

教育測驗與評量

Measurement and Assessment in Teaching (9E)

Robert L. Linn

M. David Miller 著

王振世

何秀珠

曾文志

彭文松 譯

PEARSON

教育測驗與評量

Measurement in Teaching (9E)

本書旨在介紹教學過程中測驗與評量所扮演的
明如何發展和選擇適當的測驗與評量工具以及程序
對概念的闡述清晰，取材廣泛豐富，範例多元真實，與教
學情境的連結緊密適切，並且提供許多輔助閱讀資料及相
關學習網站以幫助讀者學習，是近年來測驗與評量相關領
域中不可多得的好書。本書適合入門學習測驗與統計的大
學部學生或碩士生，而對於想要自我充實的在職教師或想
要了解測驗與評量最新趨勢的讀者而言，亦是相當重要的
參考資料。



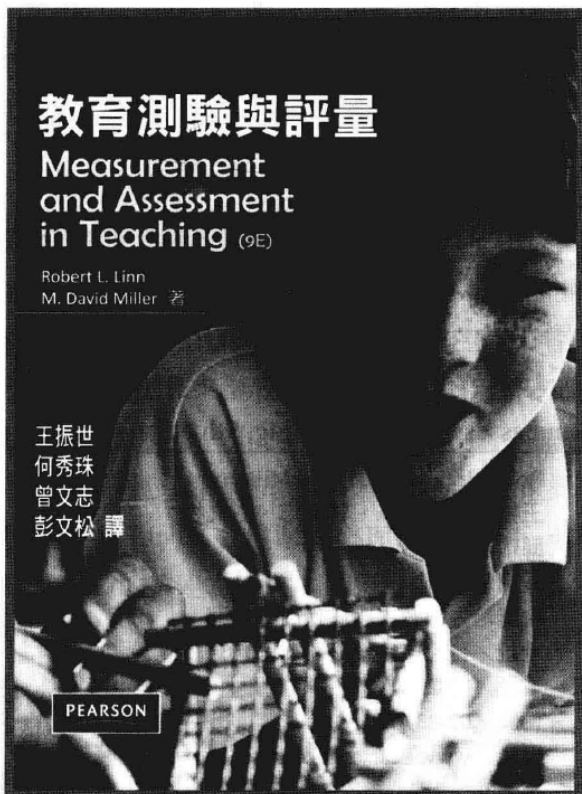
雙葉書廊有限公司

地址：臺北市羅斯福路三段269巷12號1樓
電話：(02) 2368-4198 傳真：(02) 2365-7990

Prentice Hall
is an imprint of

PEARSON

培生教育出版集團



教育測驗與評量

Measurement and Assessment in Teaching (9E)

Robert L. Linn
M. David Miller 著

王振世
何秀珠
曾文志
彭文松 譯

PEARSON

教育測驗與評量

Measurement and Assessment in Teaching (9e)

Robert L. Linn 著
M. David Miller

王振世 何秀珠 譯
曾文志 彭文松



雙葉書廊有限公司 發行

PEARSON

台灣培生教育出版股份有限公司
Pearson Education Taiwan Ltd.

教育測驗與評量 / Robert L. Linn, M. David Miller著;王振世等譯. -- 初版. -- 臺北市:臺灣培生教育出版:雙葉書廊發行, 2009.01

面;公分

含參考書目

譯自: Measurement and assessment in teaching, 9th ed.

ISBN 978-986-154-800-5(平裝)

1. 教育測驗 2. 教育評量

521.3

97020075

教育測驗與評量

Measurement and assessment in teaching, 9th ed.

原 著	Robert L. Linn, M. David Miller
譯 者	王振世、何秀珠、曾文志、彭文松
主 編	邱瓊薇
執行編輯	吳婧瑜
出版者	台灣培生教育出版股份有限公司 地址/台北市重慶南路一段147號5樓 電話/02-2370-8168 傳真/02-2370-8169 網址/www.PearsonEd.com.tw E-mail/hed.srv@PearsonEd.com.tw
發行所	雙葉書廊有限公司 地址/台北市羅斯福路三段269巷12號1樓 電話/(02)2368-4198 傳真/(02)2365-7990 網址/www.yehyeh.com.tw E-mail/reader@yehyeh.d2g.com
總 經 銷	雙葉書廊有限公司
出版日期	2009年01月初刷一刷
I S B N	978-986-154-800-5

版權所有·翻印必究

Authorized Translation from the English language edition, entitled MEASUREMENT AND ASSESSMENT IN TEACHING, 9th Edition, 0131137727 by LINN, ROBERT L.; MILLER, M. DAVID, published by Pearson Education, Inc, publishing as Prentice Hall, Copyright © 2005, 2000, 1995 by Pearson Education, Inc.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from Pearson Education, Inc. CHINESE TRADITIONAL language edition published by PEARSON EDUCATION TAIWAN and YEH YEH BOOK GALLERY Copyright © 2008.

譯者序

當雙葉書廊邀請筆者翻譯 Robert L. Linn 所著的《教育測驗與評量》（*Measurement and Assessment in Teaching*）時，我著實陷入掙扎的情緒中，但最後仍決定接受此艱難的任務。一來是因為 Linn 所著的書一直是筆者求學過程中重要的參考書籍，希望透過翻譯的工作來獲取新知，並自我成長；二來是希望透過本書的中譯本，增進相關領域大學生、研究生或在職教師對書中概念的獲益；最後則希望為大學「教育測驗與評量」課程的教科書，提供另一種選擇。

本書的第一作者 Linn，有雄厚的心理學及教育心理學背景，是 AERA、NCME、NAE 等學會的重要成員，其書籍與研究著作豐富，且曾獲得十多項重要的學術獎項，可謂是測驗與評量學界的重量級學者。此版本是本書第九版之海外版，其設定的對象是非美國本土的學生或教師。在書中，作者除了鉅細靡遺地解釋重要概念之外，並舉出相當多教室內常發生的真實案例，同時結合了教育心理學的相關知識；因此，本書可說是一本學理與實際應用兼備的教科書。

為了讓譯書的工作能順利進行，並且趕上出版的時效，特別邀請了相關領域人士共同翻譯，譯者群成員及負責章節分別為：王振世教授（第一章至第五章，第十八章與第十九章）、何秀珠教授（第十一章至第十五章）、曾文志教授（第六章至第十章）以及彭文松先生（第十六章與第十七章）。在此感謝譯者們慷慨應允，並努力配合。此外，還要感謝雙葉書廊在編輯、校對等方面的協助，讓本書的翻譯工作得以順利進行。

翻譯是一項艱難的工作，雖然譯者群擁有測驗與評量相關背景，但欲達到翻譯的最高境界——信、達、雅，卻不是一件容易的事。在翻譯的過程中，譯者群的宗旨是達意、易懂、順暢，以增加中譯本的可讀性，然而，過程中難免有疏漏或未盡完善之處，尚祈先進不吝指正。

王振世 謹識

民國九十七年十月

於新竹風城

前言

《教育測驗與評量》第九版旨在為教師介紹有效教學中，測驗與評量的重要元素。貫穿本書的中心主旨為：學生的學習評量在教學歷程中所扮演的重要角色，以及評量的效能主要取決於教師能否編製與選擇可有效測量學習結果的測驗或作業。

第九版的修訂，主要反映了從第八版到目前為止教育與評量的重大改變。這些重大的改變包括：

1. 第一章的修訂反映了測驗與評量的內容必須符合州和聯邦政府的要求，以達到學校績效，以及特別強調布希總統在 2002 年元月簽署的「沒有落後的孩子法案」（No Child Left Behind Act）之要求。
2. 增加許多課堂評量的新範例。
3. 在第四章與第五章中對效度與信度的敘述，配合了最新版「教育與心理測驗標準」（*Standards for Educational and Psychological Testing*）的主題論述。
4. 更新檔案評量（第十二章）的討論。
5. 增加或更新了許多可供參考的網址。
6. 已出版測驗的材料反映出評量的新取向與新趨勢。

本書主要是提供大學或研究所階段教育與評量課程的入門用書，並假定學生無須具備任何測驗或統計的基礎知識。書中的敘述相當直接明瞭，也採用許多範例來說明測驗與評量的概念與方法。統計概念僅在有助於瞭解和討論時提到，並且只強調概念的解釋而非計算。

如同前幾個版本，本書的焦點可以由下列幾項一般學習結果作最佳的說明：

- ◆ 了解測驗與評量在教學歷程中所扮演的角色。
- ◆ 能夠以增進發展適當測驗與評量的方式，來定義教學目標與目的。
- ◆ 了解效度與信度的概念，以及其在測驗與評量的編製、選擇、解釋與使用上所扮演的角色。

4 教育測驗與評量

- ◆ 能夠編製測量簡單到複雜等各式學習結果的測驗與評量工具。
- ◆ 能夠從教室觀察、同儕評估及自陳報告中獲得評量的訊息。
- ◆ 能夠適當地實施測驗與評量，並有效使用其結果。
- ◆ 在完全了解測驗與評量以及測量誤差之前提下，適當地解釋測驗與評量的結果。
- ◆ 了解適用於學校之各種測驗與評量工具和程序，以及其優點與限制。
- ◆ 了解測驗與評量對成績報告系統的貢獻，以及它們對教學決策的助益。
- ◆ 能夠適當地解釋與使用已出版測驗和州政府評量的結果，並針對某教學目的，從中選擇最適當的測驗或評量結果。
- ◆ 了解使用法定測驗與評量的最新趨勢與議題。

如果能閱讀其他書籍當作輔助，將能促進上述這些瞭解與能力。此外，還可藉由測驗與評量工具的編製、計分與測驗結果的分析、各式測驗與評量工具的選擇、已出版測驗及州法定評量的評鑑，以及測驗與評量結果的解釋等實際的研究作業，來輔助本書的閱讀與學習。無論如何，對本書內容的實際應用，都有助於學生學習及使用測量技巧的能力。

目 錄

PART I 測量與評量的歷程

第 1 章	教育測驗與評量：背景、議題與趨勢	1
	教育評量：改革的氣壓計和槓桿	3
	以測驗為基礎的五十年教改	4
	測驗與評量的科技進展	11
	大眾對測驗與評量的關注	14
第 2 章	測驗與評量在教學歷程中的角色	25
	教學決定需要測驗與評量的資料	26
	評量、測驗和測量	27
	評量的一般原則	29
	評量與教學過程	31
	評量方法的類型	33
	評量類型的摘要整理	41
第 3 章	教學目標與具體目標：評量的基礎	47
	教學目標就是學習結果	50
	從定義明確的目標到思考課程	51
	敘述教學目標的方法	60

第 4 章	效度	73
	效度的特性	75
	效度驗證的主要考量	77
	內容考量（與內容有關的證據）	78
	構念考量	83
	評量與效標之關係	88
	後果的考量	97
	影響效度的因素	99

第 5 章	信度及其他良好的特徵	111
	信度的本質	111
	以相關方法估計信度	113
	測量標準誤	122
	影響信度係數的因素	128
	以固定表現標準評鑑測驗／評量的信度	132
	可用性	136

PART II 課堂測驗與評量

第 6 章	課堂測驗與評量的規劃	143
	課堂測驗與評量的目的	144
	發展測驗與評量的細目	146
	選擇適當的題目類型與評量作業	155
	準備相關測驗題目和評量作業的考量	163

第 7 章	建構客觀式測驗題目：簡單形式	179
	簡答題	180
	是非題或二選一反應的題目	188
	配合題	197
第 8 章	建構客觀式測驗題目：選擇題形式	207
	選擇題的特性	207
	選擇題的使用	209
	選擇題的優點與限制	216
	建構選擇題的原則	218
第 9 章	測量複雜的成就：釋義題	235
	釋義題的性質	236
	釋義題的形式與用途	237
	釋義題的優點與限制	244
	建構釋義題的原則	246
第 10 章	測量複雜的成就：申論題	255
	申論題的形式與用途	256
	綜合比較測量的學習結果	260
	申論題的優點與限制	260
	建構申論題的原則	263
	評分標準	268
	申論題的評分原則	274

第 11 章	測量複雜成就：實作評量	281
	實作評量的類型	282
	實作評量的優點與限制	287
	建構實作作業的建議	289
	實作表現的標準	291
	評分規準和評定量表	292
	檢核表	304
	學生參與評分	306
第 12 章	檔案評量	311
	何謂合格的學生工作檔案？	312
	檔案可能的優缺點	313
	檔案的目的	315
	檔案內容項目的指導說明	319
	選擇檔案項目和自評時的指引說明及學生角色	327
	評鑑標準	331
	用於教學與溝通的檔案	333
第 13 章	評量方法：觀察技術、同儕評量和自陳報告	337
	軼事紀錄	339
	學生自評和自陳報告	345
	同儕評量	346
	自陳報告技術	355
	態度測量	356
	興趣量表	359
	人格測量	362

第 14 章	課堂測驗與評量的編輯、施測與評鑑	369
	課堂測驗的編輯	370
	課堂測驗與評量的實施和計分	380
	課堂測驗與評量的評鑑	384
	建立有效試題和作業的檔案	396

第 15 章	評量等第與成績報告	401
	評定等第和成績報告系統的功能	402
	評定等第與報告系統的類型	404
	多元化成績報告系統	409
	字母等第的分派	412
	保管紀錄和等第的軟體	421
	舉行親師座談會	422
	向家長說明標準化測驗結果	424

PART III 選擇與運用已出版測驗

第 16 章	成就測驗	433
	標準化成就測驗的特色	434
	標準化成就測驗及非正式的課堂測驗	436
	標準化綜合成就測驗	438
	特定領域的成就測驗	446
	客製化成就測驗	453
	個別成就測驗	453

第 17 章	性向測驗	457
	成就與性向測驗	457
	學業性向與學習能力	460
	學習能力的團體測驗	460
	個別測驗	467
	解釋及使用學習能力分數時的注意事項	470
	文化公平測驗	471
	區分性向測驗	473
第 18 章	測驗的選擇、實施與使用	481
	如何獲得已出版測驗的訊息	482
	選擇適當測驗	486
	已出版測驗的實施	490
	已出版測驗的結果使用	496
第 19 章	測驗分數的解釋與常模	505
	解釋測驗分數的方法	507
	年級常模	511
	百分等級	515
	標準分數	518
	側面圖	527
	技能分析	529
	常模適切性的判斷	530
	地區性常模的使用	533
	解釋測驗分數的注意事項	534

PART I

測量與評量的歷程

CHAPTER

1

教育測驗與評量： 背景、議題與趨勢

舉凡國家和州政府對學生學業成就績效評量的強烈需求、國家及國際級的評量方案、國家及州政府為學習所設定的內容及表現標準，以及全球競爭力等，這一切均顯示了社會對測驗與評量的需求與日俱增，上述因素同時刺激與反映了教育評量的新趨勢。線上電腦化測驗已開始在一些地方使用，且使用率持續增加。然而，強烈倚賴測驗與評量作為教育改革工具的現實，也引起人們對測驗與評量工具之使用和結果公平性的廣泛討論。

在美國，教育測驗已普遍使用在各級學校中，標準化紙筆測驗是州和地方教育的基本要求。這種法定的測驗與評量包含了許多形式，以期符合各種教育目標，包括機器計分的選擇式測驗、簡答或申論形式的評量，還有強調實作表現的作業（例如，一項科學實驗或創作一件藝術作品）。

在美國，大多數州已對特定學科和年級全面實施測驗，且已實施多年，各州的測驗性質和形式、被測量的年級和科目也不相同。不過，聯邦立法院於 2002 年元月 8 日簽署法案，要求州級測驗的標準應有更大的一致性。「沒有落後的孩

2 教育測驗與評量

子」(The No Child Left Behind, NCLB) 法案中就規定，於 2005-2006 學年需對所有 3-8 年級的學生實施測驗，在中學每年至少對一個年級實施數學與閱讀（或文學）測驗。從 2007-2008 年開始，州政府每年必須對每三個區段（3-5 年級、6-9 年級、10-12 年級）中的至少一個年級實施自然科測驗。

NCLB 法案旨在要求學校、學區及州政府對學生的學習成就負起責任。無法達到州所設定之「年進步標準」的學校，將被列入「需要加強」的名單中。被列為「需要加強」的學校必須規劃出改善計畫，而各學區也必須提供不同方案給所屬學校，以提升學生的學習成就。當某學校連續兩年被評為「需要加強」時，該學區就必須提供低收入戶家庭學童適當的課輔服務。學校若連續 3、4 或 5 年上榜，就會受到更嚴厲的處罰，最後可能導致學校改組。

除了 NCLB 法案的規定外，許多州各自設立對測驗與評量的要求，此額外要求可用來評鑑學校的績效。有些州並要求學生必須通過考試，才能升級或取得高中畢業證書。

此外，測驗與評量也是國家級或國際級調查研究的一部分〔例如，國家教育進步評量 (National Assessment of Educational Progress, NAEP)〕，或者被用來進行國際評量，例如國際數學和科學研究。事實上，NCLB 法案要求每州每隔一年就需參加 4 年級和 8 年級之閱讀和數學測驗。對本書的讀者而言，為了進入大學而考的 SAT 或 ACT，或為申請研究所而考的 GRE (Graduate Record Examination) 均是不久前才經歷的國家級考試。

大部分學生，特別是在美國受教育的學生，在進入大學前皆有許多標準化測驗的受測經驗。不論是探討美國教育現況或教育改革的企圖心，標準化測驗和評量都扮演了重要的角色。然而，標準化測驗卻一直是社會爭議的焦點，爭議的原因有很多，包括標準化測驗對少數族裔和女性的測量偏誤之相關指控，以及此類測驗把教學引導到過分重視事實知識和低層次能力的困境。在教育政策制訂者偏好標準化測驗，但標準化測驗又引發諸多討論和爭議的情況下，教師應熟悉標準化測驗和評量的技術，同時也應了解它被正確使用和誤用的情形。

雖然本書把重心放在教室內測驗的編製和使用，但卻從測驗和評量的法定強制性以及相關的議題和趨勢開始談起，其中有幾個理由。首先，法定的測驗和評量為教室內的評量提供了清楚的脈絡 (context)。在很多情況下，測驗與評量被

當成是教育改革的平衡桿，用來釐清課程的優先順序，並點出教學活動的重心。其次，許多與標準化測驗和評量相關的爭論性議題，若非直接與教室評量有關（例如，教學關聯性、效度和公平性等問題），就是對教室評量的優缺點引發更多的討論。

教育評量：改革的氣壓計和槓桿

雖然學生花在準備課堂測驗和評量的時間比法定的標準化測驗多，但是法定測驗卻吸引了較多來自政策制定者、媒體和社會大眾的目光。不過這也不是什麼新鮮事，事實上，這種測驗和評量在美國社會中已有相當的歷史。1992年，技術評量辦公室（Office of Technology Assessment, OTA）提報給美國國會的一份報告書就曾指出：

從早期的公立學校運動開始，美國教育學者、家長、政策制定者及納稅人就已經把這些測驗視為是多用途的工具，其功能包括：測量個人進步情況的量尺、學校改革的工具、教育機會的過濾器及全國教育情況的氣壓計。

測驗和評量無論是在性質或使用上均出現許多不同的形式。但就像 OTA 報告中所說的，它們之間存在著某些一致性。教育改革者已使用這些法定的測驗和評量結果來解決教育的缺失。但更重要的是，它們已經把這些測驗和評量作為教改的工具了。

測驗和評量吸引教育決策者的原因

1. 測驗和評量是相當便宜的。相較於增加教學時間、減少班級人數、吸引人才投入教師行列、提升教師福利或提高教師專業成長所需的經費，測驗便宜許多。
2. 測驗和評量可用法律規定的方式強制執行。比起想辦法改變每天發生在教室內的複雜瑣事，命令州或學區當局實施測驗和評量，當然是容易多了。
3. 測驗和評量的改變是立竿見影的。對於新的測驗和評量之要求和標準，可直接在辦公室內由官員制定完成。
4. 測驗和評量的結果是明顯可見的。測驗結果會定期見報，讓社會大眾知道。

就氣壓計的角色而言，測驗和評量旨在提供學生成就之目前表現和進步的情況，以及教學品質方面的訊息。若以教改的槓桿來看，測驗和評量則被當成改革教育實務的機制，用來釐清政策的優先順序，並促使當局投入更多努力在教師和學生身上。

以測驗為基礎的五十年教改

以測驗為基礎的教改可追溯至 19 世紀中期，從麻州教育督學用紙筆測驗來要求公立學校為他們的教育成果負責開始（Resnick, 1982）。20 世紀初期，Joseph Rice（1914）對幾千名學童實施拼字和數學測驗，用來探討教學時間是否有效運用的問題。第一次世界大戰後，出版了很多成就測驗組合（achievement test batteries），所採用的就是大戰期間廣泛使用的選擇題（multiple-choice）技術。雖然測驗在第一次和第二次大戰期間仍持續使用，但使用率迅速增加卻是在二次大戰之後，特別是 1960 年代到今天。

一、政府支持的補救教育：Title I

因為教育機會不均等及學生表現的嚴重落差，從 1960 年代中期起，補救教育引起了社會大眾的重視。1965 年的國民教育方案（The Elementary and Secondary Education Act, ESEA）投注整個聯邦政府之力在補救教育上，其中最著名的就是 Title I。雖然 ESEA 分別在 1994 和 2001 年有一些重大的改變（1994 年的美國學校教育改善方案及 2001 年的 NCLB），但 Title I 仍是聯邦政府持續協助初等和中等教育的最大方案。

ESEA 之 Title I 旗下及 1960 年代的其他基金會，往往受到來自國會在評鑑及績效方面的壓力，而這種壓力對測驗出版商而言無異是一種恩賜。Title I 之評鑑與報告系統（Evaluation and Reporting System, TIERS）的測驗要求，對已出版之標準化測驗的推廣有著實質幫助。捨棄選擇某些年級及一年實施一次的作法，TIERS 建議每年秋季和春季均對學生施測，以了解參與此項計畫學生的進步情形。雖然測驗的結果很少被採用，但 TIERS 的作法卻抒解了來自聯邦計畫的績效壓力。然而，對學生成就進展情形的不滿，特別是 Title I 想要服務的學生，進一步