

如何（不）個別評改專題報告

如何（不）個別評改專題報告

侯傑泰 香港中文大學教育心理學系主任

利用專題報告等作為學習或評核學生能力的方法，在香港或世界各地漸受重視及流行，但對頗多教師來說，如何評改這些報告以協助學生學習，確是一個難題。

本文主要是介紹這考核改革的潮流，並引用 Goodrich(1998)所建議的方法，描述如何以最少的批改，或甚至不作極為費時的個別評閱，但更能促進學習，也真正發揮以專題報告等學習的優點。我們希望指出很多時候學生主動個別學習的機會及時間，比被動聽教師講授更能促進學習；個別評改可能是一個高勞動成本但低報酬的教學活動，本文提出一套更省儉，且能真正發揮專題報告教學果效的方法。

表現性、真確 / 另類(performance, authentic/alternative) 評估

長久以來，評估方法極為影響及控制教學，這是不爭事實。尤其是在七十年代，美國的問責潮流導致各州制定大型強制性學科測驗，以審定各州或學區的學業成就，因為老師及學校均極為重視這些考試內容，加上傳媒大事宣傳考試成績之比較，不言而喻，這些學業測試被描繪成為‘課程磁鐵’(Popham, 1993a, 1993b)，主宰課程改革方向。不過在八十年代，公眾及學者越來越關注傳統測驗對課程改革及課堂教學的負面影響，多項選擇題及常模參照性的測驗最受批評，論者認為我們需考核及重視解難、分析及較高層的思維能力

(Cizek, 1993)。對於傳統測驗，學者亦關注如何制定成就標準(standard setting)、測量特性、公平性及如何編制成績報告等。

論者認為標準測驗應作大幅修改，以符合真正教學內容及具體行為表現，而非以常模作評核標準(Baker, O'Neil, & Linn, 1993)。這種評估方法可簡單統稱為表現性，真確或另類測驗。它集中於仔細訂定學習目標，編制測試題目，考核及紀錄學習過程、表現、成品及文件夾(portfolio)，並決定各成就等級的指標(Madaus & Kellaghan, 1993)。學生在連續多次考核情況中的表現，被視為較單一測試更為重要。舉例來說：(一) 學生需提交一份作業‘比較各種公共交通工具對社會的貢獻’，他們需搜集各種交通工具的數目、票價、對路面佔用情況、使用率等數據並作比較評論。(二) 學生需設計並進行實驗，以比較三種紙巾的吸水能力。學者希望這類表現性測驗，能測量較高層次如解難、分析及研究能力，簡而言之，表現性測驗是著重模擬及量度被試在真實情況下所需的能力及技巧。

表現性測驗雖然甚具吸引力，但亦有頗多局限及弱點，批評者要求表現性測驗能符合心理測量及統計要求，以便能準確將被試分等級，並敏感地反映能力的細微改變。它們亦應量度真正成就。再者，在訂定合格分數或表現指標時，需合乎理論及測量效度。

表現性評估亦被批評欠缺公平，所測的知識不能轉移，時間及經濟上並不可行(Bracey, 1993; Linn, 1993; Worthen, 1993)，一如 Berk (1986)所總結，“釐定表現性標準常受批評，引來甚多爭論，難以執行，而差不多完全無法自辯”(p.137)。Madaus 和 Kollaghan (1993)更為悲觀，他們認為“表現性測驗，漸離直接量度複雜的行為表現，其生命不會長久”(p.469)。

不過，為對表現性測驗有更公平的評價，我們可能需要更廣泛的評估標準，例如，Linn, Baker 和 Dunbar (1991)相信，“嚴謹評估表現性測驗需包括：期望及非期望的結果，

在特定評估所表現的能力可否轉移……，測驗的公平性……，學生解決困難時所涉及的複雜智力……，試題對學生及老師是否有意義……，判定內容質素及涵括面的基礎……及測量的成本是否合理等”(p.20)。總的來說，雖然在方法及實施的可靠性上，表現性測驗均受批評，但這方法仍在國際上漸受歡迎及流行。

評改規程(scoring rubrics)

一如批改一篇作文或一件美術作品，評核專題報告的困難在於沒有一套唯一的標準答案。我們知道評核準則極為影響學習目標，如果我們以圖畫的填色是否超越界線，或作文的錯別字次數作計分標準時，則難以培養用筆大膽的畫家或富於創意的小作家。所以當要設計專題報告的評改準則時，我們必須釐清這類報告的學習目標，一些只重視報告長度(頁數)及美術包裝的評核標準，可能會錯誤導引學生將所有精力用於美勞剪貼上。

我們現以一虛擬例子描述一專題作業的評改規程(Goodrich, 1998)，假設我們要求學生做一專題書面報告，介紹他們一種發明或設計方案，以解決某項交通、醫療、環境保護、教育等問題。表一列出其評改規程，它包含優良的要素及優劣的定義。四項要素為：發明的目的、發明的內容及如何達致其目的、發明的長處及缺點、發明與現今、過去或未來相關事項的聯繫。雖然表內列出各項得分，但只作為自評及同學間互評參考時，則無必要列出分數。若學生每年需多次呈交作業，教師亦無需依各分項評分，可省略給一個等級分數(如：甲至丙)便可。

表一位置

評改規程的功用

批改規程的功用甚多，包括(Goodrich, 1998)：

- (一) 清楚地指出老師的期望，成為學生改善報告的目標及依據；

- (二) 學生可依這些目標，更容易判斷自己及同儕報告的水平，這些自我及同儕間評估，提高學習動機及自我責任承擔，並培養自學態度；

- (三) 使用評改規程大大減低教師用於個別批改的工作量，一如下文詳盡指出，因學生掌握評核準則，教師無需個別寫上詳盡評語，若有必要，教師也可簡單引述規程上的參照編號便可，方法雖簡便，但較以往更能提供豐富的改進指引。

- (四) 教師儘量安排機會，讓優異作業得以在小組及全班中展示或報告，除可表揚卓越的同學，亦能以具體例子闡釋規程內容。

編製及使用評改規程

雖然編製一份詳盡清晰的評改規程並不容易，但規程間有甚多共通性，且每年每科所用的規程並不會太多，且可多年重覆使用，故也是甚有價值的工作，在編製及使用評改規程時可參考下述各點(Goodrich, 1998)：

(一) 展示及使用實際成品

評改規程可由教師自行編製，對能力較高及高年級學生，亦可透過同學大組討論產生，若使用討論方法時，先請學生找出同學間最佳及較差的作業，然後訂定優劣的準則及要素。

當然若組內作業並未足以完全反映各類等級時，同學亦可自行構劃出一個虛擬極為優越的報告，以填補規程內的空檔。

(二) 自評、互評及自我修改

老師及組長除了訂定及展示各規程的要素外，亦應帶領討論如何具體改善報告內容或編排的方法。如何應用評改規程甚富彈性，一般來說應包括一或多次自評互評及自我修改的機會。其中一個方法是先讓學生依據規程自我評核，讓他們在報告附頁內詳列他們對自己報告的評價及希望改善的地方。經修改後由大組同學互相評價及欣賞各人的報告，這互相觀摩的步驟甚為寶貴及重要，學習理論指出當學生能利用規程去協助及指導其他同學改善其報告時，他們的學習較為深入及全面，日後更能編製一份出類拔萃的報告。

在學生間互評後，我們亦應讓學生有機會再作修改，此時學習過程的大部份目標均已達至，而老師詳盡評改，已變得極為次要，老師所需付出的冗長時間與學生的學習回報也不成比例。當然為了鼓勵優越的報告，教師亦可能需要評定部份報告的等級。

(三) 準則及要素

在編製一份評改規程時，一般我們先找出一些最好及最差的報告，試圖列出一些分辨優劣的準則及要素（例如：結構、層次），描述最好及最差的等級（例如：『結構緊密、層次分明，構劃出一個完整理論模型』；『內容雜亂撲合，段落章節毫無脈絡關連』），然後在這兩端加插數個漸進的等級。在校內大部份報告，四至五級的區分可能十分足夠，當然為了鼓勵極為優越及需要額外照顧的同學，我們可能在兩端各加一個等級，以反映首尾 5%至 10%的同學，但我們需小心使用這些極端等級，尤其是如何以最劣等的標籤發揮其積極的鞭策作用，確是一門學問。

我們需知評改準則直接影響學生日後努力的方向，一個只注重外表美觀包裝的準則，可能令每份專題報告都變成美勞作業，同理，單強調資料性內容多寡的要求，則可能令日後報告均變成枯燥乏味的資料冊，所以如何包含及平衡所有重要標準極為重要。

(四) 宏觀及彈性標準

評改標準多是一些比較宏觀的重要準則，多以甚具彈性適用廣泛的詞藻作描述，例如，在一個交通工具的專題報告，優等的報告「能以客觀理據，指出某工具的優點，但亦能以中肯平衡態度，列出該工具的局限及弱點。」仔細的量化要求，例如：「報告需含五個長處及五個弱點」，多不適合，因為這只會導引學生在形式上滿足教師的要求。一個好的報告就像一篇好的文章或一個極富吸引力的故事一樣，永遠不拘一格，沒有一成不變的通則，故此要求學生在專題報告最尾一章或一節，作一個全面簡單總結固然是一通常慣例，但這也不能排除某學生以一故事帶出整個專題的重點，是一個更具啟發性的總結。尤其是對已熟習基本概念能力較高的學生，我們容許及鼓勵創新意念及嘗試也是極為重要的。

評改規程會依學生的能力及教學重點而不同，在低班及學習初期我們可能重視報告較基本的要素，循序漸進地我們會要求內容需要有理據支持，論點要邏輯合理。一般來說我們的評改規程盡量以正面角度描述優等報告需有的成份，而不是單臚列一些負面不應做的禁忌，正面準則能導引學生留意應加強的地方，單指出不應做的禁忌有時會令學生無從改善。

結語

專題報告在目標為本課程及其他考核改革中越來越受重視，它是一個頗能鼓勵學生以小組或個別形式有主動學習的好方法，我們不應對報告單視作一件成果，我們更應重視這作為培養學生自學態度的手段，故此提供足夠機會讓學生自評，互評及修改極為重要。考核的重要任務是協助學習，本文描述一個對老師來說相對較為省時有效率，但更能充份發

揮專題報告作為學習的方法，亦希望藉以培養學生自學態度及習慣，很多時候，學生的學習及吸收，不一定透過我們的直接課堂講授，安排及設計適宜的學習活動果效可能更為廣大持久。故此，個別批改專題報告可能是投資成本極高的工作，而課堂講授也可能是低報酬的教學活動，安排機會讓學生自評、互評及自我修改是批改專題報告關鍵所在。

參考文獻

- Baker, E. L., O'Neil, H. F., Linn, R. L. (1993). Policy and validity prospects for performance-based assessment. *American Psychologist*, 48, 1210-1218.
- Berk, R. A. (1986). A consumer's guide to setting performance standards on criterion-referenced tests. *Review of Educational Research*, 56, 137-172.
- Bracey, G. W. (1993). Assessing the new assessments. *Principal*, 72(3), 34-36.
- Cizek, G. J. (1993). Reconsidering standard ad criteria. *Journal of Educational Measurement*, 30, 93-106.
- Goodrich, H. (Ed). (1998). Understanding rubrics. In A. E. Woolfolk, Readings in educational psychology (2nd ed.). Needham heights, MA: Allyn & Bacon.
- Linn, R. L. (1993). Educational Assessment: Expanded expectations and challenges. *Educational Evaluation and Policy Analysis*, 15, 1-16.
- Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researchers*, 20(8) , 15-21.
- Madaus, G. F. & Kellaghan, T. (1993). The British experience with 'authentic' testing. *Phi Delta Kappan*, 74, 462-469.
- Popham, W. J. (1993a). Circumventing the high costs of authentic assessment. *Phi Delta Kappan*, 74, 470-473.
- Popham, W. J. (1993b). Educational testing in America: What's right, what is wrong? A criterion-referenced perspective. *Educational Measurement: Issues and Practice*, 12(1), 11-14.
- Worthen, B. R. (1993). Critical issues that will determine the future of alternative assessment. *Phi Delta Kappan*, 74, 444-454.

表一

虛擬評改規程

準則 較佳 ←..... 質素→ 較劣

目的	解釋發明的重要目的 及較不明顯的用途 (4)	解釋發明的 主要目的 (3)	解釋部份目的 , 但遺漏一些重點 (2)	沒有解釋目的 (1)
----	--------------------------------	------------------------	------------------------------	-----------------

特徵	詳列主要及隱藏的 特徵, 並解釋其作用 (8)	詳列主要特徵 及作用 (6)	遺漏部份特徵 及其作用 (4)	沒有盧列特徵 及其作用 (2)
----	---------------------------------	------------------------	-------------------------	-------------------------

批評	討論發明的長處 及弱點，並建議一些 補救方法	討論長處 及弱點	單單列出長處 或只討論弱點	沒有提及長處 及弱點
	(4)	(3)	(2)	(1)

聯繫	將發明的特徵及目的 與頗多其他相關現象 、問題連結起來	將特徵及目的 與一或兩個 相關現象、問題連結	與其他相關現象 問題的連結並 不清晰或不洽當	沒有與其他 相關現象問題 連結
	(4)	(3)	(2)	(1)

註：模仿 Goodrich(1998), 括號內為每項得分，整份作業最高分為 20。