

課程實施程度的測量

張善培

香港中文大學課程與教學學系

本文回顧及評析屬於「忠實觀」和「實用的相互調適觀」的測量課程實施程度之方法，焦點放在各測量方法的特色及限制，並指出課程學家仍需研究的方向。文獻回顧顯示，雖然「忠實觀」和「實用的相互調適觀」的支持者皆選用相似的資料搜集方法，例如觀察、訪談、問卷調查等，惟其假設和測量結果卻有差異。許多學者當測量課程實施的程度時，都忽略了課程實施建構的多度向性、搜集多方面的資料、以及嚴格測試資料的信度和效度。

引言

由七十年代初開始，課程實施(curriculum implementation)逐漸變為教育研究之其中一個重要範疇。有關課程實施的主要研究項目可分為四類：(1)測量或評鑑教師實施課程的程度；(2)探究影響課程實施的有利因素及障礙；(3)測試不同實施策略的成效；(4)發展有關實施歷程的理論，藉以解釋及推測課程實施之成功或失敗。

在教育文獻中，有關課程實施的涵義和重要性之討論可謂俯拾皆是(例如黃政傑，1991；Fullan, 1991)。但在九十年代，對課程實施分析最全面者，可算是Snyder, Bolin, and Zumwalt(1992)。根據他們的見解，學者可採用三種不同角度去分析課程實施，即忠實觀(fidelity perspective)、相互調適觀(mutual adaptation perspective)、以及課程締造觀(curriculum enactment perspective)。

持「忠實觀」的學者假設，課程實施是一個直線式過程，其理想成果便是教師依據新課程的原本意圖及計劃落實新的教學活動。若果在課室裏實施的課程愈接近新課程的原來計劃，則課程實施愈為忠實、愈為成功。忠實觀把課程當做可預測的文件或成品，而課程發展只是一個技術性問題。課程設計人員扮演專家角色，他們的主要任務是列出成功落實新課程的所有重要元素，教師只是被動的課程使用者而已。由七十年代至現在，大部分有關測量課程實施的研究均屬於忠實觀，其目的是探究如何科學地量度「課程實施的程度」(degree of curriculum implementation)。換言之，課程實施不是全有全無的現象，在有無之間，可劃分不同的程度。現將需要準確地測量課程實施程度的原因歸納如下：

- 提供數據，以便全面地評鑑某課程的成效。評鑑者首先必須了解課程的實施程度，然後才可闡明學生的學習結果與課程實施的關係。
- 是其他有關課程實施研究的必要成分。例如被認為對課程實施有利的任何因素，必須與課程實施程度具有正相關。課程實施程度也可有效地顯示某項實施策略的效能。
- 提供回饋，以便定期監察課程的運作和改進課程。
- 協助作出有關問責的決定，是判斷某課程需要被延續、擴展或終止的基礎。
- 測量課程實施程度時常需要研究員具體地列出課程的特色和運作要求，這類詳細的記錄對新的課程使用者或欲修訂該課程的人很有價值。

持「相互調適觀」的學者認為，新課程未必一定要依據原來的計劃實施，因為不同學校均有其不同的文化背景、政治情況及脈絡，所以要成功實施課程，課程本身必須具靈活性，讓教師修改課程目標、內容、活動、評估策略等。另一方面，教師本身亦可能需要改變自己原來的教學方法、價值觀等，以配合新課程的背後理念

和特色。因此，若果要成功落實新課程，課程與教師雙方均要作出調適(adaptation)，課程實施的關鍵乃在於調適，而非標準化。相互調適觀較重視教師在課程實施過程中的角色，容許學校調整新課程計劃以因應個別需要，讓教師和課程發展人員扮演共同決定者。

相互調適觀可分為兩種取向。實用的相互調適(practical mutual adaptation)跟忠實觀一樣，假設課程實施是一個直線式過程，又相信大部分調適均是可預測的，並假定調適的形式具一致性。但批判的相互調適(critical mutual adaptation)則假設課程實施是一個非直線式過程，調適是不可避免的，而且調適形式是不可預測的。

「課程締造觀」視課程為教師及學生教學經驗的總和，課程綱要、指引、教材套等文件只是工具，協助教師和學生締造課程而已。持此觀的學者相信，課程與教學是不可分割的過程，而課程是不可預先決定的。因此，持課程締造觀的學者並不研究課程實施程度的測量或課程如何被教師調適。

以上簡述了課程實施的三個角度，雖然Snyder, Bolin, and Zumwalt(1992)詳細剖析了每種角度的特色及限制，惟如何測量課程實施並不是他們文章的重點。事實上，評論或比較測量課程實施方法的研究頗少，較重要的文獻只有：Ellett, 1981; Fullan and Pomfret, 1977; Gephart, 1976; Hall and Hord, 1987; King, Morris, and Fitz-Gibbon, 1987; Leithwood and Montgomery, 1980, 1987; Revicki, Rubin, and Stuck, 1981; Scheirer and Rezmovic, 1983; Wolf, 1990; Yin, 1982。可惜以上文獻內容不是現在看來略嫌過時，就是未能詳細論述不同課程實施觀與測量課程實施方法的關係。例如Scheirer and Rezmovic(1983)回顧了一共七十四項與測量「實施程度」有關的研究，並根據五個測量準則，評鑑該七十四項研究的質素。測量準則針對研究員能否應用多種測量技巧、清晰界定實施程度的涵意、嚴格測試信度、嚴格測試效度、以及適當地抽取樣本。Scheirer and Rezmovic的最重要結論，就是研究員還未能

科學化地測量「實施程度」這個建構(construct)。但讀者要注意的是，Scheirer and Rezmovic的分析和評論局限於「忠實觀」。其實，除了「忠實觀」之外，持「實用的相互調適觀」的學者也關注如何測量課程實施的程度(Snyder, Bolin, & Zumwalt, 1992)。

本文旨在回顧過去有關「忠實觀」和「實用的相互調適觀」的測量課程實施程度之文獻，分析各主要測量方法、模式和策略的特色，探討其假設和限制，並以實例說明需要注意的事項。由於「批判的相互調適觀」和「課程締造觀」不關注如何測量教師實施課程的程度，本文將不會回顧屬於這兩類的研究。

以忠實觀為本的測量課程實施程度之方法

根據Fullan and Pomfret(1977)，「課程實施程度」是指某新課程的實際使用與原本計劃使用互相符合的程度。Scheirer and Rezmovic(1983)則認為，「實施程度」是在某一時刻已發生改變的程度，而改變方向是朝著完全和恰當地使用某新課程。

回顧文獻揭示，許多研究都採用行為觀察法(behavioral observation)測量課程實施的程度，諸如Evans and Behrman, 1977; Gersten, Carnine, and Williams, 1982; Kerr, Kent, and Lam, 1985; Morris, 1984; Wang, Nojan, Strom, and Walberg, 1984等。研究員通常首先依據新課程的原本意圖，列出將要被觀察的教師和學生行為、課室佈置與設備，以及教材的性質和使用形式等，然後設計一套評分或編碼系統，以便將觀察資料量化。因此，觀察表是針對某特別課程而設計的，並極注重資料之效度和信度。例如Gersten等人(1982)觀察了四十二位教師的課程實施情況，他們的觀察表預先列出了六項重要的教師或學生行為，觀察員以十二分鐘為一節觀察時段，記錄教師和學生正確行為出現的次數。Gersten等人的報告指出，每位教師必須每星期被觀察最少四次，才可保證觀察資料的信度。所以，觀察法是一種花費人力物力的策略，而且根據

Fullan and Pomfret(1977)的分析，課程實施最少應包含在五個度向(dimensions)上的改變，即學科內容或教學材料、組織結構、角色或行爲、知識和理解、以及價值內化，但觀察法不能直接測量教師或學生對課程的認識程度和態度。雖然有些持忠實觀的學者以「自然觀察」(naturalistic observation)形式去搜集課程實施資料(例如Sabar, 1986)，但事後仍然需要設計一套評分系統及列出所有重要變項作研究之用。在香港，Morris(1984)曾測量五十六位中四及中五教師實施經濟科課程的狀況。他利用行爲觀察法，記錄每位教師出現十三類教學行爲的時間，結果發現全部五十六位教師並無依原來課程的建議落實教學工作。但Morris沒有報告任何測試觀察資料信度或效度的結果，而課程實施的測量只局限於教師和學生的行爲變化。此外，Revicki, Rubin, and Stuck(1981)曾倡議，觀察法是測量課程實施程度的最佳方法，但Cheung, Hattie, Bucat, and Douglas (1996a)的研究顯示，觀察資料的效度與課程實施的度向有關，在某些度向上，觀察資料或會過高估計課程實施的程度。

爲了節省搜集資料的人力物力，有些研究採用了問卷調查，這不僅可以減省時間，還可抽取較大和具代表性的教師樣本，並可搜集在其他課程實施度向上的資料。例如Kimpston(1985)設計了兩份問卷，第一份問卷的目的是調查教師有沒有教授四個不同年級之語文科學習目標，而目標數目可多達二十五至三十五個。第二份問卷則測量教師有否參與實施二十八項以課程目標爲本位的工作，並搜集教師對該二十八項工作之重要性及障礙的意見。Kimpston的研究結果顯示，所有受訪的七十四位教師的課程實施程度只屬低水平，但他並沒有測試教師問卷資料之信度和效度。

Treagust and Rennie(1993)亦運用了問卷，測量六間學校實施科技教育的程度，他們設計了兩份開放式問卷，以便搜集科技教育統籌導師對文件課程和實施課程的意見。他們也利用問卷，邀請負責教授新課程的教師報告他們在教學和使用教材上有甚麼改變。此外，爲了理解學生對科技教育的態度，Treagust and Rennie在兩

個不同時間向學生發出問卷，調查學生在態度上的轉變。除了問卷之外，Treagust and Rennie還利用訪談法和文件分析法，一則可獲得更豐富和全面的課程實施資料，二則可將多來源的資料互相驗證，從而加強數據的信度和效度。可惜，資料三角測定(data triangulation)或方法三角測定(methodological triangulation)並不是Treagust and Rennie的研究焦點。

顯然，若要有效地使用問卷調查法，研究員必須能清晰、具體地編寫問卷題目和設計恰當的評定量表(rating scale)。編製優良問卷是一件很花心思的工作，為了具體表達內容，每個研究時常需要使用大量問卷題目，倘若要包括較多課程實施度向，問卷題目更可能多至不可接受的程度。Newfield(1980)曾建議應用矩陣抽樣(matrix sampling)，令每位受訪教師不需要回答全部問卷題目，而研究員仍可計算出整個教師樣本的總分。可惜Newfield未能提供實例，以便示範矩陣抽樣的可行性。

由上述分析可見，課程實施程度的測量涉及三大困難。首先，「課程實施程度」是一個多度向的建構(multidimensional construct)，與其他建構一樣(如智力、自尊、性向等)，課程實施程度是不可被直接量度的。研究員只能事先界定一些測量指標(indicators)，例如行為指標、心理指標及物質環境指標等等，以便推斷課程實施的程度。因此，測量課程實施的首要工作，就是清晰界定課程的實施度向，然後依據課程的原來計劃，設計指標以便量度在各度向上的變化。不過，課程的實施度向是很難被鑑定的(Cheung, Hattie, Bucat, & Douglas, 1996b)，而且不同課程有不同的實施度向。其次，由於課程實施的測量需要多種指標，研究員必須設計不同種類的量度工具，從不同來源搜集資料。但現今搜集課程實施資料的主要方法局限於觀察法、教師問卷調查、以及教師訪談，極需要發掘另一種有效的測量方式。第三，由於課程實施是一個多度向的建構，研究員必須設法測試問卷或訪談資料的建構效度(construct validity)。可惜實施資料的變異數(variance)通常很

小，不容易利用因素分析法 (factor analysis) 或多特質—多方法分析法 (multitrait-multimethod analysis) 測試建構效度。在下一節，筆者將討論以相互調適觀為本的測量方法，並回顧該取向的支持者如何嘗試解決上述三大困難。

以相互調適觀為本的測量課程實施程度之方法

「實用的相互調適觀」源於Berman and McLaughlin的研究，他們採用問卷及訪談，調查了293個美國聯邦政府資助的教育計劃之實施狀況，並以個案形式，詳細研究了其中二十九個教育計劃的實施過程 (Berman & McLaughlin, 1976; McLaughlin, 1976; Wirt, 1976)。該研究基於以下五種測量準則：

1. 達成某教育計劃的目的之程度。
2. 實施的教育計劃與原來建議計劃的相似程度。
3. 教師在行為上的改變程度 (四點評定量表，由「完全沒有」至「十分」)。
4. 實施的困難。
5. 倘若政府資助中斷，教育計劃被繼續實施的機會。

Berman and McLaughlin根據使用者的報告，發現約有七成計劃的目的已被達成、教師的行為改變之平均值為「中等」、實施困難的平均值介於「少許」與「中等」之間、實施的忠實程度介於「中等」和「十分」之間，而超過五成計劃將會被繼續推行 (Datta, 1980)。但Fullan and Pomfret (1977) 曾指出，以上的測量設計有兩大缺點：一方面全部測量項目只屬於使用者的自我報告 (self-report) 或感覺 (perceived) 的改變，學者對其信度和效度仍有很多爭論；另一方面，全部測量項目均不夠具體，並沒有考慮各實施度向。儘管Berman and McLaughlin的研究備受一些論者抨擊，但他們首先指出相互調適在實施過程中的重要性，被他們發現的調

適包括：學校私自增聘教師、教師執行測驗和記錄成績時需要額外教師或電腦支援、教師只教授某單元的部分課題、以及有些學校沒有在高年級推行新課程等(Wirt, 1976)。

迄今為止，Hall和他的同事創立的「關注為本採用模式」(Concerns-Based Adoption Model, 以下簡稱CBAM)是最多學者用作測量課程實施的基礎，CBAM主要有三種測量工具，即關注階段(stages of concern, 簡稱SoC)、使用的層次(levels of use, 簡稱LoU)和創新的形貌(innovation configuration, 簡稱IC)。以下各段逐一敘述每種CBAM測量工具之特色，並指出仍待研究員繼續探索的問題。

所謂SoC，是指教師對新課程的關注階段。Hall和他的同事假設個別教師在課程實施過程中，其關注變化將會經歷七個階段(見表一)。他們的研究顯示，教師起初只會低度關注新課程，接著關心自己個人，然後注意工作方面，最後關注新課程對學生的影響。若要測量教師對某項課程的關注階段，研究員可利用面談、開放式的書面報告或問卷調查，就資料的信度和效度而言，則以問卷調查最佳(Hall & Hord, 1987)。

SoC問卷包括一共三十五題，並組成七個量表，分別測量七個不同的關注階段。根據Hall, George, and Rutherford(1977)的報告，各量表的內在一致信度系數介於0.64和0.83之間，而重測的相關系數則介於0.65和0.86之間。Hall等還測試了問卷資料的效度，他們發現：(1)約七成問卷題目只與它們指定的關注階段才會產生較高的相關系數；(2)若兩個關注階段之間的距離愈遠，則它們的相關系數愈少；(3)因素分析揭示，量表能夠測量七個獨立的關注建構；(4)問卷數據與訪談資料接近一致。

由七十年代中開始，很多學者應用了SoC問卷去探究課程實施，諸如Bailey and Palsha(1992)、van den Berg(1993)、Carter and Sanders(1992)、Hall and Loucks(1978)、Hall and Rutherford(1976)、James and Hall(1981)、Kimpston and Anderson

表一 CBAM的關注階段

階段	名稱	特徵
1	低度關注 (awareness)	對新課程表示很少關注或參與。
2	資訊性 (informational)	表示對新課程具有一般認識，並有興趣了解多些有關該新課程。個人似乎不擔心新課程與自己的關係，只關注新課程本身的一般特徵、影響、使用的要求等。
3	個人的 (personal)	個人尚未肯定新課程對自己的要求，自己能否應付這些要求，以及自己在實施新課程時所扮演的角色。個人會考慮的問題包括：回報、所需作出的決策、新課程與現存學校結構的衝突、新課程與自己現在需要承擔的責任之間的矛盾、新課程對自己及同事在金錢和地位上之含義等。
4	管理 (management)	集中關注落實新課程的歷程和工作，以及使用資訊和資源之最佳方法。特別關注效率、組織、管理、時間表和所需時間。
5	後果 (consequence)	關注新課程對學生的影響。焦點放於新課程的適切性、學生成績的評鑑，以及提升學生學習成果的變革等。
6	合作 (collaboration)	關心在實施新課程時，如何與其他人士協調和合作。
7	再關注 (refocusing)	探討新課程更多普遍性優點，包括探討大量改變或由另一新課程取代的可行性，個人對其他新方案具有明確的想法。

資料來源：Hall and Hord, 1987, p. 60.

(1988)、Maloney (1993)、Marsh(1985)、Whiteside and James (1986)等。但讀者要留意的是，SoC問卷是屬於一般性測量工具，未能考慮各課程的特殊性，研究層面也只能限於教師在感受、動機及態度方面的變化。再者，Bailey and Palsha(1992)發現一個包含

五個關注階段的模式比較七個關注階段更具信度和效度，但他們只運用探索性因素分析法(exploratory factor analysis)去測試其新的SoC，若能採用驗證性因素分析法(confirmatory factor analysis)，則研究結果更具說服力。其次，Bailey and Palsha的研究樣本只包含新使用者，其結論的正確性，值得懷疑。

此外，van den Berg(1993)在荷蘭及比利時的研究顯示，SoC似乎受文化影響。他修訂了Hall在美國設計的SoC問卷，將問卷題目增至五十二題，並修改部分關注階段的次序及性質。但正如Anderson(1997)指出，van den Berg的研究涉及包含多類型創新的大規模教育改革，而Hall和他的同事的SoC研究卻針對某個別教育創新。因此，若讀者打算利用SoC問卷作為測量課程實施之工具，他們絕不可以預先假定SoC問卷資料必然具信度及效度，因為研究結果會受文化、課程性質等因素所影響。

由於SoC只測量教師的知覺、感受和態度，未能展示個別教師在學校內落實新課程的情況，所以Hall and Loucks(1977)針對教師在行為上的變化，設計了另一種測量工具，稱為「使用的層次」(LoU)。他們認為課程實施可劃分為八個層次(見表二)，而教師的使用層次最低限度要達致「例行化」，才可算已實施了新課程。

像SoC一樣，LoU是一般性工具，可被用來測量不同種類的教育創新之實施層次。為求進一步清晰界定各層次，Hall和他的同事倡議了七個度向，即「知識」、「獲取訊息」、「分享」、「評估」、「規劃」、「報告使用狀況」、以及「執行」，藉以具體地描述每一使用層次的特徵。有關以上七個度向的闡釋，讀者可參考Hall and Hord, 1987。

雖然LoU比SoC較為明確，但Hall並沒有詳細解釋如何發現該七個度向，也未能提供任何理論支持其重要性。再者，就可見的文獻來說，仍未有學者嘗試測試所有七個度向的建構效度，這類探究很有價值，因為當研究員測量某教師使用課程的層次時，她需要依據上述七個度向的資料，才可準確地決定該位教師的層次。

表二 CBAM的課程使用層次

使用的層次	使用的範圍
1. 未使用 (nonuse)	使用者對於課程改革缺乏了解，或了解甚少，未參與課程改革工作，也未準備參與。
2. 定向 (orientation)	使用者已經獲取或正在獲取課程改革的資料，且已經探討或正在探討課程改革的價值取向，及其對使用者的要求。
3. 準備 (preparation)	使用者正為第一次使用創新課程而準備。
4. 機械地使用 (mechanical use)	使用者致力於革新的短程使用或日常使用，但是缺乏反省的時間。使用上的改變，旨在符合使用者的需求，而非學生的需求。使用者基本上所試圖熟練的工作，是使用改革的課程所要求的，結果常是膚淺且不連貫的使用。
5. 例行化 (routine)	在使用過程中，使用已經成為習慣，如果有改變的話，僅是少數。很少慮及改變革新方案的修訂各革新的效果。
6. 精緻化 (refinement)	使用者依據短期或長期的結果，修訂革新的方案，以增進革新的即時效果。
7. 統整 (integration)	使用者結合自己和同事在革新上的努力，在共同影響的範圍內，給予學生集體的影響。
8. 更新 (renewal)	使用者再評鑑革新方案的品質，尋找目前革新的變通方案或重大修正方案，以增進其對學生的影響，檢視領域內的新發展，探索自己及整個學校系統的新目標。

資料來源：李子建、黃顯華，1994，頁345；Hall and Hord, 1987, p. 84.

如果想得到高信度的LoU資料，Hall and Hord(1987)建議必須採用深入的行為觀察法。可惜許多教師行為都在課堂之外發生的。因此，觀察員必須在學校內整天跟隨某一位教師，記錄她規劃新課程的情況、與其他教師談話的內容、以及在課堂如何教授新課程等。顯然，這種LoU資料搜集方法極費時，故許多研究都以「焦點

訪談」(focused interview)代替觀察，如Anderson, 1991; Carter and Sanders, 1992; Evans and Hopkins, 1988; Hall and Loucks, 1977; Maloney, 1993及Marsh, 1985等皆是。訪談首先以一開放式問題為起點，然後透過一系列有次序的問題，深入探查某特別的使用層次。每次焦點訪談需時約二十分鐘，並安排錄音作為記錄，但若要成為熟練的訪談員，需要參與三天的訓練工作坊，有關訪談詳情可參閱Loucks, Newlove, and Hall(1975)的報告。

根據Hall and Loucks(1977)的報告，焦點訪談的資料具高信度。例如他們訪問了1,381位教師，分別推行三項不同課程，而評鑑者之間的信度系數介於0.87和0.96之間。此外，為求測試資料的效度，Hall等也報告訪問了四十五位中學教師，他們首先測量每位教師實施某科學課程的LoU，然後依據LoU訪談結果，隨機抽取其中十七位教師，以便接受一整天的人種誌觀察(ethnographic observation)。研究顯示，觀察員與訪談員的資料之相關系數為0.98。可是Gersten, Carnine, and Williams(1982)的研究結果就完全不同。在上一章節，筆者已簡述了這三位學者的行為觀察法，他們的研究設計也包括LoU的訪談。Gersten等人批評LoU對課程實施不夠靈敏，在四十二位受訪教師之中，有三十七位均被評估為「機械使用者」，有三位屬於「例行化」層次，而只有兩位屬於「精緻化」層次。但當Gersten等人查閱該五位「例行化」及「精緻化」使用者的行為觀察記錄之後，他們竟然發現有三位教師的LoU被高估了，其中一位所謂「精緻化」使用者甚至從未實施過新課程！基於以上研究結果，Gersten等人懷疑LoU的效度，他們認為如果受訪教師是個能言善辯的人和認識某新課程的話，LoU的訪談法便很易出錯了。

雖然Hall and Hord(1987)曾強調，因為LoU是一個行為的概念，測量LoU不可依賴用紙筆作答的問卷，但Holmes-Smith(1993)在澳洲成功設計了一份問卷，該問卷一共有四十八題，並組成八個量表，分別測量LoU的八個使用層次。Holmes-Smith透過問卷收集

了1,508位小學英文教師的資料，研究顯示，八個量表的信度極高，信度系數介於0.86和0.99之間。他亦應用結構方程模式(structural equation modeling)，成功證明LoU的八個使用層次是按等級劃分的。但Holmes-Smith設計問卷時只包括了Hall倡議的七個度向之其中六個。他的理由是，第七個度向「執行」涉及多類型活動，故無法編製恰當的問卷題目。

其次，若同時採用LoU及SoC去測量課程實施的話，研究結果常常不一致，令課程學家進一步質疑LoU和SoC這兩個概念的效度。例如Maloney(1993)透過訪談和問卷，分別測量了十二位教師實施某新課程的LoU和SoC，其中四位被評定為「機械地使用」，但該四位教師的SoC則可以是「資訊性」、「後果」、或「低度關注」。又Maloney評定了其他三位教師的LoU是屬於「例行化」，但其中兩位的SoC被判斷為「低度關注」，而第三位的SoC則屬於「合作」。可惜Maloney並沒有嘗試解釋以上的異常研究結果，令讀者懷疑LoU和SoC資料的信度及效度。Maloney的異常研究結果並非個別例子，Anderson(1997)也指出Marsh(1987)的研究亦出現不一致的LoU和SoC結果。

Hall和他的同事創立LoU概念後，也發覺屬於同一使用層次的教師未必運用相同形式去實施課程。因此，CBAM的發展人員再加添另一種概念，稱為「創新的形貌」(IC)。根據Heck, Stiegelbauer, Hall, and Loucks(1981)之分析，若果應用IC去測量課程實施的話，研究員首先必須找出某新課程的成分(components)。所謂成分，是指該課程的特徵，特徵可包括教師的行為、學生的學習活動和教材的使用方法等。課程成分可分為兩類，第一類成分包含所有必須被實施的課程特徵，而第二類成分是課程發展人員建議使用的，但不是必須要被落實的特徵。

辨認了所有成分後，研究員便要設法找出每種成分的度向及不同的變異形式，以便設計一份IC成分檢核表(component checklist)，清楚展示所有理想的、可被接納的、以及不可被接納的課程

實施形式和調適。最後，研究員可利用訪談法、觀察法或調查法，評定個別教師實施課程的真實運作形式。例如Hord and Huling-Austin(1987)設計了一份IC成分檢核表，做為測量實施一項以目標參照的新數學課程之工具(見表三)。該檢核表包含五個成分，並透過訪談收集了十二位教師的課程實施資料，其中一項重要發現，就是有六位教師並沒有使用補充的活動教材套。

由上述簡介可見，IC與SoC和LoU的最大差異，就是IC的成分檢核表不是一般性測量工具，它是針對某項新課程而設計的，旨在描述新課程可運作的不同形式。使用IC的研究實例比SoC或LoU較少，讀者除了可參閱Heck等人(1981)之外，還可參考Carter and Sanders(1992)、Hall(1992)、Hall and Hord(1987)、Hall and Loucks(1981)、Hord and Huling-Austin(1987)、Roitman, Gottschalk, Mayer, and Blakely(1983)、Maloney(1993)、Marsh and Kennedy(1989)等人所寫的文獻。

顯然，IC比LoU能提供較豐富的課程實施資料。例如Maloney(1993)雖然評定了四位教師的LoU屬於「機械地使用」，惟各人的IC結果都不同，其中三位的實施課程形式更被認為不能被接受。此外，研究員還可以設計一套評分準則，以便計算每位教師在各課程成分的得分，例如Carter and Sanders(1992)對「理想的使用」、「可接受的使用」和「不可接受的使用」分別給予+2、+1及-1分。

除了CBAM之外，另一種較普遍被使用的模式可算是「創新的剖面」(innovation profiles，以下簡稱IP)。這模式由Leithwood and Montgomery(1980)在加拿大創立。他們批評CBAM的LoU架構太過一般性和不能提供充足訊息，以便促進教師實施新課程，可惜Leithwood and Montgomery當時不知道Hall和他的同事在同一時間也著手構思IC概念，其實，IP與IC並沒有很大差異。若要使用IP去測量某課程的實施程度，研究員首先要找出該課程的主要度向，然後描述在各度向上教師現今的實施狀況和教師需要最終表現

表三 CBAM的成分檢核表示例

(1)	(2)	(3)	(4)	(5)
←————— 可接納的變異		不可被接納的變異 —————→		
*成分一：教材的使用				
使用所有建議的教材，包括課本及補充的活動套件	使用課本及其他教師自行編製或外間出售的教材（但不使用補充的活動套件）	只使用課本	使用補充活動套件、教師自行編製及/或外間出售的教材（但不使用課本）	
*成分二：課程目標的教授				
依照次序教授所有課程目標	教授所有課程目標，但不依照建議的次序	依照書本裏編排的次序教授目標	不依照書本編排的次序教授目標	教授其他目標
*成分三：測驗				
依照建議的時間使用通達測驗	使用通達測驗，但不依照建議的時間舉行	使用課本裏的測驗	使用教師自行編製的測驗	
成分四：利用測驗成績進行編組和檢討				
依照測驗結果進行個別化教學	更改組別及學生時常被互換組別	每組學生不變，若有需要，個別學生可獲額外支援	學生不被編成小組，若個別學生有需要，可獲得額外支援	學生不被編成小組，任何檢討是以整班為單位
*成分五：記錄存檔				
記錄圖表被張貼在牆上，資料合時	記錄資料合時，但並沒有被張貼在牆上	記錄圖表被張貼在牆上，但資料已過時	不按照課程提供的記錄系統去保存檔案	

資料來源：Hord and Huling-Austin, 1987.

註釋：不可被接納的變異位於右方，可被接納的變異則位於左方。位於最左方的是課程發展人員認為理想的變異。

*代表課程發展人員需要的必要成分。

的理想實施狀況，最後研究員必須假設某一數量的使用層次，讓教師由他們現今的實施層次遞增至理想的層次。Leithwood and Montgomery (1987) 記載了如何設計IP的詳情及一些使用IP的實例。

對持「實用的相互調適觀」的學者而言，CBAM的IC及Leithwood and Montgomery的IP似乎較為有效兼實用。但根據Cheung等人(1996b)的分析，IC和IP的其中一個弱點，在於確認課程成分或實施度向的方法。CBAM建議訪談課程發展人員、協助課程實施的人員、以及教師，從而找出設計IC檢核表的課程成分。同樣，Leithwood and Montgomery(1987)提議一組教師、校長、課程發展人員和評鑑人員召開會議，商討須要包括的課程實施度向，以便製作IP檢核表。但許多研究證實，訪談及傳統的小組討論不能有效地產生一致意見，故Cheung等人(1996b)建議通過一種特別的集體討論會找出課程成分或度向。

IC和IP的另一個弱點，就是未能提供建議如何測試實施資料的效度和如何運用三角測定。Cheung, Hattie, Bucat, and Douglas (1996c)提出了七個理由，解釋為甚麼必須包括學生的報告做為測量課程實施的主要資料，他們不僅運用了驗證性因素分析法和多特質——多方法分析法，測試了1,377香港中學生的問卷資料之建構效度，還示範如何利用學生的問卷資料，以便測定教師問卷資料和觀察資料的效度。Cheung等人建議了一個包含五個度向的理念架構，做為中學化學科校內評估實驗(school-based assessment of practical work)的測量基礎。較為突破的，就是他們設計了兩款學生問卷，隨意派發給每班學生回答。每款問卷包含六十題，涵蓋了五個度向及組成了五個量表。但在兩款問卷中，有二十三題是相同的，並透過題目反應理論(item response theory)將該兩款問卷的答案變成相同測量尺度的資料(Baker, 1992)，令每位學生不需要回答全部題目。除了學生問卷之外，教師問卷和觀課表的設計也依據課程實施度向，以便進行三角測定。Cheung等人(1996a, 1996c)的

研究結果顯示，學生的問卷資料具高信度和效度，且能協助找出教師問卷資料的可能誤差和觀察法的不足處。但他們的教師樣本只有六十三位中學化學科教師，而只有其中十位教師接受觀課。其次，受訪的1,377位學生都是中六或中七學生，較低年級的學生能否也提供有效資料的問題，值得課程學家繼續研究。

結語

如何測量課程實施的程度，是研究課程實施的重點。本文的分析顯示，許多種測量課程實施的程度之方法皆可配合「忠實觀」和「實用的相互調適觀」，如觀察、訪談、內容分析、問卷調查等，但由於這兩種課程實施觀的理念不同，測量方法相同並不暗示其測量結果將會一樣。

Wolf(1990)曾指出，「忠實觀」的假設可算是教育界中一項最幼稚的想法。因此，在九十年代，課程學家多採用與「實用的相互調適觀」一致的測量方法或模式，而直至目前為止，具有較完整理念的模式仍然是CBAM。不過，SoC、LoU和IC的效度並無保證，該三類工具還需進一步發展和改良，以便精確地測量出教師實施課程的程度。儘管「忠實觀」備受學者抨擊，在特別的情況下，其測量方法可能還有價值，譬如課程需要教師在不同的時間安排診斷性的測驗(Marsh, 1992)、需要教師教授科學實驗的一些正確程序等。因此，世上並沒有一種最佳的測量課程實施的工具，選擇測量工具應以配合研究目的和所涉及的課程實施度向的性質為前提。譬如測量的目的是要找出文件課程如何被教師調適的話，研究員當然須要採用與「實用的相互調適觀」一致的測量工具了。但無論採用那個課程實施觀，研究員都必須了解其局限性。

回顧文獻顯示，近期的課程實施研究已採用較精確的程序和資料分析法，並能針對Scheirer and Rezmovic(1983)所倡議的測量方法之五大要素，即多種測量技巧的應用、清晰界定「課程實施程度」建構、信度的嚴格測試、效度的嚴格測試、以及抽取具代

表性的研究樣本。可惜直至目前為止，還未有一個測量模式可包含上述五大要素，並提供指引讓研究員參考或遵循。雖然Cheung等人(1996a, 1996b, 1996c)的測量方法已考慮了該五大要素，但其發展尚未成熟及只有一個例證，值得對課程實施有興趣的人士進一步研究。

參考文獻

- 李子建、黃顯華(1994)。《課程：範式、取向和設計》。香港：中文大學出版社。
- 黃政傑(1991)。《課程設計》。台北：東華書局。
- Anderson, S. E. (1991). Principal's management style and patterns of teacher implementation across multiple innovations. *School Effectiveness and School Improvement*, 2(4), 286-304.
- Anderson, S. E. (1997). Understanding teacher change: Revisiting the concerns based adoption model. *Curriculum Inquiry*, 27(3), 331-367.
- Bailey, D. B., & Palsha, S. A. (1992). Qualities of the stages of concern questionnaire and implications for educational innovations. *Journal of Educational Research*, 85(4), 226-232.
- Baker, F. B. (1992). Equating tests under the graded response model. *Applied Psychological Measurement*, 16(1), 87-96.
- Berg, R. van den (1993). The concerns-based adoption model in the Netherlands, Flanders and the United Kingdom: State of the art and perspective. *Studies in Educational Evaluation*, 19, 51-63.
- Berman, P., & McLaughlin, M. W. (1976). Implementation of educational innovation. *Educational Forum*, 40(3), 345-370.
- Carter, D., & Sanders, L. (1992). Evaluating the impact of staff development interventions on the process of curriculum implementation. *Curriculum and Teaching*, 7(2), 55-68.
- Cheung, D., Hattie, J., Bucat, R., & Douglas, G. (1996a). Measuring the degree of implementation of school-based assessment schemes for practical science. *Research in Science Education*, 26(4), 357-389.
- Cheung, D., Hattie, J., Bucat, R., & Douglas, G. (1996b). Identifying the dimensions of curriculum implementation: The nominal group technique. *Curriculum Forum*, 6(1), 1-16.

- Cheung, D., Hattie, J., Bucat, R., & Douglas, G. (1996c). Students' perceptions of curriculum implementation. *Curriculum and Teaching*, 11(1), 49-60.
- Datta, L. (1980). Changing times: The study of federal programs supporting educational change and the case for local problem solving. *Teachers College Record*, 82(1), 101-116.
- Ellett, C. D. (1981). *Issues related to the evaluation of program implementation in Follow Through*. Paper presented at the National Institute of Education's Follow Through Planning Conference, Philadelphia, PA, 10-11 February. (ERIC document reproduction service no. ED 243585)
- Evans, M., & Hopkins, D. (1988). School climate and the psychological state of the individual teacher as factors affecting the utilisation of educational ideas following an inservice course. *British Educational Research Journal*, 14(3), 211-230.
- Evans, W., & Behrman, E. H. (1977). Strategy for evaluating curriculum implementation. *Journal of Curriculum Studies*, 9(1), 75-80.
- Fullan, M. (1991). *The new meaning of educational change*. London: Cassell.
- Fullan, M., & Pomfret, A. (1977). Research on curriculum and instruction implementation. *Review of Educational Research*, 47(1), 335-397.
- Gephart, W. J. (1976). *Problems in measuring the degree of implementation of an innovation*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA, 19-23 April. (ERIC document reproduction service no. ED 124584)
- Gersten, R. M., Carnine, D. W., Williams, P. B. (1982). Measuring implementation of a structured educational model in an urban school district: An observational approach. *Educational Evaluation and Policy Analysis*, 4(1), 67-79.
- Hall, G. E. (1992). The local educational change process and policy implementation. *Journal of Research in Science Teaching*, 29(8), 877-904.
- Hall, G. E., George, A. A., & Rutherford, W. L. (1977). *Measuring stages of concern about the innovation: A manual for use of the SoC questionnaire*. Austin: Research and Development Centre for Teacher Education, University of Texas. (ERIC document reproduction service no. ED 147342)
- Hall, G. E., & Hord, S. M. (1987). *Change in schools: Facilitating the process*. Albany, NY: State University of New York Press.

- Hall, G. E., & Loucks, S. F. (1977). A developmental model for determining whether the treatment is actually implemented. *American Educational Research Journal*, 14(3), 263-276.
- Hall, G. E., & Loucks, S. F. (1978). Teacher concerns as a basis for facilitating and personalizing staff development. *Teachers College Record*, 80(1), 36-53.
- Hall, G. E., & Loucks, S. F. (1981). Program definition and adaptation: Implications for inservice. *Journal of Research and Development in Education*, 14(2), 46-58.
- Hall, G. E., & Rutherford, W. L. (1976). Concerns of teachers about implementing team teaching. *Educational Leadership*, 34(3), 227-233.
- Heck, S., Stiegelbauer, S. M., Hall, G. E., & Loucks, S. F. (1981). *Measuring innovation configurations: Procedures and applications*. Austin: Research and Development Centre for Teacher Education, University of Texas. (ERIC document reproduction service no. ED 204147)
- Holmes-Smith, P. (1993). *Developing a questionnaire to gauge the level of use of an innovation*. Paper presented at the annual conference of the Australian Association for Research in Education, Perth, Australia, 21-25 November.
- Hord, S. M., & Huling-Austin, L. (1987). Curriculum implementation: How to know if it's there. *Journal of Rural and Small Schools*, 1(3), 23-26.
- James, R. K., & Hall, G. (1981). A study of concerns of science teachers regarding an implementation of ISCS. *Journal of Research in Science Teaching*, 18(6), 479-487.
- Kerr, D. M., Kent, L., & Lam, T. C. M. (1985). Measuring program implementation with a classroom observation instrument: The interactive teaching map. *Evaluation Review*, 9(4), 461-482.
- Kimpston, R. D. (1985). Curriculum fidelity and the implementation tasks employed by teachers: A research study. *Journal of Curriculum Studies*, 17(2), 185-195.
- Kimpston, R. D., & Anderson, D. H. (1988). Factors affecting teachers' and principals' stages of concern over carrying out benchmark testing. *Journal of Curriculum and Supervision*, 3(4), 321-334.
- King, J. A., Morris, L. L., & Fitz-Gibbon, C. T. (1987). *How to assess program implementation*. Newbury Park, CA: Sage.
- Leithwood, K. A., & Montgomery, D. J. (1980). Evaluating program implementation. *Evaluation Review*, 4(2), 193-214.
- Leithwood, K. A., & Montgomery, D. J. (Eds.). (1987). *Improving classroom*

- practice using innovation profiles*. Toronto: The Ontario Institute for Studies in Education.
- Loucks, S. F., Newlove, B. H., & Hall, G. E. (1975). *Measuring levels of use of an innovation: A manual for trainers, interviewers and raters*. Austin: Research and Development Centre for Teacher Education, The University of Texas.
- Maloney, C. (1993). Implementing curriculum: A case study. *Curriculum Perspectives*, 13(3), 23-32.
- Marsh, C. J. (1985). Implementation of a high school geography syllabus: Issues and applications. *Educational Research*, 27(1), 30-39.
- Marsh, C. J. (1987). Implementation of a social studies curriculum in an Australian elementary school. *Elementary School Journal*, 87(4), 476-486.
- Marsh, C. J. (1992). *Key concepts for understanding curriculum*. London: The Falmer Press.
- Marsh, C. J., & Kennedy, K. J. (1989). Evaluating the implementation of federally sponsored English as a second language (ESL) curriculum materials. *Studies in Educational Evaluation*, 15, 109-139.
- McLaughlin, M. W. (1976). Implementation as mutual adaptation: Change in classroom organization. *Teachers College Record*, 77(3), 339-351.
- Morris, P. (1984). Curriculum innovation and implementation: A South East Asian perspective. *Curriculum Perspectives*, 4(1), 3-18.
- Newfield, J. (1980). Self reports and matrix sampling: A method of measuring fidelity of implementation. *Studies in Educational Evaluation*, 6, 149-155.
- Revicki, D. A., Rubin, R. I., & Stuck, G. B. (1981). Models for measuring programme implementation: A review and critique. *Journal of Curriculum Studies*, 13(1), 65-67.
- Roitman, D., Gottschalk, R., Mayer, J., & Blakely, C. (1983). Implementation of social program innovations in public sector organization: A test of the modified RD and D model. *IEEE Transactions on Engineering Management*, 30(2), 68-75.
- Sabar, N. (1986). Is the disappointment in the implementation of a new science curriculum justified?: A study quantifying descriptive data. *Journal of Research in Science Teaching*, 23(6), 475-491.
- Scheirer, M. A., & Rezmovic, E. L. (1983). Measuring the degree of program implementation: A methodological review. *Evaluation Review*, 7(5), 599-633.

- Snyder, J., Bolin, F., & Zumwalt, K. (1992). Curriculum implementation. In P. W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 402-435). New York: Macmillan.
- Treagust, D. F., & Rennie, L. J. (1993). Implementing technology in the school curriculum: A case study involving six secondary schools. *Journal of Technology Education*, 5(1), 38-53.
- Wang, M. C., Nojan, M., Strom, C. D., & Walberg, H. J. (1984). The utility of degree of implementation measures in program implementation and evaluation research. *Curriculum Inquiry*, 14(3), 249-286.
- Whiteside, C., & James, R. K. (1986). Utilizing teachers' concerns to improve microcomputer implementation. *Computers in the Schools*, 2(4), 29-41.
- Wirt, J. G. (1976). Implementing diagnostic/prescriptive reading innovations. *Teachers College Record*, 77(3), 352-365.
- Wolf, R. M. (1990). *Evaluation in education: Foundations of competency assessment and program review*. New York: Praeger.
- Yin, R. K. (1982). Studying the implementation of public programs. In W. Williams (Ed.), *Studying implementation: Methodological and administrative issues* (pp. 36-72). New Jersey: Chatham House.

Measuring the Degree of Curriculum Implementation

Derek CHEUNG Sin-pui

Abstract

This article reviews and critically analyzes methods of measuring the degree of curriculum implementation with reference to the fidelity perspective and the practical mutual adaptation perspective. The focus is on the characteristics and limitations of various methods, as well as the directions for further research. Review of the literature indicates that although proponents of the fidelity perspective and the practical mutual adaptation perspective use similar methods of collecting data, such as observation, interview and questionnaire survey, their assumptions and outcomes of measurement are different. Many past studies also neglected the dimensionality of the curriculum implementation construct, collection of data from multiple sources, and examination of the reliability and validity of data through rigorous statistical techniques.