科學教育學刊
2007, 第十五卷第二期, 215-239

Chinese Journal of Science Education
2007, 15(2), 215-239

# 過程技能之二階段實作評量：規劃、實踐與效益探究

**魯俊賢**[1]　**吳毓瑩**[2]

[1]台北市立雙園國小

[2]國立台北教育大學 心理與諮商學系

（投稿日期：民國 94 年 11 月 21 日，修訂日期：95 年 7 月 20 日，接受日期：95 年 8 月 29 日）

**摘要**：本研究之目的乃在規劃自然與生活科技學習領域中的「設計並進行實驗之過程技能」的實作評量，並且實踐與探究其效益。研究者以能力指標為本，將之轉化為教學評量目標之後，以此來設計實作任務。經過第一個班級（26 名學生）的操作以及修正之後，為有效考慮評量的教學功能，研究者最後確立二階段四要素的實作評量之實施，於第二個班級（25 名學生）再次進行。實作評量第一階段的主要工作是讓學生依據評量任務設計實驗，教師批改之後與學生共同討論並修正錯誤的設計；第二階段的主要工作是讓學生以修正後的實驗設計進行實驗操作的過程技能評量。四要素則包含實作任務（performance task）、反應型式（response format）、評分規準（scoring rubrics）、以及後設規準（meta rubrics）。透過積差相關係數、Cronbach's α 係數及 Kappa 一致性係數分析以及學生的表現，研究者認為二階段四要素實作評量可協助教師分析學生的表現結果，並由學生的表現回饋到能力指標，再從能力指標回饋教學，增進實作評量對於教學的助益。

**關鍵詞**：反應型式、自然與生活科技領域、能力指標、評分規準、實作評量

## 壹、研究目的

隨著科學教育思潮的演進，實作評量逐漸扮演了促進學習勝過檢核評鑑的角色。在過程技能的評量上，更以實作方式來回饋學習成果與教學改進。一般教師在進行過程技能的實作評量時，作法通常是從教材內涵中，將操作性或觀察性的單元提出來，判斷哪些是屬於過程技能的目標以做為評量的依據，再以此目標設計實作評量任務來評量學生。然而，過程技能的目標有哪些？規劃出來的實作任務如何反應過程技能的結構？評量結果如何回饋學生過程技能的狀況？這些具體問題，在教學現場的實作評量歷程中，一直未有效釐清。

九年一貫課程綱要（教育部, 2003）明定了各學習階段的能力指標，自然與生活科

技領域明確列出學生的「科學與科技素養」，其中「過程技能」即為第一項。其下又分有「觀察」、「比較與分類」、「組織與關連」、「歸納、研判與推斷」、「傳達」等五個次項目，並臚列出每個學習階段欲達成的能力指標。這綱要之目的在協助教師清楚地知道學生在不同學習階段應該要具備的能力，好讓教師在設計教材、進行教學及評量活動時，都有清楚的依循。同時，能力指標的出現，讓教師可以清楚地了解學生欲達成的「過程技能」有哪些，各學習階段的連接性與差異性在哪裡。然而，即便有了以上清楚的結構，這些能力指標一旦落實於實作評量上，根據研究者現場實務的經驗，往往在實施的過程時簡化步驟，而在結果的回饋上，再簡化一次能力細目，最後學生以及老師所得到的表現，通常是分數的形式或是實驗結果的檢視，而無法形成教學、學習、與評量的反饋迴路。

關於實作評量的研究成果，自然學習領域中已累積了相當的智慧。陳文典、陳義勳、李虎雄和簡茂發（1995）、李虎雄（1995）與李虎雄和黃長司（1995）與美國馬里蘭州合作，將其實作評量計畫（MSPAP）之評量任務、工具與評分手冊轉換成台灣的經驗。桂怡芬和吳毓瑩（1998）則設計「水溶液的酸鹼性質」與「四輪車與小山坡」的實作評量並探究評分規準的有效性。鄭美雪（2001）的實作評量中，設計出科學概念、技能及態度三方面之任務與規準。吳璧純（2003）所設計的「槓桿原理」實作評量，與洪文東（2004）評測國小學童對「熱的傳遞方式」的思考智能的實作評量設計，也將重點放在實作任務與評分規準上。由這些研究成果來看，關於實作評量的經驗，目前仍聚焦於實作任務與評分規準兩大核心要素。至於如何考量實作評量的品質？是否有一組教學現場可利用的具體原則與標準（後設規準）來協助教師檢核評量的運作？這是研究者從教師角度出發的關懷。

Brown 和 Shavelson（1996）及 Ruis-Primo 和 Shavelson（1996）在討論並發展自然科學領域的實作評量時，曾將其構成的雙要件：實作任務（performance task）和評分系統（scoring system），轉化成三個構成要素：實作任務、反應型式（response format）和評分系統，以利教師更清楚的規劃實作任務，反映教學目標，並且對於所觀測與評量的學生反應或實作行為，能夠更具體地描述。

基於過去實作評量的研究成果、反應型式的理論提示、以及教室現場中實作評量操作的困難，促使研究者進一步反思下述三個問題：

1. 在能力指標明確清晰的背景之下，結合理論的精進，本研究如何借重增加了反應型式以及後設規準的四要素模式來設計實作評量？
2. 實作評量如果要促進教學與學習，實施階段需要如何改進？
3. 在實踐的歷程中能力指標如何發揮指引與回饋的功能？

透過國小六年級的自然學習課程，以及歷經一個學年的規劃、兩個班級預試與正試的實踐，研究者企圖回答上述三個研究問題。

## 貳、從能力指標、評量目標到評量結構

### 一、過程技能要項之能力指標解析

自然與生活科技領域綱要中，「科學與科技素養」共有八個能力要項，每個能力要項下又分出若干個次項目，並根據學生在九年國民義務教育的學習狀況，化分成四個學

習階段。本研究則是針對第三學習階段國小高年級中的第一個能力要項「科學過程技能」，釐清其能力內涵並細膩規劃適合的實作評量型式。

研究者首先以「科學過程技能」能力要項為主，其他七項為輔，來進行能力指標內容上的分析。第三學習階段過程技能能力要項下有五個次項目，各次項目下再細分出具體的能力指標，研究者分析過程技能次項目中所有的能力指標，重新歸納整理，發現這些能力指標內有六種能力內涵。過程技能能力內涵與能力指標代碼如表 1 所示。

經過細部檢核過程技能的能力指標內容之後，研究者發現有些能力內涵可以用「紙筆測驗」的形式來評量，例如：分類的能力、本量與改變量之比較的能力等，而其他能力內涵，包括：設計並進行實驗的能力、報告的能力、整理歸納資料的能力及綜合性的能力等，則可以用口頭報告、資料整理報告、或實驗操作的方式表現出來，換句話說，這些能力適合以「實作評量」的形式來評量學生的能力，提供學生多元的表現機會，以較真實的方式描繪學生的能力。其中「設計並進行實驗的能力」乃為最重要的能力內涵。就複雜性而言，它包含九個能力指標；就評量的多元性來看，它可運用觀察、操作、設計實驗等方式表現出來；就能力的層次性來看，它展現出學生提出假設、設計實驗、進行操作、下結論等不同能力面向，教師從評量中不只了解學生操作的過程，也能了解學生的思考能力，理解過程及歸納能力。鑑於此能力的重要性，研究者以「設計並進行實驗的能力」為核心，並在其他能力要項下，擷取 2-3-1-1、3-3-0-5、5-3-1-3、6-3-3-2 等四項與此能力內容有關的指標，進行合併與分析，轉化為評量目標。

## 二、能力指標轉化為實作評量目標

能力指標只具指標性的功能，必須加以轉化，才能形成較具體的教學評量目標。葉連祺（2002）指出能力指標轉化有以下六種策略：替代（replace）、拆解（decompose）、組合（group）、聚焦（focus）、聯結（relate）、複合（mix）。研究者採用葉連祺的轉化策略，就所選擇出的能力指標進行解析。秦嗣輝（2003）曾用個案研究法探討自然與生活科技領域能力指標的轉化，其研究參與者陳老師使用「替代」、「拆解」和「組合」的轉化策略進行同一主軸的能力指標轉化，達到很好的效果。而研究者在了解了轉化策略的意義與做法之後，採取了「拆解」和「組合」的策略進行能力指標轉化。所謂拆解是使用

表 1：過程技能能力內涵與能力指標代碼表

| 過程技能能力內涵 | 能力指標代碼 |
|---|---|
| 設計並進行實驗的能力 | 1-3-1-1、1-3-2-1、1-3-3-1、1-3-3-2、1-3-4-3、1-3-4-4、1-3-5-1、1-3-5-2、1-3-5-3 |
| 分類的能力 | 1-3-2-3 |
| 本量與改變量之比較的能力 | 1-3-1-3、1-3-2-2 |
| 整理歸納資料的能力 | 1-3-4-1、1-3-4-2、1-3-5-1、1-3-5-2 |
| 綜合性的能力 | 1-3-3-3 |
| 報告的能力 | 1-3-5-4、1-3-5-5 |

註：根據九年一貫課程綱要〈教育部, 2003〉的說明，編號的第一碼為能力要項、第二碼為學習階段、第三碼為次項目、第四碼為流水號，每一組代號即為一個能力指標。

「一對多」對應化關係，將能力指標拆解成幾個互有關連的細項能力指標，以作為教學目標；組合則是運用「多對一」對應轉化關係，以一個主題結合多個能力指標，形成一個課程內容（葉連祺, 2002）。

首先，研究者針對第二個能力要項「科學與技術認知」的指標 2-3-1-1 進行「拆解」，因為這個要項的能力指標比較具體明確，拆解後將其分成「提出假設或提出問題、研商處理問題的策略」、「學習操控變因」、「觀察事象的變化並推測可能的因果關係」、「學習由變量與應變量之間相應的情形，做出合理的解釋」、「學習資料整理、設計表格、圖表來表示資料」等五部分，再與第一個能力要項「過程技能」與第五個能力要項「科學態度」中相關之能力指標進行「組合」。「過程技能」指標的 1-3-5-1 與 1-3-5-2 內涵相同，差別在於 1-3-5-1 可由學生自創表格，而 1-3-5-2 則訓練學生學習符合科學規範的資料之表述方式，只是學生呈現方式的不同，故可視為同一能力指標。第三個能力要項「科學本質」的指標 3-3-0-5 與第六個能力要項「思考智能」的指標 6-3-3-2 因內容內涵相近，採取「組合」策略轉化之，研究者再將這些能力指標轉化為具體的評量目標。其轉化策略、結果及內容以圖 1 表示之。

## 三、反應型式的考量與後設規準之設立

### (一)認知風格與反應型式

由於實作評量的現場通常皆備有各式器具讓學生選擇使用，根據 Lu 和 Suen（1995）實作的歷程將會因受評者的認知風格（cognitive style）而造成不同的結果。認知風格是指學生在面對問題情境時所偏好採取的訊息處理方式，產生在認知活動過程中外顯於行為上的特徵（張春興, 1996, 1997; Mayer, 1987/1991）。其中，場地獨立（field independent）與場地依賴（field dependent）的認知類型，與研究者在考量評量任務的決定有很大的關係。所謂場地獨立與場地依賴，是指一個人能從事某些作業而不受外界訊息干擾的程度（Mayer, 1987/1991）。場地獨立的人，知覺判斷時較少受週遭刺激的干擾，能專注，不易分心，場地依賴的人則反之，知覺判斷不能獨立，易受外圍事物干擾（張春興, 1996, 1997）。就分析訊息與學習成就的關係來看，場地獨立者傾向於以分析的方式來處理訊息，而科學與數學概念的學習大多涉及分析和解析的過程，故在數學和自然科學的學習成就上，場地獨立者的表現顯著優於場地依賴者（楊坤原, 1996; Mayer, 1987/1991），高淑芳與林麗惠（2001）的研究也指出，不論性別與種族，場地獨立的學生的推理表現、問題解決表現，均顯著優於場地依賴的學生。

上述研究發現對於研究者的啟示在於實作評量之設計，尤其需要注意是否提供恰當且不過於複雜的情境，以激發學生表現又不至於干擾學生的思考。研究者分析前人的實作評量結構中，發現實作任務、反應型式和評分系統這三個組成要素裡，反應型式是一般教師及研究者在設計實作評量時較少注意，卻是影響不同認知風格學生之表現的重要因素。所謂反應型式，如果就教師的評量設計而言，指的是教師如何布置整體評量情境，包括任務的給予形式、器材的準備方式，以及教室情境的安排，以適當激勵學生以恰當的方式解題或完成任務（Brown & Shavelson, 1996; Ruis-Primo & Shavelson, 1996）。此部份是場地依賴之學生最容易被複雜訊息干擾而無法正常表現之關鍵因素。本實作評量之理念，便著眼於反應型式的設計必須單純，以排除場地依賴學生受干擾的情況，以協助學生順利完成任務。

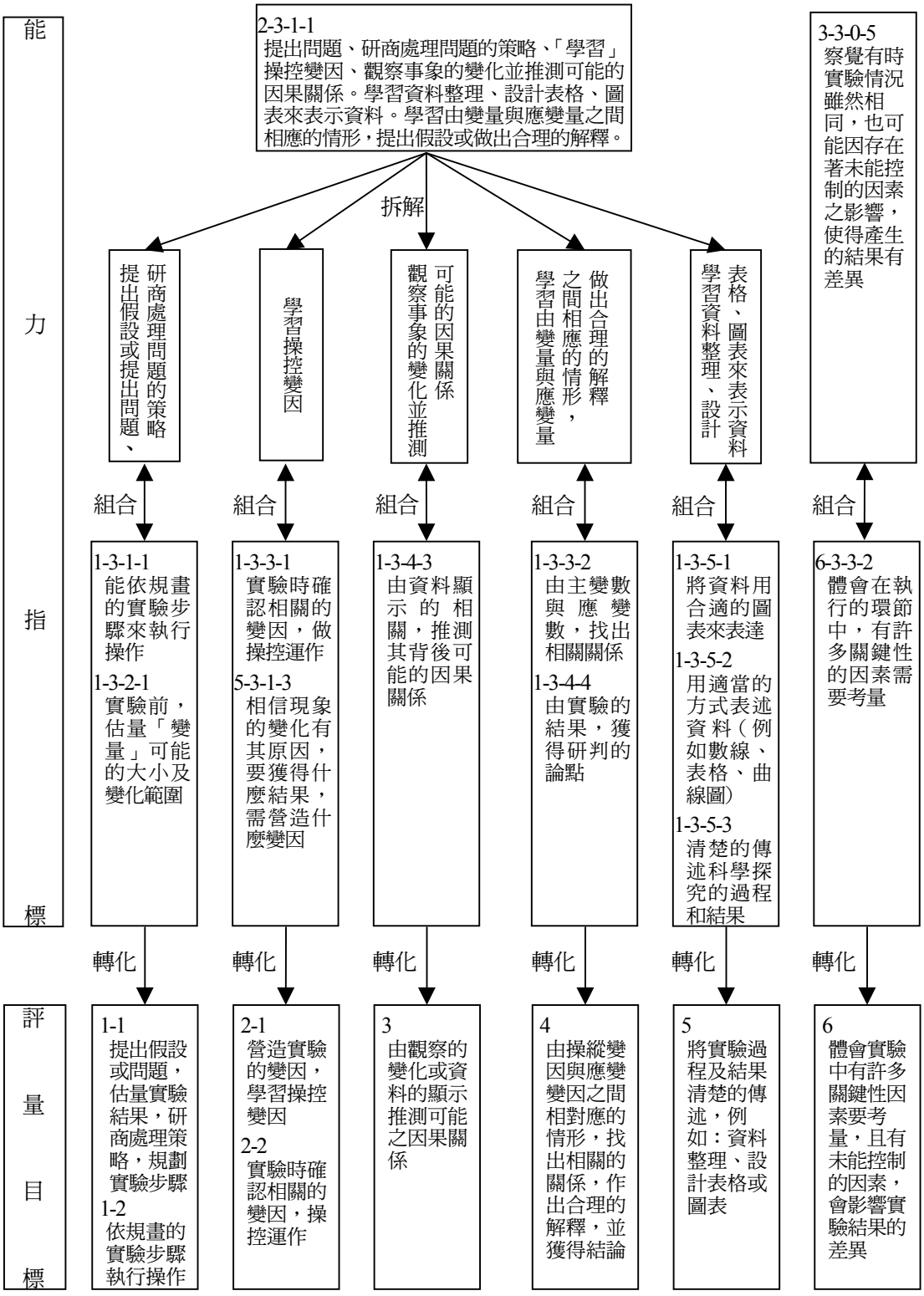



圖 1：能力指標轉化策略與結果

(二)檢視評分規準的後設規準

研究者認為實作評量的規劃與實施，應該有一個後設系統，對於此實作任務進行監控與檢視。Quellmalz（1991）提出好的實作評量應包括重要性（significance）、忠實性（fidelity）、概化程度（generalizability）、發展的適切性（developmental appropriateness）、接近性（accessibility）及實用性（utility）等六個原則。Arter 和 McTighe（2001）提出後設規準（metarubrics）的想法來確定評分規準的適用性，他們認為後設規準就是評鑑規準的規準（a rubrics for rubrics），適合教室層級的評量，不適合大規模的評量，其中有四個檢視的特徵：內容（content）、清晰度（clarity）、實用性（practicality）及技術的正確性（technical soundness）。

本研究對於檢視評分規準的做法是，首先把 Arter 和 McTighe（2001）提出的後設規準四個特徵和 Quellmalz（1991）的六個原則之內涵來進行分析。研究者發現這三位學者所提出的特性或原則，有其意義上相同之處，也有互補之處。在嘗試分析這些相近或互補的內涵後，研究者建構出評分規準的規準，整理結果如表 2 所示。

## 參、研究方法

上述文獻分析在協助研究者確立評量目標以及定稿實作評量的四要素結構：實作任務、反應型式、評分規準及後設規準。本研究之實作任務乃取材自參與研究的班級所使用之教材，目的亦是希望所得結果能夠回饋參與實驗的教師之教學與學生的學習，因而評量活動與任課教師之教學活動可以適切配合。另外，由於此實作評量乃是以自然領域的重要能力為核心，是以所規劃的整體系統，亦可在不同的自然教室中使用。

研究者以服務的學校—嘉嘉國小做為規劃與實施的學校，嘉嘉國小位於台北市的傳統社區內，相較於其他行政區，家長的平均

表 2：實作評量後設規準內涵

| Quellmalz | Arter & McTighe | 本研究的後設規準內涵 |
|---|---|---|
| 重要性：規準項目的選擇與建構，都要以其可代表認知內容和認知技能等重要結果為要 | 內容：規準的內容必須要能真正的區別品質，解釋規準中的每一項目是必要的表現品質的標準 | 內容的重要性：規準必須能區別品質，其內容包括認知表現或技能表現的結果的標準 |
| 忠實性：評分規準必須是可具體觀察、直接判斷的 | | 表現的忠實性：評分規準必須是可具體觀察、直接判斷的 |
| 概化程度：規準要能應用到其他相似的任務上 | | 概化到其他任務的程度：規準要能應用到其他相似的任務上 |
| | 技術的正確性：找出證據相信評分真的代表學生的表現 | 分數來源的正確性：詳細說明學生成就表現的證據 |
| 發展的適切性：規準需要詳細說明學生表現的成就水準 | | 發展的適切性：規準合於六年級學生表現的成就水準 |
| 接近性：使用的語言，應明確地讓教師、學生與家長了解其代表的意義 | 清晰度：規準必須能清楚地提供教師及學生解釋在規準中所陳述及說明的可能 | 結果解釋的清晰性：讓教師、學生與家長清晰地了解其代表的意義 |
| 實用性：要能提供學習表現的品質與改善作法之清楚訊息 | 實用性：給予師生做教學決定的訊息，追蹤學生重要學習結果的過程 | 回饋的實用性：要能提供學習表現的品質與改善作法，給師生回饋 |

社經地位較不高，部分學生的學習資源亦較不足。研究者以所任教的兩個六年級班級為參與研究的學生，快樂班的 26 名學生為預試的對象，並提供預試之後的回饋與改進之建議，活力班的 25 名學生為正試的對象。研究者並邀請兩位同事（皆為自然與科技學習領域科任教師）共同參與實驗，他們各有 8 年與 16 年的自然科專任之經驗。預試階段，兩位同事也以其任教的班級進行教學與評量，以提供研究者進一步修正的建議，同時在正試時，其中一位同事也是實施過程中的共同評分者。

## 一、評量規劃

### (一)確立教學單元

兩個班級所使用的教材是六年級自然與生活科技康軒版本（王美芬等, 2004），其中三個小單元內容是在引發學生的「設計並進行實驗的能力」，分別是：第一單元「物質的變化」中的「影響鐵生鏽的原因」，第二單元「大地的奧秘」中有關「流水流量與土石沖刷的關係」，和第四單元「電磁作用」中關於「影響電磁鐵磁力大小的因素」。研究者從中選擇「影響電磁鐵磁力大小的因素」做為實作評量的觸發媒介，原因如下。

1.就評量結果是否接近學生的能力來看

所選之單元「影響電磁鐵磁力大小的因素」的實驗，除了「操縱變因」和「保持不變的變因」很容易觀察之外，「應變變因」可以以量化的數據做為判斷的依據，學生可以具體的做出判斷。至於未獲選的單元「影響鐵生鏽的原因」之「應變變因」，因為鐵器（或鋼棉）生鏽必須花費至少一天以上的時間做觀察，學生必須將實驗物品帶離教室觀察，這一部分的評量必須依據學生的觀察紀錄做為依據，然而學生是否真實地完成紀錄，或是抄襲修改同學的實驗紀錄，我們無法判斷。同時，學生質性敘述的觀察結果，會對文字表達能力不佳的學生產生不公平的影響。另外，未獲選的「流水流量與土石沖刷的關係」實驗中之「應變變因」的部分，土石被沖刷的情形無法量化比較，根據我們的教學經驗，這項實驗結果並不明顯，觀察不易，學生有時無法做出明確的判斷。

2.就評量的方便性來看

「影響電磁鐵磁力大小的因素」的實驗，其影響之因素有：電池的數量、電池連接方式、線圈數的多寡、線圈數的粗細、電磁鐵鐵心的材質等，在器材準備上比較不像「影響鐵生鏽的原因」的實驗複雜多樣，也不會有「流水流量與土石沖刷的關係」的實驗的場地整理問題。但是，與「影響鐵生鏽的原因」的實驗一樣，在不設限學生的思考判斷之原則下，也可能會發生學生的假設超過我們的預期，評量時沒有器材可以操作的窘況。此部份有賴本研究所發展的二階段實作評量來解決之。

研究者認為和其他兩個實驗相比，「影響電磁鐵磁力大小的因素」的實驗較適合設計評量任務來達成評量目標，並可具體描繪學生的能力。本實作任務雖然取材自上述教材，但是所規劃的任務以及事後的研究結果，乃以能力指標為主，是故，可以施行於使用不同教材的自然教室中，透過「影響電磁鐵磁力大小的因素」實驗的過程技能之實作評量，瞭解學生在評量目標上的達成程度。

### (二)二階段評量活動的規畫

研究者認為如果實作評量只是讓學生知道過程技能優劣的結果，而沒有讓學生在操作過程中學習，那和傳統紙筆測驗相比，只有形式上的差異而已。我們假設如果某位學生在進行實驗過程評量時，其實驗設計與假設無關，或是操縱變因選擇錯誤，或是沒有

想清楚哪些變因要保持不變，那麼，之後的整個實驗過程都是有問題的，就算操作過程正確無誤，其實驗結果也是不正確。因為，他起於一個「錯誤的開始」，迄於一個「錯誤的結束」，從開始到結束，只有實作，沒有顯著的學習。

順著這樣的想法，研究者與現場老師討論之後，規劃出實作評量任務的二階段系統。具體的說，實作評量任務分成「設計實驗」與「進行實驗並下結論」兩階段：在第一階段「設計實驗」過程中，學生根據他的實驗假設思考實驗的三個變因，老師先就學生的假設與實驗變因進行評量，之後再與學生討論，讓學生訂正他的想法；第二階段則是學生根據修改過的想法進行實驗並下結論。二階段的任務規劃可以提供學生評量改進之機會，也可以讓老師根據學生第一階段的實驗假設與設計，準備第二階段相關的實驗器材，避免實驗器材準備不足之窘況。

此設計之精神乃在提供機會讓學生透過教學過程檢討或修改想法。按照這樣的規劃，依據學生表現的情形，第一階段實作評量的結果，可以分成「實驗設計正確」、「實驗設計錯誤修正後正確」及「實驗設計與修正均錯誤」等三個能力層次。而學生第二階段的表現，可依實驗操作的狀況，根據「操縱變因」、「保持不變的變因」與「應變變因」的正確性，分成八類。總和起來，學生設計實驗與操作實驗的表現可有二十四類結果（如圖 2）。這樣的分類結果相當細膩，可以讓教師清楚知道學生對實作評量任務的反應模式，亦可以讓家長以及學生自我瞭解自己的能力狀況。

### (三)實作任務的具體內容

我們考慮到認知風格對學生的影響，以及實作評量反應型式的重要，將任務內容以簡單明瞭的方式呈現，以避免場地依賴學生受到與過程技能無關的其他變項所干擾，同時提醒學生評量的方向，不要讓學生思考時偏離評量的主題，讓場地依賴型的學生不致錯亂。研究者利用教科書中「科學家小傳」的內容，編寫出適切的問題情境，內容如下：

問題情境：丹麥科學家奧斯特把導線接到電池上，意外發現會產生磁力影響指北針，之後，法國科學家安培受到奧斯特的影響，深入研究，把電線繞幾圈之後通電，磁力比較會影響指北針的偏轉，並發明了電磁鐵。安培後來又想到，如果要讓電磁鐵磁力更強，要怎麼做才好呢？聰明的你，請來想辦法解決吧！

### (四)反應型式的設計

學生在進行評量時，教師必須安排適當的評量環境，也就是 Brown 和 Shavelson（1996）及 Ruis-Primo 和 Shavelson（1996）提到的反應型式。本研究之規劃如下：

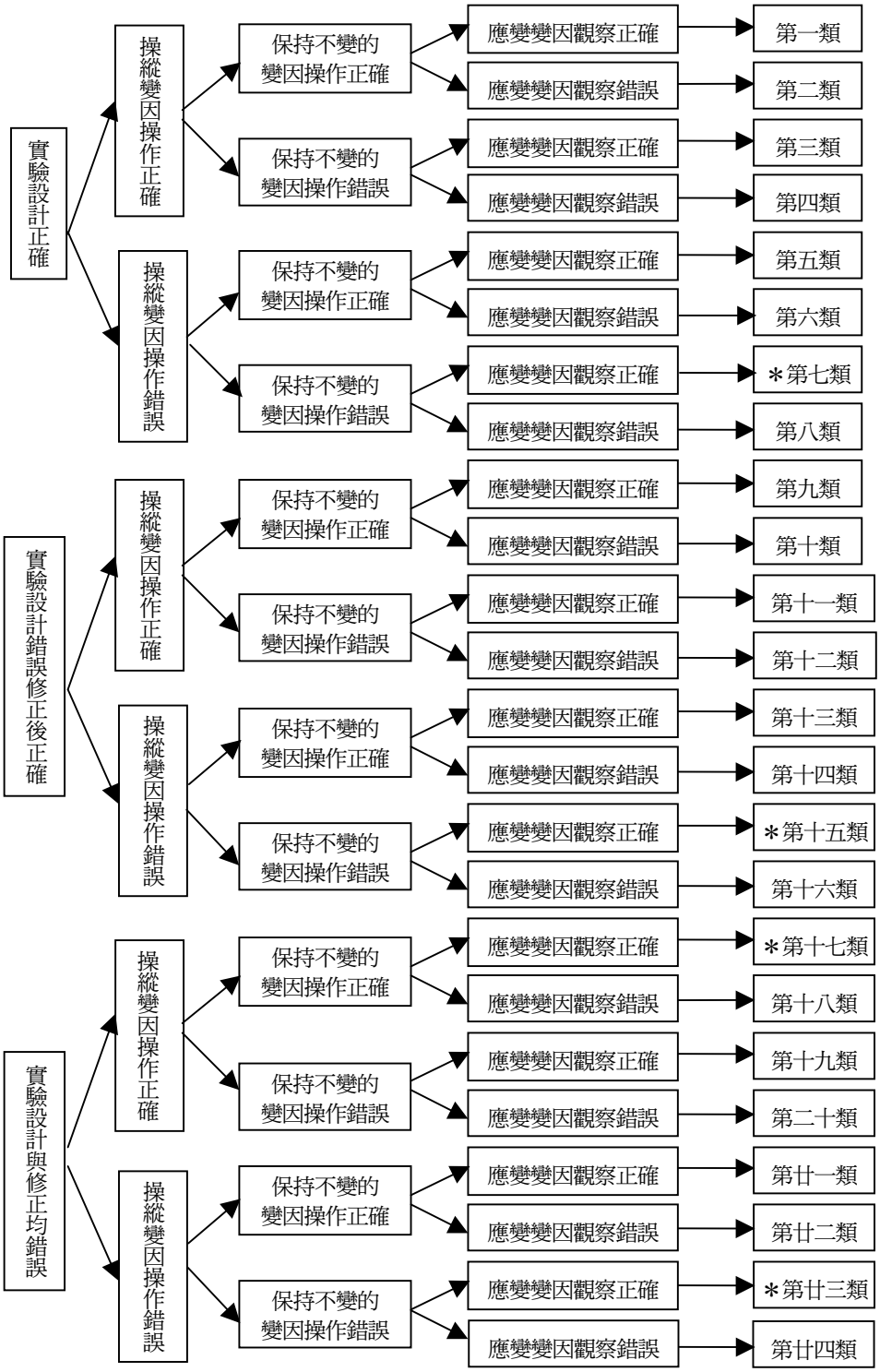
1.場地與時間

第一階段的評量活動主要是讓學生設計實驗並討論訂正修改實驗，以紙筆型式及全班性活動為主，時間為兩節，進行場地在原先上課的自然教室即可。

第二階段的評量活動就是學生操作能力的評量，因為協助進行的老師都是資深的自然科老師，同時觀察多位學生進行實驗的表現，皆相當順手。研究者規劃一次同時評量四位學生，近 30 位孩子全部進行完畢，大約需要兩節課的時間。兩個階段共需要約 4 節課的時間來完成。

2.評量教室的情境佈置

第一階段的評量實施之教室，以一般上課的教室即可。第二階段的評量教室，則在空間上、實驗桌的擺設及教室內的陳設上，都與學生的自然教室一樣，或是利用自然教室，以讓學生有熟悉的感覺。同時，安排未輪到評量的學生觀賞自然教學影片或是閱讀書籍，或是準備尚未備妥的實驗器材，不致




註：＊表示此類別為不合邏輯的表現。

圖 2：設計與進行實驗之過程技能實作評量可能表現結果之分類圖

於影響正在進行評量的學生。第二階段評量的情境佈置則是一位學生使用一張桌子，老師於四張評量桌之間走動，可以就近評量並進行觀察，對學生操作若有質疑，也可以直接對學生提出疑問，請學生說明解釋。同時，因為評量桌彼此是分開的，受評的學生也不會彼此干擾。評量教室的配置圖如圖 3 所示。

3.器材的準備

教師部份：學生實驗操作時的器材，與學生所擬定的假設和選定的變因有關。教師在學生第一階段修正實驗設計之後，可以清楚了解學生在實驗進行時所需要的器材。教師在第二階段評量前依照所有學生的實驗設計事先準備好器材。核心要項是共用的器材四份，放在四個籃子裡，再將籃子分別放置於四張評量桌上，其中包括鐵棒、鋁棒、木棒各三支，迴紋針 100 個，磁鐵一個（分辨鐵、鋁棒用），電池槽兩個。教師所準備的器材必須事先告知學生，以免學生重覆準備。

學生部份：學生必須依自己的假設纏繞漆包線於不同材料的棒子中，再將棒子取出，或是纏繞不同的圈數，及將漆包線與電線接觸之處刮乾淨，這些工作事先即已準備好。學生在尚未進入第二階段評量時，可以在教室內進行檢查或是修整。

**(五)建構評分規準**

1.評分規準的層次

研究者檢視之前轉化出的實作評量目標，從中分析出評分規準的內涵，並分類歸納出評分規準（scoring rubrics）的第一層項目，也就是實作評量的效標（criteria），再細分出規準之第二層子項目，稱為表現指標（indicators），第三層則規劃具體的評分項目，第四層則為各評分項目之下的表現水準與說明。由於評分規準乃是從評量目標出發，所以表現結果可以回頭檢視評量目標，並更進一步檢核能力指標的達成。評分規準之建構流程如圖 4。

從文獻中研究者發現吳欣黛（1998）的研究以「設計實驗，做結論」為構念中心，進行實作評量在效度上的真實性與直接性的探討，她利用「四輪車與小山坡」和「毛細現象」兩個評量任務為範圍，對學生進行實

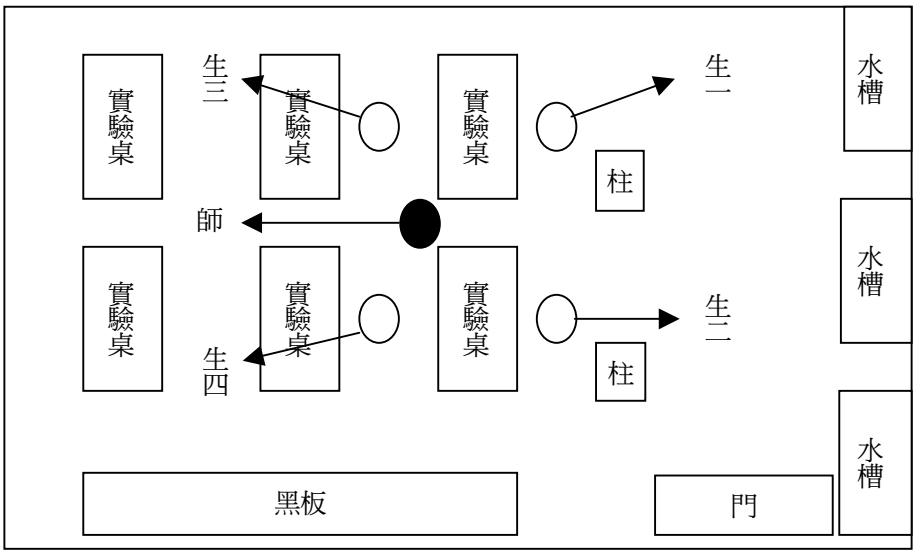



圖 3：評量教室配置圖

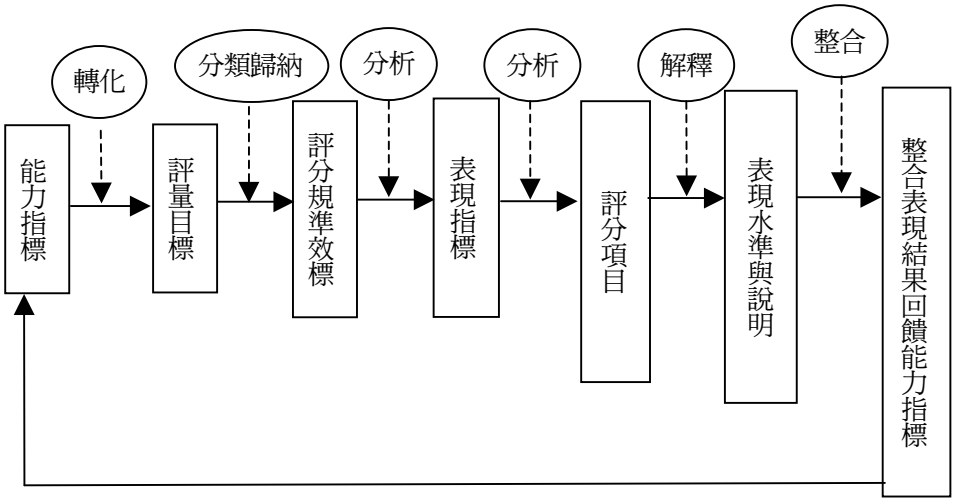



圖 4：評分規準建構流程圖

作評量，並進行研究分析，而其評分規準已建構出清楚的評量內容定義及量尺分數。本研究之過程技能「設計實驗、進行實驗、下結論」與其構念中心的內涵類似，因此亦為本研究規劃評分規準時的參考。

我們將圖 1 從能力指標轉化後的評量目標進一步整合後，歸納出四個評分規準的效標，分別是：(1)設計實驗的能力；(2)進行實驗的能力；(3)記錄實驗結果的能力；(4)驗證假設下結論的能力。效標一「設計實驗的能力」可以分出「擬定問題」、「提出實驗假設」及「設計實驗變因」等三項表現指標；效標二「進行實驗的能力」可以細分出「根據假設操縱變因」、「在實驗中控制變因」及「觀察應變變因」等三項表現指標。再根據這些表現指標分析出更明確的評分項目，例如：表現指標「設計實驗變因」可以分析出「選定操縱的變因」、「選定應變變因」和「界定保持不變的變因」等三個評分項目。整體之關係架構如圖 5 所示。

2.後設規準檢視結果

後設規準是指評分規準的規準（Arter & McTighe, 2001），本研究以評分規準的效標為主，根據表 2 的後設規準的七面向：內容的重要性、表現的忠實性、概化到其他任務的程度、分數來源的正確性、發展的適切性、結果解釋的清晰性、回饋的實用性等七項內涵，設計出二分法實作評量後設規準表如表 3 所示，以「達到」或「待修改」兩個層次來進行實作評量整體性的品質檢視。結果如表 3。

3.預試之後的修正

後設規準檢視的結果顯示在實際實施評量之後，發現有許多表現程度敘述不完整之處，教師在觀察時，無法正確的勾選適合學生表現的規準，表示所設計的規準不符合學生發展的適切性，亦無法協助判斷結果解釋的清晰性，此處仍須修改。具體而言，協助的評分者及學生認為需要修正的規準有：

(1)將效標二的部分規準更改為比較通用的敘述，關於實驗中變因的部分，我們將評分項目均改為控制變因操縱能力的品質，給予評分者使用時的彈性。另外，我們亦修改控制變因的最後一個評分項目為：作品的完成情形之程度。並將指標觀察「應變變因」

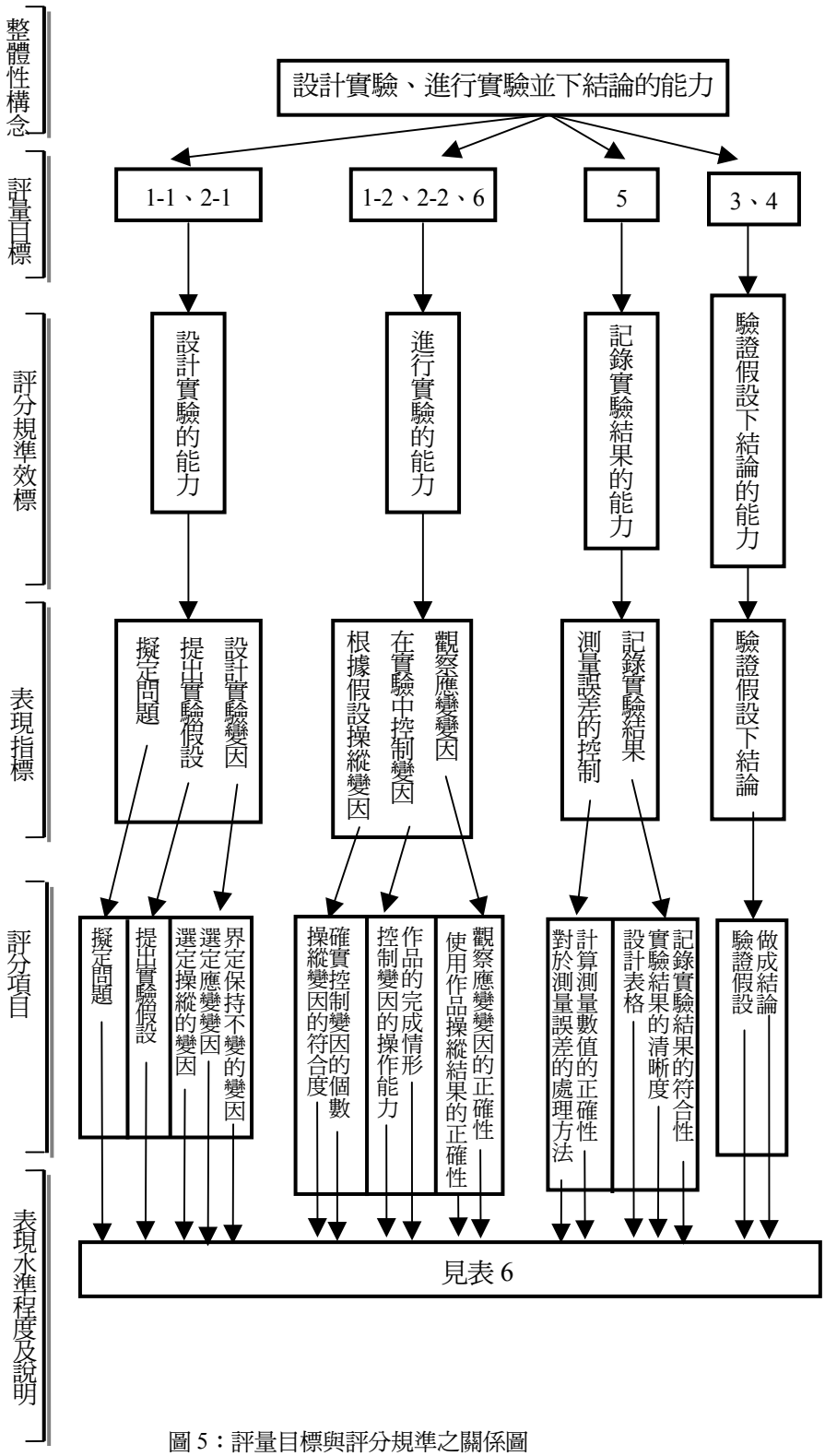



圖 5：評量目標與評分規準之關係圖

表 3：實作評量後設規準表及本研究所規劃的實作評量預試檢視結果

| 後設規準內涵 / 效標 | 內容的重要性 | 表現的忠實性 | 概化到其他任務的程度 | 分數來源的正確性 | 發展的適切性 | 結果解釋的清晰性 | 回饋的實用性 |
|---|---|---|---|---|---|---|---|
| 設計實驗的能力 | 達到 | 達到 | 非本研究討論的問題 | 達到 | 達到 | 達到 | 達到 |
| 進行實驗的能力 | 達到 | 達到 | | | 待修改 | 待修改 | |
| 記錄實驗結果的能力 | 達到 | 達到 | | | 待修改 | 待修改 | |
| 驗證假設下結論的能力 | 達到 | 達到 | | | 達到 | 達到 | |

中的「使用作品操作結果的正確性」之的評分項目改為五等第的正確程度。

(2)效標三之下「記錄實驗結果的能力」之重點在於學生對於實驗結果的「紀錄」本身是否符合操作結果，並不論此結果是否能正確地呼應實驗假設。如果紀錄與操作符合，則表示效標三已達到，至於此結果是否正確，則是效標四「驗證假設下結論的能力」之評分項目所應照顧的。是以，效標三的評分項目「記錄實驗結果的正確性」改為「記錄實驗結果的符合性」。

(3)效標三之下關於平均值的計算，此處我們並不檢核學生得到的平均值是否影響結果的正確性，只要判斷學生平均值之計算所得數據是否正確即可，至於學生得到的平均值如果有誤，將會影響下結論的正確性，這一部分在效標四中評量。所以此評分項目修正為計算測量數值的正確性。

4.評分規準定案

定稿後的評分規準之結構與內容見圖 5。研究者考慮這份評分規準是以學生的實驗操作過程技能為主，所以各項能力的配分比例，如表 4 所示。

此配分原則主要的目的是要反映能力的結構，在四個能力效標中，佔分最高的是效標二「進行實驗的能力」(40 分)，因其為過程技能實作評量的重要能力。而佔分較低的效標—「設計實驗的能力」與效標三「記錄實驗結果的能力」是主要能力的輔助面向(各 25 分)。至於「驗證假設下結論的能力」的判斷是根據實驗結果而來，佔分最低 (10 分)。效標與能力指標的對應情形如表 5 所示。

評分規準之表現指標、評分項目、表現水準與說明之最後定案如附表 1 所示。評分規準中每個評分項目的表現水準皆為五分量尺，總分為 100 分。學生在每個評分項目的能力達成程度，都可以依據水準分數與各水

表 5：學生效標表現與能力指標對應表

| 效標 | 能力指標 |
|---|---|
| 一 | 1-3-1-1、1-3-2-1、1-3-3-1、2-3-1-1、5-3-1-3 |
| 二 | 1-3-1-1、1-3-2-1、1-3-3-1、2-3-1-1、3-3-0-5、5-3-1-3、6-3-3-2 |
| 三 | 1-3-5-1、1-3-5-2、1-3-5-3 |
| 四 | 1-3-3-2、1-3-4-3、1-3-4-4 |

表 4：各項能力配分表

| 效標 | 一、設計實驗的能力 | 二、進行實驗的能力 | 三、記錄實驗結果的能力 | 四、驗證假設下結論的能力 |
|---|---|---|---|---|
| 配分 | 25 分 | 40 分 | 25 分 | 10 分 |

準的內容說明中，清楚的表示出來，除了方便教師在觀察評量時容易畫記之外，學生也可以明確地了解自己在每個評分項目上的表現情形。另外，每一個效標都有一個小計，能清楚地回饋能力結構的面向。為了方便分析學生在效標上的表現結果，研究者特別設計整體能力側面圖，呈現於研究結果一節中。

## 二、正式實作評量實施流程

### (一)實驗記錄表與評量表呈現的型式

研究者將學生的記錄表與教師的評分表合併，共有六頁，放在三張橫置的 B4 大小紙張內，以每一張兩頁的型式呈現。第一張左半邊為第一頁，是學生第一階段「設計實驗」的記錄表格，包括問題情境、探討問題、實驗假設、三項變因，第一張右半邊為第二頁，是「設計實驗的能力」的教師評分規準表。第二張左半邊為第三頁，是第二階段學生的「實驗記錄與下結論」的記錄表格，內容包括實驗結果記錄與結論，第二張右半邊為第四頁，是「記錄實驗結果的能力」與「驗證假設下結論的能力」的教師評分規準表。第三張左半邊為第五頁，是「進行實驗的能力」的教師評分規準表，第三張右半邊為第六頁，呈現學生表現的整體結果與評價。研究者特別將「進行實驗的能力」的教師評分表與學生整體表現評價放在第三張，是因為這部分的畫記只與教師有關，在讓學生了解評分規準之後，方便教師抽離出表格進行評分。

### (二)實施流程

進行第一階段評量活動時，先讓學生閱讀問題情境同時並由教師敘述問題情境給學生聽，以避免閱讀能力較差的學生不了解問題。之後，學生開始設計實驗，包括擬定問題、提出假設及選定變因，並依照變因設計實驗結果記錄表，將思考結果寫在實作評量實驗紀錄表的第一頁及第三頁上。之後，教師進行第一項能力 –「設計實驗」之批改，直接利用學生所完成的記錄表批改勾選於第一章右半邊第二頁的評分表上。教師批改之後，再利用下一節課的時間與全班學生討論實驗設計正確或錯誤之處，並鼓勵學生進行實驗設計的修正，之後進入評量的第二階段。

第二階段評量活動中，學生進行實驗、記錄實驗結果、並根據結果來下結論。評量教室中，一次容許四位學生在各自的桌上進行實驗，學生操作完畢時，將結果記錄在第二張左半邊第三頁的記錄表上，實驗各自結束後，即分別去等待教室請下一個同學進來。教師於學生操作時，進行第二項能力 –「進行實驗」的評分，並將學生的操作情形，勾選於第三張左半邊第五頁的評分表中。全部學生操作完畢大約需要兩節課的時間。之後，教師依據學生的紀錄進行第三項能力 –「記錄實驗結果」及第四項能力 –「驗證假設下結論」的評分，並將結果勾選於第二張第四頁的評分表中。

最後教師在第六頁統計各項能力的分數，及整體表現的分數，並分析學生的得分百分位置及表現側面圖，讓家長、學生及教師了解所得分數所表示的程度。詳細的實作評量實施程序及內容如表 6。

# 肆、研究結果

## 一、實作評分的一致性

研究者與一位夥伴教師為正式施測時的的共同評分者。根據評分規準表，研究者首先蒐集到兩位評分者對於學生在各評分項目上的表現類別，此為類別量尺，在探究兩位評分者的一致性上，採用 Kappa 一致性係數。

表 6：二階段實作評量實施程序及內容

| 任務規劃 / 內容 / 對象 | 第一階段 | | 第二階段 |
|---|---|---|---|
| 學生 | 設計實驗，並記錄在實驗記錄評分表的第一、三頁 | 進行實驗設計的修改 | 進行實驗，將記錄實驗結果與結論寫於實驗記錄評分表第三頁 |
| 教師 | 根據學生的記錄，評分於記錄評分表的第二頁 | 與學生討論修改實驗設計 | 根據學生的操作情形，進行評分於實驗記錄評分表的第五頁，評量後，依學生第三頁的記錄進行評分於實驗記錄評分表的第四頁，最後，統計學生成績，將整體表現結果記錄於第六頁。 |
| 實際節數 | 第一節 | 第二節 | 第三、四節 |

另外，我們亦將兩位評分者對於每個學生在評分項目下的表現，依各效標分類之後加總，得到各效標的分數以及整體分數，此可視為等距量尺，再探究兩個評量者間分數的相關上，進行 Pearson's 相關計算。最後，我們再將兩位評分者之評分結果視為一個測驗上的兩個題目，亦即此測驗由兩個題目所組成，而將評量表上的各項評分項目之結果在資料分析時，視為受評者在這兩個題目上的反應，如此則可計算兩個題目（亦即兩位評分者）之間的內部一致性 Cronbach's α 係數，等同在討論兩個評分者所評結果之間的概化程度（Shavelson & Webb, 1991）。

Kappa 一致性係數的 κ 值，Pearson 積差相關的 r 值，及 Cronbach 係數的 α 值之結果列於表 7。

Pearson 的積差相關係數 r 值與評分者間概化程度之 α 值都相當高，表示兩位老師評分結果，在各效標與整體表現上，具有高度的相關與一致性。Kappa's κ 一致性係數的結果顯示，在設計實驗的能力、進行實驗的能力、記錄實驗結果的能力及實作評量整體表現得的 z 值，均達到 .01 的顯著水準，表示兩位老師評分結果，在這些能力的評分上具有相當的一致性。但是，對於驗證假設下結論的能力，所得 z 值 = -0.025，未達顯著水準，表示兩位評分者的評量結果，如以最保守的類別量尺視之，由於評分者所勾選

表 7：各項能力效標的一致性統計係數結果

| 係數 / 效標 | Kappa's κ 值（z 值）$z_{99} = 2.33$ | Pearson's r 值 | Cronbach's α 值 |
|---|---|---|---|
| 設計實驗的能力 | .575**（5.016**） | .932** | .964** |
| 進行實驗的能力 | .424**（2.366**） | .911** | .911** |
| 記錄實驗結果的能力 | .538**（3.341**） | .864** | .864** |
| 驗證假設下結論的能力 | -.025（-0.178） | .833** | .833** |
| 實作評量整體表現 | .706**（5.517**） | .960** | .960** |

的類別不同，在「驗證假設下結論的能力」評分結果上呈現評分者間不符合的狀況。造成此現象之緣由如下：

在「驗證假設下結論」之效標中，兩位評分者之相關及一致性都有 .7 的水準，但在類別的勾選上，κ 值呈現評分者不符合之結果。在與合作教師討論之後，研究者發現這是因為兩位評分者本身對學生表現的認定有嚴格寬鬆上的不同，作為評分者之一的研究者因考量地區社經關係，一般給予學生較高的等級，而另一評分者考慮能力品質關係，給予較低的等第，是以在絕對類別符合程度的認定上，κ 值較低，未得到顯著的一致結果。但若看兩位評分者對於全體學生的評分結果，則學生表現之相對地位相當一致，亦即研究者所認為的好表現與另一評分者所認為的好表現，看法雷同，僅是研究者傾向給予較寬鬆較高的分數，而另一評分者傾向給予較嚴格較低的分數。此效標來自兩位評分者的結果顯示，在相關程度、以及評分者間的概化程度上，相當一致，但在絕對類別的相同上，則出現不符合狀況。此現象亦顯示 Pearson's r 以及 Cronbach's α 值適宜討論一致性，而 Kappa's κ 可看出評分者的價值立場。

## 二、全體學生表現的分析

全體學生的表現結果，依據圖 2 的分類方式，正試後的活力班學生的分類結果顯示，學生的表現總共出現十一類，未被歸類到的類別，是沒有這樣的表現情形。實驗設計與操作完全正確的第一類，共有兩位學生。實驗設計錯誤修正後正確，且操作完全正確的第九類學生最多，共有 7 位，表示經由二階段實作評量活動的規劃，多數學生均能透過再次修正實驗設計的機會，在實際操作時做出正確的表現。至於不合邏輯的類別有兩類學生的操縱變因與保持不變的變因操作錯誤，但是應變變因卻是正確，分別是第十五類和第廿三類的兩位。而第十七類的這一位學生，其實驗設計與修改均錯誤，但是其操作的過程均正確。以下分別探討三類學生的狀況：

(一)第十七類學生的實驗設計得分為 19 分，屬於第三級「還不錯」，雖然他的修正仍然錯誤，其錯誤在於假設與應變變因的敘述不夠清楚明確，其餘均正確。這位學生對實驗設計的理解度是足夠的，也就是說，他非常清楚實驗的假設、三項變因以及如何操作，但因為文句的敘述，使得設計部份未得高分，但是根據他對實驗的理解，實驗進行時，整個操作過程完全正確，此類學生需加強文句敘述的學習，提升語文表現的能力。

(二)第十五類學生之操縱變因與保持不變的變因有錯誤，但應變變因觀察正確的情形，這確實有些奇怪。這位學生操作實驗的能力得到滿分 40 分中的 34 分，屬於第二級的「很厲害」，她在進行操縱變因時，其界定的變因是漆包線的線圈數，但由於其纏繞的圈數不夠正確，不是如其原先所設計的 10 圈、20 圈、30 圈，而略有出入纏繞成 15 圈、19 圈、30 圈，在操作時被評分者觀察出來，而未勾選該評分項目的滿分，但這樣的問題，並不會影響實驗時，應變變因結果的觀察。而在操作保持不變的變因時，電磁鐵棒子的操縱能力的評分項目，一開始並未將棒子放入吸管中，電磁鐵並未產生作用，但因隨即修復，所以並不影響應變變因正確的結果。

(三)第廿三類的學生，其操作實驗的能力

得到第一級「非常棒」的 38 分，而其操縱變因的操作問題，也如同前一位學生的情形類似，只是纏繞漆包線圈數的問題，不影響應變變因結果的觀察。至於保持不變的變因，由於連接有一些瑕疵，但仍形成通路，所以並不影響應變變因正確的結果。這些學生的表現，只是操作的細緻性不夠，在經過第一階段評量活動之實驗設計的修正後，其實驗操作的能力仍需提升。

## 三、個別學生的表現側面圖

實作評量全部結束之後，學生的表現結果可以從百分比折線圖中一目了然。如果將這些描述上推回能力指標，我們亦可了解從能力指標的角度而言，學生的學習狀況如何。

根據評分規準表中整體評價的表現折線圖來分析，如以第廿四類學生（見圖 2 的分類）為例，他的表現側面圖見圖 6。教師可以清楚看到學生在四個能力效標中的表現情形，在設計實驗的表現上，這位學生的表現偏弱 44%，但是經過第一階段評量活動中的實驗設計討論後，修正其實驗之設計，其進行實驗的能力表現則稍高，為 55%，記錄實驗結果的能力還不錯，得分 68%，但下結論的能力則較低，得分 40%，顯然這位學生的邏輯思辨與抽象思考能力比較欠缺。

以第十一類學生為例，他的表現側面圖見圖 7。從圖中我們亦可發現學生的表現反映在效標上的情形。這位學生的設計實驗的能力很差，只有 20%，但經過第一階段活動之實驗設計討論之後，其第二階段的進行實驗之能力相當不錯，為 88%，而記錄實驗結果的能力相較而言稍弱 60%，但根據結果下結論的能力也有 80%。表示第一階段的實驗設計討論對他幫助很大，惟記錄實驗結果的能力尚須改進。

第十六類的學生之表現側面圖見圖 8。在設計實驗的得分上，她得到 60%，經過第

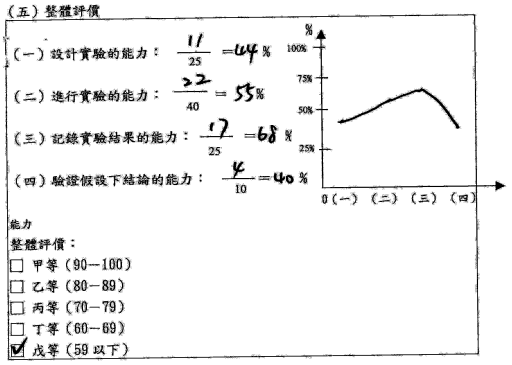



圖 6：第二十四類學生表現側面圖

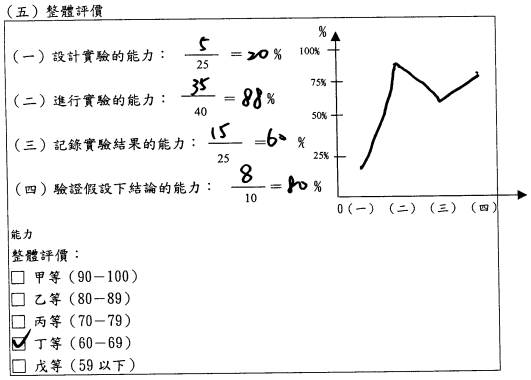



圖 7：第十一類學生表現側面圖

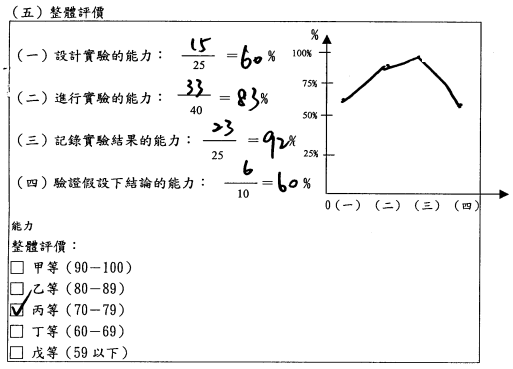



圖 8：第十六類學生表現側面圖

一階段討論修正後，實驗進行的能力得分 83%，記錄結果的能力高達 92%，從側面圖顯示他的實驗操作能力不錯，但是根據實驗結果下結論的能力較弱，表示這位學生下結論的能力需要加強。

### 四、實作評量對於整體教學與學習效益的回饋

教師可依據個別學生的表現結果，回歸到能力指標，藉以了解學生能力指標的達成情形。至於全班的教學成果，本研究列出效標與能力指標的對應如表 5 所示，研究者可以以學生的表現程度，反推至能力指標的達成情形。根據表 8 的結果及評量表中的小計分類，活力班 25 位學生中各效標的達成狀況如下：

效標一：有三位學生設計實驗正確，有 18 位學生修正後正確，總計有 21 位學生完全達到。

效標二：實驗完全操作正確的有 10 位，僅有 3 位學生的三項變因操作是完全錯誤的，其餘 12 位學生的操作則略有錯誤，學生的操作情形得分達到 80% 以上的有 20 位。

效標三：有 22 位學生的能力得分達到 80% 以上。

效標四：有 21 位學生的能力得分達到 80% 以上。

透過上述結果與表 5 的對照，加上表 8 的結果分類，以及學生表現的側面圖，研究者可以清楚了解整體學生及個別學生在能力指標上的達成情形。就活力班而言，教師之整體教學是成功的，在全班的教學上，教師進一步可強調實驗操作的精確性，加強學生在進行實驗時「小心求證」的態度，以避免操作時非必要之錯誤；而由個別學生的表現來看，教師可加強設計實驗時變因的選擇與邏輯判斷下結論等能力的指導。

## 伍、結論

### 一、二階段四要素的實作評量之規劃與實施結果

研究者經歷了一個學年的規劃與實踐之

表 8：設計與進行實驗之過程技能實作評量之結果分類表

| 類別 | 表現狀況 | 人數 |
|---|---|---|
| 第一類 | 實驗設計與操作均正確 | 2 |
| 第三類 | 實驗設計正確，保持不變的變因操作錯誤，其餘變因正確 | 1 |
| 第九類 | 實驗設計錯誤修正後正確，實驗操作均正確 | 7 |
| 第十類 | 實驗設計錯誤修正後正確，操縱與保持不變操作變因正確，應變變因觀察錯誤 | 1 |
| 第十一類 | 實驗設計錯誤修正後正確，保持不變的變因操作錯誤，其餘變因正確 | 2 |
| 第十二類 | 實驗設計錯誤修改後正確，操縱變因操作正確，但保持不變與應變變因錯誤 | 6 |
| ＊第十五類 | 實驗設計錯誤修改後正確，應變變因觀察正確，但操縱與保持不變的變因操作錯誤 | 1 |
| 第十六類 | 實驗設計錯誤修改後正確，實驗操作都錯誤 | 1 |
| ＊第十七類 | 實驗設計與修改均錯誤，實驗操作都正確 | 1 |
| ＊第廿三類 | 實驗設計與修改均錯誤，應變變因觀察正確，但操縱與保持不變的變因操作錯誤 | 1 |
| 第廿四類 | 實驗設計與修改均錯誤，實驗操作錯誤 | 2 |
| 合計 | | 25 |

＊表示為不合邏輯之表現

後，整理出可實際於教室中進行的過程技能之二階段四要素實作評量。一般教師在實施實作評量前，多以評量任務為主要考量因素，認為設計出清楚的評量任務便可進行實作評量。研究者發現在實施實作評量時，評量任務、反應型式、評分規準、與後設規準等四要素必須相互支持，同時配合二階段的分段實施，足以形成適切的實作評量。

二階段四要素實作評量的規劃，非常有利於進行「設計實驗、進行實驗、記錄結果及下結論」的評量，對教師而言，在一個完全開放學生思考的情境之下，於學生第一階段實驗設計結束後，可以從容地準備第二階段評量時的器材，也能清楚了解學生的設計，方便第二階段操作時的觀察。而就學生來講，這樣的方式不但提供學生「在評量中學習」的機會，也正向鼓勵學生動手做實驗的成功機率。

關於四要素如何相互支持，才能發揮回饋的機制，具體明確的產出改進之價值，經過本研究的實施與沈澱之後，研究者認為首先可從能力指標出發，藉由轉化以及合成與分析的步驟，形成評量目標。接著，從評量目標來確立評量單元，並規劃實作評量任務。教師在設計評量任務時，反應型式的考量亦為重要，最好規劃一個學生熟悉且訊息不至於太紛亂的環境，以免過多的線索或不熟悉的環境與學生的認知風格不符，影響思考、判斷、與實作。教師可自編、選編或選用評分規準，建構明確清楚的評分規準，讓教師在觀察學生實驗操作時有具體的依循。最後，利用後設規準來檢視評分規準，以確保整體實作評量適用的程度。同時，加入教師及學生的想法，做為改進規準之依據，以確保其品質。

## 二、透過能力指標轉化成二階段四要素實作評量之效益

從能力指標的達成情形來看，目前學生的發展狀況，操作表現的能力都很好，但設計與推理的高層次思考能力較欠缺，部分學生的設計實驗能力與根據結果推論並下結論的能力需要加強，教師在教學上需要強化歸納結果的邏輯思辨能力。當學生在進行有趣的操作實驗活動之後，更應訓練學生歸納推論的能力，否則，一場教學活動下來，學生只有在興趣與能力均高的操作活動中獲得最多，卻無法突破比較欠缺的能力培養。

對於能力結果的描述，由於本研究之實作評量有清楚的效標，是以師生都能清楚了解能力的達成程度。評分規準表中整體評價的設計，可以從學生的得分百分比及側面圖中，了解學生的表現情形，同時配合設計與進行實驗之過程技能實作評量可能表現結果之分類圖，據以回饋學生的學習情況以及教師在進行教學的改進。我們將實作評量活化，讓實作評量與教學活動充分配合，同時提供學生學習改進的證據與機會。有一位學生的家長所給予研究者這樣的回饋：「老師讓每一位學生事先知道評分標準，可以讓學生有目標去爭取分數，提高表現品質，從百分比及折線圖可以清楚孩子的能力表現，是個很難得的學習經驗。」回答了本探究之所以進行的首要動機。透過本研究所規劃的方式，冀能提供自然與生活科技領域之教師採取此模組進行實作評量，將學生的表現回饋到能力指標上，同時也讓教師透過學生的表現而知曉教學之改進。

## 參考文獻

1. 王美芬等（2004）：國小自然與生活科技六上課本。台北市：康軒。
2. 朱敬先（1993）：教學心理學。台北市：五南。
3. 吳欣黛（1998）：實作評量在效度的上的真

**實性與直接性**。台北市：國立台北師範學院國民教育研究所碩士論文（未出版）。
4. 吳璧純（2003）：以自然科實作評量探究學生的認知發展與學習成效。**教育與心理研究**, **26**, 1-22。
5. 李虎雄（1995）：國民教育階段學生基本學習成就評量研究－我們為什麼參加美國馬里蘭州學校實作評量國際共同研究。**臺灣教育**, **538**, 47-51。
6. 李虎雄和黃長司（1995）：美國馬里蘭州學校實作評量工具在臺灣施測的可行性。**科學教育月刊**, **179**, 41-49。
7. 洪文東（2004）：**國小熱的傳遞方式教學活動設計與學童思考智能之實作評量**。論文發表於九十三學年度師範院校教育學術論文發表會，899-931。屏東市：國立屏東師範學院。
8. 桂怡芬和吳毓瑩（1998）：自然科實作評量的效度探討。**中國測驗學會測驗年刊**, **45**(2), 19-36。
9. 秦嗣輝（2003）：**九年一貫課程自然與生活科技學習領域能力指標轉化之個案研究**。台北市：國立台北師範學院課程與教學研究所碩士論文（未出版）。
10. 高淑芳和林麗惠（2001）：國小學童的認知風格差異及其推理與問題解決表現之關係研究。**國立新竹師院學報**, **14**, 87-126。
11. 張春興（1996）：**現代心理學**。台北市：東華。
12. 張春興（1997）：**教育心理學 — 三化取向的理論與實踐**。台北市：東華。
13. 教育部（2003）：**國民中小學九年一貫課程綱要**。台北市：教育部。
14. 陳文典、陳義勳、李虎雄和簡茂發（1995）：由馬里蘭州的學習成就評量與其在臺灣的試測結果看－實作評量的功能與運用。**科學教育月刊**, **185**, 2-11。
15. 楊坤原（1996）：認知風格與科學學習成就的關係。**科學教育月刊**, **195**, 1-23。
16. 葉連祺（2002）：九年一貫課程與基本能力轉化。**教育研究月刊**, **96**, 60-61。
17. 鄭美雪（2001）：從美國馬里蘭州學生基本學習成就評量研究談國小自然科多元教學與評量。**現代教育論壇**, 79-101。
18. Arter, J., & McTighe, J. (2001). Scoring rubrics in the classroom-using performance criteria for assessing and improving student performance. Thousand Oaks, CA: Corwin Press.
19. Brown, J. H., & Shavelson, R. J. (1996). *Assessing hands-on science: A teacher's guide to performance assessment*. Thousand Oaks, CA: Corwin Press.
20. Lu, C.-H., & Suen, H. K. (1995). Assessment approaches and cognitive styles. *Journal of Educational Measurement, 32*(1), 1-17.
21. Mayer, R. E. (1987/1991). *Educational psychology*. 林清山（譯）。**教育心理學 — 認知取向**。台北市：遠流。
22. Quellmalz, E. S. (1991). Developing criteria for performance assessments: The missing link. *Applied Measurement in Education, 4*(4), 319-331.
23. Ruis-primo, M. A., & Shavelson, R. J. (1996). Rhetoric and reality of performance assessments: An update. *Journal of Research in Science Teaching, 33*, 1045-1063.
24. Shavelson, R. J., & Webb, N. W. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage.

附表 1：設計與進行實驗之過程技能實作評量評分規準表

(一)設計實驗的能力

| 表現指標 | 評分項目 | 表現水準之界定及說明 |
|---|---|---|
| 擬定問題 | | |
| 1-1 | 擬定問題 | ⑤能根據情境擬定探討的問題，對問題敘述清楚 |
| | | ④能根據情境擬定探討的問題，但對問題敘述不清楚 |
| | | ③有擬定問題，但與情境無關，可是敘述清楚 |
| | | ②有擬定問題，但與情境無關，也敘述不清楚 |
| | | ①未能擬定問題 |
| 提出實驗假設 | | |
| 1-1 | 提出實驗假設 | ⑤能清楚寫下實驗假設，且與探討問題相符合 |
| | | ④實驗假設寫的不清楚，但與探討問題相符合 |
| | | ③有清楚寫下實驗假設，但與探討問題無關 |
| | | ②未能清楚寫下實驗假設，且與探討問題無關 |
| | | ①未能寫下實驗假設 |
| 設計實驗變因 | | |
| 2-1 | 選定操縱的變因 | ⑤擬定一個操縱變因，且與實驗假設相符合 |
| | | ④擬定一個操縱變因，且與實驗假設相符合，但敘述不清楚 |
| | | ③擬定一個操縱變因，但與實驗假設不符合 |
| | | ②擬定操縱變因超過一個 |
| | | ①未能擬定操縱變因 |
| 2-1 | 選定應變的變因 | ⑤能界定一個正確的應變變因，且說明清楚明確 |
| | | ④能界定一個正確的應變變因，但說明不清楚明確 |
| | | ③將其他變因列為應變變因 |
| | | ②有寫出應變變因，但與實驗假設不符 |
| | | ①未能擬定應變變因 |
| 2-1 | 界定保持不變的變因 | ⑤能界定除了操縱變因與應變變因之外的大多數變因（3 個以上），且與實驗假設相符合 |
| | | ④能界定除了操縱變因與應變變因之外的大多數變因（3 個以上），但部份與實驗假設不符合 |
| | | ③能界定除了操縱變因與應變變因之外的小部分變因（1 至 2 個），且與實驗假設相符合 |
| | | ②將操縱變因與應變變因列為控制變因 |
| | | ①未能界定控制變因 |

小計：設計實驗的能力得____分（25）

☐ 非常棒（22 至 25）

☐ 很厲害（20 至 21）

☐ 還不錯（17 至 19）

☐ 普通（15 至 16）

☐ 要再加油（14 以下）

（二）進行實驗的能力

| 表現指標 | 評分項目 | 評分規準的說明及表現程度 |
|---|---|---|
| 根據假設操縱變因 | | |
| 1-2 | 操縱變因的符合度 | ⑤實驗時只操縱一個變因，且符合選定的變因 |
| 2-2 | | ④實驗時只操縱一個變因，且符合選定的變因，但操作有誤 |
| | | ③實驗時只操縱一個變因，但未符合選定的變因 |
| | | ②實驗時操縱兩個以上的變因 |
| | | ①並未操縱任何變因 |
| 在實驗中控制變因 | | |
| 1-2 | 確實控制變因的個數 | ⑤在同一個實驗中，除了操縱變因外，其餘變因全部控制好 |
| 2-2 | | ④在同一個實驗中，除了操縱變因外，有一個變因沒有控制好 |
| 6 | | ③在同一個實驗中，除了操縱變因外，有二個變因沒有控制好 |
| | | ②在同一個實驗中，除了操縱變因外，有三個以上變因沒有控制好 |
| | | ①完全沒有控制任何變因 |
| 1-2 | 控制變因的操縱能力一 | ⑤能正確操作 |
| 2-2 | | ④大部分能正確操作 |
| 6 | | ③操作錯誤但有修復 |
| | | ②操作錯誤且未修復 |
| | | ①未能操作 |
| 1-2 | 控制變因的操縱能力二 | ⑤能正確操作 |
| 2-2 | | ④大部分能正確操作 |
| 6 | | ③操作錯誤但有修復 |
| | | ②操作錯誤且未修復 |
| | | ①未能操作 |
| 1-2 | 控制變因的操縱能力三 | ⑤能正確操作 |
| 2-2 | | ④大部分能正確操作 |
| 6 | | ③操作錯誤但有修復 |
| | | ②操作錯誤且未修復 |
| | | ①未能操作 |
| 1-2 | 作品的完成情形 | ⑤能正確完成 |
| 2-2 | | ④能正確完成，但有瑕疵 |
| 6 | | ③未能正確完成但有修復 |
| | | ②未能正確完成且未修復 |
| | | ①未能完成 |
| 觀察應變變因 | | |
| 1-2 | 使用作品操作 | ⑤能正確操作 |
| 2-2 | 結果的正確性 | ④能正確操作，但有瑕疵 |
| | | ③未能正確操作但有修復 |
| | | ②未能正確操作且未修復 |
| | | ①未能操作 |

| | | |
|---|---|---|
| 1-2 | 觀察應變變因的正確性 | ⑤能正確觀察應變變因 |
| 2-2 | | ③未能正確觀察應變變因 |
| | | ①不知要觀察應變變因 |

小計：進行實驗的能力得____分（40）

☐ 非常棒（36 至 40）

☐ 很厲害（32 至 35）

☐ 還不錯（28 至 31）

☐ 普通（24 至 27）

☐ 要再加油（23 以下）

（三）記錄實驗結果的能力

| 表現指標 | 評分項目 | 評分規準的說明及表現程度 |
|---|---|---|
| 記錄實驗結果 | | |
| 5 | 設計表格 | ⑤將觀察結果用表格表示，且設計正確 |
| | | ④將觀察結果用表格表示，但部份設計有誤 |
| | | ③觀察結果未用表格形式表示，但整齊呈現 |
| | | ②觀察結果紀錄零散呈現 |
| | | ①未用任何方法做觀察紀錄 |
| 5 | 實驗結果的清晰度 | ⑤實驗結果的紀錄包含操縱變因與應變變因 |
| | | ④實驗結果的紀錄只有應變變因的結果 |
| | | ③實驗結果的紀錄只有操縱變因的項目 |
| | | ②實驗結果的紀錄分類標示不清 |
| | | ①未記錄實驗結果 |
| 5 | 記錄實驗結果的符合性 | ⑤實驗紀錄與實際操作結果相符合 |
| | | ③實驗紀錄與實際操作結果不符 |
| | | ①未記錄實驗結果 |
| 測量誤差的控制 | | |
| 5 | 對於測量誤差的處理方法 | ⑤進行 3 次相同步驟，並將測得結果加以平均 |
| | | ④進行 3 次相同步驟，但沒有平均 |
| | | ③進行 2 次相同步驟，並將測得結果加以平均 |
| | | ②進行 2 次相同步驟，但沒有平均 |
| | | ①只進行一次，且未考慮測量的誤差問題 |
| 5 | 計算測量數值的正確性 | ⑤計算的平均值正確 |
| | | ③計算的平均值有誤 |
| | | ①未計算平均 |

小計：記錄實驗結果的能力得____分（25）

☐ 非常棒（22 至 25）

☐ 很厲害（20 至 21）

☐ 還不錯（17 至 19）

☐ 普通（15 至 16）

☐ 要再加油（14 以下）

（四）驗證假設下結論的能力

| 表現指標 | 評分項目 | 評分規準的說明及表現程度 |
|---|---|---|
| 驗證假設下結論 | | |
| 3 | 驗證假設 | ⑤能將實驗結果連結到原先所寫的實驗假設上 |
| | | ④未能將實驗結果連結到原先所寫的實驗假設上，但實驗結論的描述可明顯看出和假設有關係 |
| | | ③未能將實驗結果連結到原先所寫的實驗假設上，且其描述和假設之間的關係並不明顯 |
| | | ②未能將實驗結果連結到原先所寫的實驗假設上，或兩者看不出任何關係 |
| | | ①未下任何結論無法判斷 |
| 4 | 做成結論 | ⑤能根據實驗結果做成結論，且敘述清楚明確 |
| | | ④能根據實驗結果做成結論，但敘述不清楚或不明確 |
| | | ③有下清楚的結論，但結論與實驗結果不相符 |
| | | ②有下結論，但不清楚且與實驗結果不相符 |
| | | ①未下任何結論 |

小計：驗證假設下結論的能力得____分（10）

☐ 非常棒（9 至 10）

☐ 很厲害（8）

☐ 還不錯（7）

☐ 普通（6）

☐ 要再加油（5 以下）

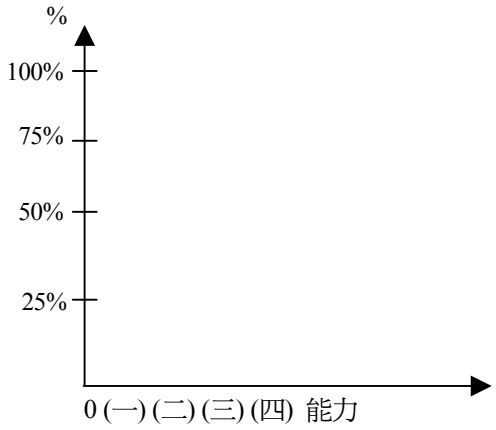
（五）整體評價

（一）設計實驗的能力：$\frac{\quad}{25}$ ＝　　%

（二）進行實驗的能力：$\frac{\quad}{40}$ ＝　　%

（三）記錄實驗結果的能力：$\frac{\quad}{25}$ ＝　　%

（四）驗證假設下結論的能力：$\frac{\quad}{10}$ ＝　　%




整體評價：

☐ 甲等（90－100）

☐ 乙等（80－89）

☐ 丙等（70－79）

☐ 丁等（60－69）

☐ 戊等（59 以下）

# Conducting a Performance Assessment on Designing and Implementing Experiments in Nature Science and Life Area

**Chun-Hsien Lu[1] and Yuh-Yin Wu[2]**

[1]Taipei Municipal Shuang-Yuan Elementary School

[2]Department of Psychology and Counseling,
National Taipei University of Education

**Abstract**

The purpose of this research was to design and conduct a performance assessment on students' ability on implementing experiments. Its effects were also discussed. Related ability indicators were transformed into assessment objectives. There were four essential elements in the assessment objectives: performance task, response format, scoring rubrics, and meta rubrics, to ensure the quality of assessment design. After the pilot test with 26 students and formal test with 25 students, a two-stage assessing procedure was applied to motivate students' learning and understanding. Pearson's r, Cronbach's α, and Kappa's κ all indicated the consistent rating results from two raters. This study concluded that the two-stage four-element performance assessment functioned as a mediator to depict students' ability structure, enhance their designing ability and provide teachers with feedback on teaching.

**Key words**: Response Format, Nature Science and Life, Ability Indicators, Scoring Rubrics, Performance Assessment