

現代測驗理論

王寶壙 編著



●心理出版社●

林幸台校閱

現代測驗理論

王寶墉編著

一般心理系列⑨

現代測驗理論

編著者：王寶墉

校閱者：林幸台

行政主編：蔡幸玲

執行主編：郭暖卿

發行人：許麗玉

出版者：心理出版社有限公司

社址：台北市和平東路二段163號4樓

電話：(02) 7069505

傳真：(02) 3254014

郵撥：0141866-3

法律顧問：李永然

駐美代表：Miss Lisa Wu

通訊處：171 Burns Ave. Lodi, NJ 07644 USA.

Tel : 201 546-5845

Fax : 201 546-7651

登記證：局版台業字第1963號

印刷者：普賢王印刷有限公司

初版一刷：中華民國八十四年五月

定價：新台幣**300**元

■ 有著作權·翻印必究 ■

ISBN 957-702-132-8

作者簡介

王寶壙

學歷 / 彰化教育學院輔導學系學士
/ 彰化教育學院輔導研究所肄業
/ 美國威斯康辛大學教育碩士
/ 美國喬治亞大學教育心理博士

經歷 / 國中、國小、啓聰學校教師
/ 彰化師範大學輔導學系助教

現任 / 文藻外語專科學校副教授兼學生輔導中心主任

作者序

心理與教育測驗的發展日新月異，理論上的發展已由古典測驗理論邁向現代測驗理論，應用上的發展也較著重測驗的連結等化、適性測驗，甚至電腦化適性測驗等課題。對於行政、研究、教學與輔導等方面提供更有效的支援。

鑒於坊間可供學生參考的有關現代測驗IRT的書籍，尚待充實，乃決定將在美國求學期間修習現代測驗理論一科和與Dr. Bashaw獨立研究所蒐集的資料，系統地編譯整理與歸納，以編寫一本淺易且具導論性質的參考書，以供研修「心理與教育測驗」時之參考。

本書前七章屬理論部份，後四章屬應用部份。但測驗理論常被指為艱深難懂，故本書之說明力求簡單扼要，盡量避免複雜的數學舉證，以期成為入門書之用。

本書之編寫，承蒙國立台北師大特殊教育研究所林幸台老師的校訂指導，謹此致謝。又蒙彰化師大陳貴龍老師及嘉義師院李茂能老師幫忙蒐集資料與校正文稿，不勝感激。筆者甫出校門，才疏學淺，經驗欠豐，又倉促付梓，疏漏之處，在所難免，懇請各界先進，不時賜予指正。

王寶壙 謹識

民國八十二年元月

校閱者序

現代測驗理論自一九五〇年代 F. Lord 發表雙參數常態肩形模式及潛在特質理論以來，已有四十年的歷史；隨著電腦的發展，一九六七年 B. D. Wright 在美國推動極受重視的 Rasch Model，亦有二十餘年；時間雖然不算長，然而有關現代測驗理論的研究與運用可說突飛猛進。我國有關此一方面的報導，最初見於林一真（民 71）的簡介，數年之後，始有臺南師範學院吳裕益、洪碧霞等人的一系列研究及研習活動，近年歸國的學人則陸續從此一角度探究有關測驗的問題（如題庫、適性測驗、及大學聯考等），有關論著則多見於中國測驗學會所編之「測驗年刊」。十年來此一理論在國內學術界已逐漸引起注意，然而其影響範圍尚未見擴大，實有繼續努力的必要。

古典測驗理論已有近百年的根基，成為國內編製或修訂測驗的主要理論依據，一般測驗編製者與使用者較習慣採用難度指數、鑑別力指數以及數種信度、效度考驗等資料做為評論測驗的指標，因此或許國人對於新理論仍存有若干疑義。誠如林一真（民 71）所言：「LTM 是那麼複雜，而各種模式所立的假設要求又很嚴，也許有人要問：何苦呢？」（p. 69）這可能是現代測驗理論尚未在國內受到應有的重視原因之一。

多年來，由於輔導的需要，測驗被大量使用，不僅因誤用、

濫用，導致測驗惡名昭彰，且所編製測驗的速度來不及市場摧毀它的速度，以致放眼所見，找不出還有多少測驗真正可以使用，加上近來因著作權問題，使得在輔導工作上最得力的工具，反而因已往過度的依賴而變成一種束縛。這個關鍵時刻，或許正是廣泛引介新理論最適當的時機。

王寶墉君自彰化師範大學的前身臺灣教育學院輔導系畢業後，擔任心理實驗中心助教，協助本人進行國內普通性向測驗（GATB）的編訂工作，基於對測驗的興趣，留學美國期間，不僅專攻心理測驗，並跟隨 Bashaw 博士從事多項此一方面的專題研究長達八年之久，有鑑於國內有關現代測驗理論尚未蔚為風尚，乃以其學習與研究之心得，編著「現代測驗理論」一書。

本書特色之一在由古典測驗理論開始，與現代測驗理論做一銜接與比較，對於熟悉古典理論者正可藉此逐步引入新測驗理論之殿堂。另一特色是文字簡單明瞭，雖然書中所引公式頗多，但並不複雜，且符號清楚，有關數學舉證部分亦力求簡明扼要，應不致造成讀者眼花撩亂之憾，對初次接觸現代測驗理論者，頗有入門引導之效。但對較深奧的理論基礎與運用部分，書中亦分別交待可資進一步參考之文獻，此為本書另一特色。

本書是國內介紹現代測驗理論的專業書籍的先驅，相信未來將有更多此一方面的論著，使國內測驗學界在數十年古典測驗理論的薰陶下，能逐漸融入現代測驗理論的方法與優點，使國內測驗的水準得以逐步提昇，而測驗學術的內涵亦可有所增進。

林幸台 謹識

民國八十二年十月

目 次

第一章 現代測驗IRT的發展	1
第一節 古典與現代測驗理論	4
第二節 古典測驗的缺點	17
第三節 現代測驗IRT的發展	21
第四節 現代測驗IRT的模式	26
第二章 現代測驗IRT的假定	31
第一節 單向性	34
第二節 局部獨立	40
第三節 其它假定	46
第四節 不變性與不定性	49
第三章 現代測驗IRT的參數	53
第一節 項目反應曲線	56
第二節 難度參數	62
第三節 鑑別力參數	68
第四節 猜測參數	73

第四章 單參數模式	77
第一節 單參數的類別與意義	80
第二節 Rasch 模式	86
第三節 Rasch 模式參數估計	92
第四節 Rasch 模式的應用	96
第五節 Rasch 模式—BICAL	101
第五章 多參數模式	107
第一節 多參數模式	110
第二節 模式選擇	113
第三節 模式假定考驗	117
第四節 模式特性與預測性考驗	123
第六章 能力量尺	131
第一節 能力量尺轉換	134
第二節 真分數	139
第三節 測量標準誤差	145
第四節 能力參數估計	150
第七章 訊息函數	157
第一節 訊息函數	160
第二節 測驗訊息函數	164
第三節 分數加權	169
第四節 相對有效性	172

第八章 參數估計法	175
第一節 聯合最大概似法	178
第二節 條件最大概似法	181
第三節 邊際最大概似法	184
第四節 貝氏估計法	187
第五節 概略法	191
第六節 電腦程式	193
第九章 評定量表	199
第一節 二分模式	202
第二節 部份給分模式	205
第三節 評定量表	212
第四節 二項式嘗試	215
第五節 Poison Count	218
第六節 評定量表實例	222
第十章 測驗等化	229
第一節 連結與等化	232
第二節 等化方法設計	236
第三節 傳統等化	243
第四節 現代IRT的等化	249
第十一章 適性測驗	255
第一節 適性策略	258

第二節 多階段測驗	265
第三節 電腦適性測驗	271
第四節 測驗偏誤	277
附 錄	283
A. 三參數項目訊息函數	285
B. 三參數對數模式項目與能力參數第一及第二階導數·····	287
參考文獻	289
一、中文部分	291
二、英文部分	293
索 引	309
一、英漢對照	311
二、漢英對照	317

1 現代測驗 IRT的發展

心理與教育測驗提供教師瞭解學生的能力、性向、成就、動機、人格與興趣等特性，以作教學、研究、行政、課程及輔導諮商之參考。目前，心理與教育測驗一科已常被列為師資培養及人文社會等相關科系的必修課程之一，各種心理與教育測驗也普遍地受到重視與運用。然而，測驗理論的研究發展日新月異，又加上快速電腦的問世，更促成了現代測驗理論研究的急劇發展，而展現出不同於往昔的新風貌。本章主要討論現代測驗理論與古典測驗理論的異同及現代測驗理論的意義及發展。

第一節 古典與現代測驗理論

古典測驗理論（classical test theory）與現代測驗理論（modern test theory）的關係猶如第一代與第二代電腦間的關係，有新有舊，有異有同。以下分別說明兩者的意義重點後再比較其異同。

壹、古典測驗理論

古典測驗理論又稱傳統測驗理論（conventional test theory），或古典真分數理論（classical true score theory）。其主要源自英國 Charles Spearman 所創相關概念的延伸。至今，有關測驗的信度，如：庫李信度、史布預測式、效度、測量標準誤、複本等觀念都是古典真分數理論的貢獻。因此，目前測驗的編製大都以古典真分數理論為依據。

古典真分數理論因其假定（assumption）較易符合一般研究資料，又稱為弱真分數理論（weak true score theory），以別於 Lord 在一九六五年以二項式分配原理所發展出的強真分數理論

(strong true score theory)，及爾後以變異數分析法來處理資料的概化理論 (generaliability theory)。以下簡要敘述古典真分數理論的假定，項目分析重點及強真分數理論和概化理論的大要。

一、古典真分數理論的假定

古典真分數理論的假定主要在界定觀察分數，真分數，及誤差分數三者間的關係。其主要者有三：

1. 觀察分數 (X) 是真分數 (T) 及誤差分數 (E) 的和。亦即

$$X = T + E$$

2. 誤差分數與真分數間無相關。也就是某群體在某測驗上所得真分數的高低與誤差分數的高低無關。

$$\rho_{TE} = 0$$

3. 兩不同測驗之誤差分數間無相關存在。亦即：

$$\rho_{TE} = 0$$

由上述的假定引申至期望值，變異數，相關之應用乃形成測驗編製中不可或缺的信度係數，如：複本，再測，折半，庫李 20，庫李 21 (難度相等)，效度係數，如：同時，預測效度，及測量標準誤等測驗的重要基礎觀念。如欲進一步瞭解古典真分數理論與信度間的關係，請參閱 Lord & Novick (1968)，Crocker & Algina (1986)，或林幸台 (民 74)，陳英豪、吳裕益 (民 75)。

二、古典測驗的項目分析

古典測驗項目分析時，選擇題目的標準常依據：

1. 項目難度 (item difficulty) 。
2. 鑑別力指數 (index of discrimination) 。
3. 相關鑑別力指數 (correlational index of discrimination) 。

項目難度指數就是答對某題目人數的百分比，也就是項目平均數 (item mean)。難度指數常以 P 表示之， $P=1.00$ 表示某題非常簡單人人通過， $P=.50$ 表示一半通過， $P=0$ 為無人答對。因此， P 介於 1 與 0 之間， P 值愈高題目愈容易， P 值愈低題目愈難。而 $q=1-P$ 表示答錯某題目人數的百分比，如在二分法計分——即答對 1 分答錯 0 分時，以 p = 平均數代入求標準差的公式可得某題目的變異數為 $\sigma^2 = P \times q$ ， σ 為 pq 的平方根，也就是答對與答錯人數百分比乘積的平方根。

項目難度的選擇得視測驗的目的與性質而定。如是教學診斷測驗應含較多容易的題目，如是大專聯考應含較多中難度的題目，以大專聯考數學科的平均分數皆在二十分以下可知其含甚多難題，而三民主義平均在七十分以上亦可知含甚多容易題，兩科都有偏頗未盡鑑別學生能力的功能，又把不同平均數標準差的不同學科分數任意相加或加權，實與雞兔同籠無異，亦是大專聯考的另一特色。但選擇題目光憑難度指數還是略嫌粗糙，應再考慮鑑別力。

項目鑑別力就是希望得分高的受試答對某題的機會要比得分低的受試答對該題的機會高，才能顯得該題確能鑑別出受試能力的差異。如高分組 (最高 27%) 在某題的通過率低於低分組 (最