

心理測驗

LEONA E. TYLER 著

鄧永秦 譯

文景書局發行

心 理 測 驗

Leona E. Tyler 著
鄧 永 秦 譯

文 景 書 局 發 行

中華民國六十七年三月初版

版 權 所 有
翻 印 必 究

心 理 測 驗

基 價：貳 元 整

著 者：Leona E. Tyler

譯 者：鄧 永 秦

出版者：文 景 出 版 社

發行者：文 景 書 局

地 址：台北市和平東路一段 105 號

電 話：3914280 • 3934063

郵 撥：一 五 七 九 一

登記登：局版臺業字第 1015 號

編譯者的話

以量數的方式 (quantitative approach) 來研究教育問題是心理及教育測驗 (Measurement in Psychology and Education) 所扮的主要角色，也是心理及教育專業訓練中主要課程之一，筆者多年來服務師範教育機構，一直擔任「心理及教育測驗與統計學」的教學，由於受每週教學二小時之限制，因此，在教材的選擇與處理方面，困難殊多，坊間有關心理及教育測驗的書籍固屬不少，但歸納起來，能符合實際教學需要者，却是十分有限，職是之故，筆者乃利用教學之餘暇，編譯此書，以爲國內教育界提供些許服務。

本書原著爲 Leona E. Tyler 所撰之“ Tests and Measurement ”，爲國外大學心理、教育心理、輔導及教育等系學生經常接觸到的參考書，內容簡潔扼要，有關標準測驗中太過理論的部份及繁瑣的測驗樣本舉例均予以省略，而特別着意於基本觀念之建立，對測驗結果的解釋及測驗發展的方向等也都有所說明。筆者曾以該書用作師專三年級測統課程有關測驗部份之教材，以現行教學時數來衡量，一學期足可授完。如大學教育、心理、輔導等系學生或學校教師及輔導工作人員想對心理測驗有所瞭解，本書亦具有很好的參考價值。

爲忠於原著起見，本書之各章節，和原書大致一樣，但是其中有些註脚是筆者增補的，基於事實的需要，特增寫第七章，期使學校教師在自編測驗時，有所遵循，爲求完備起見，彙編「國內現有測驗簡

介」與轉載師大簡茂發教授之大作：「試題分析的方法」，在此，謹向簡教授表示誠懇的謝意。

筆者才疏識淺，譯文謬誤之處在所難免，尚祈國內方家，不吝斧正是幸。

鄧 永 秦 謹識

民國六十七年三月於

臺灣省立新竹師範專科學校

原 序

本書是專為那些忽視心理測驗重要性的人而寫。我們常常發現大學裡心理系一年級的學生對數理統計敬而遠之。他們也知道主修心理學方面的課程，不可避免的需要具有量數的研究知識及測驗技能；同樣地，許多學校教師及諮商人員也知道，他們雖不是測驗專業人員，但測驗可以幫助他們瞭解受教的學生。在美國，心理測驗已廣泛應用在日常生活的領域裡，因此，無論是那個行業的人，對心理測驗都應具備若干基本的認識才行。

為了滿足以上的需要，因而執筆寫了這本小冊子。由於對象及篇幅的限制，因此，一些太過理論及標準測驗編製的方法等部份均刻意予以省略，但對測驗的基本觀念，則詳加闡述，相信讀者閱讀本書後，將會有瀏覽一般學術期刊雜誌的能力，也能明智的選擇適當的測驗及解釋個人測驗的結果。

因為作者的目的是使讀者對心理測驗的基本觀念有個正確的認識，故對有些讀者來說，也許覺得在測驗術語的解釋及計算的舉例方面，似乎是簡單了一點，但是我認為，假如讀者能領悟其中原則的話，舉一反三，應不會有什麼困難，顯而易見的，作者無意以此書涵蓋整個統計及心理測驗的全部內容，相反地，我倒希望讀者讀畢本書後，能引起對心理測驗更深一層的興趣。

Leona E. Tyler

心理測驗

目次

編譯者的話.....	1
原序.....	1
第一章 心理測驗的性質與功能.....	1
第一節 量的需要.....	1
第二節 測量的層次.....	4
第二章 基本統計概念.....	11
第一節 測量的分配情形.....	11
第二節 集中量數的測量.....	14
第三節 離勢的測量.....	16
第四節 相關係數.....	18
第五節 統計的顯著性及概率.....	22
第三章 心理測驗.....	27
第一節 測驗與測量.....	27
第二節 心理測驗的歷史背景.....	28
第三節 心理測驗的效度.....	30
第四節 心理測驗的信度.....	35
第五節 常模和轉換分數.....	38
第六節 心理測驗的分類.....	42
第四章 智力測驗.....	43

第一節	比西量表簡介	44
第二節	智商的意義	45
第三節	魏氏量表	47
第四節	團體智力測驗	51
第五節	幼兒測驗	52
第六節	智力測驗的限制和錯誤的觀念	53
第七節	智力是抽象思考的能力	54
第八節	智力的發展	55
第五章	特殊能力測驗	57
第一節	性向測驗和成就測驗	57
第二節	效度因使用的目的而不同	59
第三節	特殊能力測驗的研究	61
第四節	職業和能力的關係	64
第五節	普通性向測驗簡介 (G A T B)	67
第六節	成就測驗的用途及編製步驟	68
第七節	性向與成就測驗的組合	70
第六章	人格測驗	72
第一節	人格測驗的困難	72
第二節	人格測驗的分類	77
第三節	自陳式的人格測驗	78
第四節	投射式的人格測驗	86
第五節	其他人格測量的方式	90
第七章	教師自編測驗的原則和方法	92
第一節	教師自編測驗的種類	92
第二節	新生測驗一般的編製原則	93

第三節 新法測驗的種類.....	97
第四節 新法測驗的形式及其命題原則.....	97
第五節 論文式測驗的改進.....	104
第八章 心理測驗的應用	107
第一節 測驗和個人的決策.....	107
第二節 測驗和團體差異之研究.....	112
第三節 心理能力和年齡的關係.....	123
第四節 心理特質和身體特質之間的關係.....	125
第五節 遺傳的環境.....	126
第六節 測驗和測量的應用——結語.....	131
參考資料	132
附 錄：	134
試題分析的方法.....	134
國內現有測驗簡介.....	145

第一章 心理測驗的性質與功能

第一節 量的需要

現代科學心理學與過去哲學心理學或日常心理學之間主要區別之一即是對量與質，數字與文字同樣的重視。對剛進入心理系的學生來說，可能感覺不習慣。因為面對那些不熟悉的數量名稱，像平均數、標準差、次數分配、機率及對未能確定的事作困難的決定等等，而這些心理活動根本不存在於他們早期對心理學的概念中，更沒想到心理學須要學這些東西。他們會覺得學習這些東西有如駕車直趨目的地時，被要求改走另一條崎嶇不平且塵土甚多的道路一樣的不合理。

然而，我們知道「數量」的觀念是今天心理學重要的一部份。沒有它，心理學的進步是不可能的。量的方式使我們從實驗工作中獲得精確的結論，實驗時，我們將一些特別程序應用到動物或人類的身上，然後看其效應如何？很明顯地，一個仔細的觀察者，將會發現每一個人對刺激情境的反應都會有所不同，比方說：突然的雷聲，有的人會跑到窗口去欣賞閃電的景象，另一個人可能會將雙手掩面，以示逃避。因此在實驗的過程中，個體的反應也往往大不相同。在許多實驗裡，實驗者如對多少量？及多少個等問題無法回答時，其結果勢必矛盾與紊亂。除非我們對衆多的個體加以研究，否則，我們對人性無法

作較深的瞭解。

在研究學習的歷程時，心理學家希望知道學生們在「邏輯學」的學習上，給他們指定作業時，一開始就不使其發生錯誤較好呢？抑或先讓其做錯，事後加以改正較好？（這種情形在編寫「編序教學」計劃時，尤其重要。）如以實驗設計來回答這個問題，研究人員必須設法使受試者的各種反應以量數的方式表示出來。由於受試者對於實驗的情境反應不同是意料中的事，因此，他不會對有些受試者在情境甲得分較高，另一些在情境乙的得分較高，還有一些在甲、乙情境中得分相等的現象而感到意外。從事學習方式的比較研究，也許第一種學習方式，受試者學習的速率很快，但很容易遺忘，而第二種學習方式，雖然耗時較久，但却有利於記憶的保持。諸如此類的情形，心理學家在從事實驗研究之前必須對這些複雜的可能性有一個較清楚的瞭解。

由於以上的緣故，心理學家會應用許多不同的測量方式去衡量這些可能性。因此，他可能首先測量每個受試者達到某個水準的時間，在實驗之後他可能立即測量其思考推理的能力，一個月後再作同樣的測量。他也可能自行設計一些與推理有關的智力測驗，像解決實際問題的能力或洞悉宣傳上瑕疵的能力等。然後比較甲乙兩組的平均數，應用統計學的知識，他可以機率的方式獲得結果，以下列方式提出他的研究報告：「在我的樣本中，其顯著水準在 0.05 時以第二種方式較好。」如不以量的方式處理其研究資料，他將無法獲得結論，有了量數，他能以機率的方式報告出來，據此作出更進一步的決定——用其中方式之一去進行實驗或放棄該實驗。

測驗不僅可使心理學家們做成正確的決定，也可使個人作成明智的決定。由於心理測驗有利於決定的作成，而這方面的工具甚感缺乏

，故心理學家們乃致力於心理測驗的編造。今天智力測驗的起源可遠溯到 1905 年比奈（Alfred Binet）的量表，比奈設計該量表的動機在協助巴黎教育當局鑑別那些不管是多努力仍無法從正式教育設施中獲得裨益的兒童。從比奈時代起一直到今天，儘管智力測驗的濫被使用而遭到許多嚴厲的批評，但在許多事例中，智力測驗可使教師鑑別學生的智愚，其效果遠比其它方式為佳，卻是一個不爭的事實。

讓我們再舉一個例子來說明，在二次大戰期間，軍事當局各部門徵召兵員時，都以中上的智力，情緒的穩定，較高的文書與機械性向，良好的身心適應能力等作為選擇的條件。顯然地，均具備上項條件的人員並不很多，因此，軍事心理學家所面臨的問題是設法找出那些特別的軍事任務，需要何種的特別能力和技術，其次，他們設計一些測驗來衡量這些能力，然後根據測驗的分數來分配適當的工作。比方說，文書人員的特質是對記載的細節反應迅速，那麼以一個文書性向測驗施行於適合的人員時，其得分必較高。一個在機械性向測驗得分較高的人，應分配接受有關機械的訓練，像飛機修護之類的工作等，由於許多不同的性向待測驗及幾百萬被徵召人員的任務待分配，軍事心理學家所面臨的工作是十分繁複的。他們雖不能使每一個被指派工作的人員充分發揮其能力，然而根據客觀測驗分數來指派工作，遠比主觀的評量方式好得多。因此，選派人員參加訓練工作時，「量」是一個不可缺少的工具。

另外一種對「人」的評量方式促成了一系列測驗與測量技術的發生。由於人們對「心理健康」興趣的增加，知道心理疾病是可以治療的，於是人格測驗乃應運而生。當一個患者求治於精神病醫院或診所時，可用人格測驗來確定其人格的組型及其對治療可能的反應。我們不妨拿某精神病患的例子來作個說明，他應被診斷為「癡呆症」（

schizophrenia) 好呢抑或「焦慮反應症」(anxiety reaction) 較好？精神病醫師可能希望獲知患者對醫院的治療反應較好或對自己社區診所的治療反應較好？當患者進入醫院後，他應被指定參加團體治療(group therapy) 好呢？抑還是某些個別治療(individual treatment) 的方式較好？像這些複雜的診斷很難肯定作成，然而人格測驗却能彌補這方面的不足。

在個人處理自己的生活時，測量技術亦十分重要，有時候，在對自己問題的處理上比處理別人的問題更複雜，雖然許多個人的問題頗不易以「量數」的方式來表達。像某甲知道如果他不在公務機關就業他母親將會極度的失望，或某乙因工作關係他必須離開家裡一段很長的時間而感到憂心忡忡，假如能以「量數」的方式處理這些問題的話，則可使其變得簡易得多。比方說，一個測驗的分數，可以使一個大學畢業生決定他是繼續進研究所好呢？抑還是就業較好？另外一個測驗的分數可讓他瞭解其興趣是和生意人接近呢？抑還是和他同行的學者較接近？高中及大學裡的諮商部門每天都在應用心理測驗來協助學生作成適當決定。

由是我們可以看出，測驗技術與統計方法和普通心理學及應用心理學發生了密切的關係。現在我們將對測驗的工具及其概念，作進一步的探討，純技術方面的知識，姑且不談，但瞭解測驗的基本原理，對研究心理學的人來說，應該不會有太大的困難。

第二節 測量的層次

心理學家們使用「測量」一詞所包括的範疇甚廣。他們共同一致的地方即是「數字的使用」，因此，測量通常的定義是：「依照某些

規則，給予適當的數字符號。」這些規則不如數學知識有限的人們所想像的那樣狹窄，除非我們有充分的認識，否則，我們很可能認為所有四則的運算都應該可應用於所有的測量系統中。而認為：「除非數學的運算法則能被加以應用，否則，測量是不可能的。」以智力測驗為例，一個智商 150 的兒童比一個智商 75 的兒童聰明兩倍的說法是毫無道理的，我們可以說在抽象推理方面，這兩個小孩的反應方式有顯著的不同，但是我們無法以商數二，二分之一，或二比一的方式來表示行為的不同，用一個智力商數去除另一個商數，是沒有什麼意義可言的。

一旦心理學家知道這種現象之後，他們會發現在心理學之外也有類似的情形，且拿「溫度」做個例子，通常我們以華氏或攝氏的刻度來測量溫度的高低。假如溫度從白天的華氏八十度降到晚上的華氏四十度，我們不能夠說午夜的溫度是白天的一半，因為以華氏為依據的溫度，零度並不意味著完全沒有溫度，而它只是個任意的數字，故零度以上的溫度，不能比照測量高度或重量的方式來處理，很明顯地，智力的測量極類似溫度的測量。

對心理學家來說，以史蒂芬氏（S. S. Stevens）（註一）所提出的測量層次最為有用。依照他的看法，我們可用四種層次來劃分。每一種層次有一定的規則及限制，隨著測量層次的不同，所使用的統計方法也有所不同。如圖一所示：

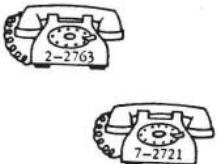
測量層次	限制	圖例
IV 等比量表 (Ratio Scales) 每個數字可視為從零點開始至該點的距離，有零點。	可作數學運算處理，可作各種統計處理。	
III 等距離表 (Interval Scales) 數字之間的距離相等，但無絕對零點。	除可作 I、II 層次的處理外，可作加減及與加減有關的統計處理，但無法作乘除運算。	
II 次序量表 (Ordinal Scales) 數字代表等級或次序。	除可作 I 層次的處理外，亦可作有關等級，大於、小於等統計處理。	
I 名稱量表 (Nominal Scales) 數字僅用來表示名稱分類或區別用。	只能以計數的方式處理及和計數有關的統計處理。	

圖 1 各種測量層次圖例說明

※高層次的測量可作低層次處理，但低層次測量不可作高層次處理，高層次作低層次處理時會喪失可貴的資料。

史蒂芬制使測量的層次符合數學運算的程序。第一種叫「名稱量表」(nominal scale)，這也是最低的一種層次。這種方式是以數字冠在人或物的身上，以資識別，這是種「類別」的分類法，像球衣上繡的號碼，是用來識別球員的，同樣地，也可用這方法來區別一群人，像以1代表男性，而以2來表示女性，這種層次的分類只具有「數數」的運算性質，我們無法以「加、減、乘、除」等計算方式來處理。

其次的測驗層次叫「次序量表」(ordinal scale)。我們將欲測量的個人，依某種特性，從最低到最高的次序加以排列，但此法仍無法顯示出其中任何二者之間的差距是多少，比方說，當一個審查小組欲就五位獎學金候選人中，依其整個資料，作一等級評定時，審查委員們可採用「次序量表」來達到目的，百分等級(percentile)分數常被用來表示測驗的結果，可以算是「次序量表」，這種測量層次，仍不具有「加、減、乘、除」的運算性質，但可應用在統計中的等級等事項上。例如，要調查一群孩子受人歡迎等級和其依賴等級的相關，就應用到這一類的量表了。

第三種測量層次叫「等距量表」(interval scale)。它可以告訴我們兩人或兩物之間的差距是多少，大多數學校的考試就是屬於這種量表，教師們通常認為考試的分數即是等距量表，他們認為比較兩個學生的得分是合理的，比方說，某項考試，甲生比乙生少12分。但此類量表，仍有一個重大的限制，即它沒有真正的零點，一位學生幾何或英文的考試得零分是可能的，但這並不表示他對該科的知識是全然不知，考試的目的在考察學生學習某項學科的知識，我們似乎也犯不著爲了這種缺點（沒有絕對零點）而捨棄考試，因為我們主要的目的在比較學習者個別之間的差異。

等距量表，雖可用加及減的方式處理，但却不能以一個數去除另一個數，因為除法須假設零點的存在，為什麼呢？假如在一次考試中，某甲得 80 分而某乙得 40 分，而主試者又多出十個簡單的題目讓他們兩人做，這時他們都做對了，於是甲獲 90 分，而乙獲 50 分。在這兩種情況中，他們的差數都是 40 分，然而其商數却不相同了，前者是 2，而後者為 1.8。因此，在任何一個考試中，我們無法指出甲的某項知識是乙的兩倍、三倍，或一半。假如試題中每一個項目的代表性（效度）相若，則我們用一個分數減另一個分數，或求其總和後再求平均值，這樣處理並不違背數學的原則，統計上將原始分數換成各種形式的轉換分數（derived score）——我們將在以後的章節中討論到——主要是以加減為主。

第四種測量層次叫「等比量表」（ratio scale），這也是最高的測量層次，我們可用加、減、乘、除等方式來運算，它具有等距量表所有的特性，另外還具有真正的零點。我們可能對此種量表甚為熟悉，因為所有的物理向度，像高度、重量、及體積等，都可用這類量表來測量，對一個誠實的屠夫來說，其磅秤指零時，表示秤盤上根本沒有放肉，假如向他買肉，我們可以相信 6 磅重的烤肉一定比 2 磅重的烤肉重四磅，我們也可以說前者是後者的 3 倍重。等比量表具有用一個數去除另外一個數的特性，進而可用比例的方式來表示。

在心理學上，除了少數的情形以外，許多測量都無法以「等比量表」來使其量化。例如，我們要測量反應時間時，我們可用普通的時間單位秒或幾分之幾秒去測量，假如要知道一位汽車駕駛人，看見紅燈而採煞車踏板的反應速率時，我們可用等比量表來測定：假若甲要 5/10 秒而乙只要 3/10 秒，於是我們可以 3：5 來表示乙比甲快多少。在這個例子中，儘管我們以比例的量表來表示，也許實際我們並