

測驗與評量

陳英豪 吳裕益 著

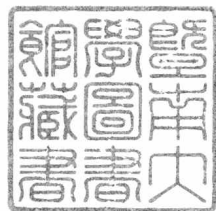
高雄復文圖書出版社

G642
831.2

港台書

測驗與評量

陳英豪 著
吳裕益



復文圖書出版社

測驗與評量

六版 2001/03
2003/03

版權所有，請勿翻印

定價：500 元

編者：陳英豪、吳裕益
發行人：蘇清足

本書如有破損，缺頁或倒裝，
請寄回更換。

出版者：高雄復文圖書出版社
地址：高雄市苓雅區泉州街5號
電話：(07) 2261273
傳真：(07) 2264697
郵撥：41299514
台北編輯部：台北市大安區溫州街48巷5號1F
電話：(02) 23695250・23695680
傳真：(02) 83691393
裝訂：台灣高揚文化事業有限公司
電話：(07) 6165206
聯合發行：麗文文化事業機構

行政院新聞局出版事業登記證局版台業字第1804號

ISBN 957-555-025-0

<http://www.liwen.com.tw>

E-mail: liwen@mail.liwen.com.tw

修訂版序

本書於民國七十一年由台北偉文圖書公司出版，書名為「測驗的編製與應用」，此書出版後，頗受心理與教育測驗課程師生之歡迎。其後偉文因故未能繼續出版，對想用本書作為教科書或主要參考書的師生造成很大不便。作者有鑑於此，乃商請復文圖書出版社負責人楊麗源先生向偉文圖書公司買回版權，重新修訂和增加部份內容，並將書名更改為「測驗與評量」。

作者所以將本書改為「測驗與評量」。是因為本書的範圍要比測驗與測量廣，改為新的書名較為妥當（請參閱本書第一章第一節對測驗、測量與評量三個名詞之界定）。本書的主要內容是「教學評量」，因此本書也可以做為新制師範學院各系共同必修課程「教學評量」之教科書或主要參考書。

本書修訂版之出版相當倉促，且因作者才疏學淺，錯誤必在所難免，尚祈諸位先進不吝指正。

陳英豪 吳裕益 謹誌——

中華民國七十九年六月十八日

序 言

教育的目的在於引發學生行為的改變，因此教師需要善於運用測驗來評量學生學習後行為改變的情形。教育的各種決策均來自班級教學結果之測量，學校及教育行政單位之入學、分組、分班、輔導等措施之實施，教學結果的診斷與補救，對新課程、新教材與教法的採擇與修訂，在在均須根據測驗的結果來做決定。

近年來由於教育目標逐漸複雜，教學內容日益繁雜，而且來自教育決策者、家長、社會與學生對教育效果的要求與批評，也日漸增多，因此學校教學結果的測量，也較以前更加複雜與困難。

雖然測驗是教學結果的回饋，是教育改進的依據，但我們卻很驚異的發現：學校督學只重一般教學的督導，而忽略了學校測驗實施情形及測驗結果之督導；師資訓練單位只重課程、教材及教法之傳授，而未重視學生對測驗應用能力之獲得；教師只忙於教學，而不注意教室測驗之效度與信度問題，更遑論應用測驗結果來改進教學了。這種只重「投入」而不管「輸出」，只問「耕耘」而不問「收穫」的現象，是目前教育忽略測量與評鑑的最好寫照。但是考試影響教師的教學與學生的學習，這是不爭的事實。因此，錯誤的考試自然導引了錯誤的教學與學習。這是我們目前教育上的一個嚴重問題，也是作者撰寫本書的動機。

測驗依其性質可分成二大範圍：教師自編測驗和標準化測驗。本書係以教師自編測驗為範圍，目的在使教師獲得編製各種測驗，選擇與評鑑測驗，以及應用測驗與解釋測驗結果之能力。

測驗的發展有幾個新的趨勢，值得我們重視：一、對「標準參照測驗」(criterion-referenced tests)的重視。由於「績效責任」(accountability)及「行為目標」(behavioral objectives)等教育新觀念的興起，引發學校對「標準參照測驗」的重視。本書除了在第一章比較了傳統「常模參照測驗」(norm-referenced tests)與「標準參照測驗」之區別外，更在第九章介紹了標準參照測驗結果的解釋，在第十一、十二章提出標準參照測驗效度與信度的考驗方法。二、重視高層次學習結果的測量。傳統的測驗大多只能測量「知識」層次的學習結果，但是現代的教育目標，在智育方面除了重視事實知識之獲得外，更強調「理解」、「應用」、「分析」、「綜合」、「評鑑」等能力之獲得與評量。本書第四章對於這些高層次學習結果的測量方法，作了一番系統而詳盡的介紹，舉例豐富，為國內目前亟待解決的問題，提供了最具體的編製方法。三、量評方法之多樣化。由於學校教育目標日益複雜，情意教育受到重視，加上對高層次學習結果測量的要求，使得教師必須以傳統紙筆測驗以外的方法，來評量學生的學習成就。因此，本書第五、六、七章分別介紹了論文式試題、實作測驗、觀察工具、口試和開卷考試的編製與實施。四、以多重層面來評量學生的能力與成就。由於受到因素分析等統計方法，及近代心理學家重視個別差異和多重能力的影響，近來的測驗逐漸重視利用各種測量的技術來評估一個人的能力，並將之安置在最適當的工作上，務期人盡其才，充份發揮個人的潛力。因此，本書第十一章特別介紹了人員分類和甄選策略，以供學校、醫療、工商機構與行政單位實際應用之參考。五、重視心理與社會因素對測驗結果的影響。近年來，個人及社會因素對測驗影響的研究資料，日益豐富，因此本書特在第八章列入「影響測驗的心理與環境因素」，以實際研究的資料，列舉了影響測驗成績的各種因素，提醒教師及測驗使用者如何正

確實施測驗，並對測驗結果作合適之解釋與應用，避免作了錯誤的推斷。

本書主要的是為各級學校教師和師專、師範院校學生而撰寫，也可供工商企業界作人員甄選及安置之參考。全書共十二章，包括各種測驗的編製與實施，測驗結果的解釋與應用，測驗的分析與評鑑等。全書力求體系完整，資料新穎，說明詳細，舉例豐富，淺易實用。

本書的寫作與編校工作，前後費時二年之久。雖然作者已盡力而為，但由於個人才識有限，書中錯誤與疏漏之處，懇請國內外先進賢達賜予指正，以供再版時之修正。

陳 英 豪

吳 裕 益 謹識

民國七十一年二月六日於

國立高雄師範學院教育研究所

目 錄

修訂版序 序 言

第一章 測驗在教學上的必要性

第一節	測驗、測量與訂量之意義·····	3
第二節	教學過程中的測驗·····	5
第三節	測驗的功能·····	10
第四節	成就測驗的基本原則·····	16
第五節	「常模參照測驗」與「標準參照測驗」的比較·····	24
第六節	本書的結構·····	27

第二章 成就測驗的編製計畫

第一節	決定測驗的目的·····	32
第二節	確定所要測量的學習結果·····	33
第三節	具體界定所預期的學習結果·····	50
第四節	列出教材大綱·····	52
第五節	編製細目表·····	55
第六節	根據細目表來編製測驗·····	56

第三章 客觀測驗的編製——「知識」層次的測量

第一節	客觀測驗類型的選擇·····	64
第二節	選擇題的編製·····	65
第三節	是非題的編製·····	92

第四節	配合題的編製	99
第五節	簡答題的編製	105
第六節	根據圖表來編製客觀測驗題	112

第四章 客觀測驗的編製——「複雜成就」的測量

第一節	單一型試題的編製與使用	126
第二節	複雜型試題的編製與使用	135
第三節	複雜型試題示例	141
第四節	各認知層次選擇題命題示例	149

第五章 論文式試題的編製

第一節	論文式試題的性質	164
第二節	論文式試題的類型	170
第三節	編製論文式試題的原則	172
第四節	論文式試題的評分方法	177
第五節	開卷考試的命題與評分	180

第六章 實作測驗的編製

第一節	實作測驗的性質	188
第二節	實作測驗的類型	190
第三節	編製實作測驗的步驟	194

第七章 觀察工具的編製

第一節	觀察評分的性質	203
第二節	軼事記錄法的使用	207
第三節	評定量表的使用	213

第四節	檢核表的使用	232
第五節	口試	237

第八章 測驗的編輯和實施

第一節	試題的編輯	245
第二節	測驗的實施與評分	254
第三節	影響測驗結果的心理與環境因素	261

第九章 測驗結果的解釋

第一節	常模參照測驗結果的解釋	279
第二節	標準參照測驗結果的解釋	313
第三節	解釋測驗結果的原則	318

第十章 測驗的分析和評鑑(一)——題目分析

第一節	良好測驗的品質	323
第二節	題目分析的功用	326
第三節	常模參照測驗的題目分析	328
第四節	速度測驗的題目分析	360
第五節	標準參照測驗的題目分析	362
第六節	題目分析實例和試題的改進	367

第十一章 測驗的分析和評鑑(二)——效度

第一節	效度的性質	381
第二節	效度的類型與考驗方法	382
第三節	標準參照測驗的效度	411
第四節	影響效度的因素	413

第五節 效度的複核與效標的混淆..... 424

第六節 效度的應用..... 426

第十二章 測驗的分析和評鑑(三)——信度

第一節 信度的性質..... 445

第二節 信度的類型與考驗方法..... 448

第三節 標準參照測驗與速度測驗的信度..... 467

第四節 影響信度的因素..... 473

參考書目 483

附 錄

附錄 A 常態分配表 497

附錄 B 范氏題目分析表..... 500

附錄 C χ^2 分配的自由度與百分點 527

索 引 529

1 測驗與評量 在教學上的必要性

- 第一節 測驗、測量與評量之意義
- 第二節 教學過程中的測驗
- 第三節 測驗的功能
- 第四節 成就測驗的基本原則
- 第五節 「常模參照測驗」與「標準參照測驗」
的比較

第一章 測驗與評量在教學上的必要性



測驗與評量在各科教學中均扮演著重要的角色。在班級教學中，我們廣泛地應用測驗來評估學生的學習成就，在個別化和編序教學中，也需要用測驗來了解學習結果。雖然測驗已廣泛地用來評量和指導學生的學習，但是有很多教師並不了解如何編製和應用測驗來增進教學效果。本書主要的目的就是在提供教師在教學情境中編製和應用測驗所需之原則和方法。

第一節 測驗、測量與評量之意義

測驗（test）、測量（measurement）與評量（evaluation）三個名詞之意義很容易使人混淆。即使修過測驗與評量課程者亦常對這三者到底有何不同感到困惑。特別是在課堂教學，我們常將「評量」視為「測量」和「測驗」的同義詞（synonym），可以互換使用。例如，當教師實施成就測驗時，我們可以說教師是在「測驗成

4 測驗與評量

就」(testing achievement)、「測量成就」(measuring achievement)或「評量成就」(evaluating achievement)。這時候我們幾乎完全沒有考慮到這三個名詞有不同的意義。

另外有某些情境下，評量被視為不需使用測量來搜集資料的那些鑑定方法(appraisal methods)之集合名詞(collective term)。在此種情況下，「評量」被視為是對學生行為「質的描述」(qualitative descriptions) (如，學生行為的軼事記錄，參閱本書第七章第二節)，而「測量」被視為是對學生行為「量的描述」(quantitative descriptions) (如，成就測驗分數)。

測驗，測量與評量三個名詞較具體的差異如下：

測驗：測量學生行為樣本之工具(instrument)或系統的方法(systematic procedure)。

測量：獲得個體具有某種特質之數量的過程，也就是以數量來描述個體具有某種特質之程度的過程。(所要回答的問題是「有多少？」)

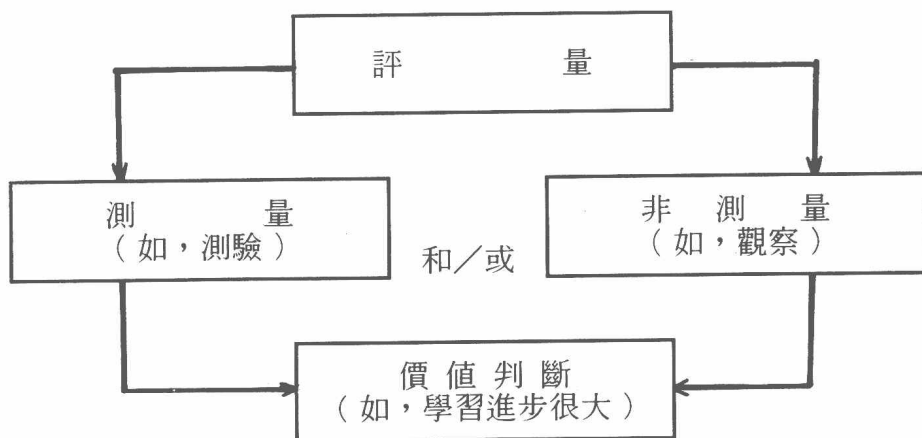
評量：用來搜集、分析和解釋有助於決定學生達成教學目標之程度的系統過程。(所要回答的問題是「有多好？」)

評量要比測量之含義廣且深入，而測驗只是測量的一類。測量限於對學生行為量的描述，測量結果通常以數字來表示(如，小華在50題的語文能力測驗答對40題)。測量沒有包括質的描述(如，小華的儀容整潔)，也沒有對測量結果作價值判斷。反之，評量可以同時包括對學生行為之量的描述(測量)和質的描述(非測量)。此外，評量通常均會對所得結果之可取性(desirability)作價值判斷(如，小華的數學計算能力進步很大)。從圖 1.1 可以看出評量之綜合性(comprehensive nature)，以及測量技術和非測量技術在評量過程所扮演之角色，評量並不一定依賴測量所得結果，即使是依賴

測量，那也遠超過簡單的數量描述。

課堂上評量（即教學評量）之主要目的是在了解教學目標達成之程度。如，學生的四則運算速度和精確性有無進步？閱讀能力有無進步？如有，那進步，多少？學生是否能更有效運用時間？學生的合羣行為有無進步等？要解答這些問題，我們需同時用到測量和非測量技術。

圖 1.1 測量與評量之關係



第二節 教學過程中的測驗

成就測驗是測量學生所學知識和技能的系統方法。雖然成就測驗強調學習結果的測量，但不表示成就測驗僅是在教學結束時才實施。儘管成就測驗經常在某一單元或課程結束時實施，用以評定學生之學習結果，此種總結性的評量僅是成就測驗眾多功能之一。就教學來說，測驗的主要目的在增進學習效果，故除了學習後之總結測驗可提供學生回饋之訊息外，學習前之安置測驗及學習中之形成性和診斷性測驗，對整個教學均有莫大的幫助。

為了使成就測驗能充分幫助學習，必須使測驗成為整個教學過程

6 測驗與評量

的一部分。在擬訂教學計畫時，就需考慮到如何實施測驗。從教學開始至結束的過程中，教師需做多種決定，測驗能提供教師作決定時所依據之客觀資料，使教師之決定更為適當，教學更為有效。

在教學過程中，教師所要作的決定可分為：1. 教學開始時的決定；2. 教學過程中的決定；3. 教學結束時的決定。茲將成就測驗在各階段做決定時所扮演的角色敘述如下：

一、教學初始

教師在進行教學前需先解決下述兩個主要問題：

(一)學生在教學初始時已具備的知識和技能的程度如何？

(二)學生已達到教學計畫中所欲達成的學習結果的程度如何？

第一個問題可由準備度測驗（readiness pretests）來解決，此種測驗在某一課程或單元開始之前實施，包括有效學習該課程或單元所需具備的知識技能。例如：學習研究法之前，需先測驗學生的統計知識；學習代數課程時，需先測驗學生的計算能力；學習德文時，需先測驗學生的英文文法。學生如果缺乏必備的知識或技能，可實施補救教學或編入特殊班級，給予特殊訓練。

第二個問題可用學習前的安置測驗（placement tests）來解決。此種成就測驗的內容包括教學計畫中預期達成之學習結果，故可能很像教學結束時所實施之測驗，但此種測驗的編製和結束時的測驗不同。因為我們的興趣僅在於了解學生是否已熟練新課程或新單元教學計畫中的教材。如果學生已熟習新教材，則我們應修改教學計畫，允許某些學生跳過已熟練的單元，或增加課程的深度或廣度，以適應其需要。

當然，並不是每個教學計畫實施前都必須實施安置測驗。如果老師與學生相處很久，非常了解學生過去的成就，那麼在新單元教學之