



# 學習進展：形成性評量與總結性評量之整合架構

張郁雯／國立臺北教育大學教育系暨教育創新與評鑑碩士班教授

## 一、前言

提到學生學習評量，台灣民眾最先想到的是升學必經的兩個大型考試：升高中的國中基本學力測驗，以及升大學的大學學科能力測驗與指定考試。這兩個考試之目的都在區分學生能力高下，進行人力篩選。因為考試結果對個人的未來教育影響極大，屬於高風險測驗（high-stakes tests）。所以學生的考試表現受到高度重視，也成為衡量教育成效之重要指標。高風險測驗使得教師必須因應外部強大的壓力，提升學生測驗成績的考量往往勝過其他教學因素的思考。因此，高風險測驗之評量形式自然成為學習評量的主流，課堂評量的角色與功能不彰，學習評量很難提供有效的訊息以協助學生改進學習。在台灣，升學篩選機制，使得總結性評量凌駕形成性評量。在美國，近幾年則是因為「把每個小孩帶上來」法案（No Child Left Behind Act, NCLB）講求教育績效，使得州政府的總結性評量左右了教學。理論上州每年測試的標的應符合各州所設立的課程目標。州考試的結果因而能夠用來回饋與改善課堂教學。然而，學者研究卻指出州考試與課程標準間是脫節的（Haertel & Herman, 2005）。這種總結性評量與形成性評量缺乏緊密連結的現象，阻礙了學生的學習進步。

## 二、學習的「惡性三角循環」 （vicious triangle）

總結性評量的大型測試，因為考生人數眾多，考慮效率與信度因素，往往採取容易評分的客觀題型，忽略評量對學生、教師和課堂學習的直接影響。如果課程能力指標並無法以客觀題型測試，這些能力的培養就因而未受到應有的重視。例如，早期國中基本學力測驗不考寫作，寫作教學的時間就大幅下滑。教學重視的是入學考試會考的內容與形式，而非課程能力指標希望學生達到的能力。考試內容太偏重零碎知識，是另一個常被提及的弊病。部分學者因而主張以改變考試作為教育改革的手段，如果教師教學的重點是測驗內容，那麼測試發展就必須確保測試內容與能力，是值得教的。最能反映這種想法的經典名言是「考什麼就學到什麼」（What you get is what you assess）（Resnick & Resnick, 1992）。臺灣在入學考試題型與形式的變革，多少是受到類似思考的影響。國中基本能力測驗近幾年將寫作納入考試，就是希望重新喚起對寫作教學的重視。雖然，透過提升命題品質的努力，的確改善過去考試過於重視零碎知識的部分問題。然而，基於大型考試講究標準化、高技術品質、時效與成本等因素考量（Shepard, 2003），測試



題型仍舊無法反應教學現場的全貌，對學生學習的助益不大。

Black, Wilson,和Yao (2011) 在「Road Maps for Learning」一文中指出外部高風險的評量，往往破壞課程、教學以及評量應有的良性交互影響，而形成所謂的「惡性三角循環」。課程羅列了眾多指標，但對教學缺乏指引。對於學生在評量上的反應，教師只能看到學生做對或做錯，評量本身也未能將教學納入考量。在高風險的測驗經常是此種狀況，結果是評量影響教學，而非是教學決定評量。教師得應付外部考試壓力，直接教學生如何考試，對學習產生不良的影響。在教學上，教師缺乏有力的工具探討學生的學習狀態，以調整或計畫下一步的教學。在惡性三角循環下，課程對教學僅具微弱指引功能，評量卻強力影響教學，但又無法對日常課堂的學習提供有力的指引。

為了改善這種課程、教學與評量缺乏整合的現象，教育學界不少學者力陳課堂評量才是協助教與學的主要評量方式 (Black, 2001; Brookhart, 2003; Shepard, 2000)，因此，應發展出不同於大型評量之理論，相較於大型評量，課堂評量具有持續更新、動態歷程、協助學生表現、立即回饋與較不嚴苛的信度要求等特性 (Shepard, 2003)。回應此一需求，Brookhart (2003) 提出所謂的課堂計量測量理論 (“classroometric” measurement theory)。美國國家研究協會 (National Research Council, NRC) 於 2001 年出版的一本名為「知道學生會什麼」 (Knowing what students know)，在此書中美國國家研究協會委員會強調需要一個新的評量系統，此一系統應具備全面性 (comprehensiveness)、整合性

(coherence) 和連續性 (continuity)。所謂的全面性指的是應使用多元評量證據來作教育決策，不應依賴單一評量分述。整合性則強調大型評量以及課堂評量所強調的學習目標需緊密連結，課程、教學和評量間應環環相扣。也就是說，課程設計所依據的學習和發展理論，也必須是教學與評量的依憑。Wilson和Bertenthal (2006) 進一步闡釋整合性的觀點，主張除了課程、教學與評量間的「橫向」整合，課室、學校、學區和州教育有相同教育目標的「縱向」整合，以及學生在一段時間如何發展出對某特定學科領域理解的「發展上」的整合。連續性則強調評量系統應於多個時間點進行評量以瞭解學生的學習成長與發展。為了因應NCLB績效政策帶來的負面影響，美國教育測驗中心 (Educational Testing Service) 成立了探索未來教育評量之Gordon委員會，其目標在探討現有的教育評量政策、實務與科技；預估21世紀未來教育最可能呈現之樣貌以及所需要的教育評量；對教育評量未來模式之設計提出建議 (The Gordon Commission, 2012)。這些關注未來評量發展的倡議與活動，都強調形成性評量的重要性，將課程、教學與評量間之惡性三角循環關係改變成良性循環，使得「課程與評量」以及「評量與教學」之間存在相互影響的雙向關係，且教學結果可以經由評量回饋到課程。在評量設計上則講求整合認知發展理論模式，最典型的產出就是學習進展之研究 (或稱學習路徑，road maps for learning)。學習進展是個假設性模式，說明長時間學習的路徑，這個學習路徑是經過實徵驗證的。其背後想法是助長學科核心能力的學習，這些核心能力對往後知識與能力的發展是重要的。



### 三、學習進展與形成性評量

課堂評量之整合性研究證實使用形成性評量（Black & William, 1998a; 1998b; Shepard, 2000）學生能獲得顯著的學習成長。什麼樣的評量才是形成性評量呢？Black和William（1998b）對形成性評量的定義如下：

在課堂中教師、學習者或同儕誘發、解釋和使用學生的成就證據以決定如何進行下一步的教學。使用這些證據所做成的決策比未使用這些證據有更佳的品質，這些作法是為形成性評量。

在這個定義中，形成性評量必須與教學有關；要使用證據不只是蒐集證據；行動者包含教師、同儕和學習者，決策品質要更好或有較好的基礎。經過多年的研究，Black和Willam（2009）提出了一個包含三個歷程的形成性評量架構，這三個主要教與學之關鍵歷程為：

- （一）學習者要往哪裡去？（Where the learner is going?）
- （二）學習者目前在哪裡？（Where the learner is right now?）
- （三）如何到達目的地？（How to get there?）

在學習過程中，教師透過評量得知學生在第一個歷程和第二個歷程間的差距，並且找出有效的方法縮短兩者間的差距。要能有效的達到這個目標，老師心中須清楚在該學習領域學習是如何發展的，才能對學生目前的學習狀態作定位，並進一步規劃教學幫助學生往下個發展里程邁進。學習進展描述在一段時間內，學生如何在某個特定領域由生手發展得越來越精熟（Stevens et al., 2007），能讓教師更清楚掌握此三關鍵歷程。

有效的形成性評量有三個關鍵要素：能提供有關學習現狀以及目標狀態間差距的學習證據；給學生回饋；以及讓學生自我評量與學習。學習進展是這三個關鍵要素的基石（Heritage, 2008）。形成性評量要有效，則必須能持續提供教師學生的學習證據，教師據此進行教學，學生學習能有所進展，然後再度以評量檢視學生的學習狀態，進行教學，使得學習不斷往前推進。教師根據學習進展產生對應的形成性評量，也根據學習進展找出學生評量結果在學習進展路徑中的位置，決定下一步的教學。

高品質的回饋訊息是形成性評量的另一要素，學習進展說明好的表現的樣貌（成功標準），教師可將學生的表現和成功標準比對，回饋給學生，告知學生需要加強的方面。同時，教師也清楚本階段的成功標準與前階段和下階段標準間的關係。整個學習歷程是透明易理解的。學生在學習初，就能知道成功標準，利於學生監控和評量自己的學習歷程，進而發展其後設認知。長期目標能讓學生預期學習結束時的可具有的能力或成果。然而，在學習過程中，學生的學習監控以及回應教師的回饋，需要的是短程目標，學習進展所揭示達成長期目標的各關鍵步驟，恰是此種自我反思所需要的訊息。例如，在國語文領域，低年級閱讀能力的一項指標為「5-1-7-2能理解在閱讀過程中所觀察到的訊息」，這是長期目標。要達到此項長期目標，需要透過幾個短期目標逐步達成。例如，一開始設立「能以六何法找出故事中的背景資料」之短期目標。當學生讀一個故事時能提醒自己故事中的主角、發生的時間與地點。而老師也可以評量學生是否能達此標準。接下來的短期目標「能以六何法分析理解故事」，能指出故事裡發生了什麼事，為何發生，結果如何。透過這樣歷程，



學生能夠具體反思其學習，教師可以給非常特定的回饋，同時也能根據評量結果計畫下一步教學。若只有長期目標，教師能判斷學生是否能理解在閱讀過程中所觀察到的訊息，若學生無法理解，就無從知道缺乏了什麼關鍵步驟可以讓學生最終能達到長期目標。

Alonzo (2011) 認為學習進展對於教師掌握形成性評量之關鍵歷程有極大的幫助。首先，教師須知道在學習歷程結束時，學生應該知道和具備何種能力。課程能力指標能提供教師這項資訊。然而，課程和教科書往往涵蓋關聯與連結度低的主題，缺乏發展的觀點，教師很難衡量學生在發展連續向度上的位置，傾向從每個年段，每個單元來看學生掌握內容的情形（亦即這個單元學生學會的多少），缺乏發展的宏觀。相較於課程能力指標重視年級內能力的敘寫（橫貫），學習進展更重視跨年級能力（縱貫）的銜接（Heritage, 2008）。以100年發佈的九年一貫能力指標國語領域中的閱讀能力為例，在一、二年級，包含了七項主要能力，第七項為「能掌握基本的閱讀技巧」。其下分為三個次指標：

- 5-1-7-1 能流暢朗讀出文章表達的情感。
- 5-1-7-2 能理解在閱讀過程中所觀察到的訊息。
- 5-1-7-3 能從閱讀的材料中，培養分析歸納的能力。

在三、四年級階段，能力擴張為14大項，第十四項仍為「能掌握基本的閱讀技巧」。其下分為五個次目標：

- 5-2-14-1 能流暢朗讀出文章表達的情感。
- 5-2-14-2 能理解在閱讀過程中所觀察到的訊息。
- 5-2-14-3 能從閱讀的材料中，培養分析歸納的能力。

5-2-14-4 學會自己提問，自己回答的方法，幫助自己理解文章的內容。

5-2-14-5 能說出文章的寫作技巧或特色。

在五、六年級，能力為10項，並沒有一個大項是「能掌握基本的閱讀技巧」，與此能力較為接近的是「能運用不同的閱讀策略，增進閱讀的能力」。其下分為：

5-3-5-1 能運用組織結構的知識（如：順序、因果、對比關係）閱讀。

5-3-5-2 能用心精讀，記取細節，深究內容，開展思路。

在七到九年級，能力項目為10項，和閱讀技巧比較相關的有兩大項，分別是「能靈活運用不同的閱讀理解策略，發展自己的讀書方法」以及「能欣賞作品的寫作風格、特色及修辭技巧」。兩者又各有其次指標如下：

5-4-2 能靈活運用不同的閱讀理解策略，發展自己的讀書方法。

5-4-2-1 能具體陳述個人對文章的思維，表達不同意見。

5-4-2-2 能活用不同閱讀策略，提升學習效果。

5-4-2-3 能培養以文會友的興趣，組成讀書會，共同討論，交換心得。

5-4-2-4 能從閱讀過程中發展系統性思考。

5-4-2-5 能依據文章內容，進行推測、歸納、總結。

5-4-3 能欣賞作品的寫作風格、特色及修辭技巧。

5-4-3-1 能瞭解並詮釋作者所欲傳達的訊息，進行對話。



5-4-3-2 能分辨不同文類寫作的特質和要求。

5-4-3-3 能經由朗讀、美讀及吟唱作品，體會文學的美感。

5-4-3-4 能欣賞作品的內涵及文章結構。

從上述的指標可以看到隨著年級上升，指標涵蓋面有變廣的趨勢，如低年級到中年級，前三個次指標是相同的，中年級又加上兩個次指標。然而，就指標數目而言，並非隨著年級漸增。高年級的大項指標數為10項，低於中年級的14項，7-9年級，大項指標數縮為8項。因此，各分項能力，在不同年級，缺乏清楚的對應。從指標的編號來看，在低年級，閱讀技巧為第七項，在中年級，則出現在第14項，在高年級，比較接近的是第5項，然而，其他大項也與閱讀技巧相關如第十項「5-3-10能思考並體會文章中解決問題的過程」。由於指標的敘寫，缺乏發展的觀點，教師不容易掌握學生能力進展，從生手如何依序發展成能手。能力指標指出教學的最終目標，但在學生學習的過程中，教師迫切想知道的是學生目前的狀態離最終目標還有多遠。也就是說，教師需要知道學生在某個領域的學習是如何發展的。

#### 四、學習進展之發展模式

美國學界對學習進展的重視，顯現在美國教育研究協會（American Educational Research Association, AERA）年會主題上。在2008年有一個海報的場次探討科學領域學習進展的發展、評量和效度驗證研究。2009年美國國家科學委員會贊助一個專門探討科學領域學習進展的會議（<http://www.education.msu.edu/projects/leaps>）。這場會議環繞著四個學習進展所面臨的挑戰議題：

（1）學習進展的定義；（2）根據學習進展發展評量以瞭解學生的學習狀態；（3）以

學習進展來解釋學生的評量表現；（4）學習進展的使用。這個會議之後，學習進展的相關論文尤其在科學領域快速成長，探討學習進展的期刊論文越來越多。2010年到2012年AERA年會均有相關論文場次探討學習進展，針對的學科內容主要是科學和數學，另外，則是探討學習進展在課室評量議題。Duschl, Maeng, 和Sezen（2011）針對學習進展進行回顧與分析，其資料來源為期刊的專刊，會議論文，35篇論文多半是2008年之後發表的。Duschl, Maeng, 和Sezen認為學習進展的發展和五個領域的研究有關：教學實驗（didaktiks and teaching experiments）、心智理論與後設認知、概念改變研究、教學法以及學習軌跡（learning trajectories），透過這些研究讓學者意識到探究學習和教學的順序路徑之重要性。

研究學習進展的學者，認為它能有效組織課程標準、評量和教學，促進學生的學習（Black, Wilson, Yao, 2011）。要發揮學習進展的功能，首要的工作是建立有效度的學習進展。學習進展的發展通常結合了「由上而下」和「由下而上」兩種方式。科學家 and 科學教育者選擇科學素養的核心知識來建構學習進展，是為由上而下取向，而透過學生學習過程的內容組織分析建構學習進展，則是由下而上取向。現有的學習進展的效度驗證模式可大略歸為三類：第一種由過去對學生學習實徵研究結果彙整而得。其效度仍有待進一步驗證，常見的作法為用此學習進展為架構，發展教學，然後檢視學生的學習成長路徑是否符合原訂的學習進展。第二種方式是以橫斷研究，調查不同年級學生在某個特定領域的能力發展情形。這種作法是將學習進展的發展與效度驗證結合。第三種則是針對不同年級設計序列性教學實踐，探討在理想的教學設計下，學生能學到什麼（Ducan & Hmelo-Silver, 2009）。由於學習進展本身



是個假設性待驗證的臆測模式，文獻上所發表的學習進展研究往往是處於不同發展階段的成果。有些仍在概念發展階段，尚未經效度驗證研究，有些則已經透過橫斷的調查研究或長期追蹤研究加以驗證和修改。

Krajcik (2011) 指出在發展學習進展時，除了著重隨著時間內容知識的成長外，更重要的是描述學生如何使用這些知識從事推理等認知活動。其次，他主張發展學習進展時，不能只依賴現有學生所接受的課程與教學。因為多半的課程缺乏發展觀點，並非建立在學習理論基礎上。學習進展研究應該能夠指出什麼樣的教學經驗能協助學生從某個層次的理解進步到下一個層次的理解。也就是說教學介入是學習進展研究的核心工作。課程和相關的專業發展都是學習進展研究的一部份。

從上述的說明可以知道，發展學習進展是浩大的工程，那麼由誰來完成這項艱鉅的任務呢？多半的學者主張應由跨學科整合的團隊，包含教師、心理計量學者、認知科學研究者、學習科學研究者以及內容領域專家。現場教師與研究者合作發展，能由教師的回饋建立合於學生學習狀況的學習進展，同時，教師也因為參與學習進展的發展，對於學習進展能有深入的理解 (Hess, 2011)。Shavelson (2009) 認為學習進展仍是假設模式，尚處於研究階段，不是相當成熟的產物。必須透過教師與研究團隊的合作進行教學實驗和行動研究，對學習進展提出修正與改進。

## 五、以學習進展為基礎的評量系統—BEAR 評量系統

研究學習進展之學者認為學習進展是一個能重新設計和整合課程、教學和評量的前瞻性策略。要能發展、驗證以及使用學習進展的前提是能夠有效評量學生在學習進展路徑上的位置。柏克萊評鑑與評量研究 (Berkeley Evaluation and Assessment Research, BEAR) 之評量系統即是以學習進展為基礎，發展評量與計分，追蹤學生朝學習目標前進的狀況 (Wilson & Carstensen, 2005)。此系統顯示如何透過學習進展之運用，緊密結合課程、教學與評量。它是以四大原則與基石循環運作的系統。四大原則分別是：發展的觀點 (developmental perspective)、教學與評量的適配 (match between instruction and assessment)、教師管理 (management by teachers)、高品質證據 (evidence of high quality assessment)，對應的四個基石：進展變項 (progress variable)、題目設計 (item design)、結果空間 (outcome space)、評量模式 (measurement model)。

第一個基石：進展變項。找出核心能力，並描繪出學生整年於此一變項上表現的軌跡。通常會分成幾個層次 (里程碑)。進展變項就如同目前國際教育評量，如國際閱讀素養調查 (Progress in International Reading Literacy Study, PIRLS)，國際學生能力評量 (Programme for International



表1 PIRLS 2006 閱讀素養表現的四個國際基準

基準	描述
進階	<ul style="list-style-type: none"> <li>◎ 閱讀故事體文本時，學生能               <ul style="list-style-type: none"> <li>• 整合文本的資訊，對主角的特質、意圖和感受做出詮釋，並提出內文的訊息做為支持證據。</li> <li>• 詮釋比喻。</li> <li>• 開始檢視和評價故事結構。</li> </ul> </li> <li>◎ 閱讀訊息文本時，學生能               <ul style="list-style-type: none"> <li>• 區分和解釋文本不同部份的複雜訊息並提出內文的訊息做為支持證據。</li> <li>• 瞭解組織特徵的功能。</li> <li>• 整合文本訊息以排出活動優先順序並能充分辯護其排序。</li> </ul> </li> </ul>
高	<ul style="list-style-type: none"> <li>◎ 閱讀故事體文本時，學生能               <ul style="list-style-type: none"> <li>• 從文本中找出相關情節並區別重要細節。</li> <li>• 能透過推論，解釋意圖、行為、事件和感受間的關係，並用文本訊息支持推論。</li> <li>• 辨識一些文本特性的使用（如譬喻等）。</li> <li>• 開始詮釋、整合文本中的故事事件和角色行為。</li> </ul> </li> <li>◎ 閱讀訊息文本時，學生能               <ul style="list-style-type: none"> <li>• 辨識和使用多種組織特性以找出和區辨相關訊息。</li> <li>• 根據抽象和嵌入訊息做推論。</li> <li>• 整合文本訊息找出主要想法並加以解釋。</li> <li>• 比較和評估文本的各部份，說出個人偏好的理由。</li> <li>• 開始了解文本要素，如隱喻或作者的觀點。</li> </ul> </li> </ul>
中	<ul style="list-style-type: none"> <li>◎ 閱讀故事體文本時，學生能               <ul style="list-style-type: none"> <li>• 找出主要事件、情節順序和相關的細節。</li> <li>• 能直接推論主角的特徵、感受和動機。</li> <li>• 開始能連結文本不同部份的訊息。</li> </ul> </li> <li>◎ 閱讀訊息文本時，學生能               <ul style="list-style-type: none"> <li>• 在文本中找出一或兩項的訊息。</li> <li>• 從文本單一部份所提供的訊息做直接推論。</li> <li>• 利用次標題和圖表找出文本的某部份。</li> </ul> </li> </ul>
低	<ul style="list-style-type: none"> <li>◎ 閱讀故事體文本時，學生能               <ul style="list-style-type: none"> <li>• 辨識明顯陳述的細節。</li> </ul> </li> <li>◎ 閱讀訊息文本時，學生能               <ul style="list-style-type: none"> <li>• 找到明顯陳述的訊息。</li> <li>• 開始做出文本清楚暗示的直接推論。</li> </ul> </li> </ul>

資料來源：整理自International Association for the Evaluation of Educational Achievement (2007) PIRLS 2006 International Report.



Student Assessment, PISA) 均將學生的表現分為幾個基準 (benchmark) 加以描述。以 PIRLS 2006 為例，學生表現分為四個基準，如表1。不同基準顯示學生提取資訊、綜合不同文本資訊、整合自身知識與文本訊息，以及批判性考量文本中訊息的能力之進展。表1的基準是涵蓋面較廣的學習進展。在科學與數學領域，常見的學習進展變項，如對「地球和太陽系」的理解，涵蓋面較窄。

第二個基石：題目設計。學生進展的評估是融入教材中的 (embedded assessment)，和平常課堂活動無法區隔。評量是教學和學習歷程的一部份，教師在學習的歷程中隨時評估學生的進展和表現。但此時，評量是課程依賴的，而非如大型測試不受限於特定課程。學習進展的研究發展出不同類型的試題結構來描述學習路徑，如，使用選擇題型，每個選項代表學習進展的一個層次 (Briggs & Alonzo, 2009)；使用開放題型外加晤談 (Carragher, Smith, Wisner, Schliemann, & Cayton-Hodges, 2009) 等。

第三個基石：結果空間。主要是幫助教師詮釋學生的表現以瞭解其學習進展。每個選項可以代表在學習進展向度上的不同位置。開放性問題的答案也可分類，找出對應在學習進展向度上的位置。

第四個基石：測量模式。Wilson 和 Carstensen (2005) 以 Rasch 模式產生 Wright maps 解釋學生的表現。特色是學生表現和任務可以放在同一量尺上，能知道學生會什麼，知道什麼，那裡有困難來解釋學生的能力。當學生能力在 Wright map 的位置接近某個題目，就表示學生答對此項題目的機率接近 50%。根據對學生反應之分析，在此圖上也可以看到擁有不同能力層次學生的分佈圖。同時，也能顯現不同题目的相對難度。不同的學者使用不同的統計技術做為其測量

模式例如，Steedle 和 Shavelson (2009) 使用潛在類別分析 (latent class analysis) 檢視學習進展各層次的診斷效度。Duschl, Maeng 和 Sezen (2011) 認為 BEAR 評量系統對學習進展的研究有非常重大的貢獻。

進展變項的想法和之前測驗的計畫藍圖是相似的概念，但在 BEAR 系統中進一步注意到評量作業使用的理由，更透過實徵的訊息瞭解學生的表現。單獨來看，四個基石在評量發展上都不是新的觀念，但結合成一個系統卻是新的嘗試。

## 六、問題與挑戰統

近年學者們對使用學習進展整合課程、教學與評量抱以厚望，相關的研究與發展持續增長中。學者們都認知到學習進展並非發展上的必然性，而是教學和學生先備知識互動的產物。同時，也瞭解沒有唯一正確的學習進展路徑。因為學習進展的假設與推論性質，目前在定義和發展學習進展上仍有待解決的歧見：

- (一) 如何選擇核心能力做為學習進展變項？什麼樣才足以被視為核心能力？回答此一問題，必須從學習研究發現找出對往後學習有重大影響的能力。如識字量會影響往後的閱讀理解。換言之，要能找出核心能力，必須有足夠的相關研究做為基礎。社會學習領域包含許多的學科，有些領域缺乏足夠的研究資料，很難找出核心能力。對於這些領域如何透過研究建立學習進展？是否需有長期追蹤研究？需要多長的研究期限？領域內容和思考能力間的整合關係，誰輕誰重？



- (二) 學習進展一年內應該有幾個里程碑（層次的數目）？這個問題涉及學習進展能提供多細緻的回饋訊息，以及在特定的一段時間內，能否有效區隔不同表現層次的差異。
- (三) 學習進展缺乏一般性。理論上，學習進展要能適用於所有教學脈絡和學習者。然而，由於每個學習者的經驗不同，單一學習成長途徑很難描述不同學習脈絡和學習者的學習進展，那麼，要如何決定學習進展的效度？若學生的學習脈絡與評量所根據的學習進展所設定的學習脈絡不同，那麼評量表現並無法正確反映學生的學生狀態。若學習進展設定較為廣泛，如前述的閱讀素養的基準，則能包含多元的進展途徑，如果學習進展描述的太細緻就比較容易受特定經驗的影響，而無法適用多數學生。
- (四) 教學的角色：學習進展與課程和教學間連結的密切程度如何？所有研究學習進展的學者均同意，學習進展並沒有發展上的必然性，也都認為學習進展不應只適用於特定的課程。有些學者強調教學介入，認為介入能促進學生學習，且是驗證學習進展定義之效度的必要過程，亦即教學是學習進展定義的一部份（Schwarz et al., 2009；Songer et al., 2009）。另一些學者則

認為，學習進展只要反映現有教學狀態下，學生的能力即可（Duncan et al., 2009; Mohan et al., 2009）。

- (五) 必要的效度證據為何？什麼樣的證據和測量可做為學習進展的效度證據？Duschl, Maeng和Sezen（2011）認為進展變項（核心能力）和學習路徑是學習進展的根本。所以學習進展研究必須能針對這兩項提出完整的說明。他們也傾向認為教學能促進學生往下個學習階段推進是重要的證據。究竟學習進展的發展應根據現有的教學實務下學生的表現，還是理想的教學狀態下學生的表現？如何在兩者間取得平衡，既有理想性又不至於不切實際。

確認教學目標為擬定教學計畫之首要工作。然而，實習教師或新手老師在寫教案時，經常是先完成教學活動規劃，再回頭看看教學活動內容可以和哪些能力指標相對應。之所以有這樣的情形，固然有許多成因，但一個重要的因素是因為這些教師缺乏學生在某個領域的發展圖像，無法清楚知道每個年級的學生在某個學習領域應有的表現。學習進展企圖透過其對課程與標準的影響，確保學生能學到適合該年齡的核心概念與能力。然而，學習進展仍處於起步階段，國外的研究也多半集中在科學與數學教育，若要能發揮其潛在的功能，未來需要有更多的研究投入於其他的學習領域之外，上述的問題與挑戰也待學界進一步的釐清。



## 參考文獻

- Alonzo, A. C. (2010). Considerations in Using Learning Progressions to Inform Achievement Level Descriptions. *Measurement: Interdisciplinary Research and Perspectives* 8(4), 204-208.
- Alonzo, A. C. (2011). Learning Progressions that Support Formative Assessment Practices. *Measurement: Interdisciplinary Research and Perspectives*, 9(2-3), 124-129.
- Black, P. (2001). Dreams, Strategies and Systems: Portraits of assessment past, present and future. *Assessment in Education: Principles, Policy and Practice*, 8(1), 65-85.
- Black, P., & Wiliam, D. (1998a). *Assessment and classroom learning*. *Assessment in Education: Principles Policy and Practice*, 5(1), 7-73.
- Black, P., & Wiliam, D. (1998b). Inside the black box: *Raising standards through classroom assessment*. London: GL Assessment.
- Black, P., & Wiliam, D. (2009). Developing the formative assessment theory. *Educational assessment, evaluation, and accountability*, 21, 5-31.
- Black, P., Wilson, M., & Yao, S-Y. (2011). Road maps for learning: A guide to the navigation of learning progressions. *Measurement: Interdisciplinary Research and Perspectives*, 9(2-3), 71-122.
- Briggs, D.C., & Alonzo, A.C. (2009). *The psychometric modelling of ordered multiple choice item responses for diagnostic assessment with a learning progression*. Paper presented at the Learning Progressions in Science (LeaPS) Conference, Iowa City, IA.
- Brookhart, S.M. (2003). Developing Measurement Theory for Classroom Assessment Purposes and Uses. *Educational Measurement: Issues and Practice*, 22(4), 5 - 12.
- Carraher, D., Smith, C., Wisner, M., Schliemann, A., & Cayton-Hodges, G. (2009). *Assessing students' evolving understandings about matter*. Paper presented at the Learning Progressions in Science (LeaPS) Conference, Iowa City, IA.
- Duncan, R. G. and Hmelo-Silver, C. E. (2009), Learning progressions: Aligning curriculum, instruction, and assessment. *Journal Research in Science Teaching*, 46, 606-609
- Duschl, R., S. Maeng, et al. (2011). Learning Progressions and Teaching Sequences: A Review and Analysis. *Studies in Science Education* 47(2): 123-182.
- Haertel, E. H., & Herman, J. L. (2005). A historical perspective on validity arguments for accountability testing. In J. L. Herman & E. H. Haertel (Eds.), *Uses and misuses of data for educational accountability and improvement* (National Society for the Study of Education Yearbook, Vol. 104, Part 2, pp. 1-34). Chicago: National Society for the Study of Education. Distributed by Blackwell Publishing.
- Heritage, M. (2008). *Learning progressions: Supportive instruction and formative assessment*. Council of Chief State School Officers, Washington, DC.
- Hess, K. (2011). It's Time for More Focus on Educator Involvement in Developing and Using Learning Progressions. *Measurement: Interdisciplinary Research and Perspectives*, 9(2-3), 152-154.



- International Association for the Evaluation of Educational Achievement (2007) *PIRLS 2006 International Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Mohan, L., & Anderson, C. W. (2009). *Teaching experiments and the development of a carbon cycle learning progression*. Paper presented at the Learning Progressions in Science conference, Iowa City, IA.
- National Research Council (2001). *Knowing what students know: The science and design of educational assessment*. Committee on the Foundations of Assessment. Pellegrino, J., Chudowsky, N. and Glaser, R. (Eds.). Board on testing and Assessment, Center for Education. Washington, DC: National Academy Press.
- Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement, and instruction* (pp. 3775). Boston: Kluwer.
- Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Acher, A., Fortus, D., et al. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, 46, 632–654.
- Shepard, L. A. (2000) The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.
- Shepard, L. A. (2003). Reconsidering large-scale assessment to heighten its relevance to learning. In J.M. Atkin & J. E. Coffey (Eds.), *Everyday assessment in the science classroom*. Arlington, VA: National Science Teachers Association.
- Songer, N. B., Kelcey, B., & Gotwals, A. W. (2009). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity. *Journal of Research in Science Teaching*, 46, 610–631.
- Stevens, S., Shin, N., Delgado, C., Krajcik, J., & Pellegrino, J. (2002). *Using learning progressions to inform curriculum, instruction and assessment design*. Retrieved on July 6, 2012 from [http://hi-ce.org/presentations/documents/Shawn\\_etal\\_NARST\\_07.pdf](http://hi-ce.org/presentations/documents/Shawn_etal_NARST_07.pdf)
- The Gordon Commission on the Future of Assessment in Education (2012). *Mission*. Retrieved on July 6, 2012 from <http://www.gordoncommission.org/mission.html>.
- Wilson, M. R., & Bertenthal, M. W. (Eds.). (2006). *Systems for state science assessment*. Washington, DC: The National Academies Press.
- Wilson, M. R. & Carstensen, C. (2005). Assessment to improve learning in mathematics: *The BEAR assessment system*. *Journal of Educational Research and Development*, 1(3), 27-50.



專 論

