

大學叢書

心理與教育測驗

孫邦正著

文通書局印行

大學叢書

心理與教育測驗

孫邦正著

文通書局印行

中華民國三十七年一月貴陽初版
中華民國三十七年二月上海第一版

大學叢書 心理與教育測驗 一冊

馬版西白報紙本

定價國幣拾肆元

(印刷地點外另加運費)

編著者 孫邦正

發行人 華問渠

印刷所 文通書局印刷廠

發行所

文通書局

上海 廣州 長沙

貴陽 昆明 成都 重慶

弁 言

本書的目的有三：

一、現在我國已出版的心理測驗和教育測驗的書籍，其中所介紹的測驗材料，大多是民國十一年前後所編造的，這些測驗材料，經過了二十年之久，到現在已不甚適用。我們若用這些陳舊的測驗作為教學的材料，在教學時教者學者都常感到索然乏味。實際上，近幾年來，國內教育研究機關已編成新的心理測驗和教育測驗約二十餘種，可以代替舊的測驗，然而一般測驗書籍中，對於這些新出版的標準測驗，尚未有作詳細介紹者，以致教學時，常苦於不易搜集材料。在這本書內，作者將我國最近出版的標準測驗詳為介紹，學者學習之後，即可實際應用，學習的興趣當較為濃厚，而今後測驗教學上的一個困難問題，也可以因之解決了。

二、近來作者擔任教學「心理與教育測驗」一學程時，常感到缺少一本較完善的參考書。現在已出版的測驗書籍雖有幾本，或者內容失之龐雜，徒眩學生的耳目；或者內容失之簡畧，不易使學生瞭解。所以作者寫這本書的時候，在選材方面頗費斟酌，務求詳畧得當。此外，在文字上，力求簡明流暢；在說理上，亦力求深入淺出，使學者易於瞭解。

三、一般測驗書籍，對於各種測驗的應用方法和編造測驗的手續，多畧焉不詳，以致學生學習之後，不僅應用測驗時方法上常有錯誤，更缺乏編造測驗的知能。在這本書內，作者對於各種測驗的應用方法和編造測驗的手續，詳為闡明，以期學者能由此獲得運用測驗和編造測驗的

知能。

本書曾在四川省立教育學院試用數次，成績尚好。書中所引測驗材料甚多，謹在此向各編造者一併誌謝。

編者識

目次

第一章 緒論	1
第一節 測驗的意義	1
第二節 測驗應具的條件	3
第三節 測驗的功用和限制	7
第四節 測驗的歷史	10
第五節 測驗的種類	15
第二章 智力測驗	18
第一節 智力的解釋	18
第二節 個別智力測驗	20
(一) 陸志韋第二次訂正比納西蒙智力測驗	20
(二) 蕭孝嶸修訂墨跋量表	27
第三節 團體智力測驗	33
(一) 廖世承團體智力測驗	34
(二) 蕭孝嶸訂正古氏兒童智力測驗	40
(三) 蕭孝嶸小學入學智力測驗	41
(四) 黃覺民訂正幼童智力圖形測驗	45
第三章 教育測驗	51
第一節 國語測驗	51
(一) 默讀測驗	51
(二) 默字測驗	68
(三) 文法測驗	70
(四) 綴法測驗	71

(五)國文常識測驗.....	57
第二節 書法測驗.....	76
第三節 算術測驗.....	77
(一)算術四則測驗.....	79
(二)算術應用題測驗.....	87
(三)算術練習測驗.....	91
第四節 自然測驗.....	94
第五節 社會測驗.....	98
第六節 史地測驗.....	104
第七節 常識測驗.....	113
第八節 英文測驗.....	118
第九節 圖畫測驗.....	124
第十節 音樂測驗.....	127
第十一節 體育測驗.....	131
第四章 其他測驗.....	149
第一節 品格測驗.....	149
第二節 職業測驗.....	158
第五章 測驗實施的方法.....	168
(一)選擇適當的測驗材料.....	168
(二)施行測驗前的準備工作.....	168
(三)開始測驗時應注意的事項.....	168
(四)測驗進行時應注意的事項.....	169
(五)測驗結束時應注意的事項.....	171
(六)施行測驗後的整理工作.....	171
第六章 整理測驗結果的方法.....	172

第一節 次數分配	172
(一) 決定全距	173
(二) 決定組數和組距	173
(三) 做次數分配表	174
第二節 均數	175
(一) 平均數	175
(二) 中數	179
(三) 衆數	181
第三節 離中差數	182
(一) 四分差	182
(二) 平均差	185
(三) 標準差	187
(四) 比較差數	191
(五) 偏態性的測量	192
第四節 相關	193
(一) 相關的意義	193
(二) 相關係數的求法	194
第五節 量數的可靠度	198
第七章 測驗結果的解釋	201
第一節 年齡量表中各種量數的意義	201
(一) 智力年齡與教育年齡	201
(二) 智力商數與教育商數	202
(三) 成業商數	203
第二節 年級量表中所用量數的意義	205
級模	205

第三節 百分量表中所用量數的意義	205
百分等級	205
第四節 T 量表中各種量數的意義	206
(甲) T B C F 分數的意義	206
(乙) T B C F 分數的求法	208
第八章 編造測驗的方法	214
第一節 編造測驗的普通手續	214
(一) 選擇測驗材料	214
(二) 編造測驗題目	217
(三) 決定配分方法	220
(四) 試用測驗初稿	222
(五) 刪訂初稿	223
(六) 編造複份	223
(七) 決定測驗時間限度	224
(八) 求常模	225
(九) 決定測驗的正確度和可靠度	231
(十) 編製說明書	233
第二節 T B G 量表的編製法	234
(一) T 量表的編製法	234
(二) B 量表的編製法	237
(三) G 量表的編製法	241
第三節 品質量表的編製法	243
第九章 新法考試	246
第一節 考試時測驗方法的應用	246
第二節 新法考試的種類和編造規則	247

第三節 新法考試的編造步驟	251
第四節 新法考試的優點及其限制	252
附本書參考書報目錄	254

心理與教育測驗

第一章 緒論

第一節 測驗的意義

(一)測驗的性質 教育上有三個重要問題：第一，教育的目標是什麼？這是教育哲學和教育社會學的問題；第二，用什麼方法去達到教育目標？這是教育心理學，教育行政學和教學方法的問題；第三，教育目標已否達到？這就是心理與教育測驗的問題了。所以測驗是教育上的一種工具，用以度量教育的結果，以爲改進的根據的。

度量物質的工具，已經有很多種了。例如權衡輕重有秤，度量長短有尺，計算時間有鐘錶。甚至於溫度的高低，電流的多寡，都有測量的工具。自然科學之所以能够日益進步，未始不由於此。我們要想改進教育方法，提高教育效率，那麼，對於度量教育結果的工具和方法，自不能不加以研究。

(二)測驗的可能 不過屬於物質方面的東西，顯而易見，度量時並不困難；而教育的結果，隱而難明，度量時卻很不容易。因此，有人對於心理測驗的可能性，發生懷疑。他們以爲智力、品格、性情、興趣、意志等抽象的對象，是否可以用科學的方法去測量？書法的優劣，文章的豪邁，學力的高低等，是否可以用數字去計算？這些都是可懷疑的問題。其實，心理測量，也和物質測量一樣是可能的。我們祇要明瞭測驗的兩個

理論的基礎，我們的懷疑就可以渙然冰釋了。現在把心理測驗在理論上的兩個基本原則畧述如下，以闡明心理測驗的可能性：

(1) 任何現象，祇要是存在的，總有數量。這個原則是美國心理學家桑戴克 (E.L. Thorndike) 於1918年提出的。他說「凡物苟能存在，必有其數量」。(Whatever exists at all, exists in some amount.) 例如智慧有高低，品質有優劣，文章有工拙，學力有高下，在高低優劣之間，必有程度的不同。所謂程度的不同，就是數量上有差異。

(2) 凡有數量的現象，都可以測量。這個原則是美國測驗專家麥柯爾 (W. A. McCall) 於1922年提出的。他說：「凡有數量的東西，都可以測量」。(Anything that exists in amount, Can be measured.) 這一條原則，在理論上是毫無問題的，但在實際上，目前尚難完全做到。教育的結果——知識、技能、習慣、品格、理想、態度、欣賞、興趣等——，我們現在還不能一一加以測量，這是不必否認的。但這是由於測驗的歷史很短，測驗的工具尚未完善的緣故，我們決不能因為測量某種教育結果的工具還未發明，而說某種教育結果是無法測量的。

(三) 測驗的方法 就上面兩條原則看來，便可以知道心裏測量是和物質測量一樣可能的。不過心裏測量的方法，和物質測量的方法不同。物質方面的東西，是固定的，具體的，可以直接去測量，例如用尺量布，用斗量米，布和米都是固定的具體的東西，所以能够直接的度量；而教育的結果——知識、技能、品格……等——是抽象的，時時變動的，所以不容易直接去測量，祇可以間接去測量。例如「智力」是一種「無聲無臭」的東西。不能直接去測量。祇可以從一個人的作業上去估定。所以智力測驗所測量的，雖名為先天的能力，實係後天的作業。由後天作業的成績去推測一人的先天能力，便是間接測量的方法。

總之，測驗是教育上的一種新工具，可以用來度量教育的結果，以

謀教育方法的改進。但測驗的歷史很短，測驗的方法，尚未達到完善的程度，然而這祇是工具完善不完善的問題，並非測驗的可能性有什麼可懷疑之處。測驗在教育上可能的貢獻很大，我們還要繼續努力研究，以求達到盡善盡美的地步。

第二節 測驗應具的條件

據測驗專家的意見，一個完善的測驗，須具備下列幾個條件。

(一)測驗的正確性須高 一個測驗，須能確實的測量到它所要測量的能力。例如一個智力測驗，應該純粹測量智力；一個算術測驗應該純粹測量算術能力。倘若一個測驗，真能確實的測量到它所欲測量的能力，可謂之「正確性」高，或「效度」高(Validity)。正確性或效度是一個測驗最重要的條件。要編造一個正確的測驗，對於測驗的材料，須經過審慎的選擇。編造算術測驗時，所用的材料，必須包括算術上各種重要的計算能力；編造歷史測驗時，所出的問題，必須是歷史中重要的部分。而且在測驗中，應當把與所測量能力不相關的分子完全除去。

(二)測驗的可靠性須高 所謂測驗的「可靠性」(Reliability)，又稱測驗的「信度」。這就是說，一種測驗，測量同一個兒童兩次或數次，若所得的結果相同，則這種測驗的可靠性很大，或信度很高。測驗的可靠性是很重要的，不可靠的測驗，就沒有多大的用處。假如用尺去量一塊布，第一次量得一丈，第二次量之為八尺，這把尺還有什麼用處？心裏測驗也是這樣的。假如用一種智力測驗去測量某兒童，今日測量的結果為智力年齡八歲，明天測量的結果為智力年齡九歲，這個智力測驗就沒有什麼功用可言了。

(三)測驗材料的難度的分配須寬廣，俾能鑑別學生的程度。例如有兩個兒童，他們的算術程度相差很微，普通的觀察不能鑑別二人算術能

力的高下，我們希望用算術測驗去測量這兩個兒童，就能夠把他們二人程度的高下分別出來。要想使一個測驗能夠精密地辨別各種高下程度的能力，則測驗材料的難度的分配須寬廣，例如一種測驗有五十個題目，若難度的分配，祇有五個等級，（即每十題代表一個難度），則所能測量的能力的程度也只有五個等級；若其難度的分配為五十個等級，（即每題代表一個難度），則所能測量的程度就有五十個等級，所以同是五十個題目，因為難度分配的不同，其鑑別學生程度的效用，也就顯然有異了。

（四）測驗的答案須簡明確定——論文式的題目，學生可以任意發表意見，各人的答案頗不一致，教師定分時，非常困難。所以測驗的答案，要簡單明瞭，祇須學生寫一個字，或一種符號，或一個數目字就夠了；並且答案的對與不對是絕對的，不應當有介乎二者之間的答案。例如艾孫編小學地理測驗，每個問題後面有四個答案，其中祇有一個是對的，其餘三個都是錯的，要被試者把對的答案指出來。這樣的答案，是很簡明確定的。

小學地理測驗（摘錄）

17. 長江發源於 (1)西藏 (2)青海 (3)新疆 (4)四川
18. 我國頂大的鹹水湖是 (1)青海 (2)巢湖 (3)洞庭湖 (4)太湖
19. 福州東南的馬尾設有 (1)兵工廠 (2)造紙廠 (3)鑄幣廠 (4)船政局
20. 我國第一大島是 (1)崇明島 (2)劉公島 (3)海南島 (4)特里屯島
21. 杭州的龍井，以產 (1)茶 (2)米 (3)麥 (4)棉……著名
22. 我國南海的國防要地是 (1)西沙羣島 (2)廣州灣 (3)廈門 (4)九龍半島

（五）施行測驗的手續和記分的方法必須規定——要求測驗的結果精確可靠，則施行測驗的手續和計算成績的方法，必須劃一。施行測驗的手續上第一個要件，就是測驗的說明須加以規定，並且說明要簡單明

瞭，使學生不致誤會。主試須依照說明去做，不得插入不相干的話。第二，測驗的時間，須加以規定，主試不得更動。計分的方法，也要規定，並須備有「標準答案」，使閱卷者有所根據，不致參入主觀的見解。

(六)測驗須有量表，以資考核。量表(Scale)是某種測驗的分數的排列，從最低的分數排到最高的分數，用年齡、年級等做單位，來表示某種能力的高下。所以量表與測驗不同，測驗是考查兒童成績的材料，而量表是評定兒童成績的工具。我們用某種測驗去測量兒童，得到一個分數，還要對照量表，視該分數在量表上的地位如何，然後始能評定他的能力的高下。例如我們用黃覺民氏訂正幼童智力圖形測驗去測量某兒童，他的測驗分數是38分；再去查量表——「測驗分數與智力年齡對照表」——，就知道得38分的，其智力年齡為96個月，換句話說，這個兒童的智力年齡是8歲，假設這個兒童的實足年齡是七歲，那麼這個兒童是很聰明的。所以一種測驗須有量表，然後評定成績的優劣纔有所根據。

一個良好的量表，須具備三種條件：(1)量表必有一個單位，如「寸」為尺的單位，「兩」為斤的單位；而且量表上每個單位之間的距離還要相等，如第一單位與第二單位之間的距離，須等於第二單位與第三單位之間的距離。(2)量表必有一個參照點，所謂參照點，就是計算的起點。例如攝氏寒暑表以零度為參照點，華氏寒暑表以冰點以下32度為參照點。(3)量表自最低一端至最高一端，應該包括全部兒童的程度。例如量表上最低的分數若為10，最高的分數為130，則最劣的兒童亦能超過10，最優的兒童，亦難達到130，然後這個量表，纔算合用。

(七)測驗須有常模，以便比較。各種測驗應當求得「常模」(Norm)，以為比較學生程度的標準。教育測驗上常用的常模，約有兩種：就是「年齡常模」和「年級常模」。「年齡常模」就是各年齡兒童在某種測驗上所得的平均分數。「年級常模」就是各年級兒童在某種測驗上所得的平

均分數。常模的功用，在解釋測驗的結果，一個兒童或一班兒童的分數，可與常模來比較，以評定其優劣。例如艾郭合編小學算術測驗曾測驗小學二年級上學期學生776人，求得其平均分數為36.60，這個平均分數就是該年級的常模。以後我們若用這一種算術測驗去測量某校二年級上學期的兒童，就可以把所得的分數與這個常模(36.60)相比較，假如兒童所得的分數比常模高，這就是表示這班兒童的算術程度，比較普通二上級兒童的算術程度好；反之，假如兒童的分數比常模低，這就是表示這班兒童的算術程度不及普通二上級兒童的算術程度。「常模」與「標準」的意義不同，「常模」是某年級學生在事實上已經達到的程度；而「標準」是某年級學生在理想上應當達到的程度。所以一個測驗的常模，經過了長久的時期，也許不適用了，它並不是可以永久不變的。

(八)測驗應當有覆份，以便應用。我們要考查兒童在一學期內的學業成績的進步情形，最好是在學期開始時舉行一次測驗，再在學期結束時舉行一次測驗，比較兩次測驗的結果，就可以看出學業成績的進步情形，但是測驗材料若只有一份，在一學期中應用兩次，難免有練習的影響，第二次的結果，就不可靠了。倘若一種測驗有兩份以上的相等的材料，則學期開始時用甲種，學期結束時用乙種，交替應用，就可以免除練習的影響。

我們說兩份交替測驗須相等，所謂「相等」，係指下列幾個條件而言：

- (1)各份測驗，須測量同一種能力。
- (2)各份測驗中所用的材料，不應當有重複的地方。
- (3)各份測驗的難度須相等。
- (4)各份測驗的可靠性須大。

第三節 測驗的功用和限制

(一)測驗的功用 測驗有什麼用處呢？簡單說來，測驗是改進教育方法的一種工具。分析起來說，測驗可以辨別智愚，分編班級，估量成績，改進教法，診斷困難，鼓勵學生作業和指導學生擇業等。茲分別說明於後：

(1)辨別智愚 人類的智力各不相同，在一百個兒童中，不特可以分爲上智、中材、和下愚三等，而且在這三等之中，還可以分成無數的等級。人類既有智愚的不同，我們就應當施以不同的教育，使各人得以充分發展其能力。但是用什麼方法去辨別兒童的智愚呢？「測驗」可說是最簡便而正確的方法。我們若用智力測驗來辨別兒童的智愚，然後就可以因材施教了。

(2)分編班級 分編班級的目的，在將學力相等的兒童編爲一班，然後再將一班中智力不等的兒童分成幾組，這樣分編而成的班級，不但教學的效率可以提高，兒童學習工作也不致互相牽掣。所以在分級時，可採用智力測驗和教育測驗作爲分編班級的工具。教育測驗的目的在求得學生現在的教育程度，以爲分級的根據；智力測驗的目的，在分別學生的智愚，以爲分組的根據。然後一級中的兒童的程度就較爲整齊，一組中的兒童的進步速率也可以一致了。

(3)估量成績 做教師的亟欲知道他的教學成績好不好，做學生的也亟欲知道他的學業有沒有進步。但是普通論文考試的方法太不可靠，不能正確地估量學生的成績，施達區(Starch)曾做了一個實驗，證明普通考試方法的不可靠。他選了一本英文試卷，請 142 位英文教員去批評，所批分數是不一致的，從 50 分起到 97 分止，各不相同，其中有 26 位教員批不及格的分數，而有 18 位教員所批分數在 90 分以上。由這個報