

師範專科學校教科書

心理與教育統計及測驗

下 冊



國立編譯館主編
正中書局印行



版權所有

翻印必究

中華民國七十二年一月臺初版

中華民國七十九年十月臺初版第八次印行

部編本師範專科學校心理與教育統計及測驗
教科書

下冊 基本定價 一元七角

(外埠酌收運費滙費)

主 編 者 國 立 編 譯 館

編 審 者 國立編譯館師範專科學校心理與教育統計及測驗教科書編審委員會

主任委員 路 君 約

委 員 方稚芳 朱進財 李富言 吳鐵雄 林生傳
林正文 林邦傑 林清山 黃元齡 黃金源
黃國彥 黃瑞煥 劉錫麒 劉寶珍 盧欽銘
謝廣全 簡茂發 蔡典謨 龔百順

編 輯 小 組 朱進財 李富言 林正文 黃元齡
黃瑞煥 劉錫麒 盧欽銘 簡茂發

總 訂 正 路 君 約

發 行 人 黃 肇 珩

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

(香港九龍油蔴地北海街七號)

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)

新聞局出版事業登記證 局版臺業字第〇一九九號 (8009) 西

分類號碼：521.89.079

編輯大意

- 一、本書依據民國六十七年三月教育部公布之「師範專科學校五年制普通、音樂及美勞等三科心理與教育統計及測驗課程標準」編輯。
- 二、本書主要供普通科第三學年上、下學期，每週教學二小時；音樂科及美勞科則於第四學年上學期每週教學三小時之用。
- 三、本書分上、下兩冊，上冊內容為「心理與教育統計」，下冊內容為「心理與教育測驗」。
- 四、本書材料之選擇與敘述，以達成部定課程標準中之五項教學目標為原則。
- 五、本書上冊每章之末，附列有習題，藉供學生練習而加深印象。本書下冊每章之末，附列有研究問題，藉供啟發學生思考。
- 六、本書上冊之末及下冊每章之末，分別開列有重要的「心理與教育統計」及「心理與教育測驗」之著作，俾供學生作進一步研究之參考。
- 七、本書若有不當之處，敬請使用本書教學之同道，不吝指教，俾便於再版時修正。

師專 心理與教育統計及測驗（下冊）

目 次

第一章 緒 論	1
第一節 測驗的定義	1
第二節 測驗的功能	3
第三節 測驗的分類	4
第四節 測驗的發展	8
研究問題	14
本章主要參考書目	15
第二章 測驗的基本原理	17
第一節 信度的意義、類型與求法	17
第二節 影響信度的因素	24
第三節 效度的意義、類型與求法	25
第四節 影響效度的因素	30
第五節 信度與效度的關係	31
第六節 信度與效度的應用	34
第七節 常模的意義、類型與求法	42
研究問題	51
本章主要參考書目	52

第三章 能力測驗	53
第一節 智力測驗	53
第二節 性向測驗	75
第三節 創造思考能力測驗	79
第四節 成就測驗	80
研究問題	87
本章主要參考書目	88
第四章 學科成就測驗的編製	89
第一節 論文式考試	89
第二節 新法考試	94
第三節 學科成就測驗的編製程序	108
研究問題	123
本章主要參考書目	124
第五章 人格測驗	125
第一節 人格測驗的意義	125
第二節 人格衡鑑的方法	129
第三節 我國國小適用的人格測驗簡介	146
研究問題	147
本章主要參考書目	148
第六章 測驗的實施	151
第一節 測驗實施的重要性	151

第二節 測驗實施的步驟.....	154
研究問題.....	161
本章主要參考書目.....	162
第七章 測驗結果的解釋.....	163
第一節 常模參照測驗結果的解釋.....	163
第二節 標準參照測驗結果的解釋.....	170
研究問題.....	175
本章主要參考書目.....	176
第八章 測驗結果的應用.....	179
第一節 測驗與教學.....	179
第二節 測驗與輔導.....	185
第三節 測驗與學校行政.....	188
研究問題.....	190
本章主要參考書目.....	191
索引.....	193

表 次

表 2-1	小學生智力測驗分數與智育成績的平均數、標準差及相關係數.....	37
表 2-2	應徵者合格率為 40 % 的泰羅二氏預期表	39
表 2-3	應徵者合格率為 40 % 的勞氏預期表	41
表 2-4	480 名學生的測驗分數之次數分配表及其百分等級之計算過程.....	46
表 2-5	小學高年級學生國語文基本能力測驗百分位數常模 (N = 480) ...	47
表 3-1	我國適用於國民小學的智力測驗.....	56
表 4-1	國小高年級自然科成就測驗雙向細目表.....	110
表 4-2	試題分析的結果.....	116
表 4-3	「范氏項目分析表」式樣	118
表 4-4	試題的鑑別指數與優劣之評鑑.....	119
表 5-1	我國國小適用人格測驗簡介	144
表 7-1	加州心理成熟測驗第二種四個因素與兩個部分的標準分數間差異之 機率	165
表 7-2	某國小二年級學生數學成績報告卡	173
表 8-1	測驗結果成績紀錄卡	181
表 8-2	某生成形成性測驗結果的錯誤分析.....	183

圖 次

圖 2-1	舊槍（左）和新槍（右）射擊之靶面彈著點的比較·····	18
圖 2-2	測驗分數之標準差與測量標準誤的關係·····	34
圖 2-3	測驗臨界分數及工作合格標準之訂定與錄取或棄卻之正誤的關係···	36
圖 2-4	測驗分數對照圖·····	50
圖 4-1	五個程度相等的學生在只有四個論文題的考試中得分互異情形·····	91
圖 5-1	社會關係圖·····	132
圖 5-2	羅夏赫墨漬測驗·····	137
圖 5-3	主題統覺測驗·····	139
圖 7-1	修訂加州心理成熟測驗第二種側面圖·····	166
圖 8-1	教學評鑑的模式·····	180

第一章 緒 論

心理與教育測驗是二十世紀一門新興的應用科學，師範專科學校的學生，將來都要從事教育的工作，教育的對象是兒童，教師不但要了解兒童的心理特性，而且要研究提高教育工作的效率，所以師範專科學校的學生，必須學習各種心理和教育的測驗。

第一節 測驗的定義

所謂測驗，是指在控制的情境中，使用一個或一羣標準化的刺激，引起受試者的反應，藉以對其能力或人格特質作定量的或定性的評量。

在上面這個定義裏，有幾個重要的概念須加說明：

一、標準化

所謂標準化（standardization），含有二方面的意義：一是編製測驗的題目要經過標準化的程序；二是測驗的使用要按照標準化的手續。編造一個測驗，舉凡測驗的設計、擬題、預試、選題、編排和建立常模等，都須經過嚴密的考驗程序。另一方面，測驗實施的步驟，包括測驗前的準備、測驗做法的說明、時間的控制、初閱、複核和換算指數等的記分、對測驗結果的解釋和應用、試場環境的布置以及主試者、襄試者和受試者之間和諧關係的建立等等，都有詳細的規定，使用測驗的人不可變更；而且使用測驗的人必須受過專門的訓練，這樣不論誰來實施測驗，都不會影響受試者的作業，受試者的分數也不因主試者的情感或偏見而轉移。測驗的編造和實施具有這樣精密的手

續，才能成爲一個客觀的科學工具。這樣的手續和過程，就叫「標準化」。

二、情境的控制

測驗旨在測量學生的能力或人格特質，這些特質不像物質的測量那樣具體而固定。影響學生心理現象或行爲特質的因素是多方面的，諸如主試者施行測驗是否按照規定的或標準化的手續，以及受試者的外在刺激，或其內發的需要、驅力、焦慮……等等因素。這些因素必須盡量加以控制，然後各受試者在同一種測驗上所得的分數，才具有相互比較的客觀意義。因此，各種測驗都規定有詳細的實施手續，這些規定都載在各種測驗的「指導手冊」或「說明書」裏，必須按照這些標準化的手續去做，才能減少非必要的干擾和影響至最低限度。

三、量化與測量

人類的智力、性向、知識、技能或人格等，都是一些抽象的特質，不能直接摸到、看到，但是這些特質會從行爲上表現出來：例如一個小孩子說話流利，應對得體，記憶力強，判斷正確，我們就據此推估這個孩子的智力高。學校裏教師根據學生學得知識和技能的多少，評定他的成就（achievement）高下，這些都是從個人具體行爲的表現，間接估量他的心理特質。人類的智力有高低，品性有優劣，能力有強弱，判斷有正誤，我們就把這些程度上的差異，予以量化（quantification）和測量；再將它和一般人的平均水準相比較，藉此對一個人的能力或人格特質作定量的或定性的評量。

綜上所述，所謂「測驗」，就是在控制的情境下，按照規定的手續，向受試者施行一套標準化的刺激，根據受試者的反應，從而對受

試者的能力或人格特質，作定量的或定性的評量的一種科學工具或方法。

第二節 測驗的功能

測驗在學校推行輔導（guidance）工作和增進教育設施的效率方面，都扮演了重要的角色，它的主要功能有三：

一、評 估

測驗的目的在衡鑑學生的心理特性，幫助教師在較短的時間內客觀地認識學生，也幫助學生自我了解。諸如辨別學生的智慧、性向，測量人格特質，藉以了解學生的個別差異，包括學生個別之間的差異和學生本身的所長與所短，以便因材施教，輔導學生充分自我發展。評量學生的成就，考查學生是否努力，使學生和家長知道子弟的學習情形，不但有助激發學習的動機，並可根據測驗的結果，作為分班編組的依據，和決定使用適合學生能力的教材和教學方法。再如比較教育實驗前後的優劣，評鑑學校教育的績效（accountability），以便發現優點，改進缺點。凡此種種，都有賴於測驗的評估（assessment），以提供客觀的資料。

二、診 斷

教師應用測驗，猶如醫師應用體溫計和血壓器診察病人的體溫和血壓一樣，目的在了解情況，幫助診斷（diagnosis），以便對症處方，藥到病除。兒童在學習和成長時期，在生活和學習適應上都難免遭遇困難，測驗用作診斷的工具，旨在發掘學生的優點和缺點，尤其著重困難原因的發現。例如教師分析在測驗上的反應，可發現學生那些

已經學習到，那些沒有學到；如果學生得分低，教師必須進一步找出什麼原因使他得分低。教師對行為發生困擾的學生，不論其為學習上的問題或行為上的問題，必須應用測驗或其他各種方法，診察其困擾的成因，進行教育諮商，然後因勢利導，方能真正解除其困難。在教育行政方面，應用測驗考驗課程設計和教材編選是否適當，教師的教學是否有效。這猶如把測驗當作一面鏡子，藉此修正或調整教學設施，都是一種診斷的作用。

三、預 測

測驗是否有用，就看它能否準確的預測受試者將來的行為。某生在某測驗上得分低，則預測他將來在某方面的成就不會高；大學錄取的學生，則預測他們比被淘汰的學生更能成功的接受高深學術的研究；許多機關應用測驗甄選工作人員，也是基於預測考取的人，將來在工作上會有更優異的表現。測驗具有預測（prediction）的功能，我們才可以了解學生未來發展的可能性，藉此對學生未來的生長作適當的安排，引導他們對升學、選課或就業作明智的抉擇，幫助他們向最有利於自己和社會的方向去發展，以求適才適所，才盡其用。

第三節 測驗的分類

測驗是因應實際需要而產生的工具，隨著應用的目的不同，產生各種不同的測驗。測驗的種類繁多，性質各異，用途亦各不相同。從不同的觀點，有各種不同的分類，茲分別說明於次：

一、依測量的行為特質分

心理學者曾把人類行為的差異概括分為兩大類，即能力上的差異

和人格上的差異。測量能力差異者稱為能力測驗 (ability test)，包括智力、性向、成就等各方面的測量；測量人格差異者稱為人格測驗 (personality test)，包括性格、興趣、態度…等各方面的測量。

(一)能力測驗

1 智力測驗：智力測驗 (intelligence test) 是測量學業潛能 (academic potentiality) 的工具，它可以幫助受試者依其潛能建立其個人的抱負水準 (level of aspiration)。現在的智力測驗往往視其測量的心理功能而命名，如稱為普通能力測驗 (general ability test)，或普通性向測驗 (general aptitude test)，或學業性向測驗 (scholastic aptitude test) 或人員分類測驗 (personnel classification test) 等等。

2 性向測驗：性向測驗 (aptitude test) 目的在測量一個人的可能發展的特殊潛能，用以預測或估計個人如接受訓練後可能在知識或技能方面達到的程度，此類測驗有：

(1)多因素性向測驗：多因素性向測驗 (multi-factor aptitude test)，內容包括若干分測驗，用以測量一個人的多方面的潛能，測量結果，用好幾個分數顯示個人能力的偏向或其所長與所短。

(2)特殊性向測驗：特殊性向測驗 (special aptitude test) 用以測量一個人在某方面的特殊能力，諸如音樂、美術、機械、語言……等特殊能力的測量。

3 成就測驗：成就測驗 (achievement test) 的目的在測量一個人經由學習或教育訓練之後的成就高低。它和性向測驗不同，測量「性向」的目的在衡鑑學生接受教育的可能性大小以及預測其未來可能的成就。測量「成就」則在考查學生學習之後的程度或成績，此類測驗又分學科測驗和綜合成就測驗二大類，前者用以測量學生在某一學科所獲得的知能，後者用以比較個人在各學科上成就的水準。成就測驗可

以幫助教師了解學生的學習狀況，作為改進教學或診斷學習困難的參考資料。

（一）人格測驗

人格（personality）係指個人身心特性的整合總體而言，其內涵相當複雜，所以人格測驗的種類也包羅萬象，包括個人的性格、態度、興趣、道德、價值觀念……等各方面。人格測驗旨在評量個人的典型行為，其反應並無對和錯或及格和不及格之分。

二、依測驗的材料分

（一）語文測驗

受試者使用語言或文字作答者，叫做語文測驗（verbal test）。如果測驗時是使用測驗題本，並以筆或答案紙作答者，也稱為紙筆測驗（paper-and-pencil test）。

（二）非語文測驗

測驗內容以圖畫、圖形、模型、實物、數目等材料為主，而不用文字者，或對不認識文字或聾啞採用手勢指導作答者，都叫做非語文測驗（nonverbal test）。有的測驗規定由受試者拼圖形、排積木、走迷津等操作演示，也稱作業測驗（performance test）。作業或操作測驗因設備限制，為防止測驗中模仿他人的動作，所以慣例上，多屬個別測驗。

語文測驗多以受過教育的識字者為實施對象。非語文測驗則以語文有困難（如測驗文盲或學前兒童，或小學低年級兒童，或聾、啞等）或語文隔閡（如外國學生）為實施對象，有時亦用以對文字測驗成績低劣者，利用非文字測驗予以複核，作為對文字測驗的輔助工具。

三、依測驗實施的方式分

(一)個別測驗

個別測驗 (individual test) 一次只能測驗一位受試者。主試者可細心觀察受試者的反應，這種面對面的測驗方式，對年幼的受試者，可激發其受測的動機。主試者須受專業訓練，適於作個案研究、診斷之用。

(二)團體測驗

團體測驗 (group test) 一次可以同時測驗許多受試者。測驗時間較經濟，可在短時間之內，收集大量的資料，便於作大規模的統計分析，但主試者多無法同時觀察每位受試者的反應。

四、依測驗結果的解釋參照標準分

(一)常模參照測驗

常模參照測驗 (norm-referenced test) 的目的在鑑別學生在團體中的相對地位的高低，所以試題的編製注重難度和鑑別度，讓程度好的學生得高分，程度差的學生得低分；有時即使學生得高分，如果團體中多數人的分數都比他更高，他在團體中的地位，也因而降低。這類測驗試題的甄選，和結果的解釋，均涉及常模團體，故稱「常模參照測驗」。此類測驗的評分方式，容易造成為分數而學習的惡性競爭，測驗分數明顯區分出勝利者和失敗者。

(二)標準參照測驗

標準參照測驗 (criterion-referenced test) 的目的在測量學生的行為，它所注重的是學生學到的有多少？沒有學到的有那些？而不是在團體中贏了多少人，輸了多少人。因此即使某個試題每個學生都答對，也不表示它是一個缺乏鑑別力的題目。學生的表現只與「標準」 (criterion) 相對照，而不與其他學生作比較，故稱「標準參

照測驗」。

第四節 測驗的發展

一、測驗產生的背景

測驗是一門新興的應用科學，其發展歷史雖短，但成長迅速，我們如果追溯它的根源，可說是由於以下四方面的研究所引導出來的：

(一)德國的馮德 (W. Wundt) 於西元1879年創立世界第一所心理實驗室。他用物理學的儀器和方法，研究人類的知覺等心理現象，引發後人以科學方法研究心理問題的興趣。

(二)英國的高爾登 (F. Galton, 1884) 爲了研究個別差異，激發了統計方法的應用。他的成就由其弟子皮爾遜 (K. Pearson) 發揚光大，使統計學成爲發展測驗的重要基礎。

(三)法國的艾斯奎勒 (J.E. D. Esquirol, 1838) 和謝奎 (E. Seguin, 1866) 爲了探索心理變態和低能者的治療方法，研究並設計了一些測驗，以辨別低能和正常兒童。他們的努力，成爲以後發展智力測驗的前導力量。

(四)美國的卡特爾 (J.M. Cattell) 於西元1890年發表「心理測驗與測量」 (Mental Tests and Measurements) 一文，首創「心理測驗」名詞，主張建立常模，以便比較測驗的結果。他認爲心理學如不立基於測驗和測量上，決不能有自然科學的準確。他的學生桑代克 (E.L. Thorndike) 受他的影響，對測驗的發展與普及，作了承先啟後的貢獻。

二、智力測驗的發展

法人比奈 (A. Binet, 1857-1911) 初習法律，後改習醫學

，最後致力於心理學的研究。世界上第一個智力測驗，就是他和西蒙（T. Simon）於西元1905年編製出來的，世稱比西量表（Binet-Simon scale）。

在此之前，比奈氏經過許多年的試驗和努力，後來法國教育部爲了鑑定那些成績低落的兒童，是由於自身智力的低下，還是由於學校教育的不當，於是比奈的研究才得應用的機會。比奈氏所編的智力測驗，成了我們現行的心理測驗的濫觴。

比西量表經西元1908年和1911年先後兩次的修訂，它的成就引起各國心理學者的興趣，於是紛起研究修訂，並且推廣應用到世界各國。現在世上許多重要的國家，都有了符合他們自己文化背景的比西量表修訂本，其中尤以美國斯丹佛大學教授推孟（L. M. Terman）在1916年修訂的斯比智力量表（Stanford-Binet Intelligence Scale）最爲有名。這個斯比量表，以後經過1937年（第二次）、1960年（第三次）、1972年（第四次）的修訂，使量表的功能更見完善。比西量表由推孟氏修訂爲斯比量表，最大的不同是把原來比西量表的心理年齡（Mental Age, MA）改變爲智力商數（intelligence quotient, IQ），以表示個人智力發展的速度。我國修訂的比西量表，即以「斯比量表」爲藍本，至今已有了第四次的修訂本出版。

三、其他各種測驗的發展

(一)早期比奈氏的智力測驗，應用上有兩點不便：一是個別測驗，每次只能測量一位受試者，費時太多；二是語文測驗，對文盲、啞吧或盲者，無法實施測驗。爲了改進這兩個缺點，於是在美國有團體測驗和非文字測驗的相繼誕生。

(二)由於智力測驗測量的結果，只以一個總分代表一個人的能力，給人一種籠統的概念，我們很難根據一個人的智力測驗的結果，對他