

教育測驗與評量

涂金堂 著

 三民書局

Education

效度、信度、標準分數等，都涉及到數字的運算，本書提供實際的計算方式，讓學習者更清楚每個概念的意涵。

3. 近年來教學評量強調多元評量，本書亦針對其中的實作評量、檔案評量與情意評量，提供豐富的實際範例，協助讀者創發出屬於自己的多元評量。



Education



ISBN 978-957-14-5150-3

(521)



9 789571 451503

NT. 380



國家圖書館出版品預行編目資料

教育測驗與評量 / 涂金堂著. — 初版一刷. — 臺北市：三民，2009
面；公分
參考書目：面
ISBN 978-957-14-5150-3 (平裝)
1. 教育測驗 2. 教育評量

521.3

98001900

◎ 教育測驗與評量

著 作 人	涂金堂
責任編輯	黃麗瑾
美術設計	陳宛琳
發 行 人	劉振強
著作財產權人	三民書局股份有限公司
發 行 所	三民書局股份有限公司 地址 臺北市復興北路386號 電話 (02)25006600 郵撥帳號 0009998-5
門 市 部	(復北店) 臺北市復興北路386號 (重南店) 臺北市重慶南路一段61號
出版日期	初版一刷 2009年2月
編 號	S 521050

3與6 教育測驗與評量

涂金堂 著

7

自序

教學評量是教師在教學活動中，常需進行的一項重要活動。適切的評量活動能夠增進學生的學習動機；不適切的評量活動則可能對學生造成無法彌補的傷害。因此，正確且清楚掌握教育測驗與評量的概念，是每一位教師或師資職前教育的學生，應該努力追求的目標。美國教師聯合會(American Federation of Teachers)，即曾於1990年提出一位適任的教師，應具備選擇適合其教學決策的評量、發展適合其教學決策的評量、進行施測與評分、善用評量結果、發展有效的同儕評分程序、與家長溝通評量結果、分辨不適當評量方法等七項有關對學生進行教育評量的能力標準。

當前中小學對於學生學習成果的評量，仍是以教師自編的成就測驗為主，如何編製一份適切的成就測驗，成為一位中小學教師極為重要的專業能力。教師若想編製一份適切的成就測驗，必須對教育測驗與評量的意涵與目的、教育測驗的編製歷程、挑選反應試題與建構反應試題的編寫、試題的分析、效度與信度的考驗、測驗結果的解釋等概念，有清楚的掌握，才容易達成。針對這些概念，本書有專章介紹，希望協助讀者更瞭解教育測驗的編製歷程。

近年來，多元評量觀念的興起，促使許多中小學的教師，除了採用傳統的紙筆評量方式，也會採用實作評量、檔案評量、情意評量等多元評量的方式，來評量學生的學習成果。為了讓讀者瞭解多元評量的實施方式，本書也針對實作評量、檔案評量與情意評量等部分，提供許多實際的範例，希望讀者閱讀之後，能更清楚如何推動多元評量。

本書能順利出版，要特別感謝三民書局編輯群的編排與校正。最後，我要感謝我的家人，對於我寫書期間，所給予的包容與體諒。

涂金堂 謹識

2009年2月

自序

第一章 教育測驗與評量的意義

- 001 第一節 教育測驗與評量的意涵與目的
- 006 第二節 評量與教學、評量與學習的關係
- 009 第三節 測量量尺的種類
- 013 第四節 教育評量的新趨勢
- 015 第五節 總 結
- 015 第一章 習 題

第二章 教學目標對教學評量的重要性

- 017 第一節 教學目標的功能與種類
- 024 第二節 認知領域教學目標的分類
- 032 第三節 教學目標的編寫
- 036 第四節 教學目標對教學評量的影響性
- 038 第五節 總 結
- 039 第二章 習 題

第三章 測驗的編製歷程

- 041 第一節 決定測驗編製的目的
- 044 第二節 發展測驗的細目表
- 047 第三節 確定測驗的題型
- 048 第四節 試題編寫與測驗編排
- 055 第五節 測驗的施測與測驗結果的評估
- 057 第六節 總 結
- 057 第三章 習 題

第四章 挑選反應試題

- 059 第一節 是非題的性質

- 060 第二節 是非題的編寫原則
- 066 第三節 選擇題的性質
- 069 第四節 選擇題的編寫原則
- 083 第五節 配合題的性質
- 084 第六節 配合題的編寫原則
- 088 第七節 總 結
- 089 第四章 習 題

第五章 建構反應試題

- 091 第一節 填充題的性質
- 092 第二節 填充題的編寫原則
- 095 第三節 論文題的性質
- 098 第四節 論文題的編寫原則
- 103 第五節 總 結
- 104 第五章 習 題

第六章 試題的分析

- 105 第一節 難 度
- 113 第二節 鑑別度
- 118 第三節 試題分析的程序
- 121 第四節 選擇題的選項分析
- 124 第五節 總 結
- 125 第六章 習 題

第七章 效 度

- 127 第一節 效度的意涵與特性
- 130 第二節 內容效度
- 138 第三節 效標關聯效度
- 141 第四節 建構效度

149	第五節	影響效度的相關因素
151	第六節	總 結
152	第七章	習 題

第八章 信 度

153	第一節	測量誤差的概念
155	第二節	信度的意涵與特性
157	第三節	常模參照測驗的信度估計
173	第四節	效標參照測驗的信度估計
176	第五節	測量標準誤
180	第六節	影響信度大小的因素
183	第七節	總 結
183	第八章	習 題

第九章 測驗結果的解釋

185	第一節	常模參照測驗的分數解釋
190	第二節	發展性常模
193	第三節	組內常模
205	第四節	效標參照測驗的分數解釋
206	第五節	總 結
207	第九章	習 題

第十章 實作評量

209	第一節	實作評量的意涵與性質
213	第二節	實作評量的作業類別
216	第三節	實作評量的評分規準
222	第四節	實作評量的實施流程
229	第五節	實作評量的優勢與限制
232	第六節	總 結

232 第十章 習 題

第十一章 檔案評量

233 第一節 檔案評量的意涵與性質

237 第二節 檔案評量的內容與類型

241 第三節 檔案評量的評分規準

243 第四節 檔案評量的實施流程

252 第五節 檔案評量的優勢與限制

255 第六節 總 結

255 第十一章 習 題

第十二章 情意評量

257 第一節 情意評量的重要性

259 第二節 軼事記錄

261 第三節 同儕評量

269 第四節 自陳式評量

273 第五節 總 結

274 第十二章 習 題

參考書目

附錄一 常態分配表

附錄二 基本統計方法的介紹

索 引



第一章 教育測驗與評量的意義

本章主要是介紹教育測驗與評量的基本概念，包括教育測驗與評量的意涵與目的、評量與教學和學習的關係、測量量尺的種類，以及教育測驗與評量的發展趨勢等。讀者若能清楚理解這些基本概念，將有助於其他章節的學習。



第一節 教育測驗與評量的意涵與目的

本節分成兩個部分，分別介紹教育測驗與評量的意涵，以及教育評量的目的。

一、教育測驗與評量的意涵

在教室的現場，一位教師每天的例行工作項目，除了教學活動占據較多的時間之外，教學評量的工作，也占了許多教師不少的時間。根據 Stiggins 和 Conklin (1992) 的研究指出，教學評量活動大約占了教師三分之一的上課時間，顯見教學評量活動是每位教師一項很重要的工作。美國教師聯合會 (American Federation of Teachers) 曾於 1990 年提出一位適任的教師，應具有下列七項有關對學生進行教育評量的能力標準：

1. 教師應有能力選擇適合其教學決策的評量方式。
2. 教師應有能力發展適合其教學決策的評量方式。
3. 教師應有能力施測、評分及解釋外在決定或教師決定的評量方法之評量結果。
4. 當教師在對個別學生、教學計畫、發展課程，以及學校革新建議等方面做決策時，應有能力善用評量結果。

5. 教師應有能力透過同儕評量方式，發展出有效的同儕評分程序。
6. 教師應有能力與學生、家長、其他人、其他教育者，溝通評量結果。
7. 教師應有能力辨別出不合倫理、不合法，以及其他不適當的評量方法與評量資料的使用。

上述的這七個能力標準，主張教師應該能選擇和發展適切的評量方法、能進行測驗的施測、能進行測驗的評分、能正確的解釋測驗結果、能採用同儕評量的方式、能向別人溝通評量的結果、能辨別評量方式是否受到誤用等能力。這些能力對一位教師確實是非常的重要，因為它們是教師在課堂上經常會使用到的能力。

以下有幾個關於教育測驗與評量的活動，經常發生在不同學習階段的班級中。

- 
- ❖ 趙老師是一所幼稚園的老師，她正仔細觀察一位大班的小朋友，能否根據作業單上指定的數字，正確無誤的排列出相同數量的硬幣。
 - ❖ 李老師是一所國小五年一班的級任導師，昨天數學課的教學單元是「公因數與公倍數」，下課之前，即預告今天數學課一上課，就會進行小考。李老師昨天晚上很認真的出了一份十道數學題目的考卷，考題是有關如何找出兩個數值的公因數與公倍數，他打算今天一開始上數學課，即來個小考，看看學生是否熟悉「公因數與公倍數」這個單元的內容。
 - ❖ 王老師、林老師、陳老師分別是同一所小型國中，一年一班、一年二班與一年三班的英文科任老師。由於這三個班級學生的英文程度參差不齊，因此，在學期開始上課之前，三位老師即決定共同採用學科能力分班的上課方式，來解決各班學生英文程度參差不齊的教學困擾。由王老師負責教導英文程度較佳的學生，林老師教導英文程度中等的學生，而陳老師則是教導英文程度較差的學生。為了有效的進行學科能力分班，三位老師編製一份「國中一年級英文測驗」，在開學的第一堂課，即以這份測驗，對這三

個班級的學生進行施測，並以該份測驗的得分高低，作為學科能力分班的依據。

- ❖ 蕭老師是高一的國文老師，今天學校舉行第一次段考，早上第一節蕭老師得去監考國文，考試的範圍包含第一單元至第五單元。在監考的過程中，蕭老師注意到平常上課很認真的同學，寫考卷寫得很順暢，而那些上課較不專心的同學，常有不知如何下筆的困窘情形發生。

上面的四個例子，分別發生在幼稚園、小學、國中與高中的教學情境中。各級學校的教師在每天的教學過程中，都可能會接觸到評量活動，顯見教學評量對教師而言，是一項相當重要的工作。

在介紹教育測驗與評量的意涵時，有必要先釐清「測驗」、「測量」、「評量」與「評鑑」等四個常容易造成混淆的重要概念。這四個概念看似相似，常容易讓剛接觸「教育測驗與評量」這門課程的初學者混淆。

「測驗」的英文詞彙是 **test**，它是指一個工具或是程序，用以獲得學習者的學習表現。在學校的教學情境中，最常見的測驗就是教師自編的平常考或段考的考卷。這些教師自編的測驗工具，常包含是非題、選擇題、配合題、填充題、計算題、簡答題、問答題……等各種的試題類型，透過學生在這些試題的得分高低情形，教師就可以瞭解學生的學習表現。

「測量」的英文詞彙是 **measurement**，它是指根據一套規則，針對行為、特質或屬性，給予一個數字的歷程。測量這個概念是很強調量化的，所以一提到測量，就知道一定會產生相對應的數字。例如，在學校的教學情境中，國二英文教師以國二標準化英文成就測驗，對全班同學進行施測，全班每位同學都獲得一個分數，這樣一個歷程就是屬於測量，此測量是在瞭解學生的英文能力。

「評量」的英文詞彙是 **assessment**，它是指透過系統化的測驗或其他評量方式，獲得學習者的學習狀況，並對學習者學習狀況進行價值判斷的

歷程。例如前面提到國二英文老師，想瞭解學生的英文學習成果，除了採用標準化的成就測驗之外，也透過觀察的方法，評估學生學習英文的用功程度，以及採用實作表現的方式，評估學生英文朗誦的發音情形。此種採用測驗、觀察、實作等多種評量方式，蒐集學生的學習成果，就是屬於評量的歷程。評量除了測量或其他的方法，瞭解學習者的學習狀況之外，還需對學習者的學習狀況提出價值判斷。例如 A 生在標準化英文成就測驗的得分為 95 分，但是 A 生上英文課的學習態度不佳，英文老師給 A 生的評量結論是英文學業成果不錯，但英文學習歷程的學習態度待加強。

「評鑑」的英文詞彙是 **evaluation**，它是指透過系統化的評估方法，瞭解某項政策或某個專案的實行績效。評鑑一詞較常使用在對機構或計畫的評判，例如某國小推動一個「兒童閱讀運動」，鼓勵每個年級的學生一學期至少閱讀十本課外讀物。校長想要瞭解這個「兒童閱讀運動」的實施成效，於學期結束後，可根據每個年級有多少百分比的學生達到一學期至少閱讀十本課外讀物的標準，來評鑑這個「兒童閱讀運動」的實施成效。

由上面的說明可知，「評量」的涵義是比較廣的，它包含由「測量」或「非測量」所獲得的資料，並且需對這些資料進行價值判斷。「測驗」則是「測量」或「評量」的一種工具。「評鑑」雖具有「評量」的功用，但其關切的對象主要是機構或計畫，而非學習者。

綜合上述有關「測驗」、「測量」、「評量」與「評鑑」的概念之後，所謂的教育測驗與評量是指「教師透過測驗或其他有系統的評量方式，蒐集、整理和解釋有關學生學習成果的資料，用以瞭解和評價學生的學習表現和學習成就」。

二、教育評量的目的

教育評量活動的實施，有其目的性。一般而言，教育評量大致可達成下列幾項的目的：

(一)協助學生瞭解自己的學習成果

教育評量可提供學生檢視自己學習成果的機會，透過教育評量的結果，可以提供學生學習回饋的資料，讓學生瞭解自己哪些教學單元的內容，學習得比較好；哪些教學單元的內容，需要再加強。

(二)提供學生生涯輔導的資料

透過教育評量，可以提供學生升學或就業所需的參考資料。例如，國中升高中職時，可透過教育評量的資料，協助學生選擇適合自己就讀的高中或高職。而高一升上高二時，也可藉由教育評量的資料，協助學生挑選適合自己性向的文組或理組就讀。

(三)促進學生的學習動機

學習活動對於學生而言，是一項需要耗費時間與精力的活動，學生若缺乏學習動機，便不會樂於參與學習活動。因此，如何引起學生的學習動機，是教師在教學歷程中，相當重要的一項工作。透過教育評量的活動，可產生一股督促的力量，促進學生的學習動機。

(四)提供教師教學計畫的參考

為了讓教學活動能順利的進行，教師在進行教學活動之前，應先瞭解學生是否具備足夠的先備知識。藉由教育評量的活動，可以協助教師瞭解學生的起點行為，是否足以學習該單元的教材。

(五)提供教師教學的省思

教師教學結束後所進行的教學省思，有助於提升教師本身的教學效能。而藉由學生的教育評量結果，教師除了可以掌握學生的學習狀況，也可以反省自己的教學活動，是否有需要調整或修正的地方。

(六)提供家長有關學生的學習狀況

學生的評量結果，可以提供家長瞭解自己子女的學習成就，家長也可

根據評量結果，與教師討論小孩在學校的學習狀況。

綜合上述的討論可知，不論是對學生、教師或家長，教育評量活動都有其重要的目的。雖然教育測驗與評量可達成許多重要目的，但教育測驗與評量也常造成許多反效果。例如教育測驗與評量容易引起學生的測試焦慮，容易根據評量結果對學生進行不同能力的不當分類，容易迫使教師接受「考試領導教學」的教學方式，容易導致家長對測驗分數斤斤計較的心態。因此，教師在進行教育評量活動時，除了瞭解其所達成的目的之外，也應該小心避免其可能帶來的負作用。

第二節 評量與教學、評量與學習的關係

評量活動與教學活動，以及學習活動都有密切的關係。底下分別介紹評量活動與教學活動，以及評量活動與學習活動的關係。

一、評量活動與教學活動的關係

傳統上的教學，教學活動與評量活動比較像是一種主從的關係，教學活動是教學歷程的主角，評量活動則是主角表演結束之後，才登場的配角。教學活動與評量活動就變成兩個互為獨立的活動，教學活動是教師教導學生知識與技能的活動，評量活動只是單純用來知道學生學會多少知識與技能的活動。

隨著教學與評量觀念的革新，教育學者主張完整的教學歷程應同時包含教學活動與評量活動，教學活動與評量活動也逐漸變成合作的伙伴關係，圖 1-1 顯示兩種活動的互動關係。

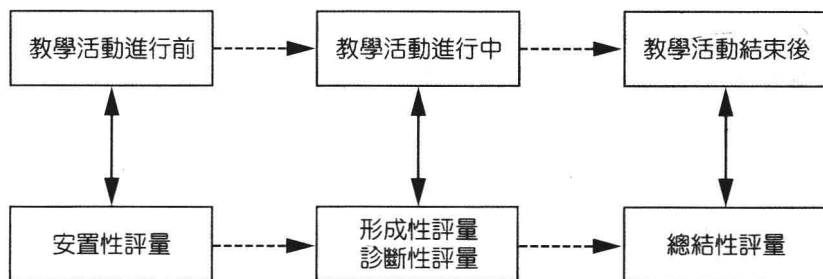


圖 1-1 教學活動與評量活動的關係

教師在教學活動開始之前，必須先瞭解學生是否具備足夠的先備知識，才能確定學生是否有能力學習該教學單元的教材。教師不論是採用課堂的提問，或是以測驗卷讓學生作答，都是屬於安置性的評量方式。由安置性評量所獲得的結果，可以協助教師瞭解學生的起點行為，可否直接進行教學活動，抑或是需要補足學生的先備知識。

教學活動進行中，教師為了瞭解學生的學習狀況，可使用小考、平時考等形成性評量的方式，評估學生是否清楚瞭解上課的內容。當教學活動進行中，發覺學生有學習困難的情形產生時，可採用診斷性評量，協助教師瞭解學生的迷思概念。

教學活動結束後，教師可選擇針對幾個不同的教學單元，進行總結性評量，用以協助教師與學生，瞭解學生的學習表現是否達到預期的目標。

由此可知，教學活動進行的前、中、後，都有與其相配合的評量活動。教師若能有效的整合教學活動與評量活動，將能獲得較佳的教學成效與學習成果。

二、評量活動與學習活動的關係

有關評量與學習的關係，Earl (2003) 認為可以分成三種不同的取向：學習的評量 (assessment of learning)、促進學習的評量 (assessment for learning)、評量即是學習 (assessment as learning)。

學習的評量取向，將評量當作是檢視學生學習狀況的一種方法，當教

學單元結束後，才會進行總結性評量。學習的評量取向之評量，其目的是作為安置、升級、獲得某種資格的判斷依據。

促進學習的評量取向，將評量視為是協助學生學習的一種方法，在教學單元進行中，就會採用形成性評量，用以瞭解學生是否有學習的困難點，其目的是協助學生獲得最佳的學習成效。

評量即是學習取向，將評量設定為是一種學習，在評量的過程中，學生可以學到如何設定目標、如何監控所設定的目標，以及如何調整或修正所設定的目標，其目的是培養學生自我學習的能力。

McMillan (2007) 將這三種學習與評量取向的特質，整理成表 1-1。

表 1-1 三種學習與評量關係取向的特質

學習的評量	促進學習的評量	評量即是學習
總結性	形成性	讓學生熱中於學習的評量本質
檢視學習成果	描述未來學習的需求	促進學生對學習的自我監控
教學單元結束後實施	教學單元進行中實施	教學單元進行中實施
經常使用常模計分方式；成績進行排名	評量作業提供教師修正教學活動	強調學生對於效標的知識，用以評鑑學習
試題從研讀的教材中命題	建議修正的教學	學生選擇修正的教學
一般性的	特定性的	特定性的
作為提供家長的報告	作為提供學生的回饋資料	促進學生的自我監控
可能會降低學生的動機	提升學生的動機	提升學生的動機
高效率，但膚淺的測驗	有深度的測驗	測驗提供學生學習
聚焦信度	聚焦效度	聚焦效度
延宕的回饋	立即的回饋	立即的回饋
總結性判斷	診斷性	診斷性

資料來源：Classroom Assessment: Principles and Practice for Effective Standards-based Instruction, by J. H. McMillan, 2007, p. 16. Boston: Pearson.

由表 1-1 可知，就評量的目的而言，學習的評量取向是屬於總結性的成果檢視，提供家長知道學生的學習成果；促進學習的評量取向是屬於形成性的學習需求探索，提供學生學習的回饋資料；評量即是學習取向是屬於鼓