

從知識到可觀察的能力： 評估學習成效的策略與建議

■ 文／黃淑玲
高雄醫學大學助理教授

教與學互為一體兩面，包含了教學目標的建立、學習者目標的建構、學習作業的設計、師生間學習重點的溝通、學生課業的輔導、學習成果的回饋等。學習成效評估的核心即在評估學習者的知識層次、因知識而產生的外顯行為，以及學習過程中的參與情形。因此，評估需要充分融合於教學活動中，而分數是知識與學習行為的總和。挑燈夜戰、考前衝刺，並無法充分說明學生獲得知識的層次，僅能測知考生在面對考試壓力時，自我情緒控制的能力！

評估學習的方式需要融入課程規劃，是一個不斷蒐集、了解學生學習進展的過程，而不是某一時間點測試學生的知識深淺、記憶強弱的作法。評估要有積極意義，該是一種為學習而評估（*assessing for learning*），並回饋整體教學的重要策略。因此，本文的重點即在探討下述兩點，並輔以案例穿插說明，試圖作進一步的闡釋與論述：

- 1.如何將評量融入課程設計中；
- 2.如何將知識轉化為可觀察、能評量的能力，並運用評分策略，達成改善教學與學習的目的。

值得一提的是，上述兩點彼此有交錯與重

合之處。案例說明中，筆者引用自己的評估設計供參，除了說明學習成效評估是課程設計的一環之外，也藉此分享知識轉化為可觀察、能測量的技能之機轉所在。同時，筆者也引用不同領域的評估策略，相互對照，試圖異中求同、同中求異。

從講授到過程導向的探索式學習

學習成效評估的論述中，已將傳統的教師為中心（*teacher-centered*）的講授方式，轉向學習者為中心（*learner-centered*）的教學模式。但是兩者除了授課方式改變以及學生的課堂參與程度不同之外，評量學生學習成果是否有差異？例如數學、化學等基礎學科，又如何讓學生有更多的參與，並評核學生的學習進展？

舉例而言，數學課程仰賴許多小考，以測知學生累積知識的程度。若採紙筆測驗，基本的作法是可以要求學生在答案卷的左頁寫出算式，相對於算式則可在右頁解釋算式背後的思考與原因。若進一步將講授式的被動學習，設計成「過程導向的探索式學習」（*Process Oriented Guided Inquiry Learning, POGIL*），引導主動學習（*active*

learning) 的型態，則可將知識建構拆解成認知 (cognitive)、社會性 (social)、情意 (affective)、後設認知 (meta-cognitive) 等四個面向，作為評量學習過程的依據，使知識成為可觀察、能測量的能力。POGIL的教學源自進階生物化學課 (biochemistry) 的課程改革 (Loertscher, 2010)，將行之多年的講授型態改為主動學習的教學法。表一「POGIL 學習成效」則依上述四個面向分列之。

表一可說是一份評分列表，而四個面向中的學習目標亦可依照科目不同而調整。比起紙筆測驗的評量，POGIL教學法中的課堂知識，成為可以被觀察的過程、學習互動的社會行為。同時，在學習成效評估上，除了紙筆測驗以了解學習者的認知層次之外，亦可在教學過程中穿插小組問題解決、學生自評與互評等評量方式，測知學生在學習行為上

的社會性、後設認知的發展情形。

多元評估 蒐集學習表現相關資料

學習是個複雜的現象，所以需要依照教學目標設計學習目標，並選擇適合的作業類型或學習活動，評估學生參與後達成學習目標的程度 (Maki, 2010, p.156)；依照評估結果也同時再度調整教學設計。圖一說明學習目標、教學目標與評估相互循環的三角關係。若三者彼此媒合，可使教、學更有效率、緊密結合。

想要徹底了解學生是否掌握教師所傳授的知識，必須考量上述表一的四個面向。因此，需要運用正確、多樣化的評估方式，彼此交叉使用，了解學生投入學習過程的情形與努力的程度。作為評分依據的學生學習資料可能因為學門領域而有別。表二依照認知

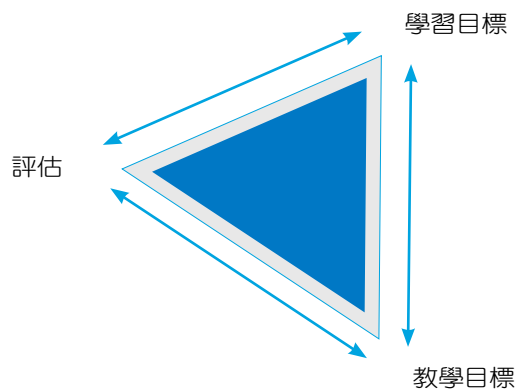
表一 POGIL學習成效

面向	目標
認知	<ul style="list-style-type: none"> • 生物化學領域的專業用字遣詞 • 運用研究方法增進解決問題的能力 • 增進閱讀領域內代表性期刊的能力 • 分析與詮釋資料 • 增進視覺與建構模型的能力 • 增進提問、假設檢定、解決問題的能力 • 強化批判思考力 • 連結新知與舊聞的能力 • 了解整體架構
社會性	<ul style="list-style-type: none"> • 能互助合作 • 能傾聽同儕意見，並從中學習 • 尊重他人的價值 • 能信守對團體的承諾
情意	<ul style="list-style-type: none"> • 能相信自己具有學習力並運用教材的能力 • 能提升面對知識差異時的包容力，並強化信實的態度 • 能訂定成長目標 • 能在必要時尋求協助
後設認知	<ul style="list-style-type: none"> • 能自我引導：主動規劃學習歷程 • 能自我回饋：回顧目標、目的與結果 • 能成為自我評估者：評估自己能力成長與缺失所在，並了解尚需持續改進之處

資料來源：“Classroom Assessment in Support of Biochemistry Course Reform at Seattle University” by J. Loertscher, 2010, in J. Ryan, T. Clark, & A. Collier (Eds.), *Assessment of chemistry* (pp. 117). Tallahassee, FL: Association for Institutional Research.

程度、對應的學生作業類別、評估方式等進行整理說明。

表二的重點在學習目標中的知識層次必須媒合作業型態，而作業型態則必須配合評分方式。例如，初階課程不適合分析、評估甚至創造的認知層次，所以評估學生學習的作業類型自不能是研究構想書或企劃案等；而重視同儕互動與合作的課程，可能多倚重小組成果作為評分依據，而不能僅以紙筆考試



圖一 教—學—評三角關係

資料來源：“Align Assessments with Objectives”
by Eberly Center of Teaching Excellence
& Educational Innovation, Carnegie Mellon
University, n.d. Retrieved from <http://www.cmu.edu/teaching/design/teach/design/assessments.html>

表二 認知—作業別—評估對照表

學習目標	作業類別	評估方式
記憶 學生能： • 想起 • 回憶起	1. 紙筆測試的題項即在要求學生能回憶起、想起資訊： • 填空 • 選擇題 • 在圖形中標示答案 2. 背誦（口頭、音樂、默寫）	1. 正確度：對、錯比例 2. 題項分析（修課班級為單位，錯誤率最高的題項為何？錯誤的答案內容是否類似？）
了解 學生能： • 詮釋 • 舉證 • 分類 • 總結／概論 • 推斷 • 比較 • 解釋	報告、口頭／書寫考試、問題、課堂討論、概念圖、回家作業等要求（口頭或書寫）： • 概述閱讀、影片、演講等內容 • 比較或對照兩種以上的理論、事件、流程等 • 運用既有的標準將案件、元素、事件等分級或分類 • 概述文件或演講內容 • 找出或辨識某概念、原則的例子或圖示	評量尺規，尺規中的評分等第能鑑別作業（品）在尺規各個向度的品質程度
運用 學生能： • 執行 • 貫徹	透過活動要求學生運用流程解決問題或完成（不）熟悉的工作項目；也會要求學生先決定那個流程最適合解決被賦予的工作項目。活動包含：問題組、展演、實驗、模擬。	正確度、流程確認單、評量尺規、表現行為分析
分析 學生能： • 分辨 • 組織 • 歸類	活動中要求學生能分辨或挑選出相關的部分，決定各項元素放在一起如何運作，或是判定被呈現出的材料中所含有的偏誤、其背後的價值觀或潛在意圖。這些可能包含：案例研究、評論、實驗室、書面報告、計畫、辯證、概念圖等。	• 評量尺規，並由授課教師、陪審團、外部客戶、雇主、實習指導員等評分 • 表現行為分析
評估 學生能： • 確認 • 評論	要求學生能依照現有的標準進行測試、追蹤、評斷或評論閱讀資料、表演活動，或產品。這些活動包含刊物雜誌、日誌、評論、問題組、產品測試、案例研究。	• 評量尺規，並由授課教師、陪審團、外部客戶、雇主、實習指導員等評分 • 表現行為分析
創造 學生能： • 發想 • 規劃 • 產出	研究計畫、音樂、表演、短論、商業企劃、網頁設計、產品設計	• 評量尺規，並由授課教師、陪審團、外部客戶、雇主、實習指導員等評分 • 表現行為分析

資料來源：“Align Assessments with Objectives” by Eberly Center of Teaching Excellence & Educational Innovation, Carnegie Mellon University, n.d. Retrieved from <http://www.cmu.edu/teaching/design/teach/design/assessments.html>

作為首要的評分方式。大學的教學評量也常提及「授課教師是否能運用作業、測驗、實作等方式合理測出我（學生）的程度」，原因就在於校方想透過學習者觀點，了解知識傳授者是否能將「知識層次—作業型態—評分方式」三者作最恰當的整合，促進前述為學習的評估（assessing for learning），達成圖一和諧的三角關係。

評分是評估成效的基礎

如前所述，分數具有指標性的意義，反應學生獲得知識的程度，了解學生的學習成果，而老師也從中得知教學成效。教師們總強調自己在教學上非常重視知識的傳遞、辯證思考、分析與整合等能力。但依據美國國家教育統計中心（National Center for Education Statistics）（Jones, Hoffman, & National Center for Education Statistics, 1995）的研究顯示，老師們並沒有將這些教學目標融入課程設計中。老師們多評量學生獲得的知識，但針對表一各個面向的發展程度，卻沒有相對應的評核工具了解學生的學習情形。

根據諸多類似教學卓越與輔助單位的實務經驗，以及實證研究（蔡佩鈞譯，2013）提及，學生如果只知道自己的分數或成績等第，如A、B、F等，並無法得知自己表現的完整面貌，例如那些面向很好可以繼續保持，那些是未來繼續努力的方向等。而這個問題在期末書面報告最常見。上表二中提及的評量尺規，其意義、目的、原理、種類、運用方法，以及帶領教學助理開發與運用此評量工具等，也有學者專著介紹（Stevens & Levi, 2005）。尺規具體的用處在於幫助授課者思

考授課內涵的知識層次、應達成的學習目標與外顯行為、達成的能力面向之優缺良窳，除了評核當下的學習狀態外，也提供明確的努力方向。

評量尺規案例與製作原理

表三是筆者參照文獻，並加以修訂編製而成的書面報告評量尺規，以說明評分者如何評量學生的知識，並將知識拆解為可觀察、能衡量的評分準則（criteria）；此準則亦可作為學生自我回饋，了解優缺的學習建議單。

為了讓知識轉化為可觀察能衡量的能力，在建構評量尺規時，筆者運用了下述的製作原理：

1. 開宗明義

說明制訂該尺規的目的與運用範圍。這個說明可以提醒評估者與受評者雙方如何看待該尺規，以及評分結果。表三提及，「由於本課程的核心能力是希望修課學生能具備（性別）專業知識，並可以進一步批判思考。因此，設計本評量尺規的目的是在衡量學生上述能力表現在撰寫書面報告的情形。本尺規所列出的評分指標皆在評核學生展現在報告中的各個能力面向。這些能力面向包含架構、辯證力、研究方法、結論與建議、資料引用，以及可促進相關領域知識發展的程度。」

2. 向度與品質描述

表三中的向度包含「論述能力與原創性」等，依照下述等第進行作業品質的描述，並給予評分。

3. 等第

包含「典範」、「適當」、「尚可」、

表三 研究生學期書面報告評量尺規

由於本課程的核心能力是希望修課學生能具備（性別）專業知識，並可以進一步批判思考。因此，設計本評量尺規的目的是在衡量學生上述能力表現在撰寫書面報告的情形。本尺規所列出的評分指標皆在評核學生展現在報告中的各個能力面向。這些能力面向包含架構、辯證力、研究方法、結論與建議、資料引用，以及可促進相關領域知識發展的程度。					
等第	典範	適當	尚可	亟需努力	學習成效準則
分數	50（每項4'）	40（每項3'）	30（每項2'）	20（每項1.5'）	
向度：表現	論述能力與原創性： 1. 書面報告的（性別）議題分析呈現出思考性、有深度、有創意，且具有延伸性； 2. 不僅陳述既定的看法或符合常情的觀點，立論深思熟慮，開展讀者對該作品的思考； 3. 能以多元觀點或第三種觀點的論述與脈絡，探討該主題； 4. 能拆解（性別）議題在不同學門領域或工作場域所呈現出的結構性或系統性問題； 5. 從書面報告中可看出撰寫者的好奇心與思考性。	1. 報告內容深思熟慮，但分析較薄弱； 2. 文中能認知第三種詮釋，亦了解文獻與真實（性別）議題的複雜性； 3. 論點合理，促進讀者思考； 4. 文章其他部分能依相關性進行統整。雖提及（性別）議題，但未主導分析的內涵； 5. 研究探討部分能整理訪談結果，並提出研究觀點。	1. 論點不夠清晰有力，較看不出創造性； 2. 觀點顯然多來自課堂閱讀與討論，或是人之常情。文章內容沒有將相關的課堂閱讀融入短文中； 3. 論點稍顯平凡，雖可接受但並未具有建設性； 4. 統整性不足，離主軸議題不夠明確； 5. 研究部分多在呈現訪談摘要。	1. 論點難以辨識，讀者需要自己推敲； 2. 多摘述他人觀點，或重複課堂討論，或將逐字稿剪貼在報告中，而缺乏分析； 3. 沒有論述，僅是說明或陳述事實； 4. 沒有統整，亦缺乏主題性； 5. 僅呈現訪談摘要，研究者鮮少提出任何說明。	條理清晰，讀者能依循研究邏輯所在；能蒐集資料，能交互援引不同資料來源的論述。
	組織架構與一致性： 1. 撰寫者掌握報告寫作的方向與架構，且標題明確； 2. 文章理路有邏輯性，使讀者容易依循研究報告中的邏輯所在，結論合理而非意料之外； 3. 次要點論述流暢俐落，能有效破題進行論述； 4. 能有效闡述研究問題、研究過程、結論或建議。	如左述，但 1. 標題不明； 2. 邏輯合理； 3. 撰寫者並無法有效破題進行論述。研究架構合理，立論前後一致； 4. 結論或建議相對不足。	讀者可感受到撰寫者掌握了寫作方向與整體架構。報告總體上可看出邏輯的脈絡。研究架構不夠緊密，其鋪陳、過程、結論與建議合理，但不具有開創性。	報告多少可辨識出主要論點。雖有研究架構，但在鋪陳、研究過程、結論或建議上，僅符合常理。	能有效闡述研究問題、研究過程、結論或建議。
	佐證、援引文獻： 1. 透過各種管道蒐集資料； 2. 能找到文獻論述的重點，與自己研究議題連結，以進行立論之鋪陳、佐證、反證等； 3. 文獻資料多元，包含課堂教材、次級資料等，並交互援引其中的觀點、特定事件、結論等，達到辯證的目的，支持自己的觀點。	撰寫者透過適當管道蒐集資料並整理；能引用文獻論述重點，結合自己的觀點。交互援引資料辯證合理，但可更緊密。	撰寫者的論點與詮釋，大致上都有文獻資料的支持，也能引用重要文章的摘要、觀點、結論，但充實度或緊密度不足。	撰寫者對於作品的論點與詮釋，多僅是整理訪談內容，缺乏引用相關文章的摘要、觀點、結論，援引之充實度顯見不足。	能從不同管道蒐集資料、能透過多元資料中尋找答案並交互援引論述，達到辯證的目的。
	語氣、調性、效度： 用字遣詞準確且高雅，達到學術著作的格調。	語句清晰可讀，用字遣詞合理，但寫作技巧可以再提升，或有些句子前後缺乏連貫性。	全文的語意可理解，但語彙直白，缺乏書寫的美感。	有時語意不明。某些句子條理不清。	條理清晰，研究脈絡條理清晰、有邏輯。

資料來源：1.筆者編製修訂。

2.黃淑玲(譯)(2013)。《新手也能上手的學習成效評估》(原著者：Barbara E. Walvoord & Trudy W. Banta)(頁146-148)。臺北：高等教育出版社。

「亟需努力」。至於等第數需要偶數還是奇數，專家學者各執一詞沒有共識，而運用者更可依照使用後的經驗加以調整修訂。筆者個人則贊成四個類別，原因在於偶數類別可以避免趨中現象；若是三個類別稍嫌精省，五個類別又過於精細難以制訂。

4. 學習成效準則

此為筆者開設此課程所屬的研究所內共同能力標準，亦是筆者因地制宜，將評量尺規的各個向度媒合成效準則，以利檢視學生所發展出的能力在各項成效準則的學習結果。

5. 分數

即依照各向度的等第進行分數配置。分數多寡依照評分者實際的配分而訂定。表三的範例是書面報告占學期總成績的50%，因此配分屬於「論述能力與原創性」的表現共有5項，在「典範」類每項4分、「適當」則3分，以此類推。而在此5項中，有些落在「典範」，或「適當」，甚至有些落在「尚可」。所以「論述能力與原創性」這個向度的得分，最高20分，最低7.5分。「組織架構與一致性」等向度皆以此類推。

運用評量尺規的意義

運用評量尺規，可避免評分者與被評者之間對於作業評分的看法不一致，且具有下述幾項積極意義：

1. 將評分的心靈活動轉化為標準化的過程

評分標準明確，個人表現到班級整體學習成效之良窳都能明確可辨。評分的標準化亦可降低評分者的工作份量，縮短作業批改的時間。同時，師生也可針對尺規內涵進行討論，讓學生能掌握自己的學習過程，自我驅

策。這一點則與下一項有關。

2. 評分結果引導教學方向與學習

教師宜教導學生如何評估自己和同儕的作業成品，可達到較高的認知層次外，更可幫助學生發展表一所提到的「後設認知」的能力面向。因此，教學的過程除了知識傳承，也包含傳達品質標準的概念，而可脫離符合考題要求的制式想法；這也就是學習成效學者所提倡的「教學是為了達成學習成效準則，並非為了考試而教學」（Teaching to the criteria, not to the test.）（Walvoord & Anderson, 2010, p.62）概念。

若擴大運用評量尺規，經過教師同儕、師生之間的溝通、共同開發的歷程，使全系或學院內運用同一尺規進行特定作業型態的評分，則很容易看出全系、院的學生之學習表現與成效。此作法的效益，小者可指引授課單元重新規劃與設計，大者可引導全系的課程改革，例如必、選修課程重新調整、增設特定課程以加強訓練、重新設計既有課程的授課目標與單元。這些評量結果若加以彙整，皆可作為日後系所自評的重要佐證資料；若到學院層級，則成為開設院級課程的依據，以利跨系選修，促進教學資源的橫向整合與運用。

評量課堂小組討論

除了書面報告，課堂討論是另一種了解學習成效的教學策略。小組討論可以得知學生如何形成問題意識、意見交換與溝通（辯證），以及整合知識的方法。但是授課教師對於這種既能觀察知識成長情形，又能促進同儕共學的課堂形式，常有捉襟見肘、不知如何評估學習成效的情形。為善用課堂小組



▲同儕討論有助於提高學習成效。(陳秉宏／攝)

討論，並測知學生學習，筆者以自己教授的通識課程作為範例說明。

課程開始之初，先以調查問卷了解學生對筆者（即授課者）設定的各項學習目標進行排序。結果發現，學生對於「能客觀批評他人意見，進行溝通與解決問題」之學習目標有極高的學習意願。筆者因而調整教學策略以媒合此學習目標，並以下列評分標準作為評估同學課堂討論的標準：

- 能確認他人的發言內容、例子或事證。
- 能將自己的發言建立在他人的意見上，例如協助提供例證。
- 能將個人意見、觀點與課堂閱讀作連結。
- 能指出他人觀點中的問題、疑點或預設立場。
- 能總結或整合各種觀點或意見。
- 能提出另一種不同觀點，並加以說明、解釋。
- 能談論指定閱讀如何影響或改變個人的看法，或解釋自己的經歷。

- 能談論指定閱讀或他人觀點如何幫助自己反思、回饋，或了解人類多元的經驗。

上述標準可視為一評分表單（checklist），引導同儕互動，並使討論內容能與課堂主題有密切的關聯性。透過學生自評、同儕互評，甚至正向表列，讓學生列舉典範組員等，都是很好的方式。

魚缸式討論法 促進後設認知發展

全班經過分組，並運用上述評分標準進行問題討論後，可運用「魚缸式討論法」的進階討論策略。意即請每組自推代表一名，並將他們組成新的討論小組，坐在全班中間，外圈則是所有各組組員。坐在中間的小組就像魚缸內的魚，魚缸外就是坐在外圈的所有組員，再次聆聽魚缸內的組員如何討論剛才的議題。所以理想的狀況是，代表們都是表現優異、足為典範的成員。經過典範組的示範討論，大家可能得到新的看法與結論。

完成小組討論以及魚缸式討論法後，再

讓學生將當天的課堂討論，依照問題寫成書面心得。如此一來，學生們面對同一個問題歷經「一般小組討論」、「魚缸式討論」、「隨堂心得撰寫」等三個不同層次、口語和撰寫並濟的意見交流與回饋活動後，不僅可達到「能客觀批評他人意見，進行溝通與解決問題」的學習目標，亦可從中了解他人多元觀點與經驗。對魚缸式討論法搭配心得撰寫的授課方式愈熟悉，學生也愈容易達成前述表一所列的學習成效。有趣的是，學生展現在書面心得中的自我引導、回饋、評估等後設認知的發展，也會隨著學期進程而愈趨明顯。

評量學生獲得知識的方式，不僅止於紙筆測驗，而是了解學習過程，並將評分作為衡量學習目標，調整教學的工具，即為增進學習的評估（*assessing for learning*）。本文以他校作法與筆者執教案例，說明評量融入課程設計、轉換知識為可觀察、能衡量的能力以評估之。

無論是研究型或教學型大學，教學的重要性不容忽視，原因在於傳遞知識與培育現代公民，為21世紀高等教育的共同課題。建立有意義（*meaningful*）、切身的（*relevant*）評估方式，使教、學與評估更具統整（*integrity*），也可說是提升教學品質、提升學習的基石。



教、學、評分三者合一 統整有效率

◎參考文獻

- 黃淑玲(譯)(2013)。《**新手也能上手的學習成效評估**》(原著者：Barbara E. Walvoord & Trudy W. Banta)。臺北：高等教育出版社。
- 蔡珮鈞譯、Sarah Richardson著(2013)。高等教育學習成果評量(AHELO)之可行性研究。《**評鑑雙月刊**》，43，19-21。
- Eberly Center of Teaching Excellence & Educational Innovation, Carnegie Mellon University. (n.d.). *Align assessments with objectives*. Retrieved from <http://www.cmu.edu/teaching/design/teach/design/assessments.html>
- Jones, E. A., Hoffman, S., & National Center for Education Statistics. (1995). *National assessment of college student learning: Identifying college graduates' essential skills in writing, speech and listening, and critical thinking*. Washington, D.C.: U.S. Department of Education.
- Little, D. (2006). Grading with rubrics: Developing a fair and efficient assessment tool. *Teaching Concerns, Fall 2006*. Retrieved from http://trc.virginia.edu/Publications/Teaching_Concerns/Fall_2006/TC_Fall_2006_Little.pdf
- Loertscher, J. (2010). Classroom assessment in support of biochemistry course reform at Seattle University. In J. Ryan, T. Clark, & A. Collier (Eds.), *Assessment of chemistry* (pp. 113-125). Tallahassee, FL: Association for Institutional Research.
- Maki, P. (2010). *Assessing for learning: Building a sustainable commitment across the institution*. Sterling, VA: Stylus.
- Stevens, D., & Levi, A. (2005). *Introduction to rubrics: An assessment tool to save grading time, convey effective feedback and promote student learning*. Sterling, VA: Stylus.
- Walvoord, B., & Anderson, V. (2010). *Effective grading (2nd ed.)*. San Francisco, CA: Jossey-Bass.