK10-3	0.536	0.627	0.832	0.987	0.100	0.246
PostK10-4	0.556	0.619	0.837	0.944	0.152	0.219
PostK10-5	0.600	0.596	0.903	0.924	0.124	0.167
PostK10-6	0.651	0.550	0.949	0.942	0.143	0.193
PostK10-7	0.699	0.503	0.982	0.914	0.141	0.202
PostK10-8	0.751	0.442	0.839	0.930	0.099	0.112

PostK10-9	0.801	0.376	0.957	0.926	0.096	0.192
PostK10-10	0.850	0.292	0.913	0.937	0.103	0.162

就選題品質來說，在難度結果方面，根據表 4-9 的難度結果欄位，數值分佈平均。但在接近 0.5 難度區間，選題難度結果就沒有前測那麼優異。在鑑別度結果方面，在難度結果 0.490 到 0.751 區間，都可達到 0.4 以上的「極佳鑑別度」門檻，最佳的鑑別度出現在難度結果為 0.536 的情況。

就選題信度來說，在難度結果方面，相較於前測結果，可以從圖 4-8 的折線走勢看出難度 R 值出現較大的波動，尤其是 PostK10-1 這組數據出現 0.683 的情況，0.850 以下的 R 值也比較多。然而，從表 4-9 的差距中位數欄位來看，PostK10-1 卻低達 0.080，甚至比難度 R 值 0.987 的 PostK10-3 還低，表示 PostK10-1 受到極端值很大的影響，信度未必不佳。在鑑別度結果方面，從圖 4-8 的折線走勢可明顯看出 R 值表現非常平穩，均在 0.9 以上，比難度 R 值表現好得多。不過，以表 4-9 的差距中位數來看，大多數來說難度數值還是比鑑別度數值來得小，這表示在 K=10、數據來源為後測的條件下，難度實驗數據的全距較大，因此 R 值跟差距中位數表現有所落差。

4.3.3 K=5 前測數據來源實驗結果

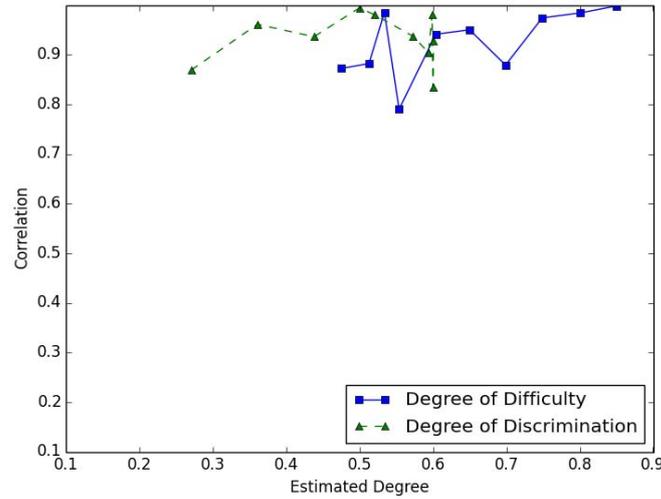


圖 4-9 K=5 前測數據來源的難度鑑別度相關係數折線圖

表 4-10 K=5 前測數據來源的各項實驗數據結果

實驗編號	預期難度	預期鑑別度	難度 R 值	鑑別度 R 值	難度差距中位數	鑑別度差距中位數
PreK5-1	0.475	0.599	0.872	0.981	0.090	0.201
PreK5-2	0.513	0.600	0.882	0.834	0.107	0.203
PreK5-3	0.534	0.601	0.985	0.928	0.088	0.176
PreK5-4	0.554	0.594	0.790	0.904	0.074	0.225
PreK5-5	0.604	0.573	0.941	0.937	0.087	0.196
PreK5-6	0.650	0.521	0.950	0.980	0.081	0.164
PreK5-7	0.699	0.500	0.879	0.994	0.038	0.115
PreK5-8	0.749	0.438	0.974	0.936	0.115	0.169
PreK5-9	0.801	0.361	0.984	0.961	0.119	0.159

PreK5-10	0.850	0.271	0.998	0.869	0.041	0.083
----------	-------	-------	-------	-------	-------	-------

就選題品質來說，在難度結果方面，根據表 4-10 的難度結果欄位，數值分佈平均。在鑑別度結果方面，在難度結果 0.475 到 0.749 區間，都可達到 0.4 以上的「極佳鑑別度」門檻，最佳的鑑別度出現在難度結果為 0.534 的情況。

就選題信度來說，從圖 4-9 折線走勢可看出，難度與鑑別度 R 值表現相近。以差距中位數來看，難度的信度又比鑑別度略好，因為 10 組實驗當中，共有 5 組的差距中位數在 0.10 以下，相對地，鑑別度則沒有任何組達到此標準。

4.3.4 K=5 後測數據來源實驗結果

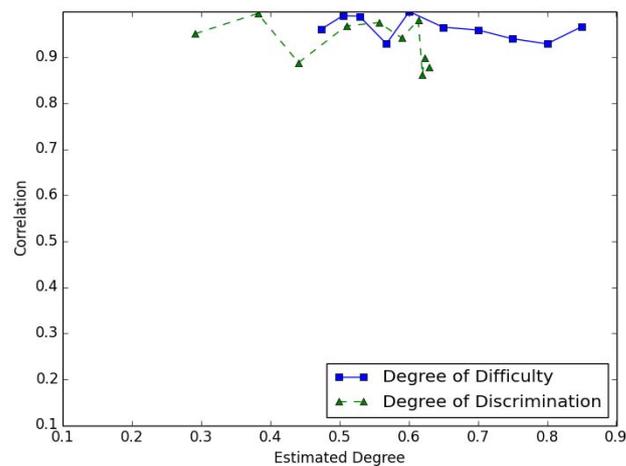


圖 4-10 K=5 後測數據來源的難度鑑別度相關係數折線圖

表 4-11 K=5 後測數據來源的各項實驗數據結果

實驗編號	預期難度	預期鑑別度	難度 R 值	鑑別度 R 值	難度差距中位數	鑑別度差距中位數

PostK5-1	0.474	0.614	0.961	0.980	0.083	0.176
PostK5-2	0.505	0.629	0.990	0.878	0.075	0.298
PostK5-3	0.529	0.623	0.989	0.899	0.045	0.162
PostK5-4	0.568	0.620	0.929	0.861	0.122	0.307
PostK5-5	0.601	0.590	0.999	0.942	0.054	0.108
PostK5-6	0.650	0.557	0.965	0.975	0.102	0.093
PostK5-7	0.700	0.510	0.959	0.968	0.099	0.137
PostK5-8	0.750	0.441	0.940	0.888	0.062	0.093
PostK5-9	0.800	0.382	0.929	0.995	0.058	0.097
PostK5-10	0.850	0.292	0.966	0.951	0.144	0.122

就選題品質來說，在難度結果方面，根據表 4-11 的難度結果欄位，數值分佈平均。在鑑別度結果方面，在難度結果 0.474 到 0.750 區間，都可達到 0.4 以上的「極佳鑑別度」門檻，最佳的鑑別度出現在難度結果為 0.505 的情況。

就選題信度來說，從圖 4-10 折線走勢可看出，難度 R 值特別優異。以表 4-11 的差距中位數欄位來看，難度的信度又比鑑別度好得多，因為 10 組實驗當中，只有 3 組的差距中位數在 0.10 以上，最大差距也只是 0.144，相對地，鑑別度則有 7 組在 0.10 以上，最大差距可達 0.307。

4.3.5 選題策略實驗結果綜合討論

綜合上述數據，首先就選題信度來說，在 4 組實驗共 40 個實驗結果之中，除了 PostK10-1 和 PreK5-4 的難度 R 值略低之外（分別為 0.683 和 0.790），其餘 38 個實驗結果的難度與鑑別度 R 值均達到 0.8「高度相關」門檻值以上，顯示本研究使用的選題策略演算法具備優異的選題信度。

分別觀察難度與鑑別度的選題信度，可以發現相較於鑑別度和難度 R 值表現差不多，但是若以差距中位數做比較，則難度的差距中位數通常比較小。這顯示難度信度方面的實驗數據的變異性比鑑別度來得低。

就 K-Fold Cross Validation 的 K 值來說，除了在 K 值為 10、數據來源為後測的條件之下，難度 R 值出現較大的變異性，其餘 3 組實驗則未反映 K 值在各種數據上變異性的差異。就數據來源是前測或後測來說，則亦無顯著的差異。

Chapter 5 結論與未來展望



本論文提出一套輔助分層教學決策的實作方式，藉由資訊技術追蹤學生學習成就以及提出組合試題方式，為有意施行這種教學方式的教師減少處理資訊的負擔。經教學實驗驗證，結合本機制的國中數學分層教學成效，確實比傳統教學優異。

從教學實驗中發現，分層教學成效與課綱單元難易度有顯著相關的交互作用，特別難的單元，分層教學的成效不彰，然而除此之外，分層教學都有很出色的成效。另一方面，分層教學成效與學習能力分群這兩個因子交互作用沒有顯著性，表示分層教學的成效並不特別明顯侷限在特定的學習能力分群。不過，比較各學習能力分群的結果，相較之下中分群進步幅度最好。而提升低分群進步幅度最好的方式則是調降試題難度。調降試題難度的方式，可利用本研究的選題策略演算法，就可自動從題庫中組合出對應條件的試題，題庫越大，就能擴充更多條件。

烏日國中事後追蹤實驗組之中 65 位同學在歷次校內模擬考與正式會考的數學科表現，從校方所提供的資料（如表 5-1）可看出，參與本研究教學實驗之後的 2 月模擬考，學生進步幅度是歷次測驗之中最大的一次。

表 5-1 烏日國中實驗組學生各次數學科測驗表現

考試時間	11 月 29, 30 日 (模擬考)	2 月 21, 22 日 (模擬考)	4 月 29, 30 日 (模擬考)	5 月 16, 17 日 (會考)
成績	11A 50B 4C	17A 47B 1C	19A 45B 1C	21A 44B
與前次相比	無資料	+6A -3B -3C	+2A -2B +0C	+2A -1B -1C



從本研究所獲得的結論中，可衍伸未來的方向。首先，本研究的實驗合作對象僅包含一個學校四個班級（一班控制組，三班實驗組），雖然已足夠完整模擬整個分層教學的流程，但是在發掘與判斷各種學生學習成就現象的層面，代表性便有限。若資源許可，能夠得到更廣大、更多樣的實驗合作對象，則有更多比較與歸納學生學習成就現象的空間，可從中觀察出更多與更具代表的影響因子與有效作法。

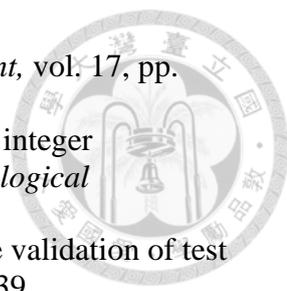
第二，未來可增加實驗次數，拉長觀察學生學習成就的時間。本研究經由不相等前後測實驗設計與二因子變異數分析結果，已經證實在剛授課完後到第一次段考的時間區間，分層教學的成效與各學習能力分群的進步情況，然而就整個國中教育制度來說，這僅是一小段時間區間的取樣。因此，時間區間拉得越長，就越貼近教育應用的真實情境。此外，增加實驗次數可更深入追蹤各學習能力分群的學習成就起伏現象，可從中觀察學生在各分群之間的升降情形。

第三、未來若能取得更多受測者的特徵，結合既有的學習能力分群，建立更細緻學習者檔案（learner profile）的分類，則可對於學習成就的各種特殊現象提出更多解釋。例如為什麼同樣教學方式與同樣學習能力分群，某些學生呈現特別好或者特別差的進步幅度，像這樣的問題就需要更多受測者特徵才足以再深入探討。

參考文獻



- [1] C. A. Tomlinson, "Grading and differentiation: paradox or good practice?," *Theory into practice*, vol. 44, pp. 262-269, 2005.
- [2] T. Hall, "Differentiated instruction," *Wakefield, MA: National Center on*, 2002.
- [3] D. Lawrence-Brown, "Differentiated instruction: Inclusive strategies for standards-based learning that benefit the whole class," *American secondary education*, pp. 34-62, 2004.
- [4] C. Tieso, "The effects of grouping practices and curricular adjustments on achievement," *Journal for the Education of the Gifted*, vol. 29, pp. 60-89, 2005.
- [5] T. Baumgartner, M. B. Lipowski, and C. Rush, "Increasing Reading Achievement of Primary and Middle School Students through Differentiated Instruction," 2003.
- [6] R. Strong, E. Thomas, M. Perini, and H. Silver, "Creating a differentiated mathematics classroom," *Educational Leadership*, vol. 61, pp. 73-78, 2004.
- [7] 韋明理, "分層教學," vol. 1, ed: 中國科教創新導刊, 2007, pp. 56-56.
- [8] S. K. Stuart and C. Rinaldi, "A collaborative planning framework for teachers implementing tiered instruction," *Teaching Exceptional Children*, vol. 42, pp. 52-57, 2009.
- [9] 齊曉峰, "分層教學的實施," *科技創新導報*, pp. 230-230, 2009.
- [10] Y. HU and X.-r. WANG, "Problems in the Stratified Teaching Model for College English and the Countermeasures," *Journal of Anhui Electrical Engineering Professional Technique College*, vol. 3, p. 029, 2011.
- [11] 潘靜, "高中英語教學中分層教學模式應用分析," *學園: 教育科研*, pp. 68-68, 2012.
- [12] Z. Yu, "An Empirical Study of College English Implicitly-graded Teaching," *Journal of Tianjin Foreign Studies University*, vol. 1, p. 008, 2012.
- [13] 李達勝, "分層遞進教學模式在桌球課程的應用與實施," *中華體育季刊*, vol. 25, pp. 564-573, 2011.
- [14] 涂金堂, "教育測驗與評量," 台北: 三民, 2009.
- [15] 陳英豪 and 吳裕益, "測驗與評量 (六版)," ed: 高雄市: 復文圖書出版社, 2003.
- [16] W. J. van der Linden and C. A. Glas, *Computerized adaptive testing: Theory and practice*: Springer, 2000.
- [17] R. D. Armstrong, D. H. Jones, and C. S. Kuncce, "IRT test assembly using network-flow programming," *Applied Psychological Measurement*, vol. 22, pp. 237-247, 1998.
- [18] K.-T. Sun, Y.-J. Chen, S.-Y. Tsai, and C.-F. Cheng, "Creating IRT-based parallel test forms using the genetic algorithm method," *Applied measurement in education*, vol. 21, pp. 141-161, 2008.
- [19] G.-J. Hwang, B. M. Lin, H.-H. Tseng, and T.-L. Lin, "On the development of a computer-assisted testing system with genetic test sheet-generating approach," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 35, pp. 590-594, 2005.
- [20] G.-J. Hwang, "A test-sheet-generating algorithm for multiple assessment requirements," *Education, IEEE Transactions on*, vol. 46, pp. 329-337, 2003.
- [21] L. Swanson and M. L. Stocking, "A model and heuristic for solving very large

- 
- item selection problems," *Applied Psychological Measurement*, vol. 17, pp. 151-166, 1993.
- [22] W. J. Van Der Linden, B. P. Veldkamp, and L. M. Reese, "An integer programming approach to item bank design," *Applied Psychological Measurement*, vol. 24, pp. 139-150, 2000.
- [23] T. L. Kelley, "The selection of upper and lower groups for the validation of test items," *Journal of Educational Psychology*, vol. 30, p. 17, 1939.
- [24] F. B. Baker, A. S. Cohen, and B. R. Barmish, "Item characteristics of tests constructed by linear programming," *Applied Psychological Measurement*, vol. 12, pp. 189-199, 1988.
- [25] R. Pagano, *Understanding statistics in the behavioral sciences*: Cengage Learning, 2006.
- [26] P. R. Hinton, *Statistics explained*: Routledge, 2014.
- [27] F. J. Anscombe, "Graphs in statistical analysis," *The American Statistician*, vol. 27, pp. 17-21, 1973.
- [28] 周新富, "教育研究法," ed: 五南圖書出版股份有限公司, 2007.
- [29] M. D. Miller, R. L. Linn, and N. E. Gronlund, *Measurement and assessment in teaching*: Merrill/Pearson Upper Saddle River, NJ, 2009.
- [30] 文科 and 王智弘, *教育研究法*: 五南圖書出版股份有限公司, 2007.

附錄一

【A 班前測】



(註) 為保護隱私，班級與每位學生均以英文字母表示。

■ 課綱單元名稱：1-1 相似形與比例線段

試題數目 \ 試題指標	難度(平均答對率)	鑑別度
5	難易適中(約 70%)	極優(約 0.5)

統計結果 \ 群組	平均數, 中位數, 內部標準差	與資料庫平均數差距, 中位數差距, 內部標準差差距
高分群	76.19, 83.33, 15.06	-19.19, -16.67, 8.21
中分群	43.33, 50.0, 8.16	8.21, -25.4, -19.63
低分群	11.9, 0.0, 14.68	-33.13, -46.88, 4.44

高分群名單：C, K, L, O, A, W, B, D,

中分群名單：G, M, S, T, U

低分群名單：F, H, P, R, V, J, E, H, N, Q

■ 課綱單元名稱：1-2 相似三角形

試題數目 \ 試題指標	難度(平均答對率)	鑑別度
5	難易適中(約 71%)	極優(約 0.54)

統計結果 群組	平均數, 中位數, 內部標準差	與均一平均數差距, 中位數差距, 內部標準差差距
高分群	88.1, 100.0, 17.17	-10.56, 0.0, 15.19
中分群	43.33, 50.0, 11.06	-35.92, -28.95, 1.22
低分群	0.0, 0.0, 0.0	-43.97, -48.02, -13.97

高分群名單：A, B, M, W, K, S, H, L, O, R, T, U, V

中分群名單：D, F, C

低分群名單：E, G, H, J, N, P, Q

■ 課綱單元名稱：1-3 相似三角形的應用

試題數目	試題指標 難度(平均答對率)	鑑別度
5	偏易(約 79%)	一般(約 0.36)

統計結果 群組	平均數, 中位數, 內部標準差	與均一平均數差距, 中位數差距, 內部標準差差距
高分群	92.86, 100.0, 8.25	-7.14, 0.0, 8.25
中分群	71.67, 83.33, 21.15	-15.83, -4.17, 8.65
低分群	0.0, 0.0, 0.0	-63.89, -66.67, -10.39

高分群名單：B, K, M, W, A, C, D, L, O, S, T, V

中分群名單：P, H, F

低分群名單：E, G, H, J, N, Q, R



■ 課綱單元名稱：2-1 點、直線與圓之間的位置關係

試題數目 \ 試題指標	難度(平均答對率)	鑑別度
5	難易適中(約 72%)	極優(約 0.56)

統計結果 群組	平均數, 中位數, 內部標準差	與均一平均數差距, 中位數差距, 內部標準差差距
高分群	97.62, 100.0, 5.83	-2.38, 0.0, 5.83
中分群	53.33, 66.67, 26.67	-34.61, -21.56, 18.85
低分群	0.0, 0.0, 0.0	-44.3, -41.66, -15.81

高分群名單：A, B, C, K, V, W, L, M, O

中分群名單：S, T, U, P, D, H

低分群名單：E, F, G, H, J, N, Q, R

■ 課綱單元名稱：2-2 圓心角、圓周角及弦切角

試題數目 \ 試題指標	難度(平均答對率)	鑑別度
5	略偏易(約 76%)	優(約 0.47)

統計結果 群組	平均數, 中位數, 內部標準差	與均一平均數差距, 中位數差距, 內部標準差差距
高分群	100.0, 100.0, 0.0	0.0, 0.0, 0.0
中分群	63.33, 83.33, 27.69	-26.79, -6.16, 22.15

低分群	0.0, 0.0, 0.0	-59.3, -60.48, -15.11
-----	---------------	-----------------------

高分群名單：A, B, C, M, U, V

中分群名單：K, L, O, S, T, W, P, D, E, H

低分群名單：F, G, H, J, N, Q, R

