School of Education
King's College
LONDON

教育署課程發展處

(譯本)

暗箱內探

透過課堂評估
提高學習水平

Paul Black & Dylan Wiliam

暗箱內探
透過課堂評估 提高學習水平
© 2002 香港特別行政區政府

Original title:
Inside the Black Box
Raising standards through classroom assessment
© 1998 Paul Black and Dylan Wiliam
http://www.kcl.ac.uk/depsta/education/publications/blackbox.html

暗 箱

利用學校教育來提高學生的學習水平是一個國家要優先處理的大事。過去十年，各國政府為實現這個目標，一直積極進行多方面的改革。各種措施，例如全國課程測試(National curriculum testing)、普通中學教育證書(GCSE)的發展、學校成績排名榜(league tables)、改善學校規畫和學校管理的新措施、目標釐定、更頻密及全面視學，都是為求達到以上目標。但是，所有措施，仍不足以構成收效的政策，因為它遺漏了某些元素。

學習是由教師和學生在課室的活動所帶動。教師為了幫助學生在現時學習及在將來更好地學習，需要費心處理各種複雜及困難的情況，又要疏導每班約三十名學生的個人、感情及社會壓力。教師若果能夠妥善處理這些教學工作，則能提升學生的學習水平。上述各種政策所遺漏的，就是對教學工作的直接支援。

就系統工程學的角度來看，現有的政策，似乎將課室當作***暗箱***看待。某些來自*外界的信息*被輸入甚或成為新要求：包括來自學生、教師、其他資源、管理規則及要求、家長對子女的擔憂、追求測驗取得高分數所帶來的壓力等等；隨之而來的是*輸出的成果*，如果一切順利的話，學生就會更有學識和能力，獲得更佳的測驗成績，而教師會大致上滿意成果，但心力卻幾乎耗盡了。暗箱內部究竟發生甚麼事呢？若果我們不探察暗箱內的現況，怎保證輸入的信息能產生更佳的果效呢？

我們往往認為，這完全視乎教師的工作表現，他們有責任令課堂內部運作得更順暢。但這答案不盡恰當，理由有兩個：其一，某些信息輸入的改變有可能造成反效果，令教師更難提高學生的學習水平；其二，將提高學習水平這個最困難的責任完全推卸給教師，既不可思議，甚至有欠公允。若制訂政策者和其他人士能提供可行的方法來直接幫助和支援日常教學工作以改善課堂學習，這些肯定是我們應致力尋求的良方。

在本文第一段提及的改革措施，沒有一項能夠提供直接的幫助和支援。無可否認，有關方面的視導的確有留意到課室情況，並且關注到重要事項，但它們的原意不是為了提高學生的學習水平提供幫忙和支援，所作出的建議只是概括性的。

本文的內容涉及暗箱內的情況，焦點放在教學活動上的一個環節—進展性評估(formative assessment)，但根據我們在文中的立論，進展性評估正是有效教學的重點所在。

論據

教與學是需要互動的，這個見解不言而喻，我們就由此展開討論。教師需了解學生的進度和學習困難，方能修訂本身的教學，去切合學生的各種需要。這些需要事前往往無法被教師預知，亦會因人而異。教師可從種種途徑得知學生的需要，例如透過在課室內觀察、與學生討論、以及留意學生的家課或課堂習作。

本文常常提及的術語「評估」(assessment)，意指教師的教學活動，*以及學生自行評估的活動，而這些*活動提供的資料，可以作為回饋，用來改進教學和學習活動。*當所得證據實際上被教師用來修訂教學工作以符合學生的需要，這類評估活動就成為「進展性評估」。*

有關進展性評估的說法其實了無新意。所有教師在教授的班級內都會作出評估。但有三個關乎這過程的重要問題，要在下文逐一解答。這三個問題是：

一：有沒有證據顯示，改善了進展性評估工作就能提高教學水平呢？
二：有沒有證據顯示，教學工作有改善的餘地呢？
三：有沒有證據顯示如何改善進展性評估工作呢？

在試圖解答這些問題之前，我們對眾多的文獻進行了一項廣泛的調查。我們查考了過去九年內的大量書籍和超過 160 本雜誌及期刊，並參考了對早期研究的評論文章。在這個過程中，探究了大約 580 篇文章或篇章。除此之外，我們從上述資料中選取了 250 份材料，準備對這問題進行長時間深入的檢視及評論。該份評論其後刊載在 “教育評估”【Assessment in Education】(Black and William, 1998)期刊內，同時刊登在該期的，還有來自澳洲、法國、香港、南非及美國著名教育學專家對我們工作的評價。

從這整份評論，我們的得出的結論，對以上三項問題所作出的回應，答案絕對是肯定的。就此，本文亦會從三個主要部分描述支持這個結論的證據的本質和確切意義。不過，我們只是會提出一個概要，希望令到文意立論有力，而論述的細節就約略帶過。我們所抱的觀點是，這些立論背後皆有證據支持，而有關論據在已發表的評論文章中很詳盡臚列，成為本文論述的基礎。

*我們深信，以下三個重要問題確立了強而有力的事實，政府機構以及教育界專業人士，若果認真想提升教育水平，就應作仔細的研究。*但我們亦要確認，重大的教育改革絕非一朝一夕的事，已有廣泛的證據顯示，教育改革要獲得成果，需要透過建基於有良好實踐基礎的專業發展計畫才可成事。故此，我們並沒有貿然下結論，進展性評估是教育的另一個「神奇子彈」，因為所涉及的事項實在繁複，又與課室活動的困難情況和推動政策的種種信念上，聯繫太密切。在文中第四及末五節中，我們會正視這些複雜的情況，並試圖勾勒出一個按照我們確立的證據所制訂的策略。

一. 有沒有證據顯示，改善了進展性評估工作就能提升教學水平呢？

一份在 1996 年發表的評論曾集中地，但未有深入地探討為有輕微殘疾兒童而設的課堂評估工作。事前，研究小組縱覽了大量的創新(Innovation)，從中篩選了 23 項(Fuchs and Fuchs, 1986)。這些創新提供了量化的証據，証明無論是參與創新的或投入不大的，學習均有得益，令研究小組感到滿意。自此以後所發表的更多文章，內容都描述謹慎進行的類似量化實驗。我們的評論選收了至少多出 20 項的研究報告—數量視乎適用選擇標準的嚴格程度。這些研究文章都顯示，凡是包括加強進展性評估工作的創新，均能產生重要及豐碩的學習成效。研究對象的年齡介乎五歲小童至大學畢業生，涵蓋眾多學科，並橫跨多個國家。

為了研究工作的種種目的，這個類別的學習成效，是將以下兩個標準作比較：

(一) 學生測驗分數的平均改善情況；

(二) 為典型學生組別進行同樣測驗得出分數的相差幅度。

以上第(一)組被第(二)組相除，得出的比率稱為效應量**(effect size)**。進展性評估實驗產生的效應量介乎 0.4 及 0.7 之間。這些效應量比大部分出現在教育性干預活動的效應量為大。以下的例子闡明這類重大成果的實際結果：

- 效應量為 0.4，將意味參加一項改革的一般學生的表現，成就與首 35%沒有參與改革的學生相同。
- 效應量增加 0.4，對參與普通中學教育證書學生的成績表現改善一至兩個等級。
- 效應量增加 0.7，若套用於數學科在最近期的國際性比較研究工作上(TIMSS-Beaton et al, 1996)，可把英國的排名，從參與 41 個國家的中間排名，提升至頭五名之內。

這類研究揭示了另一個重要特性。*許多數據顯示，改良了的進展性評估，相較其他類型的評估，更能幫助所謂學習能力稍遜的學生，因而在提升整體成績的同時，也使成績差距縮窄。*一份近期的研究報告，專門探討*學習能力稍遜*的學生和學習能力有障礙的學生，顯示出多次的評估回饋，可以協助提升以上兩類學生的學習，(Fuchs et al. 1997) 。這些學生獲得佳績，其意義特別重大，因為教育成效不彰顯是一種負累，徒然浪費才華。況且，認定自己無法再學習的學生，大部分會漠視學校綱紀，在校內搞破壞，甚至逃學曠課。若他們繼續這種惡習，加上他們離校時沒有獲得足夠的學歷令他們立足社會，這類學生很可能會與社會隔離，成為嚴重社會問題的源頭和犧牲品。

這裡得出的訊息很清晰，明顯的學習增益，皆可在我們的掌握之中。透過各種方式所取得的成就均有一個共通特性，就是加強了進展性評估工作；這個特性可作為成功的因素，或至少部分因素。但這不意味，大規模地在一般課堂內取得成果是容易的

事。我們所研究的報告中，出現了一些共通的特色：

- 這些活動皆利用新方法，去加強教師和受教者之間的回饋，這些方法需要用新的教學模式，並要有重大的課堂變革配合。
- 各種途徑的背後，蘊含着對有效學習的設想，特別是需要讓學生積極參與學習。
- 為使評估工作具備進展性功能，必須利用學習成果去修訂教學和學習，故此計畫的主要方面是讓教師落實這項工作。
- 對影響學生的動機和自尊的評估方法，以及牽涉到學生自我評估益處的，均要份外關注。

二. 有沒有證據顯示，工作有改善的餘地呢？

執行乏力

大量研究證據顯示，日常的課堂評估，受到各種問題和缺失困擾，就如以下引言指出：

「教師通常很認真地批改學生習作，但往往未能提供指引以改善工作。在少數明顯的例子証明，給分太鬆及目標不明的作業批改，反而強化低成就及低期望，教師不能充分利用有關學生表現的資料，去指導未來的工作。」
【*中學視學報告書*】(*OFSTED general report on secondary schools) 1996, p.40*

「為何理科中進展性評估的範圍和性質如此貧乏？」
(英國中學理科教師—Dawes and Singh, 1996)

「作為外在的標準，所採用的評估標準根本不具參考價值。」
(比利時小學教師—Grisay, 1991)

「他們只是口頭上說說而已，進展性評估法在現時的教育環境中不切實際。」
(加拿大中學教師—Dassa. Vazquez-Abaad and Aiar, 1993)

在英國或在其他地方所遇到的最重大難題，可以簡略歸納為三類。

其一是關於*有效學習*：

- 教師給學生的測驗，鼓勵機械式和膚淺的學習：儘管教師以為測驗可促進理解，但許多教師不覺察兩者有不協調之處。
- 同校教師之間沒有討論或分享所採用的試題和方法，更沒有對評估的目的作批判性的檢討。
- 在小學教師中尤為顯著的趨勢是重視工作量及形式，但忽略了學習的質素。

其二是關乎*負面的影響*：

- 過份重視分數和評分的功能，卻不重視提供有用的建議和學習功能。
- 採用學生互相比較的種種手法，其首要目的看來是為了競爭而非個人的改善。最終的結果是評估回饋令學習能力稍遜的學生認定他們缺乏「能力」，這令他們變得意志消沉，失去學習動機及認為自己沒有能力學習。

其三是對*評估工作的管理角色*：

- 教師對學生的回饋似是為了達到社會和管理功能，這往往要付上犧牲學習功能的代價。
- 教師通常能夠預知學生在公開考試的成績—因為學生今日經歷的考試是他們經歷過的—但他們對學生的學習需要認識很膚淺。
- 側重收集及記錄分數而忽略對學生學習需要的了解，更有甚者，有些教師對學生先前所得的評估紀錄不予理會。

當然，以上的情況不會在所有課堂教學中出現，在許多學校和課室，這些情況確實不存在。然而，上述的概括性結論，是由來自許多國家的作者，其中包括英國作者，他們透過很多學校的問卷調查、訪問及觀察所收集到的證據而得出的。

空泛的承諾

自 1988 年教學改革法案訂立以來，在英國和威爾斯出現的種種改革，深深地影響到評估工作。在法案通過之前的各種政策聲明，1988 教育部評估及測試工作小組(TGAT - Task Group on Assessment and Testing, DES 1988) 報告書的建議，以及其後陸續就政府政策發表的聲明，均強調教師所作的進展性評估的重要性。可是，大部分提供的資源，公眾和政界的注意力都集中地放在各主要學習階段最後的測試成績水平或級別上；對教師為「總結性」評估所作的各種貢獻，雖然給予一些正式的地位，但實質並未充分重視進展性評估。情況更甚的是，毫不重視教師在進展性評估和總結性評估工作的關係上不協調的情況。

在這些正式承諾之間，就進展性評估的重要性及所給予它的實際優次，確實有非常明顯的差異。最明顯的證據，見於一份由一位學校考試及評估議局(Schools Examinations and Assessment Council)成員撰寫的詳盡報道中，內容介紹該局自 1988 年成立至 1993 年間的工作(Daugherty, 1995 年) 。在這段期間，該局會會議程中明顯討論教師評估工作只出現過兩次。每次出現，只因當時的教育及科學部特別解答對總結性評估工作的提問，而教師的進展性評估工作，迄今未受到重視。故此，被委托執行推行有關評估的政府政策部門，在研究和發展教師的進展性評估工作上，都沒有訂出策略；而有關工作的出版刊物，只投入了極少的資源。

有關對教師和學校的公開評估，目的是為了透過成績表宣揚競爭信息，政治承諾被優先處理，而對進展性評估的承諾就被放在次要的位置。來自世界各地的研究者發現，諸如國家課程(National Curriculum)的種種測驗和普通中學教育證書(GCSEs)等公開考試的功能，套用美國人的說法，是「賭注極大」的考試。它們往往牽制著教學和評估的工作。其中尤其顯著的是，由於它們的功能祇是為了提供總結性成果，而非作出有建設性、有用的診斷，對於教師來說，祇能是差勁的進展性評估模式。

另一個可能性是，許多提出的承諾均假設進展性評估問題不大，因它常常出現，根本毋須正式確認它的存在。後來，學校考試及評估議會(The School Examinations and Assessment Council SEAC) 和其後繼承它的學校課程及評估局(The School Curriculum and Assessment Authority)曾經嘗試，透過製作一般的形式性評估程序指引，發表有關學生工作例證及如何評估這些例證的指引，對教師評估作出支持。後來發現，這些概括的指引不太有用，根本就不應提供，因為它們不是按照現實問題認真研究而作出的。作為例證的材料，本來非常有價值，但將國家課程標準演繹為批改學生習作的準則，對於進展性評估的發展，貢獻不大，反而加強了教師在評估工作中的總結性評估角色，而非進展性評估角色。

考慮到這點，難怪在許多對英國教育改革推行的研究報告中，都一直認為進展性評估「亟需發展」 (Russell et al. 1995)。不過，更近期的研究報告發現，進展性評估在小學有一些進展(Gipps et al. 1996)；在過去兩年，DfEE 特別將在職培訓基金分配到第二學習階段的教師評估發展工作上，令到一些地方教育當局開始透過在職培訓改善進展性評估。這種種進展無疑令人鼓舞，但仍未能矯正過去漠視情況和錯失良機帶來的負面影響。

回過頭來看，未能察覺需要充分支持進展性評估、未能承擔起這些支援發展的責任，都是嚴重的錯誤。甚至在 1988 年以前，在有關教育制度的需要中，進展性評估的角色仍很薄弱。考慮到國家課程的轉變伴之而來的沉重負累，很明顯現存的良好實踐可謂蕩然無存，更遑論出現一套更嚴苛的標準所帶來的挑戰。

三. 有沒有證據顯示如何改善進展性評估呢？

學生的自尊

「……許多學生……滿足於『混日子』……。每位想推行進展性評估的教師，必須重訂教學合約來抵銷學生養成的習氣。」
(Perrenoud 談及瑞士的學生，1991)

學生是採用評估信息改善學習的最終用家。這會導致正負兩面的情況。負面的情況已由以上的引言闡述：當課堂文化強調給予學生獎賞，貼「金星」作為獎勵，評級及班

上排名，就會驅使學生千方百計求取高分，而非留意到其分數應該反映的是他們的學習需要。一項已知的後果，是學生可以選擇的話，他們會逃避艱難的作業。他們亦會花精力和時間去發掘「正確答案」的線索。大部分學生因懼怕失敗而不願意發問。處處碰壁和成績差的學生，會因此認為自己是沒有能力，而這種信念會引致他們將困難歸咎於本身的能力缺失，於是他們「負傷退下火線」，避免花精力在只能帶來挫折的學習上，並試圖用其他方法建立他們的自尊心。成績優良的學生雖則能在這種學習文化中表現優異，但整體來說，只會促使成績低於理想的情況增加和擴大。

從正面來看，這些後果並非不可避免。我們需要的是成功文化，而背後支持的信念是所有人皆可成功。如果使用正確的方法來傳達信息，進展性評估可以是一個強而有力的武器，它一方面能幫助所有學生，而另一方面更能幫助學習能力稍遜的學生取得良好成績。因進展性評估能帶引學生集中地研習他們功課上的具體問題，又使他們明確地認識哪些地方做錯，並提供可行性目標使他們朝正確方向工作。假若學生不受一些過份渲染的觀念所困擾，例如能力、與別人競爭和比較，他們就能接受這些信息努力學習。這個信息總結如下：

- **教師對學生所作的回饋，應關乎他們的作業的具體水平、對改善的辦法給予忠告，並應避免將他們與其他學生作比較。**

學生自我評估

我們還有另一個層面要探討。許多成功的創新活動，發展了學生的自我評估及同儕之間互相評估，從而加強了進展性評估的途徑。這項工作在五歲以上的學童達到一些成功果效。進展性評估和自我評估的聯繫絕非偶然，而是不可避免的。

要作進一步的解釋，我們應要注意的是在發展中的自我評估所遇到的主要難題，不是可靠性及信賴性問題：根據研究發現，學生在自我評估和互相評估時，態度往往是誠實可靠的，他們對自己的評估可以非常嚴格，又可以很寬鬆。學生不同的表現，主要問題在於：學生對能達成學習目標的途徑有清晰的圖像，才能作自行評估。但令人詫異及難過的是，許多學生沒有這個圖像，看來他們已接受課堂教學強制性和欠缺全盤目標的練習。要克服這種凡事被動、凡事接受的學習模式，需要持續下苦功。若學生真的知道那些概括觀念，他們在學習上就更加投入和得心應手。他們的自我評估，就成為他們與教師和同儕之間討論的話題，這將進一步促使他對自己的想法反覆思量。這對良好學習至為重要。

*學生自我評估事實上是進展性評估的要素，它絕非一般奢侈品。*當任何人努力學習時，對他們作出的反饋應包括三個元素—*想要達到的目標*、有關他們*現時處境的*徵象、以及*縮窄兩者差距方法*的認識(Sadler, 1989) 。任何人都要掌握以上三個元素，才能採取行動去改善學習。

上述的論點，合乎對人們如何學習的研究所確立的共同了解：我們對新事物的了解並非是生吞活剝或孤立地儲存的，新事物必須與已有的想法相融合。新舊想法未必一致，甚至會產生矛盾，學習者必須用謹慎的行動解決這些分歧。意識到學習要有新的目標，是學習過程中一個重要的環節。

- **要使進展性評估有成果，教師應培訓學生對學習表現作自我評估，令他們能夠明白學習的主要目的，從而掌握為達成目標他們需要完成的工作。**

有效教學的開展

本文第一部分提到的研究報告清楚顯示，有效的進展性評估計畫不單止為已有的計畫加上一些觀察和測試。它們需要對教學計畫的主要組成部分謹慎監察。隨著辯論過程展開，清楚知道教師的教導和進展性評估的關係是不可分割的。

一開始，選擇堂課和家課是非常重要的。給學生的課業是否適當，要視乎是否符合學習目的。這方面能運作順暢，要視乎整個計畫是否有機會讓學生表達他們學習的進展情況，而工作計畫、討論、觀察學習活動及批改作業的設計，都可以提供這些機會；但接著最重要的是，從觀察或仔細聆聽學生所說、所寫及所行，以了解學生發展理解及表達他們的理解程度。

- **任何教學設計應提供讓學生表達他們的了解程度的機會，因為這樣將會啟動師生互動，藉此使進展性評估促進教學。**

在課堂討論中引導學生用自己的方式表達他們所知，是促進知識增長及理解的重要途徑。學生與教師的對話，讓教師能對學生的想法作出回應，並引導學生的思想方向。但從一些對討論活動的詳細記錄中看到，教師在不經意間作出的回應，妨礙了學生將來的學習。這些例子相同之處在於，教師期待學生作出一個特定回應，當出現意料之外的答案時，他們不曉得彈性處理，或沒有信心去應付。於是教師力圖引導學生去提供他們所預期的答案。教師用此方法去操控與學生的對話，會將學生與眾不同的、經過深思但反傳統的嘗試，全面被封殺。經過若干時日，學生得到的信息：他們不需要運用思考能力去想出答案；習作的目的是去完成或猜想—教師期望看到或聽到的—答案，然後表達出來，好讓教學能繼續進行。

教師和學生之間的交談的一個特點是由教師提出問題。教師用這種自然而直接的方法審視學習，往往徒勞無功。一個常見的問題是，教師不給予學生充分的安靜時間去思索問題才作答。普遍的情況是教師在提問後兩至三秒就自行說出答案，而對學生花一分鐘的思考卻不能接受，因此，學生根本無法想出答案。於是出現以下兩個後果。其一是，只有關於事實的答案才能在如此短時間內提出答案，這類問題就成為主導。

另一個後果是，學生甚至懶得嘗試想出答案—若你知道教師會在數秒之後提供答案，就沒有必要嘗試作答。另一個常見的現象是，課堂內只有少數學生會回答教師的問題，而其餘的樂意將這個機會讓給少數的學生，他們知道自己不能作出如此敏銳的反應，何況他們亦不願意冒著在眾人面前犯錯的風險。於是教師將問題降低層次，並接受少數學生的答案，這樣就能夠令到這節課順暢地進行。但其實，教師不知道大部分學生的了解程度—師生之間一問一答的對話遂成為例行公事，而學生共同投入、緊密參與學習的興趣便被抑壓了。

要打破這個惡性循環，其實有許多方法：包括讓學生有充裕時間作出回應；把學生分成一對或小組去討論他們的想法，然後讓一名學生作報告或回應；讓他們從不同的的答案作出選擇，及要求他們投票選出答案；請他們各寫下一個答案，其後讀出幾個被選的答案等等。要注意的是，要鼓勵所有學生參與，任何喚起學生深思的對話都很重要，因為只有這樣，才能令進展性評估得以開展。

- **師生之間的對話應是發人深省的，著重於引發認知能力，及引導學習，令到所有學生都有機會思考及表達他們的想法。**

課堂測驗、正規測驗或家課練習亦是促進回饋的重要方法。一個設計周全的測驗可以是一次學習機會，亦可以是一次考核機會。常設而短小的測驗，相比較長但不常設的測驗更為理想。當學生學習了一些新知識，一星期之內就要考核他們，但更頻密的測驗卻會產生反效果。測驗內容的質素，即是題目與主要學習目標是否吻合，以及信息是否已清楚傳遞給學生，都需仔細檢查。好題目難求，教師應充分合作，用批判的態度從外界資源搜羅好題目。

使用高水平的題目後，要確保回饋的質素就顯得重要。研究報告顯示，若教師只給予學生分數或等級，他們不會從習作的回饋得到益處。最差劣的情況且是，有部分學生這次得分低，上一回又得到低分，他們自然預計下次所得到的分數亦會低，因而在教師和學生之間接受一個信念，這是由於學生智慧不足所致。回饋確實可以改善學習，最好是不給予任何整體分數，而是指出個別學生的強項和弱項。故此，如何將測驗結果告知學生，使他們明白自己的強項和弱點，是一個關鍵的要點。學生應該得到途徑和機會，根據他們的困難情況對症下藥。因此，為了達致進展性評估目的，教師在完成整個教學單元時才提供測驗是徒然的，因為用測驗結果來解決學習問題為時已晚。

- **測驗和家課練習可以是學生學習的寶貴指引，但練習必須清楚明白，及配合學習目標。教師對測驗和家課的回饋，應視作每名學生就改善學習的指引，必須提供每名學生機會並幫助他們作出改善。**

以上各論點闡明，改善進展性評估絕無捷徑。唯一相同之處在於，教師灌輸學問採用的手法從一開始就要合符現實，並要正視以下的問題：「為了幫助我的學生，我對學生了解的程度是否足夠呢？」

大部分要著手的工作可能會產生困難情況。有些學生對任何改變他們已習慣的常規都會持抗拒態度，因為任何轉變都會威脅到他們，而運用個人的思考能力(不單止要更努力)的挑戰，對許多學生來說可能是件困擾的事。在學生未曾體驗過轉變帶來的好處之前，不能期望學生對學習上的轉變的價值抱有信心。

許多有待推行的改革需要花更多的課堂時間，尤其是當目標是要改變學生對學習的看法和他們的工作方式。因此，教師需要承擔風險，堅信投入更多時間會在將來獲得回報，並且明確認識到，理解膚淺的知識「直銷」和「追趕課程」是既無成效，甚至是有害的。

在這些問題之下會是兩個基本議題。第一個是*每名教師對學習信念的本質*。假設知識是要被傳遞和學習的，學生的理解力會在日後形成。教師清晰講解，學生耐心受教，就是理想教學的要素。這樣的話，則毋需進展性評估。可是，若果教師接受有大量證據所證明，知識傳遞模式是不可行的話，即使根據它本身的標準，則教師的承擔必然是透過互動來教學，促使學生將新事物、新想法融合到他們的理解內容之中。這樣，進展性評估就是一個要素，但還要將它與其他所需的特性結合，驅使學生負起責任及運用思考力。不過，這不意味個人化的一對一教學是唯一的解決辦法，反而需要的是一個具有探究性和有深層思考力的課堂文化，使學生能夠透過與教師或同窗討論中學習。

另一個爭論點是，*教師對學生學習潛力的想法*。若把它誇大使成為對比更鮮烈的描述，一方面我們有*「固定智商」*的觀點，認為每個學生的智力是一成不變的，是先天遺傳的，故此後天再無法改善，只能夠接受部分人學習能力強，而餘下的就學習能力低；而另一方面，是所謂*「未開發潛力」*的觀點，這個觀點在其他文化很普遍。這個觀點一開始就假定，所謂「能力」是後天可以學習得來的各種技能。其根本理念是，學生要有良好的學習效率，就要事先清除由以往失敗經驗造成的心理障礙，這是需要謹慎處理的。哪怕是從未經確認的認知能力受到挫折、或自信心受損、又或兩者兼有。顯然，要在這兩個極端想法之間取得平衡才有出路。*但更多跡象顯示，推行進展性評估，配合「未開發潛力」的設想，真正可以協助學生去學習，對先前成績倒退的學生幫助尤大。*

政策與實施

政策觀點的轉變

國家對評估政策的理念自 1988 年以來一直在變。最初，評估政策被推崇作為在教育界建立競爭市場的重要因素，到現時就變得不那麼迫切，重心反而轉移到為學生訂立目標，評估工作成為審核學生成就的試金石。到現在，政策的發展已到了成熟的階段，*但我們認為現在需要再向前邁進一大步，將注意力放在「暗箱」之內，從而探究評估對直接提升學習水平的潛力，並作為每名學生學習的要素。*

由於這個觀點，許多改革工作需要開展。首要的是，政策應確認課堂為提升教學水平的主要場地，故此優先要做的是必須在課堂內推行和支持改革工作。透過改革課堂暗箱的信息輸入和輸出以提升學習水平是很有用的，但不能從它們本身來判定，而它們是否有幫助，只能根據它們在課堂中的效用來判定。

我們在此所展示的理據，確認了改善進展性評估是推行課堂為本的政策的卓有成效的方法。同一個理據亦表明要做到這點，專注力不要放在教學與學習整項工程上較次或特殊的角度上。*反而應將注意力放在若干個重要元素上，那就是教師與學生互動的質素、給予激勵和支援促使學生在學習上有積極負責任態度，提供所需的幫助將學生拉出「我成績差勁」的思想陷阱，以至培養學生終身學習能力所需的良好習慣。*在能力範圍內改善進展性評估，對提升各方面的學習水平有重大的貢獻。

推行政策的四個步驟

若以上論據成立，我們跟著需要採取什麼行動呢？我們當然不能只停留在評估的層面上，而必須要有更多一般性資源作為指引(Black and Atkin 1996, Fullan 1991)。

從一個極端出發，或可要求更多的研究數據去發掘推行這類工作的最佳做法，但到另一個極端就呼籲立刻展開一個龐大計畫，涵蓋新指引、甚或規則，供所有教師落實執行。採納以上極端做法都不是明智的：顯然兩者自相矛盾。由於研究結果已有足夠數據，因此前者已沒有必要；而後者更不能成立，原因在於，對這個國家(指英國)的學校實況，我們沒有對其課堂實踐有充分了解。

*因此，進展性評估的改善工作，不是件簡單的事情。不可以在現行實踐中加上「急就章措施」，用以承諾很快便會得到成果。*相反，假若要有實質成果的承諾，這只在以下情況發生：每名教師能摸索到自己的方法，將自己的理念與課程結合，用自己的教學模式，在課室中施行。這個進程會比較慢，還需要透過持續性的專業發展和支援計畫。這絕無減低這個信息的說服力，的而且確，這是真確的，要長遠和根本改善教學與學習，只有這樣做才可以成功。一個最近進行的國際性研究，內容有關教育改革和轉變。這項大規模的調查包含在 OECD 內 13 個成員國推行的 23 項計畫，他們得出相

同信息的結論，這信息並成為各成員國訂出有效改革政策的建議。(Black and Atkin 1966)

從以上的論點，帶出以下四項發展方案。

(一) 從進展中學習

一些吸引動聽的想法，若果表達成為一般性的原則，並將這些原則轉化為常規的工作，全部推到教師的身上，就算這些動聽念頭是由廣泛的研究產生，教師也不可能實踐，因為他們的教學工作實在繁重，而只有少數對概念理解的教師才會實行。教師需要的是各種實踐示例，由他們所認同的教師現身說法，並從這些示例獲得他們一樣可以做到的信念和自信心，認定他們自己能做得更好，而且能夠從真實的示例中，了解什麼是實行之後的較佳效果。

故此，計畫不能由一個廣泛推行的培訓計畫展開，除非有合理的理據，如有足夠聲稱知道怎樣做的「培訓人員」，但這種情況並不存在。要邁開第一步，就要將本地的學校組成幾個小團體，這些學校包括小學、中學、市中心學校、市郊和農村學校。每個團體參與學校為本的進展性評估發展，並在自己團體內互相合作。在這個過程中，教師在課室內能為許多實際問題找到答案，是在這裡提出的證據所不能解答的。教師又可以將這些處理系統化，這些方法可能是關於一些根本性見解的，並被其他教師肯定認為在一般課堂情況下是合理的做法。將這項發展計畫在廣泛系列的科目中推行非常有必要，以下利用三個科目來說明：正在發展的數學教育的研究報告與語文教育的研究報告明顯有分別，而又與富創意的文學科類似的報告有異。

參與這計畫的學校需要更多的支援，如給予教師足夠的時間按照已有的理據設計學習新措施，在計畫發展的當兒對他們的體驗多加反省，以及在將來為其他人就有關培訓工作提出建議。此外，還有需要對與教師們合作的外來評估人員作出支援，協助他們發展其工作，又收集有關效益的憑證。這類例證不單對政策執行的指引有幫助，而且能將研究結果傳遞給他人。將課堂活動現場錄影下來作研究，是後一項目的的重要元素。

(二) 傳播信息

執行政策一開始就保持低調，所提供的只是一般性的鼓勵，再加上一些相關的證據，供學校按照現有做法考慮。當發展計畫開始有成效和得到更多資源，就會採取更多主動。接著就會推出一些能將信息傳播得更廣的策略，例如用撥款資助有關的在職培訓計畫。

有必要強調這發展進程必然是緩慢的。我在這裡重申，*假若要有實質成果的承諾，這只會在以下情況發生，即每名教師摸索到自己的方法，能將課程與理念結合到課室活動之中*。即使給予最佳的訓練和最有力的支持，仍需假以若干時日。

(三) 減少障礙

「在這項研究中，大部分教師夾在信念系統、制度結構、議事日程和價值觀的矛盾分歧之中，寸步難行。構成這些摩擦的源頭就是評估，它往往伴隨而來的是極度強烈的感情或感覺，如（思想）被壓制、欠缺安全感、罪惡感、挫折感、以至憤怒情緒等等。研究指出在學校展開的評估工作，絕不是一個純技術性問題；反之，它是一個影響深遠的社會及個人問題。」

(美國中學語文教師- Johnston 等人，1995)

為此，我們應探討教育制度內實際上妨礙了有效的進展性評估發展的所有特性，看看怎樣能減低它們的負面效果。最明顯的影響在於短的公開試。這些考試能夠支配教師的工作，最明顯不過的是教師操練學生為課程以外的問題提供正確答案。教師的行動既然受到支配，他們就無法重返有效的進展性評估的正常軌道。他們無法運用較佳的判斷力，去找出最好的方法來發展學生的學習能力。我們不是想爭辯說這些考試或測驗無用，實際上，它們確保公眾對學校問責上的信心扮演著重要的角色。在不久的將來，我們需要評估這些考試及進展性評估發展計畫，並要仔細觀察兩者的相互作用，試圖令公開試提供的評估模式更為有用。

所有教師必須承擔一些總結性評估，例如向學生家長報告學生的表現，及在學年完結前須完成學生的成績報告，交給學生的新教師。但為了外在原因而對學生進行的總結性評估，顯然與用於監察和改善他們的學習進度的日常持續性評估肯定不同。有些人甚至提出質詢，兩者既然分歧這樣大，就應將它們分開實行。在這方面我們是無能為力的，因為教師必定要分擔前者的部分責任，又要分擔後者的主要責任。事實上，從教師為進展性評估目的而收集的資料，透過選擇和重新演繹，他們應處於一個有利位置，能對每個學生作出公平的總結性評估(William and Black, 1996 年) 。但要教師協調他們的進展性評估及總結性評估職能，顯然困難重重。這矛盾亦顯然見於近年多個英國本土的評估教師評估工作的研究報告上，教師對兩種職能的混淆，一直妨礙了他們的教學進度。

教師在中學普通文憑(GCSE)及高等程度考試所扮演的評估角色是一個特別要討論的話題。馬卓安(John Major)在 1991 年作出不經思考的干預(引自 Daugherty 1995 第 137 頁) ，他聲稱應減少甚至限制教師將學生日常作業的評估結果加於公開試成績上一併考慮。他這個提議，就將多年來就提高教師評估質素所作的驕人進度，弄成大倒退。若能將這個倒退現象扭轉過來，將優次重新放在解決這些貢獻的責難和問題上，對發

展教師評估技巧，在促進及推動上可以有積極作用。比這個更重要的是，它可以提高教師評估工作的地位，又能幫助他們解決以上引用的「影響深遠的社會和個人」問題的言論，即眾多教師在評估角色中遇到的難題。

在上文曾經提及過，透過改善進展性評估的回饋來提高學習質素需要佔用課堂時間，當教師要在期限內「追趕」課程內容，教師就無法同時協調兩者所需的時間。在 2000 年修定課程時減少課程內容，當是一個重要的助力。

(四) 研究工作

為未來在這方面進行的研究合理化而列出研究的問題並不困難。其背後的原因是，雖然成功的創新行動的報道很多，類型又多樣化，但不能交代一些重要細節，例如實際採用的課堂活動方式、教師的動機和經驗、作為量度學生成就的測驗的性質、或參與學生的看法和期望。

當有充分理由推行這些仔細制訂的方案，我們不應斷定萬事皆期待它們的出現。我們已充分認識到有必要為積極發展工作提供理論基礎，而部分最重要的問題，只有透過實踐計畫才能得出答案。

將來可以研究的題目，例如可以是一些方法研究，用以使教師明白進展性和總結性角色兩者的關係，並知道怎樣處理；或將教師總結性評估的預測效度與校外測驗結果進行比較研究。許多評估例子需要加以整合，這對未來的發展甚為重要，因為其中許多問題是應該用基本研究加以處理的。同時，應由經驗豐富的研究人員擔當評估這些發展計畫。

我們對提升水平的態度是否認真？

我們深信我們的研究發現，應立刻成為多個承擔各種任務的機構的研究對象，例如新成立的學歷及課程局(Qualifications and Curriculum Authority)、PFSTED、地方教育局(LEAs)、教師訓練機構、研究及培訓公共機構、一系列的其他在職培訓提供者、以及學校為本的推廣計畫等等。以上概述的計畫對上述機構和組織會有特別的含義，雖然起帶頭作用的會是政府。試圖在這項計畫上考慮它們相對的角色會是言之過早，亦不按章法，雖然成功有賴兩者的互相合作。

本文辯論的要點是提高水平只能靠教師和學生在課室內直接推行改革。已有大量確實證據顯示，進展性評估是課堂活動的要素，要發展它才能提升教學和學習水平，除此別無他法。這個「表面上成立」的個案的立論基礎就是這些巨大學習成果。

我們的教育制度，是取決於許多影響深遠的新措施，這些新措施，是關注現有做法所作的回應，是基於很少證據顯示可能解決那些關注的。在我們對進展性評估的研究上，可以看到充分的理據，足以顯示一個能夠改善學習水平的改革路向的清晰指引。我們懇請在制訂國策時要抓緊良機，引領朝向這個方向邁進。

參考書籍

Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L. & Smith, T. A. (1996). *Mathematics achievement in the middle school years*. Boston, MA: Boston College.

Black, P. J. & Atkin, J.M. (1996) *Changing the Subject: Innovations in science, mathematics and technology education*. London: Routledge for OECD.

Black, P. & Wiliam, D (1998). Assessment and Classroom Learning. *Assessment in Education* 5 (1), pp. 7-74.

Daws, N. & Singh, B. (1996). Formative assessment: to what extent is its potential to enhance pupils' science being realized? *School Science Review, 77* (281), pp. 93-100.

Dassa, C., Vazquez-Abad, J. & Ajar, D. (1993). Formative assessment in a classroom setting: from practice to computer innovations. *The Alberta Journal of Educational Research, 39* (1), pp. 111-125.

Daugherty, R. (1995) *National Curriculum Assessment. A Review of Policy 1987-1994*. London: Falmer Press.

DES (1988) *Task Group on Assessment and Testing: A Report*. London: Department of Education and Science and the Welsh Office.

Fuchs, L.S. & Fuchs, D. (1986). Effects of Systematic Formative Evalustion: a Meta-Analysis. *Exceptional Children,* 53 (3), pp. 199-208.

Fuchs, L.S., Fuchs, D., Karns, K., Hamlett, C.L., Katzaroff, M. & Dutka, S. (1997) Effects of Task Focused Goals on Low Achieving Students With and Without Learning Disabilities. *American Educational Research Journal, 34* (3), pp. 513-543.

Fullan, M. G. with Stiegelbauer, S. (1991) *The New Meaning of Educational Change*. London: Cassell.

Gipps, C., McCallum, B. & Brown, M. (1996) Models of teacher assessment among primary school teachers in England. *The Curriculum Journal, 7* (2), pp. 167-183.

Grisay, A. (1991). Improving assessment in primary schools: "APER" research reduces failure rates, in: P. Weston (Ed.) *Assessment of Pupils Achievement: Motivation and School Success,* pp. 103-118. Amsterdam: Swets and Zeitlinger.

Johnston, P., Guice, S., Baker, K., Malone, J. & Michelson, N. (1995). Assessment of teaching and learning in literature-based classrooms. *Teaching and Teacher Education, 11* (4), pp. 359-371.

OFSTED (1996) *Subjects and Standards. Issues for school development arising from OFSTED* inspection findings 1994-5. Key Stages 3 & 4 and post-16, London: Her Majesty's Stationery Office.

Perrenoud, P. (1991). Towards a pragmatic approach to formative evaluation, in: P. Weston (Ed.) *Assessment of Pupils Achievement: Motivation and School Success,* pp. 79-101. Amsterdam: Swets and Zeitlinger.

Russell, T.A., Qualter, A. & McGuigan, L. (1995). Reflections on the implementation of National Curriculum Science Policy for the 5-14 age range: findings and interpretations from a national evaluation study in England. *International Journal of Science Education, 17,* (4), pp. 481-492.

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

Wiliam, D & Black, P. (1996). Meanings and Consequences: a basis for distinguishing formative and summative functions of assessment. *British Educational Research Journal, 22* (5), pp. 537-548.