

導向學習—— 教育實務匯編 的評估

梁佩雲、張淑賢 編著

梁佩雲 編譯



香港大學出版社
HONG KONG UNIVERSITY PRESS

香港大學出版社
香港田灣海旁道七號
興偉中心十四樓

© 香港大學出版社 2007

ISBN 978-962-209-841-1

版權所有。本書任何部分之文字及圖片，
如未獲香港大學出版社允許，
不得用任何方式抄襲或翻印。

本社提供網上安全訂購保障
<http://www.hkupress.org>

進穎電腦製版有限公司承印

目 錄

序	ix
鳴謝	xi
緒論	xii
一 從「對學習的評估」到「導向學習的評估」	1
二 以評估促進學習	7
三 寓評估於學習	21
1. 防止剽竊的論文撰寫安排	23
Deirdre Mary SMYTHE	
2. 兩段評改法	25
郭思豪	
3. 指導 • 預習	29
鄧城鋒	
4. 提升總結性評估的素質	31
廖佩莉	
5. 每人一問	34
關利平 (Lilian L. P. VRIJMOED)	
6. 讓學員在實務中自問自答	36
Brian W. DARVELL	
7. 對談式考卷	39
郭思豪	

8. 小組實地探究、回應及反思 羅天佑	44
9. 關鍵事件分析 Jean MORRISSEY	46
10. 挑戰成績出眾者 陳孝平、莫引鳳	50
四 學生參與：同儕評估和自我評估	53
11. 「快速反饋」的參與式學習 黎少銘	55
12. 以同儕互評確認誤解概念 鄭佩華	58
13. 態勢語言評估 鄺銳強	60
14. 同儕評核小組專題研習 鄭梁慧蓮、Martin WARREN	64
15. 小組互評在閱讀教學上的運用 簡漢乾	68
16. 寫作教學的網上互評 梁慧霞	71
17. 網上互評專題研習 楊友源	74
18. 評估同儕的海報設計 Ken VOLK	77
19. 評估小組導修的協作效能 馬宣立	79
20. 建基於「複習」的自我評估 蘇潔玉	83

21. 輔助學生確認偏誤軌跡 李潔冰	86
22. 自學自評的聆聽能力訓練 潘漢芳	90
23. 專注於評估標準與自我評估 David CARLESS	94
24. 運用錄像評估訓練口語表達能力 盧明輝	97
25. 以「知－想－學」方法評估學習需要 龍精亮、莫慕貞	100
五 反饋	103
26. 測入測出 曾君蘭	105
27. 正誤題檢測法 郭思豪	107
28. 分段式的評估與反饋 余婉兒	110
29. 「對話與靜思」：以導修回應習作初稿 張淑賢	113
30. 確保學生注意、明白並按反饋實踐 學會學習課程組，香港城市大學	117
31. 評估範例與同儕反饋 廖雅芬	119
32. 「專業發展進程圖」—— 教學實習評估 鄧怡勳、王燕茵	124
33. 專題報告發展過程中的形成性評估 林淑華	129

34. 配合個案研習的自動化反饋 Shekhar KUMTA	131
35. 電子反饋庫 梁佩雲	133
詞匯對照表	137
參考資料	149
作者通訊錄	151

—

從「對學習的評估」到 「導向學習的評估」

梁佩雲

教學趨勢轉變：評估重新定位

面對資訊科技的高速發展及全球一體化的大趨勢，世界各地均熱切關注如何能改進教育，以提升人才素質，配合不斷變遷的時代需要。古今中外，教育都離不開考試，以評定學生的學業成績，決定學生的前途。以筆試為主的考試安排，固然能發揮一定的甄選功能，但也同時為考生、教師以致教育制度帶來不少負面影響。最重要的是，考試對學生前途的影響越直接，學生要承擔的風險就越高，而教師、學校也都不敢怠慢，於是教學被迫成為考試的附庸，所謂的「應試教育」應運而生。由於一切學習都只為應付考試，而在有限的考試時間內，往往只能評核學生特定的學習成果，機械式操練和死記硬背學習內容成為主要教學方法，妨礙學生正常的學習生活；教學宗旨和目標，並不能從考試結果中體現。要應付「決定性」的考試，學生已經有甚大壓力；要接受考試失敗的事實，對大部分學生來說更是沉重的打擊。畢竟，學生在學習過程中的其他學習經歷和能力，並不能從筆試的表現反映。另一方面，即使考試高分，也不代表學生的其他學習能力較人優勝，「高分低能」的學生表現，可謂對高風險考試一記辛辣的諷刺。

針對單一考試對教育的負面影響，各界的討論焦點都集中於如何為教育實務提升素質，而最直接的考慮就是如何取替以單一考試評定學生的學習成效。結果，評估在教育過程的性質及功能正面臨劃時代的轉變。

2 導向學習的評估

首先，要更如實而全面地反映學生的不同學習能力和成效，評估工具一定不能再局限於紙筆的考試，而要採取多元化的模式，如：觀察、專題研習、口頭匯報、學習歷程檔案等，才能配合不同學習目的（李坤崇，1999, 2006; Linn & Gronlund, 2003）。其次，學生在不同學習階段的表現不同，所以評估的時間也不應鎖定於學習階段的完結，而要與教學過程結合，適時進行。最後，評估甚至超越純粹反映成績的工具角色，進而反客為主，成為引起學習動機的媒介，改進學習的動力。評估重新定義為：「通過觀察學生的表現、測驗、考試等方式，收集學生知識、能力、價值觀和態度等各方面的學習顯證」（香港教育統籌局，2002）。

換言之，評估性質的轉變，牽動整個教與學過程：評估不再脫離教學獨立，教學過程也不能再漠視評估；評估活動，就是教學活動；通過評估收集學生學習的證據，既顯示學業成績，讓學生知所改進，也可以讓教師檢討，從而改善教學——「評估、教導、學習」形成三位一體的教學循環。

探索有效的評估：「導向學習的評估」研究 (Learning-oriented Assessment Project)

釐清評估的功能和它在教學過程中的作用等理念後，接著的迫切工作，就是開發切實可行的評估模式，以達至「以評估促進學習」的理想預期。有關如何能有效評估方面的研究，不少學者早有研究，而且引起廣泛迴響，如：柏士俊 Biggs (1999) 和李坤崇 (1999, 2006) 均認為評估不應根據學生穩定的特質來把他們區分，主張採用真實、動態和多元化的評估手段；Black 和 Wiliam (1998) 分析進展性評估對不同水平學生學習的影響，香港教育處課程發展署曾把原文譯成中文《暗箱內探：透過課堂評估提高學習水平》，以便向更多中文讀者推介評估的新理念；Boud (2000) 主張培養自我評核的意識，體現終身學習；Elton 和 Johnson (2002) 質疑大學傳統評估機制的成效，建議採用學習歷程檔案；Gibbs 和 Simpson (2004) 綜述有關的文獻開發了 11 項能支援反饋改進學習的條件；Sadler (1989) 主張讓學生認清學習目標及評估準則、掌握評估能力等，都為後來的評估理論和實務的探討，奠下穩固的根基。

為了響應學界日漸重視評估的發展，香港教育學院連同香港大學、香港浸會大學和嶺南大學，曾於 2002 至 2005 年間，協辦了一項名為 Learning-oriented Assessment Project (LOAP) 計劃，計劃成員在香港本地的高等教育界舉辦了一系列的跨院校的教研活動，推動「導向學習的評估」文化，提升高等教育工作者開發有效評估實務的意識，促進學生學習。

「導向學習的評估」理念

在 Boud (2000)、Gibbs 和 Simpson (2004)、Sadler (1989) 等學者的研究基礎上，LOAP 研究小組嘗試進一步推進，建議實施「導向學習的評估」，讓評估的焦點再度集中於學生，並鼓勵學生為自己的學習負責，才能真正提升學習素質。提升學習、測量學業成績和保持學科水平，是評估的三個主要作用。一個設計周全的評估計劃，應該可以同時發揮這些作用，而三者之間必須保持平衡、綜合、甚至大部分重疊（或完全重疊）。從學習的角度而言，「導向學習的評估」尤其強調以下三個要點：

- 為學生設計切合學習的評估——評估活動可以是學習活動；
- 讓學生參與評估自己的學習表現，而自評能力對學生日後的專業發展至關重要；
- 建立完整的反饋系統，讓學生能回應反饋意見——這裏最重要的意念是「前饋」(feedforward)，因為學生可以參考反饋，學會改進表現。

「導向學習的評估」各項關鍵元素之間的關係，可以用【圖 1.1】表示：

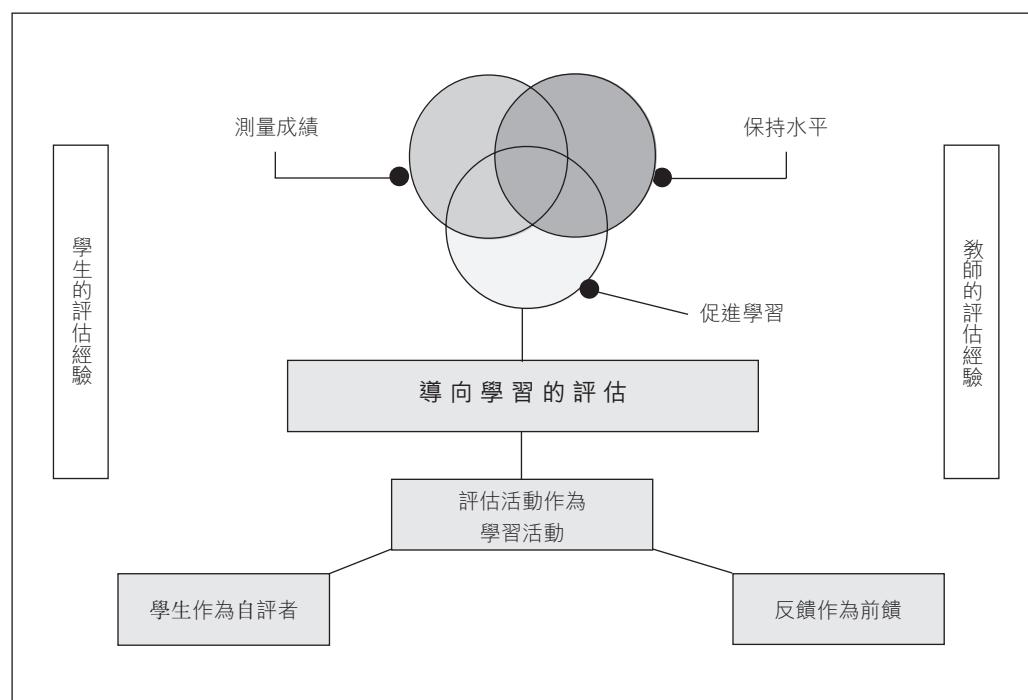


圖 1.1：「導向學習的評估」理念

4 導向學習的評估

評估活動作為學習活動

不少高等教育的研究（如 Knight & Yorke, 2003）均發現，學生嘗試學習的、學習的方式、學習時間和付出的努力，都深受他們對課程評估要求的觀感影響。因此，安排適當的評估將影響學生自訂的學習目標、學習焦點和學習時間分配。要評估能支援學習，評估課業應該要求學生付出足夠的學習時間及努力，而學生付出的時間和精力應該按照學習重點平均分配。此外，課業的設計還要顯示學生努力的素質和水平，讓學生投入有成果的學習活動，並注明高度期望學生達成的學習目標 (Gibbs & Simpson, 2004)。

學生作為自評者

在正規學習中，如果學生要成為有效的獨立學習者，必須發展自我評鑑的能力，才能監控及改善自己工作的素質 (Sadler, 1989)。在正規學習以外，他們同樣需要這種能力，才會成為有效的終身學習者和有效的專業從業員 (Boud, 2000)。因此，在任何導向學習的評估過程中，都不應忽略對學習者自評能力的培養。支援學生逐步發展自評能力，教師必須讓學生掌握要達成的工作要求，並且著學生把自己的表現水平與評估標準比較，最後付諸實踐，以實際行動縮短他們的表現與標準之間的距離。要言之，自評的關鍵作用在於讓學生認識學習要求的水平和標準，並由學生判斷自己工作的素質和改善方法，最終改進學習。

反饋作為「前饋」

雖然自評能力的發展可以推動終身學習，但一己的識見難免不足，所以通過同儕或師生互動的反饋也是學習的關鍵。對學生而言，分數的增加似乎是進步的最佳證據，然而，處理「導向學習的評估」成果，須著重質的反饋，才能讓學生了解自己的學習進程和不足之處，知道改進方向，發展判斷能力，並養成自我監督的習慣。

要推動學生改進學習，反饋應該欣賞學生的工作，讓學生認識自己做到了甚麼，有哪些地方仍須改進。此外，反饋又要連結學習目標和素質評鑑的標準，解釋評價背後的理由。解釋時還要注意採用學生能理解的形式，讓學生認真正視。無論反饋多麼簡單，學生往往需要澄清、對話和持續討論，才能確保正確無誤，從反饋中獲得啟發。當然，反饋最重要的意義是，讓學生按從反饋意見中知所改進。只要學生能顯示不斷改進的痕跡，改進的行動可以是修訂原來的作品，也可

以是下一個習作更出色的表現，因為反饋已經在改善工作及學習的過程中發揮了「前饋控制」的作用了。

有效的反饋應該適時適量，在特定的情境中（如課堂）還要多向（孫彤，2005）；趁學生對學習或工作印象猶新，反饋的迴響自然更大。除了數量外，反饋的素質更加重要：反饋的焦點應該在學習，而不在分數或學生本身；反饋應該連結習作的目的和標準；無論反饋多麼複雜或精要，都應該務求學生能夠明白、關心、並按反饋改進工作或學習 (Gibbs & Simpson, 2004)。近年，不少教師和學者還注意到情感態度與反饋成效的關係，尤其是即時當面的口頭反饋，學習者的心靈因素決不能忽視，反饋的措辭、語氣和反饋情境的氣氛等，都直接影響反饋的效果（李俊芬，2005）。

評估經驗的影響

由於「導向學習的評估」牽涉參與者的教育理念，已超越技術層面的問題。如果彼此存在經驗差異，對評估、學習及教導的信念分歧，教師和學生可以對評估有截然不同的理解和體驗。因此，設計和規劃「導向學習的評估」時，還要同時考慮師生雙方的評估經驗，儘量協助他們發展以評估提升學習的能力。

結語

從「對學習的評估」到「導向學習的評估」的最終目的，無非是讓學習者能更有效達成學習，在「對學習的評估」還沒有完全取締前，能儘量在兩者之間取得平衡，無論對學生或教師，都應該有較大裨益。在面對嶄新的學習任務時，學生必須不斷嘗試、蒐集資料、有系統地假設、並評估自己的學習進度，才能了解評估的內容及評估期望的表現。對教師而言，「導向學習的評估」理念牽涉整個教學過程的協調，即使是非正式評估如提問和觀察，教師和學生都得有意識地實踐。隨著評估活動頻次的上升，教師反饋工作增加似乎無可避免。然而，教師也必須釐清反饋的最終目的在讓學生了解學習成效，以及下一步應如何改進，不在評分，切勿把反饋演變為沒完沒了的批改；否則促進的只是機械的工作而不是學習，就完全叛離「以評估促進學習」的原意了。

參考資料

中文著作

李坤崇 (1999)：《多元化教學評量》，台北，心理出版社。

李坤崇 (2006)：《教學評量》，台北，心理出版社。

李俊芬 (2005)：以情感態度為異向的教師反饋策略模式初探，《教育理論與實踐》25，頁 43–46。

香港教育署課程發展議會 (2002)《基礎教育課程指引——各盡所能•發揮所長》，香港：香港課程發展議會。

孫彤 (2005)：教師對學生學習的反饋，《遼寧高職學報》7，頁 153–156。

翻譯著作

柏士俊 Biggs, J. (1999)：評估新取向的根本假設，輯於 Stimpson, P., Morris, P., Fung, Y., Carr, R. 主編，黃婉儀、陳瑞堅、馮施鈺珩主譯《課程、學習與評估：香港的經驗》（香港：香港公開大學出版社），頁 188–226。

Black, P. & Wiliam, D. 著，香港教育署課程發展處譯《暗箱內探：透過課堂評估提高學習水平》，<http://cd.ed.gov.hk/assessment/document/BK1.pdf>，瀏覽日期：24–06–2006。

Linn, R. L & Gronlund, N.E. 著，國家基礎教育課程改革「促進教師發展與學生成長的評價研究」項目組譯 (2003)：《教學中的測驗與評價》，北京：中國輕工業出版社。

Weber, E. 著，國家基礎教育課程改革「促進教師發展與學生成長的評價研究」項目組譯 (2003)：《有效的學生評價》，北京：中國輕工業出版社。

英文著作

Black, P. & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment, *Phi Delta Kappan*, 80 (2), 139–147.

Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, 22, 2, 151–167.

Elton, L. & Johnson, B. (2002). *Assessment in universities: A critical review of research*. LTSN.

Gibbs, G. & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education*, 1, 3–31.

Knight, P. & Yorke, M. (2003). *Assessment, learning and employability*. Maidenhead, England: Society for Research into Higher Education & Open University Press.

Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119–144.

二

以評估促進學習

張淑賢

評估的發展

評估的發展在東、西方已有一段頗長歷史。相對東方而言，評估在西方起步稍晚，但卻為現今的評估制度帶來相當大的影響。在西方，評估可以追溯到十八世紀，當時流行委任制度。選拔以引見形式，甄選的決定往往不視乎履歷或臨場表現，而在乎求職者與推薦人的知名度。十九世紀初期，由於工業資本主義經濟迅速發展，社會需要大量受過正統訓練的員工，但傳統的提名和委任方式都不能滿足這方面的需要。有見及此，有人開始以標準化的考試制度吸納來自不同社會階層和文化背景的人才，所以考試制度被視為對考生既公平而對社會又有利的選拔方法。Sutherland (1996, p. 16) 報告指出，「委任制度原先是政府部門及各行各業的招聘手法，但是後來卻逐漸被視為腐敗的化身，標準化考試應運而生，以針對此流弊」。在十九世紀，隨著學校教育的興起而英才教育廣受關注，正規的筆試制度慢慢融入英美等西方國家的教育體系。英國在 1828 年首次印製考卷，自此以後，考試制度備受肯定，更成為審定專業資格的重要關卡。美國等西方國家也大量採用測驗和考試，應用範圍甚廣，包括驗證、升級、頒發文憑、分班、學區及學校的資助分配等。這類選拔式的測試，對個人發展及前途影響深遠，因而被稱為「高風險評估」。

在東方，運用同類高風險評估屢見不鮮，中國、日本、新加坡、台灣、馬來西亞和韓國等，均長時期採用這種評估模式，明顯的例子有中國各朝代的公僕選拔。為求功名利祿，書生報考高風險試，逐級闖關，先是區試，繼而省試，最後國試。甄選過程嚴峻，通常每年只有寥寥數位考生得以取錄，獲頒一官半職。眾多考生窮畢生的努力，作無數次嘗試都不能如願，正是數十載寒窗苦讀仍一事無成、仕途落

8 導向學習的評估

泊、潦倒一生。今天國內的情況與過往沒有兩樣，教育制度依然源用高風險評估模式，教育階梯關卡重重。觀察今日的中國，可見升高中考試及入大學的高考，都是升學和雇主作專業決策的主要門檻。宏觀現今社會，幾乎所有文明國家都在廣泛採用這類考試進行評估，高風險評估其實含有莫大隱憂，不容忽視。

高風險評估的存在問題

標準化考試的原意是追求公平，換而言之，考試具有監察功能，是作為標準保證的工具。然而，經過百年的試用後，考試保證公平的觀點卻在美國受到考驗。彈劾的意見認為，測驗和考試並非對所有人都公平，因為測驗和考試往往存在偏見。學者如 Lambert 和 Lines (2000)，質疑考試成績的可信程度，事實上，從擬題到得出成績的過程中都難免會出亂子——例如：批改謬誤，錯誤評分，以及考生的健康情況也可能構成不公允的情況。值得注意的是，考試本身只是學生成就的其中一部分。標準考試只針對學生的認知水平，而不確認他們其他的成就。事實上，並不是所有學生都能適應公開考試，即使發奮學習，他們也不能在公開考試中獲取良好成績。

考試引起最大的問題是，學生經常要學習如何應付考試，教學方面側重幫助學生考試合格及提高學校的知名度。因此，這類教學經常替學生操練，要求他們死記硬背知識。教學手法不外乎頻密測驗，每星期、每個月，每學期，學校都不停地安排對準測試的演練，給學生帶來極大壓力，但卻對學習全無裨益。偏重應試的教與學，不但窒礙學生發揮創意，而且減少他們真正體會學習的機會。國內學者程方平 (2002) 指出，只靠死記硬背、被動完成的考試會壓抑學生原可在思維過程中表現出來的積極性、自主性和創造性，而考試嚴格的標準，又使學生只為考取高分數而不停操練及運用大量時間背記標準答案，在此考試制度下成為模式化的機械人。在這樣的評估制度下，學習變成只為應付考試所需，缺乏全人發展的機會。Stobart 和 Gipps (1997) 也指出，死記硬背的效果不會長久，目前還沒有足夠證據顯示強制測驗在缺乏教學的情況下可以提升學習的標準，而學習如何應付測驗的教學通常是較為偏狹的。Berry (2005a, p. 225) 補充，認為評估制度的用途廣泛，除了衡量學生成績外，還可以促進學習。

評估新視角

評估可以促進學習的概念日漸受重視。近年來，越來越多學者肯定評估對學習的影響。學者如 Airasian (2005); Black 和 Wiliam (1998); Lee (2001); Chappuis 和 Stiggins

(2002) 及 Guskey (2003) 等認為，如果評估運用得宜，可以提升學習，肯定評估與學習息息相關。評估可以量度學生的學習，也可以幫助學生獲取學習的成果。Brookhart (2005) 指出，促進學習的評估包括兩大要素——「實情」和「後續」。所謂「實情」就是從多方面搜集資料以得知學生學習的情況，包括學生各方面的表現；而「後續」是指利用那些資料來提升教與學。透過日常的評估活動如：觀察、考核等，教師可以得知學生的知識、技巧、價值觀和態度的水平。所得資料可以用作分析學生學習上的困難，為學生作出批評和建議，讓他們知道如何改進學習；資料又可用以調適課程，使之更切合學生的需要和能力，達至教學相長。簡茂發和單文經 (1995) 指出，教師從評估中，可以了解學生學習上的進程和困難，並且向學生提供進一步學習的信息，從而改進教學和修正課程設計。這項新的理念，有利確定學生在學習上的各種需要後改進，所以稱為「促進學習的評估」。

課堂通常是舉行評估活動的地方。Borich 和 Tombaris (2004, p. 1) 強調，「課堂就是學習的地方，評估定能促進學習」。課堂的評估活動可以非常簡單，例如：課堂上點名，以得知誰到了誰沒到；或用提問及反饋作為鼓勵或教育用途。要評估發揮學習的作用，課堂評估活動的設計，應以學習為本，教師應多思考用怎樣的評估方法才能帶動學習。

「促進學習的評估」在香港的發展

香港多年來採用以考試為主的教育制度，以三層架構成作為教育的基本模式，學生在接受學校教育的過程中，會經歷三次大型淘汰和分級的考試。在完成六年小學教育後，學生會根據學習能力被分配到不同等級的中學。完成五年中學教育後，學生會參與中學會考，其中約有三成學生能繼續升讀兩年的中學預科課程。在其後的中學高級程度會考中，只有四分之一的學生能夠接受專上教育。由於三次淘汰考試都對學生非常重要，香港的學校一直都以測驗和考試為教與學的目的。在社會過份強調考試的重要性以及在家長認為「多功課、多測驗、多考試就是好」的心態下，學校並不太重視教學，反而只盲目跟從考試大綱，造成課程的發展障礙 (Adamson & Morris, 1998; Biggs, 1998; Chan, Kennedy, Yu & Fok, 2006; Hong Kong Government Secretariat, 1981)。因此，近年香港政府不斷提倡教育改革，例如九十年代，政府曾提倡以目標為本的課程 (TOC) (Education Department, 1994); 2000 年又推出課程改革，並把評估與學習聯繫。官方文件(課程發展議會, 2002) 重申，評估不應被視作學習終結階段的活動，不能只為求證達標與否，而應在教與學的過程中幫助學生學習。

這個三層的教育架構曾經修訂，或許將來會再作修改。首先，學能測驗和完成小學的中學入學試已經被中學派位機制取代，免除了不必要的操練。現時的派位依據是學生就讀小五和小六時的校內評估成績，政府會以學校參與政府主辦的中文口試和數學能向測驗的整體水平表現調整學生的評估分數；其次，小三、小六和中三同學要參與基本能力評估測驗，以便教師能確定學生的學習水平，進一步輔導學生學習。最後，預計 2010 年，政府將會取消兩項高風險的公開考試，由另一項新的香港中學教育文憑課程考試取代。

政府頗有決心作全面的改革，希望能夠營造一種氣氛，讓「促進學習的評估」開花結果。課堂已經成為一個發揮這種特色的重要地方，然而，評估改革的進度仍未能大力推動課堂評估的發展。香港在目前依然有很多課堂，繼續採用評估來量度學習成績，並沒有以它作為輔導學習的工具 (Berry, 2005a)。IBM Business Consulting Service 公司接受香港政府的委託，審視香港考試局的情況。公司於 2003 年發表了一份報告，批評香港雖然經過三年的課程改革，但仍然非常依賴考試，沒有好好地利用評估來增進學生的學習。報告指出「香港一直在努力抗衡牢固的死記硬背現象，而這現象的來源正是出於高風險測驗和考試」。近來，香港有一些研究，結果也跟 IBM 的報告吻合，同樣指出評估制度在香港尚未被廣泛應用於改進學習上。Pong 和 Chow (2002) 認為，「以考試為本的文化在香港根深蒂固」。很多時，教師的首要考慮便是為學生提供應試策略，他們要從歷屆試題中找出最常重複考問的問題，然後操練學生機械式地學習標準答案 (Morris, 1985)。Bryant, Timmins, Berry, Fok 和 Ngan (Berry, 2003a) 等人的研究發現，教師並沒有採用評估來增進學生的學習。Berry 研究的結論 (Berry, 2003a) 也強調了上述的看法，注意到「香港有很多學校在評估方面頗為傳統，沒有充分利用評估來促進學習」。從上述可見，香港的評估理念普遍認為教師的任務是要知道學生接收這些教學內容的準確程度。這一來，教師在評估的角色主要是檢查成績，卻沒有以評估輔助學習。如要評估發揮促進學習的作用，首先要從教師方面入手，讓他們明白教師在「促進學習的評估」中扮演的角色，並且結合課堂實踐；否則評估只純粹是理論概念，不能改進學生的學習，也全無實用價值。

教師的角色

評估能否寓於教學，往往取決於教師和學生的課堂行為。究竟教師在評估上應該扮演甚麼角色，目前還沒有一致的看法。Wyatt-Smith 和 Ludwig (1998) 把教師的評估角色分為四類——舉行期測、輔助有關當局抽樣調查學生表現、製作學生進展圖階和校本評估。期測和抽樣調查要求學生進行一般的測驗或評估，教師在當

中的角色較為被動；而採用進展圖階和校本評估時，教師則可以有較大的自由度，因為後兩者能協助教師明白學生的發展，並輔助個別學生在不同階段成長。

Neesom (2000, p. 23–25) 指出一系列有關教師於促進學習的評估上的角色行動，包括讓學生多參與評估、示範優質作業、提供反饋和自我評估等，其中要求學生參與評估及學習的活動尤其重要。教師要與學生交流學習的不同目的，包括知識、概念、技巧和應用等各方面，幫助他們連結相關的知識，鞏固已有知識和澄清學習需要。教師的任務不再是傳送知識，而是督導、促進和輔助學生建構知識。因此，在鼓勵學生探究時，應注意如何採取多元化的教學及評估策略，例如加入專題研習和公開辯論等開放式的評估，以激發學生主動學習的動機，提供師生和學生之間互動的評估模式及學習方法（張世忠，2000）。

示範優質作業有助學生明白習作的標準要求。教師在這方面的角色包括：選取及展示其他學生的優異作品讓學生學習；請某學生訴說工作進程，以便診斷他們的偏誤，然後請學生建議改進方法。教師也可讓學生認識何謂欠佳的習作，提醒他們引以為鑑，避免犯同樣的毛病。

評分是教師工作的重要一環，但只評分而不反饋對學習幫助不大。Berry (2006b) 指出，「反饋有多個顯著的作用。視乎使用方法，反饋可以是激勵性或學習性的。」李坤崇 (1999, 2006) 認為教師應讓學生明白分數的含意，在解釋分數時，宜激發學生學習動機、強化學生的努力歷程和肯定學生的付出，不應一味苛責學生未能達到完美、不夠努力或細心。在一個促進學習的評估環境，反饋能起提升學習和輔導學生承擔學習責任的作用。反饋的可行形式包括：教師可以用引導性問題診斷學生的學習情況；分析學生完成的習作，明白學生能做到或做不到的原因；對學生的習作表示嘉許或批評，告訴學生做了或還沒有做的地方和指出更好或其他做法。Sadler (1989) 確定以下三項條件，認為可以讓學生從反饋中的得到啟發——學生必須知道：

- 甚麼是良好的表現（具備學習目標或要求標準的概念）；
- 目前的表現與良好表現的關係（學生要懂得把目前的表現與良好表現對比）；
- 怎樣才可以填補兩者之間的差距。

一直以來，自我評估都被視為幫助學習者掌握自己學習的最有效方法。能夠掌握自己的學習是重要的，因為辦得到的人，通常在學習上都會比較成功 (Berry 2006a)。無論如何，只要給予機會，學習者都能尋求和獲得新的技巧、知識和理解。教師安排自我評估時，應鼓勵學生發展自我評估的渴求和能力，掌握自己的學習 (Assessment Reform Group, 2002)。再者，Clarke (2001) 指出，「孩童的自尊

可以透過自我評估得到改善」。自我評估的形式包括：讓學生記錄自己的習作進度和審查情況；給學生時間反映，講述他們的學習情況；幫助學生明白他們的學習成績及下一步可以怎樣改善；讓學生建議如何改進；讓學生商討改善情況；以及著學生評估自己和別人的習作，並提供反饋等(Neesom, 2000)。為達到更好的效果，教師還可以把教師的反饋融入學生的自我和同儕評估。馬克·麥卡爾平Mark MacAlpine (1999)在教授畢業班(香港理工大學電機工程學系)時，就由學生講解課程內的特定內容，其他同學則按利克特量表，以同儕評估的方法匯報該生在多個範疇(即是知識、身體語言、聲調及整體效果)的表現，最後把所得分數納入期終成績。結果發現，這樣的同儕評核對大學最後一年的教與學，既有用又具建設性，還有助培養學生的自學能力，令他們判斷時更有信心。

此外，教師還有其他角色，例如監督和分析學生的習作，以改進教學。透過監督學生的學習進度，教師可以注意到學生的知識應用情況、學習態度、以及能否達到學習目標。他們可以透過課堂上的習作、默書、遊戲、觀察、設計習作和提問，檢查學生的學習進度。透過這類評估，讓學生培養判斷質量的能力，知道下一步應該怎樣做。例如，在專題研習中，教師可以用富挑戰性的問題或難題來引起學習的興趣，讓學生藉著多樣化的活動建構知識，並把知識、能力、價值觀與態度結合起來(課程發展議會，2002)。

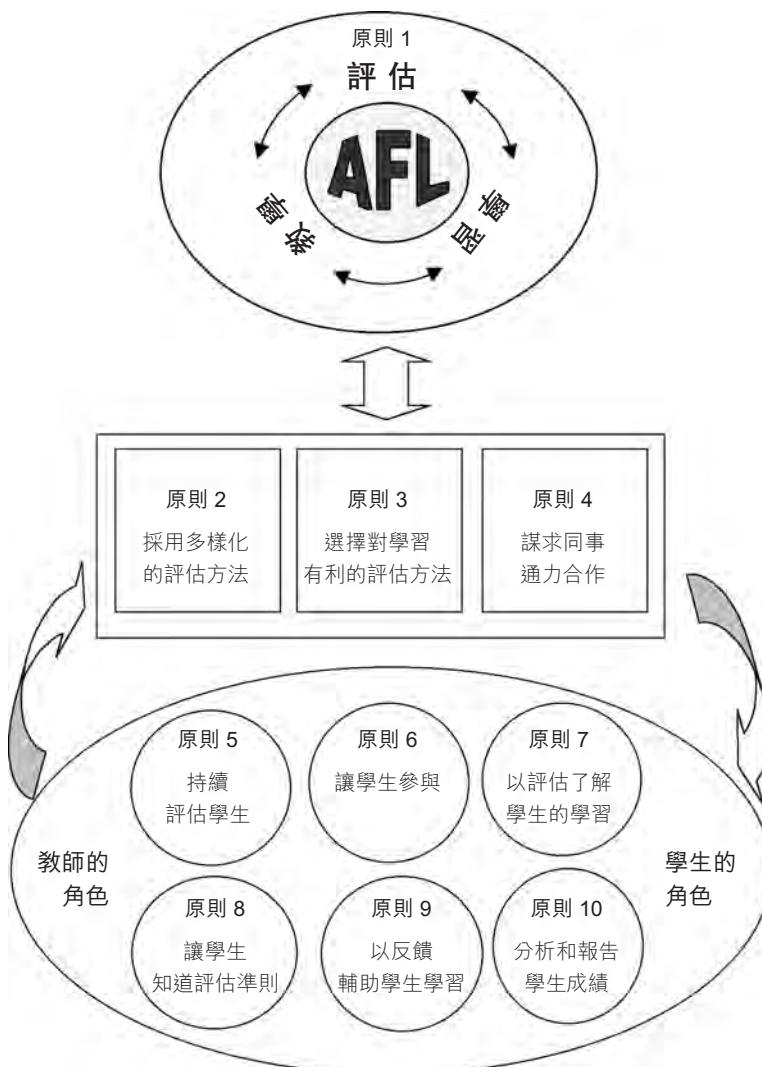
教師固然能夠運用評估來診斷學生的強項與弱點，但現今的課堂卻存在學生差異的問題，例如學習動機、學習風格、能力、文化背景、社會經濟狀況和以往的學習經驗等。當學生的學習需要日漸多元化時，如何設法滿足這些學習需要就成為一項新挑戰。Berry (2006c)指出，評估也可隨之而多元化發展，教師可以運用不同的評估策略來找出每名學生和全班的特性，以便在課堂上分組、教學、激勵和管理。最後，教師可以設計不同的評估計劃，以照顧學生的個別需要。

學生的角色

教師在以評估促進學習上的角色不庸置疑，但學生也應明白，他們在評估過程中也可以扮演一個重要的角色。學生要參與評估，採用評估資料來訂定學習目標、作改進學習的決策、理解優質作品的要求、自我評估、尋求同儕和教師的反饋，以及了解自己的學習狀況和達成學習目標的進度(Stiggins, 2005)。柏士俊 Biggs (1999)認為，自我評估能夠幫助學生有效掌握學習內容，他認為採用這個方法，需要有一份標準答案輔助。但是，有些評估習作涉及大量的主觀判斷，很難預定一份標準答案供學生自評。其實「標準」答案也有商榷餘地。究竟何謂標準？有沒有絕對的標準？這倒不如讓學生制定一套評分準則，或至少參與詮釋準則的討

論。這樣的好處是，能讓學生增強對這些準則的擁有感，而通過參與討論，他們會把標準內化，並更明白優質習作的要求。除此以外，學生還可以參與同儕評估及反思活動。值得一提的是，提供反饋並不是教師的專利，學生也肯定也可以作出貢獻。參考評估的示例說明，學生可以評估自己和同儕的習作，通過自評和同儕互評，監督自己和同儕的習作，而讓學生參與評估實在是改善教與學的有效方法（李坤崇，1999, 2006; Berry, 2005b）。

要言之，在促進學習的評估環境裏，教師和學生的角色可以按圖2.1的概念架構的十項原則調節。



十項促進學習的評估原則

雖然促進學習的評估概念為評估提供了一個宏觀的方向，但以此為底線仍然有很多需要思考的地方，我們必須詳細了解評估的各項原則，才可以化概念為行動。綜合前文討論及下列文獻，包括 Angelo & Cross (1997); Berry (2003b); Carless (2005); Goos & Moni (2001); Lynch (2001); James et al. (2002); Curriculum Development Council (2002)；謝錫金等(2003)；歐滄和(2002)；李乙明(2004)；黃銘惇和張慧芝譯(2000)、基礎教育項目組(2003)；林寶山(1998)；柏士俊Biggs (1999) 及馬克·麥卡爾平 MacAlpine (1999) 等，本文歸納出以下十項體現「以評估促進學習」的原則。

1. 評估連結教與學

在一個標準的課堂裏，評估是日常的活動。評估必須與課程的目的和教學一致，採用的評估方法也要能夠反映不同主題和課程宗旨。基本上，評估的項目要與教學的內容有關，而評估應該反映學生學習上需要的技巧。教師可以根據預先釐定的準則，以不同的途徑，觀察、搜集、記錄、分析學生的表現，然後診斷和估計他們的學習情況及能力，一方面給與有效的反饋，另一方面調整之前的教學，使教與學雙方在這個互動而循環的過程中，不斷自我完善，以求表現更加理想。教師可以利用獲得的資料，了解學生的學習進度，然後根據新的資料調整教與學活動，滿足學生的需要。

2. 採用多樣化的評估方法

評估方法多樣化，各有不同的形式、深度和寬度，有實作的，也有模擬的，可以反映學習不同的情況，評估不同的學習結果；尤其當評估在本質上是有根據時，多樣化的評估方法更可以保持學生的興趣。此外，採用多樣化性的評估方法，還可以令學生明白自己不同的學習素質，起拔尖補底的作用。

3. 選擇對學習有利的評估方法

低效的評估方法阻礙或減少學習機會，所以應該儘量少用。所謂「適當」的方法，就是能夠帶給學習者大量學習機會的方法。採用多樣化的評估策略，例如學習歷程檔案、觀察、實驗、習作、模擬、面談、表演、和演示概念構圖和字詞聯結等等，都可以加深學生多方面的學習。

4. 謂求同事通力合作

站在教育的立場，學生的學習要透過多方合作才能達成，這情況同樣適用於

評估上。合作的行動能支援有關評估的討論，加強對學習的了解。這種專業上的合作對評價標準的制定至關重要，它讓教師能共同確立多種評價標準，包括多種技能和內容，有助建立多套適用於不同層次的評估。

5. 持續評估學生

對學習進行循序漸進的評估，對學生甚有幫助。教師可以在學期裏給學生一系列合適的評估小任務，而教師也不用只根據學生的一次表現就判斷他們的整體成績。評估的工作是，給予學生運用所學，主動地展示和證明學習表現的機會，教師可以一邊觀察、判斷，一邊鼓勵、引導和指正。然而，這些評估任務的數量必須平衡，以免學生和教師都累透，最理想就是把評估結合教學活動。

6. 讓學生參與

評估不應由教師單獨進行。學生可以透過評估自己和同儕來促進學習，他們是評估的當局者，在評估和學習上並非局外人。他們要評審自己的習作、監督自己的進度、替自己制定目標和向別人展示習作。結構良好的自我及同儕評估，均能提供寶貴學習經驗，影響深遠。評估過程包括三個程序（即是：制定評估準則、確定習作是否符合準則、及判斷顯證與準則之間的吻合程度），教師可以讓學生參與其中一項或全部過程。

7. 以評估了解學生的學習

評估應該是信息化的，能夠反映學習了甚麼和還有甚麼需要繼續學習；而採用的評估方法則必須能夠評審廣泛的學習結果。本文並不鼓勵採用考試之類的單一評估方法，因為教學評估的主要目的在於改進學生的學習行為，而不是為學生評定期末成績。教師宜儘量採用多項評估方法來評審學生的表現，例如形成性評估和總結性評估一併使用，以前者提供及時和定期的反饋、後者展示最後的學習結果。

8. 讓學生知道評估準則

學生需要清楚明白評估任務對他們的要求。每項任務都必須附有評估表，要讓學生及評核者清晰掌握。教師應該在教學之前就決定成績的評估計劃，而此計劃在授課前就應該先讓學生充分了解。其中，評估的準則必須詳細、透明和平等。教師可以讓學生參與討論，甚至制定評估準則。

9. 以反饋輔助學生學習

反饋在學習的過程中是重要的，能夠給學生提供合時和廣泛的資料，幫助他

們達成學習目標。形成性評估對教學過程中監督學習進步有很大的效用，因為其目的在於提供師生有關學習結果的反饋，以促進學生學習、改進課程內容安排、探索更有效的教學方式。形成性評估的基本精神在於透過詳細的行為目標，把教學與評估緊密配合，並使之形成循環互動的關係。除此以外，學生要經常注意自己的成績和需要改善的地方，以備將來發展。他們應該有機會實踐教師、自己和同儕的建議。

10. 分析和報告學生成績

對學生在評估任務上的表現進行系統分析，可以幫助界定課程需要改進的地方。這樣做，教師在教學上會得到提示，學生在學習上最終也會受益。報告學生的成績時，教師可以採用敘述性的報告，而不採用打分數的形式或數量化的表達。其中敘述性的報告應包括有關努力、態度、舉止以及成就的陳述。這樣的好處是敘述性的描述能在減少學生家長拿學生做比較的同時，使教師有更好的機會向家長傳達學生的學習信息。

結語

評估的基本原則是把評估與學習緊密聯繫。它應該用來促進、引導和加強學習。在促進學習的評估中，讓學生參與評估活動是十分重要的，因為他們是學習的主角。教學、學習和評估三方面需要互相結合，才能提高學生的整體水平。

參考書目

中文著作

- 程方平 (2002)：《中國教育問題報告—入世背景中國教育的現行問題和基本對策》，北京：中國社會科學出版社。
- 簡茂發、單文經 (1995)，制式與非制式評量方法評析，輯於中國教育學會主編《教育評鑑》（臺灣：師大書苑），頁 61–102。
- 教育處 (2002)：校內成績評估概論，載於《校內成績評估指南》（香港：教育處），頁 10–16。
- 課程發展議會 (2002)：從四個關鍵項目學會學習：3C 專題研習，輯於課程發展議會主編《基礎教育課程指引——各盡所能・發揮所長》，第三冊（香港：課程發展議會祕書處）。
- 李坤崇 (1999)：《多元化教學評量》，台北：心理出版社股份有限公司。

- 李坤崇 (2006)：《教學評量》，台北：心理出版社股份有限公司。
- 林寶山 (1998)：《教學原理與技巧》，台北：五南圖書出版有限公司。
- 歐滄和 (2002)：《教育測驗與評量》，台北：心理出版社股份有限公司。
- 謝錫金、祁永華、譚寶芝、岑紹基、關秀娥 (2003)：《專題研習與評量》，香港：香港大學出版社。
- 張世忠 (2000)：《建構教學——理論與應用》，台北：五南圖書出版有限公司。

翻譯本

- 柏士俊 Biggs, J. (1999)：評估新取向的根本假設，輯於 Stimpson, P., Morris, P., Fung Y., Carr, R. 主編，黃婉儀、陳瑞堅、馮施鈺珩主譯《課程、學習與評估：香港的經驗》（香港：香港公開大學出版社），頁 188–226。
- 馬克·麥卡爾平 MacApline, M. (1999)：改進並鼓勵運用朋輩評估方法以評估學學匯報，輯於 Stimpson, P., Morris, P., Fung Y., Carr, R. 主編，黃婉儀、陳瑞堅、馮施鈺珩主譯，《課程、學習與評估：香港的經驗》（香港：香港公開大學出版社），頁 135–161。
- Pratt, D. 著，黃銘惇、張慧芝譯 (2000)：《課程設計：教育專業手冊 (Curriculum Planning: A Handbook for Professionals)》，台北：桂冠圖書股份有限公司。
- Weber, E. 著，國家基礎教育課程改革「促進教師發展與學生成長的評價研究」項目組（基礎教育項目組）譯 (2003)：《有效的學生評價 (Student Assessment That Works: A Practical Approach)》，北京：中國輕工業出版社（《新課程與教評價改革譯叢》）。
- White, R., & Gunstone, R. 著，李乙明主譯 (2004)：《學習理解之多元評量 (Probing Understanding)》，台北：洪葉文化事業有限公司。

英文著作

- Airasian, P. W. (2005). *Classroom Assessment: Concepts and Application (5th ed.)*. Boston, MA: McGraw-Hill.
- Adamson, B. & Morris, P. (1998). Primary schooling in Hong Kong. In J. Moyles & L. Hargreaves (eds.), *The Primary Curriculum: Learning from International Perspectives* (pp. 181–204). London: Routledge.
- Angelo, T. A. and Cross, K. P. (1997). *Classroom Assessment Techniques: A Handbook for College Teachers (2nd edition)*. San Francisco: Jossey-Bass.
- Assessment Reform Group (2002). *Assessment for Learning: Ten Principles*. Retrieved July 12, 2005, from <http://www.assessment-reform-group.org.uk>.
- Berry, R. (2003a). Alternative Assessment and Assessment for Learning. *Conference Proceedings of Societies' Goals and Assessment*. Manchester, UK: International Association for Educational Assessment (IAEA). Retrieved July 4, 2006, from <http://www.aqa.org.uk/support/iaeapapers/berry.pdf>.
- Berry, R. (2003b). Assessment methods that facilitate learning. *Conference Proceedings of Reform and Initiatives in Teaching and Learning*. Macau: University of Macau, pp. 236–241.

- Berry, R. (2005a). Entwining feedback, self, and peer assessment. *Academic Exchange Quarterly*. Fall issue.
- Berry, R. (2005b). Teachers' perceptions of their roles and their students' roles in the formative assessment process. *Conference Proceedings of Australian Association for Research in Education (AARE) 2004 Doing the Public Good*. Melbourne, Australia. Retrieved July 4, 2006, from <http://www.aare.edu.au/04pap/ber04978.pdf>.
- Berry, R. (2006a). Activating learners using the Learner Autonomy Approach: An action research on the relevance of teaching to classroom practice. *Curriculum Perspectives*, September, Vol. 26, No. 3.
- Berry, R. (2006b). Functional significance of feedback in learning. *Conference Proceedings of Australian Association for Research in Education (AARE) 2005*. Parramatta, Australia. Retrieved July 4, 2006, from <http://www.aare.edu.au/05pap/ber05533.pdf>.
- Berry, R. (2006c). Teachers' assessment practices for classroom diversity. *Conference Proceedings of International Association for Educational Assessment (IAEA)*, Singapore. Retrieved July 4, 2006, from <http://www.iaeaweb.org/conference/download/papers/Teachers%27%20assessment%20practices%20for%20classroom%20diversity.pdf>.
- Biggs, J. (1996). *Testing: To Educate or To Select? Education in Hong Kong at the Crossroads*. Hong Kong: Hong Kong Educational Publishing Co.
- Biggs, J. (1998). The assessment scene of Hong Kong. In P. Stimpson & P. Morris (eds.), *Curriculum and Assessment for Hong Kong: Two Components, One System* (pp. 315–324). Hong Kong: The Open University of Hong Kong.
- Black, P. & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80 (2), 139–147.
- Borich, G. D. & Tombari, M. L. (2004). *Educational Assessment for the Elementary and Middle School Classroom*. Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.
- Brookhart, S. (2005). Research on formative classroom assessment. Paper presented at the *American Educational Research Association (AERA) 2005 Annual Conference*. Montreal, Canada.
- Chappuis, S. & Stiggins, R. J. (2002). Classroom assessment for learning. *Educational Leadership*, 60 (1), 40–43.
- Carless, D. (2005). Prospects for the implementation of assessment for learning. *Assessment in Education*, 12 (1), 39–54.
- Chan, J. K. S., Kennedy, K. J., Yu, F. W. M. & Fok, P. K. (2006). Assessment policy in Hong Kong: Implementation issues for new forms of assessment. Paper presented at the *International Association for Education Assessment (IAEA) 2006 Annual Conference*, Singapore.
- Clarke, S. (2001). *Unlocking Formative Assessment: Practical Strategies for Enhancing Pupils' Learning in the Primary Classroom*. London: Hodder & Stoughton.
- The Curriculum Development Council (CDC) (2002). *Basic Education Curriculum Guide: Building on Strengths (Primary 1 – Secondary 3)*. Hong Kong: the Curriculum Development Council, Hong Kong Special Administration Region. Retrieved July 4, 2006, from http://cdemb.gov.hk/basic_guide/beguide_enhtml/.

- Education Department (1994). *General Introduction to Target Oriented Curriculum*. Hong Kong: Government Printer.
- Goos, M., & Moni, K. (2001). Modelling professional practice: A collaborative approach to developing criteria and standards-based assessment in pre-service teacher education courses. *Assessment & Evaluation in Higher Education*, 26 (1), 73–88.
- Guskey, T. R. (2003). How classroom assessments improve learning. *Educational Leadership*, 6 (5), 6–11.
- Hong Kong Government Secretariat (1981). *Overall Review of the Hong Kong Education System: The Hong Kong Education System*. Hong Kong: The Government Printer.
- IBM (2003). *Strategic Review of Hong Kong Examinations and Assessment Authority*. Hong Kong: Hong Kong SAR, China.
- James, R., McInnis, C., and Devlin, M. (2002). Assessing Learning. Retrieved July 4, 2006, from Australian Universities Web site: <http://www.cshe.unimelb.edu.au/assessinglearning>.
- Lambert, D. and Lines, D. (2000). *Understanding assessment: Purposes, perceptions, practice*. London: Routledge Falmer.
- Lee, C. (2001). Using assessment for effective learning. *Mathematics Teaching*, 175, 40–43.
- Lynch, B. K. (2001). Rethinking assessment from a critical perspective. *Language Testing*, 18 (4), 351–372.
- MacAlpine, M. (1999). Improving and encouraging peer assessment of student presentations. *Assessment and Evaluation in Higher Education*, 24 (1), 15–25.
- Morris, P. (1985). Teachers' perceptions of the barriers to the implementation of a pedagogic innovation: A South East Asian case study', *International Review of Education*, 31: 3–18.
- Neesom, A. (2000). *Report on Teachers' Perception of Formative Assessment*. The Qualifications and Curriculum Authority (QCA) Assessment for Learning Project. London: QCA.
- Pong, W. T. & Chow, J. C. S. (2002). On the pedagogy of examinations in Hong Kong. *Teaching and Teacher Education*, 18 (2), 139–149.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18 (2), 119–144.
- Stiggins, R. J. (2005). *Student-involved Assessment FOR Learning* (4th ed.). Upper Saddle River, NJ: Pearson Merrill Prentice Hall.
- Stobart, G. & Gipps, C. (1997). *Assessment: A Teacher's Guide to the Issues* (3rd ed.). London: Hodder & Stoughton Educational.
- Sutherland, G. (1996). Assessment: Some Historical Perspectives. In Goldstein, H. and Lewis, T. (eds.) *Assessment: Problems, Developments and Statistical Issues*. New York: John Wiley.
- Wyatt-Smith, C. and Ludwig, C. (1998). Teachers' roles in large-scale literacy assessment. *Curriculum Perspectives*, 18 (3), 1–14.

詞匯對照表

ability 能力	anchor group design 校準團體設計
ability grouping 能力分組，按學生能力分組	anchor test 校準測驗
ability parameter 能力參數	anchor test design 校準測驗設計
ability test 能力測驗	anecdotal record 軼事記錄
abuse 不正確用法	anova 方差分析
academic achievement 學業成就，學業成績	application 應用
accountability (教學) 承責	approach of assessment 評估方法
achievement 成就，成績	approximation method 概略法
achievement test 學業成就測驗，學業成績測驗	aptitude test 能向測驗；性向測驗
adaptive testing 適性測驗	arithmetic mean 算術平均數
additive effect 可加性	assessment 評估，評核
adequacy 適合性	assessment for learning 促進學習的評估
adjustment 校正，另類	assessment of learning 對學習的評估
alternate forms reliability 待選試卷信度	assessment portfolio 樣本匣評估法
alternate response item 二者選一反應題	assessment program 評估方案
alternative 待選項	assessment tools and procedure 評估工具與程序
alternative assessment 另類評量	attainment 成績
alternative assessment programme 替代評估方案	attainment test 學科測驗；學業成績測驗
alternative form 待選試卷	attitude 態度
alternative response item 對立反應題	attitude scale 態度量表
alternative testing 交替測驗	attribute data 計數資料
analysis 分析	aural assessment 聽覺評估
analysis of variance 變異分析	authenticity 原真性
analytical marking 分析等級法	authentic assessment 真實評量，真正的評估
analytic rating 分析等級，解析評分，分項性評分，分項性特級	averaging mean 總平均數
	back wash 測試回衝效應，測試對教學的反動，回沖效應

balance 平衡性	complete correlation 完全相關
banding (成績) 分組，成績組別，按能力分群	completion item 填充題
base rate 基準率	composite network design 組合網路設計
battery of test 成套測驗，測驗組題	compound binomial 複合二項式
Bayesian estimation 貝氏估計	comprehension 理解
Bayesian mean 貝氏平均數	comprehensiveness 綜合性
behavioral objective 行為目標	computer adaptive test 電腦適性測驗
bias 偏倚	computer based testing (CBT) 電腦化測驗
bilingual syntax measure 雙語句法測量	computerized adaptive testing 電腦化適性測驗
bimodal curve 雙峯曲線	concurrent validity 並存效度，同期效度
binominal distribution 二項式分配	conditional maximum likelihood 條件最大概率法
biserial correlation 二系列相關	conditional-probability model 條件機率模式
borderline group 邊緣團體	consequential validity 結果效度
branched testing 分支測驗	constant ratio model 常數比率模式
calibration standard error 校準標準誤	construct irrelevant variance 與構念無關的變異量
capacity 能量，學額，容量，能力	construct under-representation 構念的代表性低落
cascading anchor test 分段校準測驗	construct validity 構卷效度，構思效度，類目效度
central measure 集中量數	constructed response format 建構反應式題型
central tendency 集中趨勢	constructed response item 建構反應式試題
checklist 檢核表	content area 內容範圍
Chi-square test 卡方 (χ^2) 檢驗法，卡方檢定	content standard 內容標準
class interval 組距	content validity 內容效度
class mid-value 組中值	contingency table 列聯表
classification analysis 分類分析化，分析分類化	continuous assessment 經常性評估
closed-ended item 限制答案題	contrasting group 對照團體
cloze procedure 補漏測驗法	control group 控制組，對比組
cloze test 補漏測驗，完成測驗	conventional test theory 傳統測驗理論
clozentropy 補漏測驗比頻評分法	cooperative learning 合作學習
coach effect 練習效應	core entry 核心資料
coefficient of determination 決定係數	correction 校正
coefficient of variation 差異係數	correction for guessing 猜測校正
communicative testing 交際能力測驗	
competency 能力，職能，技能	

correlation 相關	degree of freedom 自由度
correlation coefficient 相關係數	dependent variable 因變量
criteria 標準	derived score 導出分數
criteria chart 標準評估表	descriptive statistic 敘述統計學
criterion 效標，標準	deterministic 決定論
criterion measure 標準量數	deviation 偏差，離差，離中差數
criterion performance 標準作業	deviation score 離差分數
criterion referenced 標準參照	diagnosis 診斷
criterion-referenced interpretation 標準參照解釋	diagnostic test 診斷測驗，診斷性測驗
criterion-referenced measurement 標準參照測量模式	dialogue journal 對話日誌
criterion-referenced objective 標準參照目標	dictation 聽寫
criterion-referenced test 標準參照測驗	differentiation 區分作用
criterion-referenced testing programme 標準參照測試方案	direct testing 直接測驗
criterion referencing 標準參照	discrete point test 離散項目測驗
criterion-related validity 效標關聯效度	discrete point testing 離散項目試驗
critical thinking 批判性思考	discriminating index 鑑別指數
culture fair 無文化偏向的	discriminating power 鑑別力
culture free 不受文化影響的	discrimination 區分度
cumulative frequency 累加次數	discrimination index 區分指數
cumulative frequency polygon 累加次數多邊形	discriminative validity 區分效度
cumulative percentile 累加百分比	dispersion 差異量，離散度，分散度
curriculum 課程	distractor 誘答，干擾項
curriculum-based assessment 課程為本評估	distractor value 誘誤值
curriculum validity 課程效度	distribution 分佈
cut-point (also cut-score) 分割點	divergence measure 離中量數
data 數據，資料	divergence tendency 離中趨勢
data matrix 資料矩陣，數據矩陣	domain 測驗範圍
decision theory 決策理論	domain of assessment 評估範疇
decision validity 決策效度	domain-referenced 範疇參照
declarative knowledge 陳述性知識	domain-referenced test 領域參照測驗
degree of difficulty 難度	domain sampling 領域抽樣
	domain specification 範疇指定
	effectiveness 效率
	efficiency 有效性
	effort 後天努力

elaboration 詳細闡述	face validity 表面效度
entry characteristic 前提特質	facility index 易度指數，通過指數
equal discrimination 等鑑別力	facility value 易度值
equal opportunity model 等機會模式	factor analysis 因素分析
equal probability mode 等機率模式	fairness 公平性
equal risk model 等風險模式	feedback 回饋，反饋
equipercentile equating 等百分等級等化	feedforward 前饋
equivalent form 等值複本	fit analysis 適合度分析
equivalent form reliability 等值複本信度	fit of model 模式適合度
equivalent form test 平行測驗	fixed branching rule 固定分支法
equivalent group design 等團體設計	fixed response format 固定反應題
equivalent test design 等測驗設計	format of test 測驗類型
error 誤差	formative assessment 進展性評估，形成性評定
error variance 誤差方差	formative test 形成性測驗，形成期測驗
essay test 作文測驗	free response item 自由反應題
essay-type question 論述題	frequency 次數
estimate 估計	frequency distribution 次數分佈，頻率分佈
estimation 估計	frequency distribution table 次數分布表
estimator 估計量	frequency polygon 次數多邊形
ethical language testing 進階語言測驗	general objective 一般目標
evaluation 評鑑	generalizability 普遍性
even distribution 均勻分佈	generalizability coefficient 概化係數
exact word method 切合原文法	generalizability theory 概化理論
examination 考試，檢驗	grade-level classroom 級別課室
examination syllabus 考試課程綱要	grading 分等，評級
examinee 考生	grid 網絡
expansion factor 展開因素	grouped data 已歸類數據
expectancy table 預期表	grouping 歸類，分組
expected a posterior (EAP) 期望事後估計	guidance 輔導
expected loss 期望損失	guideline 指引，指南
experimental design 實驗設計，試驗設計	high stakes test 高風險考試
experimental group 實驗組，試驗組	higher-order thinking skill 高層次思維
experimental method 實驗法，試驗法	histogram 直方圖，史相圖
extended response 擴展式反應	
external anchor test 外校準測驗	
external validity 外在效度	

holistic rating 總體特級	item analysis 項目分析，試題分析
horizontal equating 平行等化	item bank 題庫，試題庫
impact 衝擊	item characteristic curve 項目特徵曲線
impression marking 印象評分	item difficulty 試題難度，項目難度
independent variable 自變量	item discrimination 試題鑑別度，項目鑑別力
index of discrimination 鑑別力指數	item facility 試題難易度的
indicator 指標	item information curve 項目信息曲線
individualized testing 個別化測驗	item pool 題庫
inferential statistic 推斷統計學	item response 項目反應，試題反應
inflection point 轉折點	Item Response Theory (IRT) 項目反應理論
informal assessment 非正式評估	item validity 項目效度
information curve 信息曲線	iterative weighting procedure 迭代加權法
initial item calibration 初步項目校準	job analysis 工作分析
initial person measure 初步受試測量	joint maximum likelihood 聯合最大概似法
innate ability 先天能力	key 正確答案，關鍵碼
instruction (試卷) 指示，教學，教授，指導	knowledge 知識
integrated assessment 綜合評估	Kuder-Richardson 20 庫德—理察遜二十公式
integrative test 綜合性測驗	Kuder-Richardson Formula 庫德—理查遜公式，庫理公式
intellectual skill 智力技能	Kuder-Richardson Reliability Coefficient 庫德—理查遜信度係數，庫理信度係數
intelligence test 智力測驗	kurtosis 峯態
item analysis 項目分析	language aptitude test 語言性向測驗，語言稟賦測驗
item difficulty 項目難度	latent distance model 潛在距離模式
item facility 項目易度	latent linear model 潛在線性模式
inter-rater reliability 評分員相關信度	latent trait theory 潛在特質論
internal anchor test 內校準測驗	learning log 學習記錄檔
internal consistency 內部一致性	learning outcome 學習結果，學習成績
internal consistency reliability 內部一致性信度	learning test 學習性的測驗，學習課業
International English Language Testing System (IELTS) 國際英語能力測驗	least square estimation 最小平方法
interval scale 等距尺度，等距量表	
intra-rater reliability 評分員自身信度	
invariance 不變性	
inventory 調查表，個性調查	

level descriptor 級別指標，級別敘述語或述語	minimal guessing 最小猜測	
level of significance 顯著性水平	minimum 最低分	
licensing examination 執照考試	mode 眾數，方式，考試方式	
likelihood ratio 概似率	moderation 審核	
Likert scale 李克特量表	Modern Language Aptitude Test (MLAT) 現代語言性向測驗	
linear equating 線性等化	modern test theory 現代測驗理論	
listening test 聆聽測驗	monotone increasing 單調遞增	
local independence 局部獨立	motivation 動機	
location parameter 落點參數	motor skill 運動技能	
log-odds 對數勝負率	multidimensionality 多向性	
logical analysis 邏輯分析	multilevel format testing 多水準測驗	
logical validity 邏輯效度	multi-media laboratory 多媒介語言實驗室	
logit 邏輯子	multiple-choice 多項選擇	
loss of information 信息損失	multiple choice format 多項選擇形式	
low stakes test 低風險考試	multiple choice item 多項選擇題	
lower limit 下限	multiple correlation 多重相關，複相關	
marginal maximum likelihood 邊緣最大概似法	multiple regression 多元回歸	
marking 評分	multistage testing 多階測驗	
marking scheme 評分方案	multivariate analysis 多變量分析	
mastery testing 精熟測驗	multivariate data 多變量數據，多變量資料	
matching item 配對題	 	
maximum 最高分	N n 數值	
maximum information 最大信息	negative correlation 負相關	
mean 平均數	negative discrimination 負向鑑別力	
mean deviation 平均差	negatively skewed curve 負偏態曲線	
mean score 均數	nominal scale 稱名尺度，類別標準，稱名量表	
mean square 均方	nondichotomous scoring 非二分法計分	
mean square standard deviation 標準化殘餘均方	nonspeedness 非速度性	
measurement 測量	norm 常模，規範	
median 中數，中位數	normal curve 正態曲線，常態曲線	
metacognition 後設認知，元認知	normal distribution 正態分佈，常態分佈	
mid-score 中位成績	normalized standard score 正態化標準分數	
norm-referenced device 常模參照類		

norm-referenced interpretation	常模參照解釋	percentage score	百分比例分數
norm-referenced measurement	常模參照測量模式	percentile	百分位數
norm-referenced test	常模參照測驗	percentile deviation	百分位差
norm-referencing	常模參照	percentile point	百分點
objective	目標	percentile rank	百分等級
objective-referencing	目標參照	perfect scale	完全模式
objective test	客觀測驗	performance	表現
objective test item	客觀測驗項目	performance assessment	實作評量，表現評估
objectivity	客觀性	performance criterion	作業標準
observed score	觀測分數	performance standard	表現標準
one-parameter logistic mode	單參數對數模式	performance test	操作測驗
one-parameter normal ogive model	單參數常態肩形模式	person fit	受試適合度
open book examination	開卷考試，開卷測驗	pilot test	先導測驗
open-ended question	發揮題，開放性試題，隨意作答題	placement test	安置測驗，編班測驗，定位測驗
open-ended response item	開放反應題	pilot version	試行版本
operational version	使用方案	point-biserial correlation	點值兩列相關
optimal weight	最大加權	point measure	位置量數
option	配列答案，提供答案	point of inflexion	回折點
oral test	口試	policy of assessment	評估政策
ordering item	排列題	population	總體
ordinal scale	順序尺度，順序量表	population parameter	總體參數
outcome	結果	portfolio	代表作選，學習歷程檔案，樣本匣
paper-and-pencil language test	筆試	portfolio assessment	卷完評量，檔案評量
parallel form	平行試卷	portfolio conference	學習檔案會議
parameter	參數	positive correlation	正相關
partial credit model	部分給分模式	positive skewed curve	正偏態曲線
peer assessment	同儕評估，朋輩評估	post-test	課後測驗
percentage passing	通過百分比	posterior probability	事後機率
		practical assessment	實習評估
		practice effect	練習效應
		pragmatic test	實務的測驗，實幹的測驗
		prediction	預測
		predictive validity	預測效度

premise 前提	rank correlation 等級相關
pre-test 課前測驗	rank data 等級資料
prior probability 事前機率	rank difference 等級相差
probability 概率	rank order 名次
probable error 機誤	rank scale 等級量表
proficiency test 熟練測驗	Rasch measurement 羅序測量模式
profile 側面圖，簡介，概況，相關資料	rater 評分人
programmed testing 程式測驗	rating procedure 評分程序
progress test 進步測驗，進展測驗	rating scale 評定量表
project 設計作業，專題研習	rating 等級，品級
prompt 提示	ratio scale 比率尺度，比率量表
protocol 基礎材料，擬訂	raw score 原始分數
psuedo-chance scale level 假性機會量尺值	readiness pretest 準備度測驗
psychological effect 心理效應	reading laboratory 閱讀實驗組件
psychometric 心理測量學	reasoning 推理
psychometric paradigm 心理測驗學範式	reference point 參照點
public examination 公開考試	referenced metric 參照量尺
pyramiad testing 金字塔型測驗	regression analysis 回歸分析
qualitative analysis 質的分析	relative efficiency function 相對有效性函數
qualitative assessment 質的評估	relative variability 相對差
qualitative data 質量資料	relevance 適切性
quantitative analysis 量的分析	reliability 信度，可靠性
quantitative assessment 量的評估	reliability coefficient 信度係數
quantitative data 計量資料	reliability of homogeneity 同質性信度
quartile deviation 四分位差	representative sample 代表性樣本
question guessing 猜答	representativeness 代表性
quiz 小測	response contingent testing 反應權宜測驗
quotient 商數	response format 反應型式
r ² 決定係數	restricted response 限制式反應
random 隨機	result 成績
random error 隨機誤差	robust equality approach 強韌相等法
random sample 隨機樣本	rote memory 強記
random sampling 隨機取樣	rubric (試題) 指示，評分註釋
range 全距	running record 前科記錄

sample 樣本	Spearman-Brown formula 斯皮爾曼布朗公式
sample-free item calibration 樣本獨立項目校準	
sampling 抽樣，取樣	specific objectivity 特定客觀性
sampling error 取樣誤差	specification grid 規範網格
scaffolding 鷹架	specification table 細目表
scale 量表，尺度	specificity 特殊性
scaling test 調整測驗	specific objective 具體目標
scalogram 量表	speededness 速度性
scatter diagram 散佈圖	split-half reliability 分半信度
scattergram 散佈圖	stakeholder 持份人
scatterplot 散佈圖	standard 標準
schemata 圖式（複數）	standard deviation 標準差
scholastic aptitude 學術能向	standard error (SE) 標準誤，標準誤差
score residual square 得分殘餘平方	standard error of estimate 估計標準誤
scoring 評分，評卷	standard error of measurement (SEM) 測量標準誤
select-answer type 選擇答案式	standard nine 標準九分
selection 甄選	standard normal curve 標準正態曲線
selection ratio 選擇率	standard normal distribution 標準正態分佈
selection test 甄別試	standard score 標準分數
self-assessment 自我評估	standardization 標準化
self-rating 自我評級	standardized parameter 標準化參數
self-report 自我報告	standardized residual 標準化殘餘
self-selecting testing 自選測驗	standardized score 標準分數
self correlation 自身相關	standardized test 標準化測驗
semi-direct test 半直接方法	stanine 九點分
sequential item testing 連續測驗	statistic 統計量
short-answer item 簡答題幹	statistical analysis 統計分析
short answer question 簡答題	statistical constant 統計常數
showcase portfolio 展示型成長記錄袋，慶功型成長記錄袋，陳列案卷	statistical dependence 統計依賴
simple equality approach 簡易相等法	statistical graph 統計圖
single group design 單團體設計	statistical independence 統計獨立
skewness 偏態	statistical inference 統計推斷
skills testing 宏觀技能測試	statistical significance 統計顯著性，統計顯著度
smooth curve 修勻曲線	statistical table 統計表

statistical universe 全體事實	test 測驗，測試
statistic 統計量	test-free person measure 測驗獨立受試測量
statistics 統計法，統計資料，統計數字	test administration 測驗的施行
stem 題幹	test anxiety 測試焦慮
step function 階段函數	test battery 成套測驗，測驗組題
stochastic independence 推測獨立	test bias 測驗偏向
stradaptive testing 分層適性測驗	test blueprint 測驗藍圖
stratified sample 分層樣本	test coaching 試前集訓
strong true score model 強真分數模式	test construct 測量概念，測試結構
structured question 規劃題	test design 測驗設計
structured response item 按結構反應題	test development committee 測驗發展委員會
subjective test 主觀測驗，主觀評分測驗	test item 測驗題目，試題，測驗題，測驗項目
subtest 子測驗，分測驗	test length 測驗長度
success criterion 成效標準	test linking 測驗連結性
success ratio 成功率	test manual 測驗手冊
summative assessment 總結性評估	test material 測驗材料
summative test 總結性測驗，總結性測驗	test method 測驗方法
supply-answer type 供給答案式	test pressure 測試壓力
synthesis 綜合	test-retest 同卷重測
synthetic validity 合成效度	test-retest reliability 重測信度，再測信度
T-score T 分數	test score 測驗得分
T-test T 檢驗	test security 測驗安全性
tailoring testing 適性測驗	test specification 測驗詳述
take-home test 在家考試，課外測驗	test-taker feedback 考生反饋
target 目標	test users' handbook (manual) 測驗用戶手冊，考生使用手冊
Target and Target Oriented Assessment (TTOA) Framework 「目標及目標為本評估」架構	test validation 測驗效度，合格測驗
Target Oriented Curriculum (TOC) 目標為本課程	test-wiseness 參試技巧，應試的機智
task 任務，課業	testwiser 精於測驗者
task domain 任務範圍	tetracheric coefficient 四分差係數
taxonomy of educational objectives 教育目標分類學	three-parameter logistic model 三參數對數模式
teacher constructed test 教師自編測驗	three-parameter normal ogive model 三參數常態肩形模式

time limit 時限	validity 效度
trialling (also trying out) 試驗	variability 差異
true/false item 是非題	variable 變量
true mean 真實平均數	variable entry point 變化起始點
true score 真實分數，實得分數	variable termination criterion 變化終止標準
true score theory 真分數理論	variance 方差，差異
true value 真實值	variation 差異
try-out 預試	verbal information 言語信息
two-parameter logistic model 雙參數對數 模式	vertical equating 垂直等化
two-parameter normal ogive model 雙參數 常態肩形模式	washback 重修，引導
two-stage testing 兩階測驗	washback effect 倒流效應，回沖效應，反撥 作用，回流效應，回浪效應
two-way specification table 雙向細目表	weak true score 弱真分數
ungrouped data 未歸類數據	weighted mean 加權平均數
unidimensionality 單向性	weighted mean square 加權均方
uniform anchor test 一致校準測驗	weighted scoring 加權評分
uniform test 一致性考驗	weighting 比重
Unitary Competence Hypothesis 整體語言 能力假說	Woodcock Reading Mastery Test 伍卡克閱 讀精熟測驗
unweighted mean square 未加權均方	written test 筆試
upper limit 上限	Z score Z 分數
	zero correlation 零相關