

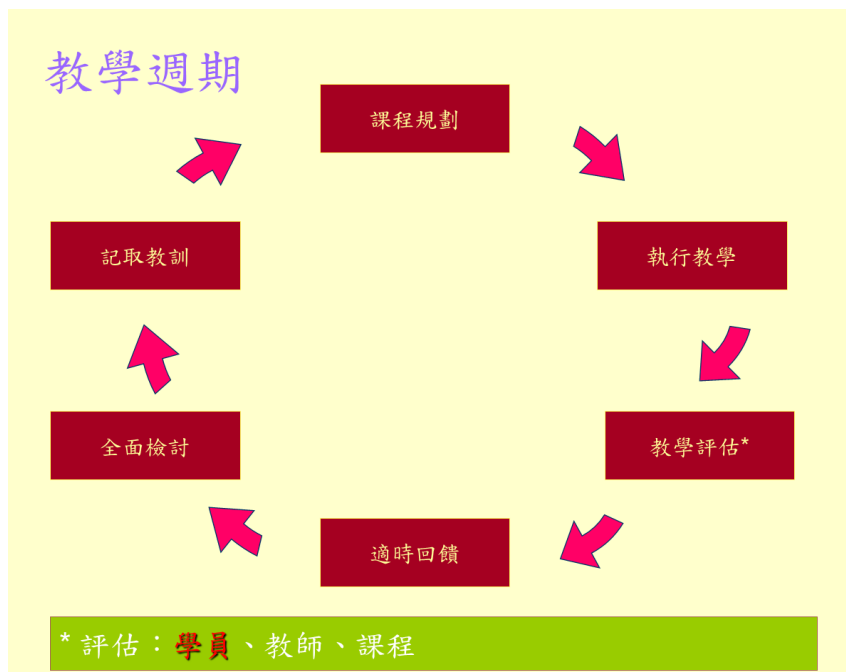
評估總論

臨床教師對學生臨床能力評估需了解下列幾個項目：

- 一、評估的基本理念
- 二、評估的類別
- 三、評估方法的選擇
- 四、評估的深度
- 五、評估的層面

一、評估的基本理念

首先要清楚的問題是「為何要評估？」現今的教育規劃也加入管理學最基本明確的思維—「Plan-Do-Check-Act (PDCA) 循環」。如下圖的教學週期所示，教學也有類似 PDCA 循環：「詳作課程規劃—依照計畫執行教學—評估教學狀況—回饋予相關人員—反思及檢討—了解得失重新規劃」，期望課程愈來愈好，學生的學習成效愈來愈佳。因此，教育課程必須包含評估方法，而評估之後必須作有效的回饋。因此，評估是教學過程中一個必要的環節。



二、評估的類別

評估的分類並沒有統一，站在不同的觀點評估有不同的分類方式。

1. 是否對受評者有助益—formative vs. summative

Formative assessment 通常在教育訓練進行期間舉行，並對受訓者作出回饋，以改善往後的學習，目的是「為了學習」(assessment for learning)，



也稱作「教育性評估」(educative assessment)或「診斷性評估」(diagnostic assessment)。Formative 一字是來自 formation(形成),便有人翻譯為「形成性評估」,可惜辭不達意,較好的翻譯是「造就性評估」。Summative assessment 通常在教育訓練結束時舉行,評估學生總體學習表現,但不作回饋,只用來判斷考生是否達到某特定水準,目的是「評估學習」(assessment of learning)。Summative 一字是來自 summation(總結),便有人翻譯為「總結性評估」,尚稱合宜。

本項分類法的重要意涵是在於評估者必須了解評估的目的來決定是否及如何進行回饋。一般而言,如果情況許可,應要儘量對受評者給予回饋,使其在評估之後有所獲益。另一方面,回饋必須時機和方法合宜才有效果,教師必須接受相關訓練才能得到效果。

2. 是否能客觀判別—objective vs. subjective

客觀式評估(objective assessment)是指考題只有單一答案的評估方法,如:是非題、選擇題、配對題等。主觀式評估(subjective assessment)是指考題可有不同正確答案的評估方法,如:簡答題、申論題,以及測驗臨床技巧常用的簡短式臨床評量演練(mini-clinical evaluation exercise, mini-CEX)和直接觀察步驟技巧(direct observation of procedural skills, DOPS)等項均為主觀式評估,甚至是客觀結構式臨床測驗(objective structured clinical examination, OSCE)也非客觀(上述三項評估方法將另文介紹)。

本項分類法的重要意涵是在於了解主觀式評估的信度可以因準備的程度是否充分而有極大的落差,假若評估的信度不佳則評估便沒有價值,甚至產生誤導。故評估者在執行任何一項主觀式評估前,必須先行接受訓練,待熟悉評估的理念、程序及標準之後,才可正式上陣執行評估。

3. 是否有固定合格標準—referencing

包括標準參照評估(criterion-reference assessment)、常模參照評估(norm-referenced assessment)及自比性評估(ipsative assessment)。

標準參照評估是依據預先訂定的標準進行「合格/不合格」的判定(如達60分為合格),不考慮考生的相對表現。這樣的評估通常包括確定學生或受訓者是否可以執行特定任務或活動而設置一個最低標準。例如駕駛考試,不論報考者的年齡或教育程度,合格標準都是一樣的。

常模參照評估則是比較全體參與者的考試表現,以某一比率成績較優的考生列為「合格」,故並沒有固定或預設的合格分數。例如大學或研究所入學考試,錄取/合格分數為正取/備取最後一名學生的成績。

自比性評估是評估考生在接受教育訓練之後進步的程度,這種評估可能涉及相同的測試。

本項分類法的重要意涵是在於了解評分標準的訂定是依評估目的而定。在一般臨床教學環境中,較常用的是標準參照評估及自比性評估。如

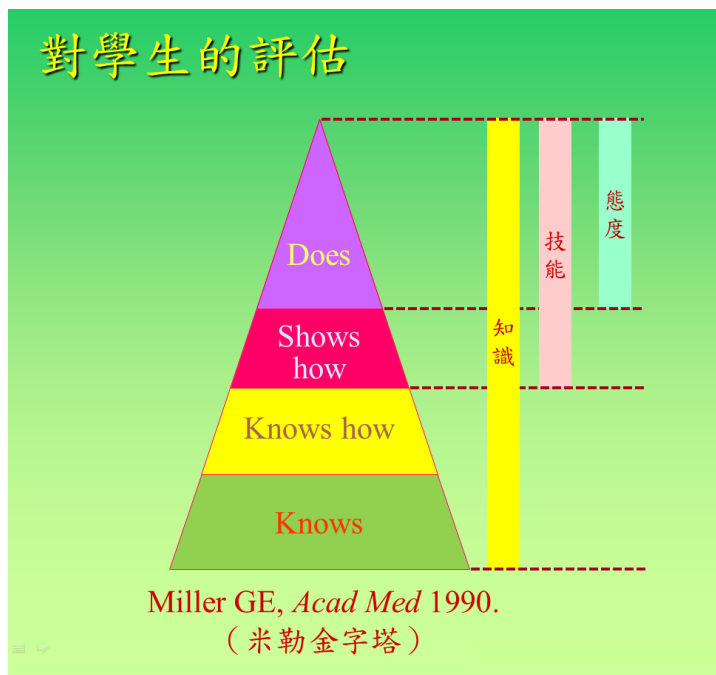
mini-CEX、DOPS 等的評分都是依據預先設定的標準，但由於臨床能力的評估大多主觀，標準參照評估中的「合格標準」必需合理、明確及建立在共識之上。至於自比性評估常以「前測—後測」作比較，但必須排除因重複考試所造成的「學習效應」(learning effect)，故不宜用相同的考題，而是使用性質、範圍及難度均相等的不同題目。

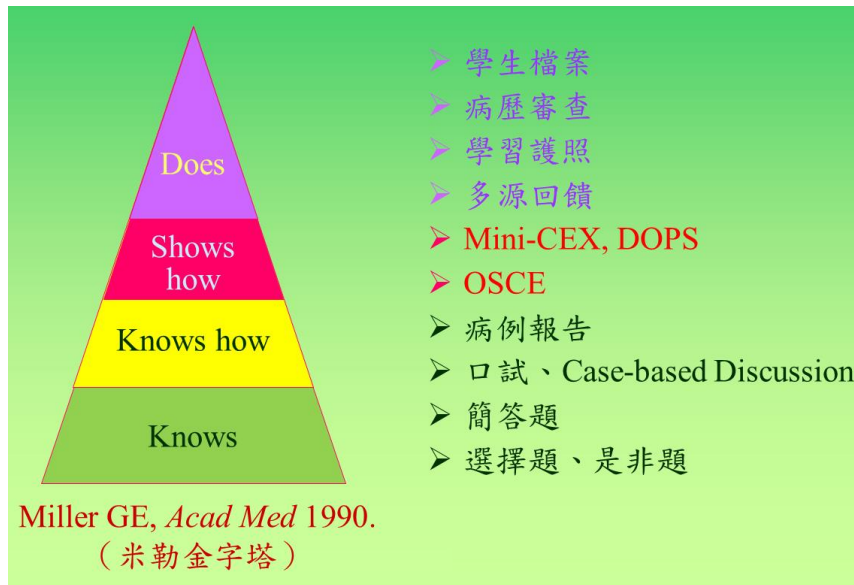
4. 是否為正式評估—informal vs. formal

正式評估 (formal assessment) 是在教育訓練計畫中明訂的評估，而非正式評估 (informal assessment) 則是指非在計畫中排定而是由老師視情況需要隨機附加的評估，通常不列入成績。

本項分類法的重要意涵是在於教師除了執行教學計畫所訂定之評估之外，隨時加入非正式評估，配合適宜的回饋(即進行 formative assessment) 將對學習有很大的幫助。例如備受推崇的一分鐘教師模式 (one minute preceptor model) 也可視為一種以學習為目的之非正式評估。

三、評估方法的選擇





談到臨床能力的評估，一定會提到 George E. Miller 醫師在 1990 年在 *Academic Medicine* 的 9 月號增刊(supplement)中被邀請撰的寫一篇回顧論文“The Assessment of Clinical Skills/Competence/Performance”。該文內容主要是對標準化病人的發展作一個概括介紹。在文中提出一個三角形的臨床評估框架(framework of clinical assessment)，並稱它為金字塔，此後大家便稱此三角形為 Miller Pyramid。而 Miller Pyramid 似乎成了評估的「金科玉律」，神聖不可侵犯，繼之不乏追隨者利用此三角形大作文章，甚至畫蛇添足。而各項評估方法屬於哪一層級更莫衷一是，特別是 mini-CEX 和 DOPS 這些所謂的 workplace-based assessment，到底算是“shows how”抑或是“does”仍是意見分歧。雖然 Miller 在發表這篇文章時還不知道有 mini-CEX（相關文章最早是在 1995 年發表）和定名更晚的 DOPS（since 2007），但從原文的“Does”是指“what a graduate *does* when functioning independently in a clinical practice”以及強調是“professional behavior”便很清楚地知道“Does”要評的是「獨自行醫時的專業行為」，無論是 mini-CEX 或 DOPS，因為老師就在身旁評分，一切行為都是“shows how”，故 workplace-based assessment 不見得都是“Does”！故上圖將 mini-CEX 和 DOPS 定位於“shows how”的層級。

美國「畢業後醫學教育評鑑委員會」(Accreditation Council of Graduate Medical Education, ACGME) 在 2000 年對六大核心能力列出 13 種評估方法(下表僅列第一項: Patient Care)，統稱這是「評估方法工具箱」(toolbox of assessment methods)。這 13 項方法是否就是評估六大核心能力最好的選擇不得而知，因為未在此列出的臨床評估方法據說超過 40 種。然而，這個表帶給我們一個很重要的觀念：每個評估方法各具所長，但只能評估其中部分，沒有一種評估方法能適合於全部能力或技能的評估。此外，每一能力的所需技巧可有兩種或更多評估方式的選擇。要使用哪些方法需由課程設計者決定，要在「評估方法工具箱」選出合宜且符合成本效益的「工具來使用」。

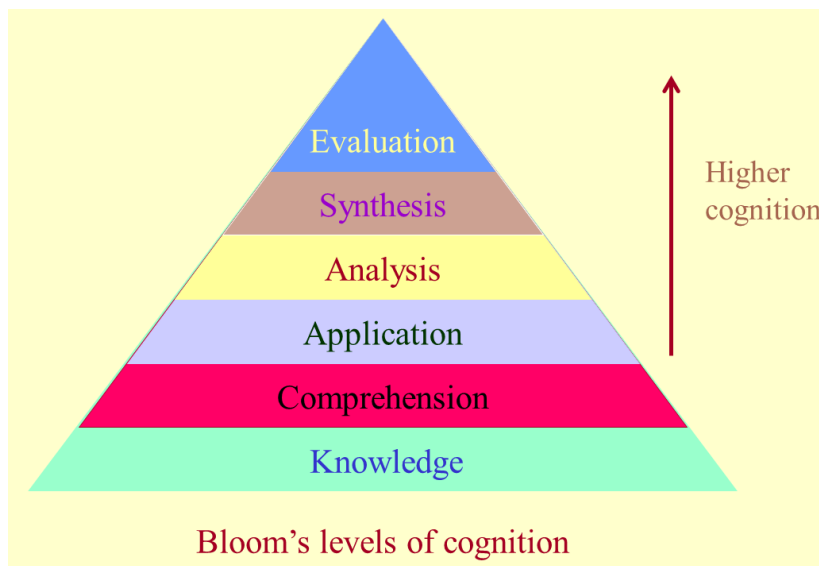
ACGME Competencies: Suggested Best Methods for Evaluation

Competency	Required Skill	1	2	3	4	5	6	7	8	9	10	11	12	13
		Record Review	Chart Stim. Recall	Check-list	Global Rating	SP	OSCE	Simulations & Modets	360° Global Rating	Portfolios	Exam MCQ	Exam Oral	Procedure or Case Logs	Patient Survey
Patient Care	Caring and respectful behaviors			3		1			2					1
	Interviewing			1		2	1		3					
	Informed decision-making		1	2			2				2			
	Develop & carry out pt. Management plans	2	1	2	3			2	3					
	Counsel & educate pt's. & families			3		1	1		2					1
	Performance of procedures a) Routine physical exam			2		1	1							
	b) Medical procedures			1	3			1	2				3	
	Preventive health services	1				2	1			3			2	
	Work within a team			3	3				1					

表格資料來源：ACGME and ABMS Toolbox of Assessment Methods。

四、評估的深度

臨床能力評估方法有數十種，有人以為愈在 Miller pyramid 下方的評估方法愈是粗淺，愈欠效度，這種想法有極大的謬誤。縱使採用同一種方法，只要調整深度，使可以適合任何一個程度的考生。選擇題就是最好的例子，由小學到大學的博士班都可以用來考學生。



每當提及學習（及評估）的深度就會聯想到 Benjamin Bloom 在 1949 年開始引述的教育目標分級（Bloom's taxonomy）。他首先提出在認知層面可分為六個等級（如上圖），其後有人加入技能和態度（如下表），跟著便不同版本及詮釋

相繼出現。與 Miller pyramid 一樣，這是「人為」的分類，不一定適用於所有領域，更不是毫無瑕疵，其最重要的貢獻是提供一個觀念——同樣一個學習議題可有不同的深度，無論教學或評估，要將目標設定深度，才能確認學習的成效。

認知（知識）	精神運動（技能）	情感（態度）
1. 知道	1. 模仿	1. 接受
2. 明白	2. 依指示來做	2. 回應
3. 應用	3. 發展精確性	3. 成為價值（了解並行動）
4. 分析	4. 接合（結合或整合相關技巧）	4. 組織個人價值系統
5. 合成	5. 自然化（反射動作，已成專家）	5. 內化價值系統（納入至行為）
6. 評量		

五、針對特定課程學習成效評估的層面

美國威斯康新大學名譽教授 Donald Kirkpatrick 在 1954 年以「四層次模式作課程評估」為博士論文主題，1959 年在 *US Training and Development Journal* 發表一系列的文章，直至 1994 年出版 *Evaluating Training Programs* 一書，此「四層次模式」才受到重視。Kirkpatrick's 4 level model 包括下列四項：

1. 反應（reaction）：參與者對教育訓練的想法和感受，通常是滿意度調查。
2. 學習（learning）：以測驗方式了解學習者學到多少，相對於 Miller pyramid 中“knows”、“knows how”及“shows how”的評估。
3. 行為（behavior）：評估在學習 3-6 個月後學員執行任務時是否運用所學習到的知識、技能和態度，相對於 Miller pyramid 中“does”的評估。
4. 結果（results）：指因教育訓練對業務產生的最終結果，如因員工服務水準或工作能力上升而導致績效改善及盈餘增加。

其後又有人將四層次加碼更多層次，但因了無新意且沒有助益，目前大多沿用原理論。而在醫院內的教育訓練，大多做到前三個層次而較少做到 level 4（結果）的評估，主要因為影響醫療品質及病人安全的原因甚多，除了教育訓練之外，制度、文化和資訊科技都扮演著同等重要的角色，只做教育而沒有其他方面的配合，將會徒勞無功。課程設計者應於規劃學習目標（learning objectives）時即訂出相應的評估層面、方法和深度。