

最近國內在推展全國性的教育評量。這是一件很重要的工作，許多國家，如美國和日本，都積極的在做，因為設計週全的評量，可以發揮多種功能，包括：

- (一)提供教育成果指標，讓國人瞭解國內學生學習的程度及變化。
- (二)鑑定各地區、各級學校以及各類學生的學習差距。
- (三)探究與學習成果差距有關的種種學校、家庭

要使教育評量發揮其真正的功能，必先從設計上及工作推展方針上下功夫。依據美國二十多年來的經驗，以及最近發展，有幾項方針值得參照，特簡述如下：

- (一)廣徵各方意見並達成共識，美國評量工作參與的人很多，無論在訂定評量目標及內容規範，在試題製作及審訂，或在評量結果分析與報告，都有委員會負責。這些委員會的成員，依工作性質而定，上自州長、州議員，

如何設計及推展教育評量，使其發揮改進教學的功能？

彭森明

、社會因素。

這些功能表面上看起來是在瞭解國內教育狀況，為教育做「把脈」的工作，但實際上可以用來促進教育與課堂教學的改革，因為瞭解狀況之後，可以進一步去探討教育的缺失；知道缺失之後，會進一步去做改進的工作。這是很自然的現象，也是很自然的期望。所以說教育評量的目的，起於瞭解教育，終於改進教育。

不過話說回來，要教育評量能達成這個目標，還不是一件容易的事。因此「如何設計及推展並運用教育評量，使其達成改進教育與教學的目標」是一個值得討論的問題。本文將分別從「評量設計及推展方針」、「評量結果分析與報導方式」，及「評量結果之應用」三方面來探討，提出個人的意見與經驗，供先進們做參考。

一、評量設計及推展方針

下至一般民眾，都有代表參與討論，當然學者專家，如測驗專家、課程專家，以及教育行政主管、校長、教師等，都分別有代表參與工作。這種方式，一方面可以使評量設計週詳，以免偏差，一方面可以讓各方真正瞭解評量的意義與目的，讓評量工作為大眾所接受，尤其是讓校長及教師所能免除疑慮，轉而熱心支持，有效運用評量結果，做改進教學之用。

- (二)評量內容要以教學目標與課程標準做規範。這是一項根本方針，否則評量會有偏差。因為評量內容，如能力層次分類及各學科內容範圍，會影響教學重點以及教材教法的選擇，所以如果評量不依照教學目標與課程標準來規範，這種評量就無法找出教育與課堂教學真正缺失，甚至會引導教育走向歧途。猶如目前升學考試試題一樣，假若只注重記憶，而缺乏思考與判斷能力的評估，學校教學為了符合升學需要，也就會偏重「死記」、





「死背」的方式，使教育無法發揮應有的功能。

因此詳細規劃教學目標與課程標準，然後決定評量內容取樣設計，是評量的第一步工作。比如美國的數學學科成果評量（Mathematics Assessment），即有全國委員會先分析各州數學課程內容與標準（即學生應學些什麼，達到何種程度），並參照由「全國數學教師協會」所起草制定的「學校數學課程與評鑑標準」，然後訂定評量規範（framework for assessment）。目前評量規範是由數學能力（mathematical abilities）與科目內容（content area）的分類，組合而成。數學能力分三層次：(一)數學觀念的認知（conceptual understanding）；(二)選擇與應用數學程序的能力（procedure knowledge）；(三)思考、分析與解決問題的能力（problem solving）。數學科目內容則包括(一)數字與計算（numbers and operations）；(二)測量（measurement）與幾何（geometry）；(三)資料分析（data analysis），統計與或然率（statistics and probability）；(四)代數（algebra），三角及微積分（functions）。由三層能力與四種科目內容組成十二項規範，內分細目，然後再決定每項規範中應有的試題。如此製訂的評量內容，非常廣泛，包羅所有課程，也包括各種能力，因此不致於遺漏任何重要教學目標。

這種方式常遇到的問題是：內容廣泛，要很多試題方能涵蓋。若要每位學生都去做這些問題，要花很長的時間，不切實際。不過評量的目的是評量整體教育成果，或整個學區或學校的成效而非鑑定學生個人成績，所以學生每人的測驗題目不需要完全一樣。依抽樣原則，可以有系統的將試題分配給各組學生去做答。如此每位學生花的時間不會太長，而所有學科

內容與能力層次都可以衡量了。目前美國評量即以此方式進行。

(三)試題方式，不應只限於是非、選擇等測驗題，應包括問答或做報告等方式，以評量學生思考過程與程序及綜合、分析、判斷、應用、能力表達等。這些內容規範會直接間接地影響教師課堂教學方式。美國以前為了評分作業方便，評量試題都以測驗題方式為主。後來發現這種方式無論如何改進，都很難測出學生思考過程，更難探測學生綜合與應用所學的能力，所以美國近年來的評量，包括大專入學用的SAT考試，都增加其他方式，以便真正測出各種層次的能力。

(四)評量除了測驗學生之外，應附帶收集有關學生家庭背景、學校環境、教師經歷與課堂教學重心與方法等，以供深入瞭解各類學生之間與各個學校之間的差異，以及影響這些差異的種種因素，目前美國做語文、數學及科學學科評量時，都有學生問卷、學校問卷，以及教師問卷調查，以收集一些有關資料供對評量結果的解釋，進一步分析，以及探究改進教學的根基。

二、評量結果分析與報導方式

評量結果的分析與報導應以「有助於改進教育與教學」為依歸。

一般常見的報導是在顯示今年學生成績程度如何，那個地區或那些學校的成績最好，或最差，這種方式對學生素質不高的學校或社會環境較差的學區，會產生壓迫感，而且往往忽略了或淹沒了這些學校與學區的真正成就，因為好學生（因家庭或其他因素造成）多的學校，即使校長與老師不花很大工夫，也很可能表現得比其他學校優越。因此評量結果的分析與報導，應脫離此種方式，以免失去了真正的目



標。

什麼樣的分析與報導方式才有助於教育與教學改進呢？目前美國聯邦及州級評量，有下列報導方式，值得參考。

(一)把評量結果與教學目標（標準）並列一表，以便比較目前學生程度與期望的，或應具有的標準，以瞭解今日教育是否與理想相差甚遠。如下表所示。美國馬利蘭州，蒙哥馬利郡的「學校學生表現成果評量」（School Performance Assessment Program）的結果，顯示此郡三年級的學生沒有一科達到標準，比如在閱讀方面（reading），所公訂的標準是：25%的學生要達到「優等」成績（excellent），70%要達到「合格」成績（satisfactory）。但1992年的結果是，只有3.4%的學生達到「優等」，和41.6%的學生達到「合格」程度，其他學科也一樣，表示還需要繼續努力。

表一 馬利蘭州，蒙哥馬利郡全部三年級學生表現成果

科目	公定標準		1993年成果	
	優等	合格	優等	合格
閱讀	25	70	3.4	41.6
數學	25	70	2.9	41.3
社會科	25	70	4.4	48.1
科學	25	70	3.9	46.7

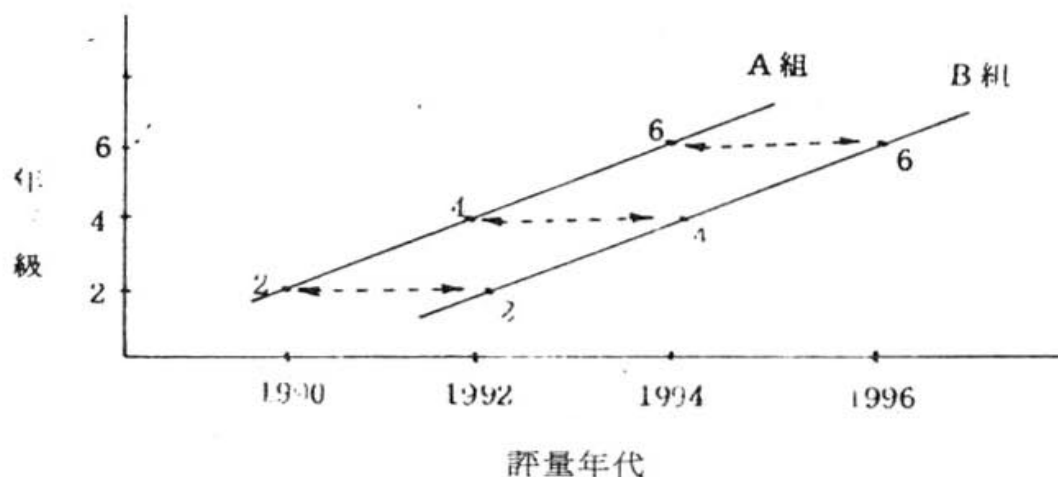
資料來源：Maryland State Department of Education, Maryland School Performance Report, 1993, P.42.

上表是以全郡為單位做統計。當然也可以為各學校或各類學生以及各部分地區製作，其目的是讓人瞭解學生程度與預期目標的差距，當然這個差距與公定標準的高低有關，因此訂標準時還得大費週章，不過一般人都認為標準不妨訂高些，讓大家都共同努力的目標。

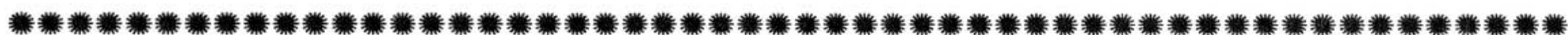
(二)與過去評量結果並列，供做比較，這種方式是鼓勵自己與自己比較，看看是否有進步，

在那一方面有進步，或在那一方面要特別加強。這種方式也比較適合評量各個學校的成果，因為各個學校的學生家庭環境不一樣，學生素質也不一樣，學校與學校之間的比較，會受這些因素的影響，因此依照學生成績訂學校之優劣是不公平的，但如果學校各自和自己過去的成績比較，則可顯示學校是否有進步，進步的程度應是考量學校師生表現的最好指標之一。

不過這種比較也不能完全忽視學生素質與環境的變化，比如比較前年九年級與今年九年級學生的成績，雖是同一學校，但這兩級學生的素質，也不一定完全相似，所以報導評量結果時，也應同時報導相關學生背景資料，以便增加瞭解狀況或程度變化。另外可能的話，評量可採用同樣學生在幾個不同階段接受測驗。如下圖所示，A組學生從1990年起在同一學校經過二、四及六年級的測驗。其結果可以前後比較，而且也可以和B組學生比較。



(三)評量結果，依內容細目，分類報導。為了瞭解那一層次的能力，那一科目需要加強，評量結果需詳細分析，比如前述數學評量規範，由三種能力，四種科目，組成十二項大綱，另在每一項大綱之內，還可以細分，因此評量若依此規範分析並報告，則很容易瞭解那些地方需要加強，對決策單位以及學校老師都很方便。



三、評量結果的應用

教育決策當局應積極採取各項步驟，應用評量結果，去改進教育。以美國馬利蘭為例，有下列方式來輔導學校改進教學措施，以提高學生學習成果。

(一)依據評量結果，鑑定那些學校成績沒有進步，甚至退步，然後要求這些學校提出改進辦法，並研討成績不良的原因，讓學校在一段時間內，自行努力，如果過了時限，仍舊表現不良，則州教育廳有權接管，由專家機構，如大學教育學院或私人顧問公司去接辦，這些機構行政效率高，受接管的學校，往往在短期內就有顯著變化，甚受家長與學生歡迎。

這種方式，在國內未有先例，也許行不通，不過教育主管單位特別加重輔助不良學校，如改進教學環境與設備，提供專家人員等，應該是可以做到的。

(二)向大眾報告評量結果，指出今後努力方向，鼓勵地方，以及家長父母合作與協助，並參與學校改進工作。

另外由美國聯邦教育部主持的全國性評量，近幾年來，特別加強考量各項教學方法，師資，以及其他可能影響學生學習的因素，供教育人員做改進教學之參考。還有主持評量工作的單位，亦有系統的將試題範例，以及各項能力定義及標準與評量結果一起公佈，讓人們，尤其是學校人員，真正瞭解評量的內涵，並提供具體改進規範。比如說學生數學應用能力欠佳，教師可以從報告中參照此項能力標準與定義，並可依照試題範例，製作相當教材在課堂內，加強此項教學與訓練。

四、結語

總之，教育評量，猶如個人健康檢查，是一項重要的工作，不僅可以瞭解狀況，也可以適時檢定缺失，以便改進。所以評量設計，得慎重周詳，不僅內容要涵蓋所有教學目標與課程標準，而且一切作業方針應以改進教育為依歸，果能如此，國內教育評量當為大眾所歡迎與熱烈支持。

參考資料

- John A. Dossey, Ina V.S. Mullis, Mary M. Lindquist, Donald L. Chambers, The Mathematics Report Card: Are We Measuring Up? (Princeton, NJ: Educational Testing Service, 1988) .
- Lee Jones, Ina V.S. Mullis, Senta A. Raizen, Inisk. Weiss Elizabeth A. Westen, The 1990 Science Report Card: NAEP'S Assessment of Fourth, Eighth, and Twelfth Grades (Washington, D.C.: U.S. Department of Education, 1992) .
- Ina V.S. Mullis and Lynn B. Jenkins, The Reading Report Card, 1971-88 (Princeton, NJ: Educational Testing Service, 1990) .
- Maryland State Department of Education, Maryland School Performance Report, 1993 (Baltimore, M.D.: Maryland State Department of Education, 1993) .
- (作者：紐約州立大學博士，現任職於美國聯邦教育署國家教育統計中心)

