

多元評量的活化、迷思、與神話

教學歷程的個案研究

林怡呈* 吳毓瑩**

本研究採個案研究法，探討教師運用多元評量策略（包括意圖、方式、與實施）之整體圖像。研究者進入主角教師之教室現場共一學期 100 節課，輔以訪談、文件分析等策略，分析歸納出主角教師實施多元評量之三種圖像：1、活化：主角教師在語文課程中選用合宜的評量，有效運用訊息與資源，結合評量與教學，並將結果回饋於學生，且作為改善教學的依據。2、迷思：主角教師在數學課程中，由於學科教學知識不足，評量規劃較貧乏僵化，對於教學無所助益，有時且誤導學生學習。3、神話：主角教師在社會課程中，寓遊戲於教學中，熱鬧有趣，但實際上未達到學習目的，多元評量如同虛幻的「神話」。研究者後續提出四個議題與讀者分享討論。

關鍵字：多元評量、教室評量、教學歷程、個案研究

* 作者現職：屏東縣富田國小教師，國立高雄師範大學教育學系博士班研究生

**作者現職：國立臺北教育大學心理與諮商學系教授

壹、研究動機與目的

多元評量在學術界中，自 1990 年代之教育改革運動以來，引發各種理論與方法的討論（如江文慈，2004；呂金燮，1999；李坤崇，1999；吳毓瑩，1998；吳鐵雄、洪碧霞，2000；莊明貞，1998；張麗麗，2000）。實務界中的多元評量亦呈現出多樣化的面貌，評量策略有更多的選擇空間，如現場老師常運用的學習單、檢核表、闖關評量、專題報告……等。但是，誠如吳毓瑩（1995）所言：「是不是以這種取代那種？或是以『好多種』取代一種便已達成多元評量所要傳達的開放意義呢？」這是值得我們省思的。張麗麗（2002）也提出相似的質疑：「相當多的教師對實作評量充滿了幻想。認為只要設計了生動、活潑與生活貼近的評量活動，學生就會在『玩』的過程中，提高學習的動機與興趣，而教師們的評量從此也可以走出傳統紙筆測驗的陰霾而海闊天空了。」凸顯出來的即是當前教師面對多元評量可能產生的迷思。職此之故，觸發我們欲了解現場教師實施多元評量整體圖像的研究動機，希冀透過對現場教師的觀察，詮釋教師在教學歷程中運用及思考評量的多層次面貌，提供教師運用多元評量之借鏡。

貳、評量在教學中之運用

評量在教學中的角色為何？對課程而言，「評量的課題賦予課程的成果與生命；只有藉著明確說明學生應該做的事，以及他們能做得多麼好，這些學生學習的敘述才有意義。」（Danielson & Abrutyn, 1997/2000, p.8）而在教學上，「一個理念再優良的教學計畫，如果少了有效的評量，就像騎匹千里馬，一路狂馳，但卻不知道方向在哪裡。」（王文中，2000）可見評量的訊息對教學活動更有著直接且及時的影響。因此，本節將針對評量與教學活動的互動關係做相關的探討。

一、教學決定需要評量的訊息

評量乃超越資料搜集技巧，是一個有系統的過程，始於確認教學目標，終於判斷這些目標達到的程度（Linn & Gronlund, 2000, p.29）。過去學生、家長、社會大眾多以為測驗卷或考試就是評量範疇。而今學習單與讀書報告在評量改革的氛圍下已被大家接受，但是這樣的評量還是不夠周全，學習活動亦需要一些非正式評量的指引。

非正式的教室觀察指引著許多教學上的決定，例如口頭問答可以帶領學生對教材做一統整回顧；教室討論可以發現學生某些迷思概念（Linn & Gronlund, 2000, p.29）。同樣地，經過對學生的個別觀察後，教師可以了解這孩子需要造句方面的協助，那孩子需要更多的數學解題技巧，而還有一個學生需要更有挑戰性的任務等。諸如此類的教學決定有些基於學生的反應、實際表現，有些來自於學生疑惑的表情、語調或行爲。這些都需要教師無時無刻的觀察，Stiggins 稱之為「審時度勢的評量」（size up assessment）（2001）。雖然觀察是一個非正式的評量，但是，它在有效教學中扮演著不可或缺的角色。

測驗或其他的測量程序並非意圖取代教師的非正式觀察與判斷，相反地，它提供更多非正式的方法讓教師獲得學生學習的訊息。教師仍然是重要的觀察者與決策者，而測量或評量程序僅是提供更有系統的證據來支持教師的教學決定（Linn & Gronlund, 2000, pp.29-30）。因此，教師的眾多教學決定，仰賴評量提供系統化以及非正式觀察之資訊（莊明貞，1999；盧雪梅，2000；簡茂發，1999）。

既然教學的訊息來自全面性、多元化的評量，那麼評量內容、方式、歷程與目標應在擬定教學計畫之初就該同步進行。思慮縝密的活動式評量，往往比靜態的紙筆測驗吸引學生的注意，即使是低學習成就的學生，在面對遊戲式的評量活動也可能激起參與挑戰的鬥志與熱情。但也並不是所有的另類評量都能激起學生面對挑戰的意願，或賦予低成就孩童信心。如林清山、林素微和鄒慧英（2001）與柯政毅（2001）的研究證實，某些以「書面寫作」為主的評量形式對教師及語文能力弱勢者是一大負擔。而吳毓瑩與桂怡芬（1997）的研究也顯示評量方式的改變雖然獲得多數學生的肯定，但也造成優秀者抱怨挑戰力不夠、無法凸顯程度之憾。這些研究證明了，另類評量不一定適合所有的學生，教

師在使用仍應多方思考學生的能力與特質。

評量在教學中的角色，超越評量結果的準確、排序、與篩選，更積極性地走向催化、鼓勵、增強或激發動機，如同 Sternberg (1998) 強調的：學生在發展專家知識 (expertise) 的過程包含五個要素——後設認知技能、學習技能、思考技能、知識、與動機，其中以「動機」為五者之核心要素。動機可以驅策其他成分，如果沒有動機，這些成分就產生不了作用 (pp.16-17)。多元評量更應提供如此的驅動力與展現的舞臺，讓各層次的學生體會到自己的價值而適才適性的發展。

二、評量與教學的關係

對於評量在教學歷程中的定位，學者們有不同詮釋，Puckett and Black (1994) 認為它是三股糾結在一起相互嵌合、彼此支持的繩索 (p.34; 侯雅齡, 2000, 頁 3339); Gross (1998) 認為它也可能是類似太極具有動感的三元素圖 (p.13)。吳毓瑩 (2003) 則認為：在後現代的跨界相互學習之情境中，它是跨越疆界的多元關係圖，如圖一所示。

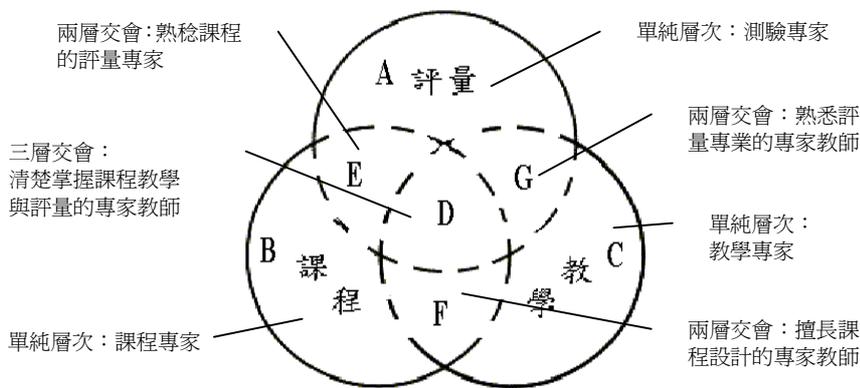


圖 1 課程、教學、與評量交錯的多元關係圖

上圖的特色在於課程、教學與評量三者的交界處均為虛線，表示當課程、教學與評量三者相遇時可能是交互融合的模糊地帶，讓人分不清這是教學還是評量，而觀點不同的人立足點也會不同。職場教師可能早上出了一張數學考卷了解學生的學習成效，立足於 A 點，目的在測驗學生的成果。下午時候，進行社會科角色扮演，立足於 G 點，目的在促進學生對於教材的理解與意義感受。教育工作者視評量之角色、目的與功能的不同來選擇適當的落點。而落點既定，對於評量結果的解釋必然也鑲嵌於此情境中。但如果落點選擇 A 點（測驗學習成就的程度），卻以 G 點的情境來詮釋（以測驗歷程作為學習動機的激發），則所做的詮釋也會有偏差。因此教師應謹慎遊走於這個多元關係圖中。

本研究傾向從教學歷程來看評量的角色，所以研究焦點置於 G 點，詮釋一個熟悉教學且公認為靈活運用評量的專家教師，如何呈現出評量在教室中的面貌，而文獻中所提及的考量與理論，亦即周邊 A 點、D 點、C 點，則是影響 G 點運用評量與解釋評量的脈絡因素。

參、研究方法與歷程

筆者以下的描述，除了是研究方法與程序的說明之外，更欲與讀者溝通本研究與文本建構的可信賴程度。蔡敏玲 (1996) 認為：「使得研究結果『值得信賴』所做的努力並不是一個與研究程序分開獨立的步驟，這項努力嵌合在研究的整個過程之中，從進入現場的第一天開始。」(頁 105) 我們將從如何結識主角老師開始描述，與讀者分享我們的詮釋與感受。為了將讀者帶進脈絡，本文的第一作者將以「我」來現身。資料來源的標示，前者是資料性質，後者則是取得資料的時間，分別是民國、月份、以及日期。

一、她是誰

透過師長的介紹我終於認識多人知悉的歡老師。將近一年的接觸，從師長眼中、同儕口中、學生轉述中以及自己的觀察中整理出歡老師的面貌。

國王的人馬－

教學現場中歡老師呈現出熱誠、開放與積極的特質，深受校長、主任器重，常被推薦代表學校至校外演講，被同事喻為國王的人馬。

熱愛教學的人－

「喜歡小孩」是支持她維繫教學熱情的動力，而此動力讓她越教越有心得。再者，兩個聰明的弟弟卻不愛唸書的影響，讓她在教學時常絞盡腦汁去思考：「怎樣才會讓學生有興趣去學。」（訪談，900608）。

面惡心善的美女－

外表豔麗卻不知如何與他人溝通的歡老師，讓人乍看之下覺得嚴肅且難以親近，有人私下告訴我：「她看起來好嚴肅，如果我是她的學生，我會很怕她……」常駐班級的實習老師－美美也說：「我一開始看到她的時候，也會覺得她比較嚴肅一點，但是相處久了你會發現她人很好……。」（訪談，910107）她是一位需要花時間相處後，才能熟絡互動的老師。我也是到了學期中時，才覺得與歡老師訪談比較親切自然。然而，關於研究以及教學專業的事務，歡老師均全力配合與協助。教學上她更是不忍見學生落後，總是利用休息時間進行個別指導。

思慮周延的掌舵者－

歡老師行事總秉持著核心信念－思慮周延、縝密計畫。在家為先生打理一切，在校體諒學生課後忙碌於各式學習的辛苦，於每週一出好當週的回家作業，以利學生彈性運用課後時間。

我的研究伙伴是如此的人，那麼在現場中的我扮演何種角色？又如何與她互動？

二、在研究現場中 - 我是誰

在現場中我這麼知覺著我自己：

有教學經驗的人－

我有五年的教學經驗，所以觀察歷程中，腦海裡相照應的是自己過去的教學圖像，有時也會另外開啓一個小視窗，浮現文獻上所看到的理論陳述。

傳遞訊息的小信差－

歡老師有時會透過我請本文第二作者修正其評分規準的擬定。這樣的角色儼然是個傳遞訊息的小信差，是現場實務與理論間的橋梁。

極盡所能保持清醒的省思者－

對於歡老師的教學或評量策略，我常像發現新大陸般的喜悅，因為許多點子對我而言很新鮮。為了避免自己過度的「擬情」而落入「好萊塢式情節」（Hollywood plot）（Clandinin & Connelly, 2000, p.181）的危險，在研究現場我努力地建立田野文本，對教學現場產生的疑問，盡可能地當天提出與歡老師澄清、說明，並於離開現場後，我們兩位作者冷靜地從中撥開精彩畫面背後理想或迷思。

三、個案研究之取向

歡老師實施多元評量之場域就是她的教室、對象是她的學生。而我們欲詮釋的是她與脈絡交織互動的情形。個案研究法（Stake, 1995）正是引領我們達到研究目的最適切的方式。

(一)資料蒐集：現場札記與訪談

我在現場共參與經歷了歡老師約 100 節課的教學，以及對歡老師、學生、現場的實習老師共約 20 小時的訪談。同時，蒐集了各項教學的資料，如教師教學計畫、教學省思日誌、學生成果、以及現場觀察記錄。為了避免自己成爲一個與研究現場「談戀愛」的研究者（Clandinin & Connelly, 2000）觀察中也許情緒上是認同歡老師的，但是田野文本記述的是我所看到的對話或事件，以幫助自己回溯省思時再度檢驗。訪談則是蒐集資料的另一種策略，也是與歡老師建立關係與信任的好時機。

(二)資料的分析

資料的分析隨時都在進行，歷程中除了先分離出清楚可辨的評量主題與評量步驟外，我的焦點著重於搜尋歡老師評量設計與教學目標結合之訊息，也同時關注學生參與評量時的反應。而在最後階段要淬煉出精華時，我首先來回審視文本，以做第一次粗略的編碼。接著回歸研究主題，反覆閱讀資料找尋解析方向。而在我欽佩歡老師的情緒中，我第一個理出的主軸是「活化」，也就是知道學生的需要及激勵學習之發生，這些證據可從教室評量的本質與目的著手。根據此方向，我們從資料中繼續尋找可能活化教學的評量主題。而在討論過程中，我們發現歡老師自信的另一面，也對照出某些時刻她迷惑與不定。在審視原始資料的過程中，同時也在回憶觀察歷程出現的質疑，確實也有著歡老師勉強擬出的粗糙評量形式，非但不能活化教學，還可能誤導學生的學習。於是「活化」之外又形成了一個「迷思」概念的主軸。除此，我們還覺察到，歡老師有些時候以爲只要富變化、熱鬧、花俏、快樂，就是多元評量的展現，這樣的思維與作法讓我們看到歡老師將多元評量以「理想情境」的方式在運作。因此「神話」形成另一個詮釋的主軸。活化、迷思、神話這三個主軸概念形成之後，我們開始勾勒歡老師實施多元評量的活生生主題。

資料歸納與分析過程中所得的資訊是飽和的，也的確有許多評量主題能分別符合三個主軸概念，但礙於字數限制，爲讓讀者貼近現場實況與理解行動者策略之運用，是故詮釋時選取典型案例/主題進行流暢的深度描述。

資料分析中會參酌歡老師的意圖（訪談得知）、學生參與評量後的反應（現場觀察與訪談）、以及我們的觀點審視，最後再給予整個評量實踐適切的詮釋，亦即此時的詮釋會同時以評量的本質與目的交叉檢視整個評量實施的歷程。

肆、解讀多元評量的活化、迷思、與神話

本文之詮釋以歡老師爲主角，多元評量策略與實踐爲主題事件，現場交織錯雜的情境所建構出來的就是本文的整體情節。這些情節經過層層解析之後勾勒出三種多元評量在教室現場中的圖像－活化、迷思、與神話。

一、多元評量活化教學 語文領域的雪人歷程寫作

評量與教學的結合，是當前評量改革的主旋律。Airasian（1994）曾言：「當教學與評量未能妥善配合而導致學生成績低落時，會影響學生的學習動機與學習策略。」（p.189）歡老師受評量改革風潮薰陶之後，致力於與教學融合的評量設計。觀察歡老師的過程中，以「雪人」之歷程寫作評量最能代表評量活化教學的圖像，亦能夠傳神表達評量與教學妥善結合、啓發學習歷程的範例。

「……評量與教學喔，我反而不會覺得很難融合，我是覺得教學和評量很難區分，很難區分是因為，你要說它是一個教學活動也可以，你要說它是評量也可以，當我從不同的角度來看時，它是不同的東西……」（訪談，900824）

有教學經驗的我總覺得教學和評量難以融合，因爲在評量的時候我不會想過要給予學生即時的指導。歡老師卻認爲教學和評量難以區分，從不同的角度觀之，其以不同的形式呈現。「雪人」的活動融

合教學與評量，左看它像教學活動，右看它卻又像評量策略；學生不當它是評量，老師卻從教學中評量學生的能力，從評量中強化學生的學習。正是文獻中所提到的位於G點、熟稔教學策略與評量活動的專家教師之精彩處。

「雪人」是一個寫作歷程的主題，歡老師透過一鉛筆素描動畫進行寫作指導。這個故事發生在一個下雪天，敘述一個小男孩在下雪午夜發生的奇遇，整部影片有背景音樂，但無對白。寫作目的是「希望能夠讓他們有所謂的過程性寫作，還有同學互相修改，還希望有各種不同的寫作形式……所以想讓他們有一個依據去寫故事……」（訪談，901127）。影片劇情在歡老師看來故事性強，故事線清楚，易激發學生興趣，更適切的是整部影片沒有對白，減少學生書寫的牽制，促進學生對詞彙的想像力，符合歡老師的教學的目的，也符合她所要評量的能力效標之一——「詞彙的運用」。

活動進行約一個月，花了三次的時間進視聽教室，每一次對歡老師來講都有不同的規劃：第一次是純粹欣賞，第二次是引導學生分段落摘取大意並完成小書，第三次則指導學生如何將大意加長，寫成一篇可閱讀的故事。因為視聽教室需要預約，歡老師事前就得清楚何時借、借多久，所以歡老師對於引導的步驟也瞭然於心。

（一）評量過程中層次豐富的提問

作文寫得長，對小朋友來說不是一件容易的事，因缺乏豐富的素材對文章做鋪陳。至於素材的挖掘與累積，有賴教師點、線、面的引導。歡老師運用了六層次遞進的目標：是教學目標亦是評量目標，一步一步引導學生完成「雪人」文章。

第一層次目標－觀察與擬稿

草稿的初擬，重點在引導學生抓出影片的大意。除了大意，歡老師要求學生還得為這些大意配上適切的插畫。對於這些插畫，歡老師仍是花費一番功夫指導：「雪人怎麼畫，那個頭怎麼畫，然後他有那些裝飾品……鼻子是用什麼做的，眼睛是用什麼做的？那個圍巾的樣式如何……手跟腳是怎麼區分開來的……」（觀察，901127）。歡老師的目的在評估學生是否觀察得仔細入微，這些評量的重點都在歡老師引導時重複提醒學生，看似教學，也像是評量規準的傳遞。對於學生的草稿，歡老師一點也不馬虎，課後一點名到桌前來，依據所說的規準檢查，分別指導修正。

第二層次目標－段落分析

確定每位學生的草稿有潛力完成小書之後，歡老師進行下一層次的引導：段落的區分與安排。歡老師帶學生重看影片，並在重要處將影片暫停，引發學生對影片中的細節做聯想，共同討論出適宜的分段點。

歡老師：「……第二段是雪人復活的那一段，那要寫到那裡？」

學生：「快要飛的地方！」

歡老師：「如果你覺得第三段是飛的過程，那你覺得到那裡該停？」

學生：「到聖誕老公公那裡……」（觀察，901204）

她不是直接告訴學生如何分段，而是拋出問題讓學生思考：「你覺得要寫幾段比較適合？」學生可以不必有共識，但必須讓老師知道分段的依據為何。因為這是第一次不同形式的寫作，歡老師的引導更是層層解析，步步提升。

第三層次目標－修辭技巧

歡老師希望學生在文章中運用修辭技巧，如「成語、形容詞、連結詞、轉折語……」（訪談，901218），所以她會針對這些評量目標進行引導討論。

歡老師：「……不要只說房子，怎麼樣的房子？不要只說平原，怎麼樣的平原，要加一兩個形容詞上去。……想想看怎麼樣形容這些飛過的房子……很多造型不同的房子怎麼形容它？」

學生：「各式各樣……」

歡老師：「自己想想看有各種形容詞可以用……。」「然後他們要穿過怎樣的森林？……」

學生：「茂密的森林……」(觀察，901127)

歡老師很清楚學生困難之所在：「孩子通常不愛寫作文，也不知道該寫些什麼，這是因為教師沒有給予足夠的引導，不只是對於寫作技巧上的引導，還包括如何引發學生的情緒來寫。」(歡老師教室日誌，901002)她藉著細膩的引導，啟發學生搜索備而不用之舊經驗，讓這些舊經驗與新的寫作形式結合，不但可以削弱學生面對寫作的困難感，還可激發學生的隱而未現的潛能。

第四層次目標－傳神意境的表達

為了讓學生用字遣詞能傳神的表達劇情，貼切的傳遞劇中人物的感覺，歡老師拋出更多耐人尋味的問題，刺激學生對影片人物的一舉一動、一顰一笑做更深入的思考與想像：

「做完雪人，小男孩為什麼站那個姿勢？表示他那時候心情是怎麼樣？」

「想想看：貓為什麼會有那樣的反應？」

「雪人為什麼想飛？」

「你看大家對小男孩做了那些友善的事？」

「你如何寫兩個人這時候說再見那種依依不捨的情況？」(觀察，901127)

為了告訴學生「句子如何描寫才會生動」，歡老師會針對某一個畫面，舉例說明：

歡老師：「要把每一個畫面的『特色』(加重音)寫出來，譬如說，剛剛那個畫面有什麼特色？」

小文：「忘了穿內褲！」

歡老師：「ㄟ！你寫忘了穿內褲就不好看了喔！你可以講說：他光著屁股迅速的把褲子穿起來，然後襪子怎麼穿呢？……」(觀察，901204)

舉例說明之時，歡老師也會使用一些較特殊的句子，如：「小男孩和雪人在房間跳舞……破壞力很強的跳舞……」，學生疑惑的說：「破壞力……??？」，不等老師解釋，一些語文能較強的小朋友就搶著說：「沒錯！破壞力！因為他們弄倒很多東西！」(觀察，901127)語文能力強者在提問時的機智反應，不但受到老師的肯定，也提供其他學生學習的機會。經過了四層提問與討論，學生蒐集到許多相關資料，所以歡老師得再進行總結性的引導。

第五層次目標－去蕪存菁

透過層層提問與激發，學生蒐集到的寫作素材不可計數，所以歡老師再加上總結的引導。

歡老師：「來！老師問你們，你們看到那麼多，你每一個細節都要寫嗎？」

小靜：「不要！這樣好像在寫流水帳！」

歡老師：「對！小靜說對了，就好像你寫雪人一直換鼻子—『雪人把把蘋果拿下來換成酪梨，把酪梨拿下來換成香蕉……』，你寫這個過程沒有意義，你可講說：『他陸陸續續拿了很多的水果來換鼻子，其中(停頓)……』」

小哲：「寫最好笑的……」(觀察，901204)

歡老師以大量的疑問句提問以及停頓，讓學生有機會發表，教師再統整處理，當遇到較艱澀的問題，歡老師更是在提問之後，給予有力的鷹架支持讓學生穩健的往上爬。眼看一篇文章即將大功告成，歡老師不忘提醒學生對文章體裁的關注。

第六層次目標－體裁風格

此次強調的體裁是「故事」寫作，所以，老師提出最後一個問題，讓學生思考－

歡老師：「你寫的是一個故事，你可不可以寫說：『其中我覺得最好笑的是……』……」（觀察，901204）

歡老師說明這是一個故事體裁的寫作，與其他寫作方法是不一樣的。也讓學生了解自己是一位寫故事的作者，而非內容主角喔！

（二）明確的質性回饋激發學生動機

作文一向是學生頗感壓力的課程，歡老師爲了讓學生不害怕、進而喜歡它，時常絞盡腦汁地設計不同形式的寫作活動。歡老師揣測學生的興趣，規劃了「雪人」活動。學生在讀寫檔案中所寫下的期末省思印證了歡老師的看法。

「在這份讀寫檔案中，我最滿意的一份作品是雪人作文¹，因為很多句子被老師畫起來，而且那次作文的分數最高。」（小逸檔案，910116）

「在這份讀寫檔案中，我最滿意的一份作品是『雪人』，因為那是我寫最多的作文，裡面還有成語。」（小寧檔案，910116）

「分數很高」常是吸引學生青睞的重要因素。對學生而言，似乎分數的增加才是進步的最佳證據。然而，對於評量成果的處理歡老師著重於質的回饋。學生的作文成品，歡老師以同儕討論的方式進行分享與回饋，明確的告訴學生優點在那裡、可以改進之處在那裡。她將每一篇文章中優美的句子、有待改進、且值得提出討論的修辭法做記號，運用實物投影機一篇一篇呈現、一句一句說明，讓學生知道小逸的比喻用得很好：「那些房子就好像是玩具們的家……」、小傳的開頭很特別：「故事發生在小男孩亮亮五歲那年耶誕節的前一天……」、小遠的排比句法用得很好：「小男孩在雪地裡走來走去，在樹上搖來搖去……」、小耀的連接詞用得很妙：「雪人一進到廚房首先摸摸水龍頭，後來玩玩肥皂瓶，結果一堆泡沫跑出來……」、小君的形容詞很豐富：「各式各樣的雪人、成群結隊的海鷗、小巧玲瓏的房子、豪華的油輪……。」（觀察，901218）

當投影機投射出小智不太有結構的句子－「……那邊有許多雪人跟科南跳舞，有的彈鋼琴、拉小提琴、跟搖鈴鼓的雪人……」，老師也請大家提供修改意見，最後大家建議他可修成：「……有的雪人彈鋼琴、有的雪人拉小提琴、有的雪人搖鈴鼓……。」（觀察，901218）歡老師總是讓優秀者有機會展現他的作品，讓需要改進者了解修正的途徑在那裡。聰明的學生還會告訴我：「老師還會強調開頭要怎樣寫比較好，還會給我們看寫得比較好的人的作文……」（小傳訪談，901225）

討論過後，歡老師亦於趕課中騰出一節課，讓學生進行同儕書面回饋活動，這樣的活動學生擁有再一次的學習機會。對於學生辛苦的回饋記錄，歡老師更是不會錯過評量的時機，在學生的回饋紙條上留下成長與進步的證據。

（三）評量鷹架型塑出難以仿效的個人風格

我全程參與了「雪人」活動，體會到如此的歷程帶給師生的收穫比命題作文還來得豐富，然而，這卻是未經歷者難以理解的。歡老師赴校外演講發表寫作活動，回來總是告訴我，她覺得外人似乎抓

¹ 研究者註：這一學期共進行了六梯次不同形式寫作活動，分別是遊記、故事寫作、書信評量、議論文、及搭配閱讀活動小說編擬。其中書信評量歷程亦可列為「活化」教學與評量的主題，其實施歷程、評分規準、修訂回饋之實踐確切，然因字數限制，無法詳加敘述。

不到精彩之處。對於只聽到陳述或僅看到教案的老師來說，更不容易成功仿效。

實習老師美美提及別班老師認真的問她：「你們『雪人』是不是弄很久？」而這句「弄很久？」是持負面的態度（訪談，910116）。歡老師接著說，有另一位同事跟她借「雪人」影片，同樣帶小朋友到視聽教室看影片，也寫一篇作文，但是那位老師卻很生氣地跑來跟歡老師抱怨：「學生都寫不出來，寫哪個什麼東西嘛！」（訪談，910116）原來這位老師平常是讓學生把作文先寫在稿紙上，經過她的修改，再謄寫於作文本子。至於「雪人」作文呢？因為她並未有如歡老師的提問引導之鷹架，且只讓學生看了一次，就讓學生書寫，效果可想而知，所以她改不下去！甚至後來不改，學生也不必再謄寫了！

對於這樣不同的觀點，我們可以想見，層次豐富與步步攀爬的評量與教學之交錯互動，是成功與否的關鍵。

（四）評量活化學習的歷程

「雪人」活動，最初的目的在於寫作評量，但是因為形式的需要，歡老師又加入教學的因子，教學與評量於是成為「雪人」活動一體之兩面，而過程中融入的檢核機制，如提問、觀察、草稿的校訂，展現了評量多元面向之特色。歷程中的引導則是喚醒實作表現所需的先備能力，也可以比擬成實作評量的反應形式之建構(response format, Brown & Shavelson, 1996; Ruis-Primo & Shavelson, 1996)。無論教學或評量，教師所給予的鷹架支持是不容忽視的，尤其面對一個具有挑戰性的任務時，學生需要不斷的去試驗、蒐集資料、建立假設、並且評估自己的工作進度（Linn & Gronlund, 2000）才容易促進表現品質。評量任務完成的同時，教學也告一段落，評量與教學同時發生，評量本身所促進的，是老師的教學能力，也是孩子的學習過程。

在歡老師的教室中，學生與學生或學生與老師的一問一答之互動過程，顯現出學習的意義，也蘊含了非正式評量的機制—提問與觀察以及口語回饋，這樣的策略呼應了Stiggins（2001）建議的「審時度勢的評量」（size up assessment）。再者，此過程亦可為正式評量（作文成績）前的引導歷程，更重要的意義在讓學生經歷故事寫作歷程的鷹架建構，而這個鷹架建構過程中，評量與回應，不斷交錯且隨著學生的能力增進而步步攀爬層次，學習就此產生，帶領學生完成有挑戰性的任務，評量與教學歷程相互激盪出活化之圖像。

二、多元評量的迷思困境 數學領域的作業單與複習

除了上述具有活化特質的評量事件，某些評量活動的規劃就無法如此妥善順利，即使歡老師盡力改善評量方式，卻仍處處呈現迷思的現象。「迷思」的主因來自歡老師的力不從心與不知所措。這些評量策略鋪陳多元評量的情境，然而卻欠缺了解促進學生學習的因子，故我們將其稱為「迷思」的評量策略，其中最甚者屬數學科的評量。

（一）情境佈題的迷思

為了迎接期中評量的到來，歡老師加倍複習，習作、作業簿、學習單、測驗卷……等，一題一題引導。有時候，即使自己是數學題目上的主角，即使老師已講解過一次，學生還是茫茫然。小嘉就是如此，面對有自己名字的數學題目：「小嘉要將 10 朵花綁成一束，一小時可做七束，現在有 420 朵花，小嘉需要多少小時才能完成？」我看到小嘉的表情，似乎臉上浮出一層無奈，是主角又如何呢？

老師請問學生怎麼算，學生們七嘴八舌：「420 除以 10 乘 7……」、「括號……」、「括號 420 除以 10」，歡老師於黑板上寫下： $420 \div (10 \times 7) =$ ，接著問是否還有第二種算法，小逸答：「 $(420 \div 10) \times 7 =$ 」對於這兩個算式，歡老師並無解釋差異，僅是算出答案讓學生知道答案是不一樣的。引導過程的提示侷限在式子的確認，但無思考的引導：「一小時做 7 束，幾小時做完？怎麼算？乘以 7 還是除以 7……？」（觀察，901016），並未刺激學生思考為什麼要乘七或除以七，因此，解題過程中對於老師的提問，學生不斷的出現錯誤的回應。歡老師心一急，又換個方式重新解釋：「……現在我有 42 元，你

做一小時可以拿走 7 元，那你要做幾小時才可以把這些錢拿光？」……所以同樣的意思，在這邊總共有 42 束花，一小時只能做 7 束的時候要除以 7……」（觀察，901016）

歡老師將學生名字帶入题目的設計是爲了佈置题目的情境讓學生進入题目核心，以利解題。此現象如同林文生所說的：「將解題當作情境佈置，將問題做爲達成課程目標的工具」（林文生，1996，頁 21），也就是學生只要會解這些题目，就視爲達成教學目標的證據。此爲工具傾向的佈題方式，教師雖然引出了三種不同的算式，卻沒有比較或解說三種算式的由來與異同，學生學習的是解題技術而不是思考。所以，當老師解說完畢後，坐在我前面的题目主角小嘉左右觀望，還是不知如何下筆解出第十題，我小聲的問她：「你會不會寫？」她有點心虛但好強篤定的說：「會啊！」然而直到下課，她仍遲遲未動筆（觀察，901015）。

評量時，學生將教學中習得的解題方式應用於許多類似題型之上，中下程度的學生，如果理解力不足、記憶力不夠強就難以在評量中獲得好成績，練習也就流於記憶；從學習單、平時測驗、到期中評量，等於再製的迷思。學生只要聽到練習或複習，便沮喪又無力。如此刻板的數學科評量相形於語文評量設計的縝密與生動，顯得貧乏且僵化，對於這樣的落差，歡老師認爲那跟自己的專長有關。

（二）內容知識的限制與迷思下的憤怒

語文是歡老師的興趣，也是擅長的學科，寒暑假時間歡老師的備課內容亦多以語文爲重心，故她可以將評量與教學做統整性的規劃，而數學是她較無興趣且生疏的科目，能做的變化相對減少，甚至連在解題的引導，她也難以想出適切的指導語或佈題情境引導學生有效的學習，評量更只能侷限在紙筆測驗內的小幅變化。「數學教學內容知識」的專業素養影響其評量策略的展現。Shulman（1987）就主張：教師本身對學科結構必須有深度的了解，他才能知道領域中重要的概念與技巧爲何？新概念如何增加（p.109）？

面對數學評量的瓶頸，歡老師有自知之明，也把握進修的機會。同時我也在下課後與歡老師討論是否有突破的可能。雖然如此，相較於語文，其數學教學或評量處理的敏銳度就有明顯的落差，數學課引導時，她的引導語句樸素而無動力，多以計算性的指示作爲引導語，難以激發學生腦力激盪的興致。如當學生對【 $420 \div (10 \times 7) =$ 】與【 $(420 \div 10) \times 7 =$ 】無法分辨時，歡老師並不覺得還可以利用圖示法、或是日常生活情境的討論等方式。

檢討考卷時，第一組的小仁一直分不清分子、分母的位置，歡老師請組長小俊負責解說，小俊站起來大聲的說：「媽媽背著小孩，所以下面是分母，上面是分子。」（觀察，910108）這樣的比喻是歡老師爲幫助同學記憶所用的聯想法。這樣的硬記是有風險的，學生若不清楚分子、分母關係的概念，媽媽與小孩的連結也幫不上忙，學生無法遷移至較深、較靈活的應用題。

重複練習之時，一些已習得的同學會發出疲累之聲音：「ㄝㄚ！」對於這些同學的不平之鳴，歡老師會說：「可是小辰還是不會算啊！」但是對於不會的同學若還發出疲累的聲音、露出不情願的表情，歡老師則會氣憤的說：「還好意思給我叫！如果不是你們不會，我需要複習嗎？……」（觀察，910108）此時全班鴉雀無聲，看得出來學生已覺察到老師的不悅，識相地乖乖端坐聽訓。

我同情學生的無辜，卻也能理解恨鐵不成鋼的教師心態，我體會到歡老師的行爲是「愛的另一種模樣」。

（三）教學知識的不足強化了迷思

當課程內容有實物可操作時，歡老師也會設計與實作有關的評量，如：時分的計算，請學生帶鬧鐘，實際操作鐘面；重量的測量，用磅秤實作；面積計算，利用小方格紙替代，模擬估測。然而，在解題方面，歡老師能夠設計的實作評量方式就有限，難以發揮。這樣的限制來自於教學知識的不足，Fennema and Franke（1992）認爲數學教學知識是數學教師應具備的教育專業知識（knowledge of

pedagogy)，了解數學概念如何形成，另外，需了解到自己「會」數學與「會教」數學是有區別的，也才能了解學生如何思考數學（引自柳賢，2002，頁3）。學科知識的不足導致課程規劃的缺失，再加上教師數學教學知識的薄弱，在評量上就呈現出無力感與僵化的面貌。從專長的限制到教學約制，再到評量的瓶頸，歡老師的數學科評量呈現出無奈的迷思畫面，即便數學題目以孩子為主角，學習動機仍無法提振，評量走入了死巷。

三、多元評量的神話圖像 社會領域的臺灣地理大王

歡老師對於多元評量的想法持相當開闊的態度，只要某個方式能夠幫助學習，它同時也可能是評量的手段。爰此，歡老師設計的許多教學活動相當生動有趣，這些活動也理所當然被賦予「評量」的期待。以歡老師的觀點視之，因著激起學習動機的意圖設計活動式的評量，實施之後，學生樂於參與，老師也有成就感，師生皆大歡喜。以觀察者的點視之，這些評量活動確實有助於學習的氣氛，但是若從「了解」的角度來檢視這些活動，干擾「了解」的無關因素似乎過多，而模糊了評量或是學習的焦點。歡老師若以為並期待這些評量活動在激發學習動機之餘，尚能發揮促進學生「了解」的實質功能，則就落入把多元評量視為萬靈丹的「神話」之誤解。

社會領域的「臺灣地理大王」之學習，足以代表歡老師「神話」了多元評量的具體實例。

「大富翁這個，我其實沒有設到那麼嚴謹，我是讓他們以遊戲性質為主，我只是當作是複習功課用，我評量的目的之一是幫你複習……」（訪談，901029）

社會領域學習內容有些需要記憶，在沒有複習的情況下，久了就忘記。月考前，為了增進學生的記憶，歡老師規劃了「臺灣地理大王」活動。

（一）激活學生熱烈迴響的闖關評量

「臺灣地理大王」的內容設計是以臺灣、金門、馬祖的地理環境為主。關卡的設計以社會課本重要內容為題目，學生依答對的項目多寡積點。有些課本上的內容，歡老師認為不需要背或者是不容易背，她則將其設計為「休息：念一念」或「送分題」一唸完就可以得5點。

闖關過程中運氣比較差的學生還會遇到「危險地帶」，當不幸走到這一格時，學生剛剛所得的點數全被大水沖走了，就得從0開始計分，有些學生因此扼腕不已，直呼倒楣。除了令人害怕的危險地帶，還有令人振奮的「超值區」，計分為原始分數的三倍。當學生走到此格時多會發出歡呼聲：「YA！……太爽了！……讚啦！……」（觀察901030）。這一節課教室裡熱鬧非凡，驚訝聲、叫好聲、拍手聲不斷，放眼望去沒有一個學生不全神投入。

闖關大富翁的活動，歡老師的原始意圖是為了解讓學生記住課本的內容，既然要讓學生記住課本的內容，歡老師細心的想到，課本寫一大堆，學生彼此檢核答案時得花時間一頁一頁找，而且還有找錯之危險，於是她又細心的為學生設計一張「答案卡」，讓同學一目了然。思慮周延的歡老師，就是這樣規劃著評量步驟，連答案卡都不會忽略。學生在歡老師的安排之下，從闖關評量中不知不覺的記住了平常難以記住、也不願主動去背誦的社會知識。小君開心的說「不但可以玩遊戲，又背得起來，又可以蓋好幾個章！」（訪談，901204）。

歡老師原本期待藉此活動「活化」學生的學習，也將此視為是多元評量的一種。從學生的角度看這個活動，對其熟記課文確實有正向的推進作用。過程中不但加深內容印象，也充滿樂趣，讓教室的充滿快樂的學習氣氛。然而，從「評量」的角度看這樣一個書面闖關評量，我們有著與歡老師不同的見解。

（二）「了解」學生學習成效的證據不足

歡老師曾言在設計評量時她的首要考量是學生不排斥且能快樂參與，所以有些評量是比較偏遊戲性質。在她的信念裡，學生快樂學習是教學、評量首要考量，快樂學習之後才能事半功倍。闖關評量為

小學中、低年級之最愛，原因無他，就是有趣、快樂。然而評量形式變得多樣而有趣絕不是目的，因為好的評量是目標依賴的（吳鐵雄、洪碧霞，2000）。有具體明確的目標，教師才知道學生究竟會了什麼，還有那些需要努力，失去了目標的指引，老師對學生的了解可能會變成「誤解」，誤解之後的危險就有誤導教師教學決定的可能。

歡老師視「臺灣地理大王」為評量的策略之一，老師加分的標準以學生在活動得到的點數多寡決定。第一輪的比賽中，首先得到 50 點者受封為「臺灣地理大王」，也會獲得老師平時成績加分鼓勵。但是仔細檢視得分標準，可以發現學生得分高低絕大部分與運氣有關，如走到超值區者，輕而易舉就可以得到以倍計算的點數，反之，不幸走到危險地帶時，分數被大水沖走，也只能自認倒楣。因此，點數獲得的多寡與是否熟記課文並無必然的關係。

在回歸評量的本質和目的上，無論是Linn and Gronlund（2000）說的：「廣泛的收集學生學習的訊息，以做助於學習歷程的價值判斷」（p.31），或Wiggins（1988）所言：「以有系統的方式培育學生『自我修正』（self-correct）的能力，而不是作為學習結束時的稽核工具」（引自張麗麗，2001，頁3）。評量不脫「了解學生」、「幫助學習」兩個目的。「臺灣地理大王」是一個能夠幫助當下記憶的活動，只是活動的歷程中，教師無法兼顧各組學生的表現情形，之後收集到的僅是學生活動歷程累積的點數，這些訊息對於老師了解學生學習情況的助益是微弱的。在老師了解學生學習成效的證據不足之下，老師無從肯定學生的學習成就，及提供適性的補救教學，相對的，「雪人」寫作評量，有確切的口語互動、評量目的、書面成品的產出，更有成品的追蹤修訂。相較之下「臺灣地理大王」所產出的結果點數讓老師僅知道甲生得了 30 點，並不知道他是對臺灣的山脈分佈地點不了解，還是對臺澎金馬的位置排列不清楚。再者，若僅以得分點數做為鼓勵的依據，學生著力的可能就不是把內容背起來，而是如何才能避免走到危險地帶，如何才能走進超值區以獲得高點數。如此一來，幫助學生記憶的美意也會有所減損。活動之後老師也沒有足夠的訊息參考，給予學生有助於學習的回饋，活動結束就是一切的終點。在無法提供充足的了解證據之下，亦造成無法提供教學回饋，那麼此項活動若欲以「評量」稱之，評量實質功能在此就受到了約制，無從彰顯，也無發揮的空間，流於視多元評量為萬靈丹的「神話」了。

（三）神話的暫時快樂

評量活動熱鬧、有趣是師生樂於見到的表象。就學生的反應而言，他們玩得不亦樂乎，熱烈的迴響，這些活動的魅力是無法抵擋的，研究者亦不否認它也可能有機會激發學生的學習動機。但是，活動過於花俏，恐有淺化知識之虞，使得學科知識殘缺不全（歐用生，2002）！教學活動、評量作業、計分規準如果與欲評估的能力面向脫節，極易造成評量的預期目的不是霧裡看花，就是如天邊彩虹一般，遙不可及（張麗麗，2002）。這樣的憂慮與質疑，其實歡老師是清楚的，活動過於快樂，可能會迷失評量目的，是故，她也認為此活動有再修正的空間。有趣、快樂是評量附加價值的追求，但若喧賓奪主搶了評量目標的任務，就會模糊了評量的焦點。歡樂有趣多元的評量型式，包裝了內涵不足的學習鷹架，此處我們看到了歡老師神話了評量的形式與花樣，而模糊了評量作為學習促進的工具。

伍、結論與討論

一、結 論

我們在此提供一個圖像，讓讀者了解一位經驗豐富的老師處理多元評量時候各種糾結的力量，時而神話、時而迷思、時而展現活化的催化機制。

（一）悠遊於虛幻的神話中

當老師將多元評量視為萬靈丹的時候，就有將之神話化的可能。對歡老師來講，多元評量是一個廣義的概念，這個概念支持她追求另類評量的設計，她期待每一個評量都能激發學生學習的興趣、並於評量時體會快樂的滋味。然而，當快樂蒙蔽評量目標致使評量失去方向時，就徒留快樂而難以獲致

了解或協助學生學習的目的。評量要發揮協助學習或教學的功效，就必須緊扣教學目標，否則就會流於空泛而削減支持學習的力量。因此，目標不在多、不在廣，而在反問自己「什麼是我們希望學生具備的能力？我們又如何知道學生是否具備了這些能力？」（張麗麗，2002）職此之故，我們期待教師除了涉獵寬廣的實務知識外，還能對理論知識有所了解，如此，教師的教學實踐才可能是「方以智」的功夫，又是「圓而神」的藝術（甄曉蘭，2000）。

研究者亦於教學現場觀察到某些親、師、生總把闖關評量視為多元評量的代名詞，評量遊戲化的結果即可能神話了闖關的價值。那就像為多元評量打造一個虛幻的神話天堂—也許我們真的如此期待它如天堂一般的美好，可以讓學生快樂參與、有效了解學生能力、還可以幫助學生學習。但是過度完美的想像，也許下一刻下一秒迷思就登臺演出。

（二）墮入濛霧中的迷思

迷思來自教師心有餘而力不足的茫然、與不知所措的迷惘。歡老師是語文領域的專家，數理領域的生手，規劃的語文科評量，歷程完整，兼顧了解學生與協助學生學習的評量功能，但是她卻無法設計出如語文一般豐富的數學科評量，僅能沿襲紙筆測驗或稍加變化。即使意識中知道還有努力的空間，即使她知道學生無法快樂學習，仍然難以突破，不知從何思考、從何著手，其良善的意圖就因為缺乏學科知識的目標，如同歡老師沒想到討論亦可以有數學感覺的引導，就在這樣的學科內容以及教學內容之專長限制中，迷失了。世上沒有完美無缺的評量設計，也沒有放諸四海皆適用的評量方式，一位有心嘗試多元評量者也必須是個願意持續學習者，多元評量不是一項技術，習得就可終身受用，它是一個理念、一個精神，只有不斷精進本身對於學科的知識、以及教學的知識，使得從濛霧中撥雲見日找到活化的出口。

（三）撥雲見日的活化

多元評量如何能擺脫神話的誤導或迷思的茫然，對教學發揮活化機制呢？從歡老師的評量規劃，我們看到活化機制蘊含在評量歷程中多方資源的運用—包括人員的運用（班群、學生、家長）、媒介的運用（錄影帶、VCD）、時間的運用（晨光時間、午休、寒暑假）、進修機會的掌握等。再加上對於學科的熟稔而產生的引導策略，將評量支持教學的功能發揮得淋漓盡致，雖然活動重複出現（如觀看數次「雪人」影片），但一次一次精緻化的鷹架，讓學生一次一次地深入探索自己潛在的能力，使學生在每一次的引導中習得更高層次的能力。評量歷程的縝密規劃則是歡老師活絡教學歷程的鎖鑰，歷程中掌握規準的擬定、檢核機制的督促與教學目標的完善結合，活化的機制就在這樣一連串的引導與一層層的深化中發酵。

多元評量活化教學是一個理想也是一個目標，但是過高的期待、不切目標的作為就容易流於神話，缺乏豐富的學科知識就可能落入迷思。透過歡老師的個案研究，我們認為技術的仿效容易，但是精神的掌握就得看個人的用心與功力。

二、討論

研究歷程中，有許多值得深思的議題，在此提出探討並與讀者共勉於評量活化教學之路上。

（一）期待優秀而非優異的行動者

研究結果發現，歡老師可以讓語文科的評量活化教學，卻無法活化數學科的評量。在語文引導方面她很優秀，但並不是優「異」，她只是比一般教師多想一點、多走一點，她不認為自己是個特殊的行動者，亦不願被視為特例，誤以為她得天獨厚或天賦異秉，使得有今日的作為與名氣。她其實就與一般教師沒有兩樣，在實施多元評量的歷程也會有神話、迷思、活化糾結的時候。只是其在有限資源的協助下，能妥善運用手邊的資源使成為有助於教學的利器，創造生動的教學與評量。這樣的成就不代表他人無法超越，而欲凸顯的是「有為者亦若是」的希望，只要有心為之，甚至可能臻於比歡老師更

佳之境界。

（二）多元評量呈現個人專業素養的實力與風格

歡老師告訴我：「你就要回學校教書了，你可以把你收集到的評量策略或學習單回學校用用看，看看有什麼不同……」（訪談，910411）當我身爲一個旁觀者觀察歡老師施展多元評量策略時，心覺佩服卻也興起仿效之意，雖然也耳聞他班老師仿效失利的實例，我仍好奇於在取法專家教師策略時所需考量的班級文化差異爲何？爲此我將雪人之引導寫作帶入自己新接班級的作文課，並曾被中國時報報導過（賈先發，2004），更利用電腦融入資訊教學，提供學生在網站上互相觀摩作品或意見交流的機會，然而學生寫出來的作文還是與歡老師的學生有天壤之別。主要差異在於歡老師的學生已受過一整年的語文指導，再者，歡老師對於語文教學的興趣與進修是熱切且與時俱進的，在其熟稔語文教學策略之際，同時也能構思出檢核此學習成效的評量機制。教師精熟的教學領域不同，規劃出來的教學與評量策略即會有差距。

（三）行動者迷、研究者清

歡老師有一次至苗栗演講多元評量實務（910410），她將「雪人」的評量規劃、實施結果與在座老師分享，現場老師有人反應：「那樣太浪費時間了！影片看一次和看五次有什麼不同，學生寫出來的東西還不是一樣……」（訪談，910411）。面對這樣的反應，歡老師回臺北後向我吐露了無奈的感慨以及多元評量的窒礙難行。然而，我卻持不同的看法，我認爲歡老師尚未能將其評量的精髓與主軸以適切的方式傳達給聽眾；再者，歡老師並未意識到「雪人」評量的成功，在於那層次豐富的提問鷹架，這是她不自知的關鍵點，身爲研究者的我卻可透析出那六個層次的提問所形成的非正式評量之功力，了解精彩面向之所在。演講後的隔天她看到我對她的詮釋之後高興的表示，如此的分析，讓她洞察到自己行動的核心概念。

歡老師是一個專家教師，專家教師擁有熟稔的教學技巧與評量策略，這些知識都內化在其心中。但內化之後的知識就變得自動化，自動化可以讓人不假思索提取知識即時運用，自動化之後對於自己後設思考的敏銳度也相對減低，因爲一切行爲在其看來都視爲理所當然。此時，研究者的介入正是挖掘、釐清、汲取這些行動者自己都難以說明的寶貴結晶時刻。這些鑄鑄之後的結晶可以讓讀者清楚了解行動者內在的結構，也可以反映給行動者做後設省思的參考，個案研究的價值就在這裡發生效用。

（四）有效化證據之期待

若以 Messick（1992, 1995）的效度漸進矩陣來看，多元評量對於學習的「活化」，可以歸類在效度漸進矩陣中，測驗使用的後果基礎上（吳毓瑩，2004，頁 86）。若就《教育與心理測驗歷程的標準》（Standards of Educational and Psychological Testing）中所探討的五類證據而言（內容、反應歷程、內部結構、其他變項關係、以及施測後果）（AERA, APA, & NCME, 1999, pp.11-17），亦可屬於測驗實施之後的後果影響面向上。由於本研究乃從教學角度出發，「有效性」證據並不是本文關懷的重點，然而評量對於教學的活化與評量對於能力詮釋的有效性，是探討學習歷程的兩個重要關鍵。有心探究多元評量的研究者尚可進一步深入解析，以回應眾人對於多元評量在效度議題上的關注。

致謝 本論文之完成，感謝國科會研究計畫 NSC 91-2413-H-152-011 之支助，更感謝歡老師對我們知無不言言無不盡的交心。雖然文中只有一位主角，但是這位主角，讓我們看到現場老師豁達的胸襟與溫暖的懷抱，不只教書、教人、更是教心。

參考文獻

- 王文中 (2000)。專文推薦：擴展多元智慧評量。載於郭俊賢、陳淑惠譯，**落實多元智慧教學評量**，頁 13-15。臺北：遠流。
- 江文慈 (2004)。教學評量改革的難題分析：技術、文化、政治、與後現代觀點。**課程與教學季刊**，7 (3)，1-16。
- 呂金燮 (1999)。實作評量——理論。載於王文中、吳毓瑩、呂金燮、張郁雯、張淑慧著，**教育測驗與評量——教室學習觀點**，頁169-203。臺北：五南。
- 李坤崇 (1999)。**多元化教學評量**。臺北：心理。
- 吳毓瑩 (1995)。開放教室中開放的評量：從學習單與檢核表的省思談卷宗評量。**開放社會中的教學**，頁 93-100。臺北：國立臺北師範學院。
- 吳毓瑩、桂怡芬 (1997)。形成性評量效度的驗證及教師的角色——以自然科平時評量為例。載於黃政傑主編，**當代師資培育的課程教材教法**，頁 177-198。臺北：漢文。
- 吳毓瑩 (1998)：我看、我畫、我說、我演、我想、我是誰呀？——卷宗評量之概念、理論、與應用。**教育資料與研究**，20，13-17。
- 吳毓瑩 (2003)。多元評量之解毒與解讀。**課程與教學季刊**，6 (1)，133-154。
- 吳毓瑩 (2004)。效度。載於於王文中、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**。臺北：五南。
- 吳鐵雄、洪碧霞 (2000)。教學評量的反思與建言。**測驗與輔導**，159，3323-3324。
- 林文生 (1996)。一位國小數學教師佈題情境及其對學生解題交互影響之分析研究。國立臺北師範學院國民教育研究所未出版碩士論文。
- 林怡呈、吳毓瑩 (1993)。多元評量概念在課程標準演變中的定向與意義。**教育研究資訊**，11 (6)，3-32。
- 林清山、林素微、鄒慧英 (2001)。國小高年級數學溝通檔案評量發展初探。收錄於中國測驗學會與臺灣師範大學主編，**第五屆華人社會心理與教育測驗學術研討會：多元文化中的測驗發展論文集**，頁 1-12。臺北：中國測驗學會。
- 柯政毅 (2001)。**數學歷程檔案評量實施下學生數學溝通能力——以六個個案探討**。國立新竹師範學院數理教育研究所未出版碩士論文。
- 侯雅齡 (2000)。如何有效運用多元評量的結果。**測驗與輔導**，159，3336-3341。
- 柳賢 (2002)。**數理科教師教學能力指標與評鑑工具研究**。上網日期：民國 91 年 5 月 22 日。網址：<http://www.nknu.edu.tw/~edu/item/item3-article7.htm>。
- 莊明貞 (1998)。真實性評量在教育改革中的相關論題：一個多元文化教育觀點的思考。**教育資料與研究**，20，19-23。
- 莊明真 (1999)。多元的動態評量與教學——從維高斯基觀點談起。**教師天地**，99，25-31。
- 張麗麗 (2000)。「什麼樣的檔案才值得…？」——談自主學習檔案的基本要素。**屏縣教育**，4，5-12。
- 張麗麗 (2001)。**跨世紀教學評量改革：實作評量現況、模式的建立與實踐**。國科會專題研究計畫成果報告。編號：NSC89-2420-H-153-002-F-19。
- 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。
- 賈先發 (2004)。分享經驗北斗國中小結盟。**中國時報**，5月15日，C2 版。
- 甄曉蘭 (2000)。新世紀課程改革的挑戰與課程實踐理論的重建。**教育研究集刊**，44 (1)，61-89。
- 蔡敏玲 (1996)。詮釋性研究的一個可能方式：我如何建構婷婷和穎的故事。載於黃政傑主編，**質的教育研究：方法與實例**，頁 71-113。臺北：漢文。
- 歐用生 (2002)。建立二十一世紀的新學校——校本課程發展的理念與實踐。收錄於歐用生、莊梅枝主編，**邁向課程新紀元 (十)：反省與前瞻——課程改革向前跑論文集**，頁 6-26。臺北：中華民國教材研究發展學會。
- 盧雪梅 (2000)。多元教學評量的理念和實務。**教育研究月刊**，76，57-66。

- 簡茂發 (1999)。多元化評量之理論與方法。《**教師天地**》，99，11-17。
- Airasian, P. W. (1994). **Classroom assessment**(2nded.). McGraw-Hill, Inc.
- American Educational Research Association , American Psychological Association, and National Council on Measurement in Education. (1999). **Standards for educational and psychological testing**. Washington, DC: American Educational Research Association.
- Brown, J.H. & Shavelson, R.J. (1996). **Assessing hands-on science: A teacher's guide to performance assessment**. Thousand Oaks, CA: Corwin Press.
- Connelly, F. M., & Clandinin, D.J. (2000). **Narrative inquiry: Experience and story in qualitative research**. San Francisco: Jossey-Bass Publishers.
- Danielson, C., & Abrutyn, L. (2000)。《**檔案教學**》(鄭英耀、蔡佩玲譯)。臺北：心理。(原著出版於1997年)
- Fennema, E. & Franke, M.L. (1992). Teachers knowledge and its impact. In D. A. Grouws (Ed.), **Handbook of Research on Mathematics Teaching and Learning** (pp.147-164). New York: Macmillan Publishing Company.
- Gross, S. J. (1998). **Staying centered : Curriculum leadership in a turbulent era**. Washington, DC: ASCD.
- Linn, R. L., & Gronlund, N. E. (2000). **Measurement and assessment in teaching**(8th ed.). Upper Saddle River, NJ : Prentice-Hall.
- Messick, S. (1992). Validity of test interpretation and use. In M. C. Alkin (Ed.), **Encyclopedia of educational research** (pp. 1487-1495). New York : Macmillan.
- Messick, S. (1995). Validity of psychological assessment : Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. **American psychologist**, 50(9), 741-749.
- Puckett, M. B., & Black, J. K. (1994). **Authentic assessment of the young child—Celebrating development and learning**. New York: Macmillan College Publishing.
- 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.
- Shulman, L. S. (1987). Knowledge and teaching : Foundations of the new reform. In A.C. Ornstein & L.S. Behar-Horenstein (Eds.), **Contemporary issues in curriculum** (pp.103-121). Needham, MA : Allyn & Bacon.
- Stiggins, R. J. (2001). **Student-involved classroom assessment**(3rded.).Upper Saddle River, NJ:Prentice Hall.
- Stake, R. E. (1995). **The art of case study research**. London : Sage.
- Sternberg, R. J. (1998). Abilities are forms of developing expertise. **Educational Researcher**, 27(3), 11-20.
- Wiggins, G. (1989). A true test : Toward more authentic and equitable assessment. **Phi Delta Kappan**, 70(9), 703-713.