

# 自然學科成就測驗的評量— 馬里蘭州及臺灣學生學習成就評量計畫

\*曾文雄 \*\*吳美麗 \*\*\*卓娟秀  
\*國立臺灣師範大學生物系  
\*\*台北市立師範學院數理系  
\*\*\*台北市立師範學院附屬小學

## 一、前言

國內中、小學自然學科的教學成就測驗，大都是教師自編的選擇、是非等類型的非標準化測驗，而測驗的分數則用以比較學生個人成績的高低，在班級或年級所佔的地位，以決定學生個人的成就水準，給予名次，此種對測驗結果的解釋屬於常模參照評量的方式。另一種較少被應用的為標準參照評量的方式，即根據教學目標或內容指標，評量學生對教材內容精熟的程度，以判斷學生學會什麼，還沒學習到什麼，以瞭解教學的效果，檢討教學的缺失，做為改進教學及輔導學生學習的依據（Gronlund, 1982）。

國內目前正在推展全國性的教育評量計畫，訂定學習目標及內容規範，建立教育成果指標，鑑定各地區、各級學校、及各類學生的學習差距，這些目標的評鑑工作，需採用適當的評量方式。教育部為此主辦「國民教育階段學生基本學習成就評量研究」計畫，為使研究順利進行並獲得教育界人士的支持，先後在八十一年十二月（國立臺灣師範大學，民81）及八十二年十二月（國立臺灣師範大學，民82）舉辦兩次學科基本學習成就評量國際研討會，邀請美國、日本及我國學者專家，教育行政人員，國民中、小學校長參加，以溝通觀念，建立共識，協助推動研究工作。自八十二年度起委託國立臺灣師範大學、國立台中師範學院辦理，為求國際間之共同研究比較，第一階段與美國馬里蘭州教育廳共同合作進行「馬里蘭州學校學習成就評量計畫」MSPAP（The Maryland School Performance Assessment Program）（簡茂發等，民83a.；李虎雄等，民84）。

MSPAP採用標準參照評量方式，本文在闡述此種評量應用在國內自然學科的施測，並以生物科為例，及探討相關的問題。

## 二、測驗架構

### (一) MSPAP

馬里蘭州為提昇州內各學校的學習成效，實施四種學習成就評量，包含一種全國性及三種該州自行設計的評量工具（Gabrys, 民82），MSPAP 即為自定的一種，在每年的五月，對全州三、五、八年級的每一個學生進行測驗，來評定各校是否達到該州教育委員會所規畫的目標，以提供高品質的教育環境，造就能適應二十一世紀社會需求的學生。測驗的內容包涵數學、閱讀、寫作、科學、語言使用及社會，測驗時間為五天，每天105分鐘，所需時間總共不超過九小時（MSPAP, 1993a）。此種測驗對五年級學生實施的數學及科學測驗，在八十三年擴展到中華民國臺灣地區及德國的巴登—吳登堡地區實施，以增進國際間對基本能力評量方面的知識交流及合作發展（MSPAP, 1993b）。

### (二) 國內測驗

因應國情的不同，在臺灣地區將科學再分為物理、化學及生物三科。施測樣本為全省十五所學校的六百名五年級學生，每校祇取一班，學校涵蓋北、中、南、東部之智、仁、平勇、偏勇、特勇五類。國內試題僅選擇MSPAP的部分試題含數學7題，物理、化學及生物各3題。試題分為三組群（5A, 5B, 5C），5A及5C試題含數學兩題，物理、化學、生物各一題，5B則數學有三題，其餘科目題數與5A, 5C相同。試題經翻譯成中文，並略加修訂，在兩天內完成施測，各校祇測驗一組群試題，在相同日期及時間內，由各班級導師擔任監試進行測驗，每一組群試題共有五個學校學生受試（簡茂發等，民83a）。

### (三) 測驗的內容指標

由於此測驗的目的在評量學生的基本學科學習的精熟程度，所以試題內容的設計，以學習結果（outcome-based）為基礎，每一年級訂定內容指標，試題內容即根據內容指標而擬定，評量結果也以是否達到內容指標為依歸。科學的內容指標分為六項：(1)科學概念，(2)科學本質，(3)思考習慣，(4)態度，(5)科學過程，(6)科學應用。各項的內容指標細目又依年級而異，隨年級的增加，指標的層次由淺而深，五年級的六項內容指標及細目參見表一。

### (四) 試題設計

有鑑於是非、選擇等試題型式不易評量思考過程、綜合、分析、判斷、應用等能力，

表一 五年級科學課程內容指標

- |  |   |   |
|--|---|---|
| <p>(一) 科學概念</p> <ol style="list-style-type: none"><li>1. 收集證據以解釋對生物間相互作用及相互依賴的現象，所觀察的結果。</li><li>2. 在物理和化學領域中可觀察的交互作用和變化。</li><li>3. 指認、探究使地球成為太陽系中獨一無二行星的自然資源與天然過程。</li></ol> | <p>(二) 科學本質</p> <ol style="list-style-type: none"><li>1. 瞭解科學知識長久以來持續發展。</li><li>2. 科學知識以證據為依據。</li><li>3. 科學知識使我們能作預測。</li><li>4. 在發展物理模型時顯示創意。</li><li>5. 發展可驗證的假設。</li><li>6. 認識不同組群、不同地點、不同時間收集的資料相互比較非常重要。</li></ol>   | <p>(三) 思考習慣</p> <ol style="list-style-type: none"><li>1. 發展檢驗方法以尋找答案。</li><li>2. 提供證據以支持答案。</li><li>3. 產生「如果……會如何？」的疑問。</li></ol>   |
| <p>(四) 態度</p> <ol style="list-style-type: none"><li>1. 知道科學家是探討日常問題的。</li><li>2. 每個人都能從事科學活動。</li><li>3. 鍥而不捨。</li><li>4. 熱衷接受具挑戰性的工作。</li></ol>                               | <p>(五) 科學過程</p> <ol style="list-style-type: none"><li>1. 使用發展性的適當工具和材料以顯示：<ol style="list-style-type: none"><li>(a) 下操作型定義</li><li>(b) 質疑</li><li>(c) 發展可驗證假設</li><li>(d) 設計公平公正的驗證方法</li><li>(e) 用公制單位作測量</li><li>(f) 整理並呈現資料</li><li>(g) 解釋證據和推論</li></ol></li><li>2. 用口頭或文字解釋發現。</li></ol> | <p>(六) 科學應用</p> <ol style="list-style-type: none"><li>1. 使用科學的知識和現有的科技解決實際問題。</li><li>2. 描述地區性的環境問題。</li><li>3. 使用科學知識以決定自己怎麼處理這個問題。</li><li>4. 向別人描述自己的科學知識怎樣幫自己下決定。</li><li>5. 設計和執行處理地區性環境問題的行動計畫。</li></ol> |

所以MSPAP的試題全部採用問答式。其試題的設計流程為，首先提供一種情境，然後提出一系列的問題，要求學生進行一系列的活動，來逐次解決問題，回答問題。評量則根據學生在活動解題過程中，對科學概念、過程、應用等的瞭解與表現能力。命題的情境，在物理及化學題目，通常以日常的事例引發，並設計「實作」試驗，令學生由實作中學習，獲取答案；至於生物題目，則由於學科性質的特性，很難在施測的短時間內，經由「實作」試驗呈現出生命現象的內涵及變化，所以生物題目的命題缺少實作試驗，通常以文字敘述或圖表來替代實作，直接呈現出長時間的「實作」結果。學生對問題的回答型式，則要求以文字敘述或以圖表來表達。

#### (五) 生物試題剖析

生物的三个試題，A卷為「動物物種的生存」，B卷為「浮萍的難題」，C卷為「爺爺的湖」。每卷各有3種相關資料：資料本、作答本及指導手冊（簡茂發等，民83b）。資料本及作答本在施測時發給學生，指導手冊則僅供教師使用，提供準備實作，及施測監試時的指導用語和有關事項。資料本提供學生有關試題的文字敘述、圖表或實作指導，問題及回答則列在作答本內。茲將生物的三个題目及問題剖析如下，以便了解此種命題及評量方式和其理念：

##### 1. A卷「動物物種的生存」

本試題為有關「動物物種的生存」系列試題的一篇，篇名為「鳥類的糞粒」，主要目的在使學生由對鳥糞特性的了解，進而全盤了解貓頭鷹的習性及生態。資料本中介紹鳥類糞粒的特性，不同型式的鳥糞，檢查貓頭鷹糞粒的實驗步驟，貓頭鷹糞粒內含物的圖片八張，老鼠、地鼠和田鼠三種鼠類的特徵及生態習性，圖文並茂，資料相當豐富。作答本有八個活動及問題，其中三個為評量語言使用的活動問題，祇有五個為生物問題，分別為根據資料本中提供的資料，做綜合、比較及歸納的功夫，定義鳥糞及鳥糞如何產生，檢查鳥糞的步驟，繪製貓頭鷹的食物網及能量流動的方向；由鳥糞內含物的圖片，進行觀察、比較及推理，辨識出鳥類食物的名稱及數量。推測鼠類可能會消失的原因及如果鼠類滅絕後，貓頭鷹的食物可能發生的變化等有關生態系消長的問題。全部測驗時間為40分鐘，由於資料本的文字敘述資料豐富，語文閱讀能力較差的學生，不容易掌握作答時間。

##### 2. B卷「浮萍的難題」

資料本僅提供一個水族箱的內容物在3週中變化的表格，表中有3個縱欄，第一欄為「平均水溫」，第二欄為「植物—漂浮的浮萍」，第三欄為「動物—孔雀魚、螺」；

橫欄亦有3欄，分別為第1週、第2週及第3週，以數字表示縱橫、交叉格子中，每天4個時刻的平均水溫，植物、動物數量的變化，植物形態及動物行為的變化。表中內容相當精簡，主要在評量學生解讀表格的能力，及綜合、研判水溫、植物、動物間互動及因果關係的能力。作答本有五個活動及問題，分別為將表格的水溫數據繪製成折線圖，提出水溫變化、浮萍增殖及螺死亡的假設，及提供證據以支持自己所提出的假設，提出臺灣地區的淡水池塘，如果浮萍迅速增殖所產生的環境問題，和如何應用科學知識，以解決此種環境問題。

### 3. C卷「爺爺的湖」

資料本呈現兩張折線圖及兩張湖的剖面示意圖，折線圖中表示出自1953年至1993年的40年間，每年流入湖中的沉積物數量和陽光自湖面可照射到湖水中的深度；湖的剖面圖中標出1953年及1993年湖中沉積物數量、水量、動物、植物的相對變化。

本試題在學生開始作答活動前，有類似物理、化學科目試題的「實作」活動，但僅由教師做示範操作，在教桌上事先備好兩個廣口瓶，內含沙及水，分別示意為1953年及1993年的湖中內容物，示範時將表示1993年的瓶中加入少許沙並用力搖晃，使瓶內沙水混濁，並逐漸產生沉積。隨後師生共同討論植物生存所需東西的清單，再由學生個別回答問題。作答本有四個問題，包括解讀折線圖中的資訊，並將沉積物和陽光兩個變數的資料綜合的能力之評量；食物鏈中的一個變數發生改變時，對食物鏈的影響；預測受到污染時，湖可能發生的變化和利用科學知識技能，來設計和解決地區性環境污染問題。

綜觀三個生物題目，此種命題評量有四項特色：

(1) 題目內容和課本教材無直接關連。題目內容從日常環境中取材，學生祇需學習到科學概念、本質、過程，培養思考習慣，科學態度，就足以應用科學知能，解決問題，因此教學活動無需強調死背科學知能，而以整體性能夠達到內容指標為目標。

(2) 測驗的題目形式為問答式，而不採用選擇、是非題，以方便評量較為深入的科學概念，思考、分析、綜合、判斷、應用等能力。

(3) 題目及作答以各種資訊形式呈現，有文字、圖片、表格、圖表，不侷限於文字的敘述方式。

(4) 經常以地區性的情境導引出問題，佐以「實作」或「演示」，再陸續提出一連串相關的問題，啟發學生的思路，最後以解決問題為目的，建立一套面對問題，以科學解決問題的完整流程。

### 三、評 量

#### (一) 評量模式

由於試題都是開放性的問答題，因此沒有固定的標準答案，凡是合理的答案都是正確的，而學生的答案都是以文字或圖表來陳述，所以建立一套明確的評分標準及程序十分重要。針對每一個問題，先擬定可能的作答類型，每一類型給予一個代號，再參考 MSPAP 的給分標準及我國各學科參與研究教師的意見，訂定評分手冊，供評量教師遵循使用。為避免評量尺度寬嚴不一，對評分手冊的規定解釋不同，會影響評量信度，因此每一位評量者都需接受嚴格的評量講習及考試，MSPAP 的評量資料有評分手冊 ( Scoring Guide )，訓練手冊 ( Training Sets ) 和資格測試手冊 ( Qualifying Sets ) 三種 ( MSPAP, 1993c )，並委由評量公司 ( Measurement Inc. ) 的專家以其規定的評分模式訓練，考試及格後，方能進行評量工作。

#### (二) 評量結果

測驗的結果，學生對問題精熟的程度，應以表一的內容指標為依據，每題計分以答對的學生百分比來計算，以符合評量目標，例如某校全班學生有百分之八十答對折線圖，通過內容指標中第五項「科學過程」中的整理，並呈現資料的能力。至於個人成績的計算，以便探討與學習成果有關的因素時，則需用項目回答理論評分 ( Item Response Theory Scaling )，訂定學生在全部學科或學科內主要項目的程度指標，此種指標再依客觀標準訂層次，如分為滿意、中等、優等三個等級，看有多少比率的學生，達到各種程度。標準的訂定，則反映社會、家長及教育人員對教育的期望與可行性，可隨客觀的條件而做彈性的改變 ( 彭森明，民 82 )。

由於國情不同，我國的測驗結果暫不以 MSPAP 的內容指標為評量依據，也未進行項目回答理論評分，而另以「解決問題」的內容指標為評量依據，個人成績也依我國自行訂定的評分手冊計分。「解決問題」過程中，各項心智運作能力可分為八大流程：(1) 認識情境，發現問題，(2) 提出問題，確定問題，(3) 提出假設，形成策略，(4) 落實計劃，(5) 執行計劃，(6) 整理分析，解釋資料，(7) 規則性應用，及(8) 模型理論。各流程再依情況細分為各種能力如表二所示。生物的 A、B、C 三卷綜合起來，共有 24 小題的問題，各問題所評量的能力，如依照「解決問題」的模式歸類如表二的右欄所示。每一小題給予 4 個字代號，第一個字為英文字母，代表卷別，第二個數字代表測驗的排序別，第三及第四個數字代表問題號碼及問題下再分的小問題號碼，例如 ( C553 )，代表 C 卷的

表二 解決問題之實測

解決問題過程中的各項心智運作能力		評量題目
1. 認識情境，發現問題。	1.1 察覺問題	
	1.2 認明問題	
2. 提出問題，確定問題。	2.1 界定問題特質、現況、目標	
	2.2 認識相關變因及其輕重關係	
3. 提出假設，形成策略。	3.1 預想可能的情形	(C521) (C533)
	3.2 選擇策略、方法、工具	
4. 落實計劃（規劃工作、安排流程	4.1 安排工作流程、估量進度	(A451) (B441)
	4.2 安裝選購儀器、安排工作環境	
	4.3 預想執行情況、設計圖表、安排相關支持	(B452) (C541)
5. 依計劃執行	5.1 照指示施測、觀察	
	5.2 操作、調整儀器、克服執行上的困難	
	5.3 描述觀察的現象及記錄 (a)科學性描述，掌握關鍵變因 (b)實作將資料登錄 (c)「操控變因」之因了解與操作	
	5.4 處理井然有序	
6. 整理資料分析、解釋。	6.0 了解、整理、歸類資料	(A411)
	6.1 將資料轉換成數據	
	6.2 運用圖表、列表、數線去表達數據	(A461) (B411)
	6.3 評析資料之可信度、涵義 (a)取舍資料之判別 (b)對資料之誤差、可信度、適用圍之評估 (c)評估資料趨勢 (d)評估均值、不變異、變異值 (e)與自己假設、別人資料相比較評估可信度	(A471)
	6.4 解釋相關關係、規則性 (a)定性研判 (b)定量規則性	(A421) (B421) (B422) (C511)
	6.5 形成新的問題及新的研究	
7. 規則性應用	7.1 規則性，應用於相似情境中	(A431)(A441)(A472)(A482)(B431) (B432)(B451)(C531)(C532)
	7.2 將規則性應用於不同情境，有如通則	(A481)
	7.3 使用於規則有如「定律」	
8. 模型理論	8.1 瞭解模型 (a)了解模型涵義 (b)模型及實驗之代表性、可用性 (c)了解模型之局限性	
	8.2 提出模型、理論	
	8.3 賦予 $y=f(x)$ 意義，用來解釋資料	

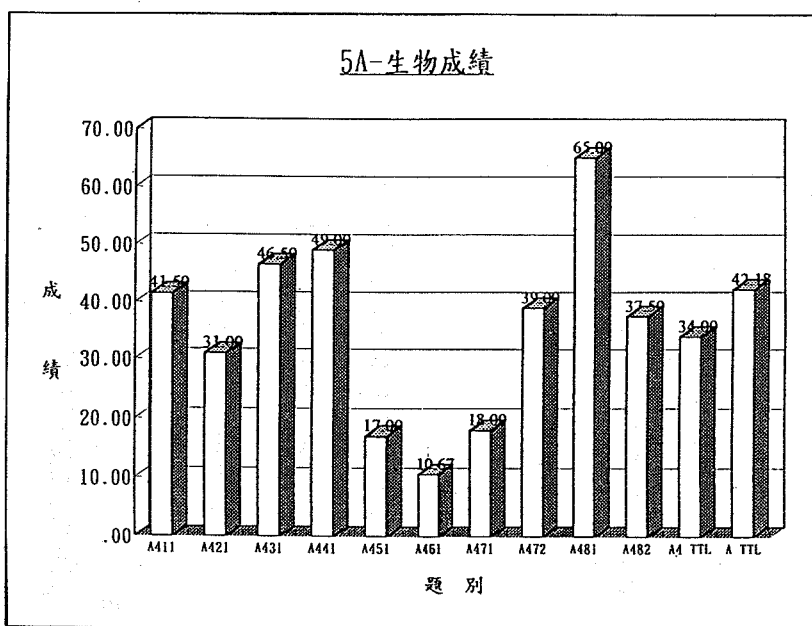
第五個測驗之第五個問題的第三小題。

每一小題因內容多寡及難易不同，所給的分數由 1 分至 3 分不等，為客觀比較各小題的回答正確率，每小題的得分以下列公式計算：

$$\text{得分} = \frac{\text{平均分數}}{\text{最高分數}} \times 100$$

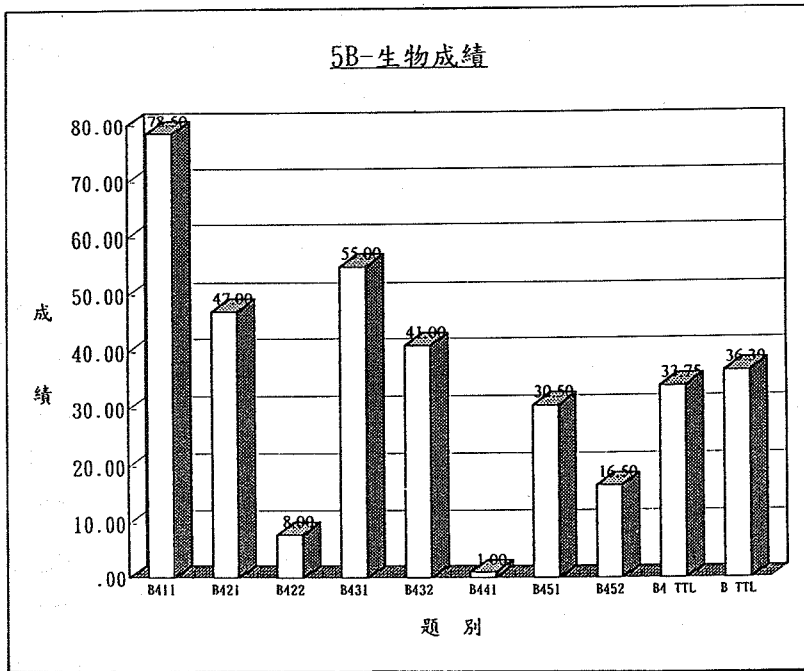
例如 (A 441) 的最高分數為 2 分，而五所學校學生該題的平均分數為 0.98 分，所以該題的平均得分以百分制計算為  $(0.98/2) \times 100 = 49$  分。生物科各小題得分，該測驗總得分及該卷總得分的柱形圖，A 卷見圖一，B 卷見圖二，C 卷見圖三。

由表二得知，全部 24 個題目中，以 (7.1) 的應用規則性於相似情境中有 9 題，佔 37.5% 為最多；其次為 (6.4) 的解釋相關關係、規則性，有 4 題，佔 16.7% 居次；有 2 題的有 (3.1)(4.1)(4.3)(6.2) 等細目；有 1 題的有 (6.0)(6.3)(7.2)。其餘流程的細目，均無任何評量問題可以歸項，一方面是表二僅就生物的 24 個問題分類，題目太少；原來的科學題目是不分科的，而在臺灣的施測又祇採用 MSPAP 的部分試題而已；另一方面是生物科學的特性之限制，如「模型理論」、「提出問題」、「確定問題」等，在國小階段而言，其層次偏高。

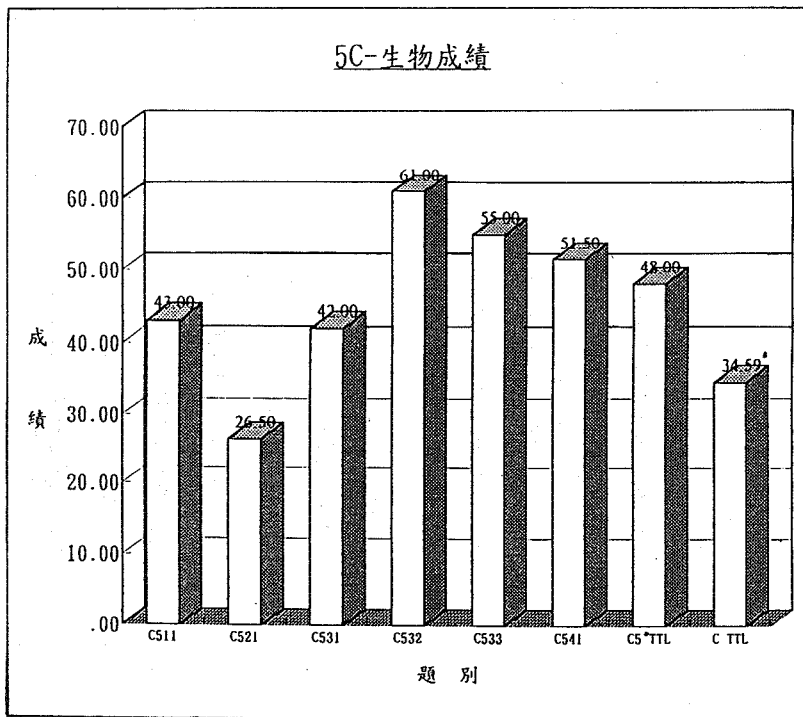


圖一 A 卷生物科成績





圖二 B卷生物科成績



圖三 C卷生物科成績

由圖一、圖二及圖三中得知，各問題得分的多寡，差別很大。由高而低排序，(B 411) 78.5分，(A 481) 65分，(C 532) 61分，佔全部試題的12.5%。有15題得分在60分至20分之間，佔62.5%。有6題得分在20分以下，佔25%，(B441) 得分最低，只得1分。各題得分的多寡與「解決問題」步驟的分類沒有呈現一致性，例如歸類在(6.2)的運用圖表、列表、數據去表達數據能力的有(A 461)及(B 411)兩題，(A 461) 得分10.67分，(B 411) 則為78.5分。究其原因為「解決問題」的模式，其步驟分類尚欠週全或定義不夠明確，題目在歸類時錯誤，或係由於即使歸類正確，但題目本身有深淺難易之別，所以得分多寡不同。

各題得分的內部一致性用克郎巴赫係數(Cronbach's  $\alpha$  Coefficient)表示。A卷10題， $\alpha$ 值為0.6647，標準化題數 $\alpha$ 值為0.6641；B卷8題， $\alpha$ 值為0.6381，標準化題數 $\alpha$ 值為0.6318；C卷6題， $\alpha$ 值為0.6435，標準化題數 $\alpha$ 值為0.6825。

A、B、C三種試題，每一種試題均有五所學校的學生施測，評量結果顯示，B卷及C卷的各五所學校間，學生得分無顯著差異；A卷則有顯著差異，以薛費氏(Scheffe's test)事後檢定顯示城市學校得分高於鄉鎮學校。A、B、C卷得分，男、女性別無顯著差異。

#### 四、討論及結論

本研究的第一階段着重在MSPAP評量工具做為國際間共同研究的可行性，就中文翻譯、施測過程及評分而言，在我國的研究結果，顯示正面的意義(李虎雄等，民84)。施測後，訪談六所學校的教師及學生，師生均對此種測驗方式予以肯定，尤其是學生，大都表示喜歡此種不限制在課本內容的試題，令他們覺得比較沒有考試的壓力感。但對教師而言，雖然讚賞此種測驗工具，但因評量的工作負荷加重，今後的教學活動將更為任重道遠，需要加強新知及再教育。

我國正在發展的學科基本學習成就評量，其目的是在瞭解全國學生在各學科的程度和考核學生程度的升降情況，因此必需訂定學生學習成果指標，此種評量以標準參照評量的方式來解釋測驗的結果，較為適當。國內的自然學科測驗習慣上以常模標準測驗方式來解釋測驗結果為主，今後宜兼顧標準參照評量的方式，在試題內容與測驗方式上改進，需加重對分析、研判及思考能力的評量，測驗方式宜增加問答式的問題，要求學生寫出一段短文或以圖表來回答。

評量的設計和作業方針上，依照美國的教育評量二十多年的經驗，應廣徵各方面的

意見，包括專家學者、教師及社會人士，以達成共識，減少阻力，尤其在實施新的教育評量時，尤需各方的參與。評量內容要密切反映課程標準，而課程標準及內容指標，應依時代的變遷與進步，時時做修正，才能教育出適應未來世界的畢業生。評量的結果，要提供給教師及教育工作和決策人員改進教學之用。

MSPAP在我國的施測，生物科的成績不盡理想，各種解決問題的能力尚無法達到令人滿意的程度。但因參與第一次測驗的學生人數太少，祇有600名，且因實際作業的困難，無法選取具有代表全國學生的隨機樣本，所以測驗結果祇能提供參考，無法做推論至全國學生。另一方面就技術層次分析，我國的教材與美國的教材不盡然能互相銜接，內容指標在各年級也不盡相同，因此依照美國的內容指標而製作的試題，不一定適合我國學生。翻譯成中文的試題，在語意、結構上，有時難免辭不達意，令有些學生無法瞭解。我國學生一向習慣於是非選擇題的測驗方式，對於文字、圖表、圖形的瞭解及表達能力缺少訓練，可能影響其作答能力。

今後國內宜多發展以標準參照評量來解釋測驗結果，根據學習成果指標來設計試題，針對學生熟知的國內情境來命題，而不以課本的內容命題，以帶動教學活動的改進。施測的學生數目宜增加並具有代表性，以便瞭解全國學生在各基本學科上的基本成就及歷年升降的情形，做為瞭解國內教育，改進國內教育的客觀依據。

## 五、後 記

本研究計畫為教育部主辦，臺灣省政府教育廳、台北市政府教育局及高雄市政府教育局協辦。研究委員計有：簡茂發、李虎雄、陳昭地、林保平、王淑真、陳文典、陳義勳、吳碧霞、黃長司、黃萬居、鄭美雪、曾文雄、吳美麗、卓娟秀、張武昌、朱玲玲、林秋麗。

## 參考資料

1. 李虎雄、黃長司（民84）美國馬里蘭州學校實作評量工具在臺灣施測的可行性。科學教育179：41-49。
2. 國立臺灣師範大學（民81）學科基本學習成就評量研討會（I）論文彙編。
3. 國立臺灣師範大學（民82）學科基本學習成就評量研討會（II）論文彙編。
4. 彭森明（民82）學生學習成果評量與分析—美國經驗。學科基本學習成就評量研討會（II）論文。

5. 簡茂發等(民83a)教育部八十三年國民教育階段學生基本學習成就評量研究。研究報告。
6. 簡茂發等(民83b)MSPAP評量工具中文版三套九冊：指導手冊(5A, 5B, 5C)、資料本(5A, 5B, 5C)、作答本(5A, 5B, 5C)。
7. Gabrys, R.E. (1993) The Maryland School Performance Program, An Approach to School Accountability. In: International Symposium on Basic Learning Achievement.
8. Gronlund, N.E. (1982) Constructing Achievement Tests. (3rd ed.) Englewood Cliffs, N.J. : Prentice-Hall.
9. MSPAP (1993a) Test Administration and Coordination Manual.
10. MSPAP (1993b) Proposal to Administer Mathematics and Science Sections of the 1993 Maryland School Performance Assessment Program to Grade 5 Students in Taipei, Taiwan, Republic of China.
11. MSPAP (1993c) Scoring Guide, Training Sets, Qualifying Sets for 5A, 5B, 5C.

★

## ● 更正啓事 ●

本刊第179期「科學史在科學教學的角色與功能」一文中，作者告知在第19頁第一行第二字之後，應增〔kuhn(1962)的理論則是“常態科學”的工作。「根本重建」則牽涉到概念架構的改變，以〕kuhn的觀點……。

特此更正排校失誤之處，謹向讀者深致歉意。

編輯室

★