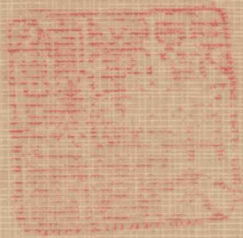


S 014942

教育測量

潘友昌編著

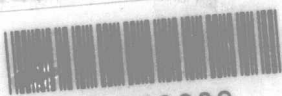


正中書局印行

G40
8811

教 育 測 量

潘 友 昌 編 著



S9002909

正 中 書 局 印 行

版權所有



翻印必究

中華民國五十六年八月臺初版
中華民國六十六年五月臺四版

教 育 測 量

全一冊 基本定價 一元三角

(外埠酌加運費匯費)

編 著 者 潘 友 昌

發 行 人 黎 元 譽

發 行 印 刷 正 中 書 局

(臺灣臺北市衡陽路二十號)

海外總經銷 集 成 圖 書 公 司

(香港九龍油麻地北海街七號)

海 風 書 店

(日本東京都千代田區神田神保町一丁目五六番地)

東 海 書 店

(日本京都市左京區田中門前町九八番地)

新聞局出版事業登記證 局版臺業字第〇一九九號(6041)辦
(500)

弁 言

- 一、本書以「教育測量」爲書名，我們解釋教育一詞從其廣義，乃包括全部行爲測量，心理測量也在內。
- 二、這一個領域的中文書籍甚少，又不能完全因襲外文資料，因爲時代、社會、文化及科學等實際狀況，都直接影響教育測量的實施。職是之故，本書不列舉各種測驗之樣本及說明；以本書作爲參考書自無問題，如以本書作爲大專學校教本，可自行補充當地之現行測驗，最好能選用標準測驗，中文測驗或有不足，再行以外文者補充之。
- 三、本書採用之若干特殊材料加註來源於括號內；爲期簡潔，以數字代書名，頁數隨後，以冒號分開，並以 100 冠中文書之首，200 冠西文書；如某書在參考書目之編號爲 210，該項材料在該書之 51 至 53 頁，則列爲 (210 : 51—3)。
- 四、全文圖表標識，係以章來分，如圖四、一，即是第四章之圖一，其所以沒有圖一、一者，係因爲第一章至第三章均無圖。
- 五、本書之專門名詞及西人姓名均不附原文，以資醒目，書後附有漢譯西人姓名對照表及中英名詞對照表，以便參考。
- 六、本書筆者僅一人，以習慣關係，行文時，間有用「我們」之處，當荷諒解。疏漏之處，在所不免，仍祈 高明不吝指正，幸甚。

潘 友 昌 謹 識

12	目次	
12	第一章 教育測量發展簡史	1
87	第一節 十八世紀以前的情形	1
73	第二節 十九世紀的教育成績測量	3
47	第三節 二十世紀以來的教育測量	4
47	第四節 十九世紀智力測驗的萌芽	6
87	第五節 二十世紀智力測驗的發展	6
77	第六節 人格測驗的發展	7
63	第七節 若干重要文獻	8
78	第八節 中國教育測量的發展	9
78	第九節 一九五〇年以後的發展趨勢	11
78	第十節 現行測驗概述	12
40	第二章 測量的記號	16
150	第一節 教育上的變數	16
331	第二節 測量的尺度	17
331	第三節 測量記號的類別	19
331	第四節 選用測量記號	23
161	第三章 事象與測量	26
134	第一節 可測量的教育事象	26
134	第二節 可測量的度的性質	29
134	第三節 推斷的度和構想	32
134	第四節 選擇度的原則	35
134	第五節 教育事象的共同的度	36
134	第四章 測量與方法	39
134	第一節 測量方法的功能	39
134	第二節 方法的基本性質	40
134	第三節 測量方法的判斷標準	45

第五章 觀察法	51
第一節 觀察法的標準分析和充分取樣	51
第二節 適合於觀察法的測量方式	53
第三節 觀察法的利弊得失	54
第四節 觀察法的指導原則	56
第五節 觀察法所用的表式	58
第六節 觀察法的效度和信度	67
第六章 作品分析法和自由反應法	70
第一節 作品分析法和自由反應法的通性	70
第二節 自由反應的引出	72
第三節 作品及自由反應的評閱方法	77
第四節 作品分析法和自由反應法的實用性	86
第七章 導向反應法	87
第一節 問題的種類及特性	87
第二節 導向反應測驗分數的性質	93
第三節 編製導向反應測驗	94
第四節 導向反應測驗的實施	120
第五節 標準測驗	122
第八章 論文式考試	128
第一節 論文式問題的類別	128
第二節 論文式考試的檢討	131
第三節 論文式考試和導向反應式測驗的比較	134
第四節 編製論文式問題的改進	136
第五節 評閱方法的改進	139
第六節 幾個原則性的建議	141
第九章 測量結果的整理和解釋	144
第一節 次數分配	144
第二節 集中量數	152
第三節 離中量數	158

第四節 百分位數和百分等級.....	167
第五節 關係量數.....	173
第六節 標準分數.....	179
第七節 測驗分數的信度和效度.....	182
第八節 常態分配曲線之應用.....	188
第十章 測量、評價和分數.....	197
第一節 測量與評價.....	197
第二節 評價標準	199
第三節 評價的步驟.....	201
第四節 學校分數和成績考查.....	203
參考書目.....	208
附 錄 標準分數在常態分配曲線圖之面積及縱坐標.....	213
漢譯西人姓名對照表.....	223
中英名詞對照表.....	226

附 圖 目 次

圖 四、一	測驗長度和信度係數關係圖	50
圖 五、一	觀察用之體育查核表	59
圖 五、二	小學低年級晨間整潔檢查查核表	59
圖 五、三	體育事蹟紀錄表	60
圖 五、四	評定量表舉例	60
圖 七、一	選擇一個答案式的導向反應問題舉例	88
圖 七、二	寫出一個答案式的導向反應問題舉例	89
圖 七、三	分子排列式的導向反應問題舉例	90
圖 七、四	標準答案紙板	118
圖 七、五	漏空標準答案紙板	118
圖 九、一	50名學生歷史測驗分數分配之直線圖	149
圖 九、二	50名學生歷史測驗分數分配次數多邊圖	150
圖 九、三	50名學生歷史測驗分數累積分配圖	152
圖 九、四	正偏斜分配曲線：均數、中數及衆數之相對位置圖	158
圖 九、五	負偏斜分配曲線：均數、中數及衆數之相對位置圖	158
圖 九、六	離中量數在常態分配曲線所佔之面積圖	167
圖 九、七	由累積分配曲線求百分位數圖	169
圖 九、八	常態分配百分曲線圖	170
圖 九、九	常態分配曲線	180
圖 九、十	標準分數與均數間之面積	189
圖 九、十一	標準分數爲正值之較大面積和較小面積	189
圖 九、十二	標準分數爲負值之較大面積和較小面積	190
圖 九、十三	常態分配曲線圖 ($X=29.75$; $S=6.75$)	192
圖 九、十四	男女生論理記憶測驗之理論次數分配曲線	194

附 表 目 次

表 一、一	現行測驗分類	13
表 一、二	現行測驗出版國別	14
表 一、三	現行測驗絕版測驗比較	14
表 七、一	逐題分析表片斷	100
表 七、二	逐題分析法之問題登記卡片式樣	100
表 七、三	甲乙二生答對各題之或然率	102
表 七、四	各種導向反應題之機會因素，作答時間及難度比較表	104
表 八、一	論文式考試與自編導向反應式測驗：桑代克及黑根兩氏所作之比較研究結論	134
表 八、二	論文式考試與自編導向反應式測驗：羅斯氏的優點及缺點比較表	135
表 八、三	論文式問題暫行評定表	138
表 九、一	學生英文字彙測驗分數之簡單次數分配表	145
表 九、二	歷史測驗分數之次數分配	147
表 九、三	50名學生歷史測驗分數分配	151
表 九、四	均數計算表	153
表 九、五	分組量數中數計算表	156
表 九、六	分組量數四分位數計算表	160
表 九、七	簡單量數平均差計算表	161
表 九、八	分組量數平均差計算表	162
表 九、九	簡單量數標準差計算表	163
表 九、十	分組量數標準差計算表	164
表 九、十一	應用查氏複檢法之計算表	165
表 九、十二	百分位數計算表	168
表 九、十三	百分等級計算表	171
表 九、十四	簡單量數相關係數計算表	174
表 九、十五	積差法相關係數計算表	176
表 九、十六	近似常態之次數分配表	180

表 九、十七	學生二人測驗分數和標準分數比較表.....	181
表 九、十八	兩組學生論理記憶測驗次數分配.....	193
表 十、一	美國 258 所優良中等學校對學校分數之功用之次數分配	204

第一章 教育測量發展簡史

人類行為測量的發展，可劃分為三個主要階段；十八世紀以前，十九世紀，和廿世紀等。十八世紀以前的教育測量十分粗率，更談不到精確，不過，個體能力差異的事實，遠在數千年以前即已受到了重視。十九世紀的科學已有長足的進步，統計方法已可作為研究用之工具，奠定了日後教育測量方法的更新和改良的基礎；此階段的測量，無論在智力，在教育成績，或在人格等方面，都有新的發明，將分別加以敘述。二十世紀的統計方法更為進步，各種各類的測驗大量編製和流通，各級學校也紛紛採用新的測量方法。

第一節 十八世紀以前的情形

一、最早的口試：遠在舊約聖經時代便有下列的一段記載：基列人把守約但河的渡口，不容以法連人過去。以法連逃走的人若說：「容我過去」，基列人就問他說：「你是以法連人不是？」，他若說：「不是」，就對他說：「你說『示播列』」，以法連人因咬不真字首，便說：「西播列」，基列人就將他拿住，殺在約但河的渡口。那時以法連人被殺的，有四萬二千人。（舊約聖經：士師記12:5—6。）

上述的口頭測驗是可資查考的最早的紀錄。孔子（紀元前551—479）和蘇格拉底（紀元前469—399）也用口頭詰問的方法作為教學之方法。此後，口頭問答方式的教學活動便相沿成風，至今仍多所採用。

二、最早的筆試：筆試之採用當較口頭詰問之測量為晚，遠在紀元前二千二百年左右，中國之西周選士制度即已有良好的規模（101:95）。三代以上，家有塾，黨有庠，鄉有序，國有學，州長，黨正，遂師，鄉大夫，為其地之教師。居處相邇，耳目相

習，平居之時，於羣士之道藝德行，孰優孰劣，知之最稔。及大比之日，書其賢者與能者。蓋教之有素，非漫然決優劣於一二日之間也(103:419)。當時情形雖無確切之資料可資佐證，其必有測量之方法，殆無疑問，而此種方法當非口試，馬丁認為當時選士之方法為筆試(223:20)。

三、個別差異的確認：柏拉圖（紀元前427?—347?）以為人的心理可以析成智，情及意欲三大部。據此去看，社會結構也不外三種階級：一是哲學家亦即統治者，潛心於知識的追求，是主於智；一是軍人，潛心於武術，是主於情；一是工商，潛心於財貨，是主於意欲(102:31)。柏拉圖認為各人應適合其階級，予以教育，以充分發展其人格(226:62)。

四、人格型之劃分：希臘醫生希波克拉梯斯（紀元前約460—370）在紀元前四世紀，即依照人體體質將人格分為四大類：多血質，神經質，膽汁質和黏液質等，有四液說之稱。此種學說雖未盡如理想，但亦曾流行了若干世紀之久。

五、最早的教育測驗：早期測量教育成果所用的方法，並非口試亦非筆試，其方法與今日之實作測驗有些相似。諸如原始部落或初民社會教導其第二代之方式，對於狩獵、捕魚、作戰、禮俗等方面之考察，係在青年晉入成人之盛大莊嚴之儀式中，對各種技能和能力，以及勇敢和堅毅等氣質，均藉各種實際行動一一予以測量和評價。

古代斯巴達人尚武好戰，訓練青年之重心則以體魄鍛鍊及堅忍禁慾為主，在紀元前五百年即有考試之舉，以查驗受訓練後之成果(235:16)。古代雅典人崇尚運動及藝術，以運動競技之方式測量運動能力之發展，以比賽或競賽之方式測量寫作及歌詠等能力。

六、最早的學校測驗：中世紀時歐洲的大學已經用口試作為教育測量的方法。波羅家大學（在意大利中北部，以城市之名為

校名，二次大戰時，被嚴重破壞，該校創辦於十一世紀。）在1219年，巴黎大學在十三世紀末葉，學位候選人均須經過口試，以答辯其論文。不過，採用筆試作為學校之測量方法，却遲了四五百年之久，至1702年，英國劍橋大學始採用筆試，當為學校筆試之肇始(227:2—3)

第二節 十九世紀的教育成績測量

一、科學制度：科學制度至本世紀已屬強弩之末，考試方法一仍舊貫，未有改進，測量標準則以詞藻華麗，修辭（姑借用此近代名詞）工整為尚，對於內容涵義則列為次要。近四千年悠久歷史的考試任官制度，在測量方法上多率由舊章，一成不變。

二、美國早期的教育測驗：依照可靠之記載，美國波士頓在1845年即舉行學童之筆試。波士頓學校委員會對所屬各校學童依例每年舉行口試一次；後來學童人數日形增多，口試所費之時間極多而且工作亦甚繁重，逐漸成為例行公事，主考人員敷衍了事。該委員會之小組委員會研究結果，乃決定於1845年起以筆試代替口試。考試科目計有：算術、天文、地理、文法、歷史及自然哲學（即今日之自然科學）。波士頓考試制度在美國教育史上佔有輝煌的一頁。

當時的麻省教育廳廳長是曼·郝瑞斯，他對波士頓的新猷大加讚賞，並在「普通學校雜誌」發表專論，強調筆試實較口試為優，並列舉筆試之優點如下(210:37)：

1. 筆試是不偏不倚的。
2. 筆試對學生來說是公平而精確的。
3. 筆試比舊法考試更為徹底。
4. 筆試可免除教師在無意中影響考試之成績。
5. 筆試可瞭解學生受教之情形。
6. 筆試可消除各種偏袒及成見。

7.筆試之結果可公諸於大眾。

8.筆試問題之難易可由公眾人士評審。

此種改革對近代教育測量概念上的貢獻至鉅，在測量工具及方法上仍未能有所改進。

三、早期的新法測驗：英國小學校長費希在1864年編定了各種測量簿在格林尼治醫院學校應用，這是最早的教育成績測量工具。這種測量簿的種類甚多，計有書法、拼字、數學、航海學、古文常識、文法及作文、法文、通史、繪圖、及應用科學等。書法及繪圖之測量及評價則偏重於質的方面，且有「標準樣本」可資比較，各樣本皆定有等次，學生作品與之比較後，決定該作品與何者最為接近，該樣本之等次即為該作品之分數或等次。其他各科則着重於量的測量，注意實作作品中之錯誤(223:23)。

費希的測量簿是近代教育成績量表的雛形，他的觀念也是近代的測量方法的萌芽；可惜他生不逢辰，當時並未受到重視，是不是他的觀念發生得太早了一點！

四、美國的新法測驗：美國人賴斯是比較測驗的真正發明家，他在1894年便發生了比較測驗的概念，更進而編製了一套拼字表，奠定了新法教育測驗的基礎。該項拼字表曾在美國不同地區的若干學校中應用，於1897年將應用結果及分析報告在美國全國教育協會視導委員會發表，其結論為：學生每日練習拼字三十分鐘為期八年，其學習結果並不優於每日練習拼字十五分鐘為期八年之成績。一時引為異端邪說，除攻擊賴斯本人外還抨擊該項拼字表。十餘年後，客觀的教育測量方法始受到普遍的重視，賴斯之功也。(223:24)

第三節 二十世紀以來的教育測量

一、教育測量專著首次出版問世：桑戴克之巨著「心理及社會測量之理論」於1904年由哥倫比亞大學師範學院印行(246)，

以心理測量及教育測量為主要內容，對後世之影響極大，該書以及桑戴克之學生二者對日後教育標準測驗之發展和普及，具有承先啓後之作用。

二、教育成績標準測驗之印行：斯通受業於桑戴克，在1908年印行算術理解測驗，是為標準化之測量工具問世之始。次年桑戴克之兒童書法量表刊行，是為第一個標準化的教育成績量表，在1909至1915六年間，全套算術測驗出版，尚有分別測量英文作文，拼字、繪畫及書法等能力之量表陸續印行。在早期印行之測量工具中，僅有二種係屬測驗，餘者皆為量表，足堪玩味。

當時教育界人士多反對標準測驗，甚至嘲諷測驗編製者。然而標準測驗仍繼續發展，當不外受下述三種事實的影響：1.關於學校分數正確性與真實性之各種重要研究，一致認為當時之學校分數過分主觀而不正確，勢須有能精確測量教育成績之良好工具。2.由於學校調查運動的影響，促使以客觀方法測量學生能力的努力更趨於積極，也使測驗之編定及使用得到了良好的激勵。3.各大學、教育廳、以及較大學區之研究機構，都積極擴大研究教育測量問題，對教育測驗之普及具有極良好之影響。早期之標準測驗及量表多僅適用於小學，不數年，中學用及大學用之測驗亦相繼出版。

三、自編客觀測驗的提倡：麥考爾於1920年發表專論，建議教師不必完全仰賴標準測驗，可自行編製各種教學用之客觀測驗。四年後，魯治之「筆試之改進」問世，專門討論自編測驗問題，迨1936年，調查結果顯示約有百分之七十四之中學教師自行編製測驗，此外，尚有百分之十教師兼用論文式考試及測驗(228:6)。

四、評價工具及技術之發展：隨着筆試方法之改良和進步，技能，實作作品以及行為測量等評價工具亦有長足之進展，諸如查核表、評定量表、徵詢表、事蹟紀錄表等工具均相繼發明；此

外，面談、個案法、社會測量法以及觀察法等技術都已應用，使測量與評價能更趨於客觀、精確及可信。

第四節 十九世紀智力測驗的萌芽

一、個別差異的科學研究：高爾敦(1822—1911)的「遺傳的天才」於1869年出版，將個別差異科學研究之結果公諸於世，並制定測量方法以研究人類的身體品質及運動能力，高爾敦在教育測量方面最偉大的貢獻並非個別差異，而在於應用統計方法。

二、統計方法的應用：高爾敦發明了一套「標準分數」體系，首創以圖形及線條表示測量結果的關係，氏不僅對教育及心理測量有不可磨滅的功績，在教育研究的科學方法方面亦厥功甚偉。皮爾遜(1857—1936)繼之將統計方法更形發揚光大，為今日教育測量奠定了極穩固的基石。

三、測量智力的開端：美國醫生柴勒早在1887年即使用簡單測驗並定有各種標準，以判斷三歲以下兒童之智愚及心理成熟情形；當時雖未使用心理年齡一語，但已寓有心理成熟指數之意味。及後，開泰爾首先使用心理測驗一詞，時為1890年。本世紀末葉，開泰爾等人採用身體特質，感覺敏銳度及運動能力等項目以測量智力，但其結果未臻理想，且有負相關之趨勢(216:58)。同時，法人比奈(1857—1911)等人也從事類似之研究，於1895年發表十種測驗，以區分各種不同之心理能力，該等測驗以測量較複雜之行爲為主，與開泰爾之重點不同。

第五節 二十世紀智力測驗的發展

一、個別智力測驗的編定：比奈及西蒙二氏的智力量表在1905年編妥，其主要目的在於發現智力低下的兒童，以便予以特殊教育。該量表以實際年齡為依據，應用各種不同方式的測驗，使各問題的難度依照年齡而增加，以便解釋不同兒童的相對智力。

比奈西蒙智力量表分別於1908年及1911年各修訂一次，但基本觀念及精神並未改變，直至今日，此項量表仍為個別智力測驗之典範。1908年修訂之量表，發明心理年齡之概念及其計算方法，對日後智力測驗之發展貢獻殊多。

比奈西蒙智力量表亦盛行於美國，至1916年，杜滿(1877—1955)等人從事修訂，定名為斯丹福修訂比奈西蒙量表，後又於1937及1960年分別修訂，創造了智力商數的概念，此項智商係表示兒童實際年齡與心理年齡之關係(216:101)。

二、團體智力測驗的使用：心理學家均熱心於團體智力測驗之編製，歐悌斯首先於1917年編製成功，是為陸軍甲種測驗，用以甄選第一次世界大戰美國應徵士兵，參加測驗者必須有英文之讀寫能力；另有陸軍乙種測驗，可供文盲和不瞭解英語之人士使用，是為非文字團體智力測驗之始祖。一次大戰之後，教育界人士亦廣泛採用此等測驗，而編印團體智力測驗之風日益增長，顯示對智力測驗已有普遍之興趣，也受到廣泛的注意。總之，測驗技術已大有改進，惟在測量方法上則無新猷可言。二次大戰時所用的陸軍普通分類及陸軍個別心理能力測驗，與早期的陸軍甲乙種測驗相較，亦無何特殊的革新。

三、能力傾向測驗的編造：明斯特堡(1863—1916)曾於1913年編造兩種能力傾向測驗：一為電話接線生能力傾向測驗，一為營業車司機能力傾向測驗；此後二十餘年間，關於機械、音樂、藝術、文書工作等之能力傾向測驗，以及中學和大學各學科的能力傾向測驗，都相繼問世。斯比爾曼於1904年在美國心理學月刊發表「普通智力」論文一篇，主張智力可分為普通因素和特殊因素兩部分，對能力傾向之研究及此種測驗之編造有深遠的影響，有時即稱能力傾向測驗為特殊能力測驗。

第六節 人格測驗的發展