科學教育學刊
2005, 第十三卷第一期, 1-27

Chinese Journal of Science Education
2005, 13(1), 1-27

科學教學評量的反思與實踐：教師行動研究

林曉雯[1]　**陳佩君**[2]
[1]國立屏東師範學院
[2]高雄市立中山國小

（投稿日期：民國 92 年 9 月 25 日，修訂日期：93 年 3 月 25 日，接受日期：93 年 5 月 7 日）

摘要：本研究探究一位國小自然科教師進行行動研究以實施多元評量的歷程中，如何藉由反思，指引行動以獲得教學專業的發展。資料來源包括：教師研究日誌、教學錄影帶、相關文件、教師對學生訪談記錄，以及行動研究小組會議之記錄。針對個案教師進行行動研究過程中對教學評量的理解與實務的演變進行分類、比較及信效度的建立。研究者據此資料形成四個主張，以詮釋個案教師多元評量的歷程的反思與實踐。(1)反思評量的目的，引導教師建立評量制度、破除分數迷思，體認教學目標、課程與評量交互檢核的角色。(2)反思能力的教學與評量，教師全力投入能力指標的解讀、教學與評量，重新反思能力的意義與分類。(3)科學本質的反思與再建構促使教師建立科學探究歷程教學評量模式。(4)反思學生後設認知的能力引領教師再建構自我評量的意義。

關鍵詞：行動研究、國小自然科教學評量、教師反思、專業成長。

壹、緒論

國內一波波教改的推動，從解開僵化的束縛、注重適性發展、到十大基本能力的訴求中，可看出教育目標愈著重個性、潛能，培養帶得走的多元能力（教育部, 民 87）。評量與教學息息相關，在教學改革的歷程中，如果我們仍沿用傳統的封閉式測驗，教學上無可避免的，將會導向迎合封閉式測驗的教學型態（黃秀文, 1996）。因而，基於教育目標的多元，評量的改革是必要的。

透過評量引導教學內容的過程來推動教改是當今許多教育改革的重要方式，因此評量的改革成爲世界各國及我國教育改革的重要項目之一（吳鐵雄和洪碧霞, 2000）。現代科學教學評量呼籲教師能評量最有價值的事項、系統的知識、理解與推論、評量學習成就與學習機會，以及邀請學生參與評量等（王澄霞, 1997）。無疑的教學評量能力是教師重要的教學能力之一，全美教育基金會出版的「美國國家科學教育標準」（National Science Education Standard）

（NRC, 1996）一書，明確訂定科學教育的評量標準，以爲科學教師參考。然而大部分的自然科教師仍慣於使用紙筆測驗的形式，教學與評量並無緊密的結合，評量以給分爲主要目的，評量的內容以事實知識爲主，極少注意信、效度的議題（Stinggins & Bridgeford, 1985; Stinggins, Conklin & Bridgeford, 1986）。國內大部分的教師對新的評量認知似乎不佳，偶爾會使用新的評量方式（楊銀興, 2000），可見教學評量是一重要且值得研究的主題。

貳、文獻探討

在此教育改革聲浪中，教學理念、策略的改變促使教師需不斷的進修，進行終身學習。教師可藉參與相關活動如演講、大學進修、專題討論、電腦網路遠距教學、工作坊、或研讀研究期刊等各式各樣的進修機會促成專業發展。然而這些進修均屬外在政策或學者專家制訂方案或策略，基層教師被動執行的取向，這種由上而下或由外而內的改革取向，較忽略現場教師的信念、想法、知識體系，故較難獲得現場教師的認同，不易在教室中產生長期且重要的改變（Duffy & Roehler, 1986）。

近十年有愈來愈多的教師投入教師即研究者的行列，著重以行動研究促使教師對自己的教與學進行探究，進而促成教師專業成長（例, 歐用生, 1999; 蔡清田, 2001; Hensen, 1996; Tobin, 1989）。Hensen（1996）指出教師進行行動研究，使教師對改變的內容、範疇、策略、進度更具選擇與主導權。行動研究可分爲個人獨立的自我研究，與同儕或他人合作進行研究（Calhaun, 1993），無論何種形式，對參與研究教師之教學可產生一些正面的影響，例如，教師們指出他們對教學實務有更深一層的認識，願意嘗試新的教材或教法，及不斷學習如何教學（王靜如, 2000; 林曉雯, 2001; 郭重吉, 1999; 楊宏衍和段曉林, 2001; 熊同鑫, 1999, 2001; Briscoe & Wells, 2002; Feldman, 1996; Tabachnick, 1999），甚至參與教師對自己都有更深一層的瞭解（熊同鑫, 1999; Baumann, 1996; Tabachnick, 1999）。

教師進行教學行動研究雖然名爲研究，實則教學本身即爲一種實驗性的探究活動（Stenhouse, 1975），教師進行教學探究活動時，教師個人所具有的專業知識、能力及經驗對教師專業成長皆有重要影響。教學探究活動不但提供教師形成教與學的知識的機會，並達到教師增權的效果（Cladinin & Connelly, 1995）。教師行動研究是一種自我反省及探究的歷程，教師運用各種不同的詮釋研究法，理解並解決各種關於個人教學情境及學生的問題。教師可試用新的理念，評估其可行性及反思其個人關於教與學的信念，讓教師能批判性的檢視其教學實務；試行的結果如果更能符合其教學情境或學生需求，則此理念將成爲教師進行教學決策的知識基礎（Richardson, 1994, 1996）。當老師經由研究獲得愈多知識，他們就愈能自主地且不斷地檢視其提供給學生的學習經驗，並質疑理論與實務的當然性（Lytle & Cochran, 1990）。研究亦發現：當個體越能獨立於外界的知識權威，而趨向於自行建構個人知識，則其愈能主動投入，進行研究，並爲其個人經由行動研究歷程中所發展之知識所影響（King & Kitchener, 1994）。

反思是行動研究中教師專業成長的主要原動力（Elliott, 1991）。Schon（1983）對反思實踐的系統分析影響甚巨，以行動研究的原則觀之，他的「隱含式行動中知曉」（Tacit knowing-in-action）說明例行事務，是一種和教學任務相關的精鍊知識，涵蓋覺知的問題、有關問題解決條件的資訊與解決問題步驟的知識。其提出的「行動中反思」（Reflection in action）說明教師面臨新且複雜的情境或例行

性行動爲令人困擾的問題所干擾時所進行的行動，是一種對情境所進行的反思對話。當教師需要明白的用口語或文字建構或形成知識時，會由行動中抽離，並對行動進行反思，以增加分析、重組知識的能力，並增加知識的可溝通性，這便是「對行動反思」(Reflection on action)。行動研究者需具備「隱含式行動中知曉」、「行動中反思」、「對行動反思」的能力，才能有效的進行日常實務工作、因應複雜的問題與改變實務。

由於教師是學校的一份子，教師產生的知識不可避免的受到學校中其他教師、行政人員、家長及學生影響，因此「合作」是行動研究中的重要部份。在與他人合作中，教師即行動研究者可獲得支援，可分享經驗或給予協助，這些都是行動研究成功的重要影響因素(甄曉蘭, 1995; Hesen, 1996)。更重要的是，教師在合作的歷程中能互相討論行動研究中「改變」對教與學的眞實意義 (Briscoe & Peters, 1997)。成員間經由互相討論及對話可促成認知結構的成長，或對研究中的相關問題有新的解決策略，這點呼應了 Vygotsky (1978) 心智發展的看法，他認爲人類的高階層心智功能源自社會歷程，心智功能在與他人互動的社會活動中，藉著語言傳遞意念、啓發抽象思考、進而促進心智功能的發展。教師可與學校的同事一起合作進行行動研究，然而學校同事間有時因時間不易磋合或理念的差異，不一定容易建立合作的途徑，此時源自學校以外的支援便具有決定性的影響力 (Little, 1990)，如大學的研究者可能是行動研究者在研究歷程中的重要資源與合作伙伴(例, 王靜如, 2000; 林曉雯, 2001; 郭重吉、江武雄和張文華, 1999; Briscoe & Wells, 2002; Watts, 1985)。

傳統評量是植基於二十世紀初的心理學理論—行爲主義。行爲主義認爲可以將能力分解以精熟學習，因此評量便將能力一一分解，並一再反覆進行測驗；又因能力是非情境化的，因而評量可以抽離情境來舉行。建構主義則主張學習是學生自己主動建構他們的知識，並強調較高層思考能力的培養、注重過程而非內容 (Cunningham, 1998; McGhee, 1998)。基於建構主義的主張，評量與學習過程應是融合無間的，是一種動態不斷重新建構的，以協商達到共識的過程；其中參與評量的人員以及評量內容，都不是絕對或一成不變的，而是漸續生成，以求獲得更深入且更寬廣圓融的結果 (張文華和郭重吉, 1995)。

現今的學者多建議教師使用多元化評量，其精神在於重視學生個人內在的學習改變、瞭解個人與環境的互動關係，著重個體之間的差異性，給予合適的評量程序，藉此讓每個學生都體會到自己的價值而適才適性的發展 (侯雅齡, 2000)。綜合學者們 (王文中, 1999; 李坤崇, 1999; 張惠博和黃文吟, 2000; 鄭淵全, 2000; Glaser, 1990; Hart, 1994; Puckett & Black, 1994) 對多元化評量的看法，可知多元評量具備下列特色：(1)就評量的模式而言，它強調教、學、評量同時並進，認爲評量可以看成一種教學或學習活動。(2)評量的目的是了解、支持、激勵學生學習，以導引學生有更佳的表現、具備更多的能力。(3)評量的內容強調學生學習的過程，能實際應用的能力、努力、進步的情形，且是課堂中習得的。(4)評量方式是多元化的動態方式，有時學生可選擇評量類型。(5)評量設計較傳統評量耗時、費力、較主觀；使用較持續、彈性，在非標準化的眞實情境中實施。(6)結果評定方面，其評量基準多元、廣泛；評分者可爲教師、學生、家長…；成績單除單一分數或等第外增列文字描述。(7)在評量情境中，教師爲參與者，學生則扮演主動的參與者、能自我調節的學習者。

近年來相關的評量實徵研究，在教師教學上，Howe 等人 (1995) 之研究中，發現這種

評量方式的改變能鼓舞教師採用適當的教學策略，有利於教師們設計較爲深入的課程及教學計畫。美國加州的教育部門對新式評量進行一項合作性計畫研究（California Assessment Collaborative, 1993），亦顯示教師比較能夠了解學生學習的多樣化及個別差異，能使教師對學生表現建立較適切的期望。國內研究亦發現實施實作評量有助於教師診斷學生的學習困難，並讓教師從以往單向性的教學習慣，轉變爲提供學生機會以思考課本的內容以及討論實驗的程序、評量標準（桂怡芬, 1996）。

在學生學習方面，美國加州教育部門的研究（California Assessment Collaborative, 1993），發現新式評量因爲學生實際投入學習活動的機會增加，學習的成績比較好，並能增加學生自我評量的能力，提高學生爲自我學習負責的意願。國內相關研究，例如田麗娟（2001）於國小自然科採用學習檔案評量，發現學習歷程檔案所呈現的內容是豐富而多元的，可以使學生瞭解自己的學習興趣、困難及成長歷程，修正上課態度，並得到學生的正面回饋。辛慶偉（1998）亦同樣提出學生能經由檔案展現出其多元的面向。戴君佩（2001）以準實驗研究法進行檔案評量對學生學習動機的影響，提出檔案評量比傳統紙筆測驗更能增進學生的學習動機。在實作評量方面，陳聖泰（2000）在國小自然科「電磁鐵」單元實作評量應用之研究中，提出學生對實作評量方式具正向態度。桂怡芬（1996）所作的研究發現，學生的學習態度變得比較積極，操作實驗的能力進步了，也改變了準備考試的方式及學習的焦點。另外，在台北市國民小學試辦教學及評量改進班之評鑑研究中，曾嘉琪（1996）的研究指出教評班學生比普通班學生更喜愛學習、更主動學習、更能夠自我悅納，在人際間更感到和諧融洽。這些研究均顯示實施不同於傳統的新評量方式，對於教師的教學角色、教學計畫、教學策略，以及學生的學習角色、學習態度、學習動機上，均有正面的助益，能促進教與學的品質。

基層教師面對實施多元評量以呼應課程與教學的革新，教師要問的問題是：我要評量什麼？我要選擇哪一種評量方式？我的評量眞的評測到學生的學習情形嗎？我的評量能幫助學生學習與回饋教師教學嗎？而要回答這些問題的最好答案，莫過於讓教師在自己的教室中進行研究來尋求解答。

叁、研究目的

本研究旨在探討個案教師於所任教的中年級自然科教學中進行教學評量行動研究歷程中，如何藉由反思，指引行動以獲得教學專業的發展。文間同時描述個案教師實踐多元化評量所遭遇的問題與因應策略、實踐的歷程與轉變、及實施後的影響；據此個案教師得以澄清對多元化評量的意義、建構教師個人的專業實踐理論，以解決實務中的問題，進而提昇學習與教學的品質。研究結果除可提供教育現場教師進行教學改革的參考，此研究亦可了解教師如何在教室中進行行動研究時面對的問題，及教師如何持續不斷的進行探究，以促成教師專業的成長，即對「教師即研究者」的觀點有進一步的理解。

肆、研究方法

一、個案教師王老師的背景

個案教師王老師，女性，係師院數理教育學系畢業，在實習那年曾擔任一個半月的自然科任教師後，一直擔任級任老師的職務。擔任級任老師時，王老師的教學重心是「班級經營」，王老師與家長們共同合作，以「主動創新、快樂學習、愛我自己、感恩他人」爲教學

主要目標。王老師與學生間保持著亦師亦友的關係。王老師曾參加「國小多元化評量種子教師研習」，並將其部分理念與技術在教學中實施；教學評量大部分以紙筆測驗、習作爲主要的評分依據。

任教十年後，王老師進入師院數理教學碩士班就讀，陸續接觸了科學教育的相關理念，例如建構主義，對這些理念很感興趣；再加上碩士班同學中有二位可稱得上是自然科專家教師的互相切磋下，王老師毅然決然地投入自然科的教學。於是，王老師於所服務的國小主動爭取兼任組長的職位，以獲取擔任自然科教學的工作。成爲自然科教師後，王老師以準備課程、設計活動爲努力的方向，期許自己能營造有趣、探究的學習情境，引導孩子們主動建構知識及培養科學素養。此外王老師亦先後參加「提昇國小教師自然科教學成效」教師成長團體（林曉雯, 2000），及國教輔導團自然與生活科技組，希望能發展兼顧理論與實務的專業智能，以持續從事自然科的教學與探究。王老師考量行動研究能促成教學的改變，故決定投入行動研究的行列。

二、研究情境

王老師所服務的學校位高雄市市郊約 100 班的大型學校，學生都爲常態分班。學生家庭社經背景大部分屬中上階級，家長們大都關心孩子的教育情形，但仍較重視智育成績。一般而言，教師們認爲本校學生較市中心學校的學生的生活單純、可塑性高，但過於活潑好動，教師認爲須加強學生的生活教育。現任校長今年剛到任，採溫和的形式作風，整個學校正在規劃發展的方向與特色。雖面臨教改的風潮，但學校大部分教師仍依照過去的方式教學，只有少數教師會在自己的班級進行個別、小部分的試用，例如三年級部分的級任老師們開始試行統整課程，是學校裡最有活力的一群教師。學校中的自然科科任教師多兼任行政人員，忙於行政業務，自然科教師間也少因教學或課務進行討論與溝通，目前尚無新的教學措施或改變。由於學校班級數多，校舍不夠使用，以三年級自然科爲例，是由四位任教三年級自然科的老師輪流使用一間自然科專科教室，因此三年級學生的自然課輪流在自然科教室進行學習活動，其他時間在原班併桌上課。

在 89 學年度，王老師擔任三年級四個班的自然科教學。王老師在這四個班級，均同樣實施多元化評量，所有的學生均爲王老師的研究伙伴，但在時間考量下，只對固定一個班進行教學錄影。在 90 學年度時，由於王老師兼任輔導員，只教一個班級，因而便繼續教授原觀察班級四年級的自然課程。這個班級的學生程度中庸、整齊，沒有資源班的學生，學生的上課情緒與秩序穩定。王老師進行教學研究流程前後歷經約一年半，教學是由三年級上學期始，至四年級下學期第一次學月考試的期間（如表 1）。

三、研究者的角色與合作關係

王老師參加本文第一作者輔導成立之「提昇國小教師自然科教學成效」（林曉雯, 2000）之教師成長團體，與團中老師進行自然科教學理論之研讀、教學經驗交流、教學觀摩，其後王老師以教學評量爲主題與本文第一作者進行合作行動研究。第一作者的主要任務是透過每次的研究會議，協助王老師發展行動研究的任務，協助詮釋資料，提供研究建議或諮詢，並指導寫作。會議進行中王老師說明教學計畫、實施情形、觀察或收集資料的結果、遭遇的困難進行說明；會議中視需要播放教學錄影帶、呈現訪談筆錄、相關文件以輔助說明。第一作者及討論會其他成員則提出問題協助王老師釐清問題及行動背後的意圖、確認計畫的可行性、凝聚觀察的焦點或拓展觀察的深度及廣

表1：王老師教學研究時間與單元名稱

時 間	授課年級學期	單元名稱
890901~900131	三年級上學期	座標的方向
		奇妙的光
		植物的根莖葉
		動物的身體
		溫度與保溫
		水的三態
		空氣的流動
900219~900630	三年級下學期	養蠶
		認識人體
		植物的花和果
		水流的作用
		天氣的觀測
		輕與重
900901~910118	四年級上學期	雲和雨
		植物的蒸散教作用
		月亮出來了
		陰溼環境中的生物
		溶液
		電路
910218~910430	四年級下學期	熱的傳導
		時間的測量

度、提供其他可能的解決方法，促使王老師對各項行動進行反思，以及效化研究資料等。王老師對研究進度、方向、投入、設計、分析、繕寫都有很深的投入，由於王老師的獨立研究能力、反省與實踐能力使本個案在此團隊衆多同時進行的行動研究中成果豐碩。王老師進行行動研究中，研究所老師、同學、學校及輔導團同事、研習會或工作坊的演講人皆成爲諮詢的主要對象，王老師完全以一個學習者的角色投入此研究。在本研究中，王老師以教師研究者的理念進行研究，即教師不但是教學實務工作者、應用者、亦是批判反省的實踐者。

四、資料蒐集

本研究持續兩年，期間收集的相關資料包括：成長團討論會現場筆記、教師教學日誌、教案文件、相關報告、學生或家長的相關資料，教學錄影等。試圖由個案教師的觀點，探討促使王老師在進行行動研究的歷程中理解的改變的重要影響因素。

(一)成長團討論會現場筆記

八十九年九月至九十一年六月間教師成長團持續每週進行一次研究討論會，會議中研究者與個案教師個別或成長團成員多人同時參與討論，以探究王老師對評量進行改變的原因、行動的假設、教學情形及由資料分析中所獲得的結果及省思。此外討論會能協助王老師釐清行動研究的問題、促進計畫的周全性、提供成員間分享的機會、收集資料的再確認、評量與對不同層面進行反思。這些會議討論資料以現

場筆記的方式收集。

(二)研究日誌

研究日誌是王老師進行行動研究最重要的收集資料方法之一。王老師在日誌中記載(1)教學前各教學單元的教學設計、教學策略、評量計畫編製與實施使用、後果運用的構想與過程。(2)課堂事件發生的情境、非結構式的課堂觀察(如：學生的學習情形)、實踐上的感想、發現、困境與反思。(3)閱讀相關文獻的心得、突來的想法…等。(4)同儕對話：個案教師把握機會不定期地與專家及同儕進行對話，以澄清、分析自己的想法或問題，或是從中獲得新的觀點、批判、支持或建議。對話方式除了面對面談話外，亦透過電話或電子郵件進行。通常王老師會當場做簡單的記錄，事後儘快的詳實筆錄在日誌中。(5)學生訪談資料：王老師欲瞭解或確定學生的感受、觀點、解釋…時，會視目的而採隨機或立意取樣，進行非正式訪談或正式訪談，並於事後儘速記錄。隨著時間的進展，它呈現了個案教師在不同階段對研究主題之不同知覺、洞察與實踐的發展過程。

(三)文件資料

文件資料包含研究日誌以外的其他教師文件、學生文件、家長文件等三種類型的資料。教師的文件資料包括：教學計畫或準備資料、教具、教學評量設計或流程、評量用具(如評量能力貼紙)、評量結果分析、給家長的信…等。可以看到教師的試題編製與設計情形、教學評量設計或流程、評量結果使用的情形、對遭遇問題所作的部份因應措施。此外，個案教師的行動研究報告在本文中有重要意義，它提供視爲個案教師實施多元評量行動研究中產生研究問題、計畫、實施、評量、再計畫的歷程，是教師描述其專業成長的完整圖像與最佳範例。學生的文件資料包括：學習狀況通知單、能力記錄表、小組學習態度行爲評量表、實作評量表、學習單、評量單、試卷、自評互評表、檔案、作文、提問紙箋、查詢資料、習作、…等。家長的文件資料包括：成績單上的意見欄、家長給老師的短箋。從學生、家長的文件中，可以瞭解學生的學習狀況，對實施多元化評量的意見與感受。

(四)教學錄影

將個案教師從89年9月到90年4月期間，大部份的教學單元進行錄影，錄影機架設在教室後方，拍攝個案教師大部分的教學情形，及部分學生的上課情形。錄影的焦點著重在蒐集教師在評量上的實施使用及後果運用之情形，在教學上使用的教學策略、教師扮演的角色，以及學生對評量及教學的反應等。教學錄影帶的資料主要是協助個案教師較完整的回憶教學實況、觀察學生表現，以便較詳實地繕寫教學日誌，並爲研究討論會的討論及反思的參考素材，亦是建立研究效度的策略之一。

五、資料分析

將蒐集到的資料整理成一個個小單元，或將重要的地方用顏色做記號。將小單元或做記號處的每個段落(包含情境)歸入預設的類別中，或衍生新的類別，並不斷修正比對；每一個小單元或段落可能同時歸於不同類別。這些類別是評量計畫編製、評量實施使用、評量後果運用、教師教學活動、學生學習情況、學生感受、遭遇問題、研究者的省思與成長，而在這些類別下面又有子類別，以評量後果運用爲例，其子類別是評量規準、批改、回饋等。將各類的資料進一步的歸納、編碼、分析其特性。不斷的進行批判分析，並透過在實務中的反思與檢視，進行更進一步的測試，比較各類資料間的關係。主要的議題主張及支持的証據因而產生。依據不同的資料來源檢視其支持或拒絕這些主張，並在逐次的會議中討論、修正，以反映行動研究促成專業成長的意義之瞭解與建構，並進一步重新檢視資料分類及形成主張的

適當性。主張形成是在主題、類別、暫時性主張間來回進行演繹性思考，並以三角校正方式確認檢驗之。即循(1)重複閱讀已收集之資料，(2)尋找可能的主題、解釋、概念，(3)形成暫時性主題，(4)建立分類架構，(5)分析比較類別間的關聯性，(6)形成主張，不斷循環的過程中，最後提出本文的主要主張。

六、研究信效度

王老師從事教育行動研究，秉持倫理性、嚴謹性與邏輯性、實用性、及審美性（陳惠邦, 1998)等四個效標，不只強調技術方法的有效，更重視其目的企圖與結果的影響；王老師經常以上述的四項效標作自我檢核，並詳述研究歷程，以透過協同反省及研究結果的報告呈現，讓批判朋友及讀者來共同檢核。

質化研究相信實體是多元而複雜的，並強調研究者與參與者間的密切關係，因此在本研究中，研究者進行達兩年的長期合作互動，持續收集、分析王老師的教學與評量實施歷程、產生的影響，並在研究中以開放的心態，記錄相關的資料，包括反面或未曾預期的發現，採用不同的方法、資料來源、理論來交叉驗證資料「可信性」。研究者透過自我反省及與研究會議多位成員對話批判來加以檢核，以減少研究者個人的主觀與偏見，以提高研究的「中立性」與「驗證性」。並將所有資料詳實記錄，並將原始資料、資料分析和統整的結果，妥善保存，以供他人的稽核，以提昇「可靠性」。此外，研究者詳實介紹個案教師的背景經歷與教學理念、研究學校與班級等情境，並將個案教師的信念、行動計畫與歷程、自我與協同省思的內涵，明示化，以提昇研究之「遷移性」。

伍、結果

研究結果以四個主張呈現個案教師進行行動研究歷程中如何藉由反思，指引行動以獲得專業知識的發展。

一、主張一、反思評量的目的，引導王老師建立評量的技巧與制度以破除分數迷思，體認教學目標引導課程與評量及三者互相檢核的角色，若沒有高品質的教學活動，更奢言高品質的評量

擔任自然科任教師後，許多自然科教學問題不斷產生，出乎王老師的料想，也打亂了原定的計畫，這些問題包括：上課流程不順暢、學生不配合老師的指示，上課秩序亂七八糟，王老師自述道：

第一次帶三年級的自然，三年級的學生與過去帶班級都是高年級的學生差很多，不會察言觀色，明明對學生表現不滿意，已經氣得要命了，學生都不知道老師在生氣了。還有學生上課一直哭，還哭一整節的，想要裝作沒聽見，但哭聲大到讓你無法上課。（890910 討論會筆記）

教學的問題，只覺得不順，但未釐清問題核心，分析王老師的教學，似乎教學前的計劃與教學實施才是重心，此時的評量像是教學的附屬品，是老師在教學後應盡的義務，爲的是要打一個成績以對學校、家長有個交待。在教學中實施的評量，大都是以分數作爲控制秩序的利器，亦或是將它設計成有趣好玩的競賽方式，目的是讓學生專心且反覆練習教學內容。這時評量的主要內容是以學習態度及課本習作上的知識認知爲主。王老師認爲這些評量最好能採用多種方式的評量，這和王老師此階段對多元評量的體認：「多元評量就是用很多種評量方式來打成績」相當一致。

王老師開始對自己的評量感覺有些不大對，最初是在教授三年級上學期的兩次月考的考前，王老師發現自己都會一再爲孩子複習、一再強調一些重要的概念…，這種現象一直延續整個學期，因而引起王老師的關注並思考：

為什麼會這樣重複出現反覆練習的情形？到底

自己心中擔心的是什麼？再三思索後，我不得不承認自己是怕孩子考不好；而成績不好，可能的原因之一是老師、家長、孩子對分數的迷思：大家均習慣高分，甚至是滿分，大家會拿成績與以往、與其他科目做比較，以判定教或學的品質。(900303 研究日誌)

上述的反思促成王老師在下學期行動研究的焦點：「破除分數的迷思」。另外，王老師也發現三年級小朋友在各科的成績平均還蠻高的（大約九十以上），因而王老師會配合各科的情況，調整給分的標準，以不至於給得太低。但結果卻變成有些上課時頗令王老師困擾的孩子，其秩序方面的成績雖然不高，但經過與其他評量項目平均後，學月成績可達九十幾分。而學期結束時更發現每班的期末總成績有百分之九十以上都得優等（九十以上），這讓王老師覺得十分訝異與不妥，因爲這樣的分數不就代表孩子在自然科上的表現優異，在學習行爲上不用做任何修正了！但卻認爲有許多孩子還可以再努力，甚至有一些孩子是十分需要「加油改進」的。這迫使王老師不得不去檢視自己的「評分制度」！

又在一次成長團成員間的教學觀摩中，王老師觀察到教學演示的教師陸老師在批改完學生的學習單後，會發還給學生，課堂中檢討學生的答題情形或指引可能的修正方向，並讓學生有再訂正學習單、繳回批改、再調高分數的機會。王老師想到自己對於學生學習單的處理方式，都只是打完分數發下去、再稍微向全班說明大家寫的情形而已。這對王老師有如當頭棒喝：「教學評量應當是如此才對，但是自己好像一直都沒有這樣做，爲什麼？自己好像覺得評量重要的是可以打分數。」(900110 研究日誌）這讓王老師察覺自己的評量重心有所偏失。

自然科教學中，秩序問題一直困擾著王老師，在教學上王老師也覺得不甚理想，但由於評量問題具體、不斷地浮現，讓王老師不得不去思考：怎樣的評量作法可以更符合自己心中「教師不做無意義機械式的複習，讓孩子主動求進的動力」？如何讓親師生從現行教育環境的分數迷思中解套？如何更清楚地表達孩子的學習情形？評量的目的又是什麼？

研究團成員們對「分數的迷思」這個主題有踴躍的發言，最後的共識是：老師所評的分數，與家長孩子們心中所解讀的可能不同！王老師認爲表現好的孩子，其家長孩子們看了成績後可能覺得還須多加油；而王老師心中認爲應多努力的孩子，其家長孩子們看了成績後可能覺得表現不錯了。會造成這樣誤解的原因之一，是因爲各科老師間的評分標準不同，以及親師生間對於「好」的標準不同。不是自然科的成績比較低（高），便是代表孩子在自然科的學習上比其他科差（好），更何況這個分數是如何產生出來的，各科間的算法並不相同，有些科目甚至是一再地進行平時測驗，一再檢討、練習…，所以更是不能比。

經過與研究團體成員對話及不斷省思之後，王老師寫下希望能眞實呈現孩子的學習情形是：

希望家長與孩子們看到成績後，能知道自己真正的學習情形，…而得到鼓勵或再努力的訊息。(900302 研究日誌)

「擬定評量標準」是王老師破除分數迷思的第一個策略：

當自己重新出發去了解、去學習多元化評量後，發現評分時是有判準的，會按照幾個類別（等級）來區分學生的學習狀況，...多元評量應是從各個面向、用適當的標準來表達學生的學習情形(920302 研究日誌)

在擬定評量標準後，到底各等級要算幾分？要如何將之轉換成分數呢？在研究團隊所提供的衆多參考資料中，王老師選擇參考李坤崇（1999）提出的「核算等級計分」的方式，

並將「所有評量項目化爲分數後再加權平均」，以消彌分數會隨著評量項目多寡不同，而呈現出分數高低不同的情形。

分數已有它代表的意義與標竿，研究團成員指出：

但將許多項目平均後得到的分數，帶給學生的是什麼？代表的意義是什麼？家長與孩子們仍然是看不出來好在哪裡？不好在哪裡？（900227 討論會筆記）

因此王老師決定用等級來呈現結果：

用分數呈現學習結果，只是像法官一樣給學生一個判定，對於學生後續的學習沒有什麼實質上的幫助。為讓孩子清楚自己的學習狀況，知道是否達到學習目標，在學習單、成績單上，均用等級符號來表示，不再打一個分數，也不再將數個項目加總後再平均來呈現。（900227 教師文件）

不過，爲配合現今的評量生態，王老師還是決定在學月後算出一個分數來交給班級老師。

在評量標準、計分方式與結果呈現擬定了後，王老師評量系統的架構也成形了大半，不過尚缺少一個重要部分，那就是：到底要評什麼？評量是針對教學目標來進行檢核，那麼自然科的教學目標有哪些？經過研究討論會議的討論，王老師認爲：這些教學目標需完整、符合現今科哲潮流、我國國情、時代脈動，並劃分出各年段深淺不同的目標的，當下非九年一貫課程綱要裡的能力指標莫屬。因而能力指標便成爲王老師教學與評量上的指引。

三下的課程一是「認識人體」與「養蠶」兩個單元，其中「養蠶」單元王老師覺得很適合進行「檔案評量」，他的的理由是：

養蠶活動需長期的飼養觀察，很適合將這一段時間蠶的變化、飼養情形蒐集起來…，並以此單元來介紹『檔案』的觀念，教孩子如何做檔案。（910403 研究日誌）

王老師歸納學者們（吳毓瑩, 1999; 李坤崇, 1999; 鄒慧英, 2000; Popham, 1997; Stiggins, 1997; Tombari & Borich, 1999）所提出的檔案評量設計流程：(1)擬定檔案評量的目的：檢視學生的學習成果，並呈現學生學習歷程進步、努力與成就的觀察與紀錄，以成果評量爲主兼有過程檔案的特色。(2)擬定學習目標，發展檔案項目：由於「養蠶」單元目標繁多，且編制檔案項目費時，實施又迫在眉睫，因此王老師採邊教學邊編製的策略。(3)擬定評量標準。(4)規劃相關細節並製作使用說明。

檔案實施時王老師先寫給家長一封信說明檔案評量的特色以取得家長的信任，並循序漸進地先配合教學活動繕寫學習單，再告訴小朋有要製作檔案及檔案評量的目的及意義，遇有一些情況需家長或其他老師配合時都會事先溝通。

實施過程中亦發展評語貼紙給予學生更多的回饋，若有學生表現未達王老師期望時，王老師會舉小朋友的優良作品爲參考範例，進行說明，以幫助學生學習，並肯定表現優良學生的努力。對於少數無法完成學習單的學生，王老師採個別課後協助的方式進行輔導與補救。在此過程中，王老師發現學生展現提問、記錄、表達能力，且部分學生在老師鼓勵後，會自動蒐集相關資料、物品放在檔案裡。

爲了讓家長瞭解學生學習的情形，王老師依能力指標、測量項目與學生學習狀況製作學習通知單。爲了讓家長孩子了解該評量項目爲何，王老師在其後附上簡短的文字說明其學習目標，以及孩子的學習狀況是五等級中的：「做得很好」、「完全做到」、「大部分做到」、「做得很少」還是「沒做或缺交」。

檔案評量結束後，王老師在研究者建議下，形成訪談大綱，收集學生對於教師實施檔案評量的感受，發現學生喜歡用等級來呈現學習狀況，且認爲檔案評量能幫助他們學習，覺得這種方式很有趣，並不太難，例如：

甲生：可以瞭解蠶寶寶的身體。

乙生：蠶我喜歡，只要常常看蠶的話，就可以

知道上面的題目怎麼寫。

丙生：寫檔案可以多瞭解蠻，才不會零零散散一大堆，做一個像這樣的檔案，就讓我們覺得好珍貴的感覺，…只要用心觀察就可以寫得出來。（900302 研究日誌）

由於研究團隊的成員並無教師教授三年級，因此曾建議王老師找同年級的自然老師一起合作設計編製檔案評量。不過在學期一開始王老師與同事談起時，很快的就發現大家的看法不同，在教學上的實施不一樣，在課程即將實施，時間緊迫的情形下，無法共同討論共同設計，因而王老師只好單打獨鬥。這也讓王老師覺得如果無法分工合作，要「改進教學、落實想法」，是需要付出大量的時間精力的。

而在檔案評量實施完後，王老師又遇上了另一個難題：標準難訂、批改費時。量化的規準與批改是容易的，但是質化的規準難訂、判斷學生屬於哪一等級需要思考，因而批改一個班的一份學習單，往往要花上一段時間。

在這個階段裡，王老師制定了評量標準、擬定分數轉換方式，並以等級來呈現結果，來破除「分數的迷思」之問題。在教學目標上，王老師不再僅以課本上的認知、操作技能…爲滿足，而以九年一貫第二階段的能力指標，作爲自己努力的方向。養蠶檔案的實施，受到家長的肯定，成爲王老師往後可以全力投入的後盾，而檔案評量裡包含許多的目標，讓小朋友培養並展現各樣的能力，小朋友也從中能體會到老師的用心，這是讓王老師最高興的。在這過程中，王老師發現評量標準很難訂定、批改費時，是檔案評量實施時會遇到的難題。

二、主張二、反思能力的教學與評量，王老師全力投入能力指標的解讀、教學與評量，卻發現課堂中能力似乎被分割成細小獨立的部分而看不到全貌，引領教師重新思考能力的意義與分類

課程來到「認識人體」單元，由於王老師想不出來這一個單元要如何設計，所以在教學與評量是以課本上既有的操作活動、習作、月考來評量，而在月考前又同樣是經歷「強調、複習」的窘境。

在這一課裡，有「製作並操作手臂運動模型」的教學活動，王老師便讓學生自己製作、操作，而後打成績：

小朋友翻到課本第 15 頁…，15 頁上面有四個步驟對不對？待會你就照著做，有不懂的問老師，它上面寫的很清楚，你照著做有不懂的問老師。…做完的小朋友，你自己操作看看，看怎麼讓手臂動，對，自己操作看看，待會來給老師打成績的時候，老師要看你怎麼讓它動的喔。（900320 教學錄影）

除了評課本上既有的操作活動外，王老師尚用了如同上學期般的小組個人學習態度、習作與學月紙筆測驗的評量型式。王老師的教學情形是以發表及競賽進行小組加分方式，來進行課程：

請一位小朋友當模特兒站在黑板前，老師問全班他是屬於動、植物或都不是，在得出人算是動物後，老師讓各組討論人的外形分為哪些部分（提示鳥分成哪些部份？昆蟲呢？），而後抽問小朋友回答、下結論。接下來，老師再問人的「頭、軀幹、四肢」上有什麼？一樣先分組討論、抽問、別組補充。（900227 教學錄影）

不過這樣的教學活動，學生不是很有興趣，所以王老師便有了如下的思考：

今天的活動有很多學生分心、講話，是不是教學活動不吸引學生？如設計較有趣的活動可改善嗎？要怎樣進行才好呢？（900227 研究日誌）

王老師於是改由各組默選一動物，討論後寫（說）出其與人的不同，讓大家猜那是什麼動物。改成猜測競賽、小組加分的方式後，再上別的班時，學生變得很有興趣、很專心了。但是，王老師省思：

在這一個單元中，我教了什麼能力？似乎大部份都算是科學知識的介紹，再加上個製作及操作手

臂運動模型而已。如果沒有教任何能力，當然也無法進行評量，而最終學生學到的只有知識概念，其他能力要如何培養出來呢？如果每課都這樣教，那麼「培養學生帶得走的能力」變成只是願景，「科學素養、能力指標」只是教條規臬…。(900405 研究日誌)

在「認識人體」的單元裡，王老師察覺到自己的教學及評量的方式，仍然是以認知爲主，而其他的能力則幾乎沒有觸及，到底「在課程中能力要如何教學、如何評量？」，成爲王老師現在迫切需要探究的課題。

從三年級下學期將「能力指標」設爲教學目標後，王老師，便一直對「能力指標的這些能力要如何進行評量？」的問題困擾與不解，它常常是王老師向同儕求教的話題。由於九年一貫課程尚未實施，王老師尋找了國內文獻未看到相關的資料，在有機會時向學者、教授請益，得知目前「能力指標如何評量」尚未有具體發展。因此王老師只好從實作評量…等的研究與文獻中，自己摸索、發展出自己的能力指標評量理論。

在文獻中，王老師發現桂怡芬（1996）等是在學月考查時來進行實作評量，但認爲這樣的方式須耗費大量的時間精力，似乎僅能偶一爲之，恐難達成王老師想要「長期落實在實際教學中」之理念。此外，王老師認爲每次的教學活動，很像是一次馬里蘭州實作評量（陳文典、陳義勳、李虎雄和簡茂發, 1995）所要評的內容，因此就想：爲什麼不結合課堂中的活動來進行呢？而且，這樣做後，又符合 Wiggins（1990）所說的「希望學生會什麼，就測他什麼」，也就是評量主要嵌入在教學活動中來進行。

此時的王老師認爲多元評量是：

因為教學目標是多元的，需用多種方式方能適切掌握學生的學習情形，因而需要有多元評量…(900404 研究日誌)

因此，王老師的評量實施，依教學目標的不同，而有不同的教學活動，並隨著不同的教學目標需求，發展出了四套評量招式：「探究活動、過程技能、情意及學習態度、科學認知」的評量，與一套評量暗器「能力軼事記錄」，隨後又開創出「另類的學月評量」。

在實施了這些的評量活動後，王老師發覺科學素養的培養已開始落實在教學中，伴隨而來的是時間不足的壓力：

能力指標的教學評量中自己不斷的思考如何在教學中來培養能力？如何評量？由於要進行評量，因而科學素養的培養不再只是高遠的目標，而是一定要落實的。…科學素養的培養需要時間讓孩子從中去探索…，而評量內容、評量標準、評量實施、評量回饋…等也都需要時間說明，所以我常覺得時間緊迫，常在算距月考還有多久，上不上得完。(900515 研究日誌)

此時王老師亦體認教師角色更難扮演，但成長與收穫也相對地增多：

教師的地位更形重要，正處於重要關卡之處，是學習者與知識之間的仲介者（吳毓瑩, 1996），這就是自己一直覺得很難的原因；老師要設計課程與評量、控制好班上秩序、引起學生動機、接好學生拋出的東西，決定如何回應…，而最後又要達到效果…一路走來，雖然有許多的挫折與無力感，但回頭看看留下的足跡：自己在設計、思考、實務能力上有不少的成長，倒是又給了自己不少鼓舞與勇氣；尤其看到同儕、學生與家長的回饋，會覺得再多的辛苦都是值得的。(900607 研究日誌)

到這個階段結束時，評量的技術、技巧都發展得差不多了，但王老師自己仍覺得在對學生的「評量回饋」上沒有處理好，例如評量前有許多時候未事先告知孩子評量規準，而事後傳達時又覺得有些孩子沒在聽、效果不佳，若時間不足時，此部份的時間又常被刪除掉。

關於事前評量規準要不要告知，也常令王老師感到迷惑：

如果要他們觀察一朵花，學習單上把要他們觀察的項目、標準都寫了（或事先先告訴小朋友），那不是以老師為中心，只是要孩子來配合做到這些嗎？如果在上了課（比較了兩朵花的不同：已分析到顏色、形狀、數量、氣味…）後，不先說什麼直接讓他們觀察應該是比較合適的，但是這又沒做到文獻上所說的：應事先告訴學生明確的標準…。（900420 研究日誌）

而事後的回饋，有時會令王老師感到它是例行公事：老師這樣做好像是覺得它是一件義務與責任，學生則表現得心不在焉、莫不關心，這似乎不是自己當初所熱切希望進行的方式。

王老師亦反思到教學中能力的培養似乎被切割了，而在教學的引導上、教學的成果上也有待改進。第一次經驗到自己似乎將「能力」切割成獨立的個體分別進行評量與教學，是在成長團體的教授、教師觀摩王老師的班級教學，會後的討論會促使王老師省思到的：

覺得自己當初設計是依一個一個能力單獨來看，似乎沒有思考到建構、科學本質真正的核心…。好像自己不重視他們有沒有概念改變，有沒有從預測、實作、寫出結論等一連串的過程中得到什麼，好像只要他們單獨會預測、實作、寫出結論便可以了。(900503 研究日誌)

在上觀測天氣的單元時，王老師在上課後又反思到：

自己的評量似乎太強調零碎行為的測量（如風向風力的測量）而把一些重要的科學本質忽略了，…好像有些課程，沒有什麼連貫，只評某項能力，而非整個學習行為的統合，如自己似乎不重視學生先前概念、有無改變，只要要評的能力做了就好。（900601 研究日誌）

王老師自己心中是期待那些能力（如提出適當的實驗設計）在教學中有出現就好，而對於整個教學活動的進行，卻未全盤考量。

小組活動如何評量個人表現是另一個讓王老師困惑的問題：

在許多活動中，王老師是讓小組共同寫一張學習單，而這張小組的學習單如何評量個人的部份，是王老師感到疑惑的，如果把所有的評量項目擺在一起來評個小組成績，再讓組員互評得出個人在小組的表現排序，再以小組成績為基礎依個人表現排序來加減分，即得個人成績。這樣雖公平地區分出小朋友的表現而給個分數，但評量並非僅是給個能區分出小朋友表現的分數，重要的是知道誰達到了什麼學習目標、誰需要什麼協助。(900424 研究日誌)

在這個階段，王老師的重心是進行由上往下的「能力指標」的教學與評量，在找不到可作爲參考或指引的情況下，王老師只好自己開闢途徑：決定採用嵌入在教學中的評量活動；並隨著不同的教學目標，例如：科學認知、探究、過程技能、情意及學習態度等設計不同的評量方式。

分析教師的教學錄影、學生文件，發現教學與評量活動，涵蓋了部分能力的培養與評量，例如：提問能力、預測、創思、解決、歸納、傳達能力、批判能力。而且由王老師對七位學生進行的訪談資料，幾乎所有學生都喜歡這些評量方式，認爲可以幫助學習、讓他們瞭解自己的學習情形；只有一位學生表示：要自己思考，想不出來。又全班的學生問卷也可看出 63 %小朋友認爲這些方式激勵與提昇學習動機、興趣與成果，促使自己用心學習。此項作法並獲得家長的肯定，例如一位家長回饋：「這種評量方式讓小朋友有思考的能力，不再是死讀書，謝謝老師」(9205 家長文件)。不過，伴隨而來的能力歸類統計、時間進度壓力、仲介角色扮演，是王老師要跟著去面對、去思索、去同步成長的。

除了評量的設計、實施、運用，及其他專業能力的提昇外，尚有王老師在目前即使是花了心力也無法解決的問題，這些「評量回饋未

做好、能力培養被切割、小組評量難運用」的問題，只好留待下一階段，看能否有所行動與突破。

三、主張三、科學本質的反思與再建構促使王老師建構科學探究歷程教學評量模式，並對能力指標有更深入的理解、詮釋與評量

王老師表示，雖然自己上過些科學本質的相關課程，但對科學本質的體認一直是零碎與片段的，就像瞎子摸象的故事中沒有看過大象的瞎子，雖然有人告訴你大象像樹幹、也像扇子、也像繩子、也像牆壁、也像矛，但樹幹、扇子、繩子、牆壁、矛到底要擺放在哪裡呢？他們間相對的關係仍很模糊，瞎子是拼湊不出大象完整的圖像，自己則似乎只記得科學本質中的暫時性、可試驗性、重現性…。但在一次「科學本質」的課程探討中，王老師有了許多突破性省思：

以前我的教學活動是有「預測」，有「實驗後的想法或發現」…，表面上看來是有些符合科學本質的，但它們是片段的，局部的，自己是把「預測」看成是能力指標中有的，所以要做的…。…其實「實驗設計是孩子們用來驗證其想法的創思」、「實驗結果產生後並非是來傳授科學知識，而是用來檢驗預測是否正確、提出解釋、發現新問題的重要階段」…自己以前的評量好像流於條文般地要一一去達成與檢核！不過再細想，以前雖是沒有全套暢通的流程，可能只是片面的考慮、局部的使用，但這些是可以修改的，以前老師與學生的經驗正好成為往後的基礎。(901017 研究日誌)

在這樣的領悟之後再看「能力指標」時，王老師發現原來每一條能力指標便是在科學探究活動中須用到的各種思考、技能…等，雖然在課程綱要裡，它們是依「思考知能、過程技能…」等科學素養項目來歸類；其實每一條能力指標都有其扮演的角色與位置，是會隨著探究活動的進行而依序出現的。

而教學上必須是「學生主動投入探索」才有可能引導出符合科學本質的探究活動，因而王老師便融合過去成長團推行的「學習環教學模式」(林曉雯, 2000) 與「科學本質」，並列出對應的能力指標與評量型式，經過時間的精煉與討論後，王老師最後發展出「科學探究歷程教學評量模式」。在此教學評量模式中，科學探究是一個無止盡的循環歷程，一個探究活動引出的問題，可以成爲下一個探究活動的起點。在各歷程步驟間，王老師對應列出了九年一貫課程第二階段的能力指標及評量型式。此時王老師對多元評量的體認是：

多元評量是因應學科本質的課程目標，採用能真切了解各種學習目標的方式、人員與時機，以瞭解孩子的學習情形、引導後續課程與教學的計畫與施行…」(910118 教師文件)

因應學習目標的不同，期望這些評量型式能眞切評量出該目標的多元評量型式。其中在科學探究歷程中的「心智活動」(思考智能與部份的過程技能)，王老師設計「評量單」讓孩子記錄下其想法；「實驗操作」則由教師當場「觀察記錄」；於全班討論時孩子所展現的「各項能力」，王老師發展出「能力軼事記錄」來捕捉孩子的表現；在一個段落後欲對「科學認知、各項能力…」進行中結性評量時，則採用另類的「學月評量」型式；整個歷程中的學習態度、主動參與、傳達溝通等能力，則可以用「組內自評互評、小組競賽、能力軼事記錄」來了解孩子的學習情形。此模式於「雲與雨」、「植物的蒸散作用」中逐漸成形，於「月亮出來了」、「陰濕環境中的生物」、「溶液」、「電路」等單元中實施。

有幾條能力指標如何解讀？仍是王老師的疑惑：

能力指標如何解讀？我解讀出來的意思與原來相同嗎？當初會寫出這些能力指標，應該是它們分別表示了在科學的過程中，所會運用到的能力吧？是不是像我最近所理解的：「歸納研判與推斷」是

屬於實驗後的能力…等。(901103 研究日誌)

雖然後來王老師已能從科學本質中，領悟到大部份能力指標的意思，在解讀這些能力指標時王老師會參照前面的子題，如「歸納研判與推斷」、「傳達」；再看看低、高年級的相同子題下的能力指標是指什麼，來增進理解並減少曲解原意的情形，但是王老師仍有一些問題，例如少數幾條能力指標王老師仍不解其意，王老師心中暫且留著這樣的疑惑，看看以後是否能再釐清。

另外，王老師還發現了「操作、裝置實驗器材設備」的能力在能力指標中找不到歸屬的問題，但在多次的活動下來，王老師發覺小朋友「實驗操作」是重要的技能之一：

「實驗操作」應該也是很重要的，像小朋友在「植物的蒸散作用」中，水沒有加一樣多、試管口的膠泥沒有封好，這都會影響實驗的結果。雖然說設計實驗、創思能力最是重要，但如果實驗過程中不精確，也無法分析出正確的結論，或是又要重新操作…；所以，這方面的能力應該也是要加進去的。(910305 研究日誌)

王老師並增加了「中結性評量」(王老師自創此名詞，強調屬「小總結的形成性評量」，因其結果可做爲教師後續教學引導與課程安排，以便與「總結性評量」區隔)的步驟，來評量其他多種能力，爾後經由多次理論與實務的對話中完成「科學探究歷程教學評量模式」(如表2)。

這樣的「科學探究教學評量模式」的大架構定了後，課程設計、教學引導與評量便朝著這樣的模式進行，由於在進行過程中，王老師能隨時評量出學生的學習狀況、學生會有進一步的問題出現，因而後續的課程、教學與評量，便須依據教學現場的評量結果與學生的反應來引導、調整：

當評量嵌入在教學中，當每次的教學活動，就是一次的實作評量時，那麼它就能立即反應出「評量即是老師所重視的重點」、「評量學生真實的表現」，也就是它是如吳毓瑩（2001）所說的「真實評量」，而且，它不就可以立刻蒐集到學生學習情況的資訊，可以知道學生學得如何，可以立刻的反應在教學的引導上、在給學生的回饋上、在課程的編排上，而每次的評量又都是學生一次練習的機會。(910130 研究日誌)

亦即，整個科學探究教學評量模式，在一開始的情境佈下後，整個課程、教學與評量，便是依著師生的互動呈現動態的發展了。

在每個單元中「科學探究教學評量模式」裡的各個步驟，所佔的比重不同，會隨著學生的學習情況與課程內容，而有各自的重心：

在「植物的蒸散作用」單元，重心是偏在讓小朋友去比對實驗結果與預測，並看看各組的結果相不相同、提出批判…。在「月亮出來了」單元，重要的是看小朋友能否從二個月的觀察資料中整理出規則，所以這裡花比較多的時間。…「溶液」單元的最後一個探究活動，本來還要在各組報告完後，大家針對每二組的結果進行歸納、提出結論、想法、批判、問題…等，但因時間來不及便沒進行，不過，它不是這一個單元最重要的重心，省略沒關係。(901228 研究日誌)

「時間來不及、課程進度上不完」，應該是一直會面臨到的難題，王老師對這個難題的態度是：

時間不夠，是永遠會遇到的問題，我們只能面對它、想策略去解決，而不能說它是個問題就不做了…。(910613 研究日誌)

「課程、教學與評量合而爲一」、「抓住每個單元的重心」、「評量單由小組共同完成」，「分組同步進行不同的探究活動」是王老師使課程進度快一些的解決策略。王老師亦發現當學生漸漸熟悉科學探究教學評量模式、漸漸培養出能力後，也會使課程進行得更順暢、更快。

表2：「科學探究歷程教學評量模式」

階段	科學探究歷程	第二階段的能力指標	評量型式
投入	***情境佈題*** ↓ ***引出想法*** ↓	→（1211）觀察.（1231）通則性、（1221.1222.1224）比較分類.（6211）提出問題 →1234 推理 先前概念；普遍性假設	→能力軼事記錄 →評量單
探討	確定問題 預　測 ↓	→1232 關連（預測）	→評量單
	探究活動：設計實驗 ↓ 實驗操作 ↓ 觀察結果(蒐集資料) ↓	→6221 創思（辦法） →1233 控制變因 →6222 創思（設備） →1235 裝置操作 →1211 觀察 1221.1222 比較（度量） 1251 傳達（表格）	→評量單& 老師觀察記錄或檢核表
解釋	組內比較：預測與結果 ↓ ***(探究後：)*** 歸納 批判 想法 發現 ↓ 組間比較：結果&歸納 獲得結論（概念） ↓	→（1231）通則性 1241 歸納 6211 批判（提問） 6212 批判（質疑） 1234 推理 1223 比較（原因） 3202 科學本質 5212 發現樂趣 →22**科學認知	→評量單& 能力軼事記錄 →習作.學月評量
精緻化	研判.推斷 ↓ 科學應用	→1242 研判推斷 →7201 科學應用	→小組競賽 學月評量
評量	↓ 中結性評量	→22**科學認知 其他多種能力	→學月評量& 態度量表…
	整個歷程	＊6231 主動參與 ＊學習態度（含器材整理） ＊傳達	組內自評互評 小組競賽 能力軼事記錄
	其他	＊科學應用＊科技發展	

註：***標楷體斜體字***：代表教師要特別去營造情境、引導學生主動投入學習
新細明體正體字：代表學生進行的活動
（能力指標編號）：代表該項能力可能出現，但情形較少且較不易評量
能力指標編號：代表王老師自己增列的能力

評量規準的擬定與批改呈現「上下交融」的情形，某些評量項目在發展穩定後，則能達到「由上而下」的境界。例如在學月評量時，王老師發現有些評量規準已能事先做初步的擬定，再於批改時做修訂，這樣的規準擬定與批改，不再是單純的「由上而下」或是「由下而上」，而是呈現「上下交融」的情形：王老師心中先有概略的標準，再依照學生的作答情形於批改時逐漸修訂發展而成。而在組內自評互評各個項目的評量規準中，王老師已發展出固定的評量規準，這樣的評量規準擬定與修改，便已達到「由上而下：事先擬定而後依照規準評定」的境界。

此時從評量中得到的訊息，不只回饋在該項「評量內容」、「教學引導」中，亦回饋在後續的「課程設計與編排」上，故促成王老師的專業成長。以往，王老師從評量中所得到的訊息，便會回饋在該項「評量內容」上，例如：說明評量標準、批改並告知其學習表現，並給予肯定或勉勵；亦會回饋在「教學引導」上，例如發現小朋友不了解或操作錯誤時，會再進一步地與小朋友討論：

在觀察登錄（評量）各組實驗操作的情形時，發現小朋友們用的水量不同，為使兩組容易比較，便使大家先暫停下來討論此一問題，並最後決定都用 100ml。(910701 教師文件)

但在這一階段裡，王老師更將從評量中得到的訊息，回饋在「課程的設計與編排」上，以掌握孩子對月形的迷思概念或從活動中產生的進一步提問，來設計編排後續的課程：

因為各組小蘇打粉與硼酸溶解的情形不同，這已與接下來要進行的課程「溶解量與相關因素、濃度…等有關」，並且，在讓小朋友說說看他認為可能的原因時，小朋友也提到了有關的變因，因而下面的課程便打算接續這個時候的問題來繼續發展。（910701 教師文件）

這些評量得到的訊息與反思，回饋到王老師的專業成長中，並更了解「科學本質」與「建構的意涵」，以及解決教學現場中的問題。

省思中王老師自認仍未解決給學生的評量回饋之問題。王老師在向小朋友說明評量標準，讓小朋友知道要怎麼寫比較好，但總覺得在向學生說明評量標準時，學生沒有專心聆聽，王老師感到自己的評量回饋「還是做不好」！

能力指標需要逐條來評嗎？這也是王老師一直疑惑的：

能力指標需要這樣分條來評嗎？但不這樣評難以看出小朋友的各項的能力情形、進而給予協助，不過我目前尚未做到掌握表現較差者給予協助的階段；但如這樣評，它的信效度可能有問題，且成績統計上費力費時…（910128 教師文件）

另外一個影響的相關因素，就是成績統計處理的問題，如果依照王老師的想法：「孩子是否培養出那項能力是要累計多次表現來看」，那麼，就需要有適當的成績統計軟體來配合處理才行，但王老師找不到適當的軟體可支援。

在這個階段中，王老師的教學因爲「科學探究歷程教學評量模式」建立與實施，貫通了科學本質、能力指標與評量的策略，而逐漸掌握到科學教育探究教學的特色與重心。但是，當這樣的模式一再重現，課程進行到較長時間的批判對話時，卻發現教學中有不少的孩子已無法進入學習情況了：

這幾次上課發現有的孩子們已經沒有跟上節奏了，以前在小朋友間進行較久的批判對話時也會有這種情形；自己是不是太強調邏輯—數學智慧了？自己本來就是邏輯—數學智慧比較強，好像自己忘了顧及一個班級中，每個孩子的優勢智能不同，喜歡的學習風格不同，而自己忘了設計不同的學習方式，只以一種學習方式，當然有些孩子無法持續保持秩序與專注度。而對於思考能力較差又好動的孩子，是否可以多提供更多操作體驗的機會？（910114 研究日誌）

「如何在大班中，依孩子的能力與性向的不同，而提供多樣式的學習機會，來滿足每個孩子、讓他們有不同的發揮」，這正是王老師最初欲實施的多元智慧的議題，而現在又要重新回頭面對此一問題。

在這個階段裡，王老師接觸到「科學本質」的課程，了解了科學探究各步驟的意義後，發現每條能力指標不是分割的、獨立的，其所指的內涵，便是在科學探究歷程中須有的思考、技能、認知…等，因而王老師融合了建構精神與科學本質、納入了九年一貫課程第二階段的能力指標，並搭配上一個階段所發展出來的評量型式，形成了王老師個人的「科學探究歷程教學評量模式」。在這個模式裡，評量便在教學歷程中進行，王老師從評量中得到的訊息，回饋在該項「評量內容」、「教學引導」、後續的「課程設計與編排」、及王老師的「專業成長」上，使得課程、教學與評量是依著學生的學習狀態來做動態的調整，以達成教學目標。不過王老師仍未解決評量「回饋」的問題，並產生了「能力指標需要逐條評嗎？」、「缺乏適當的成績統計處理軟體」的疑惑與困難，也省思到「自己未提供多樣式的學習機會，來滿足孩子的個別差異」。

四、主張四、反思學生後設認知的能力引領王老師再建構學生自我評量的意義

王老師在閱讀文獻中體悟到自我評量的重要性：當孩子有「自我評量」的能力，便會有「自我學習」的動力，而當然也便會學習到各種能力：

法國著名精神科醫師克利斯多夫．安度禮（Christophe Andre）及法蘭蘇亞．都羅爾（Francois Lelord）在其「自我評量心理學」一書中提到：「對自己的看法和判斷會對我們的心理平衡有很大的影響。如果自我評量是肯定的話，即能以比較安定的感覺度過，做事比較積極，不論遇到什麼樣的困難都能勇敢接受挑戰。相反的，如果自我評量是否定的話，即會感到不安，在日常生活中會事事感到不如意，事事無法順暢。因此，自己如何對待自己，如何看自己-也就是知道如何評量自己，成為一件極為重要的事。為使自己的人生過得充實，為使自己能做到主體性學習，必須會自我評量。（910216 教師文件，引自柯啟堯（2001））

在四下的第一次自評裡，王老師觀察到小朋友的情形：

我們班上有一個小朋友很愛講話、非常愛動，可是啊如果讓他專心起來，他提了很多問題、思考都很好，可是他常常是管不動，他是不自主的。那次讓他們自評，我非常的驚訝他自評的結果，他當然是給自己是整組最低分，這個部分他很公正，可是他所有的評價都比別人高，所有的評分點都很高。我就想他要是不知道自己那麼不好，覺得自己還不錯，他怎麼會改進？（910412 教師文件）

因而，王老師領悟到：

除了讓學生學會自我評量以外，更不能讓學生自我看待的標準與社群有太多落差，要讓孩子有比較自己、師長、同儕的標準與評量的機會。這應該就是多元對話：「營造出學習社群中，所有人員可以相互對話、交換意見」（吳毓瑩, 1999; 李坤崇, 1999; Airasian, 1996; Linn & Gronlund, 1995）的另一個重要性所在！…而後更重要的是引導學生能進一步去朝著做不好的部份努力，使其自我評量能漸趨向「自我肯定」！如能做到這樣，那麼，不論學習或處事的目的主要是為了達到「自我肯定」的評價，

這樣的動力是內發的，而不是靠外在獎懲的控制！（910424 研究日誌）

培養出孩子能帶著走的「自我評量」的能力，是很重要的，當孩子有「自我評量」的能力，便會有「自我學習」的動力，而當然也便會學習到各種能力！體悟到評量最終的目的是要培養出學生「自我評量」的能力之後，王老師發現使用的評量方式雖然與以前的做法相同，但用意、目的是不同的：

以前小組討論的結果，讓學生評分，是為了討論的結果要給分數，所以讓他們來評，以提高學生參與的機會，不但能讓學生能專心，又能達到評分人員多元。現在讓學生參與評量的主要目的是讓他們「學習評量」。我想，從中他們可以比較出彼此間評分的差異，去學習拿捏標準。（910304 研究日誌）

另外，王老師也會去構思一些不同的做法，如學習狀況通知單上，王老師加上了讓小朋友自己自評的部份。王老師會讓小朋友先自己評量學習單而後王老師再修改。王老師認爲透過這樣的多元對話—評量人員提出自己的意見、交換看法，小朋友就可以在自己、老師、同儕的標準比對中，逐漸形成自己的評量能力。這應該就是「在營造出的學習社群中，所有人員可以相互對話、交換意見」（李坤崇, 1999; 吳毓瑩, 1999; Airasian, 1996; Linn & Gronlund, 1995）的目的。

在瞭解自評的重要性後，王老師也同時澄清了自己覺得「評量回饋」做得不好的問題根源是在：以往是「老師希望孩子做到」，並非是營造情境讓「孩子自己想要達到」。由於以往的回饋，大部份都是王老師來「告訴」孩子們標準是什麼、希望他們改進，而這種對學生而言的「被動的要求」效果自然不好，所以王老師一直覺得「回饋」未做好；未來，王老師要營造「主動的需求」，讓學生「投入」去評量、往「自我肯定」的評價去努力，那麼，這樣才能達到孩子「自主學習、樂在學習」的境界。

在「養蠶」檔案之後，王老師就未讓孩子做過檔案評量了，因爲王老師覺得如果以自然科而言，似乎在「專題研究」時才比較用的上，其他情形不太適用。但是在領悟到「自我評量」的重要性後，王老師才看出檔案評量的重要性

像檔案那樣，讓小朋友蒐集每樣的表現？如此他們可以了解自己的情形、增進自我概念，可以力求達到各能力指標，又可以往自己專精的地方去發展。（910412 研究日誌）

檔案評量除了可以增進孩子的自我概念外，它又含有人文關懷在裡面，又可以讓孩子省思、監控自己的學習情形。這時候，王老師認爲檔案評量是培養孩子正確「自評能力」、引導孩子努力趨向「自我肯定」的適當的評量方式。

同時王老師亦領悟到「小組表現」、「個人表現」各有其重要性與功能，教師須搭配交替使用，並且體會到小組表現的自我評量，能幫助孩子學習在社會群體中的角色扮演。在三年級下學期時，王老師便對「小組活動如何評量個人表現」感到疑惑，在此階段時，王老師已掙脫了這樣的困境，認爲小組共同完成的評量單並不一定需要算成個人成績，「小組合作」、「個人表現」各有擅場，因而，王老師的歸納是：小組合作探究，可以從中進行社會建構、學習民主溝通的方式，小組的評量單，可以讓老師掌握全班大致的學習情況，如果全班學生差不多都會，就可以帶過，如果大部份學生都反應出不太會或錯誤的情形，再共同討論。而個人探究，則可以練習或展現個人的能力，從個人的評量單中，老師可以掌握個人的學習情形，可以看出每個孩子的想法與思考、興趣與態度、擅長及不足之處。這兩者需要交替運用，除了可以在時間上、興趣上做調配外，重要的是他們可以從中「評量」出自己與別人的不同、

學習別人的長處，可以較快達到「自我肯定」的目標！

另外，融合「眞實評量」(吳毓瑩, 2001)與「自我評量」的體會，王老師又有了如下想法：

孩子們在小組裡的角色與關係是否正是如同以後在社會上一般，所以小組的合作學習，是孩子們學習往後在社會上所扮演角色，例如領導者、執行者…的機會，如果在小組中是各自運用自己的強勢智慧的話，那麼就另有一番功用了，比如說評量裡可能要讓他們記下組員的表現與貢獻，而後再來分析，讓他們可以常發揮，也可以有時要他們練習弱勢的智慧、學習別人強勢智慧的做法。這次讓他們自己來寫對組的貢獻好了。這樣做應該也較符合利用評量的正面力量來協助他們，而不是負面打分數箝制的力量。(910402 研究日誌)

因此，王老師在四年級下學期第一次學月學習狀況通知單裡，也加上了讓孩子省思「自己在小組中的表現與貢獻」的部份，讓孩子們開始「察覺」自己在小組中的表現。

此階段中王老師領悟到培養出孩子「自我評量」的能力，是實施評量的更高層目的，也澄清了自己一直覺得「給學生的評量回饋做不好」的問題根源，並轉而認爲檔案評量是適當引導孩子們學習自我評量的方式，也開始著重讓孩子學習未來在社會環境中的群己角色。

陸、討論

本研究探討一位國小自然科教師在發展其計畫、實施、反思以實施評量改革，進而促進教學改進、提升教與學的品質的歷程中，影響教師的行動研究歷程及專業成長的重要因素－反思對教師教學實務與專業發展的影響。在過去的許多與教育改革或教學改變有關的研究顯示，教改的成效多半不彰或維繫不久(Duffy & Roehler, 1986)。王老師勇於挑戰個人的教學，一方面源於對自我教學的需要與期許，另方面歸之於相關的指引、支援與資源，讓王老師勇敢的選擇並投入教學改變。教師的改變始於教師的需求，以發展解決教室教學問題的策略，教師個人是否有「改變的需要」是促成改變的最重要因素(Tobin, 1989)。否則教師很容易在使用新策略或新方法時遭遇困難或挫折即便放棄，不願持續改革。王老師進行教學評量是以「如何讓孩子努力、改進？」的問題爲始，以學生的成長爲考量，誠實面對學生的反應，支持王老師面臨疑惑、困境、阻礙時都能堅持到底，盼得柳暗花明，終於邁向「如何培養出學生擁有自我評量、趨向自我肯定的能力？」的評量目標。教學情境中的狀況讓教師感受到不滿意，是促使教師持續反省、進行行動的主要因素，但這樣的不滿意是以「以學生的成長」爲主要的考量，而不是老師個人的喜好、方便、容易入手，也不是外在的掌聲。例如，在這個行動研究的開始，王老師便一路得到許多家長、學生、同儕的肯定與鼓勵，但是王老師並未將之視爲終點的訊息，而是把它當做往前邁進的後盾與動力。王老師認爲眞正要考量的是以眞實、充分的學生回應爲基礎，省思自己「是否做到引導孩子有意義的學習？」、「是否能協助孩子有更多的成長？」。在這研究歷程中，王老師眞實的面對自己從學生的反應中，身處種種困難與疑惑中，如「我掌握到能力指標的涵義？」、「我做好給學生的回饋？」，看不到目標與出路，不知道要往哪裡走，但王老師在熟讀相關論述，以相關理論爲基礎，堅信努力方向是正確的，這樣做是值得，只要堅持、努力實踐，終究會撥雲見日、柳暗花明。在王老師的例子中，他看到評量改革可以幫助學生學學習與教師教學的遠景，因此他願意不斷的投入與付出。這點和 Fullan (1991) 在研究教師改變中指出，教師衡量改變的得失，若非較目前的方案有效，或在可預見的未來是有利的，

否則教師不會冒然改變的論點不謀而和。

釐清問題、尋求支援與資源，將問題一一破解則是王老師的行動秘笈。教學情境中各種因素交錯影響，使得評量問題更趨複雜。但王老師並未將這些困難視爲不可跨越的阻礙，而是以一個教師即研究者的身份，將教學中遭遇的問題一一分析、努力、突破，並獲得教學專業的成長。以王老師在第二階段面臨「能力指標如何評量」的問題爲例，要發展這樣的評量型式，須先對能力指標進行解讀；而在課堂上進行時，又需處理班級的秩序問題、孩子們之間吵架、不合作的問題…等。這些都需先處理、解決，才能進行評量活動。

回顧整個研究歷程，王老師是在三年級上學期課程中時察覺到教學現場中的評量問題，而開始去重新認識並實踐「多元化評量」。有時候一開始關心的焦點問題，到後來才發現它只是問題產生的表象，其實如果澄清情境，原來的問題就解決了。就像王老師在研究歷程的前半段，常常被「秩序」的問題所困擾，雖然使用了不少策略，但常只是治標而不治本。等到王老師解決了教學的問題—體會到科學本質的涵義、能引導孩子投入探究後，發現「秩序」便不太成爲問題，要留意的只是在活動轉換時，凝聚孩子的注意力。

又如「分數的迷思」與「結果的解讀」等問題，在王老師建立了評量制度後，很快便能夠解決，但關於「如何讓孩子努力、改進」的回饋問題，雖然在短時間內，似乎看到了狀況的改善，但歷經數個月之後，王老師還是覺得給予學生的回饋沒有做好，一直到最後王老師澄清了原來「如何讓孩子努力、改進？」的問題是「如何培養出學生擁有自我評量、趨向自我肯定的能力？」後，心中的疑雲才得已解開、心中的不安才得以釋懷，此次的行動研究，才在此暫告一段落！

在釐清上述多元複雜的問題後，王老師展現了多重向度的解決問題能力。對於「能力指標如何解讀？」、「如何評量？」、「如何促進自然科學生適性發展？」等開創性問題，由於研究當時並未獲致直接相關的研究或文獻，王老師從「科學本質」的探討課程、「實作評量」、「多元智慧」的相關文獻中，加以融會、轉化、應用以解決此類問題。

在「如何給學生評量回饋？」、「能力指標須逐條評嗎？」、「小組探究結果如何評量個人表現？」等偏屬於意義性的問題上，很難在短時間內想出因應的方法，因而王老師是將它們放在心中，不斷地反思、與同儕探討…，後來在因緣際會、時機成熟時，便開悟體會到其涵意。

在「小朋友不會寫、沒有寫檔案」、「時間不足」、「試題編製不佳」、「戶外學習情況難以控制」等困難上，王老師是發展技術性的策略：「舉例說明」、「抓住各單元重心、同時進行多個實驗…」、「調整評量標準、再訪談小朋友」、「使用組內自評互評」來解決這些困難。以上教師發展多元的解決策略，呼應過去研究發現，研究結果亦支持相關研究文獻中所述，即教師投入行動研究能促使教師解決所面臨的教學專屬問題（例, 歐用生, 1999; 蔡清田, 2001; Hensen, 1996; Tobin, 1989）。而當教師進行行動研究時，不只變得更能分析、反省、批判自身的教學與行動，更具備問題解決的技能(Briscoe & Wells, 2002)。

而相關的進修活動，如研究所的課程所介紹的相關理念在教學多年後爲王老師帶來嶄新的視野與願景，成長團員間的教學實務上的激勵，讓王老師有感：有爲者亦若是。定期的研究討論會中團員間分享、展示與說明、互相切磋、所營造的支持教學改變的環境促使王老師理念的釐清、規劃計畫、深層的反思，更重的是在內心覺得懷疑、不安時組員間的相互扶持，發揮了穩定情緒的功效，才能再接再厲，

鍥而不捨，終能撥雲見日。這亦說明行動研究歷程中合作的重要性（王靜如, 2000; 林曉雯, 2001; 郭重吉, 江武雄和張文華, 1999; Briscoe & Wells, 2002）。

王老師亦察覺到書寫的價值：「書寫研究日誌，雖然會花很多的時間，但常常在書寫的當中，會讓自己省思到更多事情，很多時候問題的澄清或領悟，是發生在書寫的時候，透過文字的整理，可以將事情做進一步的整理與釐清。」透過撰寫的過程，王老師再反省自己所理解的內隱知識，是否與所呈現的方式與內容相符；並誠實地將整個行動研究的過程詳細地描述出來，特別是反省的部分（Elloitt, 1991; Feldman, 1994），並期能因報告的呈現，引起專業社群的共同思考與批判性對話，而彼此間獲得深入的理解與啓發（陳惠邦, 1998）。

王老師由評量技術改變入手，體認目標、教學、評量三者緊密的關係，沒有高品質的教學，遑論好的評量，轉進教學後，省思能力指標的意涵，進而促成對科學本質、科學教學理念的轉變，建構自己的教學評量模式，最後體會教學評量的深層意義，即培養學生後設認知的能力，培養學生終身學習的心智習性。以課程發展的結果觀之，歷經以活動、內容、歷程、心智習性爲結果的四個階層（Costa & Kallick, 2000）。也歷經 Habermas（1971）所謂技術旨趣、實務旨趣、解放旨趣的不同階段，建構其個人的專業知識。行動研究中的各項反思呈現了 Schon（1983）對反思實踐的系統分析，隱含式行動中知曉、行動中反思與對行動反思。由於教師對自己教學行動的反思，潛在的實踐理論得以被揭露與發展（Elliott, 1991）。然而整個歷程中巨大轉變的重要關鍵點，在教師對於教學信念、科學教學目的、教師角色的反思與再建構改變，促成教學實務的重大變革：探究教學評量模式的建立，以及後設認知能力的培養，這與許多科學教學評量成功案例的歷程相似（例如, Briscoe & Wells, 2002; Kamen, 1996; Tobin, 1989）。

柒、啓示與建議

本研究中王老師以評量爲研究的起點，促使她重新思考自然科教學的目標、課程的內涵，與評量的意涵，以提供學生更多或更佳的學習機會，進而經由評量檢驗教學的成效。王老師融合過去所學與行動研究中所獲致的經驗、技能與知識，以進行教學評量的工作，在此歷程中，無疑的教師專業獲致長足的增長。此結果不但能鼓勵教師以自己教學現場的關注點或問題爲起點，與志同道合者組成合作團隊，進行技術、實務、目的、信念的反思與再建構，建立個人之教學模式，更能發展教師的專業知識，以提昇學生的學習成效。

此結果亦可提供學者專家思考如何提供適當的協助，以促成教師反思，進而引導自然科教學的改變，及教師教學專業之成長，爲教學革新貢獻一己之力。可能策略包括：進行研究歷程的分享及鼓勵教師書寫研究日誌或書面報告以達成釐清、反思自然科教學的目標、課程的內涵，評量的意涵，與個人信念的目的。並探索教師的教學需求以引發教師的研究、協助組成志同道合的成長團體以達相互扶持之效、協助教師爭取支援與資源以解決問題。

本研究仍有許多未竟之處仍待後續研究，如研究歷程王老師展現歷經以活動、內容、歷程、心智習性爲結果的四個階層的課程發展順序，是否能以其他的方式促進教師跳躍式的發展歷程，或直接以最高層次的心智習性進行，值得進一步研究。而王老師似乎以技術、實務、解放先後不同階段，建構其個人的專業知識的歷程是否是必經的順序與歷程？或是有其他不同的發展取向，亦是下一步研究的焦點。此外，可探討不同學科的教師、擔任不同年級的教

師、其他背景的教師（如初任教師、非主修科學之教師，不同信念之教師）進行行動研究時其教學反思的特色。另教學評量改革的成效如何？研究者可針對學生學習成效、學習興趣、學習態度、自我學習的瞭解等，進行大量的、有系統的設計與研究，以提供相關專業人士之參考。

致　　謝

本研究承國科會科教處給予經費補助（NSC 90-2511-S-153-004）（NSC 91-2511-S-153-005），研究助理、參與教師在進行研究及資料整理時齊心投入與配合，使本研究得以順利完成，僅在此致上最高的謝意。

參考文獻

1. 王文中（1999）：測驗與評量的意義與趨勢。載於王文中、呂金燮、吳毓瑩、張郁雯和張淑慧合著。**教育測驗與評量：教室學習觀點**（pp. 1-20）。台北市：五南。
2. 王澄霞（1997）：科學科之教學與評量標準。**國立編譯館通訊**, **10**(1), 39-41。
3. 王靜如（2000）：國小教師教學改變之研究。**屏東師院學報**, **13**, 281-315。
4. 田麗娟（2001）：**國小自然科評量方式之行動研究－以學習歷程檔案評量爲例**。台北市：國立台北師範學院碩士論文（未出版）。
5. 辛慶偉（1998）：**國小自然科卷宗評量建構效度之探究**。台北市：國立台北師範學院碩士論文（未出版）。
6. 李坤崇（1999）：**多元化教學評量**。台北市：心理。
7. 歐用生（1999）：行動研究與學校教育革新。**國民教育**, **39**(5), 2-12。
8. 吳毓瑩（1996）：評量的蛻變與突破--從哲學思潮與效度理論思考起。**教育資料與研究**, **13**, 2-15。
9. 吳毓瑩（1999）：卷宗評量。載於王文中、呂金燮、吳毓瑩、張郁雯和張淑慧合著。**教育測驗與評量：教室學習觀點**（pp. 235-258）。台北市：五南。
10. 吳毓瑩（2001）：**後現代社會中多元評量之解毒與解讀**。2002 年 1 月 30 日，取自 http://www.trd.org.tw/Cpast/9010/901102/5-1.doc。
11. 吳鐵雄和洪碧霞（2000）：教學評量質變的反思與建言。**測驗與輔導**, **159**, 3323-3324。
12. 林曉雯（2000）：建構主義教學策略—學習環的基本理念及國小自然科學設計舉隅。**科學教育**, 11, 43-51。
13. 林曉雯（2001a）：國小自然科教師試行「學習環」之合作行動研究。**屏東師院學報**, **14**, 953-986。
14. 林曉雯（2001b）：提昇國小自然科教學成效之研究：自然科教學實務改進及影響因素之研究。**國科會專題研究報告**。計畫編號 NSC 89-2511-S-153-009。
15. 侯雅齡（2000）：如何有效運用多元評量結果。**測驗與輔導**, **159**, 3339-3341。
16. 張文華和郭重吉（1995）：科教革新中評量理念的重建。**教育研究雙月刊**, **45**, 23-30。
17. 桂怡芬（1996）：**自然科實作評量的效度探討**。台北市：國立台北師範學院碩士論文（未出版）。
18. 黃秀文（1996）：從傳統到變通：教學評量的省思。**國民教育研究學報**, **2**, 1-26。
19. 張惠博和黃文吟（2000）：科學學習的評量理念。**科學教育**, **231**, 49-57。
20. 柯啓堯（2001）：自我評量和交互評量的意義與功能。**翰林文教雜誌網路版**，18 期。2002 年 1 月 28 日，取自 http://www.worldone.com.tw/magazine/magazine.htm。
21. 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和運用。**科學教育**, **185**, 2-11。

22. 陳惠邦（1998）：**教育行動研究**。台北市：師大書苑。
23. 教育部（民 87）：**國民教育階段九年一貫課程總綱綱要**。台北市：教育部。
24. 郭重吉、江武雄和張文華（1999）：中學數理教師在職進修課程設計之行動研究(1)--教師轉變與課程調整。**科學教育學刊**, **5**(3), 295-320。
25. 陳聖泰（2000）：**國小自然科「電磁鐵」單元實作評量應用之研究**。桃園縣：中原大學碩士論文（未出版）。
26. 曾嘉琪（1996）：**台北市國民小學試辦教學及評量改進班之評鑑研究**。台北市：國立政治大學碩士論文（未出版）。
27. 楊宏衍和段曉林（2001）：合作學習一高中化學教學之行動研究。**科學教育學刊**, **9**(1), 55-77。
28. 楊銀興（民 89）：**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**。台北市：國立臺灣師範大學碩士論文（未出版）。
29. 鄒慧英（2000）：多元化的檔案評量。**國教之友**, **52**(1), 16-23。
30. 甄曉蘭（1995）：合作行動研究－進行教育研究的另一種方式。**嘉義學院學報**, **9**, 297-318。
31. 熊同鑫和王振興（1999）：行動、反思與成長：一位自然科教師的自我教學研究。**科學教育學刊**, **7**(1), 17-34。
32. 熊同鑫、王振興和陳淑麗（2001）：以國小高年級自然科爲核心之課程統整設計研究。**科學教育學刊**, **9**(2), 123-145。
33. 蔡清田（2001）：教師如何進行教育行動研究：「教師即研究者」的理想與實踐。**國教之友**, **52**(3), 3-18。
34. 歐用生（1999）：行動研究與學校教育革新。國民教育, **39**(5), 2-12。
35. 鄭淵全（2000）：國民小學教學評量之評析。初等教育學報（新竹師院）, **7**, 105-123。
36. 戴君佩（2001）：**國民小學自然科卷宗評量對三年級學生學習動機之影響**。台中市：國立臺中師範學院碩士論文（未出版）。
37. Airasian, P. W. (1996). *Assessment in the classroom*. New York: McGraw-Hall.
38. Baumann, J. (1996). Conflict or compatibility in classroom inquiry? One teacher's struggle to balance teaching and research. *Educational Researcher*, *25*(7), 29-36.
39. Briscoe, C., & Peters, J. (1997). Teacher collaboration across and within school: Supporting individual change in elementary science teaching. *Science Education*, *81*(1), 51-66.
40. Briscoe, C., & Wells, E. (2002). Reforming primary science assessment practices: A case study of one teacher's professional development through action research. *Science Education*, *86*, 417-435.
41. California Assessment Collaborative (1993). *Charting the course toward instructionally sound assessment*: *A report of the alternative assessment pilot project*. San Francisco: California Assessment Collaborative (ERIC Document Reproduction Service No. ED 373 098).
42. Clandinin, D., & Connelly, F. (1995). Teachers' professional knowledge landscapes: Secret, sacred, and cover stories. In F. Connelly & D. Clandinin (Eds.), *Teachers' professional knowledge landscapes* (pp.1-15). New York: Teachers College Press.
43. Costa, A. L., & Kallick, B. (Eds.) (2000).

Discovering & exploring habits of mind. Alexandria, Va.: Association for Supervision and Curriculum Development.

44. Cunningham, G. K. (1998). *Assessment in the classroom*: *Constructing and interpreting texts*. London: Falmer.
45. Duffy, G., & Roehler, L. (1986) . Constraints on teacher change. *Journal of Teacher Education, 37*(1), 55-58.
46. Elliott, J. (1991). *Action research for educational change*. Philadelphia, PA: Open University Press.
47. Feldman, A. (1994). Erzberger's dilemma: Validity in action research and science teachers' need to know. *Science Education*, *78*(1), 83-101.
48. Feldman, A. (1996). Enhancing the practice of physics teachers: Mechanisms for the generation and sharing of knowledge and understanding in collaborative action research. *Journal of Research in Science Teaching*, *33*(5), 513-40.
49. Glaser, R. (1990). Toward new models for assessment. *International Journal of Educational Research*, *14*(5), 475-483.
50. Habermas, J. (1971). *Knowledge and human interests*. Boston: Beacon.
51. Hart, D. (1994). *Authentic Assessment*: *A handbook for educators*. Parsippany, NJ: Dale Seymour Publications.
52. Hesen, K. T. (1996). Teachers as researchers. In J. Sikula, T. J. Buttery, & E. Guyton (Eds.), *Handbook of research on teacher education* (2nd ed.) (pp.53-66). New York: Simon & Schuster.
53. Howe, M. E., Thames, D., & Ward, C. (1995). *Teacher perceptions toward the interpretation of results from the new norm-referenced portion of the Mississippi assessment system. A Pilot Study*. Paper presented at the 24th Annual Meeting of the Mid-South Educational Research Association, Biloxi, MS. (ERIC Document Reproduction Service No. ED393 846).
54. Kamen, M. (1996). A teacher's implementation of authentic assessment in an elementary science classroom. *Journal of Research in Science Teaching*, *33*(8), 859-77.
55. King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgment*: *Understanding and promoting intellectual growth and critical thinking in adolescents and Adults*. San Francisco, CA: Jossey Bass.
56. Linn, R. L., & Gronlund, N. E. (1995). *Measurement and assessment in teaching*. Englewood Cliffs, NJ: Merrill.
57. Little, J. (1990). The mentor phenomenon and the social organization of teaching. In B. Cazden (Ed.), *Review of research in education*, Vol. 16 (pp. 297-351). Washington, DC: American Educational Research Association.
58. Lytle, S. L., & Cochran-Smith, M. (1990). Learning from teacher research: A working typology. *Teachers College Record*, *92*(1), 83-103.
59. McGhee, T. J. (1998). *Utilization of authentic assessment in Georgia's elementary schools*. Unpublished doctoral thesis, The University of Georgia.
60. National Science Council. (1996). *National science education standards*. Washington, D.C.: National Academy.
61. Popham, W. J. (1997). *Classroom assessment: what teachers need to know*. Boston: Allyn & Bacon.
62. Puckett, M. B., & Black, J. K. (1994). *Authentic assessment of the young child celebrating development and learning*. New York: Macmillan

College Publishing Company.
63. Richardson, V. (1994). Conducting research on practice. *Educational Researcher*, *23*(5), 5-10.
64. Richardson, V. (1996a). From behaviorism to constructivism in teacher education. *Teacher Education and Special Education*, *19*(3), 263-71.
65. Richardson, V. (1996b). The case for formal research and practical inquiry in teacher education. In F. Murray (Ed.), *The teacher educator's handbook* (pp.715-737). San Francisco, CA: Jossey-Bass.
66. Schon, D. A. (1983). *The reflective practitioner*. London: Temple Smith.
67. Stenhouse, L. (1975). *Introduction to curriculum research and development*. London: Heinemann.
68. Stiggins, R. J. (1997). *Student-centered classroom assessment*. Upper Saddle River, NJ: Merrill.
69. Stinggins, R. J., & Bridgeford, N. (1985). The ecology of classroom assessment. *Journal of Educational Measurement*, *22*(4), 271-286.
70. Stinggins, R. J., Conklin, N., & Bridgeford, N. (1986). Classroom assessment: A key to effective education. *Educational Measurement*: *Issues and Practices*, *8*, 5-14.
71. Tabachnick, B. R., & Zeichner, K. M. (1999). Idea and action: Action research and the development of conceptual change teaching of science. *Science Education*, *83*(3), 309-22.
72. Tobin, K. (1989). Teachers as researchers: Expanding the knowledge base of teaching and learning. In M. L. Matyas, K. Tobin, & B. J. Fraser (Eds.), *Looking into windows*: *Qualitative research in science education* (pp. 1-12). Washington, DC: American Association for the Advancement of Science.
73. Tombari, M., & Borich, G. (1999). *Authentic assessment in the classroom*: *Application and practice*. Upper Saddle River, NJ: Merrill.
74. Vygotsky, L. (1978). *Mind in society*: *The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
75. Watts, H. (1985). When teachers are researchers, teaching improves. *Journal of Staff Development, 6*(20), 118-127.
76. Wiggins, G. (1990). *The case for authentic assessment*. ERIC Digest. Washington, DC: ERIC Clearinghouse on Tests, Measurement, and Evaluation. (ERIC Document Reproduction Service No. ED328 611).

Reflective Practices on Science Assessment: A Teacher's Action Research

Sheau-Wen Lin[1] and Pei-Chun Chen[2]

[1]Department of Science Education, National Pingtung Teachers College,

[2]Kaohsiung Municipal Chung San Primary School

ABSTRACT

The objectives of this study were to explore a science teacher's reflective practices on science assessment during action research. The researchers gathered and analyzed data from teacher's journal, documents, teaching videotapes, interviews, and field-notes of research team meetings. Triangulation was used to enhance the validity of findings. Findings were presented in the form of four assertions: (1) reflection on objectives of assessment led the teacher to establish assessment routines, to break the score myth, and to clarify the interaction of instructional goals, curriculum, and assessment, (2) reflection on the ability indicator system of the new curriculum in teaching and assessment led to rethinking the meaning of science literacy, (3) reflection and reconstruction of what science was led to developing a "instruction and assessment model", (4) examination of students' self knowledge making led to the reconstruction of self assessment.

Key words: Action research, Elementary Science Classroom Assessment, Reflection, Professional Development.