

# 行政院國家科學委員會補助專題研究計畫成果報告

## 微積分學習之多元化輔助教材的研發與評量之研究 子計畫二：教學評量的研究

計畫類別： 個別型計畫          整合型計畫

計畫編號：NSC 89 - 2511 - S - 032 - 005 -

執行期間：89年8月1日至90年7月31日

計畫主持人： 王健華      淡江大學教育科技學系

計畫參與人員：尹維洸      淡江大學教育科技學系

游昇達      淡江大學數學系

本成果報告包括以下應繳交之附件：

赴國外出差或研習心得報告一份

赴大陸地區出差或研習心得報告一份

出席國際學術會議心得報告及發表之論文各一份

國際合作研究計畫國外研究報告書一份

執行單位：淡江大學教育科技學系

中 華 民 國 90 年 10 月 30 日

# 微積分學習之多元化輔助教材的研發與評量之研究

## 子計畫二：教學評量的研究

計畫編號：NSC 89-2511-S-032-005-

執行期限：89 年 8 月 1 日至 90 年 7 月 31 日

主持人：王健華 淡江大學教育科技學系

計畫參與人員：尹維洸 淡江大學教育科技學系  
游昇達 淡江大學數學系

### 一、 中文摘要

本研究為配合總計畫多元教學教材之發展，將學習成就評量定為教學評量研究之第一階段，且決定發展為網路適性測驗之模式。計畫年度內，為求時效，在教材發展子計畫進行之同時，本子計畫以網路適性測驗模型之建立以及試題開發雙軌進行。

在適性測驗模型建立方面，兩種適性測驗模式，AMT (Adaptive Mastery Test)、以及 SPRT (Sequential Probability Ratio Test)、分別依據其運算法則成功轉化為電腦程式，並以模擬方式加以測試，而以 SPRT 較適合本研究之規模。本研究並以電腦模擬，修改原型之 SPRT，其能設計合理之評等級距。

在題庫建立方面，本子計畫配合教材發展，陸續設計配合教學單元之測驗題。題目來源部份為數學系助教所自行命題，部份系取材自原文教科書翻譯而來。原文為計算之試題，均改編為選擇題之型式。由於 SPRT 假設題目難易一致，篩選難易適中題目之法則也已規畫，待測驗題發展至足夠題數，即可進行篩選與正式測試。

**關鍵詞：**適性測驗、SPRT (Sequential Probability Test)、受測態度

### Abstract

The first stage of this study focus on the development of an adaptive testing model for assessing the learning achievement. In the meantime, the test items for the item pool are also developed.

For developing an adaptive testing prototype, two available models are evaluated and programmed to simulate the real situation. These two models are: Original SPRT (Sequential Probability Ratio Test) and AMT (Adaptive Mastery Test) by employing IRT. The Original SPRT is selected as it requires no subjects for pilots. SPRT model is also modified by this study so that grading learning achievement into categories becomes possible.

For filling the test item pool, multiple-choice questions are developing, in accordance with the development of instructional unit. Items are either developed by graduate assistant of the Math Department, or by translating from English textbook. In order to meet the assumption of equal difficulty of the SPRT model, a procedure for filtering extreme questions are also developed. The filtering process and the test for effectiveness of the model are able to carry out in the second stage, as long as sufficient test items are gathered.

**Keywords:** adaptive testing, SPRT (Sequential Probability Test), test attitude

## 二、緣由與目的

本研究之總計畫旨在利用多元化教學方式，改變傳統完全由教師講授之微積分教學法，以期能提高學習興趣與教學成效。而多元化學習重要管道之一即為網路。因此本子計畫第一階段之研究目標，教學評量中之學習成就評量，亦在求突破傳統紙筆測驗之方式，改採網路適性測驗。電腦適性測驗，基本上具有即時評估、縮短測驗時間、個別進行不受時間限制、作弊不易等優點，尤其如以網路作為工具，更可以突破空間的限制，因此雖然一般適性測驗模式均須以測驗題進行，計算、問答等題型均無法適用，但如出題適當，仍然可以具有相當準確性。

所有適性測驗模式均需借重電腦的即時運算能力，而網路為即時傳輸最為便利之工具。網路化之適性測驗，在本計畫中，既可以作為教學前的前測 (pre-test)，也可以作為單元單元教學的 embedded-test，更可作為單元教學完成之後的後測 (post-test)。而且，無論單元教材是否利用網路傳輸，網路化之適性測驗均可獨立運作或溶入網路學習過程中運作，因此本子計畫之教學評量，在第一階段成就評量方面決定採用網路化之適性測驗。

在適性測驗眾多的模式之中，項目反應理論 (Item Response Theory, IRT) 是最為廣泛應用的適性模式，IRT 也幾乎成為適性測驗的同義詞 (Bunderson, Inouye, & Olson, 1989)。但是 IRT 依據所使用的模式，需要 200 至 1000 名受試者以校準題庫中的項目參數 (Weiss & Kingsbury, 1984)，如此龐大的人力、物力的需求，使得以 IRT 模式為基礎的電腦適性測驗幾乎不可能應用於教室中的日常測驗情境。

本計畫選定引用 SPRT (Sequential Probability Ratio Test) 作為發展的基礎。SPRT 早由 Wald (1947) 所發展出來，直至 1969 年，Ferguson 才將其應用來判定學生的精熟程度 (mastery or nonmastery)。SPRT 模式不考慮項目特徵，(項目難度、鑑別力，與猜測因素)，不需依據受測者的成就水準選擇項目 (Frick & Plew, 1989)，因此不需要繁複的程序判定題目

的項目特徵，題目係隨機選取，相當適合本計畫的規模 (參見表一)。

表一 兩種適性測驗模式的比較

(摘自 Welch & Frick, 1993)

適性測驗模式	項目選擇方法	校準題庫所需人數	項目參數的計算
SPRT	隨機出現	不需校準	不需要
IRT	依據受試者反應	200-1000 人	需要

雖然 SPRT 沒有複雜的計算，然 Frick et al. (1989) 的研究顯示，當受測者能力分布並未集中在精熟程度的臨界點 (mastery cut-off) 上時，SPRT 模式的測量準確度是可以信賴的。

由於原型之 SPRT 僅能判定精熟與否，本子計畫亦研究將其改良成為能夠判定學習成就級距 (例如: A, B, C) 之改良型 SPRT。改良之成果請參見下節敘述。

本子計畫之另一重要目的即為發展測驗所需之題目，因 SPRT 採用隨機選題，因此判定測驗終止所需之題數較 IRT-based 模式為多，尤其當區分成就至較高等級之時，所需之題目可能更多，加上為求滿足 SPRT 模式題目難易均等之假設，僅有 1/3 難易適中之題目得以置入題庫，因此測驗題之發展為重要工作項目。

## 三、結果與討論

本子計畫在年度內已依據 SPRT 原型之演算程序 (algorithm) 完成該適性模式之電腦模擬程式，並已將題庫系統建構完成，待教材發展子計畫發展出足夠測驗題，此一原型模式即可正式運作測試。此外，本計畫並研究修改此一原型模式，突破其只能作為精熟測驗的限制。利用改變「精熟上標」與「精熟下標」之設定值，可使得判斷「終止測驗」之臨界值亦隨之改變。由於達到越高之「精熟上標」所需的題數越多，因此當受測者達到較低之「精熟上標」時，電腦程式並不終止測驗，持

續進行直至觸及該等距之「精熟下標」，則判定「挑戰該等級失敗」；或超越最高之「精熟上標」，則判定「挑戰最高等級成功」。電腦模擬之各等級所需之測驗題數如附表。

模擬程式所得之資料，清楚顯示出在設定不同之「精熟上標」與「精熟下標」值，以及設定「上標」與「下標」之間的不同級距，對於判定不同等第「通過」與「失敗」所需之答對與答錯題數間之關係。有關判定等第所需之題數，從附表中可歸納出如下之規律：要判定通過越高等級以及設定越接近之級距，需要越多之題目；反之，要判定沒有通過越低等級以及設定差距越大之級距，所需之題目越少。然而，在實用上，目前仍需研究的是，如何判定那一等第相當於傳統紙筆測驗的多少分？且其準確度為何？其結果是否為學生信服？這些問題，均需積極在第二階段的研究中尋求解答。

有關題庫之發展，本子計畫配合教材發展，陸續設計配合教學單元之測驗題。題目來源部份為數學系助教所自行命題，部份係取材自原文教科書翻譯而來。原文為紙筆計算之試題，均需改編為選擇題之型式。部份翻譯之題目，原文提供有難易等級之區分，然此難易等級，並不一定適合我國學生，而由數學系研究生助教所自行發展之題目，則並無難易等級之區分。由於 SPRT 假設題目難易一致，因此過度簡單或困難之試題均不宜置入題庫。本計畫亦發展出一套簡易之流程，用以篩選出難易適中題目。此項法則，不需住眾多之預試對象，一般而言，30-50 人所作之預試，即具有參考價值。此項流程極為簡單，一般教師均可自行運用。此法則為：針對每一題，計算該題答對人數與總人數之比值，取比值在中間 1/3 之試題置入題庫，以期接近 SPRT 模式之假設。

值得討論的是：在期末考等「後測」中，因內容範圍較大，也許還能發展出數百題以供篩選，但若作為 embedded test，因其僅包含一個或少數幾個教學單元，因此勢必無法發展出上百題測驗題。以微積分教學而言，解決此一困難的方法之一，是在原有的模式之內，再設計一個

程式，使電腦在一定範圍內可以隨機設定某些問題的「數據」，然後由電腦依隨機選取之數據即時依題意計算出一組三錯一對的答案以供考生選擇。如此，題庫的規模將可相當可觀的提昇，而且只要設定的數據範圍適當，即使不同數據所產生出來的題目，其難易也相似。此項程式之設計，亦將是本計畫第二階段的主要目標之一。

#### 四、研究成果

本計畫最重要之成果為：

- (1) 突破大學基礎數學教學成就評量均需以紙筆、定時、定點進行的傳統，首度嘗試以即時評量之網路化適性測驗為主要評量工具，計畫第一階段已完成適性測驗模式之雛型，並經過電腦模擬，證實其運作完全符合模式之運算法則 (algorithm)。第二階段正式測試，除將分析此模式對大學微積分評量之精準程度外，更將配合問卷，分析受測者受測態度之傾向。這模式將提供另一途徑供大學基礎數學教育之學習成就評量。
- (2) 突破國內傳統適性測驗大部份均以 IRT 為理論基礎之設計模式，採用較易操作之 SPRT 模式。此模式雖不是新發展出來，但利用此一模式之測驗，不需大量樣本以測定題目之參數，沒有繁複的運算法則，適於中小規模之測驗。而大學微積分之教學成就評量通常即為中小規模之測驗。
- (3) 嘗試將原型之 SPRT 改良成為具有評定等級功能之適性測驗模式。改良之雛型亦已發展完成。第二階段亦將測試其應用在微積分學習成就評量中之精確程度，延伸原型 SPRT 模式僅能以二分法判斷精熟與否的單一功能。
- (4) 嘗試將大學微積分學習成就評量之測驗題設計為選擇題，並將在第二階段以電腦在設定範圍內隨機選定問題中之數據並即時產生答案。如此將可大幅降低學生猜題或作弊成功之機率。

## 五、參考文獻

- Bunderson, V., Inouye, D., & Olson, J. (1989). The four generations of computerized educational measurement. In R. L. Linn (Ed.), *Educational measurement*. New York: Macmillan.
- Ferguson, R. (1969). *Computer-assisted criterion-referenced measurement* (Report WP-41). Pittsburgh: Pittsburgh University, Learning Research and Development Center.
- Frick, T., Plew, G. & Luk, H. (1989, March). *EXSPRT: An expert systems approach to computer-based adaptive testing*. Paper presented at the Annual Conference of the American Educational Research Association. San Francisco.
- Wald, A. (1947). *Sequential analysis*. New York: Wiley.
- Wang, A. C. (1992). An in-depth analysis of computer-based adaptive testing methods. Paper presented at the Annual Conference of the AECT, Hong-Kong.
- Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21, 361-375.
- Welch, R. & Frick, T. (1993). Computerized Adaptive Testing in Instructional Settings. *Educational Technology Research and Development*. 41(3). 47-62.

附表

<b>達到精熟</b>	[精熟：0.65] [非精熟 1：0.4] 組距：0.25 [非精熟 2：0.2] 組距：0.45	[精熟：0.75] [非精熟 1：0.5] 組距：0.25 [非精熟 2：0.3] 組距：0.45	[精熟：0.85] [非精熟 1：0.6] 組距：0.25 [非精熟 2：0.4] 組距：0.45	[精熟：0.95] [非精熟 1：0.7] 組距：0.25 [非精熟 2：0.5] 組距：0.45
<b>錯誤 0 題</b>	7 [0.25] 3 [0.45]	8 [0.25] 4 [0.45]	9 [0.25] 4 [0.45]	10 [0.25] 5 [0.45]
<b>錯誤 1 題</b>	8 [0.25] 4 [0.45]	9 [0.25] 5 [0.45]	12 [0.25] 6 [0.45]	16 [0.25] 9 [0.45]
<b>錯誤 2 題</b>	9 [0.25] 4 [0.45]	11 [0.25] 6 [0.45]	15 [0.25] 8 [0.45]	22 [0.25] 12 [0.45]
<b>錯誤 3 題</b>	10 [0.25] 5 [0.45]	13 [0.25] 7 [0.45]	17 [0.25] 10 [0.45]	28 [0.25] 16 [0.45]

<b>達到非精熟</b>	[精熟：0.65] [非精熟 1：0.4] 組距：0.25 [非精熟 2：0.2] 組距：0.45	[精熟：0.75] [非精熟 1：0.5] 組距：0.25 [非精熟 2：0.3] 組距：0.45	[精熟：0.85] [非精熟 1：0.6] 組距：0.25 [非精熟 2：0.4] 組距：0.45	[精熟：0.95] [非精熟 1：0.7] 組距：0.25 [非精熟 2：0.5] 組距：0.45
<b>答對 0 題</b>	6 [0.25] 4 [0.45]	5 [0.25] 3 [0.45]	4 [0.25] 3 [0.45]	2 [0.25] 2 [0.45]
<b>答對 1 題</b>	7 [0.25] 5 [0.45]	5 [0.25] 4 [0.45]	4 [0.25] 3 [0.45]	2 [0.25] 2 [0.45]
<b>答對 2 題</b>	8 [0.25] 7 [0.45]	6 [0.25] 5 [0.45]	4 [0.25] 4 [0.45]	2 [0.25] 2 [0.45]
<b>答對 3 題</b>	9 [0.25] 8 [0.45]	7 [0.25] 6 [0.45]	5 [0.25] 4 [0.45]	3 [0.25] 3 [0.45]

Alpha = 0.05

Beta = 0.05