

大學生專題報告Rubrics之發展與成效評估： 以人力資源管理相關課程為例

柳玉清*

摘要

研究者採用 Rubrics，分別發展人力資源管理課程的書面報告大綱和口頭專題報告之評分規準。研究參與對象為研究者任教課程之51位大學生，分成十組。結果發現：在二位教師之間、以及教師與學生之間的評分者間信度頗高，而且評分標準越具體、可量化，評分的信度相對提高。雖然以 Rubrics 評分的相關係數略高於傳統評分方式，只是此差異極為微小。問卷調查顯示：學生對於 Rubrics 評量方法傾向認同與肯定，學生也願意接受同學之間的互評。本研究整理研究發現以及學生意見，提出對於未來研究與使用 Rubrics 的具體建議。

關鍵詞：Rubrics、評分規準、學習成果評量、實作評量、信度



DOI : 10.3966/199679772016063301003

投稿日期：2015年11月2日，2016年1月21日修改完畢，2016年5月16日通過採用

*柳玉清，玄奘大學應用心理學系助理教授，E-mail: yuching@hcu.edu.tw

壹、緒論

臺灣近年來由於高等教育的快速擴張，廣設大學院校，大學錄取率已經超過九成以上，使得原為菁英教育的高等教育轉變為大眾化教育。雖然學子進大學不再需要擠窄門，且國民的教育程度普遍提高，但也帶來大學生素質低落的隱憂；加上畢業生被企業界詬病就業力不足，造成高學歷卻高失業率的現象（曾慧青，2009），社會大眾開始關注學生在校學習成果，以及大學文憑所代表之實質意義與價值。

環顧歐美先進國家，美國早在1970年代已產生類似問題，許多學者建議實施學習成果評量，以進行品質管控（彭森明，2010）。歐洲各國也明顯感受到學生成就不如預期，因而提倡以學生學習成效為基礎的教育，經濟合作暨發展組織（Organization for Economic Cooperation and Development, OECD）更倡議與發展「高等教育學習成效評量」（池俊吉，2011）。因此，重視學生學習成效與評估，實為目前國際高等教育評鑑的主要發展趨勢之一（蘇錦麗、黃曙東與浮絲曼，2011）。我國教育改革方向，也由重視資源投入轉為學生學習成效，這可從教育部針對大學院校進行之第二週期評鑑重點——強調學生實質學習成果的評量——得知強化教學與做好學生學習成效評量，是教學品保的必要措施。

傳統的課堂學習評量，多以紙筆測驗為主，主要評量認知能力與記憶成果，其目的是為學生打分數或評等第，以檢定學生是否通過該科之學習（彭森明，2010）。然而，現今評量的功能，不只是判斷學生是否達到教學目標，還需要兼具教學功能，協助教師改進教學，進而促進學生達到學習目標（趙建豐與林志隆，2005）。尤其某幾類的學習成果，如：溝通技能、心理動作技能、問題解決能力以及情意特質等，強調實際的行為表現，比較無法藉由客觀式紙筆測驗作正確測量，而是需要教師根據學生的表現過程或最後所完成作品的品質來評量，這類評量方式稱之為「實作評量」（余民寧，2011）。實作評量包含實際操作、口頭報告、科學實驗、數學解題、寫作、作品展示……等（余民寧，2011；李坤崇，2006；吳清山與林天祐，1997），其目的是希望了解學生能否將學習結果應用於真實情境的表現，又因為教師提供一模擬真實的測驗情境，以此得以推知學生在真實情境下的實際表現技能為何，故實作評

量又有「真實評量」之稱，可以提供更豐富的訊息，且較精確的評量學生學習成果（吳清山，2012；Montgomery, 2002）。

學生學習成效反映在畢業生的就業力，為促使高等教育提供符合產業升級發展所需質量之人才，教育部青年發展署（改制前為青輔會）於2006年進行大專畢業生就業力調查結果發現，雇主認為最重要的八項核心就業力技能，包括：一、良好工作態度；二、穩定度與抗壓性；三、表達與溝通能力；四、專業知識與技術；五、學習意願與可塑性；六、團隊合作能力；七、基礎電腦應用技能；八、發掘及解決問題能力。除了專業知識外，這些企業需求的核心技能，往往不適宜採用紙筆測驗來衡量（Riebe & Jackson, 2014）。因此，為提升學生學習成果的評量成效，應採用多元方式——包含紙筆測驗以及實作評量，有系統地評量學生多元學習成果，以檢視學生整體素質與能力。

基於前述所提，表達與溝通能力是企業重視的核心能力之一，高等教育應加強培訓。目前大學課程中常見的分組口頭報告，確實可以提供學生共同合作、彼此溝通以及利用口語展現成果的機會。因為口頭報告是依據學生表現評分，屬於實作評量，而評分規準（Rubrics）則是近期美國教育界推廣的實作評量方式。

研究者教授人力資源管理相關課程，為增進學生蒐集、組織、統整資料，以及溝通、表達之能力，規定學生分組，以口頭搭配PPT（PowerPoint）進行專題報告，作為本課程多元評量方式之一。幾年下來，發現學生常掌握不到老師要求的重點，而且傳統給一個總分的方式，造成學生只在意成績及格與否或分數高低。研究者以自身實務工作上遭遇的問題，經由參加教師研習以及閱讀文獻，了解Rubrics為一多元化之實作評量方法，為了讓學生更了解教師評分重點，且藉由參與評量的過程，能更投入作業的準備，達成教師期待的表現水準，因而嘗試發展適用於研究者任教課程的專題報告評分規準，以行動研究方式，進行教學評量之改進。

本研究目的有四：一、因應課程需求發展專題報告適用之評分規準；二、進行教師之間以及教師與學生之評分者間的信度驗證；三、探究 Rubrics 評量方法是否較傳統打整體印象分數的方式更增進評分者間信度；四、利用問卷蒐集學生對使用 Rubrics 評分方式的態度與意見。期望藉由研究結果，發展出更精確的評分規準，運用多元方法對學生的

學習成果進行評量，而所得評量結果可回饋學生，協助其改進學習不足之處，同時也能提供教師教學改善之參考，強化以學生學習成果導向的多元評量。

貳、文獻探討

一、Rubrics的源起與發展

Rubrics，應用於教育評量領域的中文譯名為「評分規準」（李坤崇，2006；趙建豐與林志隆，2005），或「評分量尺」（蘇錦麗等人，2011），它代表一組清楚界定的、教師對於學生表現期望的標準，可以做為指導學生為評量作準備的規範，也是學生學習品質評定的準則（Andrade, 2005）。

近代在1962年，首次有論文提到將 Rubrics 應用於高等教育情境中（蘇錦麗等人，2011）；1978年，紐約州教育部門使用 Rubrics 作為評量學童繪畫發展程度的一種評分工具；1979年則用在寫作能力的評量（Selke, 2013），之後逐漸將Rubrics應用於幼稚園至高中學生的評量。21世紀開始，美國高等教育認可機構以及各大學院校教師，紛紛運用 Rubrics 於學生學習檔案評量，甚至包括整個學習課程或方案的評鑑。例如：美國大學與學院協會（Association of American Colleges & Universities, AAC&U）於2007至2009年間，邀請超過100個高等教育機構進行的VALUE（Valid Assessment of Learning in Undergraduate Education）方案（AAC&U, 2010），共發展了16項重要學習成果評量 Rubrics，可見 Rubrics 已成為美國高等教育評鑑學生學習成效之重要工具。在臺灣重視學生學習成果評鑑之今日，也開始推廣 Rubrics 評分方法。

二、Rubrics的要素與類型

Rubrics 有三個基本的組成要素：其一是評量的準則（criteria），也就是評分者判斷學習成果優劣的標準或指標；其二為品質的定義，以具體且詳細的敘述，定義出學生須達成的技能或知識程度之內涵，以此

來區分學生實作表現的優或劣；其三是計分的方式，根據評量結果轉換成等第或分數（Reddy & Andrade, 2010）。Rubrics 通常製成評分表格形式：最左欄是評量的指標項目（即評量的準則），最上方由左到右列出品質由優至劣的等第（相當於計分），而中間交集的各方格則是不同學習表現程度的陳述（即品質的定義，參見表1及表2的形式）。

Rubrics 可分成二大類型：整體型（holistic rubrics）與分析型（analytic rubrics）。整體型的評分規準是將學生學習過程或作品視為一個整體來作評量，而分析型則是將學習成果分為幾個部分，逐一根據各個標準加以評分，然後再將評量的結果加總計分。整體型的評量較省時，比較適用於創新且沒有正確答案的學習成果，通常做為總結性評量。至於分析型的評量方式，則是學生的學習成果可以區分成幾個特定、可觀察的標準或指標，雖然評分時較花時間，但是能產生多向度的評分結果，反映出學生學習成效的輪廓，提供學生了解本身學習的優點與不足之處。如果要進行形成性評鑑，最好採用分析型評量方式（Mertler, 2001）。

三、Rubrics對教師教學與學生學習的幫助

傳統作業的批改是教師依據其個人標準給分，教師或許有評語，但是學生要自行去解讀與理解，而且教師的評語通常是在學生交了作業之後才得到的，學生無法事先獲知教師評鑑的準則。而 Rubrics 是事先訂定的評量標準，代表教師期望學生達到的成果，教師依據標準進行評量，相較於傳統個人主觀的判斷，評分上相對容易達到一致性與公平性。如果教師能在事前清楚說明評分標準，甚至讓學生參與制定標準，Rubrics 便成為指導原則，學生在做作業之前，即可針對要求的標準或是評量的重點去努力，以達到教師期望的目標。此外，在完成作業的過程中，學生也可以先行自我評量或同儕互評，以發現不足之處並盡早加強改進，如此亦可達到形成性評鑑目的，形成持續修正與進步的循環（史美瑤，2012；Andrade, 2007; Jackson & Larkin, 2002; Mertler, 2001; Moskal, 2000）。

美國相關研究發現，在大學或研究所學生對於 Rubrics 的評價大多是正向的（Reddy & Andrade, 2010）。除上述優點，有學生反應 Rubrics

提高了學習參與，導引他們產出更佳的學習成果，得到更好的成績，甚至降低對做作業的焦慮感。

對於教師而言，謹慎發展的 Rubrics 可以幫助教師澄清與定義精熟/優異的標準，也有助於教師在課程設計時強調這些標準（也是教學目標的所在）。通常教師認為使用 Rubrics 讓評分更具一致性與客觀性（信度提升），然而有些教師卻反應使用意願不高，或許是不了解 Rubrics 除了評分，亦可增進教學成效（Reddy & Andrade, 2010）。因此，雖然發展評量準則需花相當時間，然而一旦建立 Rubrics, 就可以對所有學生重複使用（Jackson & Larkin, 2002; Montgomery, 2002; Yoshina & Harada, 2007）。不過，研究也顯示：僅僅採用 Rubrics 評分並不足以提升學習的成效，最好是讓學生參與評分規準的制定，同時採用形成性評量策略，要求學生在學習過程中以此自評或與同儕共評（Reddy & Andrade, 2010）。

四、Rubrics的信效度

Rubrics 作為一套評量工具，其本身的信效度值得驗證。大多數研究以探討信度為多，效度則相對較少（Reddy & Andrade, 2010）。信度以評分者信度為主要關注焦點，包含不同評分者評量的一致性（inter-rater reliability），以及同一位評分者對同一作業評分的一致性（intra-rater reliability）。設計良好的 Rubrics 應該減少評分的不一致，然而，要達到此目標，評分規準的敘述要清楚具體，還需要加強評分者的訓練，提供其評分結果精準程度的回饋。

研究指出：運用評分規準，教師和學生評量結果的一致性頗高（Dunbar, Brooks, & Kubicka-Miller, 2006）。有關效度（如內容效度、建構效度、效標效度）的研究非常少，有研究顯示：評量標準的各水準的文字描述若清晰且適當，可增進評量的精準度，亦即提升評分規準的效度（Green & Bowser, 2006; Lapsley & Moody, 2007; Moni, Beswick, & Moni, 2005）。

研究結果也發現：（一）評分者訓練能讓評分者在評分時更有信心，訓練的成效在於降低評分者給分的隨機誤差（random error）；（二）評分者訓練能降低評分者整體嚴格度的變異，特別是能降低非常

極端嚴格或寬鬆的評分傾向，但是明顯的評分者差異依然存在（藍珮君，2012）。

參、研究方法

本研究起因於教師為改善評量所進行的行動研究，主要目的之一為發展符合課程需要之學生學習成果評量Rubrics，並驗證其評分者信度；二為瞭解學生對於使用Rubrics的回饋意見，以做為改進評分規準之參考。採用的研究工具、研究對象、評分方式說明如下。

一、研究工具

（一）Rubrics評分規準

本研究所發展的 Rubrics 有二種，用以評量學生所做的期中書面報告大綱與期末口頭專題報告，其同時也是研究工具，用以驗證評分者信度。

1. 期中書面報告大綱Rubrics

研究者為督促學生按部就班進行專題報告作業，特別要求學生於第九週期中考時，先交期末專題報告的報告大綱。所謂報告大綱，即是預定於期末進行專題報告的內容摘要（包含題目、想學習的重點、報告內容之綱要以及參考資料來源）。

有關報告 Rubrics 的撰寫，係研究者參考國內外有關 Rubrics 發展的原則、步驟與範例（如李紋霞，無日期；Halonen et al., 2003; Kerby & Romine, 2009; Moni et al., 2005; Skillings & Ferrell, 2000），加上研究者身為教師的要求（例如：聚焦人力資源管理特定主題，深入探討、呈現具體的實際案例、制度、做法或是研究），以及對學生學習成果的期望（能清楚條列報告重點、內容豐富、且能依照APA格式列出至少三項參考資料來源等），作為評分的標準。報告大綱評量標準有三：(1) 主題聚焦且有實際案例、(2) 大綱清楚條列與內容豐富程度、(3) 參考資料來源數目與符合APA格式程度。以上三項準則，依據表現程度給予「優」（90-100分）、「良」（75-89分）、「普」（60-74分）、「待改進」（0-59分）等不同程度的分數。報告大綱Rubrics初稿交給三位專家審

閱（二位精通 Rubrics 之教育學者，一位曾任教研究者學校並教授人力資源課程之A老師），三位學者專家均回饋寶貴意見，但修正的幅度不大。

此外，為讓學生了解 Rubrics 並且參與評量標準之制定，任課教師曾在期中考前，在課堂上解說 Rubrics 的定義、用途、形式、發展步驟與相關範例，以及報告大綱 Rubrics，說明老師使用此評分規準的目的與做法。研究者逐項說明報告大綱評量的標準後，請學生檢視，是否有修正意見。有學生提出對於第二部分「內容」評分標準，希望將不同等級字數的要求稍降一些。因此，經綜合三位專家與學生意見後，修正後的版本如表1，此標準即作為教師與學生共評的評分規準。

2. 期末口頭專題報告Rubrics

研究者歸納前述所參考的相關範例，認為口頭專題報告評分重點有三：(1) 主題與內容、(2) 表達方式、(3) PPT的製作。其中，「主題與內容」下分四項：(1) 聚焦人力資源管理特定主題，深入探討；(2) 呈現具體的實際案例、制度、做法或是研究；(3) 組織架構是否完整且具連貫性和邏輯性；(4) 提出分析、比較、評論、與建議。「表達方式」下亦分四項：(1) 口語說明是否清楚有條理、掌握重點、易於理解；(2) 音量與語調；(3) 目光接觸，適當運用手勢；(4) 時間掌控。最後，「PPT的製作」評量有三項：(1) 字體、排版、色彩等視覺效果；(2) 適當圖表或影片；(3) 錯字多寡。以上11項準則，依據表現程度給予「卓越」、「良好」、「尚可」、或「待改進」的等級。而其相對應的分數為（9-10分）、（7-8分）、（6分）、（1-5分）。有關分數修訂為1~10分，是因為一位審查專家以及期中使用報告大綱評量後學生反應，若以1~100分評分會較複雜，建議改為1~10分。

口頭報告 Rubrics 初稿同樣交給前述三位專家審閱，經綜合三位專家意見，且在課堂上（第十週與十三週－即報告的前一週）與學生討論後，修正後的版本如表2。

（二）Rubrics態度問卷

為了瞭解學生對於使用 Rubrics 的看法，研究者針對文獻所提 Rubrics 的特點，初步編擬包含封閉式以及開放式問題之間卷。封閉式問題採Likert氏五點量表，涵蓋的面向有：公平客觀、參與感、學習回

饋（知道評量重點、本身優缺點以及改進之處）、使用 Rubrics 偏好程度（包含教師與同儕互評）。問卷初稿亦先交給前述二位精通 Rubrics 的教育學者審核，經修正部分文字與刪除某些問項後，二次問卷共同題目有八題（參見表9），期末口頭報告再增加二題：「對同學報告進行評分，讓我更真實的體驗擔任評量者的角色」、「進行小組互評時，我會仔細評估、比較後，才給分」。

開放式問題在期中報告大綱詢問使用的心得與建議等，期末的問卷則另外增加指出哪些項目較容易評分、那些較難等二項問題。問卷於期中、期末考時發給學生填寫，用以蒐集學生的意見。封閉式問題得到的量化資訊進行平均數、標準差等描述統計分析，以了解學生平均之態度；開放式問題得到的質性資料，能更進一步了解學生真實的心得感受，先將學生反應編碼，再經歸納分類後，整理出學生主要的看法與意見。

二、參與研究對象

參與本研究之學生來自研究者任教的北部一所私立大學，大多數是社會科學學院大三與大四學生，分成十組，每組4~6人，男女比率是2比3。參與期中書面報告大綱評量的學生共51人，而參與期末口頭報告評分人數，則因學生每週出席、遲到早退狀況不同而有差異，扣掉報告組別成員，實際評分學生人數在32~39人之間。期中與期末填寫問卷人數均為51人。

參與評分之教師有研究者（任課教師）與前述擔任審查之A老師。研究者執行學生書面報告大綱和口頭報告評分工作，而A教師僅參與期中書面報告大綱之評分。

三、評分方式

（一）期中書面報告大綱

無論是教師或學生，均要求閱讀完一份大綱後，先依循傳統改作業方式，給予一個整體分數，之後再依據 Rubrics 評分表格（如表1），每組逐項評分。研究者蒐集所有資料後，再將各項 Rubrics 評分的分數

（即「主題」、「內容」、「資料」等三項），依據原先設定的權重（分別佔30%、50%、20%），計算加總後成為該組作業的分數（此即表3、表4的「報告大綱_Rubrics總分」欄）。

（二）期末口頭專題報告

每組報告時間要求以25分鐘為原則，當台上一組學生報告完畢後，任課老師以及學生（非報告組學生），先依照傳統方式先給予一個整體分數，之後再依據 Rubrics 評分表格（如表2，每人一張），在該組相對應的欄內逐項評分，每組評分時間約5分鐘。共有十組學生報告，分三週進行。

最後研究者蒐集所有評分表，再將各項評分的分數（即主題與內容、表達方式、PPT的製作等三項），依據原先設定的權重（分別佔40%、30%、30%），計算加總後乘以10，成為該組報告的分數。

在評分過程，教師與學生評分之分數作為檢驗 Rubrics 信度之量化數據。

肆、結果與討論

一、人力資源管理報告Rubrics的制定

研究者根據身為教師對於課程的要求，初步擬定書面報告大綱與口頭專題報告的評分 Rubrics，交給三位學者專家審查，研究者也採納其意見加以修改，之後再經過與學生討論，完成此二份 Rubrics 評分規準，如表1與表2。因為研究者參考國內外相關的 Rubrics 範例，加上三位學者專家的審查，以及學生的意見，此二份 Rubrics 堪稱具有一定程度之內容效度。

表1

報告大綱Rubrics評分規準

指標表現 評量標準	優 (90~100分)	良 (75~89分)	普 (60~74分)	待改進 (0~59分)
一、主題與想 學習的重點： 相關且聚焦與 有實際案例 (30%)	與本課程有關 且聚焦在一特 定主題，探討 實際機構的制 度、案例、實 務做法，或是 統整研究論文 做報告。	與本課程有關 且有聚焦相關 重點，有案例 或整理研究論 文。	與本課程有關， 但主題缺乏聚 焦，或好幾個 主題但關聯性 不大，未能深 入探討。	與本課程不 太相關，或 範圍太大無 重點，沒有 實際機構的 制度、案例、 實務做法， 或是沒有整 理研究論文 做報告。
二、內容： 清楚條列與豐 富程度 (50%)	有組織有邏輯， 清楚條列式列 出報告的重點， 內容豐富詳 盡。(至少 250字以上)	有邏輯的條列 出報告重點， (200~250 字左右)	條列出報告重 點。(100~ 200字)	內容很少， 沒有清楚條 列出來，敘 述沒有組織 或無邏輯， 或是字數100 字以下。
三、參考資料 來源： 數目與符合 APA格式程度 (20%)	依據APA格式， 清楚條列出參 考資料來源， 至少三項。	與APA格式不 一致，但有清 楚條列出參考 資料來源，至 少三項；或依 據APA格式， 清楚條列出參 考資料來源， 至少二項。	依據APA格式 條列出參考資 料來源，至少 一項；或沒有 依據APA格式， 但列出二項。	沒有列出參 考資料來源。

表2

口頭專題報告Rubrics評分規準

標準	評分	卓越 (9-10分)	良好 (7-8分)	尚可 (6分)	待改進 (1-5分)
壹、 主 題 與 內 容 40%	1-1	聚焦人資特定主題，深入探討。	主題與人資相關，但沒有聚焦並深入探討。	主題與人資雖有些相關但是較廣泛，沒有聚焦，不夠深入。	主題與人資不相關。
	1-2	完整且詳盡的呈現具體的實際案例、制度、做法或是研究，內容豐富。	有呈現實際案例、制度、做法或是研究，但內容不夠深入與詳盡。	僅呈現一個案例、制度、做法或是研究，且內容是研究，且內容較少。	資料內容非常貧乏或是沒有案例、制度、做法或研究。
	1-3	報告內容有完整的組織架構，且具有連貫性和邏輯性。	內容架構堪稱俱全，整體連貫性與邏輯性尚可。	內容架構尚可，但整體連貫性與邏輯性較差。	內容架構不完整，且缺乏連貫性與邏輯性。
	1-4	有深入與精闢的分析、比較、評論與建議。	有進行分析、比較、評論或建議，但不夠深入與精闢。	有進行初步的分析、比較、評論或建議。	沒有提出分析、比較、評論或建議。
貳、 表 達 方 式 30%	2-1	口語說明清楚有條理、能掌握重點，讓人易於理解	口語說明有分段落，能讓聽眾理解。	口語說明少部分讓人抓不到重點，或不太理解。	說明較無條理，無法表達意思，聽眾較難理解或不知重點。
	2-2	音量、語調清楚、有抑揚頓挫。	音量、語調清楚，偶而照稿念，卻不至於沉悶。	音量、語調尚可，一半時間照稿念，有些沉悶。	幾乎照稿念，聲音太小或語調單調。

(續下頁)

參、 PPT 製 作	2-3 持續和聽眾保持目光接觸，並經常在適當時機運用適當手勢促進溝通。	和聽眾常有目光接觸，偶而在適當時機運用手勢輔助說明。	有時會與聽眾目光接觸，或曾經有一兩次以手勢輔助說明。	幾乎沒有目光接觸，肢體僵硬或姿勢不恰當。
	2-4 時間掌控在25~27分鐘	時間在27~30分鐘或23~25分鐘左右	時間在30~35分鐘或20~23分鐘左右	時間超過35分鐘或少於20分鐘
	3-1 字體清晰、讓人一目了然、排版色彩適當，視覺愉悅	字體與排版普通，沒有明顯疏失。	字體與排版尚可，但有些畫面字體太小、畫面文字太多、色彩不清晰，畫面被切掉等疏失。	字體太小或太多、沒有分段或某些字在畫面外、被遮住或難以辨識，色彩不清晰分明…等明顯疏失。
	3-2 有適當圖表/影片，有助理解、增加吸引力。	有圖表/影片，但效果普通或圖表/影片太多。	有圖表/影片，但是有些不相關。	沒有圖表/影片或圖表/影片不適當，無助於理解或不吸引人。
	3-3 沒有錯字。	有1~2個錯字。	有3~5個錯字。	有6個以上錯字。

二、Rubrics之評分者信度驗證

(一) 期中書面報告大綱之評分者間信度

1. 教師之間評分的相關係數與一致性百分比

期中報告大綱的評分，任課教師與A教師在十組總分的相關為.884，可謂高相關，而各個分項目（主題、內容、資料）的相關也很高，分別為.82、.80、與.97（見表3）。

若依據二位教師評分，將表現歸類於「優」（90-100分）、「良」（75-89分）、「普」（60-74分）、「待改進」（0-59分）等四等級，二位教師在每組等級一致性的百分比為60%；各分項評分的等級一致性

百分比分別為：「主題」50%、「內容」60%、「資料」70%，可見二位教師評等的一致性並非很高，尤其是對於報告的主題和內容部分，可能因為A教師雖參與審查修訂Rubrics過程，但事前未與研究者共同練習，故評分者之間的信度未能提升。然而在資料部分評分一致性較佳，或許是因為此指標可數量化，較具體之故（見表3）。

表3

二位教師在期中書面報告大綱評分者間信度

	報告大綱 _主題	報告大綱 _內容	報告大綱 _資料	報告大綱 _Rubrics總分
Pearson相關係數	.82	.80	.97	.88
評分等級一致性百分比	50%	60%	70%	60%

2. 學生與教師評分的相關程度

學生與教師評分分數的相關，係將兩位教師在十組報告大綱評分的平均數，與學生評分的十組平均分數（去除學生自評分數，因為統計分析發現學生自評分數有偏高傾向），求Pearson積差相關，得到 .96。此外，三個分項指標的相關分別是：「主題」.97、「內容」.93、「資料」.97，以上係數均顯示師生之間評分呈現高相關現象（如表4），可見師生評分的差異性不大，而且指標較具體的「主題」、「資料」項目，評分一致性較高。

表4

教師與學生在報告大綱評分之相關係數

	報告大綱 _主題	報告大綱 _內容	報告大綱 _資料	報告大綱 _Rubrics總分
教師與學生Pearson相關係數	.97	.93	.97	.96

註：教師評分人數為2人，學生為45~47人（學生共51人，扣除自評人數每組4~6人，故人數不一）

（二）期末口頭報告評分之信度檢驗

依據口頭報告大綱 Rubrics，任課教師與學生在11個指標評分（給分在1~10分的範圍，共10組報告）的Pearson積差相關數值為 .81，呈現高

相關，11細項的個別相關係數介於 .53~.96之間（詳見表5），最高的是時間掌控項目（因為有具體測量每組報告時間），較低的是對於報告內容的組織架構、連貫性和邏輯性的評量，表達是否有條理、掌握重點、易於理解，以及字體排版與錯字多寡等項目，這些項目可能較抽象，故一致性偏低；錯字數目雖然是可以數量化，但是可能是學生對於錯別字的偵查能力較低，因此和任課教師評分差異較大。

若是統計學生與教師在每一指標的評分完全相同的一致性百分比（即：內容、表達、PPT製作等三大項之下11小項的分數），由表6結果可見：學生與教師評分完全一致的比率不高，大約二至三成左右（16.5%~38.3%）；若將分數歸類為卓越、良好、尚可、待改進等四等級，則等級一致性比率之平均值大約是五成（21.2%~77.9%）；如果計算教師與全體學生評分分數相差在 ± 1 分之內的比率，則提高至近七成的比率（57.8%~80.8%），亦即任課教師與學生的評分分數70%差異在 ± 1 分之內，二者評分一致性尚可接受。從各分項目來看，主題與人資有關（內容_1）、目光接觸和運用適當手勢（表達_3）、時間掌控（表達_4）、以及錯別字字數（PPT_3）等較具體的項目，評分一致性之百分比比較高（詳見表6）。且整體而言，學生和教師對於報告內容評分一致性較口語表達和PPT製作為低。

由以上二種作業評分一致性信度結果可知，運用 Rubrics 評分標準，無論教師之間，或是教師和學生之間的評分一致性頗高，亦與國外研究結果相呼應，尤其是具體可量化指標，可見利用 Rubrics 標準來評量學生學習成果具有相當高的一致性（Dunbar et al., 2006），本研究所發展的 Rubrics 評分規準具有良好之評分者信度。

表5

任課教師與學生在口頭報告評分之相關係數

評分項目	相關係數
內容_1	.76
內容_2	.78
內容_3	.55
內容_4	.72
表達_1	.63
表達_2	.76
表達_3	.76
表達_4	.96
PPT_1	.53
PPT_2	.79
PPT_3	.54
11項指標	.81

表6

任課教師與學生在口頭報告評分一致性百分比率

評分項目	分數完全一致 之百分比	同等級一致性 之百分比	分數相差±1 之百分比
內容_1	30.7	51.5	72.7
內容_2	22.6	37.3	62.3
內容_3	25.1	44.2	69.1
內容_4	16.5	21.2	57.8
內容項目平均	23.7	38.6	65.5
表達_1	27.2	49.5	68.4
表達_2	25.9	43.5	65.9
表達_3	32.4	47.5	73.1
表達_4	29.5	73.2	74.6
表達項目平均	28.8	53.4	70.5
PPT_1	22.8	42.6	65.1
PPT_2	24.5	35.0	60.8
PPT_3	38.3	77.9	80.8
PPT項目平均	28.6	51.8	68.9

三、Rubrics評分與傳統評分之間的相關

（一）期中書面報告大綱評分

二位教師以 Rubrics 評分的相關是 .88，以傳統打分數的相關是 .85，而二位教師與學生以 Rubrics 評分的相關是 .96，以傳統評分的相關為 .95（見表7），雖然根據 Rubrics 標準，似乎比傳統評分得到的評分者間相關係數稍高，然而對於提升評分者間信度的增幅極為微小。

另一方面，計算教師自身以 Rubrics 和傳統評分所得分數之間的相關，發現任課老師以二種評分方式所給分數的相關為 .98，而A老師的相關為1.0（四捨五入結果），顯然教師本身所評分數的相關（評分者內評分相關），比教師之間依據 Rubrics 評分的相關（評分者間評分相關）.88還高出許多，換言之，評分者內的評分一致性比採用 Rubrics 評分標準所得的評分者之間的信度還高。

表7

教師與學生以Rubrics和傳統方式評量報告大綱之相關係數

	Rubrics評分	傳統評分
二位教師之間	.88	.85
教師與學生之間	.96	.95

（二）期末口頭報告評分

任課教師與學生在期末口頭報告所評分數，以 Rubrics 加權計分得到的總分相關為 .89，依據傳統方式所評分數的Pearson積差相關為 .85，二者均呈現高相關（詳見表8），雖然以 Rubrics 評分所得到的相關較高，但此差異卻極微小（僅0.04），似乎使用 Rubrics 標準並未顯著增進評分者之間評分的一致性。

另一方面，計算任課教師本身以傳統評分和以 Rubrics 評分的相關為 .91，學生以此二種評分所得分數的相關為 .99，由表8相關係數來看，以 Rubrics 評分規準得到的評分一致性（.89）較傳統評分略高（.85），但是評分者內的相關（如任課教師本身所評 Rubrics 分數與傳統評分相關為 .91、學生二種評分的平均分數相關為 .99），仍是略高於評分者之間以 Rubrics 評分的相關（.89）。

表8

教師與學生在口頭報告以Rubrics及傳統方式評分分數之相關

	傳統評分	Rubrics評分
教師與學生評分分數相關	.85	.89

由於 Rubrics 事先具體且詳細的定義出學生須達成的技能或知識程度之內涵，教師依據嚴謹的標準進行評量，相較於傳統個人主觀的判斷，理論上在評分時相對容易達到一致性與公平性，據此推論，評分者之間以 Rubrics 所得分數間的一致性或相關性會大於傳統評分之間的一致性或相關性。然而，綜合期中報告大綱以及期末口頭報告二次評分結果發現：評分者本身 Rubrics 分數與傳統評分分數的相關，卻有高於不同評分者之間以 Rubrics 評分相關的現象，也就是評分者內的分數一致性高於評分者間分數一致性。此現象是否導因於評分者整體印象對於分數的評定佔有很大的影響力（例如：月暈效果的評量偏差或是評分者評分的傾向），以至於評分者本身分數的相關（intra-rater reliability）高過於不同評分者之間分數（inter-rater reliability）的一致性？這是 Rubrics 相關文獻未見探討的，本研究發現此現象，值得進一步評估與檢驗。

四、學生對使用Rubrics的意見

（一）封閉式問題

本研究以問卷來蒐集學生使用 Rubrics 的態度與意見，封閉型問卷係Likert氏五點量表，數值介於1~5，3為中間值，4為同意、5為非常同意。統計期中與期末學生對於使用 Rubrics 評分的問卷結果（八題共同題目見表9），每一題目的平均數均超過3.0，故整體而言學生對使用 Rubrics 持正向的態度，也就是較多的學生認為 Rubrics 比傳統打總分方式更客觀公正，是有效、多元的評量方法，學生樂意教師採用 Rubrics 評分方式，也願意同學之間彼此互評。獲得學生特別肯定的是：預先設定的評分標準能讓學生知道評量的重點是甚麼、可以讓學生對課程設計更有參與感、減少因為老師主觀造成的評分偏見、評分結果可以知道自己報告的優缺點是什麼、以及清楚知道未來應該加強或改進之處。本研究發現學生對於使用 Rubrics 的看法也與前述文獻結果一致，可見

Rubrics 不但適用於實作評量，也有提升學生學習動機與促進學習成效之功用。

此外，由於甄選與評估是人力資源管理的重要工作任務之一，任課教師利用Rubrics 評分方式讓學生體驗擔任評量者角色，在期末問卷中增加二題詢問：「對同學報告進行評分，讓我更真實的體驗擔任評量者的角色」，學生平均反應分數（五點量表）為4.0、標準差 .57；而「進行小組互評時，我會仔細評估、比較後，才給分」平均數4.1、標準差 .57，由此可以得知，大多數學生是認真地扮演公正的評量者角色。

綜合二次問卷共同題目，進行配對樣本 t 檢定， p 值結果如表9最右欄，可見學生在期末第二次使用 Rubrics 的正向感受較期中第一次使用有顯著增加情形，也就是更認同 Rubrics 比傳統打總分方式客觀公正、能幫助學生知道評量的重點、希望老師多利用 Rubrics 評分、更喜歡同學互評。此結果與文獻中一般大學生對 Rubrics 持正向態度且接受同儕互評的結果具有相似性。

表9
學生對於使用Rubric評量方式的態度

題目	期中報告大綱		期末口頭報告		p 值
	平均數	標準差	平均數	標準差	
1. Rubrics評分標準比傳統打總分的方式更客觀。	3.41	.638	3.84	.681	.000
2. Rubrics評分標準比傳統打總分的方式更公平。	3.35	.688	3.72	.607	.002
3. Rubrics評分標準讓我知道評量的重點是什麼。	3.96	.799	4.32	.683	.014
4. 依據Rubrics評分標準所得到的結果，讓我知道自己報告的優缺點是什麼。	3.88	.864	4.02	.685	.182
5. 依據Rubrics評分標準所得到的結果，讓我知道未來應該加強或改進之處。	3.92	.688	4.00	.670	.322
6. 老師和學生一起討論學習成果的評量方式與標準，可以讓學生更有參與感。	4.04	.848	4.10	.814	.679
7. 我希望老師多利用Rubrics方法來作為評分的標準。	3.39	.723	3.64	.749	.018
8. 我喜歡同學互評的評分方式。	3.27	.802	3.34	.717	.010

（二）開放式問題

開放型問題主要詢問學生使用 Rubrics 的心得感受以及建議。因為期末口頭報告評分標準有11項，研究者想了解同學評分難易處為何，因此期末問卷加入詢問那些評分項目較容易以及較困難、為什麼等二個問題。學生填答的質性資料，採取歸納與分類方式，綜合整理結果，敘述如下，標楷體字為學生回答之原文，括弧內英文數字為學生代碼。

1. 期中報告大綱

（1）使用 Rubrics 心得

多數學生認為這樣的評分方式有標準較客觀（51位學生中有22人明確表達），例如：

我覺得它清楚的界定評分標準，打起分數來會有一個標準，也許能避免評分時的誤差（S112）。

我覺得Rubric的評分標準可以給評分的人一個中立的評準，就不會太容易產生主觀，因為大家都可以用這個指標做為參考，就可以變得比較公平（S126）。

其次是有15人表示同學互評可以看見其他同學的報告，因此更能了解自己和他人的優缺點，加上實際參與評分，更有參與感。例如：

我覺得這是一種與以往不同的評分方式，就學生也能參與評分而不再是只有老師評分，而且還可以瀏覽其他組別的報告內容，可以知道別組的報告方式，不再是井底之蛙，不知道他人的想法，也算是一種學習交流（S113）。

很特別的方式，讓我感受到實際參與了這堂課，透過評分可以了解到自己組別與其他組別的優缺點，對於我們的長處可以保持做得更好，對於短處能確實比較我們跟其他組的差異，做更好的改進與加強（S119）。

也有學生表達 Rubric 評量可以事前知道評分重點，有助於做作業的

準備方向，例如：

利用Rubric的評分可以讓在評分時更能了解到以後做報告時，題目給了什麼，那麼我們應該把重點放在哪裡，分數才會顯得比較高（S120）。

雖然多數學生感覺 Rubric 評分標準很客觀公平，但仍有部分學生覺得會有主觀因素影響，例如：

個人覺得這個方式很公平，但是很難評出一個標準的分數，因為每個人的標準不一樣（S101）；

這種評分方式雖然有著客觀標準，但仍離不開主觀看法與刻板印象，若與過去的傳統單一人憑印象相比已經客觀許多（S104）。

或許是如此，加上有些同學表達擔心自己經驗或專業不足，因此怕給分不夠客觀，例如：

第一次幫同學打分數，感受很不一樣，不確定自己有沒有資格（S144）；

幫別人評完分數以後，感覺自己的評分方式還有待加強（S108）；

原來評量的面向和重點有很多，需要不斷練習，增進評分的客觀標準化（S141）。

只有四位同學認為 Rubrics 評分方式比較耗時（S111、S124、S130、S150）。不喜歡的只有二位：很複雜，看不懂怎麼評分（S134，或許這位同學因為常缺課，老師在講解以及和同學一起討論未出席所致）；多餘，給同學評分這個舉動是沒必要的（S136）。

從以上反應得知學生進一步的感受，對於使用 Rubrics 是正向與肯定，但是第一次擔任評分者，比較擔心評分是否夠客觀公正。

(2) 學生對於 Rubrics 的建議

經歸納學生填寫資料，將具體建議條列如下：

- a. 最好提供不同得分的報告做為參考範例（S105、S118、S133）；
- b. 評分時可以和其他同學討論，會比較公正客觀（S130、S140）；
- c. 建議不要給自己的組別評分（S103）；
- d. 評分標準再多增加一些分項目，敘述更具體詳細（S122、S148）；
- e. 建議老師多花時間詳細說明評分標準（S151）；
- f. 評分標準不要由老師事先擬好，而是讓同學討論後，共同決定那些標準與分數範圍（S147）；
- g. 增加評語欄位，提供老師對於評分者是認真或流於主觀，提供分析的參考（S113、S126）；
- h. 增加情感因素分數，以降低同學之間情感造成的偏差分數（S109）；
- i. 建議分數以0~10給分，因為以0~100給分範圍太廣，要花較多時間斟酌分數（S113）。
- j. 建議以檢核表形式，條列出所要評量的項目，讓學生一項一項檢查，有符合就打勾，最後再加總計算分數（S112）。

從學生的反應中可以發現：不少學生對自己評分的信心不足，希望老師給予不同分數的範例做為參考，因此，如文獻所示，評分者需多訓練，才能在評分時更有信心（藍珮君，2012）。

2. 期末口頭報告

(1) 較容易理解的評分標準與評分項目

學生認為較容易評分的部分有：報告時間掌控（因為教師在每組報告完畢後，即將該組報告時間寫在白板上）、PPT字體與排版、適當圖表/影片、錯字數目、以及口語說明是否清楚有條理、掌握重點等項目。檢視這些標準都是較易感受到的口語表達或是具體圖像，尤其是可以數量化的標準，如時間與錯別字數目等。

(2) 較難理解與評分的項目

這部分的反應集中於主題與內容的評分標準。絕大多數學生認為「報告內容有完整的組織架構，且具有連貫性和邏輯性」、「內容是否有深入與精闢的分析、比較、評論、與建議」很難有一個具體的評判標準。學生表達因為是初學此課程，本身對於人力資源管理知識不足，故較無能力判斷報告是否聚焦人資特定主題、內容是否有深入與精闢的分析、比較、評論、與建議（S238、S240）；學生也反應因為許多組報告時沒有列出目錄大綱，加上每一組的主題不同，當同學報告了二十幾分鐘後，前面的內容也忘的差不多了（S214、S233），因此要確實分析報告是否有完整架構較困難。

由此二問題發現學生認為量化的、較具體的評分項目較容易掌握，但是對於抽象的組織架構、連貫性和邏輯性，以及有關分析、比較、評論等，因為受限自己本身還是在學新手，對專業知識不足，所以會沒有信心，也可能影響評分的客觀性。因此，如要讓學生參與評量，Rubrics的指標宜具體量化。

(3) 使用 Rubrics 心得感受

期末所蒐集有關心得資料，因為學生主要反應與期中報告大綱相似，為節省篇幅不一一陳述。不同之處是有學生表示上課會更專心、有參與感，甚至被尊重的感覺，例如：

會因為要評分，而去認真地聽同學報告，會覺得在課堂上更有參與感（S221）。

我覺得加深我的注意力，因為聽完後要評分，所以比一般會來得專心（S226）。

在我們這組報告時，他人也比較會專心聽講，好做評分，有被尊重的感覺（S214）。

同時，學生也體驗了擔任評分者的角色，而顯得更慎重：

這是很新鮮的體驗，同時身為評分者與被評者能更了解評分的角度，不是單一的，因為身歷其境才更感受到將心比心的心情（S238）。

讓我真實的體驗到了擔任評量者角色的重要性，客觀性，能夠有客觀的觀點來進行小組互評（S249）。

(4) 學生對於 Rubrics 建議

期末問卷蒐集到學生的具體建議較少，主要建議有以下幾點：

- a. 分再綜合老師的評分，以求得最終成績（S211、S248）。
- b. 公布他組評分結果，讓學生可以知道他人對自己的看法（S214）。
- c. 評分項目再細一點（S212）。
- d. 評分項目再簡化、簡單明白些（S232、S234、S235、S236）。
- e. 主觀的項目可以少一點（如美觀之類的）（S227）。
- f. 先利用影片教導我們評分時的方式與過程（S226）。
- g. 除了評分的項目之外，建議可以在後面自己附上具體的建議或是想法（S231）。
- h. 可在評分表上，寫下每組報告的主題，這樣報告到後面也較不會不清楚報告主題，而造成評分項目中的「壹、主題與內容」的評分不公（S233）。

經歷二次評分過程，學生對於 Rubrics 評分方式是肯定而正向（如前節所述）。而在開放性問題的回答中，大多數同學在意的是自己本身以及其他同學是否能真正公平客觀的評分，主要是認為人都是主觀的，不可能完全客觀。其實 Rubrics 評分也不可能避免個人主觀影響，只是與傳統評分方式相比，事先詳列評分標準，可以提升客觀性與公平性。這一點在日後教學中，應先和學生溝通，建立正確認知，否則，學生似乎期待的是一個絕對公平客觀的評量工具。

另一現象是學生對於自己評分能力沒有信心，希望老師給予範例，並詳加說明。因此，教師最好能提供不同分數的範例，做為樣本讓學生參考。

伍、結論與建議

一、綜合以上研究結果，所得結論如下：

（一）因應課程需要，已初步建立一套明確之評分規準

編擬之報告大綱與口頭報告 Rubrics，經過相關專家學者審閱以及學生的參與討論，已初步發展為一套明確的評分準則，具有內容效度。加上問卷結果顯示學生普遍認為 Rubrics 相對傳統評分方式更為公正客觀、以及清楚知道未來應該加強或改進之處，有助於提升學習成效。因此，本研究所使用之 Rubrics，確實為評估大學生在人力資源管理相關課程專題報告的有效評鑑工具。

（二）本研究發展之評分規準具備良好之信度

分析，發現報告大綱三項評分規準在二位教師之間、以及教師與學生之間的評分者間信度頗高，而專題報告的11項 Rubrics 在任課教師與學生之間的評分者間信度也超過0.8以上，顯示均有良好信度。且研究結果顯示，若評分標準越具體客觀（例如有量化之參考依據），評分的信度將更加提升，也會更公正客觀。

（三）Rubrics 的評分一致性略高於以傳統打整體印象分數的方式

本研究結果發現，無論教師之間，或教師與學生之間，以 Rubrics 評分的一致性略高於以傳統打整體印象分數的方式，只是此差異極為微小。

（四）學生普遍對於 Rubrics 評量方式持正向態度

1. 學生表示因為 Rubrics 有具體評分標準，將減少因為老師主觀造成的評分偏見，相對傳統評分方式更為公正客觀，因此樂意教師採用 Rubrics 評分，也喜歡同學之間彼此互評。
2. Rubrics 評分結果可以讓學生知道自己報告的優缺點是什麼，以及清楚知道未來應該加強或改進之處，有助於提升學習成效。
3. 老師和學生一起討論學習成果的評量方式與標準，可以讓學生事

先瞭解報告要求的重點與品質，學生也感到有參與感。

（五）可量化的標準較易評分，無法量化或較不具體的評分項目則感到困難

例如可觀察到的口語表達或是PPT的具體圖像，或是可量化的標準，如時間與錯別字數目等，較容易評分。但若是評定報告內容的組織架構、連貫性和邏輯性、是否有深入與精闢的分析、比較、評論、與建議等，則因為難有一個具體的指標，加上學生是初學者，專業知識不足，對學生而言是較困難的評分項目。

二、本研究根據課堂實際操作時學生反應以及問卷資料分析結果，針對未來研究以及使用 Rubrics，提出以下建議：

（一）建議日後邀請其他教師應用此規準評分，以驗證其信效度

本研究所蒐集之資料來自研究者任教班級，教師評分（期中書面報告大綱）之間的相關，僅有二位教師，而專題報告 Rubrics 則無其他教師參與評分，雖然目前所得數據呈現高相關，為求持續驗證此評分標準之信效度，日後可以在相關科目邀請其他教師的利用此規準評分，不斷修正與改進，累積更多經驗與研究成果，擴大使用範圍。

（二）評分者內的分數一致性高於評分者間一致性，值得進一步研究與檢驗

評分者本身以 Rubrics 評分的結果和以傳統方式所得分數的相關，有高於評分者之間以 Rubrics 評分相關的現象，也就是評分者內的分數一致性高於評分者間一致性。此現象是否導因月暈效果的評量偏差，或是評分者評分的傾向，值得進一步研究與檢驗。

（三）加入評分者與被評學生之熟悉程度或好感程度測量

雖然學生們接受同儕互評，惟有學生擔心與被評分同學的熟悉程度或好感度可能會影響評分，日後研究可以探究此變項是否影響同儕互評的客觀性。

(四) 對於課堂使用Rubrics的建議

1. 建議讓學生參與 Rubrics 標準制定過程，以及提供更多訓練：若要學生應用Rubrics互評，Rubrics 評分標準最好由師生共同商定，教師應鼓勵學生投入參與訂定評分標準的討論，共同決定標準項目與分數範圍，讓學生有機會從更廣泛的視野來了解學習的目標以及作業的特色。另一方面，老師需花時間仔細解說評分標準，最好能呈現不同得分的報告或作業做為參考範例，增加練習機會，以確定學生清楚了解，才能讓學生評分更為一致。
2. 評分時間要充足：由於本研究專題報告的評分項目有11項，每組報告完後，雖然給學生5分鐘時間評分，仍有部分學生感覺時間不足，或許是初次使用 Rubrics，比較不熟練，因此，給予較無經驗的學生評分，時間應再充裕些。
3. Rubrics 每一項分數以0~10給分，或分成數個等級評定為佳：研究者在期中使用的報告大綱 Rubrics 有三項評分標準，分數為0~100分，學生反應以0~100給分範圍太廣，要花較多時間斟酌分數。因此，建議以0~10分給分，或依據文獻所建議的，以偶數等級評等，例如卓越、良好、普通、待改進等四等級。
4. 採用檢核表形式：列出所要評量的項目，讓評分者一項一項勾選，最後再加總計算分數。
5. 增加評語欄位：評分者可以在此處寫下具體的建議或是評語，有助於學習者改進，同時反應評分者的觀點。

誌 謝

本研究承蒙玄奘大學提供經費補助。

參考文獻

- 史美瑤（2012）。提升學生學習成效：評估表格（Rubrics）的設計與運用。評鑑雙月刊，40，39-41。
- 【Shi, M.-Y. (2012). Improve students' learning effect: Design and application of Rubrics. *Evaluation Bimonthly*, 40, 39-41.】

池俊吉（2011）。大學校院推動學習成效為本教育應有之認知與作為。

評鑑雙月刊，**33**，28-32。

【Chi, J.-J. (2011). Promote learning effect as the basic cognition and action in college education. *Evaluation Bimonthly*, 33, 28-32.】

余民寧（2011）。**教育測驗與評量:成就測驗與教學評量**（三版）。臺北市：心理。

【Yu, M.-N. (2011). *Educational test and measurement: Achievement tests and teaching evaluation* (3rd ed.). Taipei, Taiwan: Psychological.】

李坤崇（2006）。**教學評量**。臺北市：心理。

【Li, K.-C. (2006). *Teaching evaluation*. Taipei, Taiwan: Psychological.】

李紋霞（無日期）。**學習成果導向評量：Rubrics運用與設計**。取自ctl.scu.edu.tw/photo/photo_td/activity/1010606/ppt.pdf

【Li, W.-X. (n.d.). *Learning result evaluation: Application and design of Rubrics*. Retrieved from ctl.scu.edu.tw/photo/photo_td/activity/1010606/ppt.pdf】

吳清山（2012）。善用評量方法提升學生學習成就。**教育人力與專業發展**，**29**(4)，1-4。

【Wu, Q.-S. (2012). Using evaluation to enhance student's learning achievement. *Educators and professional development*, 29(4)，1-4.】

吳清山、林天祐（1997）。真實評量、實作評量、卷宗評量。**教育資料與研究**，**15**，67-69。

【Wu, C.-S., & Lin, T.-Y. (1997). Authentic evaluation, performance evaluation, and portfolio evaluation. *Educational Resources and Research*, 15, 67-69.】

教育部青年發展署（2006）。**在正式教育中提升就業力：行政院青輔會大專畢業生就業力調查摘要報告**。取自www.center.nkfust.edu.tw/paper/jobsurvey.pdf

【Youth Development Administration, Ministry of Education. (2006). *Increasing employability in formal education: The survey on college graduates' employability*. Report of National Youth Commission, Executive Yuan. Retrieved from www.center.nkfust.edu.tw/paper/jobsurvey.pdf】

- 彭森明（2010）。大學校院如何推展學生學習成果評量。評鑑雙月刊，**24**，28-34。
- 【Peng, S.-M. (2010). How to promote the evaluation on student' learning result in college. *Evaluation Bimonthly*, 24, 28-34.】
- 曾慧青（2009）。大學畢業生就業問題與對策。取自<http://www.npf.org.tw/post/3/6075>
- 【Zeng, H.-J. (2009). *Employment problems and policies of college graduates*. Retrieved from <http://www.npf.org.tw/post/3/6075>】
- 趙建豐、林志隆（2005）。評分規準（Rubric）的類型、發展方式與使用原則。國教天地，**161**，99-107。
- 【Zhao, J.-F., & Lin, Z.-L. (2005). Types, developing methods, and application of Rubrics. *World of Compulsory Education*, 161, 99-107.】
- 藍珮君（2012）。以多面向Rasch測量模式分析TOCFL口語測驗評分者訓練效果。取自http://www.sc-top.org.tw/download/research/Pei_Jun_Lan_04.pdf
- 【Lan, P.-J. (2012). *Using many-facet Rasch measurement to examining rater training effects of TOCFL Speaking*. Retrieved from http://www.sc-top.org.tw/download/research/Pei_Jun_Lan_04.pdf】
- 蘇錦麗、黃曙東、浮絲曼（2011）。評分量尺rubrics在大學生學習成效評估之運用。教育研究月刊，**207**，18-31。
- 【Su, J.-L., Huang, S.-D., & Fou, S.-M. (2011). Application of Rubrics in the evaluation of college student's learning effect. *Journal of Educational Research*, 207, 18-31.】
- Andrade, H. G. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching*, 53(1), 27-30.
- Andrade, H. (2007). Self-assessment through rubrics. *Educational Leadership*, 65(4), 60-63.
- Association of American Colleges & Universities (2010). *VALUE rubric development project*. Retrieved from <http://www.aacu.org/value-rubrics>
- Dunbar, N.E., Brooks, C.F., & Kubicka-Miller, T. (2006). Oral communication skills in higher education: Using a performance-based evaluation rubric to assess communication skills. *Innovative Higher*

- Education*, 31(2), 115–28.
- Green, R., & Bowser, M. (2006). Observations from the field: Sharing a literature review rubric. *Journal of Library Administration*, 45(1–2), 185–202.
- Halonen, J. S., Bosck, T., Clay, S., McCarthy, M., Dunn, D. S., Hill IV, G. W., McEntarffer, R., ... Whitlock, K. (2003). A rubric for learning, teaching, and assessing scientific inquiry in psychology. *Teaching of Psychology*, 30(3), 196–208.
- Jackson, C. W., & Larkin, M. J. (2002). RUBRIC: Teaching students to use grading rubric. *Teaching Exceptional Children*, 35(1), 40–45.
- Kerby, D., & Romine, J. (2009). Develop oral presentation skills through accounting curriculum design and course-embedded assessment. *Journal of Education for Business*, 85, 172–179.
- Lapsley, R., & Moody, R. (2007). Teaching tip: Structuring a rubric for online course discussions to assess both traditional and non-traditional students. *Journal of American Academy of Business*, 12(1), 167–172.
- Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation*, 7(25). Retrieved from <http://PAREonline.net/getvn.asp?v=7&n=25>
- Moni, R. W., Beswick, E., & Moni, K. B. (2005). Using student feedback to construct an assessment rubric for a concept map in physiology. *Advances in Physiology Education*, 29, 197–203.
- Montgomery, K. (2002). Authentic tasks and rubrics: Going beyond traditional assessments in college teaching. *College Teaching*, 50(1), 34–39.
- Moskal, B. M. (2000). Scoring rubrics: What, when and how? *Practical Assessment, Research & Evaluation*, 7(3). Retrieved from <http://PAREonline.net/getvn.asp?v=7&n=3>
- Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435–448.
- Riebe, L., & Jackson, D. (2014). The use of rubrics in benchmarking and

- assessing employability skills. *Journal of Management Education*, 38, 319-344.
- Selke, M. J. G. (2013). *Rubric assessment goes to college : Objective, comprehensive evaluation of student work*. Lanham, MD: Rowman & Littlefield Education.
- Skillings, M. J., & Ferrell, R. (2000). Student-generated Rubrics: Bringing students into the assessment process. *The Reading Teacher*, 53(6), 452-458.
- Yoshina, J. M., & Harada, V. H. (2007). Involving students in learning through rubrics. *Library Media Connection*, 25(5), 10-14.

Development and Evaluation of Rubrics for Assessing the Performance of College Students in Human Resource Management Courses

Yu-Ching Liu*

Abstract

Two sets of rubrics were developed for assessing the performance of students in Human Resource Management courses. The rubrics were reviewed by 3 experts and discussed with students in class. The final versions of the rubrics were used for assessing students' performance. In addition, a questionnaire was designed to obtain students' opinions on using the rubrics for assessment. The participants were 2 college professors and 51 college students. They were divided into 10 groups. The results showed high interrater reliability among the professors' scores, as well as between the professors' and students' scores; interrater reliability results were higher when the criteria were concrete and quantifiable. Despite the use of rubrics resulting in slightly higher interrater consistency than the use of traditional grading methods did, the differences were negligible. However, the intrarater correlations (scores graded using 2 methods by the same person) were higher than the interrater correlations (scores graded using the rubric method by different people). The results from the questionnaire revealed that the students had positive attitudes toward the uses of rubrics for classroom assessment and peer assessment. On the basis of these findings, I offer specific recommendations for future studies of the application of rubrics to classroom assessment.

Keywords: rubrics, learning outcome assessment, performance assessment, reliability

DOI: 10.3966/199679772016063301003

Received: November 2, 2015; Modified: January 21, 2016; Accepted: May 16, 2016.

* Yu-Ching Liu, Assistant Professor, Department of Applied Psychology, Hsuan Chuang University, E-mail: yuching@hcu.edu.tw

