

遵照部頒課程標準編著

師範學校及鄉村師範學校

教育測驗與統計

編著者 王書林

正中書局印行

本書於廿六年八月十八日領有教育部師字第十四號執照

翻印必究

中華民國二十四年七月京初版
中華民國三十五年十月滬三版

師範
鄉村師範

教育測驗與統計

全一冊 定價國幣一元四角

(外埠酌加運費匯費)

編	著	者	王	書	林
發	行	人	吳	秉	常
印	刷	所	正	中	書
發	行	所	正	中	書

(116)

目 次

第一章 緒論

第一節	測驗的意義…	1
第二節	測驗的功用及限制…	9
第三節	測驗的歷史…	14
第四節	測驗的種類…	23

第二章 智力測驗

第五節	智力的解釋…	28
第六節	個人智力測驗…	32
第七節	團體智力測驗…	51
第八節	測驗實施的方法…	63

第三章 教育測驗

第九節	國語測驗…	69
第十節	書法測驗…	89

第四章 教育測驗(續)

第十一節	算術測驗…	95
第十二節	自然科測驗…	110
第十三節	史地測驗…	111

第五章 教育測驗(續)

第十四節	英文測驗	...	...	...	...	...	117
第十五節	圖畫測驗	...	...	...	...	...	124
第十六節	音樂測驗	...	...	...	...	...	126
第十七節	常識測驗	...	...	...	...	...	128
第十八節	學校調查用測驗	...	...	...	...	...	130
第六章 新法考試							
第十九節	新法考試	...	...	...	...	...	133
第七章 其他測驗							
第二十節	體格測驗	...	...	...	...	...	144
第二十一節	品格測驗	...	...	...	...	...	148
第二十二節	職業測驗	...	...	...	...	...	158
第八章 普通統計及報告方法(續)							
第二十三節	次數分配	...	...	...	...	...	167
第二十四節	平均數	...	...	...	...	...	174
第二十五節	離中差數	...	...	...	...	...	184
第二十六節	百分數	...	...	...	...	...	200
第九章 普通統計及報告方法(續)							
第二十七節	相關	...	...	...	...	...	206
第二十八節	常態曲線	...	...	...	...	...	221
第十章 普通統計及報告方法(續)							
第二十九節	表列法	...	...	...	...	...	229

目	次	3
第三十節	圖示法… … … … …	233
第十一章	編造測驗的方法	
第三十一節	編造測驗應注意事項… … … …	250
第三十二節	<i>TBCF</i> 制 … … … … …	256
第三十三節	品質量表的編造法 … … … …	284
附錄		
中西人名對照表		
中西名詞對照表		

目 次

表 一	一年級生的工作品質與智齡	...	...	...	...	11
表 二	智力分級	...	...	...	...	36
表 三	預備測驗的 T 分數	...	...	...	...	44
表 四	各年齡的 B 改正數	...	...	...	...	46
表 五	廖氏團體智力測驗正測驗題數與練習題數的比較	...	...	...	...	52
表 六	艾氏英語測驗量表乙第一類的常模	...	...	...	...	123
表 七	陳鶴琴小學常識測驗第一類的內容	...	...	...	...	129
表 八	學校自用普通教育測驗的內容	...	...	...	...	131
表 九	七位國文教師給分比較	...	...	...	...	135
表一〇	五十個學生測驗的成績	...	...	...	...	169
表一一	次數分配表示例	...	...	...	...	172
表一二	中數的求法示例	...	...	...	...	177
表一三	均數的求法示例(量數已歸類 組距爲1)	...	...	...	...	179
表一四	均數的求法示例(量數已歸類 組距大於1)	...	...	...	...	180
表一五	均數的簡捷求法示例(一)	...	...	...	...	181
表一六	均數的簡捷求法示例(二)	...	...	...	...	183
表一七	上下四分點的求法示例	...	...	...	...	186
表一八	平均差的求法示例(一)(量數未歸類)	...	...	...	...	188

表一九	平均差的求法示例(二)(量數已歸類)	...	...	188
表二〇	平均差的簡捷求法示例	...	...	191
表二一	標準差的簡捷求法示例	...	...	195
表二二	百分數求法示例	...	...	202
表二三	等級相關求法示例	...	...	208
表二四	等級相關的簡捷求法示例	...	...	209
表二五	積矩相關的求法示例(事實未歸類者)	...	...	213
表二六	積矩相關的簡捷求法示例(事實未歸類者)	...	...	216
表二七	常態曲線標準差值與縱線比例對照表	...	...	223
表二八	常態分配面積與橫距的關係	...	...	226
表二九	我國初等教育與各國比較表	...	...	230
表三〇	民元以來全國中等學校校數與學生數的趨勢	...	...	231
表三一	全國初等教育概況統計分表	...	...	232
表三二	求 T 量表方法示例	...	...	259
表三三	T 分與百分數對照表	...	...	260
表三四	擴充 T 量表方法示例	...	...	261
表三五	四種量表的比較	...	...	263
表三六	求 B 分的方法示例	...	...	269
表三七	實足年齡與 T 分對照表	...	...	273
表三八	智商與 B 分比較表	...	...	274
表三九	廖世承團體智力測驗 T 分與 G 分對照表	...	...	277

表四〇	各年級的分數次數表…	278
表四一	G 分數與 T 分數對照表…	279
表四二	C 校正數與距開學月對照表…	280
表四三	C 校正數測驗月對照表…	280
表四四	T 量表與年齡量表各種數量對照表…	282
表四五	專家評判若干篇作文的結果示例…	284
表四六	機誤值對數表…	285
表四七	表示將百分比的差別化爲機誤差別的方法…	286

圖 目

圖 一	難度量表圖	9
圖 二	常態分配圖	17
圖 三	142 教師對於一本英文試卷所給的分數分配	136
圖 四	次數分配直方圖	173
圖 五	次數分配多邊圖	174
圖 六	標準差量表圖	198
圖 七	百分曲線	203
圖 八	相關係數圖	207
圖 九	等級相關圖	211
圖一〇	積差相關分布圖	218
圖一一	常態曲線與次數多邊比較圖	224
圖一二	某城教室光線與標準教室光線的比較	235
圖一三	國立中央大學十三年度與十八年度學生數的比較	235
圖一四	浙江全省中學校教職員資格百分比圖	236
圖一五	一年來各省市幼稚園小學校數增減比較圖	236
圖一六	各省市小學校數比較圖	237
圖一七	濟南高等國民小學校各年級男女學生人數的分配	238
圖一八	民國十八年度全國各級學校經費百分比比較圖	239

圖一九	某地 2229 個在學兒童由體格檢查所發現的不良 狀況 240
圖二〇	二十年度國立中央大學實驗學校課程的支配 ... 241
圖二一	完全中學校的系統組織圖 241
圖二二	全國中等學校分布圖(一)學校數 242
圖二三	國軍遺族學校歷年添置圖書比較圖 243
圖二四	曲線代表一類事實圖 245
圖二五	不正確的聯絡曲線圖 245
圖二六	正確的聯絡曲線圖 246
圖二七	表明數目過多的曲線圖 246
圖二八	對數格線紙上曲線圖 247

第一章 緒論

第一節 測驗的意義

一 教育測驗的基本概念

教育事業中有三個大問題：1.在兒童方面我們究竟應該造成何種變化？2.怎樣去造成這種變化？3.這種變化究竟已否造成？第一個問題是教育目標問題，就是教育哲學，教育社會學的問題。第二個問題是教育方法問題，就是教育心理學的問題。第三個問題是教育結果問題，就是教育測量學的問題，也就是本書所討論的問題。

從有教育以來，就有考試的制度。不過舊的考試方法有許多缺點（見第六章），所以測驗代之而興。測驗（test）是教育上一種新方法，——一種標準的客觀的考試方法。它所測量的方面，或為教育成績，或為智力，品格等特性。但是任何一種新的方法，開始時，總為一般不了解的人們所盲目地反對或盲目地贊成。故自測驗方法發明以後，信仰的人以為它是可以解決教育上一切的問題，反對的人以為它是無稽的東西。其實這兩

種人都未明瞭測驗的基本概念。

測驗的基本概念是什麼呢？

甲 凡是存在的東西必有數量 美國的心理學家桑戴克 (Thorndike, E. L.) 說：「隨便什麼東西，祇要是存在的，總存在於數量之中。」世界上一切現象，如高低，美醜，輕重等，都不能一律，都有程度的不同。所謂程度的不同，就是數量的不同。教育事業當然不能例外。兒童的智力 (intelligence) 是個別的，有智愚的分別；學生的成績也是個別的，有優劣的區分。所謂智愚，所謂成績優劣，都是程度的不同。

乙 凡有數量的東西都可測量 麥柯爾 (McCall, Wm. A.) 說：「隨便什麼東西存在於數量之中的，都可以測量」。不過反對心理測量者，以為有些特性是可以測量的，有些特性是不能測量的。這是一種錯誤的見解。凡是有數量的現象，都有被測量的可能性。不過有些現象，我們已有完善的工具；有些現象，測量的工具，尚在試驗之中，或尚未發明，我們決不可因某種現象的測量工具尚未完善或尚未發明，就說它是不能測量的。

丙 被測量現象的必具條件 可以被測量的現象，應具下列三種條件：

子 現象的存在性 現象的存在，有些可以看見的，聽到的，接觸的，嗅出來的，或嘗出來的，有些是思維得來的。

丑 現象的公同性或同一性 物質現象中的同一性，許多

是顯而易見的，如高度，美醜，輕重等。心理現象中的同一性，有許多不是這樣的顯明，例如某甲的算術能力如何，我們不能一望而知。

寅 現象的相等距離的可能性 有許多存在的現象，其同一性已爲人所共認，但是仍舊不能測量，因爲工具尚未完善。換言之，就是測量的單位尚未求得。單位的根本條件，是要有相等距離。所謂相等距離，就是第一單位與第二單位間的距離等於第二單位與第三單位間的距離（關於單位之討論，參閱二，乙）。

丁 心理測量採取間接測量法 許多物質的測量，因爲根據於空間或時間，其單位爲人所共見。心理的測量（mental measurement）往往採取間接測量法（indirect measurement），所謂間接的測量，就是測量心理過程的外表的、物質的、身體的，或生理的因果。例如智力測驗所測量的特性是先天的能力，此種能力，不能直接測量，祇可從其所表現作業中估定。譬如欲知一個三歲兒童的智力若何，我們示以一張圖畫，要其列舉圖中的物件。此種物件的名稱 兒童都是後天學習得來的。不過我們由統計的結果，知道凡普通三歲兒童都應該知道；不知道的，在心理的機能上或有缺陷。由後天的作業成績推測其先天的能力，這就是間接測量法。這種方法，應用時一不留心，將失了真義。譬如圖中有一汽車，而被試者或從來未見過

汽車，測驗的效用就失掉了。所以編造測驗時，何種問題可用，何種不可用，非經過詳細的考慮與精密的統計不可。

二 測量的要素

一種測量，不論是物質的測量或心理的測量，必具有下列幾個要素：

甲 參照點 (reference point) 無論那種測量，都須有一參照點。參照點即計算的起點，經度的參照點爲英格蘭的格林尼治 (Greenwich)，緯度的參照點爲赤道。參照點有兩種：1. 絕對的零點；2. 人定的參照點。輕重的參照點，爲恰好一點沒有的重量，是絕對的零點之一例。教育測量中所用的參照點，都屬於第二類，因爲智力或國文能力等特性的絕對的零點甚難決定。惟用人定的參照點，大家必須商議一個統一的辦法，否則，各種量表中分數的意義不同，沒有方法來比較，至少比較的步驟太繁了。譬如有人以南京爲緯度的參照點，則現在所謂北緯十度的地方，決非以南京爲參照點的北緯十度的地方了。現行教育測驗有以一年級學生平均作業爲參照點者，有以某年齡兒童平均作業爲參照點者。我國測驗，大部採用麥柯爾的 T 量表法，乃以十二歲兒童平均作業爲參照點。

乙 單位 (unit) 單位的根本條件，上面已經說過。惟現行教育測量中所用單位，不但是不合相等距離的標準，且沒有

統一的單位。教育測驗中所最常用的量表有四種：1. 年齡量表 (age scale), 2. 年級量表 (grade scale), 3. 百分量表 (percentile scale), 4. 差異度量表 (variability scale), 每種的單位不同。

年齡與年級量表均以年月為單位, 所謂智齡 (mental age), 教齡 (educational age), 級模 (grade norm) 等就是 (智齡等討論見第二章)。此種單位的缺點, 則在單位價值的不相等。據我們現在所知, 智齡四歲與智齡五歲間 (或一年級成績與二年級成績間) 的差別大, 智齡十四歲與智齡十五歲間 (或五年級成績與六年級成績間) 的差別小。 (關於此點, 本書中不能詳細討論, 讀者可參考王書林心理與教育測量。)

百分量表是以百分數 (percentile) 為單位 (百分數討論見第八章)。概括言之, 一個兒童在測驗上得百分點三十者, 表示在類似的 (同年齡或同年級) 兒童中, 有百分之三十人不如他。百分數的缺點亦在單位的不相等。據統計學中的討論, 在量表的兩極端, 兩個百分數間的距離大, 在量表的中段, 兩個百分數間的距離小。

差異度量表有兩種, 一為作業差異度 (variability-of-performance) 量表, 二為評判差異度 (variability-of-judgment) 量表。我國所採用的 T 量表屬於第一種。品質量表 (product scale) 如綴法量表, 書法量表等屬於第二種。它們的單位或為

標準差，或為機誤（標準差與機誤的求法見第八章）。至於 T 等於十二歲兒童某種特性的分配的標準差的十分之一，其意義及編造方法，將在第十一章中詳細討論，現在暫不敘述。

三 測驗的條件

甲 效度 (validity) 物質可以直接測量，而心理方面的特性祇能間接測量，所以必須設法引起被試者的行為，然後測量才可進行。測驗即引起行為的工具。惟各種測驗方法，不能絕對正確。因為心理的現象，變化萬出，因子複雜，欲測量全無差誤，幾不可能。差誤可分為常差 (constant error) 與變差 (variable error) 兩種。若目的在測量算術能力，而問題的文字深奧，則發生常差。常差使測驗的效度降低，故測驗第一個條件必須有效。所謂有效的測驗，就是一個測驗能夠真確的測量所要測量的東西。普通決定測驗的效度方法，應用統計學上的相關法 (the method of correlation 求法見第九章)，把測驗的分數與效標 (criterion of validity) 求一相關。所謂效標即求效度的標準。好的效標的求得極其困難，因為最好的效標，是所欲測驗的真實量數；此種量數往往不易求得。若此種完全效標可以得到，那又何必測驗。效標既不完全，則測驗的真正的效度不能決定。所以不但測驗本身是間接的測量，就是測驗的效標也是間接的。

乙 信度 (reliability) 常差使測驗的效度降低，變差使測驗的信度變小。測驗中變差的來源很多，不勝枚舉。例如甲乙二兒童，甲見過火車而未見過輪船，乙則相反。今若以輪船為測驗問題，則乙成功而甲失敗；若以火車為測驗問題，則乙失敗，而甲成功。甲乙二生兩次的相差，都由於變差。又如甲乙二生同受測驗，而乙生的鉛筆忽斷，於是甲多做了三題。這三題的相差，亦由於變差。所以變差使兩次測驗的結果不同。因而降低測驗的信度。一個可信的測驗是指以一個測驗測量某人的某種能力時，假使施行測驗的手續與記分方法相同，今日量之如此，明日量之亦如此；或者施以兩種相同的量表，結果相同。表示測驗的信度方法，亦用相關法。這個相關係數，專門名詞稱為信度係數(reliability coefficient)。

丙 常模 (norm) 上面已經說過，測驗是用一羣標準的刺激以引起人們的行為，我們再根據所引起的行為以估計人的某種特性。估計時最要的方法是參考他人的行為以定其優劣。所以測驗中必有常模。常模不是理想，乃是許多人在某種測驗上所得的平均分數。常模有兩種：1. 年齡常模（智齡等），2. 年級常模（級模）。所以求常模的方法是很簡單的。我們祇須將測驗去測量許多兒童，求某年齡或某年級兒童所得的平均分數，這個平均分數就是某年齡或某年級的常模。常模求法雖然簡單，但是求得常模時所用被試人的選擇，卻是一個重要問題。