

大學用書

心理與教育測驗及統計

編著者 汪 俊 爵

正光書局出版

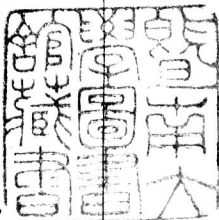
016777

G 449
886

汪 俊 爵 編 著

大 專 用 書

心理與教育測驗及統計



正 光 書 局 印 行

石 景 雲 先生 贈 書

月 日



S9002764

編輯大意

1. 本書分上下兩篇，上篇爲測驗部份，下篇爲統計部份。
2. 本書內容材料測驗部份與統計部份平均分配，各爲五章，總計十章，以免偏頗之弊。
3. 本書測驗部份理論材料與實例材料兼重，統計部份敘述統計與推論統計俱備，以求完整。
4. 本書用作課本或準備考試自學進修均可。
5. 本書材料可適合四學分至六學分教學需要，宜參酌實際教學時間靈活運用。
6. 本書每章末均列有重要練習作業，以增進學習效能。
7. 以所學有限及編寫時間迫促，書稿未能仔細推敲思考，疏漏之處，敬請海內外家，不吝教正。
8. 書末所列參考文獻，僅爲主要參考資料，欠週漏列，在所難免，虔請諒是幸。

汪俊爵

民國七十一年九月於台北市

大 專 用 書

心理與教育測驗及統計

目 次

上篇 統計資料的來源——測驗的實施

第一章 測驗的意義與要件

一、測驗的意義·····	1
二、測驗的種類·····	2
三、測驗的功用與限制·····	3
四、測驗的要件·····	5
五、測驗的發展·····	9

第二章 智力測驗與性向測驗

一、智力測驗·····	18
二、性向測驗·····	47

第三章 成就測驗

一、綜合教育測驗·····	53
二、科別教育測驗·····	54

第四章 人格測驗

一、評定量表·····	76
二、自陳量表·····	84
三、情境測驗·····	96
四、投射測驗·····	97

第五章 標準測驗的實施與教師自編測驗

一、標準測驗的實施·····	104
二、教師自編測驗·····	107

下篇 測驗結果的分析整理——統計方法

第六章 統計學的意義與應用

一、統計學的意義與發展·····	119
二、統計學在教育方面的應用與限制·····	122
三、統計數字·····	124
四、次數分配表的編製·····	126
五、次數圖的繪製·····	132

第七章 平均數

一、平均數的意義及種類·····	139
二、算術平均數·····	139
三、中位數及四分位數·····	145
四、衆數·····	151
五、幾何平均數·····	153
六、調和平均數·····	156
七、各種平均數的比較·····	157

第八章 差異量數

一、差異量數的意義與種類	164
二、兩極差	164
三、四分位差	164
四、平均差	165
五、標準差	171
六、均互差	174
七、相對差數	176
八、各種離差的比較	177
九、偏度	178
十、峯度	180

第九章 相關

一、相關的意義	183
二、相關圖及相關表	184
三、相關係數的計算	189
四、迴歸方程式	198
五、估計標準誤	202
六、其他相關法	205

第十章 推論統計撮要

一、統計推論的涵義與方式	220
二、二項分配	224
三、常態分配	225
四、 χ^2 分配	242
五、 F 分配	247
六、 F 分配與變異數分析	252
附錄一：常用對數表	274
附錄二：倒數表	277

附錄二(續)平方及平方根表·····	286
附錄三：常態分配機率表·····	309
附錄四：七分配機率表·····	313
附錄五： X^2 分配機率表·····	316
附錄六：F 分配機率表·····	319
附錄七：最近兩年高考教育統計學試題及解法·····	325
附錄八：最近兩年普考教育測驗及統計試題及解法·····	328
附錄九：主要參考書目·····	330

心理與教育測驗及統計

上篇 統計資料的來源— 測驗的實施

第一章 測驗的意義與要件

一、測驗的意義

測驗是動詞，也是名詞。如果說要認識個體瞭解個體，必須測量考驗個體的性行特質，測驗就是動詞。如果說要適應個別需要增進教育效能必須藉重各類測驗，測驗就是名詞。比較完整的說，「所謂測驗，就是應用標準化之客觀性的刺激，引導被試者表現某種反應，以度量其智力、性向、學業或人格等心理特質的工具」。標準化的重點，為測驗之實際程序的一致性，好像測驗時間的限制，測驗場所的安靜舒適，測驗指導語的允當劃一，主試被試之和諧關係的建立，測驗常模的規定等，均須標準化，唯一可變的主體，祇有受測之不同的個人。客觀性的重點，為測驗題的取捨，難易順序的排列，分數的評給，結果的解釋等，均須具體明確，不受主持人員或評閱人員主觀的影響。今天我們說教育工作是科學性的工作，教育學識為科學性的知識，多半得力於測驗的應用，因為科學的要件，為測量，數量與證驗，這就非要藉重測驗不可了。麥柯爾（Mccall）說：「心理測量與身體測量是大概相仿的，假使有人說檢查身體不量脈搏，血壓與體溫，我們一定稱之為瘋子，所以如果有人說不要心理測量時，我們也應該稱之為瘋子了」。司開德（Scates）說：「沒有了測量，我們教育上所

2 心理與教育測驗及統計

期盼的科學化的工作將大部分蕩然無存。數量為真理本質的一部分，也是科學所不能缺少的」。由於測驗為對心理特性之客觀性與標準化的測量工具，可作數量上的描述，故為奠定教育科學化的基本要件。

由於心理測量工具的未臻完善，不像物質測量那樣的正確可靠，難免令人懷疑。其實心理測量與物質測量是同樣的可能的，桑代克（Thorndike）於一九一八年說：「任何存在的事物，均可用數量表示」（Whatever exists at all exists in some amount）又說：「要澈底明瞭某一事物的底蘊，知道其量與認識其質同屬重要」。此外麥柯爾（Mccall）於一九二三年說：「凡有數量的，皆可測量」。（Anything that exists in amount, can be measured），以上均證實心理特性是可以測量的，而由於尚未達到盡善盡美的境地，故須吾人繼續研討發掘，其未來發展，實屬不可限量。

上項說明，雖表示心理測量物質測量同屬可能，但由於二者性質不同，是以測量方法亦異。因為物質現象比較具體固定，故可用直接法測量，而心理現象比較抽象變動，不能直接測量，可用間接法測量，即由被試者的行為表現，以間接評量其心理特質，所以在精確度方面，就難免與直接測量有所不同了。

二、測驗的種類

（一）按測驗內容或目標分

1. 智力測驗——評量被試者的一般能量。
2. 性向測驗——評量被試者某一方面的特殊能量。
3. 成就測驗——俗稱教育測驗，評量被試者的學業成績，如國語測驗，算術測驗等。
4. 人格測量——評量被試者的興趣、情緒、態度、德行等。

（二）按測驗的方法分

1. 個別測驗——同一單位時間內，祇能測驗一個被試者，時間方面不太經濟，而主試者訓練有素時，測驗結果較為精密正確。

2. 團體測驗——同一單位時間內，可測驗若干被試者，時間經濟，但對於被試的活動，難作切實控制，故有時效能較差。

(三) 按測驗的材料分

1. 文字測驗——測驗材料全用文字表示或部分用文字表示，均為文字測驗。此項測驗，對於測驗內容文字不熟識者，無法施行。

2. 非文字測驗——或稱圖形測驗，係以圖形為測驗材料，對於識字或不識字者均可適用。

(四) 按測驗的工具分

1. 紙筆測驗——係將測驗印於紙上，或錄於唱片上（或錄音帶上），受試者應用紙筆作答。

2. 動作測驗——係由受試者用手操作物體，圖片、木塊或器械等。由於器材的製作與購置攜帶等方面的困難，以及同時作業時難免彼此觀察模仿，故普通多為個別測驗，或小規模的團體測驗。

(五) 按測驗的功能分

1. 速度測驗與難度測驗——速度測驗為題目多，難度一致（較易），時間短，無人能於限制時間內全部做完。難度測驗為題材由易向難排列，時間寬，無人能規定時間內全部做對。事實上，普通測驗多兼具速度與難度兩項功能，只是重點的偏向不同罷了。

2. 預測測驗與成績測驗——前者功能在預估受試者未來某方面的可能成就，後者功能在考查受試者方面的當前成績。

3. 普遍測驗與診斷測驗——前者的功能為瞭解受試者個人或團體之某方面的大概情況，後者的功能為探究診斷受試者個人或團體之某方面的優點缺點，以為補救教育的依據，一般診斷測驗中常附有練習測驗，藉之克服診斷發現之被試者的缺點，以促進學習的進步。

三、測驗的功用與限制

(一) 功用

1. 評量心性：

4 心理與教育測驗及統計

如智力測驗可辨別智愚，通常多應用智力商數表示：

$$\text{智商} = \frac{\text{心齡}}{\text{生齡}} \times 100 ; \text{或爲智商} = \frac{\text{智齡}}{\text{實齡}} \times 100$$

又如教育測驗可評量學習結果，性向測驗可選拔特殊人才，人格測驗可瞭解人格傾向等。

2. 考查教學成果與學習努力情形

$$\text{成業商數} = \frac{\text{教商}}{\text{智商}} \times 100 = \frac{\text{教齡}}{\text{智齡}} \times 100$$

(附注：教育商數 = $\frac{\text{教齡}}{\text{實齡}} \times 100$ ，可表示受試者教育程度的高低)

3. 診斷困難——最明顯的爲教育測驗中的診斷測驗，可測驗知被試者的學習困難所在，而對之對症下藥，給予合適的練習題材，以解除原有的困難。此外智力測驗，人格測驗等，均具有此項功能。

4. 實驗研究——所有的教育實驗研究工作，不論是單組法，等組法或輪組法，例如對於不同教學方法效果高低的實驗，均須藉重教育測驗，智力測驗等爲工具，否則就是盲人騎瞎馬而要茫然無序不知如何措手足了。

5. 其他——如分編班級、指導升學，輔導就業等，必須應用各類測驗才能允當合適。

(二) 限制

(1) 本身方面的限制

1. 測驗的編造比較費時費事，且於不同時地情況下不宜應用，如改編則甚麻煩。

2. 心理特性之內容確定困難——例如智力的原素，情緒的內涵，態度的項目等，均難確定。

(2) 應用方面的限制

1. 心理特質的變動性較大，例如人格特質，預測效能不太大。

2. 受試對象與常模取樣不易完全適合，故應用常模作爲比較標

標準時，不一定完全適合。

3. 易以測驗項目為學校教材，有時對教育的正常化可能會有妨碍。

四、測驗的要件

我們使用已編好的標準測驗，或自行編製測驗，應審查其必備要件情形，才能運用方便，結果平實正確，穩妥可靠。一個優良的測驗，須具備下列有關要件。

(一) 效度

效度 **Validity** 又稱正確性，係指一測驗能正確的有效的測量出某項特性的程度而言，也就是一個測驗名實相符的程度。評量測驗的效度，由於各類測驗的性質不同，可用不同的方法考核。

1. 內容效度 **Content Validity**：

即一測驗內容，是否包含了所應有的事項，此在成就測驗方面，應用較多。例如——小學算術測驗，是否加、減、乘、除各種試題均有，是否整數，小數，分數各類試題均有，是否各不同難度試題均有等。（與內容效度相對的為表面效度 **Face Validity**，嚴格的說表面效度並不能稱為效度，而為測驗表面的親切感，例如某算術測驗有一試題為某童每天拍皮球 25 下，一星期共拍皮球幾下？或某童每天跳繩 25 次，一星期共跳繩多少次？該題對兒童合適，對成人就缺少表面效度了。如改為某人每月工作 25 天，七個月共工作若干天？就對成人比較合適了）。

2. 預測效度 **Predictive Validity**：

係指一測驗結果與將來發展情形的相關度而言。例如入學考試某科成績與入學後該科成績的相關，智力測驗結果與將來學業成績的相關，人格測驗結果與日後人格特質表現的相關，性向測驗結果與以後實際工作表現的相關等，均為預測效度。

3. 對照效度 **Concurrent Validity**：

6. 心理與教育測驗及統計

係一測驗與效標（公認的有效標準）的相關度。例如新編智力測驗與比納西蒙測驗的相關，智力測驗與學業成績的相關，教育測驗與同類科目在學成績的相關，或短的測驗與長的測驗的相關，團體測驗與個別測驗的相關等，均為對照效度。

4. 特質效度 Construct Validity :

為理論上一測驗所能測量到之特質的正確程度而言。例如智力測驗結果與年齡的相關，教育測驗結果與年級的相關，分測驗與總測驗的相關，某項目在高分與低分被試者間通過人數的百分比是否不同（普通以最高分與最低分各四分之一被試者，或各百分之二十七被試者為準）等，均為特質效度。人格測驗，最為重視特質效度，例如就人格特質中的「焦慮」一項而言，人們遇到危險（如電震時焦慮增加），神經質的人容易焦慮，好焦慮的人比較深思寂靜，期望高求勝心切等，可由實測考驗和應用統計方法計算群體間的差異情形。

(二) 信度

信度 Reliability 又稱可靠性，為某一測驗對被試者所得結果的穩定或一致度。

1. 重測信度 Retest Reliability :

同一測驗材料兩次結果的相關度。

2. 複本信度 Equivalent Form Reliability :

一測驗與其複本結果的相關度。

3. 折半法 Split Half Method :

先求一測驗之奇數項目與雙數項目結果的相關度，再用斯布公式（Spearman and Brown Formula）求信度係數。

$$r = \frac{nr'}{1 + (n-1)r'}$$

r 為信度係數， n 為測驗項目的倍數， r' 為奇雙項目的相關。此間 n 為 2，故上式可改為：

$$r = \frac{2r'}{1+r'}$$

上式假定兩半測驗結果的差異程度相似，但有時亦未必如此。故可改用古特曼氏公式（Guttman's Formula）計算：

$$r = 2 \left(1 - \frac{s_1^2 + s_2^2}{s_t^2} \right)$$

s_t 為全測驗的標準差， s_1 與 s_2 為各半測驗的標準差。

4. 庫李法（Kuder-Richardson method）：

庫李二氏公式係就一次測驗結果考核項目間的一致性（Inter-item Consistency），以評核測驗的信度。

$$r = \left(\frac{n}{n-1} \right) \frac{s^2 - \sum pq}{s^2}$$

n 為測驗的項目數， s 測驗總分的標準差， p 為某一項目通過人數的百分比， $q = 1 - p$ ，各項目 p 與 q 的乘積相加為 $\sum pq$ 。由於測驗編製過程中對各項目難度分析時，已將 p 及 q 求得，故應用較便。

上項折半法與庫李法，對速度測驗均不適用，均易高估速度測驗的信度。由於折半法和庫李法與受測者錯誤題數有關，而速度測驗的信度是應當根據作業的速度計算的，是不應根據錯誤計算的。速度測驗的信度，可由重測法或複本法求得。如用折半法，應將奇數項目偶數項目各成一套，全部測驗時間勻分為二，連續實施。此外可將全部測驗時間勻分為四，每四分之一時間之際，由各受試者於所做項目上作一記號，然後就第一第四部分與第二第三部分核計信度係數。

就以上所介紹之測驗效度與測驗信度言，有效度的測驗，必定具有信度，而有信度的測驗，不一定有效度，沒有信度的測驗，則是沒有效度的。又效度係數，一般皆小於信度係數。此外在適當範圍內增加測驗的長度，信度亦隨之增加，頂高可增至 1；而效度經改進達到某最高限度時，則無法隨長度繼續增加：

$$\text{長度與信度公式：} r_n = \frac{nr}{1 + (n-1)r}$$

r 為原來信度係數， n 為長度倍數， r_n 為長度 n 倍後的信度係數。

$$\text{長度與效度公式：} r_n' = r_v \frac{\sqrt{r_t'}}{r_t}$$

r_n' 測驗長度增加 n 倍後的效度

r_v 測驗原來的效度

r_t 測驗原來的信度

r_t' 測驗增加 n 倍後的信度

由於 r_v 恒小於 r_t （即效度係數的數值小於信度係數的數值，縱然 r_t' 高至 1， r_n' 亦不可能達到 1，而有其一定限度，是以測驗長度的增加，對效度的影響較小。

(三) 常模 Norm

所謂常模，係指一測驗對相同情況受試者所得之平均結果而言。普通可分為年齡常模與年級常模兩種，智力測驗須應用年齡常模，教育測驗常應用年級常模。常模的合用度，常隨所據個體的多寡與符合隨機取樣的標準程度而定。智力測驗方面的心齡（即智齡）就是一種年齡常模，而智齡的單位不相等，有隨年齡縮小的趨向，此由於智力於年幼時發展速，年長時發展緩，是以對應實齡而來之心齡亦隨之縮小，好比五歲時較常模落後或超前一年，其相差較十歲時落後或超前一年為大。

年齡常模或年級常模的應用，多限於兒童或青少年，對於成人往往難以適用。同時與年齡或年級無關的心理特質，如人格特質，常無法用年齡或年級表示。為了補救此項困難，另有百分位數常模與標準分數常模，俟後於下篇統計部分敘述，此間從略。

(四) 複份

係指一測驗有兩套以上相等的材料而言，必須內容相似難度相同

始可。可彼此核對，或比較真確的測出不同時期的實質變動，此在成就測驗方面，最為需要。

(五) 方便經濟

1. 測驗的說明（指導語）是否簡單明瞭？
2. 測驗的印刷排版是否清楚醒目？
3. 測驗的做法是否簡便確定？測驗中的分測驗是否太多太繁？
4. 測驗時限的規定，是否長短合適？
5. 測驗的評閱是否方便？
6. 測驗的試題紙與答案紙是否分開？
7. 測驗結果應用的建議，是否明確有效？
8. 測驗的價格是否經濟等。

五、測驗的發展

任何學識的肇始與形成，均是有本有源而逐漸成長的，測驗的發展，亦屬如此，祇是歷史稍短罷了。茲分智力測驗，教育測驗，人格測驗與我國測驗的發展四方面撮要介紹如下：

(一) 智力測驗的發展

智力測驗的產生，是由於適應社會需要而來的。人類的能力，有高低的不同，有智愚的差異，我國孔子說：「唯上智與下愚不移」。又說：「中人以上，可以語上也，中人以下，不可以語上也」。希臘的柏拉圖於理想國一書中，也將人類的秉賦分為三等，分別接受適當的發育，以從事不同的適合個性的工作，均為對人類智力差異的重視。近代教育方面對適應個別差異的特別注意，期能人盡其才，才盡其用，更加速了測驗的進展。

德國心理學家馮德（W. Wundt）於一八七九年在萊比錫創立心理實驗室，使得心理學從哲學的領域邁入科學的道路，從文字的敘述，而兼重數量的表達，實為智力測驗的先導。不過當時所研究的問題，多限於視覺，聽覺等反應速度的測量，偏重於行為一致性的尋求，而

未能顧及個別差異的事實。英國人類學家高爾登（F. Galton）專心於遺傳的研究，倡導優生學說，致力於人與人間差異的探對，應用統計方法研究人與人間的個別差異，於一八八三年以感覺辨別力的測驗（例如辨別熱、冷、痛的能力）作為評定智能的工具，創立以圖示法表示相關，其對統計分析方面的貢獻，由於門人皮爾遜（Karl Pearson）和斯皮門（Charles. E. Spearman）的廣續推展，更為發揚光大。美國心理學家喀推爾（J. Mckeen Cattell）師承馮高二氏的研究於一八九〇年提出「心理測驗」一詞，為此一名詞的開創者。而真正編造客觀智力測驗的第一人，則為法國的比納（A. Binet），他於一九〇五年與西蒙（T. Simon）合作發表「比納西蒙量表」，實為近代評定智力的第一個量表，所以有人稱比納為智力測驗的鼻祖，可說是名實相符。以後於一九〇八年和一九一一年經過兩次修正，使之較前完善。

一九〇五年量表 本量表計有測驗題三十個，可從多方面觀察智力，較以前心理學者只單獨測驗感覺反應速度或感覺辨別能力為正確可靠。同時本量表的測驗題，由易至難排列，可分別高低不同的智力。

一九〇八年量表 該量表較一九〇五年量表改善進步，將原有測驗擴充為五十九個，並將測驗題按照難度依年齡分組，組別自三歲組至十三歲組。同時將測驗結果以年齡表示，稱為智力年齡（Mental age），智力年齡的概念從此確定。

一九一一年量表 本量表與一九〇八年量表不同處有二，一是刪除不當的舊有題目，增加了新的測驗材料，並將題目次序重加排列，一是每年齡組的題目，一律改為五個，使得計算智力年齡時較前方便。

比西量表於法國刊行後，首先傳入美國，其中最具有貢獻的，為美國司丹佛大學（Stanford university）教授推孟（L. W. Terman）於一九一六年發表司丹佛修正量表（Stanford Binet Scale），共有測驗題九十個，其中五十四個為比西量表所有的，另三十六個為