

(5) 分析而不解說因果關係

只對數量關係進行分析，因果關係的解說由當地的管理者、研究者和教師進行解說，將更能因地制宜，進行適切的診斷，能更準確地判斷自己地區或學校各學科的整體水準及優弱勢，得以有針對性地進行改進和指導。

(6) 即時回饋

各地教育行政部門方能依據測試結果，即時展開針對性的對策，進行課堂教學改革活動。

(7) 提供學生學業品質分析報告

提供呈現學生在各學科上的總體表現及其在不同學科內容領域和能力維度上的優勢和不足，並分析了學生學業成績和學習環境因素的關係的學生學業品質分析報告，方利於課堂針對性教學。

(8) 試題內容生活化

日本學力調查，將試題除「知識」外，同時將「活用」列入試題，「活用」的試題主要考查學生能否將知識與技能等活用於實際生活中的能力以及解決課題的構思、評鑑、改進等的的能力。

(9) 檢測認知發展水準

結合索羅分類法 (SOLO Taxonomy) 設計題目選項，以測驗的方式瞭解學生解決問題時表現出的認知發展水準，更精細區分學生能力。

四、本建議方案之思路與國際趨勢

在前文，本研究已詳盡描述國內相關的品管方案(例如：攜手計畫)與美國、英國、日本等國家之學力品管架構，而面對如此多元與豐富的資訊，要如何有效的規劃或者從哪裡開始執行?則變成一個重要的課題。研究團隊在依循 Shavelson, McDonnell, Oakes, Carey 與 Picus(1987)的想法下，以他們所認為監控系統的資訊來源是包含下列五種，作為參考建議，更指引本研究建置學力品管之思路，作為未來執行的方向，陳述如下：

(一) 採用現行可得資料(Status quo)

對於監控系統的資料來源，Shavelson, McDonnell, Oakes, Carey 與 Picus(1987)初步的想法很簡單，他們認為如果現行的資料已經能夠滿足或回答所需的問題時，研究者可直接採用，不必另行開發或建置，但這方面的資料，來源可能非常複雜(或時常變動)，必須視當下時間，所能夠獲得哪些訊息為主。

(二) 機制的整合(Patchwork)

Shavelson, McDonnell, Oakes, Carey 與 Picus(1987)所提出的另一項選擇，是

認為研究者能考慮有效的整合目前現存的各式品管機制，而這個概念是有點像將上述可能雜亂無章的資訊來源，以系統化、制度化的方式，進行整合的動作，亦即只納入定期性出版或調查的方案，例如：研究者能考慮整合 NAEP(National Assessment of Educational Progress)、TIMSS 與 PISA 等方案，以形成另一種整合性品管監控機制。

(三)週期性的研究(Cyclical studies)

在上述方案中，Shavelson, McDonnell, Oakes, Carey 與 Picus(1987)認為可能受限於時間、經費等因素，導致每年能搜集到的訊息，會很表面或簡略，因此，他們認為可以透過週期性的研究，分年度針對不同的主題進行更深入的訊息搜集，例如：在過往時，研究者可能受限於某些因素，以致無法同時(或深入)針對學生成就、教師或學校背景變項等逐一進行評量與調查，但利用週期性的研究模式，研究者可考慮將時程切割成幾年完成，在第一年，研究者可只針對學生成就進行評量(例如：採用實作評量取代紙筆評量，以獲得更深入訊息)，而第二年，再執行教師、學校背景訊息之調查等，如此，研究者每年就能以單獨、較充裕的資源深入調查所需的訊息。

(四)援用現行方案(Piggyback)

這項方案主要是 Shavelson, McDonnell, Oakes, Carey 與 Picus(1987)對於上述機制整合的另一種想法，他們認為當各方案無法整合(或整合難度過高)時，研究者是可以單獨採用(或必要時，加以擴增)一項完整的方案，例如：研究者能考慮只採用 NAEP 評量方案，以作為品管機制，但其中，為滿足自身所需，可視情況加以作些微的調整。

(五)獨立方案(Independent)

Shavelson, McDonnell, Oakes, Carey 與 Picus(1987)的最後一項建議，就是在上述訊息來源，都無法滿足研究者所需時，可以考慮自行建置一套獨立的品管機制，但此時研究者不僅需要衡量其必要性外，還必須審慎考量要如何重新建置之問題。

本研究團隊綜觀國內現有相關方案與機制後，認為臺灣目前不論是自行建置或國際的參與，都存在著學生接受過多的測驗、評量之疑慮，加之各縣市、中央單位仍不斷的推出新式品管方案，致使現行可用的訊息或資訊，可謂十分豐富，因此，團隊成員認為實無再獨立建置之必要性，而在進一步思考其它可行的層面時，研究者認為若以機制的整合為考量時，實務上，是有其執行的難度，主因在於現存的評量機制，不論在目的、決策單位、未來規劃上，多存在歧異性。綜合上述，研究團隊會是以援用現行方案，再加以擴充、修訂為建議目標。

在確立執行的大方向後，其細節的規劃，則是本節下一步描述的重點，而在上述的文獻探討中，研究者大致能獲得品管機制的實體架構，但對於其意涵，則

有必要預先加以陳述。研究團隊是根據 Koretz(2008)、Koretz 與 Hamilton(2006)等人對於國際趨勢與潮流的剖析，作為借鏡，而在本建議方案中融入這些想法，其中，大致能歸納出以下幾個方向，陳述如下：

(一)增加教育人員與學生的利害關係

1980 年左右，對於教育成效的評定，並不是那麼倚賴在高風險測驗 (high-stakes exams)，但 Koretz 與 Hamilton(2006)認為近年則逐漸朝向以測驗結果檢視成效為導向，導致利害關係(stake)之重心移轉至教育人員與學生身上，間接使得教育人員跟學生所必須承受的壓力增加，其中，最重要的例子，就是美國 NCLB 要求學校每年都要有足夠的進展(AYP)。

(二)發展多元題型

由於成本與便利性等因素的考量，使得選擇題型在 20 世紀大型測驗中所運用的比率，始終高居不下，但 Koretz 與 Hamilton(2006)認為在 1980 年末與 1990 年初，有另外一項呼聲不斷的傳入，那就是許多教育學者與改革者開始強調使用實作評量(performance assessment)作為檢定學生能力的工具，因為，他們希望能評量較複雜、進階的知識或技巧，同時，期望著能推升評量引導教學之功能，在如此訴求下，傳統選擇題型所扮演的影響力，也逐漸受到影響，例如 NAEP 在 1996、2000 與 2003 年數學評量中，也納入近一半的建構反應試題(constructed response items)(NCES, 2004)，至今，雖然受限於經費因素，許多測驗始終是以選擇題型為首要，但實作評量或其它多元題型之訴求，仍不斷受到重視。

(三)依賴以標準為基礎的測驗結果呈現方式

在 1980 年代後期，決策者與教育學者開始訴求建立一套應用在所有學生的標準，以期能向學生、教師溝通所期望的成就水平，而這項想法不久即獲得美國聯邦立法同意，法案 The Goals 2000: Educate America Act of 1994 的成立，就促成內容標準(content standard)與表現標準(performance standard)的使用，接續，在由 National Council on Education Standards and Testing(NCEST)規劃進行(NCEST, 1992)。在這項改革下，研究者對於測驗結果呈現亦逐漸由傳統常模參照模式(norm referenced)，轉變成以標準參照(standard referenced)的解讀方式，強調在某個水平上，學生學會了什麼知識、技巧或能力，至今，例如 NAEP 就已完全採用標準參照的方式，進行測驗結果之詮釋(請見網站 The Nation's Report Card：<http://nationsreportcard.gov/>)。

(四)客製化的測驗與標準模式

對於內容標準與表現標準的建立，所強調的是由各州自行發展，Koretz 與 Hamilton(2006)認為此舉直接促使著各州對於客製化測驗的需求，他們期望能適切的聯結課程、測驗與標準之關係，以進行學生測驗結果之詮釋。而此項進程，

在伴隨著 NCLB 的鼓吹下，更是快速的融入於各州的活動中。

(五)納入學習不利學生的需求

近年來，隨著大型評量之推動，對於標準或施測流程的一致性要求，亦逐漸的推升，但其中，對於學習障礙學生或英語受限的學生，是否也具備同樣要求，則變成另一個關注重點。在隨著美國 Title I of the Improving America's School Act of 1994 立法的成立，強調接受經費補助的學校，都必須將學習不利的學生納入評量架構後，間接促使學校開始針對不同需求之人士，發展適合或需經修訂、調適的測驗工具，接續，加上 NCLB 的推展，又重新要求在測驗上，要呈現出不同種族、社經、學習不利與語言受限等四個群體之結果，再再促使納入這群學習不利學生的需求。

(六)強調群體的測驗結果報告

伴隨著以測驗為基礎的責任績效系統建立，其內容所強調的都是以學校層級的獎賞或懲罰為核心，因而，Koretz 與 Hamilton(2006)認為間接促使各州亟欲建立一套能提供學校或年級層級等群體的測驗結果解讀系統，以彰顯其執行成果，而此舉亦逐漸推升對於群體測驗結果解讀之重視。

(七)強調報告學生學習成就的進步量

在早期，對於學生、學校表現之評估，多是以跨區域(cross-sectional)的資料比較為基礎，但時至今日，Koretz 與 Hamilton(2006)認為注意力已開始轉移，逐漸強調著學生隨時間表現或分數的進步，例如：在大型的 K-12 評量中，則多是以監控學生改變為目標。而這種改變的呈現方式，除了 NCLB 所採用的世代比較模式(cohort to cohort model)，也可以是以縱貫性研究(longitudinal approach)或稱增值模式(value added model)，進行個別學生長期性的追蹤，並評量其成長，亦或是採用準縱貫性研究(quasi-longitudinal approach)(Linn, 2000)，例如：以今年某學校四年級學生的平均分數，對照該校前一年三年級學生的平均表現，以進行改變量的監控。但不論採用何種研究方法，可以肯定的是決策者對這項趨勢的重視，是不斷的在增長中。

綜合上述，大致可以發現目前國際的潮流與趨勢，不外乎是上述幾點要件，而研究團隊就是在這些想法的導引下，規劃本研究建議之學力品管架構與細節，茲將詳細流程陳述如後。

五、本研究初步建議方案

從上述國內外現行的品管機制或措施中，大致可從其經驗中，粹取出學力品管機制中幾項重要的元素，包含有受輔對象的篩選、評量工具、基本學力標準、教學人員、配套措施、補救教學、成效考核工具等等，而此要素即為本研究建議方案之組成元素，在上述理論、實務基礎下，架構出整個面向，進一步詳細說明