

過程技能能力要項藉由能力指標聚焦並轉化成 實作評量設計之行動研究

魯俊賢¹ 陳美英² 吳毓瑩³

摘要

本研究之目的乃在呈現三位自然與生活科技領域國小教師與一位師資培育機構教師如何透過行動研究，共同合作解決現場問題做為行動的核心，完成實作評量之規畫與設計，以評量學生之設計並進行實驗之過程技能的能力。行動團隊首先發覺自然與生活科技領域中過程技能能力要項下，有許多能力內涵，其中設計並進行實驗之能力包括的能力指標範圍最廣、內容也最繁複，且縱跨不同的學習階段，行動團隊亟思統整以周延描述學生的實驗能力。團隊一方面在行動中透過文獻理論的說法，激盪出問題解決策略，另一方面，也是關鍵的方式，透過團隊成員間的腦力激盪，將豐富的實務知識結合理論，規畫出實作評分規準以化解疑惑。行動過程有三階段：首先，從過程技能能力要項的分析開始，聚焦於相關的能力指標，將之轉化為評量目標。接著第二階段則從目前使用的教材中選擇符合評量目標的活動，設計實作評量任務。並於第三階段依評量目標及任務內容建構評分規準，包括效標、表現指標、評分項目、表現水準以及各水準的內涵說明。此規準的建立，可以讓教師在實施實作評量時，解決在教學現場如何利用實作評量結果回饋學生學習以及教師教學的問題，並回答家長對於評量公平性的質疑。基於行動研究在成果分享時，容易疏漏行動歷程的遺憾，我們將自己在遭遇困境後，建構實作評量的歷程透過本文之整理，期望藉由行動過程之經驗分享，協助同儕將此歷程應用在類似的問題與情境上，達成教師研究社群的專業實踐。

關鍵詞：能力指標、實作評量、評分規準、行動研究、專業實踐

1 魯俊賢，台北市雙園國小教師
電子郵件：ericlu@tp.edu.tw

2 陳美英，台北市雙園國小教師
電子郵件：cmy4444@tp.edu.tw

3 吳毓瑩，國立台北教育大學教育心理與諮商學系教授
電子郵件：yuhyin@tea.ntue.edu.tw

投稿日期：2005年3月25日；採用日期：2005年10月31日

An Action Research of A Performance Assessment Design on Assessing Students' Ability of Conducting Experiments

Hun-Hsien Lu¹ Mei-Ying Chen² Yun-Yin Wu³

Abstract

The purpose of this action research was to form a performance assessment under the objectives of experimental ability indicators in the domain area of physical sciences from 9-year-integrated curriculum. The action team was composed of three elementary school teachers in physical sciences and one professor in teachers' training institute. They had a on-the-field problem about performance assessment: how can it be designed in the way that it comprises fairness, convenience, and authenticity. They started from analyzing ability indicators, then allocating appropriate tasks. Scoring rubrics were designed based on the analyzing results of ability indicators. The team set up a complete performance assessment module on the one hand to resolve their own problem, on the other hand to demonstrate teachers' professional praxis.

Key words: ability indicators, performance assessment, scoring rubrics, action research, professional praxis

1 Chun-Hsien Lu, Teacher, Taipei Municipal Shuang-Yuan Elementary School

E-mail: ericlu@tp.edu.tw

2 Mei-Ying Chen, Teacher, Taipei Municipal Shuang-Yuan Elementary School

E-mail: cmy4444@tp.edu.tw

3 Yuh-Yin Wu, Professor, Department of Educational Psychology and Counseling, National Taipei University of Education

E-mail: yuhyin@tea.ntue.edu.tw

Manuscript received: Mar. 25, 2005; Accepted: Oct. 31, 2005

壹、前言

本文所探討的問題是在教學現場發現的，並因問題的存在，而有行動團隊的成軍，以實際行動進行研究並得到成果。文章的順序是以時間的順序為主軸，呈現出我們合作行動研究的過程與結果。首先在貳——「蓄勢待發」中，描述九年一貫課程的能力指標對教學的影響，在與同儕老師分享實驗進行之過程技能能力要項評量的難處時，浮現出實作評量的方案，然此方案沒有人力構思，而在我們的領域教師編制減少一人的時候，竟成為解決現場問題的生機。我們於參——「成軍——行動策略」中，敘述行動團隊的成軍過程，並介紹行動策略、研究場域與團隊成員，以及分析架構與行動歷程。於肆——「激盪——在行動中與文獻交會」中，描述我們在行動中收集理論知識，行動團隊閱讀理論之後，將我們自己隱而未彰的隱形知識與顯性知識交會激盪之後，將省思所得化為實際行動，形成本研究的行動智慧。伍——「行動結晶」，呈現我們在行動中與文獻交會之結果，把行動智慧的結晶清楚地呈現出來。

貳、蓄勢待發

一、問題的浮現

九年一貫課程綱要（教育部，2003）中明定了各學習階段的能力指標，自然與生活科技領域更明確列出學生的「科學與科技素養」，「過程技能」為其中的第一項。然而，過程技能的能力指標要如何評量呢？是不是所有的過程技能都可以透過某種評量方式評量出來呢？要如何做才好？這是我們現場教師的急迫問題。

近年來，因為傳統客觀式紙筆測驗的限制，使得另類評量（alternative assessment）逐漸受到重視。一般而言，研究結果多半擔心如教室中僅實施單一客觀式紙筆測驗，會造成學科知識窄化，忽略技能與情意的內涵，無法促進問題解決等較高層次的思考（李坤崇，1999；吳欣黛，1998；桂怡芬，1996；張麗麗，2002；蘇義翔，1998）。另外，Darling-Hammond（1994）及 Janesick（2000）亦提到，在美國，傳統評量最大的缺點是在不同學生族群中、非英語系的學生及低收入戶的學生間造成不公平的結果。但是從我（本文第一作者）任教學校的生態看來，家長們對於傳統評量的信任，仍是無法改變的，多數教師不覺得客觀式紙筆測驗應該被取代，這表示家長及教師認為紙筆測驗有其存在的價值。而在傳統評量與另類評量之間，要如何取捨？以我和美英老師（本文第二作者）的做法為例子，對於學生認知表現的部分，以紙筆測驗為主，輔以書面報告、觀察紀錄等任務來補充，而紙筆測驗評量不到的能力，例如過程技能、情意表現，就以實作評量（performance-based assessment）的方式來進行。

然而這些年來，我們在實施實作評量時，遇到了一些問題，這些問題可以從家長、學生及教師本身三方面來看。

就家長而言，家長基本上對於實作評量的實施大多給予正面的肯定，認為實作評量給予學生多元表現的機會，評量方式活潑也深受肯定。但是家長最常質疑的就是它的公平及客觀，部分家長想要明確清楚的知道老師如何進行評量，且評分標準在哪裡？會不會因為老師對學生的偏見或月量效應而產生不公平的現象。教師面對這樣的問題，往往無法具體客觀地提出有效的證據來說明。

就學生而言，學生很喜歡以實作評量的方式進行測驗，感覺沒有壓力，在活動中就可以完成一項考試，這和許多研究的結果類似（李淑娟，2003；席家玉，2002；陳怡如，1999；陳學淵，2003；楊銀興，2000；蔡菁玲，2001；簡琇芳，2002）。我在實施實作評量時，往往因為時間的因素，並顧及評量過程不具系統性，因而不敢大意公布評量結果即時回饋學生，學生在接受評量時也不清楚知道自己有哪些問題需要再學習，要如何改進。雖然學生表達喜歡實作評量，但由於我的評分原則不具體，造成對學生學習上的回饋沒有實質幫助。

就教師而言有兩方面的問題。首先，進行一項實作評量任務，有一些技術上

的問題需克服，例如：經費及設備上的問題，耗費時間施測等（桂怡芬，1996；彭森明，1996；盧雪梅，1997）。教師進行實作評量設計時，大都以容易及方便施測為主要考量因素，在進行實作評量時，因為過於簡便的評分原則，教師也會質疑評量是否客觀公正，但是如果評分原則過於繁瑣，又不方便實施，客觀公正與方便實施如何能兩者兼顧？另外，我們設計的實作評量，難道真的就表示是學生能力的結構嗎？吳毓瑩（2004）指出，評量編製者在取樣內容時，需考慮受試者在面對評量時的反應歷程（response process），是否為編製者心中以為的反應歷程，二者必須是一致的。我以往設計出的實作評量任務，是以一個專家的立場設計的，但是這樣的任務可以反映出學生的反應歷程嗎？

因為實作評量無法立即回饋學生的表現，評量的公平性受到質疑，所以，如何訂定一個公正客觀、可以給學生回饋又容易實施的評分規準，既能具體解答家長的質疑，又能讓學生瞭解學習的能力面向，提供學生瞭解與檢視自己表現的資訊，回饋學生的評量表現，並讓教師清楚知道評量是否合乎學生的反應歷程，這些似乎是我在實施實作評量所應解決的問題。張麗麗（2002）亦曾提及教師在實施實作評量時，常缺乏編製作業任務以及系統性計分之技術協助，這與我在現場遭遇的問題不謀而合。因此，在與毓瑩老師（本文第三作者）將問題攤開來檢討後，我們便著手邀請學校夥伴，一同來思考可採行的策略，以解決現場問題。

二、危機乎、轉機乎、生機乎

九十三學年度對於我們行動研究的現場，嘉嘉國小的自然領域而言，出現了一個危機。六年級有一位自然老師申請退休獲准，行政決定將四位領域老師減為三位，對我們而言這是一個危機。最後的決定是各自保留自己原先任教的年級（三、四、五年級），而六年級的六個班級，就每人分擔兩個班級。因此我們開始規畫協同的方式，從教學準備到教學活動設計，甚至教學活動實施時的分享，到教學活動後的省思，都可以一起討論，這樣似乎可以改進我們之前的互動模式——大家各教各的，常常分享教學經驗卻不能立即給予實質的幫助。我們或許可以形成一個更佳的夥伴關係，一個更有效能的領域團隊。

因為達成了這樣的共識，這個團隊的任務與夥伴的使命感更形堅固，讓我們的實作評量問題產生了一線生機。在我的提議下，我們決定以六年級的教材為範圍，尋找一個能力來嘗試規畫，討論建構實作評量任務及評分規準，如果這樣的模式能夠成功，我們不但可以解決教學現場的問題，更能因此自我培育教師的專業能力。

參、成軍——行動策略

二〇〇四年七月上旬，在兩位夥伴教師的同意下，我們的行動團隊（以下簡稱行動團隊）順利成軍，短短的時間內讓兩位教師願意一起行動，除了我們有過多次不錯的合作經驗外，也表示她們有強烈的解決問題動機。我希望透過「合作行動研究」（co-operative action research）（甄曉蘭，1995）的方式，從實際的行動中來解決問題，藉著討論、行動、反省的過程，不斷地建構概念，期能建立出一個進行實作評量的模式。

一、研究場域

嘉嘉國小位於台北市的傳統社區，成立於二〇年代，是一所有著輝煌歷史的學校。全校目前普通班有三十一班，特殊教育班有十四班，教師數約有九十餘位，由於學區內人口外移情形嚴重，班級數一直在遞減中，目前普通班維持穩定的三十多班。隨著教育改革步伐的加快，嘉嘉國小資深教師退休的速度，在近五年達到高峰，同時也聘進了許多新進教師，並在九十二學年度止畫下休止符。也因此，學校新進教師與資深教師，在經驗的傳承上所出現的斷層現象，將要進入磨合期。年輕教師雖然教學經驗不足，但是充滿熱忱，教師教學討論的風氣也逐漸展開。學校教師已取得碩士以上學位或正進修學位中的教師約有十餘位，教師的研究風氣正在改變中。

二、團隊成員

行動團隊的夥伴們，常常在正式與非正式的場合彼此討論，交換教學心得，並分享新的想法，而且有過多次的合作經驗，並得到不錯的成績。我們三位老師曾經共同編撰「嘉嘉國小校園植物圖鑑（上下冊）」、師大舉辦的「教學模組設計」特優、教育部舉辦的行動研究優等及台北市行動研究入選。相信我們的合作經驗及夥伴的專業能力，對於本研究有絕對的幫助。

(一)我

我帶過兩屆高年級學生，由於自己具有理科背景，對自然科學又很有興趣，所以卸下行政工作之後，主動向校長爭取擔任自然科科任教師，轉眼間，自然科科任教學工作也步入第十二年了。在智育掛帥的時代裡，自然科也曾經是家長們所重視的「顯學」，但在當時，我就常和學生溝過程技能的重要，教學活動也比較重視實際操作。因此，學生都期待上自然課。教育改革高唱的現在，我更重視學生形成性的表現，相信他們不會排斥接觸自然科學。希望透過這次的研究機會，能夠激發同事提升專業能力的興趣，或是激發同儕間的影響力，試圖建立一個專業的研究社群。

(二)美英老師

美英老師從師專畢業至今，教學生涯已經進入第十八年了，她創新的教學理念及有效的班級經營策略，常是許多新進教師學習模仿的榜樣。美英老師擔任自然科的教學已第八年，她常和學生一起在生活中發現自然，並帶著學生動手做，一起觀察，共同體會發現的樂趣。美英老師覺得自然科的評量不應只著重於紙筆考試，所以，她擔任自然科教學的第一年，就非常重視學生過程技能之能力，每學期都會有二至三次的實作評量，她希望能藉著這個評量，讓不同能力的學生都有表現的機會。美英老師覺得教師平時教學工作量很大，若還要在繁重的工作中抽出時間做研究，必定會影響教學工作，不過若透過合作行動研究的方式來進

行，便可以相互截長補短，集思廣益，她期待在研究過程中也能對自己有所助益，並得到不同的體驗。

(三) 毓瑩老師

毓瑩老師在師資培育機構任教也有十二年了。與上述三位現場教師一同經歷了開放教育、小班教學、多元智慧、九年一貫等近十年來的教育改革思潮之洗禮。在我第一次與毓瑩老師提及心中的困惑時，毓瑩老師一掃我的陰霾，擊掌說道：「為何我們不一起來建構一個適合現場狀況的實作評量？」毓瑩老師的明快與行動力，也是促成我們行動團隊積極成軍的原因。由於毓瑩老師已指導過多篇實作評量的研究，也撰述不少評量相關的議論文獻，但是，正如毓瑩老師說的：「我以前與現場教師合作的實作評量都偏向心理計量的思考。而有了成果之後，往往僅是成為案頭上的一篇期刊文章，對於現場的影響似乎不大。」毓瑩老師有一股不能落實教室中的實作評量之遺憾。她的遺憾與我們的困境，二者相乘，成就了今日的合作行動之成果。

(四) 莉莉老師（匿名）

莉莉老師調入嘉嘉國小就擔任自然科的教學，當年擔任自然科教學的原因，是因為只剩下一個自然科任缺，就這樣地走入了第十六個年頭。對非理科背景的莉莉老師而言，當初這樣的安排，確實讓她緊張了許久，對自然科的教學準備與教學掌握上都必須重新調整。自然科學的專業地位早已穩固的莉莉老師認為自然科學的探究必須從動手做中體會，她非常重視學生的過程技能表現。早期對於過程技能的評量，是以學生在課堂操作時的表現，隨時觀察評分，之後會以活動單的形式，或設計比較開放性的紙筆測驗，來評定學生的過程技能。由於莉莉老師曾經透過行動研究的方法來解決現場教學的問題，所以她很期待以合作行動研究來發展一些策略，解決現場問題，提升教師專業。然本文在撰寫之際，莉莉老師認為自己在撰稿與分析上並未有貢獻，因此主動提出不成為共同作者。

三、分析架構與行動歷程

(一)分析架構

本研究以開發實作評量工具為主要目的，有三個目標需要完成：(1)過程技能能力要項透過能力指標轉化為實作評量目標、(2)實作評量任務定向及設計、(3)建構評分規準。為了要達成第一個目標，我們必須探討與分析「過程技能能力要項」及「能力指標轉化評量目標的策略」，並分析「實作評量內涵」、「教材」及探討「設計實作評量任務的方式」，以達到第二個目標。討論「建構評分規準的方式與原則」與「評分規準的形式」以達成第三個目標。我們在行動過程中一方面尋找文獻、閱讀文獻，並於理論知識中融入實踐知識，同時在團隊教師的修正中，釐清了「能力指標轉化評量目標之方式」、「過程技能實作評分規準建構之可行方式」。行動架構如圖 1。

(二)行動歷程

確立了方向後，我們便開始一連串文獻閱讀與實際行動相互配合的歷程，進行實作評量任務與評分規準之規畫。我負責尋找文獻並將這些資訊帶入行動團隊中進行分享，利用領域會議時間或週三下午研習的時間，作為我們討論的時間，且訂出每次會議的討論主題，以便於有效控制討論進度，並透過腦力激盪、實作、觀察、相互回饋討論、文件分析等方式進行。詳細行動的時間及討論主題與方式如表 1。

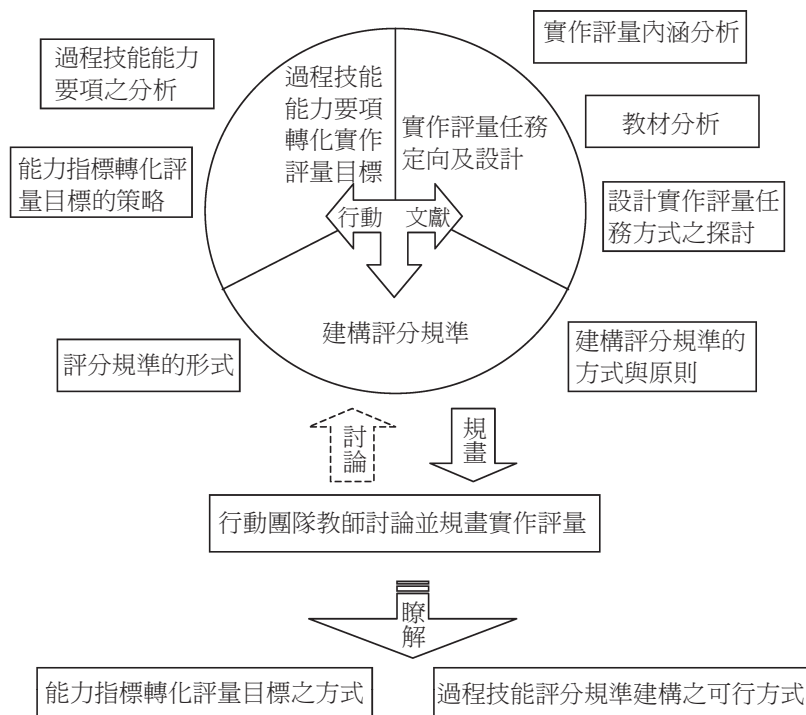


圖 1 過程技能能力要項藉由能力指標轉化實作評量之分析架構

肆、激盪——在行動中與文獻交會

我們採取行動與文獻同步進行的方式，將文獻活化，成為我們思考的重要支柱。在此先呈獻我們閱讀的結果，至於行動與文獻交會的成果，則列於伍——行動結晶之中。以下之三階段的閱讀成果，緊扣住三階段的行動目標以及結晶。行動歷程中，我們的閱讀層次與行動結果相互呼應與支援。

表 1 行動歷程之時程與進度

日期	主題及事項	進行方式	參與人員
2004.8	● 文獻蒐集與閱讀	閱讀與經驗整理	我
2004.8.23 (一)	● 文獻閱讀心得分享	分享	美英師、莉莉師、我
2004.8.27 (五) 下午	● 實作評量內涵及實際經驗分享 ● 確立研究內容及進程 ● 分析架構說明	分享、腦力激盪	美英師、莉莉師、我
2004.9.1 (三) 下午	● 實作評量理論及實務經驗 ● 能力指標轉化策略	文獻、理論與實務交錯分析	毓瑩師、我
2004.9.3 (五) 領域會議時間	● 分析第三階段能力指標 ● 過程技能能力要項下能力指標的選擇 ● 能力指標轉化策略	課程內涵與教學目標分析、實作	美英師、莉莉師、我
2004.9.8 (三) 下午	● 能力指標轉化任務目標 (一)	任務目標草稿成形、討論、修改	美英師、莉莉師、我
2004.9.10 (五) 下午	● 能力指標轉化任務目標 (二)	任務目標初稿討論、修改	美英師、莉莉師、我
2004.9.17 (五) 領域會議時間	● 設計實作評量任務討論 ● 實作評量任務定向 (一)	腦力激盪、實作任務草稿成形	美英師、莉莉師、我
2004.9.22 (三) 下午	● 實作評量任務定向 (二)	實作任務修改、討論	毓瑩師、我
2004.9.24 (五) 下午	● 實作評量任務定向 (三) ● 實作評量任務初步規畫	實作任務初稿討論、修改	美英師、莉莉師、我
2004.10.1 (五) 領域會議時間	● 評分規準內涵 ● 評分規準表形式	結合任務目標與歷程討論評分規準	美英師、莉莉師、我
2004.10.6 (三) 下午	● 評分規準探討	評分規準效標修正與討論	毓瑩師、我
2004.10.13 (三) 下午	● 建構評分規準	評分規準初稿討論	美英師、莉莉師、我
2004.10.20 (三) 下午	● 建構評分規準表	評分規準初稿再討論	美英師、莉莉師、我
2004.10.29 (五) 下午	● 再探實作任務目標、歷程、與評分規準	實作任務試做、觀察、反省	美英師、莉莉師、我

表 1 行動歷程之時程與進度（續）

日期	主題及事項	進行方式	參與人員
2004.11.5（五） 下午	●實作任務目標、歷程、與 評分規準交互關係與系統 成形	討論與修訂	毓瑩師、我
2004.11.17（三） 下午	●與實務經驗交互檢驗、修 改	討論分享	美英師、莉莉師、我
2004.11.26（五） 下午	●前瞻下一學期的進度與實 施進程	相互回饋、分享	美英師、莉莉師、我

一、能力指標轉化的意義與策略

在進行教學評量前，我們深知教師必須根據教學目標與教材內容來決定評量目的。若無明確的評量目的，將難以辨認與定義實作的表現行為或作品，不易建立適當的評分規準和計分程序。從清楚界定評量的目的開始，才能進而發展出評量任務及評量形式。我們首先碰到的第一個問題是：在九年一貫課程中，實驗進行的過程技能之能力指標需要轉化為更具體清楚的評量目標嗎？如果需要，要如何轉化？我們試著從文獻的理解開始，了解能力指標轉化的意義與策略。

（一）能力指標轉化的意義

九年一貫課程強調培養學生具有帶得走的能力，能力指標可以是教材統整規畫的基礎，亦是課程設計的核心，因此，如何詮釋能力指標，將能力指標轉化為教學目標及教材內容，是九年一貫課程中非常重要的課題。許民陽（2002）指出教師在設計活動時，應先解讀能力指標，再將之轉化為適當的教學目標與設計合適的教材內容，透過適當的教學方法與歷程，使學生習得科學與科技知識及培養基本能力，達成預定的學習目標，以符合九年一貫課程的基本精神。

高新建（2001）認為能力指標對於教師而言，依舊和教育目標或課程目標類似，都是比較抽象而遙遠的理想，所以，能力指標需要經過一個或數個轉化的步驟，才有可能轉化成為教學材料或教學活動的形式。因此，教師應具有解析與轉

化能力指標之能力，才能依轉化之結果，設計教學活動，檢視教學活動是否達成能力指標，作為評量之依據。

(二)能力指標轉化的策略

葉連祺（2002）認為能力指標轉化為教學活動有以下幾種策略：

1. 替代（replace）：利用「一對一」對應轉化的關係，以某主題物替換原有能力指標內的關鍵詞，形成教學目標。
2. 拆解（decompose）：使用「一對多」對應轉化關係，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。
3. 組合（group）：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。
4. 聚焦（focus）：由多個「具關鍵性的一對一」對應轉化關係所構成，係選取某能力指標的某部分或全部為主軸，以其為教學焦點，逐次擴大發展其他活動，可運用認知層次如觀察、紀錄、敘述、比較、分析等，作為擴展的參考。
5. 聯結（relate）：聯繫「多組一對一」對應轉化關係，先以某個能力指標和主題成為發展活動的起點，再不斷聯結其他不同學習領域或思考層面（如人、事、時、地、物），構成一個課程內容。
6. 複合（mix）：是適度擇取前述五項策略的某幾種或全部，形成複雜的轉化關係，進而發展出一個或多個教學活動。如圖 2。

策略	替代	拆解	組合	聚焦	聯結
對應轉化關係	一對一	一對多	多對一	多組一對一	多組一對一
關係圖示					

圖 2 能力指標轉化策略之比較

資料來源：九年一貫課程與基本能力轉化（葉連祺，2002）。

能力指標詮釋或轉化時，可以有不同的轉化策略，因此能力指標如何經由詮釋、適度轉化為課程以達到培養能力的要求，都是國內亟待研究及解決的問題（許民陽、林麗詩，2004）。我們團隊所採取的方式，請見伍——行動結晶之第一部分。

二、實作評量的任務定向

有了評量的目標之後，實作評量可以怎麼做呢？有沒有一定的標準程序可依循？學者們對於實作評量的定義不一，同樣的，實作評量的做法也沒有一定的標準程序可依循。因此，我們綜合歸納學者們的看法與建議後（李坤崇，1999；呂金燮，2004；吳璧純，2003；莊明貞，1996；張麗麗，2002；盧雪梅，1997；Airasian, 1997; Brualdi, 1998; Linn & Gronlund, 2000），整理出以下對自己的提醒：

- (一)在設計任務前，教師應先評估可利用的時間、可獲得的教學資源、在整個問題中學生扮演的角色、學生在進行問題解決時需要多少材料和資料、學生將需要應用到哪些知識或技能的問題。
- (二)在設計任務時，應利用題目引導學生運用所學的知識，不只是使用教師所給予的線索，不經思索的完成一個任務。簡言之，題目的設計，要選擇一個可以讓學生發揮所學的評量情境，讓學生使用所欲評量的知識或技能，以表現出學習的成果。
- (三)對於實作任務難易度的考慮，可由問題的結構化程度、情境的新奇度、解題的途徑和表達的方式等來決定。一般而言，若教師是剛開始使用實作評量，可由中度挑戰的實作任務開始，慢慢地隨著學生對挫折容忍力以及接受延遲報酬增加之後，再逐漸提升實作任務的挑戰性。
- (四)實作任務所激發的學生表現，要與評量目標所指涉的效標相符，換言之，此任務本身必須對於效標具有代表性，可讓學生於完成任務時，達到或是朝向效標所要求的方向精進。
- (五)教師規畫合適的評量情境，讓學生選擇表徵的方式（口語或圖形）以表現

所學會的內容與概念，儘可能地減低與評量目的不相關的知識和技巧之需求。如有關連，例如實作任務需要應用到先前的知識時，則儘可能在評量進行之前，融入教學進度中，以確認實作任務所需要的能力是「實驗進行之過程技能」，非自然科學內容知識。

從圖 1 來看，評量目標、實作任務及評分規準三者相互影響。評量目標決定實作任務，也是建構評分規準的前導引領；實作任務從評量目標而來，並決定評分規準的內涵，而評分規準對於學生學習的幫助，則有賴於明確的評量目標與有意義的實作任務。三者相互支援，乃是我們行動團隊的前進方向。

三、實作評分規準之建構

(一)設計實作評分規準的方法

余民寧（2002）、陳玫良（2002）、陳明印（2002）及 Airasian（1997）提出設計評分規準的具體步驟。首先，教師必須先決定評量的目的是屬於過程或是結果或是兩者兼具。其次，在確定評量的目的之後，教師必須知道所觀察的內容或範圍，並列出觀察與評量的重點，說明這些規準的意義。接下來，與學生共同討論評分規準，確定學生是否真正瞭解其含義，使學生也能建構評分規準，也可以和同儕教師共同決定評分規準。最後，教師進行實作評量之前，必須詳細檢視細節行為的項目，並進行修正。

張麗麗（2002）認為在界定規準的過程中，可透過文獻探討、腦力激盪、分析學生作品等方法，最後，試用並修正評分規準，直至其可行且能反應評量所測的重要能力為止。

(二)發展實作評分規準的依循原則

Quellmalz（1991）提出一個健全的評分規準應包括重要性、忠實性、概化程度、發展的適切性、接近性及實用性等六個原則。我們綜合 Linn & Gronlund（2000）的原則，整理如下圖 3：



圖 3 訂定高品質評分規準

資料來源：評量改革的應許之地，虛幻或真實？——談實作評量之作業與表現規準（張麗麗，2002）。

1. 重要性（significance）：能力的表現不只在於內容的廣度，也要看過程的深度。實作評量的規準必須能描述學生在學科或真實世界中，其重要的學習表現（Quellmalz, 1991）。評分規準中項目的選擇與建構，都要以其可代表認知內容和認知技能等重要結果為主要考量（Linn & Gronlund, 2000）。
2. 忠實性（fidelity）：實作表現的規準應符合忠實性的原則，應與典型情境下產生的表現有相同的評分規準，所以不僅任務與條件要儘可能接近真實生活的自然，對於表現的期望及品質層次也要儘可能真實（Quellmalz, 1991）。評分規準所描述的規準特質必須是可具體觀察、直接判斷，此為自然而真實的基本條件（Linn & Gronlund, 2000）。
3. 概化程度（generalizability）：所謂概化程度是指規準要能應用到其他相似的任務上，也就是說評分規準的呈現要具有一致性，可應用於同一類的實作情境，也就是說評分規準可應用於同一類的任務、情境或條件（Quellmalz, 1991）。
4. 發展的適切性（developmental appropriateness）：評分規準需要詳細說明學生表現的成就水準，及發展上的重要事件，同時要注意所使用的語言，必須具建設性（Quellmalz, 1991）。對於同一評分規準在不同層次上的特質，也應該清楚地界定（Linn & Gronlund, 2000）。
5. 接近性（accessibility）：在評分規準中使用的語言，應明確地讓教師、學生與家長瞭解其代表的意義。簡單說明的評分標準，除了可以讓學生評

估自己的表現情形、認同改善學習的可能性，也可讓家長瞭解學生的進步情況（Quellmalz, 1991）。

6. 實用性（utility）：評分規準要能提供學習表現的品質與改善作法之清楚訊息（Linn & Gronlund, 2000）。評量規準要能描述學生完成任務的優點及缺點，並聚焦在能以教學和實務改善的實作表現特色上。描述的詞語應傳達適合教學的層次，也就是要讓學生瞭解其意涵（Quellmalz, 1991）。

上述的提醒，促進了我們任務擇定的品質以及評分規準的結構。我們的成果請見伍——行動結晶之第二部分。

（三）實作評分規準的結構層次

為有效利用評分規準之結構來描述能力的表現，一般而言，文獻皆指出評分規準有三個層次，但名詞用法不一。我們於團隊中重新釐清其內涵，並認為最適合現場使用，以評量「設計並進行實驗之過程技能」的評分規準，乃是四個層次，與文獻相較，則在具體指標之下，加入評分項目。四層次分別為：(1)第一層次：所欲評量的效標（criteria），乃為最主要的能力構成面向或特質；(2)第二層次：效標之下有具體的表現指標（indicators）；(3)第三層次：具體表現指標之中則有評分項目（items）；(4)第四層次：評分項目之下則有表現水準（performance levels）的界定以及內涵說明。評分規準可以量化或質化或質量化兼具，此部分通常是最瑣碎細節的步驟，現場老師往往略過此部份，而以專家自由心證的方式來評定。而今我們既以合作的方式來規畫一個完整的實作評量模式，大家分工了之後，便也不覺此部分之繁瑣，反而收到彼此監督品質與相互支援之效。

伍、行動結晶

以下將行動團隊呼應肆——文獻激盪的結果，分段說明我們的成果。

一、行動第一階段：能力要項的聚焦以及能力指標的分析與轉化

如肆之一所述，「實作」與「科學與科技素養」在自然與生活科技學習領域中佔有十分重要的份量，然而，哪些「科學與科技素養」適合用「實作」的方式達成，從第三階段能力指標中，分析能力指標及其內涵，這是行動團隊首先要思考的問題。之後依據分析出的能力指標內涵，考慮哪些內涵可以用紙筆測驗來評量，哪些必須用實作評量的方式進行，接著要將這些能力指標轉化為具體的評量目標，以作為建構評分規準之依據。以下敘述能力指標分析的過程，及使用轉化策略轉化能力指標的結果。

(一)能力要項的分析

自然與生活科技領域中，科學與科技素養的能力要項有八項。我們認為，第二項「科學與技術認知」的指標內容比較具體，大部分可以「紙筆測驗」的方式進行評量，而其他七個能力要項的指標內容，則是「紙筆測驗」比較不容易評量的，對於這些紙筆測驗比較不容易評量的能力指標，是否可以用「實作」的方式來評量呢？這是我們必須進行瞭解的。

首先，我們以第一個能力要項「過程技能」為主，其他七項為輔，來進行分析。由於團隊教師所共同合作教學的年級是六年級，而六年級也是過程技能能力最成熟的階段，因此，我們將六年級所屬的第三學習階段的過程技能能力要項之下的各能力指標依其內容做分類，歸納整理出幾個重要的能力內涵如表 2。

表 2 第三學習階段過程技能能力要項之各能力指標內涵歸納表

能力內涵	能力指標
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

從中我們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。至於其他四項過程技能能力內涵包括：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，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。此部分能力如果以「紙筆測驗」的形式，反而不能真實地評量出學生的表現，相對地，我們認為以「實作評量」的形式來評量學生，可以提供學生多元的表現機會，能真實地評量出學生的能力。此四個過程技能能力內涵，如本文在第貳部分蓄勢待發中所述，實作評量的實質回饋功能，一直是我們在現場的瓶頸，而其中，尤以「設計並進行實驗能力」之實作評量，是自然與生活科技老師之最愛與最怕。經過分析與分類之後，我們亦發現「設計並進行實驗能力」此能力內涵中，能力指標涵蓋的廣度最大，所包含次項目的內容亦最繁複，此亦為我們目前最需要增進專業評量能力的領域。

在其他七個能力要項之下，我們亦發現 2-3-1-1、3-3-0-5、5-3-1-3、6-3-3-2 等的能力指標與「設計並進行實驗的能力」內涵有關，所以我們亦將之納進，與表 2 目前的能力指標共同進行分析轉化為實作評量目標之後，再依據評量目標決定適合的實作任務。

其中我們所整合的能力指標共有十三項，分別為：（指標內容請見圖 4）

1. 來自過程技能能力要項：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。

2. 來自科學與技術認知能力要項：2-3-1-1。
3. 來自科學本質能力要項：3-3-0-5。
4. 來自科學態度能力要項：5-3-1-3。
5. 來自思考智能能力要項：6-3-3-2。

(二)能力指標轉化為實作評量目標

我們採用葉連祺（2002）的轉化策略，試著就所選擇出的能力指標進行解析，並轉化成實作評量目標。秦嗣輝（2003）用個案研究法探討自然與生活科技領域能力指標轉化的研究中，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。而我們在瞭解了轉化策略的意義與做法之後，採取了「拆解」和「組合」的轉化策略進行能力指標轉化。

首先，行動團隊針對 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，拆解後將其分成五部分（詳見圖 4），再與相關之能力指標進行「組合」。其中 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 訓練學生學習符合科學規範的資料之表述方式，可視為同一能力指標。3-3-0-5 與 6-3-3-2 因內容內涵相近，採取「組合」策略轉化之。詳細的轉化策略及結果以圖 4 表示之。

二、行動第二階段：確立實作評量的單元與任務

(一)以現場實用性為關懷

我們從六年級上學期採用的康軒版（王美芬等，2004）來分析，共有三個小單元內容很適合我們分析出的評量目標，分別是：第一單元「物質的變化」的「影響鐵生鏽的原因」，第二單元「大地的奧秘」的「流水流量與土石沖刷的關係」和第四單元「電磁作用」的「影響電磁鐵磁力大小的因素」。這三個小單元，可以評量學生設計實驗的能力、判斷變因的能力、進行實驗的過程技能及由實驗結果進行推論並歸納結論的能力，並且可以訂定出一套評分規準，用來進行

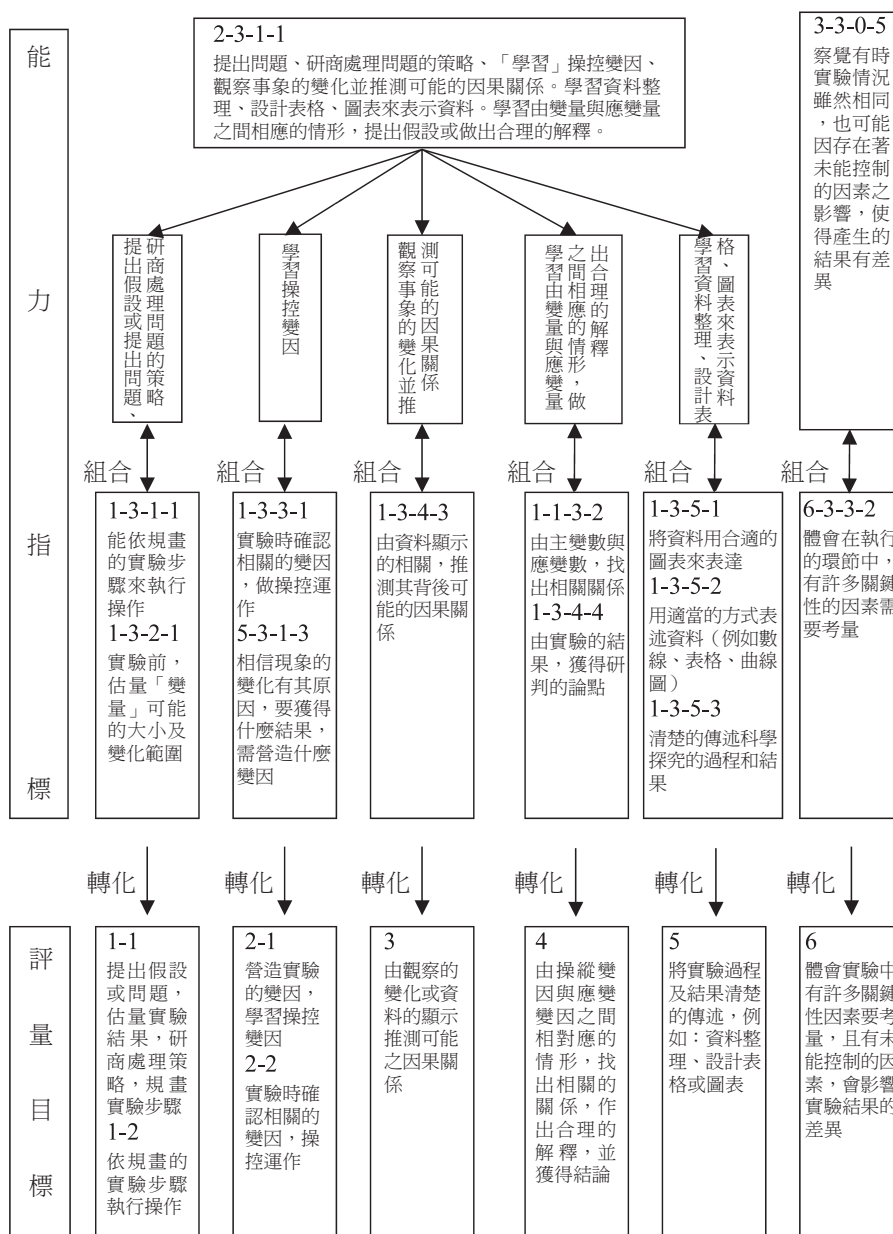


圖 4 第一階段行動結果：能力指標轉化策略與產生之評量目標

這三個實作任務的評量。討論到此，對於評量任務有了初步的聚焦。文獻指出，要選擇對學生有意義、讓學生有參與感的任務，以及有利學生發揮所學的評量情境，讓學生運用所欲評量的知識或技能，以表現出學習的成果（呂金燮，2004；Brualdi, 1998），我們一致認為可以規畫出這樣的任務。幾天後，我們發現一個問題：

根據美英老師的經驗，實施實作評量的過程很耗時，如果一個學期要實施三次實作評量，一定會影響到正常的教學進度，我們有必要進行三次評量嗎？莉莉老師附和地表示，如果一個評量任務就可以達到我們所要達到的能力指標，在九年一貫課程授課時數減少之下，似乎沒有針對同一目標進行三次評量之必要（會 20040917）。

夥伴們所言很有道理，這個問題給我們再次釐清問題的機會。我們目前並不是以「構念中心」（construct-based）為基礎做探討，行動的目的並沒有要探討所欲評量的構念之能力結構的有效性與概化性，所以不需要規畫三個實作任務來進行探討，也沒有必要比較三個任務之間的關係。以實務者的角度出發，就實際場域來規畫，才是解決現場問題之道。而我們教學場域所遇到的問題就是如何規畫評分規準來實施實作評量以瞭解並回饋學生之設計並進行實驗的能力，所以，行動團隊決定訂定一個符合評量目標的實作評量任務，將研究焦點放在評分規準的建構上。

（二）選定適合的單元

我們採取各自選定一個單元進行分析，再一起討論的方法，繼續聚焦我們的實作評量任務。之後，行動團隊得到以下的共識：（會 20040922）

1. 就評量的真實性來看：

這三個小單元都可以真實地評量出學生的過程技能，包括思考能力、操作能力及做結論的能力，也都能設計出具有擬真實性的實作評量任務。

2. 就評量回饋學生學習表現的程度來看：

必須設計出一個具擬真實情境的任務，問題情境夠具體明確，並建構評分規

準給評分者容易評分，並清楚地呈現給學生，這三個單元都可以設計出這樣的實作任務。

3. 就評量的客觀公平性來看：

- (1)「影響鐵生鏽的原因」的實驗的「操縱變因」和「保持不變的變因」可以清楚客觀地觀察，但是「應變變因」的部分，因為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，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，而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。
- (2)「流水流量與土石沖刷的關係」的實驗，其「操縱變因」和「保持不變的變因」可以客觀公平地觀察，但是「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，學生無法做出明確的判斷。
- (3)「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」可以客觀公平地觀察之外，「應變變因」可以以量化的數據作為判斷的依據，學生可以具體的做出判斷。

4. 就評量的方便性來看：

- (1)「影響鐵生鏽的原因」的實驗，其影響之因素很多，在不設限學生的思考判斷之原則下，評量時在器材的準備上會較費時，如果學生的假設超過我們的預期，可能會發生評量時沒有器材可以操作的窘況。
- (2)「流水流量與土石沖刷的關係」的實驗，其「保持不變的變因」在操作上比較困難也較難控制，解決之道就是事先準備好定量（數量與重量）的土沙石，但是這樣的評量，在器材準備上既耗時又麻煩，還必須隨時注意實驗場地整潔的維護及整理。
- (3)「影響電磁鐵磁力大小的因素」的實驗，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」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。

綜合以上討論的結果，行動團隊認為「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，到此，我們的實作評量任務，有了更具體的聚焦。我們的實作評量任務定向在：康軒版（王美芬等，2004）六上第四單元「電磁作用」，評量學生進行「影響電磁鐵磁力大小的因素」實驗的過程技能。然而，這個實驗仍有其評量上的限制必須克服，當學生的想法超過我們的預期時，實驗器材的準備要如何克服，我們的評量任務要如何進行，還有哪些問題需要思考呢？

（三）確立二階段實作之進行

如果實作評量的結果，並不能給予學生立即性的改進回饋，那麼我們何以要這麼大費周章的實施實作評量呢？我們的實作評量可以怎麼做，才能讓學生真正能「從評量中學習」，而不是「學習只為了評量」。如果我們的實作評量只是讓學生知道過程技能的結果，並不能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。

在一次的會議中討論到這樣的問題，如果有一位學生的實驗設計與主題無關，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，他之後的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，我們要期待他能下什麼結論呢？莉莉老師認為：

如果我們把學生評量初期可能發生的錯誤事先做好準備，並作為教學的一部分，讓學生有機會去改正他的錯誤，學生就能「從評量中學習」了。（會 20040924）

我們順著這樣的想法，討論出實作評量任務的兩階段規畫，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，學生根據他的實驗假設進行實驗設計。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，再進行實驗操作，如此有機會讓學生透過教學過程改正錯誤的想法。此外，也可以讓老師根據學生的實驗假設與設計，準備相關的實驗器材，如此，可以事先掌握學生的想法，避免實驗器材準備不足之窘況。

三、行動第三階段：評分規準的建構

(一)建構的方式與過程

由於評定量表（rating scale）的形式可以用在過程的評量上，所以行動團隊打算將評分規準設計成評定量表的形式，讓教師在進行評量觀察時方便勾選出適合描述學生表現的水準點數。我們檢視之前由能力指標轉化出來的實作評量目標，從中分析歸納出重要的能力效標，再細分出表現指標。而在我們模擬學生的表現時，發現各表現指標之下，事實上還有一個層次，我們稱之為評分項目。例如，紀錄實驗結果的表現指標下，仍有操作與紀錄的一致性、實驗結果的清晰度以及設計表格等三個評分項目，在各評分項目之下，我們再進一步設定表現水準以及該水準的表現內涵之說明。建構評分規準之分析過程如圖 5 所示：

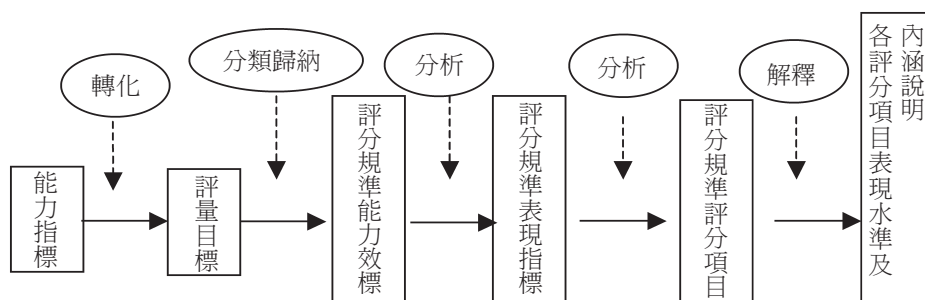


圖 5 評分規準建構分析流程圖

我們想要從文獻中找尋已有的評分規準，可以作為參考的範例，再透過夥伴教師的腦力激盪，建構出一個方便可行的評分規準。我們發現吳欣黛（1998）研究中的評分規準，已建構出清楚的評量內容定義及量尺分數，我們可以修正其評分規準，作為我們評量的工具。修正結果關係如圖 6 所示：

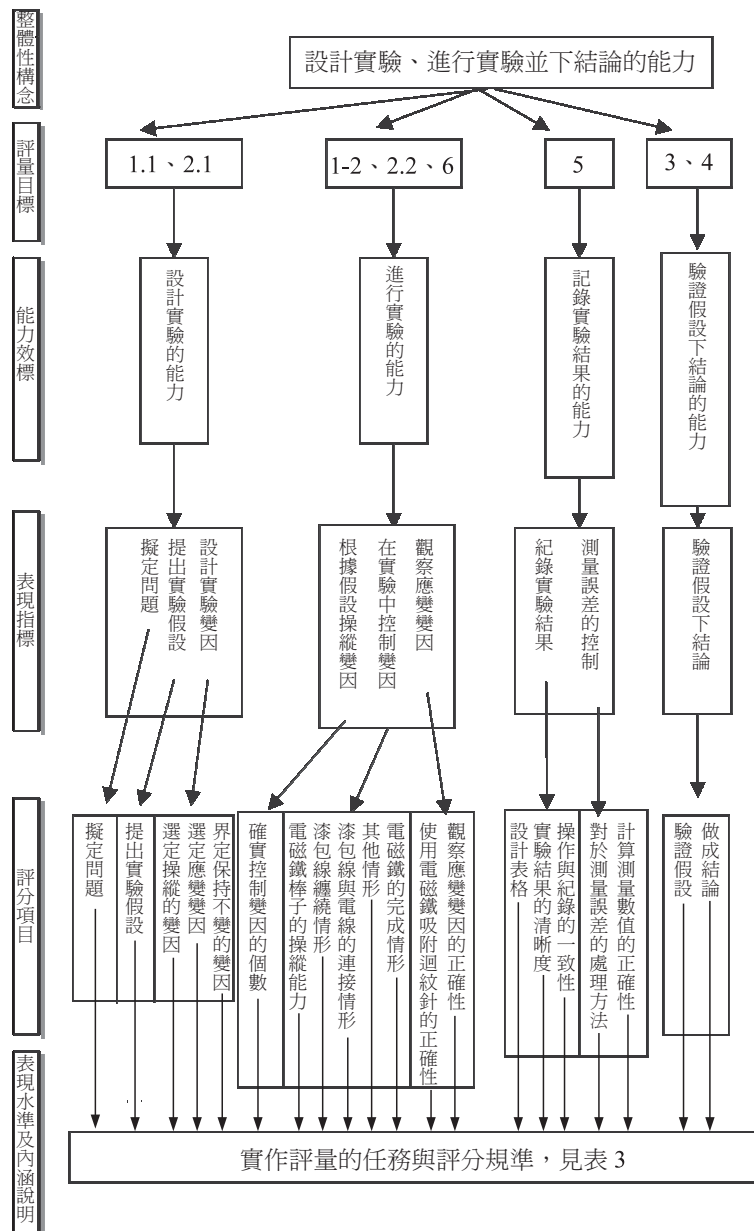


圖 6 第二階段行動結果：評分規準項目與評量目標之關係

(二) 評分規準的品質精進

依據以上的建構過程，行動團隊設計出的評分規準如表 3。在評分規準中，學生每個評分項目的能力達成程度，都可以從規準說明中的五分量尺得分，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，學生也可以明確地瞭解自己在每個評分項目上的表現情形。

另外，每一個能力效標都有一個小計，可以讓學生及教師瞭解各效標的表現情形，也能清楚地回饋能力結構的面向。行動團隊考慮這份評分規準是以評量學生的實驗操作的過程技能為主，所以各項能力的配分比例，以第二項能力——「進行實驗的能力」佔了四十分為最高，而第四項能力——「驗證假設下結論的能力」之判斷依學生實驗的結果而來，所以只佔十分，其他兩項能力則各佔剩餘比例的一半，即第一項能力——「設計實驗的能力」及第三項能力——「紀錄實驗結果的能力」各佔二十五分。

此評分規準除了可以看到得分之外，我們配合質性的評價描述，對學生每個能力效標的表現給予回饋，依學生的得分區隔出五種表現，分別是：「非常棒」、「很厲害」、「還不錯」、「普通」及「要再加油」。以第一項效標能力為例子，其區隔學生表現的方法是，這項目的滿分為二十五分，學生得分在二十二分以上表示「非常棒」，二十二分的計算方式為 25×0.9 ，表示達到滿分的 90% 程度；學生得分在二十分至二十一分之間表示「很厲害」，二十分計算方式為 25×0.8 ，表示達到滿分的 80% 程度，以此類推。其他各個效標項目的計算法相同。

最後的「整體評價」以學生各個評分項目之得分相加，最高可得一百分，配合學生期末成績表現通知單的符號，採甲、乙、丙、丁、戊五等第方式呈現，甲等的學生分數範圍為九十至一百分，乙等的學生分數範圍為八十至八十九分，以此類推。另外，我們亦將學生在四個效標所組成能力結構面向的得分情形，依表現程度將每個能力效標之得分以一百分為分母，重新計算學生得到多少百分比的分數，此處以數字輔以側面圖形來表示學生的表現，方便學生瞭解自己的能力狀況，同時家長亦能清楚地看到孩子效標能力的表現，而教師也能明確地對學生所反映出的表現做出適合的後續教學之判斷。

表 3 「怎樣增強電磁鐵的磁力」實驗進行能力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	評分規準的說明及表現程度
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），但部分與實驗假設不符合 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）		
<input type="checkbox"/> 非常棒（22 至 25） <input type="checkbox"/> 很厲害（20 至 21） <input type="checkbox"/> 還不錯（17 至 19） <input type="checkbox"/> 普通（15 至 16） <input type="checkbox"/> 要再加油（14 以下）		

表 3 「怎樣增強電磁鐵的磁力」實驗進行能力實作評量評分規準表（續）
（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 2-2	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 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 2-2 6	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 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 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 2-2 6	電磁鐵棒子的操縱能力	⑤能準確地將棒子放入吸管中 ④棒子未能完全放入吸管中，但仍能使電磁鐵產生作用 ③棒子未能完全放入吸管中，未能使電磁鐵作用但立即修復 ②棒子未能完全放入吸管中，未能使電磁鐵產生作用且未修復 ①未選擇棒子放入吸管中
1-2 2-2 6	漆包線纏繞情形	⑤纏繞整齊 ④大部分纏繞整齊 ③大部分纏繞不整齊，重新纏繞 ②大部分纏繞不整齊 ①纏繞不整齊
1-2 2-2 6	漆包線與電線的連接情形	⑤連接完全，形成通路 ④連接不完全，但仍形成通路 ③連接不完全，形成斷路但立即修復 ②連接不完全，形成斷路且未修復 ①未注意連接情形
1-2 2-2 6	其他情形 （教師略述）	⑤ ④ ③ ② ①
1-2 2-2 6	電磁鐵的完成情形	⑤形成通路，且連接正確 ④形成通路，但連接有瑕疵 ③形成斷路，但馬上檢查並修正 ②形成斷路，且未修復 ①未注意完成情形

四
擇
三

表 3 「怎樣增強電磁鐵的磁力」實驗進行能力實作評量評分規準表（續）

觀察應變變因		
1-2 2-2	使用電磁鐵吸附迴紋針的正確性	⑤能正確利用磁極吸附迴紋針 ④能正確利用磁極吸附迴紋針，但吸附不確實 ③未能利用磁極頂端吸附迴紋針，但隨即修正 ②未能利用磁極頂端吸附迴紋針，但不至於影響結果 ①未能利用磁極吸附迴紋針
1-2 2-2	觀察應變變因的正確性	⑤能正確數出迴紋針數量，並確實紀錄 ④能正確數出迴紋針數量，但未紀錄 ③未能正確數出迴紋針數量，但有紀錄 ②未能正確數出迴紋針數量，且未紀錄 ①不知要數迴紋針數量
小計：進行實驗的能力得__分（40） <input type="checkbox"/> 非常棒（36 至 40） <input type="checkbox"/> 很厲害（32 至 35） <input type="checkbox"/> 還不錯（28 至 31） <input type="checkbox"/> 普通（24 至 27） <input type="checkbox"/> 要再加油（23 以下）		

（三）紀錄實驗結果的能力

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，但部分設計有誤 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未紀錄實驗結果
5	操作與紀錄的一致性	⑤實驗紀錄與實際操作結果相符合，且和正確結果相符 ④實驗紀錄與實際操作結果相符合，但和正確結果不符 ③實驗紀錄與實際操作結果不符，但近似正確結果 ②實驗紀錄與實際操作結果不符，也和正確結果不符 ①未紀錄實驗結果

表 3 「怎樣增強電磁鐵的磁力」實驗進行能力實作評量評分規準表（續）

表現指標	評分項目	評分規準的說明及表現程度
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 ④計算的平均值正確，但未用四捨五入法表示迴紋針數量 ③計算的平均值有誤，但不影響結果 ②計算的平均值有誤，且會影響結果的正確性 ①未計算平均
小計：紀錄實驗結果的能力得_____分（25） <input type="checkbox"/> 非常棒（22 至 25） <input type="checkbox"/> 很厲害（20 至 21） <input type="checkbox"/> 還不錯（17 至 19） <input type="checkbox"/> 普通（15 至 16） <input type="checkbox"/> 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

表 3 「怎樣增強電磁鐵的磁力」實驗進行能力實作評量評分規準表（續）

表現指標	評分項目	評分規準的說明及表現程度
小計：驗證假設下結論的能力得_____分（10）		
<input type="checkbox"/> 非常棒（9 至 10）		
<input type="checkbox"/> 很厲害（8）		
<input type="checkbox"/> 還不錯（7）		
<input type="checkbox"/> 普通（6）		
<input type="checkbox"/> 要再加油（5 以下）		

（五）整體評價

（一）設計實驗的能力：25%	
（二）進行實驗的能力：40%	
（三）紀錄實驗結果的能力：25%	
（四）驗證假設下結論的能力：10%	
整體評價： <input type="checkbox"/> 甲等（90 - 100） <input type="checkbox"/> 乙等（80 - 89） <input type="checkbox"/> 丙等（70 - 79） <input type="checkbox"/> 丁等（60 - 69） <input type="checkbox"/> 戊等（59 以下）	

表 3 分成四個能力效標的分析性評價，及學生表現的整體評價。我們在行動的第一階段分析過程技能能力要項之後，將能力指標轉化成評量目標，並將評量目標歸納成四個進行實驗時的能力效標。在行動的第二階段確定這次的評量任務之後，我們依據任務的內涵，配合四個能力效標，分析出表現指標以及評分項目，並在行動的第三階段，彼此分享教學評量的經驗，不斷的進行討論及修正，逐漸建構出表現水準五分量尺的界定及內涵說明。

表格設計以轉化後的評量目標為依歸，但在使用表格觀察學生表現時，則以具體的表現指標為觀察項目，教師使用本表格時，可以清楚地知道，學生在進行操作實驗時，就具體指標以及評分項目而言，表現如何，是否達成目標及具備了

能力。教師於評分時觀察學生的操作表現給予勾選適當的分數，並將表現的得分統計在各個能力效標的小計中，讓學生清楚瞭解自己在該項能力的表現情形。最後的整體評價，呈現各項能力效標的得分百分比，瞭解自己在各項能力的表現，並以五等第方式表示本次設計並進行實驗之過程技能的整體表現情形。

(三)二階段實作規畫

我們採取兩階段任務規畫，目的是給學生修正設計實驗之機會，並讓教師準備實驗器材。

1. 第一階段實作：需時兩節課

第一節課中，首先我們讓學生閱讀以下一段問題情境，並由教師加以說明。學生有一節課的時間來設計實驗，包括擬定問題、提出假設及選定變因，並將思考流程寫在學習單上。教師收回學習單後，進行第一項能力效標——「設計實驗」之批改。教師批改之後，利用下一節課的時間與全班學生討論並鼓勵學生進行實驗設計的修正，並提醒同學根據自己的實驗設計準備恰當的實驗器材，例如：依自己的假設纏繞漆包線於不同材料的棒子中，或是纏繞不同的圈數，及將漆包線與電線接觸之處刮乾淨。

2. 第二階段實作：需時兩節課

第二階段時期，學生可根據自己設計的實驗進行操作、紀錄結果與下結論。教師可以同時評量四位學生，以有效利用評量的時間。一次評量四位學生的狀況，我們曾模擬試作過，由於我們幾位都是資深的自然科老師，一雙眼睛關照四個位置的學生進行實驗，相當順手。同學的進度亦有快有慢，提早結束者會退位讓下一位開始進行。由於學生的假設是不受限制的，在控制變因的操作時，評分表中特別預留一個保持彈性的評分項目，教師可依據現場情形在四個評分項目中選擇三個項目進行評分，以符應不同的設計。在學生操作完畢紀錄完成之後，教師依據學生的記錄進行第三項能力效標——「紀錄實驗結果」及第四項能力效標——「驗證假設下結論」的評分。最後統計分數分析學生整體表現並描繪側面圖。

陸、結語—我們行動任務的價值

行動團隊在建構評分規準的同時，不斷以 Quellmalz (1991) 的所提示的優良評分規準六個原則進行檢視，期能符應這些原則。我們期望在接近真實的情境中，評量出學生真實進行實驗的能力，並使用淺顯易懂的語言詞彙，讓學生、教師及家長清楚地看到學生的能力表現、瞭解規準所代表的意義，而不只是學習表現的缺失。國內已有許多論文以實作評量進行研究，大多研究的重點在實作評量實施過程及成果上，有些則對於實作評量的編製過程及所遭遇的困難加以討論。然而，詳細觀之，這些實作評量工具的編製，有些雖有經過信、效度的檢測，但是，對於實作評量評分規準的建構過程，並未加以釐清。就一個實務工作者而言，如何編製一份有創意、且能現場實施、並且讓學生及家長一目了然表現狀況，而又賦予學習意義的實作評量，應該是相當重要的。實作評量的研究已備受重視，然而，一般教師如何結合各階段能力指標進行評量，在執行的技術與能力上，仍有精進之空間。本研究乃以自然與生活科技領域中，設計並進行實驗之過程技能的能力指標為核心，藉由評分規準的建置，設計出一個具回饋性的二階段實作評量任務。我們期許學生於評量進行過程中，能因為二階段實作評量的實施，可於實驗設計階段有回饋修正的學習，而我們更期望同儕在合作行動過程中，彼此滋養智慧，達成教師研究社群的專業實踐。

參考文獻

中文部分

- 王美芬等(2004)。**國小自然與生活科技六上課本**。台北市：康軒。
- 余民寧(2002)。**教育測驗與評量：成就測驗與教學評量**。台北市：心理。
- 吳欣黛(1998)。**實作評量在效度的上的真實性與直接性**。國立台北師範學院國民教育研究所碩士論文，未出版，台北市。
- 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。
- 吳毓瑩(2004)。**效度**。載於王文中、呂金燮、吳毓瑩、張郁雯、張淑慧著，**教育測驗與評量：教室學習觀點**(頁47-80)。台北市：五南。
- 呂金燮(2004)。**實作評量——理論與應用**。載於王文中、呂金燮、吳毓瑩、張郁雯、張淑慧著，**教育測驗與評量：教室學習觀點**(頁173-239)。台北市：五南。
- 李坤崇(1999)。**多元化教學評量**。台北市：心理。
- 李淑娟(2003)。**實作評量對國小五年級學生科學態度與自然科學習成就之影響**。國立台中師範學院教育測驗統計研究所碩士論文，未出版，台中市。
- 席家玉(2002)。**國民小學生活課程實作評量之研究**。國立台中師範學院自然科學教育研究所碩士論文，未出版，台中市。
- 桂怡芬(1996)。**自然科實作評量的效度探討**。國立台北師範學院國民教育研究所碩士論文，未出版，台北市。
- 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**。國立台北師範學院課程與教學研究所碩士論文，未出版，台北市。
- 高新建(2001)。**基本能力的課程與教學轉化**。**台北市立師範學報**，32，237-254。
- 張麗麗(2002)。**評量改革的應許之地，虛幻或真實？——談實作評量之作業與表現規準**。**教育研究月刊**，93，76-86。

- 教育部（2003）。**國民中小學九年一貫課程綱要**。台北市：教育部。
- 莊明貞（1996）。國小社會科教學評量的改進途徑——從「真實性評量」實施談起。**教育資料與研究**，**13**，36-41。
- 許民陽（2002）。**自然與生活科技第二階段教科書對能力指標落實初步分析**。國立編譯館，國民中小學九年一貫課程綱要能力指標研討會。
- 許民陽、林麗詩（2004）。自然與生活科技教科書之能力指標應用及評量設計初探。**教育研究資訊**，**12**（1），77-96。
- 陳怡如（1999）。**實作評量在國小數學科之應用研究**。國立台中師範學院教育測驗統計研究所碩士論文，未出版，台中市。
- 陳明印（2002）。基準評量表及其在實作評量上的應用。**研究資訊**，**19**（1），60-77。
- 陳政良（2002）。評量規準（Rubrics）在生活科技教學評量上之運用。**生活科技教育**，**35**（1），2-9。
- 陳學淵（2003）。**國中教師發展實作評量之探討——以自然與生活科技與教師為例**。國立彰化師範大學科學教育研究所碩士論文，未出版，彰化市。
- 彭森明（1996）。實作評量理論與實際。**教育資料與研究**，**9**，44-48。
- 楊銀興（2000）。**傳統評量與新式評量之比較及國小教師對實施新式評量相關問題覺知情形之研究**。國立台灣師範大學教育學系博士論文，未出版，台北市。
- 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。
- 甄曉蘭（1995）。合作行動研究——進行教育研究的另一種方式。**嘉義師院學報**，**9**，298-318。
- 蔡菁玲（2001）。**國小一年級數學領域實作評量之行動研究**。國立花蓮師範學院國小科學教育研究所碩士論文，未出版，花蓮市。
- 盧雪梅（1997）。實作評量的應許、難題和挑戰。**現代教育論壇**，**4**，3-9。
- 簡琇芳（2002）。**實施實作評量對國小學童自然科學習成就之影響**。國立台中師範學院自然科學教育研究所碩士論文，未出版，台中市。
- 蘇義翔（1998）。實作評量的理論與啟示。**測驗與輔導**，**149**，99-102。

西文部分

- Airasian P. W. (1997). *Classroom assessment*. New York: McGraw-Hill.
- Brualdi, A. (1998). *Implementing performance assessments in the classroom*. (Report No. EDO-TM-98-06) Washington, ERIC Clearinghouse on Assessment and Evaluation. (ERIC Document Reproduction Service No. ED423312)
- Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review*, 64, 5-30.
- Janesick, V. J. (2000). *Curriculum trends: A reference handbook*. Santa Barbara, CA: ABC-CLIO.
- Linn, R. L., & Gronlund, N. E. (2000). *Measurement and assessment in teaching* (8th ed.). Upper Saddle River, NJ: Prentice-Hall.
- Quellmalz, E. S. (1991). Developing criteria for performance assessments : The missing link. *Applied Measurement in Education*, 4(4), 319-331.