

華超編

現代師範
教科書

教育測驗網要

商務印書館出版

0005468

自序

教育測驗，在教育上的位置，已經教育家多數的承認。從改進社編行各種測驗之後，測驗的呼聲，更一天高似一天。但是，常聽得人說，「知其然，而不知其所以然。」這是於測驗的使用上，改良進步上，阻害不少。編者編本書的主旨，就在把教育測驗中，重要的假定，原理，編造測驗具體方法和如何利用等等，約略介紹於開始研究測驗的學生。

本書一部分，是十三年夏季，在蘇，滬兩教育局暑期講演會講演的結果。關於理論和應用方面，是根據伊里諾大學教育教授兼教育研究所主任孟羅 (Walter Scott Monroe) 所著教育測驗之理論和教授結果之測驗二書。關於編造法，當然要取最新的，把哥倫比亞師範院助教授麥柯爾 (William A. McCall) 所著的教

教育測驗法和實驗教育一書中所談的做根據（各書原名詳各章參考書中）

本書主旨雖在引導初學的人，但亦希望校長和其他主持施行教育測驗的教育行政當局，獲得相當的見解，熟悉測驗的基本原理和如何利用之道，使教育測驗盡充分的功效。

統計方法，本是編造或批評研究教育測驗中重要的部分。但是，因為時間關係，本書中並沒加上討論，只得待將來再說。讀者最好參考相當書籍（見附錄參考書報）。

編者對於本書所根據的書籍的原著者，深切的感謝。因為沒有這類作品，本書實無從告成。增刪或修改原著的地方，由編者負責。

中華民國十三年八月二十五日

華
超識於上海

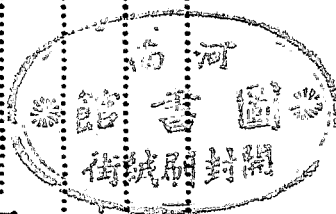
目次

第一章 教育測驗的沿革

- (1) 導言.....一
- (2) 測驗的類別.....一
- (3) 標準測驗的起原.....三
- (4) 其他事件.....三
- (5) 研究機關.....四

第二章 教育測驗的性質

- (1) 敘述的種類.....一九
- (2) 用於分量敘述的量表.....一九



(3) 標準單位.....二二

(4) 從作業測驗心理能力.....二二

(5) 作業的三方面.....二三

(6) 教育測驗中的假定.....二四

第三章 教育測驗的結構

(1) 普通結構.....三一

(2) 客觀記錄.....三三

(3) 難度解釋.....三四

(4) 測驗類範.....三五

(5) 勢力測驗和速率測驗.....二七

第四章 教育測驗的功用

(1) 教育上的測度	四一
(2) 行政方面的活動之須要測驗者	四二
(3) 教授方面的活動之須要測驗者	四七
(4) 研究方面的活動之須要測驗者	五三
第五章 教育測驗的編造法	

(1) 應注意的要點	五七
(2) 難度測驗的編造法	五八
(3) 速率測驗的編造法	六八
(4) 品質測驗的編造法	六九
(5) 彙選測驗的編造法	七七

第六章 最近 T, B, C 量表的編造法

(1) T 量表..... 八三

(2) B 量表..... 八七

(3) C 量表..... 九九

第七章 教育測驗的實施法

(1) 施行前的準備

(a) 標準測驗的選擇..... 一〇五

(b) 主試者的訓練..... 一〇七

(c) 測驗委員會的組織..... 一〇八

(2) 施行後的利用

(a) 分數的解釋..... 一〇九

(b) 錯誤的範類..... 一一〇

(c) 圖表的利用·····	一一一
(d) 補救的手續·····	一一二
(e) 就年級標準診斷·····	一一三
(f) 就成業率診斷·····	一一三
(g) 普通診斷和精密診斷·····	一一五
(h) 分析的診斷·····	一一六
(i) 補救教授的計畫·····	一一七
(j) 就學習手續診斷·····	一一八
(k) 設定直接的教育目標·····	一一九
附錄中西參考書報·····	一二三

現代師範
教科書

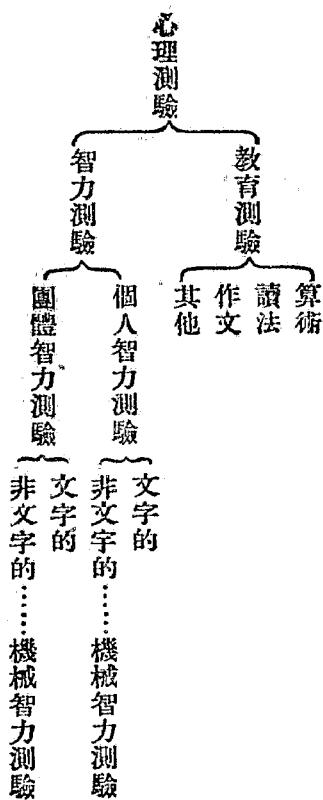
教育測驗綱要

第一章 教育測驗的沿革

導言 研究科學，宜略知某科學的背景。例如，研究物理學，宜略讀物理學史，研究數學，宜略讀數學史。教育測驗，在現今雖不能說是完全的科學，比不上物理學，數學等的那樣精密，但是要研究教育測驗，教育測驗中，過去的事實——怎樣發生，怎樣改良等等——亦須得知道一個大概情形。所以本書先講教育測驗的沿革，再進而論他的性質，結構，功用，編造等等。

測驗的類別 學校中所謂測驗，除體格檢查外，可通稱爲心理測驗。倘要分開來講，可分做教育測驗和智力測驗。教育測驗，不用說，就是算術，讀法，作文等等測驗。智力測驗又可分做個人的和團體的兩種。個人智力測驗，是說一個主試者

於一時間內，只能測驗一個被試者；團體智力測驗，是說一個主試者於一時間內，可以測驗一羣的被試者。個人智力測驗裏頭，又可分做文字的和非文字的兩種。前者是用文字，有許多不便的地方；後者是補救前者的缺點，改用圖畫等等。這一類（非文字的智力測驗）裏頭，有專測被試者於使用機械方面的能力——對於機械各部分之觀念是否準確等等——的，叫做機械智力測驗。團體的智力測驗，亦可以分做文字和非文字的兩種。試看下表，自能格外明瞭：



爲何教育測驗和智力測驗，可通稱做心理測驗？因爲教育測驗所測學生學業的成績，智力測驗所測學生天賦或遺傳的能力，二者都是屬於心理方面的。並且教育測驗和智力測驗也不能有絕對的區別。教育測驗中的問題，非有相當的智力不能答，智力測驗中的問題，大部分要有學校訓練，家庭，社會經驗等等纔能應答。所以有人主張測驗，只分做個人和團體兩種，不主張有教育和智力的分別。

標準測驗的起原 從有學校以來，就有學生成績的考查（成績考查何以必要，何以廢不掉，此地不多說）這就是測驗的一種形式。但是，用標準測驗以考查學生成績，乃是新近的發展。這種新工具——標準測驗——和普通考試的異點，簡單的說來，有二種。一，客觀的，就是他所測得數量，不因人而異。申說一句，不同的主試者，於同一情形下，測驗同一學生的能力，所得的結果，可以約略（approximately）相同。二，標準的，就是他有一定的標準，足以解釋所測得的結果。例如，測驗

某級學生算術能力，測驗的分數是四十。這四十可以和某級應達的標準分數相比較。這樣，就可以斷定某級學生算術能力，在標準以上，或在標準以下，或剛達標準。

用標準的，客觀的測驗以考查學生成績，當最初發展的時期中，貢獻最大的，是心理學家和學校行政當局。心理學家，當研究人心 (human mind) 的機能時，往往在學校科目中，規劃各種測驗。這種測驗，雖亦非常簡陋，但通常具有客觀性質。他們並不訂立標準，因為標準於心理研究上，沒有甚麼效用。學校行政當局則羣趨於課程，學校組織，以及教授方法等等問題的研究，但又缺乏科學根據，沒有科學方法，所以大家的意見不能一致。最近，學校行政當局，要根據科學方法以解決這些種問題。於是發生客觀的測驗的編造。

標準的，客觀的測驗的發展，既然是新近的事情，而貢獻者，為數又極多，勢難

一個一個的盡說出來。以下所說的，不過是發展中最可注意的人和事，又因爲在美國標準測驗最發達，所以又偏重於這一方面。

菲曉 (Reverend George Fisher) 西紀一八六四年，英國菲曉氏，有見於教員給分數沒有一定的標準，於是搜集學生成績樣本，收來之後，分別優劣，彙成一書，叫做尺度書 (scale book)。凡評判某學生的成績，就把他的成績和尺度書上的樣本，兩相比較而評定優劣程度。這書雖很簡陋，實開近來教育測驗中量表的端緒。但是，因爲當時適當的統計方法，還沒有發達，這種研究於英國，就無從繼續研究下去。到後來，葛爾敦氏 (Sir Francis Galton) 著有遺傳與天才一書，出來之後，統計方法逐漸完備，於是這種新工具的研究，再得上軌道。

拉愛斯 (J. M. Rice) 就美國而論，標準測驗的發展中，首先應當稱述的，是拉愛斯博士。氏第一次的探究（一八九四年至九五年）是學生拼法或默字能力

的測驗。這測驗合生字五十而成。彙五十字爲一表，寄給各地視學員，分途實行測驗。後來，視學員裏頭有二十個人，把一萬六千個學生的試卷，寄回來評閱。氏就更進一步，另編一拼法測驗，把各個生字擺在句子中間，親自測驗大批學生。據他測驗的結果，證明八年中每天學習拼法四十分鐘的學生，拼法能力並不優於每天只學習拼法十分鐘的學生。氏的研究報告，發表之後，頗遭當時教育家的反對，以爲學生拼法能力，怎樣能把少數規定的字，測驗得出。但是，事實上，這是可能的。因爲學校裏普通的考試，也不過從教材中選取若干重要的事項，列爲少數問題來考驗全部的成績。再後來，氏更進而研究算術和語言能力的測驗。

桑代客 (E. L. Thorndike) 在拉愛斯之後，而於標準測驗的發展中，貢獻最大的是桑代客教授。當拉愛斯第一次研究拼法測驗的時候，桑代客正專攻統計學的方法。氏又喜歡心理的研究，結果，於一九〇三年，發表各種算術能力之關係

(Relation between the Different Abilities in the Study of Arithmetic) 一文。氏用來研究這關係的工具，就是算術測驗，但是，桑氏研究的動機是心理學的，而拉氏的研究，在解決學校行政問題——時間分配，教授法等等。

一九〇四年，氏著心理及社會測驗 (Mental and Social Measurement) 一書。這書除敘述統計手續外，現今編造測驗所根據的重要原理，亦在討論之列。一九一三年，這書又重加修正，出版。所增加的，雖不過是解釋原理的具體問題，但是，這書在現今仍舊是研究教育測驗的，所必讀的書。一九〇九年，氏曾把所著書法量表 (Thorndike's Handwriting Scale) 提出於美國科學改良會。這量表的編造法，見一九一〇年三月的師範院季刊 (Teachers College Record) 中。從這量表發行以後，各家書法測驗和英語作文測驗等編造，都應用這方法量表的編造原理，或稍稍加上一點修改。

斯東(C. W. Stone) 氏是桑代客的學生。當拉愛斯研究拼法測驗時，斯東也於算術範圍中，提出兩個問題，做研究的對象。

- (1) 小學前六年的算術教學，結果的性質是怎樣。
- (2) 各種算術工作的過程和所產生的能力，關係怎樣？

氏研究上述兩種問題時，就編造算術四則測驗和應用題測驗各一種。這兩種測驗，規定怎樣試行的說明和怎樣記分的方法。主試的人雖不同，倘依照所規定的進行，測驗同樣學生，結果可以相仿。所以，可以說他有客觀性質。從初步研究告成之後，應用題測驗，就訂定標準。凡學務調查和教員考查學生成績等等，都用這測驗。

克的斯(S. A. Courtis) 斯東氏的算術測驗，限於小學六級，而克的斯氏的研究算術測驗，涉及小學各級和中學校。克氏的規劃測驗，用意不在於心理學，也

不在於學校行政，主要的目的，在扶助教員教授學生。他所著算術測驗甲組，於一九〇九年告成。合八種測驗而成。前四種是加減乘除，第五種是膽寫數目，第七種是包括四則，餘二種是應用題測驗。全組測驗，備有說明書，且規定時限，定有標準。教員要考查學生算術能力進步不進步，就可應用這種測驗，再和所定標準比較，診斷學生的需要，在什麼地方。一九一三年到一四年間，氏又另編算術測驗乙組。這組測驗有四，就是整數的四則。通常，所謂「克的斯標準算術研究測驗」(Coulter's Standard Research Tests in Arithmetic) 就指這組測驗而說。

哀夏 (L. P. Ayres) 氏於教育測驗發展中的貢獻，是他的書法量表 (一九一二年) 和拼法量表 (一九一五年)。他的書法量表，和桑代客的書法量表不同。這量表定學生書法的等級，用「清楚」(legibility) 做準繩，而桑氏的量表，用「美觀」「齊整」等做準繩。定「清楚」的程度，法將許多學生書法樣本，使許多人誦

讀，而計算每一樣本所須平均時間。易讀而平均時間少者，「清楚」的程度高；不易讀而平均時間多者，「清楚」的程度就低。這種編造法，現今編造書法量表的，都不用。因據研究的結果，這種量表並不十分優於桑代客的量表，且編造的方法亦不如桑氏法的簡而易行。至於哀夏的拼法量表，在美國卻頗通行。

勃金漢 (B. R. Buckingham) 氏也是桑代客的學生，於一九一三年，發表所著拼法量表的編造法。這量表現今不很通行，但是，於教育史上占一重要位置。因為，這量表代表一新式工具。量表中所選各字的難易，看各級學生拼對的百分數而定。然後再依照難易，列爲一表。最容易的（學生中幾無一不能拼對者）列在前，較難的次之。逐漸增加難度到頂難的（極少數學生所能拼對的字）擺在量表的另一端。凡學生拼法能力的高下，就看他在量表上所達的難度而定。

這種編造測驗的原理，已經爲大多數編造測驗者所仿用，而尤以華狄

(Woody) 的算術測驗，霍志 (Hotz) 的代數測驗，很蒙 (Hemmon) 和勃郎 (Brown) 的拉丁語測驗，屈拉布 (Trabue) 的語言測驗，範威極南 (Van Wagenen) 的歷史測驗是頂顯著。

比納 (Binet) 以上所說的，都限於學生成績的測驗和這種測驗所用的工具。學生校課成績和他的天賦智力，有密切的關係，所以普通智力的測驗，亦宜略說一說。講到普通智力的測驗，法國心理學家比納是最著名。他在一九〇四年，受法國公衆教育部的聘，做低能兒班委員會委員之一。要組織低能兒班，就發生一個問題——兒童的低能，非低能，究竟怎樣斷定？為解決這問題，比納就在一九〇五年，協同醫生西蒙 (Simon)，編造一組測驗，就是現今所謂比納西蒙智力測驗。這測驗的特點，在於用年齡做標準，用心理年齡（亦叫做智齡）解釋學生作業的優劣。到了一九〇八年，這測驗又經比納修正。

比納測驗出世後，美國心理學家很注目。訂正或改良比納測驗的人逐漸多起來，而尤以哥德（Goddard）和推孟（Terman）二氏是最著名。推孟的訂正本，在美國更風行，現在已由編者譯成中文，歸商務印書館出版。

團體智力測驗（一九一八年）用比納測驗法測驗學生，必須把學生一個一個的測驗，費的時間多，效用因而狹隘。於是就發生團體智力測驗——於同一時間內，能測一羣學生——的編造。當美國加入歐戰（一九一七年）之前，美國心理學家已做種種試驗，要想把測驗同時可以施行於一大羣的人。等到美國既加入歐戰之後，軍事委員會決定用心理測驗來甄別招募來的兵士。這時候，推孟的學生渥的斯（A. S. Otis）已有一種團體智力測驗告成。軍事委員會就採用這個測驗，結果極好。以後，團體智力測驗的數目，一天多一天，但大多數和烏的斯氏原著經軍事委員會採用的，形式相仿。

其他事件 教育測驗，既然是新近的事情對於這裏頭的個別事實，也極難一一估量他的價值。而最可注意的，是再生分數（或第一次分數 derived score）和從成績分數與智力分數比較而得的成業率（achievement quotients）的導入。再生分數，是不用學生實得的分數作為分數，而把統計中所規定的分數作為分數。例如，改進社所編各種默讀測驗和智力測驗等，既得學生答對的題數後，再要查照各測驗中所規定的T分數。這T分數（亦叫做總成績，亦叫做總能數）就是再生分數。至於刺激或促進教育測驗發展的，有教員分數的研究和學制效率的調查。從前，以為學生成績的優劣，得由教員所給的分數決定。經過多數人研究——其中以凱萊（E. J. Kelley）的研究尤為著名，氏著有教師分數（Teachers' Marks）一書——證明教員給分數不準確。教員分數既不準確，就有創造標準測驗，使考查學生成績比較滿意並準確的需要。用教育測驗，考查學制效率，在美國

始行於一九一一到一二年紐約公立學校的調查。在這調查中，克的的算術測驗甲組，曾應用於三萬左右的學生。從這次調查以後，學校調查，差不多沒有不用教育測驗的。

研究機關 教育測驗的發展，因研究機關的設立而進步格外順利。這種研究機關，大別有二。一由學區設立，主要的目的，在指導並促進測驗的應用。這種研究機關，或附設於視學員 (superintendent) 的辦公機關中，或另設一局所，而受視學員的支配。還有一種，由州立大學或州教育部設立。如一九一三年烏克拉克麥大學 (University of Oklahoma) 所設立的測驗和效率研究所 (Bureau of Measurement and Efficiency)，一九一四年，印第阿那大學 所設立的合作研究所 (Bureau of Cooperative Research)，剛薩斯州立師範學校附設的教育測驗和標準研究所，納勃拉斯卡大學附設的教育測驗研究所。而費斯康新州教育部，

於一九一五年亦有類似的研究所設立。

這種研究機關，於教育測驗的推行上，很有影響。因為有了這種研究機關，視學員和教員等可以熟悉教育測驗，並於測驗之施行上，受相當的訓練。主任這種研究機關的人，又常編造，發行新測驗，供給各地方應用。美國教育測驗，能發展得這樣快，不得不歸功於這類研究機關。

吾國從一九二二年，中華教育改進社聘哥倫比亞大學師範院教授麥柯爾（W. A. McCull）來主任編造各種測驗後，現在已完成的標準測驗，不下十數種。舉其主要的如下：

陳氏小學默讀測驗，中學默讀測驗，初小默讀測驗，俞氏小學算術應用題測驗，小學混合四則測驗，陸氏訂正比納西蒙智力測驗，廖氏團體智力測驗，查氏教育測驗，俞氏增訂算術練習測驗，德氏非文字智力測驗，機械智力測驗，安氏英文

測驗等。而學校中應用這種測驗的，爲數不很多。主要原因，大抵（1）由於經濟困難，（2）不明測驗的真相，持懷疑的態度，（3）缺乏促進或宣傳的研究機關。

參考書：

Monroe: Theory of Educational Measurement.

G. M. Whipple: Classes for Gifted Children, Chap. II, pp. 14-15.

Seventeenth Yearbook of the National Society for the Study of Education (美國

教育研究會年報)

陳鶴琴——智力測驗法第三章。

問題和討論：

- (1) 教育測驗和普通考試的異點何在？
- (2) 教育測驗在美國，何以特別發達？
- (3) 教師所給分數，在那幾方面是不準確？

- (4) 團體智力測驗和個人智力測驗，那種可靠，並述其理由。
- (5) 教育測驗能不能完全替代普通入學試驗？
- (6) 現在教育測驗的反對論調有幾種？用何法得反證其不能成立？

第二章 教育測驗的性質

敘述的種類 我們對於物理的事物 (physical objects)，應用兩種敘述。或由東西的顏色，組織，形式，優美等等敘述某種東西；或由東西的大小，輕重，高低等敘述某種東西。前一種，叫做性質的敘述；後一種，叫做分量的敘述。所謂測驗，簡單的說來，就由分量上（或用分量的語詞）敘述東西罷了。無論什麼時候，一件東西的特性的總量，已經決定之後，就可以說這件東西的特性，已被我們測驗了。

用於分量敘述的量表 所有分量的敘述，必須有分量的量表或尺度 (scale of quantity)。量表中，最普通的，定有出發點。從這出發點，排列若干單位。這出發點亦叫做零點。（通常指完全不含有單位中的特性而說）應用這量表，以測驗東西時，就看這東西距離零點有多遠。例如，敘述一樹之高度是三十尺，意思就是說這

樹從地上的零點起，在直線量表上，占三十個單位（尺）。

分量的敘述，須有分量的量表；而量表上所用的量（magnitude），又必須和所要測的量同一種類。例如，我們不能把方尺做單位的量表，來測驗一東西的重量。更進一步，所測得的量，必須前後一致。這兩個條件，於物理的測驗上是必要，容不着多說。在此地慎重的提出來，因為心理測驗也必須滿足這兩個條件。

除上述之外，用於分量敘述的量表，還有一種。就是選擇同範類（type）的東西，依照量的增進，逐一排列而成的量表。量表上的東西，用數目標明，從量的小的起，到量的最大的止。使用這種量表以測驗東西，就看這東西和量表上那一個東西最相近。量表上東西的位置，就準確指出某東西的位置。例如，行書測驗，就把學生的成績和行書量表上的字，逐一對比，看他 and 量表上那一度的字最相近。倫和量表上四十五度的字最相近，這學生的行書成績，就是四十五分。倫和五

十度相近似，就算五十分。

標準單位 敘述物理的事物，所用單位的量（或大小），常以習慣而確定，所以可叫他為標準單位。例如，一公尺，一公斤等，常常有確定的意義。但是在心理測驗中，標準單位還沒有確定。有些人用一級的變化率（variability）做單位；有些人用各級混合的變化率做單位，有些人用判斷的變化率（variability of judgment）做單位；有些人，用成人作業的變化率（variability of adult performance）；還有些人用比較的價值（relative worth）做單位。而現今改進社所編各種測驗，遵照麥柯爾的主張，用十二歲學生的標準差（S. D.）的十分之一做單位（這指編高小用的測驗說，倘編初小用的測驗，則以八歲學生之標準差之十分之一為單位，編中學用的，以十六歲學生之標準差之十分之一為單位。）編造測驗，單位之確定，雖不是重要的事，各人可以用自己選擇的單位，測得的結果仍舊可準確。

不過爲易於比較測驗結果和使人容易了解起見，必須設定公認的普遍使用的標準單位。

從作業以測驗心理能力 學生心理能力的高下，只能從心理能力所表現的作業 (performance) 估定。所謂能力指動作的能力 (ability to do)。倘能力而沒有一種機能，申言之，倘不能產生一種動作，則這個能力就不是能力。根據這定義，能力的測驗，實在就是用分量的語詞，敘述能力的機能，於規定情形下，產生的作業。

或有主張心理能力，不是我們所能測驗的，我們所能測驗的只是作業。就實用上說來，這種區別，無足重輕。照前所說，心理能力的意義，從他（心理能力）所產生的作業得來（能力必有動作）；當這種能力於規定情形下（即使學生工作於同樣情形或條件之下），表現他的機能時，能力的測驗和作業的測驗，實異名而

同意。可以叫他是測驗能力，也可以叫他是測驗作業。因各種作業必有相當的能
力程度存在。

作業的三方面 作業有速率 (rate)，確度 (accuracy) 和難度 (difficulty) 的三方面。這三方面包括於下列三個問題中：『學生工作，每分鐘是多少？』對於規定的工作，做對的有多少，做錯的有多少？』他所表現的工作能達到什麼難度？』要完全敘述作業的三方面，必須對於這三問題，一一解答。至於能力表現的作業和能力自身的關係，要看作業當時的情形而定。例如，作業時，學生的努力不努力，身體狀態，情緒狀態，注意集中不集中，主試者的態度，測驗的方式，測驗內容的說明，學生是不是熟練測驗的手續，事前曾受過人家的指使或特別預備否，時間的長短等等，都大有影響於作業的結果。所以這些種情形的敘述或規定，也是測驗中的主要部分。這作業的三方面，不是同樣重要，所以某種測驗，以速率為主；

某種測驗，以速率和準確爲主；某種測驗，兼顧速率，準確，難度三種。

教育測驗中的假定 簡單的說，心理能力的測驗，不外乎由分量上，解釋心理能力於規定條件下，所產生的作業。細分起來，心理能力測驗，有兩個步驟。第一，就是得到學生的作業（把測驗應用於學生）。第二，解釋作業的結果，就是應用測驗中的量表。心理能力，產生作業，故假定於同樣情形下，這心理能力，能產生同樣的作業。現在把心理能力測驗中，幾個重要的假定，略說一說。

(I) 假定作業和所測能力間，有不變的機能關係 (constant functional relation) 這假定的意義，是說能力的變化，產生相當作業的變化（能力高者，作業之結果優；反之，能力低者，作業之結果劣）。換句話說，作業程度的不同，由於能力程度的不同。再，這個假定包含於解釋作業結果的普通常模 (common norm) 中。例如，假定初小三年級算術測驗的常模是三十，這常模既經定奪之後，凡初小

三年級學生算術成績的解釋，就看他比較常模的高下而定，至於主試者和測驗地方等等的不同，不再過問。倘沒有這個假定，則這個人所測得的或者要異於那個人所測得的，在這地方所測得的，或者要和在別地方所測得的全不一樣。結果，就無從比較。

(2) 假定作業是某種心理能力表現機能時所產生的結果，而不是別種心理能力的結果。這假定包含於上一假定中。照上面所說，倘作業和某種能力間，有恆久不變的機能關係，則這作業必不為別種能力所影響，可以無疑（別種能力的影響，和所測能力產生的變化，同時變化者，除外）。普通說來，各種不同的能力，都是互相獨立，所以他的影響不十分相關（correlate）。倘這話是確實的，那麼所得到的作業，一定是某種心理能力表現機能時的結果。就是有別種能力表現影響，影響也一定是很微弱的（不變的機能關係大受阻害時，除外）。

(3) 假定能得到某種能力的機能表現。這假定包含於標準的、客觀的測驗之使用中。因為我們應用一種測驗於大多數的學生，我們就假定得到各個學生的同一能力的機能，且更假定各個學生有適當表現自己能力的機會（事實上無適當表現能力之機會者，除外）。申言之，標準測驗的應用，習慣上含有一種意義，就是無論什麼時候，我們要得到某種能力的機能，這能力的機能就可以得到，所不可少的，不過是把測驗試用於學生罷了。

(4) 假定測驗情形 (testing conditions) 不但於某時間內可以控制 (controlled)，就當異時，異人的時候，也可以控制。這假定已完全包含於第一假定中。因為這假定非常重要，故另立一條。照上文所說，測驗情形，含有測驗說明，時限問題，學生態度和努力，測驗室光線，甚至當時的溫度和氣壓等等動因 (factor) 都有影響於學生的作業。倘使我們不能控制這些種動因，使各學生於異時，異人

(指主試者說)的時候，都受同一的影響(同一情形下工作)則作業和能力間不變的機能關係，將因阻害而分離。

(5) 於智力測驗中，假定被測驗的學生，有得到測驗中規定能力的同等機會。普通智力測驗，雖在測驗學生的天賦(或遺傳)能力。但是，這由間接測驗後天(或獲得的)能力而定。因為智力測驗中，大多數問題，都須要學生的學校訓練和普通經驗。測驗中某種特殊問題所須要的能力，各學生容或沒有得到這能力的同等機會，但就測驗問題全體而論，凡生活於同樣普通環境中的學生，都得假定他們有同等機會。

上述假定近似真實。上所說的幾個假定，只能說是逼近於真實(approximately true)。所測能力和作業間，完全不變的機能關係，是永遠不能實現。有時候，別種能力已減少到最低限度，而這能力仍有影響，或阻害某能力和作業間的機

能關係。例如，某種算術測驗中，學生書寫的能力是有影響於他的算術能力的。有時候，因為不能完全控制測驗情形，這機能關係也受阻害，而當異時和異人（主試者）的時候，是最顯著。

所有假定，不能完全實現，故所測得的結果和結果的解釋都不能完全沒有錯誤。幸而，近來標準測驗，一天一天的改良，統計方法也一天一天的精密，各種假定，或者可以逐漸實現。

參考書：

- McCall's How to Measure in Education 第九章，二九一—三〇〇面。
Monroe's Theory of Educational Measurement 第二章。

問題和討論：

- (1) 何謂教育測驗？

- (2) 能力測驗含有幾種假定？
- (3) 主張心理能力不能測驗者，根據何種理由？
- (4) 何謂標準測驗？
- (5) 何謂標準單位？
- (6) 由測驗上觀之，能力之重要特性是那幾種？
- (7) 說明書中規定事項，何以必須絕對遵守？有無通融餘地？試述其理由。

第三章 教育測驗的結構

普通結構 從實際上說來，所有教育測驗都由若干部分組合而成。這些部分 (divisions) 亦叫做問題或練習 (exercises)。倘使要精密一點，這些問題又可分做若干部分。所以測驗中的問題是測驗全體中的一個單位。各種測驗的類範 (type) 就由這些問題性質，問題排列，提出問題的方式等而定。有些問題，是難的；有些是容易的。有些問題，要所測驗的能力於逼近正常的情形下 (approximately normal condition) 表現出來；有些，要在人為的 (artificial) 情形下表現出來。有些問題，須要時間不過幾秒鐘，有些，要幾分鐘。有些問題，須要被測驗者書寫；有些，則只要打個記號。

就算術四則測驗而論，選擇可以表現某種能力的機能的問題是很簡易的。

並且這種問題是可以於普通情形下，使所測的能力表現機能。通常，編造測驗者，可以無限的選擇許多同類問題，來測驗同一能力（不過各問題所須的答案是不同，所以不能說絕對的測驗同一能力）。關於這一類的測驗，可參考孟羅氏算術診斷測驗（Monroe's Diagnostic Tests in Arithmetic）或克萊佛蘭算術測驗（Cleverland Tests in Arithmetic）。

就拼法測驗說，把所拼的字夾入句中使學生拼，又當注意力完全集中於句子的意義時，就可說拼法能力在正常情形下表示。但是，因為要許多學生拼同樣的字，事實上萬不能編造一套問題使學生於完全的正常情形下拼出來。倘使拼法測驗中的一組字，由主試者一個一個口裏頭說出來，叫學生拼，學生拼時要留心上下文，注意力主要的集中於某字之拼法上，則拼法能力更不能於正常情形下表示出來。所以，拼法測驗中的問題總帶有一些人爲的意思。

客觀的記錄 在某種學校科目中，能力的機能的正常表現，不能發生客觀的記錄。遇到這種事例時，必須用一種人爲的問題，來測度這種能力，使他有個客觀的記錄。例如，於正常情形下，閱讀過程之結果，是看不出來的（不能加以考察的）。有幾種結果，如身體態度，面容的表示，以及其他器官作用的表示，容或可以看得出，但是這些種行動，對於測驗的目的上，沒有什麼大功用。所以，測驗閱讀能力，必須另行計畫一種問題，這種問題不但須要閱讀能力，並且須要可以看得出，容易加上記錄的別種能力。這一類的閱讀測驗，或者是使學生看了一段文字後，再對答這一段文字的問題，或者是使學生，將一段文字，自己簡單的複述出來。這兩種測驗都是使被試者不但表示默讀能力，並且表示默讀能力以外的各種能力，以便易於觀察。

至於問題的種類，很多。如選擇法，是非法，問答法，訂誤法等等。中華教育改進

社編的測驗，都用選擇法。各法具體的例子，和各自的功用，此地不多說。

難度之解釋 問題的難度，是問題的一種特性 (characteristic)，當問題很難時，不準確的答案的百分數就大，當問題不很難時，不準確的答案的百分數亦就小。換句話說，問題的難度，是看答對的答案的百分數而定。〔這就大較而說，此外亦視所定答對(或準確)的標準而異。〕倘使，我們對於被測驗者的能力的分配，預先立了一個假定，並且定一個能力的零點(或出發點)，則一個問題的難度就可以用能力分配的變化率 (variability) 的單位 (或能力之差異數量之單位) 來解釋。通常都用概誤數 (P. E.) 來解釋。用概誤率表示某個問題和別個問題的難度的不同時，是說某一個問題有百分之五十個人能答對，某個問題只有百分之二十五個人答對 (此指一個 P. E. 而說)。有時亦用標準差 (S. D.) 來解釋。用標準差時，則某一個問題有百分之五十人答對，某一個問題，只有百分之一

五·八七人能答對（此指一個 S. D. 說）所以，我們往往說某一個問題的難度是 2.7 P. E. 或 6.3 P. E. 或說 3.2 標準差。倘用麥氏教育測驗法第二十三表查，則每個問題的難度更容易決定。這表是用標準差解釋難度的。

測驗之類範 倘使一個測驗裏邊的各個問題，難易程度是同等的，又這些問題須要被測驗者發生的行動是勻遍的（uniform），這種測驗，就可叫他做勻遍的測驗（uniform test）。例如，克的斯氏的算術測驗，B 組（Courtis Standard Research Tests, Series B 卽正數四則測驗）

倘使一個測驗裏頭問題的難易是不一致的，從容易的逐漸到難的，則這個測驗，可以叫他做分度的測驗（scaled test）。通常，編造這種測驗，常選擇難易程度不一致的，就是從最易的到最難的各種問題，並且要各個問題的難度逐一增加——從百分之九十九人到百分之一人能答對。華狄氏算術測驗（Woody Arith-

metric Exercises) 就是這一種的測驗。中華教育改進社編造的測驗亦是屬於這一類的測驗。

倘使測驗中，排列各個問題並不注意於難度大小（即先易而後難）則這種測驗可以叫做非規律的測驗（irregular test）。普通選擇這類測驗的問題時，極容易的和極難的都刪而不用，並且於選擇時，不拿難度做基礎而拿問題的別種特性做基礎。倘使測驗中各問題的難度相差不大，則這種測驗就是上述勻遍的測驗的變相。如却士氏語言文法診斷測驗（Charters' Diagnostic Language and Grammar Test）和孟羅的標準算術推理測驗（Monroe's Standardized Reasoning Tests in Arithmetic）等。

此外，還有輪迴或循環測驗（cycle test）和螺旋測驗（spiral test）。前者（循環測驗）是把各種問題分爲若干類（type），每類問題見過一次之後，再經

一定的幾個階段或步驟又發見一次，這樣每類問題，照規定的階段發見出來。某一類的問題的難度，是假定相同的。例如依里諾亞標準的代數測驗 (Illinois Standardized Algebra Tests) 可以代表這一類測驗的編造。後者（螺旋測驗）是上述分度的測驗的變相，將全體測驗分爲許多小測驗 (sub-tests)。各小測驗裏邊的問題，難度是差不多一致的。而各小測驗和小測驗間，則難度不一樣。但是，當編製的時候，着重之點是問題的內涵而不是問題的難度。通常，計算這種測驗的分數，各小測驗，是分別計算的。再這種螺旋測驗的問題就大體上說，雖分難易，但是分得不甚細，所以實際上是勻遍測驗之彙選測驗 (battery of uniform tests)。

克萊佛蘭調查測驗 (Cleveland Survey Tests) 就是這一類測驗的例。

勢力測驗和速率 勢力測驗 (power test) 這個名稱，常指上述的分度測驗而說。用分度測驗來測驗學生，則學生的能力數常由他所達某種難度的問題和

做正這種問題的規定的準確度而定。就通常說，使學生做這種測驗時，是不限時，儘他做下去，待他完全不能做對時爲止。桑代客字彙測驗 (Thorndike Visual Vocabulary Scale) 和範滑起能的讀法測驗 (Van Wagenen Reading Scales) 是這種測驗的例子。倘把勢力測驗這個名稱的意義稍加變更，則無論何種測驗施行時，倘不加時限，多可叫做勢力測驗。施行後，被測驗者能力之高下，就看他作業的性質或準確不準確而定。

速率測驗，就是測驗被測驗者工作的速率。這種測驗，通常亦可測出被試者工作的性質。測驗工作的速率，或者是把各個人須要的時間，分別的記錄出來；或者，規定一個時限，使得所有學生在規定時間內，都不能做完測驗中的問題。有幾種速率測驗，准被測驗者，到達規定時間，將問題（規定時間到達時所做的那個問題）打個記號後，再繼續做下去，直做到完結。如上述克的斯氏標準研究測驗，

就是速率測驗的一種。通常，上述的各類測驗，多有一個時限，但除非是勻遍的測驗（測驗中問題的難度是相等的）和勻遍測驗相差「不遠的」非規律的測驗以及輪迴測驗，則所得速率的數量不很準確，沒有多大意義。

以上專就問題的性質和排列的形式把測驗分成勻遍的、分度的、不規則的等等。倘就問題的功用說，又可把測驗分做三類：（1）成績測驗（achievement test），即專供評判成績用的測驗；所有學校中用的測驗差不多全是成績測驗。（2）診斷測驗（diagnostic test）即在診斷學生於某科中的弱點而加以補救的一種測驗。（3）預斷測驗（prognostic test）即預斷學生學習某科，能否有某科必要之學習能力。例如洛果氏的預斷測驗（Roger's Prognostic Test）。這測驗編造的用意是預備測驗學生學習中等數學的特殊能力。這種測驗亦可以稱做特殊智力測驗，以別於普通智力測驗。

參考書：

Monroe's Theory of Educational Measurements, Chap. IV, pp. 56-64; Chap. VII, pp. 148-152.

Clifford Woody: Measurements of Some Achievements in Arithmetic, Part II.

問題和討論：

- (1) 已出版之標準測驗，於論斷速率一方面是否準確？
- (2) 已出版之標準測驗，於診斷一方面，是否能盡最大之效用。
- (3) 編造測驗何以須將問題分別難易，又每二題間難易之差必須約略相等？
- (4) 選擇法是否屬最可靠的有無缺點？
- (5) 現在測驗中記分公式是否尚有弊病如何改良？
- (6) 最少限度之時間，如何規定？

第四章 教育測驗的功用

教育上的測度無論是筆試，或別種粗草的方法，或用客觀的標準的測驗，主旨都在增進學校工作的效率。要增進工作的效率，非先把學生的能力，下一個準確的評判不可。準確評判的工具，當然要推標準的測驗。

學校工作裏頭，許多活動，須要評判或測驗學生能力的爲數很多。有些活動，須測驗學生學習能力（capacity to learn）方面的各種事情；有些，須知道學生的成績（achievement）的；有些，須知道學生從事於某種學習範類（type of learning）的效率的程度（或度數）的。概括說來，學校活動裏必須應用測驗的，有下列三種：

- （1）行政方面的活動。
- （2）教授方面的活動。

(3) 研究方面的活動。

這幾種活動，在學校各級中並不是同等的重要，而須先搜集或測得的各方面事實亦並不屬於同一種類。以下，將把這些活動在各級中所處的地位（重要與否）和各自所須要的各種事實，略加討論。這樣我們，才可指出學校各部分最重要的各種測驗。

(1) 行政方面活動之須要測驗的

分級和升級 通常一個學生進學校讀書，是把他插入相當的年級。在學年終了時候，凡是學生在自己級裏做滿意的工作的，都升入上一級。有時候，一個學生可以連升二級，（就外國說）有時候可以降低一級。有些學制規定，當學期中亦可升級。學生的升級，在以前是，到現在仍舊是把學生的成績做根據。近幾年來，在外國已逐漸提倡把普通智力或學習能力做根據。無論把成績或普通智力做根

據這兩方面都須要測驗——成績測驗和智力測驗。

以上是講升級，再就分級說。一級裏的學生可以分爲若干組。凡學生的學習能力差不多相同的，可以集爲一組，以便教授。倘經費充足，人材易找，則可另設優秀兒班或低能兒班。把一級分成若干組，或特設優秀兒或低能兒班，都須根據學生的學習能力或普通智力——就是都須應用智力測驗。但是，成績測驗，在這方面，亦不可忽略的。因爲有些兒童，就他的成績說，優於他的普通智力所指示的（這大都由於學習的環境特別好，或應用自己知識的機會特殊好）有些兒童，他的成績沒有達到他智力所應達的地步——智力好而成績不好。所以成績測驗和智力測驗在分級上都屬不可少。現在改進社編的各種測驗，除規定各級總成績（T）外，還有同年程度或聰明數（B）一種項目，並且訂定了一種求努力數的公式，用這種測驗，來做升級或分級，實在是很容易的事情。聰明數（B）的解釋如

下：

B 分數之解釋：

B 分數	{	0—19.9.....確定低能
		20—29.9.....近低能
		30—39.9.....笨
		40—49.9.....普通聰明
		50.....常態
		60—69.9.....較高聰明
		70—79.9.....超等聰明
		80—10.0.....天才

教育和職業指導 這種活動和上述學生的分級和升級有密切的關係，就小學前四年以下講，無所謂教育指導，實際上不過是分級和升級的事情。到了初中以上，要行選科制，這時候學生在規定的科目中有選擇的自由。這時候，倘任他

選擇，每每發生弊病，勢非加以指導不可。指導者要指導，總須有些根據，才不致妄斷某人宜選那些種科目，某人宜選那些種科目。根據最可靠的，莫如智力測驗所得的結果——從智力測驗中所指示我們的，來定某人是近於那些種學科。同時，成績測驗和各方面所指示我們的事實，自然亦不可忽略。至於職業指導，自然應當在職業範圍中搜集知識，如農工商各界目下情形，最需要的是那一種人材等等，調查好了，指導者才可指導學生；但是學生自己智力的高下，學業成績的優劣，卻是職業指導最有價值的一種基礎。

學校效率的估量 通常，大家都承認說，要學校效率增加得高，必須每隔若干規定時期，有一次檢查。所謂學校，實質是一種製造廠。學生就是原料；學校的房產地，設備，教科書等等就是廠的本身；各教員和視學員等就是這教育廠裏的工人。這廠企業的效率，就看他的產品——就是學生的成績。但是，亦必須考察用

來工作的原料的性質。所以，估量學校的效率，我們必須測驗學生的智力和他們的成績。從別一方面看，學習的方法，是形成成績裏的要素。就中等學校講，這學習方法實占成績中最重要的要素之一。所以，除測驗普通智力和成績外，還須測驗學生在各種不同的學習範類中所達效率度數。

要很有效率的，調查一學校或一學區裏的教授情形，必須先估量目下學校的效率。譬如一個視學員，想變更現在的教授或改革現有學校的組織，他必須預先估量現有種種手續所產生的效率是怎樣。

估量學校的效率，通常叫做學務調查 (survey)。學務調查中應用測驗的，在吾國始於一九二三年。所用的測驗是查氏教育測驗和德氏非文字智力測驗。教員之甄別，教員成績的好壞，一部分是看學生的成績定的。但是，學生性質的好壞（譬如工廠裏原料性質的好壞）亦應行顧及，否則我們評判教員的教

授效率，恐不可靠。一個教員，教質地不好的學生，萬不能望他使這學生產生很優秀的成績。所以甄別教員效率，學生普通智力和學業成績的測驗都是不可少。但是，同時還須注意一件事，就是估量教員效率時，除測驗學生智力和成績外，還須注意教員的品性 (Character)，人生觀，對於工作的興趣等等。因為學生所受的教育，除教員正式的教授外，教員的品性，人生觀等等，都對於學生有很深微的影響。

(2) 教授方面的活動之需要測驗的

就成績診斷學生 通常，教授分爲兩種範類。普通的教授 (General instruction)，是把普通的原理做根據。譬如，對於一羣學生，我們都用同樣的教法，同樣的解釋，做同樣的工作。我們把這一羣學生當做一個單元 (unit)。但是，就學生以前的經驗講，學習的能力，各人都是不相同。結果，往往有些學生，在這普通的教授中，得不到什麼幫助 (或好處)，別的學生在這普通教授中，已充分滿足他們的需

要。

補充普通教授的不足，必須另有一種教授，把各學生於教授上特殊的需要做根據。這種教授，可叫做補救的教授 (remedial instruction)。要發見各學生特殊的需要，勢必先對學生成就，做個特殊的測驗。這種特殊的測驗，用意在找出沒有成就的學生，尤其需要找出不能成就的是那幾種特殊的事情。所以，這種測驗亦叫做診斷測驗 (diagnostic measurement)。

要使所測得的成績，有個適當的解釋，所以又必須把學生的智力亦測驗一下。倘使，智力測驗的結果，一個學生的學習能力很高，那麼，他的學業成績亦應當高。反過來說，倘使他的學習能力很平常，我們就不應當希望他的學業成績要高。換句話說，我們只能希望學生的學業成績和他的學習能力相稱 (commensurate) 學習能力 (或智力) 低的，決不能希望他有高的成績。所以，補救教授的功効，

只限於富有必要的學習能力而學習失敗的學生（即學習能力還好而學習不佳的學生）。倘使一個學生的成績，已達他的學習能力能達的地位（即和他學習能力相稱），而教員仍努力的行補救教授，則這種努力徒然費掉，沒有相當的報酬的——產生不出什麼好結果來。

診斷的機會 就成績以診斷學生，要看學校的性質而定。小學裏頭，各科目的教授有延長到幾年之久（例如讀法，算術等等）。在這幾年中，教員繼續的，反復的把各科的技能，灌輸或訓練學生。診斷的機會就很多。機會既然很多，所以不妨在每一學年末診斷一次。而中學裏頭，有些科目一年就教完，隔一年這科目不再教授，那麼診斷的機會就少，所以宜每學期舉行一次診斷。

診斷了學生的成績，教員可以自己發見自己的缺點，可以自己知道自己的教授效率，並可以做將來教授進行上的指導。最好在學年始和學年末，各舉行診

斷一次，這時診斷要用標準測驗（現時出版的標準測驗，不全是有診斷的功用的，但是爲目前計，只得用這種標準測驗。）其餘在學年中間，可以仿照測驗卷的形式，編製各種問題來考查學生。據心理學家說，大約每五星期，做一次總復習是最有效率。那麼我們就不妨從五星期以內所授的各教材中，選擇問題，編成測驗卷的形式來試驗學生。

診斷和教授的內容 現在出版的各種教育測驗的標準，是根據大多數學生的結果定的，所以可用來評判一個人或一級成績的好壞。但是有些學校不很信仰這種測驗，說測驗中的問題，不全是學生所學習過的，怎麼能斷他的好壞。並且所選的問題也不一定是某科中公認的最少限度的知識。這兩種見解，很有些真理在裏頭。然而改良社的標準測驗發行以前，沒有公認的，全國普遍使用的各科教科書可以做選擇問題的根據，各科的內容應當怎樣，最少的限度，也沒有決

定，所以只得自由選擇重要的問題編爲測驗卷，試驗大多數的學生，看這些學生的成績定出標準，供大家應用。現在，新學制小學和初中各科的內容和畢業的最低限度都已經各專家議決。（參考新學制課程綱要）所以最好是根據新學制教科書中教材選出問題，並參酌畢業最低限度等，編造各科測驗，定出各級和各年齡的標準來。倘使用這樣編造成的標準來診斷學生，所得結果，似乎比隨便造的標準測驗，使人滿意些。

就學習的手續診斷學生 一個學生在學校的工作，須要各種不同的學習範類（different type of study or learning）。例如，在書法，讀法，算術以及其他類似的科目中，我們希望學生在這各科中得相當的技能（skill），以便去實做時或對答對問題時可以順利（不費思索，做得很快）。在別的科目中我們希望他能解決問題，獲得抽象或概括的意義。還有別的科目中間，我們希望他得到理想

(ideals)。各科的目的不同，須要的學習活動 (learning activity) 亦就不一樣。就記憶、事物說，須要某種學習的手續，就得到理想說，又要別種學習手續。要學習有效率，學生必須得到一種善良的學習技術（或方法）。他們必須知道要做的工作的進行方法。

學校工作的目標固然是很多，把學習方法訓練學生，實是學校工作目標中最重要的一種。而在中學校，這是更屬重要。中學是大學的預備，倘使一個學生已經得到很好的學習方法，他就可以說得到很重要的預備，非但對於升學上，可說是得一個重要預備，就是對於參加成人的活動上，亦可以說是得一個很重的預備。

從上述種種情形看來，學校裏教授，要有效率，必須就學生在各學習範疇中，診斷他們的學習效率。學生必須是有效率、學習效率很好的學生，在學校方面纔能希望他達到學校的目的。

在小學方面，學生的學習，多半是須要獲得技能的學習。這種學習手續，比較的簡單，通常略加觀察，就可以知道一個學生是否依照善良的手續學習。所以在小學方面須要診斷學生的學習效率的工具，還不十分顯著。在中學方面，情形就不同了，學生要在比較複雜的學習範疇中進行，因為學習的範疇比較的複雜，不容易僅由觀察來定一個學生的學習是有效與否，不能斷定他所用的方法是妥當的方法——能解決「問題」的方法或能得到「理想」（高尚觀念）的方法——與否。所以在這一方面須用測驗來診斷學生的學習習慣（study habits），比較在小學方面，要大得多。

（3）研究方面的活動

由測驗結果決定應採用的手續。所謂應用科學的態度於教育中間，就是說無論教授方法，教授策畫，課程編制，教科書以及其他教育中各項手續的取舍，

都應把他們成就或達到教育目標的效率，做基標。所以，決定什麼方法或手續是教育中最好的方法或手續，必須把這方法實地試驗後所產的結果，測驗一下。例如決定設計教學法的價值，必須把設計教學法應用於一羣學生，看他所生的結果是怎樣，然後方能定他的價值的高下。又如比較設計教學法和道爾頓式的教學法的價值，必須用兩組學生（一組用設計教學法，一組照道爾頓制進行）試驗，經過若干時期，各用標準測驗測度一次，看兩組成績的高下定他們價值的高下（這種試驗的詳細方法，參考麥可爾的實驗教育第二章）。又如，現在大家知道編制學級，一級的人數愈少愈好，二十至三十人者比四十人以上是好，也是經過試驗纔證明的。推而至於全校或一學區的組織的好壞，或兩種組織的比較，均須實地試驗後，決定。

研究方面的活動，不但必須知道學生的成績，並且須知道學生的智力，所以

學業成績測驗和普通智力的測驗都不可少。研究這件事，不用說，在小學和中學裏頭，都屬同等的重要。但是，在小學方面比在中學方面，研究的機會多些，因為現在出版的測驗工具，小學方面的，比中學方面多。

結論 學校裏的活動，凡屬須要評判學生能力的，可說用得着智力測驗。學生學業成績的解釋宜參照智力測驗的結果。學生在各學習範類中的學習效率的解釋，亦宜參照智力測驗結果。而當分級和升級時，除根據成績外，更宜看智力測驗中各學生的智力高下。

成績測驗在小學和中學都很重要。但是，在小學方面，宜就成績診斷學生在各科中的弱點加以補救；在中學方面，宜注重學生在各學習範類裏的學習效率，以便有個好好的預備。

參考書：

McCall: How to Measure in Education 第一章至第六章。

Monroe: Theory of Educational Measurement 第三章。

華超譯推孟氏訂正比納西蒙智力測驗第一章。

問題和討論。

- (1) 普通智力與成績之區別何在？
- (2) 普通智力與特殊智力之區別何在？
- (3) 診斷的測驗與普通測驗之區別何在？
- (4) 何謂預斷測驗？
- (5) 學校活動中須要測驗學生能力的是那幾種？
- (6) 小學校中最重要的測驗，屬那一範類中學呢？
- (7) 何謂補救教授？試舉補救教授之具體例子。
- (8) 爲何教育測驗與視學員之職務有關？

第五章 教育測驗的編造法

應注意的要點 編造教育測驗時，應當注意的地方，很多。現在先揭出幾個大綱，然後再進而分述各類測驗的編造法。

(a) 作業和所測能力間，必須維持恆久不變的機能關係。這條件是最重要。以下幾個條件中，有幾個都是從屬於這個條件的。這條件亦可以說是包含於第二章教育測驗的假定的第一條中。

(b) 所編測驗，必須使學生有表示自己能力的充分機會。

(c) 編的測驗，倘供考查學生成績用的，那麼所測作業應當和公認的教育目標相調和。

(d) 所測作業，應當是易於獲得的，記錄的並容易加以敘述的。

(e) 所編測驗，應當規定產生作業的情形。就是說，應當把測驗時情形，規定得明明白白，使得不同的主試者，於不同的地方，不同的時間，測驗起來，都在同樣情形下進行。

(f) 所編測驗，應當規定學生的作業，可以從作業的重要方面敘述。

(g) 偷編的是分度測驗 (scaled test) 或難度測驗，測驗中問題的難度須逐漸增加，每兩個問題之間，難度之差相仿。

這七條不能完全滿足，只能部分的實現。但滿足的條件愈多，所得的結果一定更好。所以這七條，可以說是編造測驗的指南。以下專就難度測驗 (difficulty test)、速率測驗 (rate test)、品質測驗 (quality scale or product scoring scale)、如作文、書法等測驗) 和彙選測驗 (battery of tests) 的編造法，簡略的說一說。

1. 難度測驗的編造法

(1) 決定所要測驗的心理特性 (mental trait)，並且把這種特性，嚴密規定他的意義。

(2) 決定測驗的形式和內容，使得可以測驗我們所要測驗的特性，並且僅限於這種特性；被試者對於測驗中各問題只能有一個準確的答案，（全對或全不對）使得記載成績時容易記錄。

(3) 決定所要測驗的能力的範圍。如初小一級至二級或三級。這是由編造者自己定奪。

(4) 參考前人已編造的測驗，以便決定問題的難易，測驗的說明和測驗的適宜的排列。例如自己要編算術測驗，可參考現成的標準算術測驗就是了。

(5) 倘使沒有現成的測驗，可供參考，則自己先預備一點說明和若干初步的問題。把這些問題施用於若干學生。選擇學生時，在規定的能力範圍以內，

最好的和最不好的學生，都要選得來。（這是因為要定測驗難易程度大概怎樣。）

(6) 問題的難易程度，大概知道了，然後費一點心思來規劃測驗，務必使測驗各方面都沒有缺點（如不易記錄，不易印刷等等）。排列問題，由最容易的到較容易的，一直到最不容易的。選擇時，最容易的差不多要個個人都能答出，最不容易的，差不多要一個人都答不出。預備問題的數目，大概比最後用的多四分之一。倘所編的測驗，是為診斷用的，則選擇的題數，只要足夠最後的應用，就是了。

(7) 預備以下各種要學生填的項目：(1) 姓名，(2) 男或女，(3) 年齡，(4) 生月，(5) 生日，(6) 所在學校，(7) 所在年級，(8) 測驗日期等。

(8) 再預備施行時對學生講的例子和說明。再預備些主試者所應守

的條式——這於各標準測驗說明書中都有，此地不多說。

(9) 把測驗施用於幾個聰明的成人。試用後，請他們批評。發見缺點，就隨時更正。

(10) 再把測驗施用於大約一百十個學生——這一百多個學生，要於規定能力之範圍內，各部分都有。最好的和最不好的都要有。施行時，不加時限，使得他們把每個問題都做到，或使他們完全表見他們的能力。學生交卷時，把各個學生所須的時間記在他的試卷上。

(11) 把測驗中的問題，定出一個準確或標準答案來，並想定一種可以幫助校閱的方法和怎樣記分的公式等等。

(12) 把各個問題，逐一記錄。對者，用√；錯者，用×；未做者，用○或。

(13) 刪去問題中雙關的或不易評判的和其他不完備的問題。

(14) 把一百十本卷子，抽去十本，餘下一百本。但是最好的卷子或最不好的，不抽去。

(15) 計算各人做對的單數題共多少；再算各人做對的雙數題，共多少。

(16) 把這兩組分數，做一個相關圖。招集那使兩組分數相關低減的學生來，討論一下。再檢查這些學生做或未做的許多問題，有沒有含糊不妥當的地方。倘使發見了缺點就較正他，或竟把他刪去。

(17) 把各學生所得總分數和在別種標準（如同樣的標準測驗等）所得的分數，做個相關圖。做好了，再討論，加上校正。

(18) 招集最聰明的學生，問他們爲什麼有許多問題要答錯。倘使發見問題本身上有缺點，再加上校正，或把問題刪去。

(19) 把餘下來的各問題分別學生姓名，一一記入下列表格中。這樣把一

百本卷子都記好。

姓名	問題數。															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
王 某	×	×	√	√	○	×	√	√	○	√	√	×	×	√	○	其他
李 某	√	√	×	×	√	√	×	×	○	√	√	×	○	○	其他	
總 對	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
標 差	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	

(20) 從上表計算做對每一問題的學生數和百分數。因為學生一共一百人，總對欄內所記的數字就是所須要的百分數。但是當學生有五十對五十的機會，對正一個題目時，則這欄所記的還不是百分數，必須從每題總對數中減去總錯數，然後用一百除，所得的商纔是對正的百分數。

(21) 把百分數，轉爲標準差(S. D.)價值或T難度。這用麥氏教育測驗第二十三表，或麥氏實驗教育第七表(即本書第二表)。

(22) 照標準差價值，把問題重行排列。數量小的在前，大的在後。

(23) 參酌各學生試卷上所記錄的時間和最後測驗的時間，決定問題的數目，以便最快的學生於規定時間到達時，亦不能完全做完。規定時限時，應當顧及測驗的實用和可靠兩方面。通常說來，凡可靠度(自己相關度)須達百分之八十五。別種事情可不問，凡簡短的測驗，可靠度不會高。所以，倘使自己相關度太低，則須把時限延長，或增加測驗的問題，或把一個很長的測驗，拆做兩個測驗。把兩個測驗的結果，平均一下，作爲學生的成績。

(24) 選擇那已經定妥的幾個問題。怎樣去選擇呢？就是各個問題的T難度(或標準差的價值)的增加，務必相等。就是，先選一問題有百分之九十九人

做對，一個一個選下去，到只有百分之一人能做對的問題，倘使所有的問題是太易或太難（就是不能從百分之九十九，九十八，九十七，逐漸到百分之一，中間有空缺的地方），則必須另行想法，求出合用的問題來加進去。有時候，決定測驗的最後內容時，診斷或別種目的，應當比問題的難度或時間限制，多占勢力（格外加以注重）。這全由編造測驗者自己判斷，以定測驗的內容究竟可以變更多少。

(25) 看測驗的構造和說明是不是還有缺點。倘使發見，就盡力改正。

(26) 把測驗稿正式付印，作為最後應用。

(27) 要證驗假定的時間，是不是一個適宜的時間，再把測驗應用於規定範圍中能力最高的學生。測驗開始後，每隔若干分鐘，使學生於正在做的一題上，打個圈或別種記號。看有幾個最快的學生已做完，便叫別的學生一起停止。

把這些最快的學生所須要的時間，記錄出來。

(28) 計算每一個時間階段內，每個學生做對的題數，共有多少。這樣，把各個學生做對的總題數，都算出來。

(29) 規定最後須用的時限，要在規定時限內，最快的學生不能完全做完所有的問題，而能力最高的學生亦已完全發揮他的能力。然後參照這兩種目的，定一個時間的最小限度。

(30) 把正式的測驗，施用於二千（愈多愈好。華狄的算術測驗定出標準，曾測驗過二萬多學生）左右的學生——在規定能力範圍以內的學生。選擇學校，務必隨意選擇，不能加上主觀的意見，這就是說，要使所選的學校，好的，壞的都要有，這樣求來的標準，才可以供普遍的應用。在所選的學校中，而又在規定能力範圍以內的各級學生，都應加以試驗。

(31) 二千多人測驗過後，把試卷校閱，計算每個學生的總分數（即共對多少題）。當記分數時，凡做對一題者作一分算。有些人喜歡看學生答案的好壞，給二分，一分或零分。但是這種記法很費時間與心力，而結果並不十分增加記分的準確度。通常，問題範圍較大的，可以析做二個以上的問題，叫學生答，每對一問，給一分。所以給分數，不要依照問題的難度為比例（易者，少給分數；難者，多給分數）。倘使這樣，於計算時非常不便，且容易發生錯誤。

(32) 做一個各年級分數的分配表；再做一個各年齡分數的分配表。把次數分配，填入距度中。通常每一距度，以一為單位。

(33) 看所編的測驗是為初級小學，高級小學或中學校用，然後選擇八到九，十二到十三，或十六到十七歲的學生。把這些學生做對的分數（或原得分數）變成T分數（或總成績）。變T分數時，用麥氏實驗教育第七表，或教育測

測驗第二十三表（即本書第二表）

(34) 倘 T 分數的範圍太狹時，便擴充 T 的範圍。怎樣擴充 T 分數的範圍，參考下章，T, B, O 量表的編造法。

(35) 造一個 B 的量表（即同年程度或聰明數的量表），方法見下章。

(36) 造一個 O 的量表（即年級程度或分組數的量表），方法見下章。

(37) 預備一本說明書，——說明怎樣使用測驗，記算成績等——以便和測驗同時發行。這可參考已出版的各種說明書。

II. 速率測驗

除所有問題的難度須彼此一致或差不多一致外，編造速率測驗之第一，第二，第三，第四步驟和編造難度測驗相同。除逐漸增加各個問題的難度外，照上述難度測驗之第五，第六步驟做去。除須規定時間限制外，第七，第八，第九步驟亦和

難度測驗相同。再選測驗中幾個代表的問題，照上述第十一到十九步驟做去，看測驗的問題是不是合於我們希望的難度。總之，除論及難度的事情以外，以後的步驟都是和上述的一樣的。

因為速率測驗，通常產生兩種分數：試做分數和做對分數。所以，我們可以把這兩種分數，各編 T, B, C 量表；或只把做對分數，編 T, B, C 量表，或把兩種分數適當均衡一下。

III. 品質測驗（如書法，作文，圖畫等）

(1) 照難度測驗的第一，第二步做，不過評判學生在品質測驗中成績，是注重在全體上而不甚注重於各部分的。第三，四，五，六各步亦同，不過講到難度的是不適用。第七，八，九，十各步驟亦是一樣，不過要有一個時限；而施行作文或圖畫測驗時，在規定時限到達前幾分鐘，須先給學生一個警告。

(2) 把第十步驟再應用於同一團體的學生，以便得到兩組數量。譬如第一次測驗一百十人；第二次再測驗這一百十人；結果得到兩組數量（或兩種分數）。

(3) 把兩組數量，照第十四步做，就是各刪去十個數量（或十本卷子）。

(4) 請三個以上（愈多愈好）有能力的評判者，把第一組一百本卷子評閱。最壞的，打個(1)；次壞的，打個(2)。這樣一直類推下去，把一百本卷子打分數。然後把三個評判者所給每本卷子的分數，平均一下，得每本卷子的最後等第。

(5) 把第二組一百本卷子，亦照上述(4)的辦法做一下。

(6) 把兩組分數做個相關的圖。或把兩組分數利用標準測驗所得的分數，各做一個相關的圖。或把一組的分數利用標準測驗所得的分數，做個相關

圖，亦可。倘使自己相關度太低，便延長時間，或預備二套以上的測驗，平均一下，當做一個測驗用。詳細，看上述難度測驗第十六及十七兩步驟。

(7) 校正測驗的構造和施用時的各種說明。然後把測驗正式付印。印好後，施用於二千多個學生。選擇學生和學校，應注意的地方，詳上難度測驗第三十步驟。

(8) 看規定能力的範圍怎樣，然後選入到九，或十二到十三，或十六到十七歲的卷子（編的測驗為初小用者，選入到九歲學生的卷子；為高小用者，選十二到十三歲學生的卷子；為中學用者，選十六到十七歲學生的卷子）。通常，選十二到十三歲學生的卷子，對於各程度的能力都可適用。把選好的卷子，不顧成績好壞，每本做一個記號。

(9) 把選好的卷子，依照成績的好壞分為十堆——A（最壞的）B（次

壞的) C, D, E, F, G, H, I, J (最好的)

(10) 再把A的一堆,照成績的好壞,分成五堆——a (最壞的), b (次壞的), c, d, e (最好的)

(11) 照上第十步的方法,把其餘十堆,各分做五小堆。

(12) 把Aa堆裏卷子,再依成績好壞,順次排列。

(13) 照上第十二步方法,把Ab, Ac, Ad, Ae, Ba, Bb, Bc等五十堆,各依好壞順次排列。

(14) 把Aa裏邊最好的幾本卷子和Ab裏邊最壞的幾本卷子,兩相比較。比較後,倘使排定的等第不準確,便在兩鄰接的等第裏邊,更正等第先後。同樣,再比較其餘的四十八鄰接的等第。倘發見不確當處便更正。

(15) 用一記錄紙,把所有的卷子,依照好壞順序把各卷的數目記下來。再

把最壞卷子，打一個(1)字；次壞的卷子，打一個(2)字。這樣，把所有卷子都打好分數。

(16)以上第九，十，十一，十二，十三，十四和十五各步驟，須分請三個有能力
的評判者做。做的時候，要各人彼此不知道各人所打的記號（或某卷是何種
等級）。

(17)把每本卷子，三個人所給的分數，平均一下，得一平均分數。照平均分數
的大小，依次排列。

(18)算某號卷子的分數的超過數和二分之一的到達數的百分數是九
九·八六五的。照本書第二表換算，這號卷子，便作為有二十分價值的卷子。再
算某號卷子的分數的超過數和二分之一的到達數的百分數是九九·三八
的。這號卷子，依表換算，便有二十五分的價值。其餘別種的百分數，換算成S. D.

(或T分數)參考下面的例子(共十三本卷子)倘使我們只要選入本作爲最後的應用,則用相當於T分數二十,三十,四十,五十,六十,七十和八十的各百分數,就是了。倘使須用的卷子數要在十三本以上,作爲最後的應用,則查本書第二表,那幾個百分數是相當於那幾個T分數。查定,再在卷子中選擇。茲舉百分數與T分數對照的幾個例於下:

百分數	99.865	99.38	99.72	99.32	84.13	69.15
T分數	20	25	30	35	40	45
百分數	50	30.85	15.87	6.68	2.28	.62
T分數	50	55	60	65	70	75
						80

(19)十三號(不必一定是十三本)的卷子已經選出來之後,再依照好壞,在每號之前後,各選五本。這樣十三號卷分爲十三組——N, O, P, Q, R, S, T,

U, V, W, X, Y 和 Z —— 每組有十一本。把每組中十一本卷子，故意的混亂一下（即不照好壞順序排列）。

(20) 請多數評判者，把 N 組裏頭的卷子，依照好壞排先後，並請他們另用一紙把卷子的號數記出來，同時於號數旁，最壞的記 1 分，最好的記 11 分。這時，前三個評判者所給的分數，可以用來比較對照。

(21) 把其餘十二組的卷子，照第二十步方法，一個一個的記出號數和分數。

(22) 把多數評判者所給每本卷子的分數，平均一下，得每本卷子的平均分數。

(23) 在 N 組中，參照各平均分數（即共有十一個平均分數），把最中心的一本，選出來。這一本卷子，是最可以代表二十 T 分數的（即值二十 T 分數）同

樣，再在 O, P, Q 等組中，各選最中心的一本，而各給相當 T 分數二十五、三十、三十五等等。這十三本卷子，聯同各自的 T 分數，就形成品質測驗的量表。無論那一學生的書法（或圖畫或作文）成績，要評他的好壞，只要在這量表上，逐步比對，看和量表上那一種相近，倘和二十 T 分數的一種相近，則這學生的書法分數就是二十分，倘和量表上三十五 T 分數的相近，那末他的分數就是三十五分。

(24) 請至少三個評判者，用已成的量表，把二千多本原來的卷子中沒有評閱過的，一本一本的評閱，打上分數。然後把三人對每本卷子所給的分數平均一下，得一平均分數，這平均分數就算每本的卷子的 T 分數。

(25) 做一個分數的年級分配表。再做一個年齡分配表（十二至十三歲不必再計算）。

(26) 做一個 B 量表(同年程度量表)和一個 O 量表(卽年級程度的量表)。再預備一本說明書和量表同時發行。

(27) 倘純粹爲研究起見另有一種較費時，較準確的編造法，見麥氏教育測驗第九章，第四節。

倘依照這書(麥氏教育測驗)中說的去編造，則從第八到第二十三步驟都可省去。做本節中第二十四，二十五(十二到十三歲分數的年齡分配亦要計算)步驟，再照難度測驗中的第三十三，三十四，三十五，三十六步驟做去。

IV. 彙選測驗

(1) 難度測驗中，從第一步驟到第二十五步驟，適用於無論何種測驗的編造，所以編造彙選測驗(一種測驗中含有幾種形式的測驗，或一種測驗中含有測驗數種能力的測驗，如廖氏團體智力測驗，陳氏初小默讀測驗等)就

把各難度測驗、速率測驗或品質測驗，照上所述，預備到第二十五步驟。倘使彙選測驗中，雜有品質測驗的，則必須另外編造一種評判品質的量表——這量表要根據一千多個不加選擇而年齡在八到九，十二到十三，或十六到十七歲的學生的成績編成。

(2) 把彙選測驗中各個部分的測驗，應用於一百多個學生後，所得的結果，一一的批評，凡不適用的刪去。倘使各個部分的測驗，只經過了初步的手續，便把這些測驗，應用於一百多個學生。

(3) 決定了最後所用的問題數目，計算一百個學生在各個部分測驗中所得的總分數。

(4) 做各個部分測驗中一百個學生分數的次數分配表。

(5) 把各個次數分配表，算出標準差來。

(6) 倘使彙選測驗中各個部分測驗，要占同等的勢力，選一個可以使標準差大小差不多相同的乘數。例如：

標準差	4	2	8	11
乘數	1	2	$\frac{1}{2}$	$\frac{1}{3}$

倘使各個部分測驗，不須占同等勢力，就選一個可以使標準差顯出我們所希望的比率的乘數。選的乘數，要很容易計算的，免得多費勞力。

(7) 把彙選測驗正式印刷封面上添進乘數，形式如下：

測驗	原得分數	乘數	均衡分數
1	1
2	2
3	$\frac{1}{2}$
4	$\frac{1}{3}$
總得		

(8) 把彙選測驗中的難度測驗，定出一個時間的最小限度。方法，見上述難度測驗中第二七，二八，二九各步驟。

(9) 把彙選測驗應用於二千多個學生。

(10) 計算每個學生在各個部分測驗中所得的總分數。

(11) 把每個學生所得總均衡分數算出來。參考(7)。

(12) 做均衡分數的年級和年齡分配表。

(13) 選入到九歲，十二到十三歲，或十六到十七歲學生，把他們各人的均

衡分數，變換成T分數。

(14) 倫T量表的上下端，不夠用，便擴充T量表的範圍(方法見後)。

(15) 造一個同年程度的量表(方法見後)。

(16) 造一個年級程度的量表(方法見後)。

(17) 預備一本說明使和彙選測驗同時發行。

參考書：

麥氏實驗教育第五章，一三〇—一三九頁。麥氏教育測驗法第十章，二七二—二八〇；第九章，二六三—二七一頁。

問題和討論：

- (1) 除上述七種應注意事項外，編造測驗時是否尚有別種事項應加注意？
- (2) 編造測驗時，被測驗者人數何以愈多愈好？
- (3) 求標準時，倘有意選擇學生，其影響如何？
- (4) 以十二歲兒童為標準，較用每一年齡各求一標準，孰妥？並述其理由。
- (5) 使用作文、書法等量表，宜如何練習，始不發生弊病？

第六章 最近T, B, C量表的編造法

(一) T量表

T量表之編造法，已經於難度測驗中略已說過。詳細情形，可以參考麥氏教育測驗第十章。下面一個表，就是解釋T量表之做法的。表中第一縱行，是做對的問題數，第二縱行是指不加選擇的十二歲學生的數目。編造的測驗倘供普通的，通常可用十二歲（十二至十三歲間）學生做被測驗者。倘用別種年歲的兒童，則於表上須另加 T_{11} 或 T_{13} 或 T_{16} 等注解。倘使為實驗用，實驗者就可用一羣或數羣實驗的學生。表中第三縱行，是各做對題數之超過數加二分之一的達到數的人數。譬如做對題數超過三十三的是零（就是說沒有人）而做對三十三題的人數之半是0.5。零和0.5加起來，是.5。所以第三縱行末，是0.5。又如，做對題數超過三十二

表
顯示怎樣把總分數分成尺度

做對題數	十兒 二童 歲數	半 加 好	半百 加分 好數	量表分數
0	3	498.5	99.7	23
1	1	496.5	99.3	25
2	2	495.0	99.0	27
3	1	493.5	98.7	28
4	2	492.0	98.4	29
5	2	490.0	98.0	29
6	2	488.0	97.6	30
7	2	486.0	97.2	31
8	4	483.0	96.6	32
9	2	480.0	96.0	32
10	2	478.0	95.6	33
11	10	472.0	94.4	34
12	3	465.5	93.1	35
13	8	460.0	92.0	36
14	8	452.0	90.4	37
15	13	441.5	88.3	38
16	15	427.5	95.5	39
17	18	411.0	82.2	41
18	28	388.0	77.6	42
19	26	361.0	72.2	44
20	34	331.0	66.2	46
21	40	294.0	58.8	48
22	40	254.0	50.8	50
23	41	213.5	42.7	52
24	37	174.5	34.9	54
25	31	140.5	28.1	56
26	35	107.5	21.5	58
27	24	78.0	15.6	60
28	26	53.0	10.6	62
29	21	29.5	5.9	66
30	14	12.0	2.4	70
31	3	3.5	0.7	75
32	1	1.5	0.3	78
33	1	0.5	0.1	81
34	0			85
35	0			90

是 1.5。2 和 1.5 加起來是 3.5。第三縱行中別的數目的計算，類推。因為造這尺度時用末第二的數目是 1.5。又如做對題數超過三十一的是 2。做對三十一題人數之半是 1 人，而做對三十二題的人數之半是 0.5。1 和 0.5 加起來是 1.5。所以第二縱行

的學生共五百人，所以第四縱行各數量，是把五百去除第三縱行各數量所得的商數，在表中稱爲百分數。倘使把第四縱行顛倒一下子，則第一縱行和第四縱行就成爲一個「百分量表」。第五縱行是T分數。用表一中的百分數，照表二計算就是了。例如百分數是99.7，照表二，T分數（或S.D.）是82.6，或簡寫爲23。

T量表範圍之擴充 根據十二歲兒童所做的T量表的闊度（或範圍）在沒有經驗的人看來，以爲太小，不夠應用。事實上，卻不然。例如對於連續不斷的機能（continuous function）像讀法等科目，這種量表可以測驗初小一年級到大學中大多數的學生。自然，這兩極端的測驗，比近於十二歲兒童分配中心的，是較不可靠。對於非連續不斷教授的機能（non-continuously-taught）像代數一類的科目（有時讀法亦屬非連續不斷的機能）這種量表自應把十二歲兒童所達到的範圍擴充一下。擴充的方法，是把表一中的過程，複演一下，就是用小學和中學中

表一 顯示在零點以上某百分數之標準差距離。每一標準差之價值用10乘以以免小數。零點定在平均數下5個標準差。表中標準差等於T。

標準差	百分數	標準差	百分數	標準差	百分數	標準差	百分數
0	99.999971	25	99.98	50	50.00	75	0.62
0.5	99.999963	25.5	99.29	50.5	48.01	75.5	0.54
1	99.999952	26	99.18	51	46.02	76	0.47
1.5	99.999938	26.5	99.06	51.5	44.04	76.5	0.40
2	99.99992	27	98.93	52	42.07	77	0.35
2.5	99.99990	27.5	98.78	52.5	40.13	77.5	0.30
3	99.99987	28	98.61	53	38.21	78	0.26
3.5	99.99983	28.5	98.42	53.5	36.32	78.5	0.22
4	99.99979	29	98.21	54	34.46	79	0.19
4.5	99.99973	29.5	97.98	54.5	32.64	79.5	0.16
5	99.99966	30	97.72	55	30.85	80	0.13
5.5	99.99957	30.5	97.44	55.5	29.12	80.5	0.11
6	99.99946	31	97.13	56	27.43	81	0.097
6.5	99.99932	31.5	96.78	56.5	25.78	81.5	0.082
7	99.99915	32	96.41	57	24.20	82	0.069
7.5	99.9989	32.5	95.99	57.5	22.66	82.5	0.058
8	99.9987	33	95.54	58	21.19	83	0.048
8.5	99.9983	33.5	95.05	58.5	19.77	83.5	0.040
9	99.9979	34	94.52	59	18.41	84	0.034
9.5	99.9974	34.5	93.94	59.5	17.11	84.5	0.028
10	99.9968	35	93.32	60	15.87	85	0.023
10.5	99.9961	35.5	92.65	60.5	14.69	85.5	0.019
11	99.9952	36	91.92	61	13.57	86	0.016
11.5	99.9941	36.5	91.15	61.5	12.51	86.5	0.013
12	99.9928	37	90.32	62	11.51	87	0.011
12.5	99.9912	37.5	89.44	62.5	10.56	87.5	0.009
13	99.989	38	88.49	63	9.68	88	0.007
13.5	99.987	38.5	87.49	63.5	8.85	88.5	0.0059
14	99.984	39	86.43	64	8.08	89	0.0048
14.5	99.981	39.5	85.31	64.5	7.35	89.5	0.0039
15	99.977	40	84.13	65	6.68	90	0.0032
15.5	99.972	40.5	82.89	65.5	6.06	90.5	0.0026
16	99.966	41	81.59	66	5.48	91	0.0021
16.5	99.960	41.5	80.23	66.5	4.95	91.5	0.0017
17	99.952	42	78.81	67	4.46	92	0.0013
17.5	99.942	42.5	77.34	67.5	4.01	92.5	0.0011
18	99.931	43	75.89	68	3.59	93	0.0009
18.5	99.918	43.5	74.22	68.5	3.22	93.5	0.0007
19	99.903	44	72.57	69	2.87	94	0.0005
19.5	99.886	44.5	70.88	69.5	2.56	94.5	0.00043
20	99.865	45	69.15	70	2.28	95	0.00034
20.5	99.84	45.5	67.36	70.5	2.02	95.5	0.00027
21	99.81	46	65.54	71	1.79	96	0.00021
21.5	99.78	46.5	63.68	71.5	1.58	96.5	0.00017
22	99.74	47	61.79	72	1.39	97	0.00013
22.5	99.70	47.5	59.87	72.5	1.22	97.5	0.00010
23	99.65	48	57.93	73	1.07	98	0.00008
23.5	99.60	48.5	55.96	73.5	0.94	98.5	0.000062
24	99.53	49	53.98	74	0.82	99	0.000048
24.5	99.46	49.5	51.99	74.5	0.71	99.5	0.000037
						100	0.000029

的九歲和十六歲（在十二歲以上或以下，不拘那歲兒童都可，只要近於十二歲就是了）兒童所得的T分數算出來，再把這些T分數和十二歲兒童的T分數合起來。表二就是顯示這種聯合方法的。

三
表
顯示擴充T量表之範圍

做對題數	T ₉	T	T ₁₆	最後T量表
0	32			22
1	36			26
2	40			30
3	43	33		33
4	46	35		35
5	48	38		38
6	50	40		40
7	52	43		43
8	54	45	34	45
9	58	48	37	48
10	61	50	40	50
11	65	53	42	53
12	70	56	45	56
13		59	47	59
14		63	50	63
15		67	53	67
19		71	56	71
17		75	60	75
18		80	65	80
19			70	85
20			76	91

(二) B量表

此地要預先解釋的，是T量表怎樣能包含年齡量表 (age scale) 中所有各種商數或比率 (quotient) 的特長。表四就是年齡量表和T量表的對照表。年齡

表 四

顯示年齡量表和 T 量表之對照

年 齡 量 表	T 量 表
C. A. = 實足年齡	C. A. = 實足年齡
M. A. = 心理年齡	Ti = 智力總能數
E. A. = 教育年齡	Te = 教育總能數
R. A. = 讀法年齡	Tr = 讀法總能數
Ar. A. = 算術年齡	Ta = 算術總能數
其他	其他
$I. Q. = \frac{M. A.}{C. A.}$ = 智力商數	Bi = 智力聰明數
$E. Q. = \frac{E. A.}{C. A.}$ = 教育商數	Be = 教育聰明數
$R. Q. = \frac{R. A.}{C. A.}$ = 讀法商數	Br = 讀法聰明數
$Ar. Q. = \frac{Ar. A.}{C. A.}$ = 算術商數	Ba = 算術聰明數
其他	其他
$A. Q. = \frac{E. A.}{M. A.}$ = 成業商數	F = Te - Ti = 努力數或效率數
$R. A. Q. = \frac{R. A.}{M. A.}$ = 讀法成業商數	Fr. = Tr - Ti = 讀法努力數
$Ar. A. Q. = \frac{Ar. A.}{M. A.}$ = 算術成業商數	Fa. = Ta - Ti = 算術努力數
其他	其他

量表的詳細解釋，可以參考麥氏教育測驗第九章。表中B和F兩種記號是教育測驗一書出版後纔發明的。

T_i 是某種智力測驗的T分數（總成績或總能數）。 T_e 是幾種教育測驗的平均T分數。 T_r 是讀法測驗之T分數。 T_a 是算術測驗的T分數。各種F（努力數）可以從各相當公式求得。當 $F_o = F_i$ ，結果是正的，這個學生或這班學生在教育上的進步，可以說是比有同樣智力的一個學生或一班學生好。反之，就可以說不好。

B是由T得來的。例如，計算 B_i ，必須要有智力測驗的各年齡的T量表，就是須要 $T_{10}, T_{11}, T_{12}, T_{13}$ 等等量表。倘使一個學生，年齡是10歲，他的 T_i （智力總能數）就是 T_{10} （十二歲兒童之總能數）的分數，但是他的 B_i （智力之聰明數）就是 T_{10} （十歲兒童之總能數）的分數。倘使他是13歲，他的 T_i 分數，仍舊是 T_{10} 的分數，而 B_i 是把 T_{13} 的分數算分數。倘使他是12歲，他的 T_i 分數是把 T_{12} 的分數算分數，

B_i 也把 T_1 的分數算分數。一個學生的 T_1 分數應隨年齡而俱進，但是 B_i 分數，是比較的不變動（假定天賦智力之聰明程度是不變的）。倘使他是個普通十歲的兒童，他的 B 是 50；當他到十一歲的時候，他的 B 仍是 50（只要他常是保持普通的地位，他的一生 B 總是 50）。 B_r 之計算，和 B_i 之計算同；不過計算 B_r 要用讀法測驗罷了。表中 B_e 是 B_r 和別種教育測驗中所得 B 的平均數。

倘使能得到各年齡不加選擇的兒童（即每一年齡均有可以代表該年齡之兒童），則各年齡 B 量表的編造，不過是重複 T 量表的編造（即與編造 T 量表同）。但是，此地有種困難——得不到代表各年齡的兒童。因為十三、十四、十五歲兒童中，一部分最聰明的兒童已在中學校，或已經完全脫離學校；而七、八、九歲兒童中一部分愚笨的兒童還沒有升入一年級，或都在一、二年級中，不使用文字的測驗來測驗他們（現在規畫的測驗，大都數只能用於三年級以上，三年以下不

能用)。結果，要編造各年齡的T量表，實際上不容易做到。

各年齡的T量表，既然因困難做不到，改用下面比較簡單的手續，使他產生和各年齡的T量表相類似的結果。分做十二項，敘述於下：

(1) 做一個和表五相同的年齡分配表。

(2) 算出每一年齡中學生的總數。把這總數寫在下面(參考表五)相當的縱行中。

(3) 造一個以十二歲兒童為根據的T量表。把這個T量表的價值，寫於第二縱行(參考表五)。

(4) 計算年齡最小的學生總數之半數。表五中七歲學生三十五，一半是十七·五。

(5) 於年齡最小的一縱行，從底下向上加起來，加到超過半數的一數為止。

表 五

顯示七至十七歲兒童答對題數(見第一縱行)和相當的量表分數(見第二縱行)。

答對題數	量表分數	7	8	9	10	11	12	13	14	15	16	17
0	23	1	3	1	2	1	3	5				
1	25	2	3	3	4	1	1	0				
2	27	2	3	2	1	1	2	0	1			
3	28	3	0	6	3	1	1	0	0	2		
4	29	0	5	5	5	1	2	0	0	0		
5	29	2	5	9	6	1	2	2	2	0		
6	30	2	6	6	5	1	2	1	1	0	1	
7	31	0	10	6	3	5	2	2	0	0	0	
8	32	1	8	9	6	4	4	0	1	0	0	
9	32	2	10	5	5	2	2	1	0	0	0	
10	33	2	6	15	8	6	2	3	2	0	0	
11	34	2	11	20	5	4	10	1	0	1	0	
12	35	2	9	21	12	3	3	6	2	1	0	
13	36	4	14	25	12	4	8	3	1	1	0	
14	37	1	12	23	17	12	8	4	1	3	0	
15	38	2	13	21	25	15	13	12	5	2	0	
16	39	0	17	25	23	22	15	6	4	3	0	
17	41	2	17	34	24	31	18	14	4	4	0	
18	42	1	5	29	25	20	28	19	11	5	1	
19	44	3	3	20	27	32	26	26	21	3	0	
20	46	0	4	22	33	42	34	26	19	5	1	
21	48	1	4	18	25	35	40	32	28	10	2	
22	50		2	6	30	40	40	35	25	6	1	
23	52		2	6	27	32	41	42	24	9	2	
24	54		1	8	16	29	37	42	38	8	1	
25	56			3	17	22	31	46	24	16	2	
26	58			6	9	16	35	39	23	18	1	2
27	60			0		11	16	24	17	3	2	
28	62			2		3	13	26	23	5	1	
29	66				7	3	21	19	12	5	0	
30	70				2	4	14	11	7	2	1	
31	75				1		3	6	5	1		
32	78					0	1	1	3			
33	81					1	1	2				
34	85											
35	90											
學生總數	B量表分數	35	173	347	399	426	500	452	303	118	16	2
B量表分數	T量表分數	68	59.5	53.5	53	52.5	50	48.5	44	38	28	21
B校正數	B校正數	34	36.0	38.0	44.0	48.0	50	52.0	52	54	52	58
	B校正數	34	23.5	15.5	9	4.5	0	-3.5	-8	-16	-24	-37

把這超過半數的一數用二除，再將得數加於超過半數的一數下各數之和。這就是左邊T分數（量表分數）的超過數加二分之一的到達數。舉個例來說，七歲兒童的半數是一七·五。從下面數上去， $1+0+3+1+2+0+2+1+4+2+(2+2)=17$ 。這17就是T分數34的超過數加二分之一的到達數。

(6) 用十二歲兒童的總數除前項 (5) 超過數加二分之一的到達數。十二歲兒童共五百，所以 $17+500$ 得百分之4。³ 照表二，把這百分數轉為T分數，得68（參考表五下B量表分數）。倘使所有七歲的兒童都經過測驗而得一T₇的量表，則做對十一題的T分數，大概在六十八左右。

上述的手續，是假定七歲兒童中，沒有一個能比這三十五個受過測驗的兒童的好的一半更好。這個假定，可以說是合理的（reasonable），而應用於八歲，九歲，十歲和十一歲的兒童，更可以說是合理的。以上的手續更有一個假定，就是，十二

歲不加選擇的兒童既有五百個，那麼低年級中七歲的兒童亦有同樣的數目存在。

(8) 把十二歲兒童的T分數，寫於七歲的T分數之下；如表五，把34寫於68之下。

(9) 從T₁分數中減T₂分數。餘數，得正34（參考表五）。這餘數就是聰明數或B量表的校正數。這樣，倘使一個七歲的學生，做對測驗中九個問題，照表五第二縱行，他的T分數是32。他的B分數，是32加上校正數34，得66。這B分數66，指出這個學生比普通學生好16點，或照表二計算，七歲學生中只有百分之五·四八比他好。

(10) 把上述(4)，(5)，(6)，(7)，(8)，(9)各項手續應用於十二歲以下各年齡的兒童。十二歲兒童的B校數爲零。再把十一歲兒童做例子，計算的步驟如下：(3) $426 \div 2$

$$= 213. \quad (b) 1 + 0 + 6 + 4 + 3 + 13 + 16 + 16 + 22 + 29 + 32 + 40 + (35 \div 2)$$

$$= 199.5. \quad (c) 199.5 \div 500 = \text{百分之} 39.9. \quad (d) \text{百分之} 39.9 = \text{十一歲的} T$$

$$52.5. \quad (e) 52.5 - 48 = B \text{校正數} 4.5.$$

(II) 計算十二歲以上各年齡的 B 校正數，和計算十二歲以下的相類似。不過，有個不同之點，應當加以注意，就是表五中失去比較好的學生而不是壞的學生（此言十二歲以上學力優秀者已升入中學，不在測驗範圍內，表五中僅屬一部分較不優秀之學生）。所以求得超過數加二分之一的到達數之結果後，再須加上所失去的一部分比較好的兒童數。例如，十三歲兒童中失去的比較好的兒童是 48 (500 - 452)。既知道失去的比較好的兒童是 48，然後再依下列步驟計算：

$$(a) 452 \div 2 = 226. \quad (b) 2 + 1 + 5 + 11 + 19 + 25 + 24 + 39 + 46 + 42 + (42 \div 2)$$

$$= 235. \quad (c) 235 + 48 = 283. \quad (d) 283 \div 500 = \text{百分之} 56.6. \quad \text{百分之} 56.6 =$$

十三歲的 T 48.5. (f) $48.5 - 52 = B$ 校正數 -3.5 . 有了這 B 校正數, 十三歲兒童的 B 分數就可算了。例如, 倘使一個十三歲兒童得 T 分數 40, 他的 B 分數就是 $40 - 3.5 = 36.5$ 。

所有各年齡的 B 校正數, 見表五最末一橫行中。因為七歲, 十六歲, 十七歲這三個年齡的人數太少, 所以這三年齡的校正數不可靠。

以上是把每一年齡兒童總數的半數 (half-sum) 詳細一點說, 超過數加二分之一的到達數) 求出 B 校正數。據麥柯爾說從這法求出來的 B 校正數和用下二法求出來的 B 校正數相差不遠: (1) 從人數行自下向上數到總的四分之一 (quarter sum) 這是指計算十二歲以下各年齡說) 或自下向上數到總數的四分之三 (three-quarter sum 這是指計算十二歲以上各年齡說); (2) 用 T_{12} (十二歲的 T) 計算各年齡的估計真平均分數 (estimated true mean score 此法

詳見於麥氏教育測驗第二十表。下列表六，就是表示用這三種方法所得的B校正數是約略相同。表中左邊I，指用總數半數法；II，總數四分之一或四分之三法；III，估計真平均分數法。

(12) 最後一步，要決定七歲和八歲，八歲和九歲，九歲和十歲，十歲和十一歲中間的B校正數。這可用簡單的插入法 (interpolation) 來決定。例如，七歲或九十個月的B校正數是三四，八歲或一百零二個月的B校正數是二三·五。那麻一〇二——九〇，相當於三四——二三·五；一個月便等於10.5 + 12，就是〇·八七五。九十

表 六

顯示用總數半數法，總數四分之一或四分之三法，估計真平均分數法，所得的校正數。

年齡	7	8	9	10	11	12	13	14	15	16	17
I	34.0	23.5	15.5	9	4.5	0	-3.5	-8	-16	-24	-37
II	33.5	24.0	16.0	8	4.5	0	-3.5	-7	-12	-22	-37
III	15.0	9	4.0	0	-4.0	-10

表 七

表示由實足年齡，將T分數轉為B分數。

實足 年齡 年月	加於 T分 數上	實足 年齡 年月	加於 T分 數上	實足 年齡 年月	加於 T分 數上	實足 年齡 年月	加於 T分 數上
7- 6	34	10- 2	11	12- 8	- 1	15- 2	-13
7- 8	32	10- 4	10	12-10	- 1	15- 4	-15
7-10	31	10- 6	9	13- 0	- 2	15- 6	-16
8- 0	29	10- 8	8	13- 2	- 2	15- 8	-17
8- 2	27	10-10	8	13- 4	- 3	15-10	-19
8- 4	25	11- 0	7	13- 6	- 4	16- 0	-20
8- 6	24	11- 2	6	13- 8	- 4	16- 2	-21
8- 8	22	11- 4	6	13-10	- 5	16- 4	-23
8-10	21	11- 6	5	14- 0	- 6	16- 6	-24
9- 0	19	11- 8	4	14- 2	- 7	16- 8	-26
9- 2	18	11-10	3	14- 4	- 7	16-10	-28
9- 4	17	12- 0	3	14- 6	- 8	17- 0	-31
9- 6	16	12- 2	2	14- 8	- 9	17- 2	-33
9- 8	14	12- 4	1	14-10	-11	17- 4	-35
9-10	13	12- 6	0	15- 0	-12	17- 6	-37
10- 0	12						

個月既等於B校正數正三四，九十一個月就等於三三·一二五（或爲便利計，作爲三三·三）。九十一個月以上的月數，照同法計算。算到一百〇二個月時，須再用二三·五——一五·五推算。上面表七，就用這簡單的插入法算出。七歲以下和十七歲以上的B校正數，可用估計法推出。有了這種表，就可以看學生的實足年月多少，再從他的T分數中加上或減去相當的B校正數，而算出他的B分數來。

(三) C量表

T量表，測驗總能力，有絕對的意義。B量表，測驗聰明程度，就是和年齡比較的能力。C量表，是要自動的就學生某種特性（如書法能力，算術能力等）指出他應歸入校中何級，並測驗他和年級比較的能力。一個學生的工作，就他的年齡論，可以很好，但是就年級論，卻不好。反之，就年級論，一個學生的工作可以說是好的，而就他的年齡，卻是不好。C量表就是從年級上定學生的適當地位。現在把C



量表的造法，略述於下。

(1) 照表五年齡分配，造一個年級分配表。

(2) 從T分數(即量表分數)行和各年級的次數(即人數)行，計算每級的平均T分數。倘使測驗的學校，是規定半年一升級(half-year promotion)，則須計算每半年級(half-grade)的平均T分數。這求得的各級平均T分數，就是年級常模。茲假擬各年級常模如下：

年級	2A	2B	3A	3B	4A	4B	5A	5B	6A	6B	7A	7B
常模	26	30	33.7	37.3	39.6	41.8	44.9	48.0	50.9	53.7	56.0	58.3
年級	8A	8B	9A	9B	10A	10B	11A	11B	12A	12B		
常模	59.6	60.9	61.5	62.1	62.9	63.6	64.5	65.4	66.8	68.1		

(3) 假定測驗是在陽歷六月施行的。測驗既然是在六月施行，2A 剛完結二年級的一半，2B 已達二年級的全部，其餘各級類推。這樣，用數字來表示則 2A 應寫做 2.5，2B 應寫做 2.99 或 3.0，3A 寫做 3.5，3B 寫做 4.0，4A 寫做 4.5，4B 寫做 5.0，餘類推。倘使施行測驗適當半年升級 (mid-year promotion) 之後，則 2A 應寫做 2.0，2B 應寫做 2.5。餘類推。

(4) 把一個年級作十個學月算，用插入法來定一個年級中每十分之一的常模。因為年級 2.5 相當於常模 26，年級 3.0 相當於常模 30，用插入法，則年級 2.6 (即二年級的六學月) 得常模 26.8，2.7 得常模 27.6。餘類推。表八 (假定是算術測驗， T_a 是算術之 T 分數， G_a 是算術年級) 就是用這種插入法推算的結果。在這表中，用 G (年級地位) 代表年級，用 T 來代表常模，因為常模就是各級的平均 T 分數。並且照常識估計，把表的下部推廣，表的上部亦已隨意推廣，使最高可能的分數和 200

相當。

現代教科書
教育測驗綱要

表 八

顯示 T 分數與年級地位之對照

Ta	Ga	Ta	Ga	Ta	Ga	Ta	Ga	Ta	Ga	Ta	Ga
22.8	2.1	42.4	5.1	58.6	8.1	63.8	11.1	72.5	14.1	84.5	17.1
23.6	2.2	43.0	5.2	58.9	8.2	64.0	11.2	72.9	14.2	84.9	17.2
24.4	2.3	43.6	5.3	59.2	8.3	64.2	11.3	73.3	14.3	85.3	17.3
25.2	2.4	44.2	5.4	59.5	8.4	64.4	11.4	73.7	14.4	85.7	17.4
26.0	2.5	44.9	5.5	59.6	8.5	64.5	11.5	74.1	14.5	86.1	17.5
26.8	2.6	45.5	5.6	59.9	8.6	64.7	11.6	74.5	14.6	86.5	17.6
27.6	2.7	46.1	5.7	60.2	8.7	64.9	11.7	74.9	14.7	86.9	17.7
28.4	2.8	46.7	5.8	60.5	8.8	65.1	11.8	75.3	14.8	87.3	17.8
29.2	2.9	47.3	5.9	60.8	8.9	65.3	11.9	75.7	14.9	87.7	17.9
30.0	3.0	48.0	6.0	60.9	9.0	65.4	12.0	76.1	15.0	88.1	18.0
30.7	3.1	48.6	6.1	61.0	9.1	65.7	12.1	76.5	15.1	88.5	18.1
31.4	3.2	49.2	6.2	61.1	9.2	66.0	12.2	76.9	15.2	88.9	18.2
32.1	3.3	49.8	6.3	61.2	9.3	66.3	12.3	77.3	15.3	89.3	18.3
32.8	3.4	50.4	6.4	61.3	9.4	66.6	12.4	77.7	15.4	89.7	18.4
33.7	3.5	50.9	6.5	61.5	9.5	66.8	12.5	78.1	15.5	90.1	18.5
34.4	3.6	51.5	6.6	61.6	9.6	67.1	12.6	78.5	15.6	90.5	18.6
35.1	3.7	52.1	6.7	61.7	9.7	67.4	12.7	78.9	15.7	90.9	18.7
35.8	3.8	52.7	6.8	61.8	9.8	67.7	12.8	79.3	15.8	91.3	18.8
36.5	3.9	53.3	6.9	61.9	9.9	68.0	12.9	79.7	15.9	91.7	18.9
37.3	4.0	53.7	7.0	62.1	10.0	68.1	13.0	80.1	16.0	92.1	19.0
37.8	4.1	54.2	7.1	62.3	10.1	68.5	13.1	80.5	16.1	92.5	19.1
38.3	4.2	54.7	7.2	62.5	10.2	68.9	13.2	80.9	16.2	92.9	19.2
38.8	4.3	55.2	7.3	62.7	10.3	69.3	13.3	81.3	16.3	93.3	19.3
39.3	4.4	55.7	7.4	62.8	10.4	69.7	13.4	81.7	16.4	93.7	19.4
39.6	4.5	56.0	7.5	62.9	10.5	70.1	13.5	82.1	16.5	94.1	19.5
40.0	4.6	56.5	7.6	63.0	10.6	70.5	13.6	82.5	16.6	94.5	19.6
40.4	4.7	57.0	7.7	63.1	10.7	70.9	13.7	82.9	16.7	94.9	19.7
40.8	4.8	57.5	7.8	63.2	10.8	71.3	13.8	83.3	16.8	95.3	19.8
41.2	4.9	58.0	7.9	63.4	10.9	71.7	13.9	83.7	16.9	95.7	19.9
41.8	5.0	58.3	8.0	63.6	11.0	72.1	14.0	84.1	17.0	96.0	20.0

(5) 做一個 O 校正數表，使得 G 轉變成 C。C 的校正數見下。這些校正數，對於編造小學或中學用的測驗都能適用，就是編造測驗時期雖不同，亦得適用這校正數。

陽曆月終	1	2	3	4	5	6	7	8	9	10
O 校正數	.4	.3	.2	.1	0	-.1	-.2	-.3	-.4	-.5

參考書：

McCall, W. A.: A Proposed Method of Scale Construction, in Teachers College Record, Vol. XXII, pp. 31-35.

McCall, W. A.: How to Experiment in Education, Chap. V, pp. 98-126.

問題和討論：

(1) 試述 T 量表之優點。

- (2) 以十六歲或十二歲或八歲學生的T作標準，此種簡易的編造法，有無缺點，並如何補救？
- (3) 倘作各年齡的T量表，於事實上辦得到否？於勞力與時間上經濟否？
- (4) 每一年齡有五百人代表，此五百人足夠訂定測驗標準否？
- (5) 測驗問題何以必分為難易等級？理想的測驗問題最易者百分之九九人能答正，最難者百分之一人能答正，事實上能達此目的否？
- (6) 所謂不加選擇的兒童，是何意義？
- (7) 問題既照難易排列（先易後難），倘一學生做對 1 4 7 8 9 13 14 17 各問題，又一學生做對 1 2 3 4 8 9 12 15 各問題，同得八分，果公平否？
- (8) 學校中採用測驗單位（如T）記分，比用普通記分法，優點何在？
- (9) 試選二十個問題，應用於四級學生，編個T量表。

第七章 教育測驗的實施法

關於教育測驗的施行方法，各種測驗說明書中，都已規定，不必多說。此地所要說的，是施行前如何準備，施行後如何利用。分兩項，敘述於下。

(一) 施行前的準備

標準測驗的選擇 怎樣選擇測驗，是施行測驗前一個先決問題。學校各級學生能力參差不一，這是大家都知道的，但是一級裏，就全體而論，究竟是普通能力，是優秀能力還是低等能力，教員不能準確知道。優秀的，低等的，普通的究竟各有幾個，教員更不能準確知道。教員倘使要因材施教，使各個學生各盡所能，那麼不得不選擇智力測驗。因為智力測驗的結果，一級智力率的平均數的高下，就可定這級學生能力的高下。決定智力測驗後又須進一層考量，究竟是要做精密的

診斷，還是普通的診斷。倘使要精密一點，就須選個人智力測驗；否則，團體智力測驗亦行。講到教育測驗，亦須決定你校中那幾級須要是那幾種教育測驗。據編者個人看來，凡一級中年齡和班級不相稱（年齡或太大或太小）的學生愈多，須施行個人智力測驗，以分別各人能力的高下。一級中學業進步遲緩的時候，智力測驗和教育測驗宜同時並用，以診斷不進步的原因是由於學生的不努力還是由於學生天賦能力的不好。現在把已出版的標準測驗列爲一表以供選擇。

名稱	類數	適用範圍	須要時間
小學算術應用題測驗	四類	三年級至七年級	三十分鐘
小學常識測驗	二類	初小三年級至高級二年級	三十分鐘
小學默讀測驗	五類	同上	二十五分鐘
初小默讀測驗	二類	一年級至四年級	十八分鐘
小學默字測驗	二類	初小三年級至高小二年級	約五十分鐘

算術四則測驗

二類

小學三年起至初中二三年級

十二分鐘

教育測驗

一類

小學三年起至初中止

六十分鐘

小學算術混合四則測驗

二類

三年級至初中一年級

二十分鐘

算術練習測驗

二組

小學用

無定

小學社會自然測驗

二類

無定

比納西蒙智力測驗

一類

三歲至十七歲

無定

機械智力測驗

一類

五級至十二級

無定

調查用非文字智力測驗

二類

小學三年級至高中三年級

三十分鐘

團體智力測驗

二類

國民三年至初中二年

量表甲第一類二十三分
量表乙第一類二十分

英文測驗

一類

高中三年以下習過英文二年
以上者

約一小時又十一分鐘

中學默讀測驗

二類

中學用

二十五分鐘

國語教科書閱讀測驗

二類

初小一年級至四年級

約二十分鐘

主試者的訓練——以上所說須要時間，是指學生對答問題的時限，其餘分

發試卷主試者的解釋等等尙不在內。至於主試者的態度等等影響於被試者成績很大。所以主試者，事前須有相當的訓練。大概，施行教育測驗，主試者於事前做過同類測驗四五次的實地練習，把說明書中應說的話，詳細學過後，已不至發生甚麼重大弊病。而施行個人智力測驗，主試者非有充分的經驗，往往不能得準確的結果，所以要特別的加上練習。而教育測驗中，如國文、書法、圖畫等測驗，量表的準確使用，亦非事前有充分的訓練不辦。

測驗委員會 施行測驗於一個學校之中，範圍還小，可由該學校的教師自己主試。倘作大規模的測驗，如一鄉或一學區的測驗則須另組委員會主持。爲使測驗時情形彼此一致，使結果可以互相比較起見，測驗委員會人數愈少愈好，並且要選對於測驗富有經驗的人充當。倘能由委員長一人擔任，別的委員，幫同助理一切測驗事務的進行，那更好了。視學員或教育委員而於測驗有經驗的，當然任

委員長，以便易於和各地教師接洽——教師的同情，是施行測驗上最重要的事。而當施行個人智力測驗，主試者最好能說本地話，以免言語隔閡，結果不準確。

(II) 施行後的利用

我們對於標準測驗，應當看做是學校裏有價值工作的一部分，不應當看做無足重輕或敷衍門面的事情。所以學校裏經費無論如何支絀，每年應當規定一筆測驗費——少則數元，多則數十元。但是已經施行之後，必須利用由測驗得來的消息——學生的需要，弱點等等——始所費的金錢，勞力和時間纔得到相當的報酬，纔合於經濟。

分數的解釋 標準測驗所產生的分數，必須就怎樣利用該測驗的方法解釋。利用測驗的方法很多，現在不過舉示最普通而又屬初步的手續。通常分數算好之後，須將分數彙集起來，做成一分配表，這是解釋分數的第一步。從這分配表，

計算一班或一級的分數。這一級或一班的分數，通常用平均數（中數亦可）代表。這平均數不但是解釋班級的成績有用，就是對於解釋個人的分數亦有用。

錯誤的範圍 標準測驗所得的分數，雖比通常考試所得分數準確，但亦不能免錯誤——變錯（variable error）和恆錯（constant error）。測驗的變錯，在一級或一團體的平均數中，往往互相消除（cancel each other）即言各人分數對於真正分數差異有大有小，平均後，影響微小。故平均數較各個的分數可靠些。就二十五人一級的平均數而說，這一級平均數的變錯大約相當各個人分數錯誤的五分之一。倘有一百人的一級，測驗所得的平均數的變錯，只有各個人分數的錯誤之十分之一（如是人數愈多，平均數愈可靠）。有些測驗，編造時的用意，原來是在獲得一級或一團體的平均數量（如華狄算術測驗“Woody's Arithmetic Scales”）。倘使用這種測驗來評量各個學生的能力，則測驗的變錯往往比

較的很大。老實說，心理能力的測驗，變錯比物理的測驗大得多。要準確知道測驗的錯誤，究竟有多少，真不容易。我們只能指出分數中的百分之幾（通常百分之五十）是超過某限度，百分之幾，在某限度以下。倘就個別的分數（各個學生的分數）而論，我們完全不能斷定錯誤多少。所以解釋個人的分數時，應當特別留意。

解釋分數時，還須注意測驗的恆錯。變更規定的測驗條件，或使所有學生同受某原因的影響，都可以產生測驗的恆錯。測驗所定的標準，是由規定的測驗條件得來的，倘使任意變更條件，或導入別種動因，使學生分數增多或減少，則分數的解釋便不正當。

圖表的利用 把分數和標準比較時，最好用圖表。要把測驗結果表白於大多數人的，更宜利用圖表。圖表的編製和怎樣利用，可參考孟羅教授結果之測驗第二章及第九章，或亞歷山大學校統計及公開辦法第十一章，或麥柯爾教育測

驗法第十三章（第二、三書有譯本）。

補救的手續 測驗所得的分數最好由學生的需要方面解釋，以便採行補救手續。例如，倘使就教育指導以解釋一個學生的分數，則解釋時應當由這個學生宜習或不宜習某種學校科目着眼。這不是僅僅說明學生成績在測驗標準上或下或剛達標準，所能決定的。解釋者當這時候，不但要能領會所行測驗的性質，並且要於學校的手續各方面都有充分的經驗。進而論之，一個解釋者，除非於教授方法上有充分的研究，不能由補救的教授上解釋學生分數。解釋者亦不能由升級或分級上解釋分數，除非他於學校這一方面的組織富有經驗。

分數的解釋和計畫補救教授，可以說是標準測驗使用中，最難的一部分。但是，實際上，是最重要的一部分。沒有這一部分，教育測驗是沒甚價值。所以一學區中行過測驗之後，最好招集和測驗有關係的教員討論怎樣解釋分數。老實說，教

員在這方面比在施行測驗方面更須有經驗的測驗專家的扶助。

就年級標準診斷 通常診斷學生成績，多把學生分數和年級標準比較。並看分數分配的形式是怎樣。倘使一級的平均數在年級標準之下，就可說這級須要增加教授。或者，這一級全體須要較多的訓練，或者須要別種訓練。有時候，便可說某種能力不達標準，是由於不充分注重的緣故。倘使一級的平均分數在測驗標準之上，或各個學生分數的分配形式太散漫，則可就學生於教授上的需要，另作解釋（參考孟羅教授結果之測驗第五章算術診斷與補救教授）個人分數的解釋，類推。

就成業率診斷 上述的診斷是有價值的診斷，教員能實地做去，於教授的效率定能增加。不過，這種診斷忽視學生的學習能力。照上述的方法診斷，凡屬於同級的學生，都把他們和這級的標準比較。實際上，學生雖同屬一級而學習能力

各有差別。所以同級二個學生可以得相同的分數，但是學習能力不同，各人須要的補救教授亦不同。例如，一個聰明的學生和一個通常的學生於算術測驗中所得的分數相同，並且兩個人所得的分數都剛達標準。這裏頭，通常的學生已盡他的能力（已達我們能希望於他的程度），但是聰明的學生卻沒有盡他的能力。他的分數應當要多一點，因為他的學習能力高。倘使要使他的成績和他的智力程度相稱，應當另外想出補救教授。

要正當解釋分數，並對各學生的智力或學習能力加以相當的考慮，所以試行教育測驗後，須另行施行智力測驗。把教育年齡用智力年齡除，得一成業率。成業率一百，表示學生的成績和他的智力相稱，低於一百（例如百分之八十）則表示他的成績低於他的智力年齡所應達的，高於一百，則表示他已超過智力年齡所應達的。

最好，求出每一種智力年齡，於某種教育測驗中應得的分數，然後列爲一表，則解釋分數時更便利。

普通診斷和精密診斷 一種學校科目中的成績，是很複雜的，包含許多能力的，而這許多能力又比較的不相關聯 (relatively unrelated)。所謂學生成績的普通診斷，是不過指出那沒有成就我們所定普通目標的學生。倘進而討究一個學生的特殊短處或弱點，那就要有精密的診斷了。這精密診斷，看我們編造的成績測驗的精密程度而異。普通和精密兩種診斷是互相補充的，可以把算術測驗做解釋的例子。

用普通算術四則測驗（於加減乘除中各選若干問題，編爲測驗）可以說是含有診斷作用，因爲他能檢別那於算術四則中普通不及格或不能成就的學生。例如，克的斯算術測驗乙組可以說有診斷性質，因爲他能指出學生在整數四則

中各別的成績。但是，他不能表示一個學生的特殊短處；僅能指出這個學生在四則問題中的普通能力的缺陷程度或範圍。

要斷定一個學生究竟不能做對的是那種特殊範類的問題，必須使用一組精密的測驗，例如孟羅算術診斷測驗 (Monroe's Diagnostic Tests in Arithmetic) 或克萊佛蘭學務調查用算術測驗，或克的斯算術練習測驗。所以，學生成績的測驗，最好要用能獲得學生成績詳細情形的那種測驗或彙選測驗。自然，教員因為時間與經濟的關係，都不免於這種精密診斷的使用，多少受限制。大概說來，施行測驗，不妨先用普通的測驗，發現了學生的普通缺陷，再用精密的測驗來補充。

分析的診斷 僅看學生的分數，不能知道為什麼他是失敗，或不能成就，亦不能知道他所犯的錯誤是什麼樣的特殊錯誤。所以施行診斷測驗之後，須把學生作業，詳細分析考究為什麼做錯或做錯的是那幾種。這實在是已經超出通常

所謂教育測驗的使用範圍以外。但是這種分析的診斷，是不可少的，對於那成績不良的學生利益更大。這種分析的診斷，可以根據標準測驗中學生的作業，亦可以從學生平時的作業情形着眼（參考孟羅教授結果之測驗第一三八至一五二面）。

補救教授的計畫 補救教授的詳細討論，不在本書範圍以內。此地不過指示應持的方針。大概一個學生的成績不能和他智力年齡相稱（不能達他智力年齡應達之標準）是有多種原因。除身體上的缺陷和別種影響外，重要原因在他已受的教學中。他或者已受的教學不充分，須得增加教學，或者須換別種教學。或者，因為他對於某科目中必須先行具備的知識技能還沒有充分，不能活用。總之，無論什麼原因，教員的任務，在規定補救教授，消除這些種不良原因。

標準測驗，不過把須要補救的學生指出來，並沒有指出學生成績不良的原

因。要確定原因並規定補救方法，須要教員的無窮策畫和他於教學原理上的修養。換句話說，要依靠他的教授經驗和平時於學習過程上的研究。詳細的情形，非此地所能說。

就學習手續診斷 就學習手續診斷學生，和就成績診斷學生，根本上沒有什麼大區別。事實上，有效率的學習能力是成績中許多形式之一。所謂學習手續診斷，主要的任務就在指出那沒有得到有效率的學習方法的學生。此地所要注意的，不是某學生有成就，某學生沒成就，是某學生有正當的學習手續（足以使他於某科中有成就的學習手續），某學生沒有這種正當的學習手續。一個學生可以有好的學習算術的手續，但是同時因為研究太少或並沒研究，他的算術知識就很少。反之，算術知識很多的學生，不一定就是算術學習手續很好的學生。學習方法既是形成成績的要素，有效率的學習法，實是學校中最重要目標，必須

使他實現。所以學習手續的診斷，在成績測驗中是必要。而當小學末期和中學時期，更宜注重這種學習方法的診斷。可是現在出版的測驗專門有這種功能的，還沒有。

設定直接的教育目標 把教育目標普通的敘述出來，還不足以規定學生日常的工作，使他們有個確定的目的。必須規定學生人人能明白的，詳細的，直接的 (immediate) 教育目標。而教育測驗中，所定的常模或標準，是代表應達的能力程度，並且有確定的意義，所以這些常模可以作為直接的教育目標用。

例如，我們假定的克的斯的算術測驗乙組的常模，代表學生在整數四則中應達的能力程度，這常模就可用在整數四則中的教育目標。學生人人可懂得這常模的意義，因為他們知道他們的目的，要在某時限內做若干某類的問題。這某類問題中要若干做對。圖畫，讀法，書法以及其他學校科目中，凡有製定的標準測

驗的，都可以利用規定的常模，作為直接的教育目標。設定這種教育目標，實是創造或引起動機的很有效率的方法。教員倘能把確定的直接的目標定妥，並努力使學生達到這目標，結果，學生的成績常有顯著的進步——不一定要增加教學時間，只要努力訓練學生達到這目標。

參考書

Monroe's *The Theory of Educational Measurements*, Chap. X.

Monroe's *Measuring the Results of Teaching*.

華超譯惟孟氏訂正比納西蒙智力測驗第六、七、八章。

問題和討論

(1) 解釋學生的成績為何須參照智力？

(2) 試於已出版之教育測驗中任選一種測驗，並於智力測驗中任選一種。施行後，就年級標

準及成業率解釋學生分數。

(3) 施行上述測驗後，試對補救教授如何進行，略敘述之。

(4) 某級於學期開始時，某種測驗結果平均得五十二分。迨學期終了，再行某種測驗，平均得五十四分。此四分數是否為該級真正之進步？

(5) 施行教育測驗與個人智力測驗孰難？試述所以然。

(6) 試將俞氏小學混合四則測驗與德氏算術四則測驗施行於某一級。比較該二種測驗所得之平均分數。並述何種測驗較適於補救教授。

(7) 試作一學區有初小學生二千人之測驗計畫。再作一個二萬學生之測驗計畫。進行的手續或步驟務求詳細。

(8) 何謂診斷？有幾種？

(9) 測驗之錯誤與解釋之錯誤，區別何在？

(10) 施行測驗而並無特定的目的，有益否？

附錄中西參考書報

(I) 西文

關於理論方面著

W. A. McCall: How to Measure in Education.

W. S. Monroe: The Theory of Educational Measurements.

關於測驗之應用與編造著

W. S. Monroe: Measuring the Results of Teaching.

G. M. Wilson and K. J. Hoke: How to Measure.

W. A. McCall: How to Experiment in Education.

關於測驗統計著

H. O. Rugg: Statistical Methods Applied to Education.

- E. L. Thorndike: Theory of Mental and Social Measurement.
G. Udny Yule: An Introduction to the Theory of Statistics.
Willard C. Brinton: Graphic Methods for Presenting Facts.
Carter Alexander: School Statistics and Publicity (此書已由朱斌魁譯)。

物理學

- S. A. Courtis: The Gary Public Schools, Measurements of Classroom Products.
Yakum and Yerkes: Army Mental Tests.
G. M. Whipple: Classes for Gifted Children.
National Society for the Study of Education, Year Books.
Journal of Educational Psychology (Warwick & York, Baltimore).
Journal of Educational Research (Public School Publishing Company, Bloomington, Illinois)

Robert R. Rusk: *Experimental Education* (此已由陶孟和譯)。

Frank N. Freeman: *Experimental Education*.

劉湛恩——*Non-Verbal Intelligence Test for Use in China*.

(II) 中文

陳鶴琴廖世承合著——*智力測驗法*。

陸志韋著——*高中心理學第三六至三七課*。

朱斌魁著——*測驗與統計名詞漢譯*。

廖世承著——*中學教育第十二又十三章*。

俞子夷編——*測驗統計法概要*。

費培傑譯——*兒童心智發達測量法*。

鄒恩潤譯——*職業智能測驗法*。

戴應觀譯——*教育心理的實驗*。

華超譯——推孟氏訂正比納西蒙智力測驗。

薛鴻志編——教育統計法大綱。

中等教育第一卷第二期教育測驗號。

中華教育界第十一卷測驗號。

小學校第十一、十二各號。

俞子夷——問答法選擇法是非法的測驗那一種可靠（初等教育一卷二期）。

俞子夷——行了設計法以後學生讀書能力要低麼？（初等教育一卷三期）。

陳啓天——教育事實的統計分類法（中華教育界第一卷十二期，再參考第十卷第九期第

十一期，十二卷一期及四期）。

標準測驗簡說（新教育七卷四期）。

孟憲承——智力測驗之論爭與教育學說（教育雜誌第十五卷第一期）。

薛鴻志——關係數量之研究（同上，第一期）。

- 廖世承——應用科學原理改良入學考試的方法（教育雜誌第十五卷第十期）。
- 錢希乃——麥柯爾測驗編製的T B C F制（教育雜誌第十五卷第九期）。
- 陳鶴琴——中小學默讀測驗編造順序（教育雜誌十六卷五號）。
- 趙崇華——教育心理測驗之實施法（同上，第十五卷第十二期）。
- 俞子夷——中數計算法說明（同上，第十五卷第二期）。
- 薛鴻志——計算中數之法則。
- 華超——全國小學調查測驗之我見（同上，第十卷第九期）。
- 杜佐周——國文教育問題及兒童讀書能力測驗（同上，第十五卷第七期）。
- 華超——優秀兒之特殊班（初等教育二卷三期）。
- 朱光潛——智力測驗法之標準（教育雜誌十四卷五期）。
- 劉建陽——教育測驗（同上，第五期）。
- 解中蓀——測驗之種類（同上，第一期）。

廖世承——五項測驗報告（同上，第二，三期。）

郝耀東——美國陸軍適用的智力測驗（同上，第五期。）

錢鶴——日本久保氏之皮奈西門智力測驗改訂法（同上，第十又十二期。）

羅廷光——小學算術學力測驗法（教育雜誌第十三卷第三期。）

太玄——亞克斯布利基士分數式測定法的解釋（同上，第五期。）

李勉韶——學業成績考查法（同上，第五期。）

趙欲仁——常識測驗報告（同上，第八期。）

趙欲仁——測驗式的學業成績考查法（同上，第八期。）

智力試驗之報告（教育雜誌第十三卷第二號。）

慈心——教育心理學之最近傾向（同上，第三號。）

俞子夷等——崑山初小算術測驗編造法（教育雜誌第十六卷第四號。）

薛鴻志——T B C 量尺之編造法（同上，十六卷六號。）

- 陳鶴琴——初小默讀測驗編造順序（新教育八卷第三期）
- 陸志章——編造測驗自檢雜問（教育彙刊二卷一期）
- 陳鶴琴——矯正機遇錯誤的公式（教育彙刊二卷二期）
- 吳定良——美國團體智力測驗之種類（心理第二卷一號）

新教育需要的要品

中學 小學 應用 **測驗**

中華教育改進社編製

測驗的功用，在以客觀的標準，辨別兒童的智愚庸能，及其興趣之所在，而不致失實。凡學校舉行入學試驗，以及分別班次甄別成績，都應採用。已出下列各種：

001104

商務印書館發行

每類包	含測驗	二十五	份外答	案及成	績表	又各編	說明書	一冊每	冊五分	至二角
-----	-----	-----	-----	-----	----	-----	-----	-----	-----	-----

小學默讀測驗	五類各四角
初小默讀測驗	二類各五角
中學默讀測驗	二類各四角
小學默字測驗	二類各一角
小學常識測驗	二類各四角
本國史測驗	一類 六角
算術四則測驗	二類各二角五分
小學算術應用題測驗	二類各五分
小學算術混合測驗	二類各五分
英文測驗	一類 五分
團體智力測驗	二類各五分
機械的智力測驗	一類 六角
調查非文字的智力測驗	一類 五角
教育測驗	一類 五角
全國小學調查測驗	一類 壹角

實 際 的 教 育 狀 况

英文 **中國全國小學校概况** 一册 一元五角

本書為中華教育改進社舉行全國小學教育測驗後之報告，用科學的方法，指示全國小學教育的能力，實為教育界中罕有之作。

中國教育一瞥錄 一册 一元二角

王卓然編 本書係編者隨同美國孟祿博士調查我國各地教育時之見聞實錄，除記述各地學校調查與風俗習慣外，并詳載孟祿在各地發表之意見。

中國教育統計概覽 一册 二角

本書對於各省區高等中等初等之教育，均有精密之統計表，統計之事項，有教職員數、男女學生數、歲出額、各生平均費等。

(以上均屬中華教育改進社叢書)

英文 **中國最近教育狀況** 一册 一元八角

凡關於職業教育、師範教育、女子教育、高等教育、中等教育、初等教育、體育狀況、文化運動、新學制問題、課程研究等，均將其最近狀況，詳細記載，此書亦由中華教育改進社編輯。

江蘇教育行政概况 一册 三角

蔣維喬編 本書分六章，將江蘇全省教育行政機關之變遷、教育經費之籌措，以及中小義務實業專門與社會各種教育之狀況，詳述無遺。

商 務 印 書 館 出 版

簡 易 師 範 用 的
現 代 師 範 教 科 書

新學制系統改革案，爲補充初級小學教員之不足，得酌設相當時期之師範學校或師範講習科，在義務教育推行期間，此項簡易師範科，實爲必要。本館特編輯現代師範教科書一套，專供簡易師範之用，書名如下：

- | | | | | | | |
|------|-------|-------|------|-------|-------|-------|
| 現代師範 | 現代師範 | 現代師範 | 現代師範 | 現代師範 | 現代師範 | 現代師範 |
| 倫理學 | 教育心理學 | 哲學綱要 | 小學行政 | 學校管理法 | 各科教學法 | 教育學原理 |
| 一册二角 | 一册三角 | 一册四角半 | 一册五角 | 一册五角半 | 一册四角 | 一册四角 |

元(1749)

Modern Normal School Series
Outline of Educational Measurement
The Commercial Press, Limited
All rights reserved

中華民國十四年一月初版

* (現代師範教科書) **教育測驗綱要(二册)**

(每册定價大洋肆角)

(外埠酌加運費匯費)

編纂者 華 超

發行者 商務印書館

印刷所 上海北河南路北首寶山路 商務印書館

總發行所 上海棋盤街中市 商務印書館

分售處 商務印書分館

北京 天津 保定 奉天 吉林 龍江
濟南 太原 開封 西安 南京 杭州
蘭谿 安慶 蕪湖 南昌 漢口 長沙
常德 衡州 成都 重慶 廈門 福州
廣州 潮州 香港 梧州 雲南 貴陽
新加坡 張家口

此書有著作權翻印必究

