

萬 有 文 庫

第一集一千種

王 雲 五 主 編

教 育 測 驗 概 要

華 超 著

商 務 印 書 館 發 行



教育測驗概要

華超著

師範小叢書

編主五雲王

種千一集一第

要概驗測育教

著超華

號一〇五路山寶海上	人	行	發
五雲王			
路山寶海上	所	刷	印
館書印務商			
埠各及海上	所	行	發
館書印務商			

版初月四年十二國民華中

究必印翻權作著有書此

The Complete Library

Edited by

Y. W. WONG

OUTLINE OF EDUCATIONAL MEASUREMENT

BY HUA CH'AO

PUBLISHED BY Y. W. WONG

THE COMMERCIAL PRESS, LTD.

Shanghai, China

1931

All Rights Reserved

自序

這本小冊子是根據麥何爾博士的教育測驗 (How to Measure Education) 的上半部和參考孟羅的教育測驗理論編譯而成的。目的在介紹教育測驗的理論和應用方面。講到測驗的怎樣編造及種種統計的方法都是見教育測驗的下半部，本書並沒有把他介紹。我所以敢於這樣「斷章取義」有兩種原因，第一因為原著的性質是可以這樣分開，第二是因為教育測驗介紹到本國來雖然有好幾年的歷史，但是他的理論和實用方面還沒有得到一般從事於教育的人的深切瞭解。所以想於這方面多介紹些文字，希望多數人明了測驗的真面目以後可以盡量的應用他。

測驗的高深學理——無論是教育測驗或智力測驗或特殊能力（或傾向）的測驗——在現在都還是一種假定，和物理化學方面的最高假定相同，沒有一種實驗可以證明。所以在測驗這門學科最發達的國家如美國，有權威的教育家還時有反對論調發生。在我們教育幼稚的國家裏頭要想推行這種新式工具，自然是更不容易。我這本小冊子倘能對於內地的教育界多少激起一

點研究的興趣，那末我所化於編譯這本小冊子的勞力就可算獲得了相當的酬報。

一九，一二，五，華 超識於上海。

教育測驗概要

目次

第一章	測驗在教育上之地位	一
第二章	測驗和分別班級	一五
第三章	測驗在教育診斷上之功用	三九
第四章	測驗在教學上之功用	七五
第五章	測驗和教學效率的估量	一〇二

教育測驗概要

第一章 測驗在教育上之地位

理論一 凡存在的東西都是有量的

學生成績的測驗，內容很是複雜；但是也有他的基本的意義。倘不把這基本意義認清楚，便容易發生兩種教育的「不幸事件」：一種是過度的熱心於教育之科學的測驗，一種是過度的反對這種測驗運動。現在把這種基本意義歸納為若干種理論，以便一般沒有充分時間來研究測驗的人可以藉此認識測驗在教育上之真正地位。

第一種理論，就是凡存在的東西都是有量的。這個理論，無須加以多大說明，因為有健全頭腦的人都是承認的。但是為明顯起見，不妨再說一說。譬如一個教師教了學生，學生所起的變化一定是有量的。做教師的極少願意承認自己的努力一點沒有產生量的變化。若果他承認是這樣，則徒

然證明他對學生的教學是沒有效率，而我們所持的理論卻仍舊可以成立。若說教育的努力，在性質之改善，這話是有一部分的真理，但是要知道沒有一種性質是屬「非量的」[non-quantitative]。

理論二 凡是有量的都能够加以測驗

這個理論，也無須多大的解釋。譬如一個人對於日落時景象的愛好，對於音樂的鑑賞，是多少可以測驗的。又如一個人的創發力，比較價值的評判，領袖才能，以及對於文學的鑑賞等等，——這都是教育努力的目標——也是可以測驗的。我們目前的問題不是這個理論之能成立和不能成立的問題，乃是在這個理論下我們的測驗能不能比較以前主觀的評量更為準確，更為可靠。好在科學一天進步一天，測驗中所使用的工具，所謂 *Test* 者，各方面已證實比主觀的評價準確得多了。

理論三 通常說來教育上的測驗是和物理科學上的測驗相同的

這兩種測驗，基本上是相類似的，因為兩者都是測驗物理的表現（physical manifestation）。加減能力的高下和動機的好壞，都不是把寒暑表插入學生的精神媒介體中所可測驗的，但是

我們可以由他的外表的行爲和他的內在的情形，加以測驗。現在心理學家可以由呼吸曲線 (breathing curve) 來斷定證人說的話可靠不可靠，除非這個證人是慣於說慌的。

在最遠的將來，我們對於學生的心理活動的機制或者可以攝成影片，但是就現在而論，只得由心理活動所產生的結果來測驗；遇到必要的時候，再由結果來推想學生的內在情形。

理論四 凡物理的科學中的測驗都不是完全的

物理的測驗，通常是比教育的測驗較為準確。物理科學中所用的測驗工具，儘有是現在教育測驗專家所望塵莫及的。但是這種優秀點，並不是由於兩類測驗性質上根本不同而起。物理測驗實際上往往也免不掉教育測驗中所常犯的錯誤。不過他們的錯誤，數量極小，致通常人不能辨別出來罷了。不但物理科學是這樣，一切科學中都有錯誤，其中竟有爲專家所不能超脫的。桑代客 (E. L. Thorndike) 曾說：①

『這些教育結果的測驗和物理的測驗（如大小、密度、速度、溫度以及電量等等）的互相對比，毫不出奇。因爲溫度的零點的確定不過幾年以前的事，而溫度表的單位的定奪也有賴於多種

內在的，極精微的假定。我敢擅斷說一句，這種假定任何人都不能很清楚，的很精細的說個明白。在現在，我們的教育結果的測驗，固不能作為讀法教學法或算術教學法等等的取舍標準，但是其中是有許多已比主觀的評量如法庭判斷某一商標是侵害他一商標權的評量優勝得多了。」

○見美國全國教育研究會年報第十七卷，第二編，第十六面。

但是教育測驗中的缺點，通常說來總比物理學，化學以及其相類似的科學中的缺點是顯著些。有些感情用事的人極端相信標準測驗的完全，是不合於真理的。標準測驗的功用是有許多確定的限度。所以要發起測驗的宣傳運動，與其說是因為測驗的十分完美，不如說是有鑒於現行各種測驗種類不多或尚有缺點須加改良。

教育測驗通常遇到一種反對，就是說僅僅測驗了學生的全部能力的一部分。就這點而論，物理的測驗更不如教育測驗。大多數物理的測驗如長度，廣度，重量和溫度等等的測驗都是極狹隘的，但是不失其為有用。對學生全部能力的測驗固然一定是有用的，但是倘使我們有了這種全部能力的測驗我們也必然進而為能力之分析的測驗。譬如作文測驗是有用的，但是分析作文的要

素的量表也是一定有用的。做教師不但從「全體」上教授學生，他也從小處着想。倘使測驗是能夠扶助教學的，則全部能力測驗和分析的（或逐步的）測驗是同等有用。

理論五 測驗是科學的教育之發展上所不可少的

各種自然科學所以有極迅速的進步，就是由於有準確的測驗。有人說過，肥皂之使用量是一國文化之指數。測驗之準確度，也可以說是某一種科學之指數。試問沒有密達，沒有格蘭姆，沒有安丕（電流之單位），沒有弗打（電動力之單位），沒有歐穆（電阻力之單位），沒有瓦特（力之單位），七四六瓦特為一馬力，科學還成個什麼樣子。教育之被擴於科學範圍以外，固有種種原因，缺乏相當測驗實是其中原因之最最大的一個。我們要把教育發展成爲一種科學，是非謀測驗工具之進步，之準確不可。

列却特 (Richards) 在美國科學改進社 (American Association for the Advancement of Science) 的演說辭中，曾有下列一段話：『好多年以前，柏拉圖已經承認沒有權衡和尺度無論何種藝術便空無所有。近代科學所以着重於量的張本就是這種注重權衡的返響；一種科

學於量的上面格外準確，就變爲更科學的。」

果真，測驗之和教育如同一對學生的孩子，這對學生子的頭髮已被母親編結在一起，他們無論那一個不能單獨行動只能有共動的作爲。

有些人以爲教育苟發展而爲一種科學，將侵害生活中的情緒及優美的知覺。這種見解正如恐怕提倡了婦女參政權或工業中雇用了女子，將毀損男子的勇敢一樣可笑。要知道這些優良的性質在人性中已根深蒂固，決不是科學所能破壞。科學之基本的功能在幫助我們以最簡捷最經濟的方法到達我們的目標，不論我們的目標是物質的或是精神的。科學之所以每被人看做是唯物的，主要原因就因爲只有那探求物質目標的人已善意的去獲得了科學的幫助。幾百年以來，教育上的變化真是千端萬緒。教育上有許多大思想家從柏拉圖、裴斯泰勞齊、福祿培爾以至現在各國的教育名家。舊秩序去了，新秩序起而來替代。但是似乎沒有人知道究竟舊的好還是新的好。就事實而論，竟有人懷疑以爲我們是在一個圓圈中行動。這些教育領袖並沒有解答了問題。他們所探討的問題，在別人心目中並沒有發生。他們只提起了預備試驗的問題。各個教育問題之最後的

解決，一定要交託給教育測驗的專家，一定要等到教育發展而成爲一種科學。

○見一九一九年正月三日科學 (Science) 中 The Problem of Radioactive Lead 篇。

理論六 除測驗外教育中尚有幾種不可忽視的動因

鼓吹教育之科學的測驗的人竟承認這個理論，這在旁人看來，大可驚怪。其實不然。教育中至少還有三種重要的動因，就是（1）學生，（2）方法和材料，（3）目標。做教師的必須明瞭學生的心理，訓練學生時所取之適當目標，以及使學生從開始能力 (initial ability) 到達希望的目標時所必須使用的適當的方法和材料。測驗對於這些教學過程，有些什麼關係，試讀下文便知。

理論七 倘使學生的開始能力是不能加以測驗的則我們對於學生是無從瞭解

上文已經說過，做教師的對於學生須有一種最直接的可能的知識，以便應用適當方法和材料去幫助他極迅速的達到所希望的目標。倘使只知道學生所有的能力是什麼，我們只是部分的了解學生罷了。部分的了解，只知道學生有某種能力，並沒有知道某種能力之量有多少。他這種能力是够得上得博士學位麼，還是僅够進幼稚園學習，我們都無從說起。所以我們不但要知道某

種性能的存在，並且要知道這種性能在何等量數以內存在，知道得愈準確愈好。這樣看來，測驗對於心理學之實際知識是必不可少的。

理論八 不可把捉的目標是沒有價值的目標

關於教育目標，我們至少要解答三種事情，第一就是，目標的價值是怎樣？第二就是這個目標的地位是怎樣？第三就是學生的行動是趨向於目標還是和目標背道而馳？這三個絕對重要的問題的解答，都離不了測驗。現在假定有個教學目標，在使學生產生書法能力。這個目標是完全不可捉摸的，除非確定了他的地位。就是要確定了書法能力究竟要多少，書法的速度是怎樣，書法的品質是怎樣，有了一個地位這個目標纔是能把握的目標。更有進者，沒有確定的地位，目標的價值也無從斷定，因為價值隨量之多少而不同。「要多少」這三個字實含有測驗的意義和須要測驗的意義在裏頭了。至於學生的行動是否合於目標這個問題，尤其非測驗不能解答。

理論九 各種教學方法和材料的價值是不知道的除非測驗了他們的效力

方法和材料之主旨固然在幫助學生向着目標進行。但是這個方法和材料是否實踐了這種

目的，我們要施行測驗之後纔可定奪。有時候一種方法我們以為是很有益的，就決定採用，但容或可以是弊端百出。譬如採用了一種方法，這種方法不但不引致學生向目標進行，並且足使繼任的教師採用別種方法時發生困難。好幾種方法和材料集在一起，而要下一個比較價值的評判時，也非使用測驗，決定每一方法每一材料對於學生的效力不可。這就是說教育中對於各種最經濟的教育情景之實驗的選擇都離不掉測驗。

教育中和日常生活中，到處都是測驗。測驗不是新奇的或突如其而來的事情。從早上起來到晚上睡眠的時候，測驗都伴隨我們。指示路程遠近的標誌，錶的指針，廚房中的不值錢的杯子，草地上樹影子的延長，一年四季的遞傳，凡此種種都是表示測驗之無所不在。教育的全過程中有測驗，生活中也無處不是測驗。教育中除測驗以外固然還有別的東西，但是這別的東西和測驗脫離了關係，便是沒有價值的。

理論十 成績測驗應在教育視察之前

初等教育之主要目的在使學生產生某種變化。為達這個目的起見，所以我們有種種工具，如

書本、圖畫、地圖、玩具以及各種教授方法，聯同使用這些工具的教師。這些工具真的能產生我們所希望的變化麼？此地我們有兩種測驗的方法。一種是間接法，就是測驗某種書本之應用，或某種圖書儀器之設備，或某種教授能力（如激起學生動機或使學生有創發性之能力），或某種專業訓練（指教師說）之存在是否於學生的變化為不可少的條件。不可少的條件不知道，則我們就無從斷定不可少的條件多滿足了，自然能產生所希望的變化。一種是直接法，就是直接測驗學生的變化，換句話說就是測驗學生的成績。間接法可以確定教學法，教學工具和學生變化的因果關係，但是迂緩而費時直接法是最經濟又最合實用。所以舉行教育視察之前應當先施行成績測驗。

理論十一 測驗不是教育上新奇的事情

從反對測驗和極力宣傳測驗兩方面看來，測驗好像是絕對新奇的事情。其實不然。教育上的測驗並不是突然而來的，是有他的逐漸的演化。曾有人說過：從事實的立場說，教育測驗是古代的；從過程的立場說，教育測驗是中世紀的；從科學的立場說，教育測驗纔是近世的。這幾句話已把測驗的演化概括無遺了。幾百年以來，中國的考試制度和近來的學校考試都是一種測驗。不過他們

都是些因襲的形式的考試，並沒有準確的標準足以和現在教育測驗比擬罷了。就是論標準測驗的產生也不是一朝一夕的事。以美國而論，從拉哀斯 (Rice) 博士於一八九四年把許多學校中的教學結果做一個比較的測驗之後，方纔產生測驗中的所謂常模 (norms)。桑代客教授利用客太爾，福爾董的等距理論 (Cattell-Fullerton equal-distance theorem) 方纔產生教育成績測驗中的一種量表單位 (scale unit)。科學的教育測驗從此有了一個發端。一九〇八年斯東 (Stone) 氏的算術測驗公表於世；一九〇九年桑代客又發表書法量表。隨後桑代客的學生對於統計技術和教育量表的構造又多所研究。於是希來格斯 (Hillegas) 勃金漢 (Buckingham)，脫拉別 (Trabue) 和華迪 (Woody) 等等又各自造成作文量表，拼法量表，語言量表，和算術四則測驗。一九一一年至一二年紐約市的學務調查全用克的斯 (Couris) 的算術測驗，頗為時人所信從。從此以後，測驗的應用範圍愈加擴大，幾乎凡學務調查都信賴測驗是估計學校效率的主要工具。多數大學設有教育測驗一學程。有幾個大學且特設研究部和各學校共同研究教育測驗。著名的洛克弗樂基金社 (Rockefeller Foundation) 對於教育測驗的運動又極力協助。到了美

國加入歐戰時候，測驗的研究和測驗的貢獻乃大白於世。

我國從中華教育改進社聘了麥柯博士來編造測驗之後，教育測驗也有過一番運動，和一番研究。現在出版的測驗無論是教育的和智力的，爲數也在三十種以上。國內也曾經發生反對和贊成的論調。可惜研究的空氣不如美國來得持久，來得遠大。

理論十二 測驗不會使教育或教育家機械化的

有人以爲測驗是贊成教育中機械的和保守的方法，而不看重教育中激進的方法。老實說來，測驗苟能適當的使用，對於兩種方法都沒有一點偏頗。要定這兩種方法中那一種比較的有效率，最後的不偏不頗的判斷，將全賴測驗。但是，我們對於這兩種方法現在都不能加以評判，因爲目前所有的測驗對於教育的保守主義者和激進主義者雙方重要的目標有幾種還不能加以測驗。概括的說來，目前的標準測驗，苟其使用得不適當，很足以使教育機械化；但是適當的使用了測驗，大可以補救教育之太機械化。這話的成立的理由怎樣，請讀以下幾章的文字，便可瞭然。

至於說測驗將剝奪教師的自由，將使教師機械化，這話更不可信。做教師的應當主宰工具不

應當爲工具所主宰。一個農夫如果自己不能知道他爲他所養的牛工作，還是他所養的牛爲他工作，則這個農夫可稱全無主宰能力。我們不希望有像這種農夫的教師。萬一真有這種教師，測驗也不能加害於他。因爲這種教師將和測驗完全無關。測驗有一定的手續，一定的目標，手續雖繁，目標雖遠，可是怎樣利用，做教師的儘有自由伸縮的餘地。

理論十三 測驗不會破壞學生的個性

測驗不會產生一種死板的「一致」。換句話說，測驗不會破壞學生的個性。有人以爲作文量表，書法量表，或圖畫量表掛在教室中學生的面前，或由教師用來測驗學生的成績，將阻害各學生的個性。這種恐怖心理大可不必。倘使做教師的專教學生依照作文量表上的文字來作文，或專依書法量表上的書法來寫字，則這種恐怖心理未嘗沒有一部分的理由。可是人人都知道作文量表或書法量表是測驗的工具，並不是用以模仿的規範。量表上的樣本所表現的逐漸增進的品質，不過是表示使學生企求的一種目標。就事實而論，這種量表適當使用之後，很能增進或鼓勵學生的個性。因爲把確定的目標擺在學生面前，學生到達目標的興趣便會增進。堅強的興趣，照心理學的

原則說，是各種產品的惟一源泉，也就是各種創發性和個性之最最有效的源泉。做教師的苟能明瞭這一點，則量表的使用，不但沒有害處，並且足以助長學生個性的發展。

討論或練習：

- (1) 你讀了這十三種理論之後，對於教育測驗有何種見解？
- (2) 那一種理論是最有力量，試述其理由。
- (3) 你對於最後一種理論有何意見？
- (4) 什麼是開始能力？

第二章 測驗和分別班級

一 由智力測驗分別班級

分級的基礎和目標 可以用來做分級的基礎的，主要的測驗有三種。這三種主要的測驗就是：智力測驗，教育測驗和教授測驗或教師分數 (pedagogical measurements or teachers' marks)。醫學的診斷，固然也有用為分別班級的基礎的。例如把診斷結果是貧血的兒童，集合為一班或一級。又如年齡的大小，也可以作為分級的基礎。但是我們此地所要討論的只限於上述三種主要的測驗。看學生在測驗中的各科的平均分數怎樣，然後把他插入相當的班級。

分級的第一個基本的目標，在把教育地位同等的 (equal educational status) 學生集合在一起。把同性質的學生集在一起，各學生的需要就十分相近。教師對於他們措施容易適合。教許多學生如同教一個學生。不致因能力不同，時時要改變措施。這樣，做教師的可以減少許多麻煩。同時學生卻可以得到比較滿意的進步。倘若性質不同的學生成爲一集團，則將如弗朗村 (Franzen)

所說：「使低能力的起迷惑，使高能力的生厭倦。」

分級的第二個基本的目標，在把進步速率同等的 (progress at equal rate) 學生集在一起。分級的時候，倘使只顧到學生教育地位的同等，而不注意於學生進步速率的同等，則定期性的重行分級或分組，在事實上為必要。為免去麻煩起見，所以分級時對於教育地位和進步速率須要同時兼顧。要對進步速率的同等上有圓滿的措施，必須於學生的能力外，再考慮到各個學生的興趣、勤勉以及生理學的限制等等。

以上是說分級時適當的目標。至於不適當的目標，最好借用吉特的話：①

「有時候學校當局見於學生的父母是有勢力的，不便反對他們的意見，即便將學生升上一級或一班，雖然這個學生在下一級或一班中還沒有做完他所應做的功課。」

「有時候低級中的學生人數太多，教師為年齡較小的學生留餘地起見，即便將某種學生升上一級，雖然明知這種學生的能力是不能做上一級的功課。」

「有時候某級中的教師對於低能兒或不守秩序的學生覺得麻煩，極願將這種學生的教育

交託別人，而其出路只有升上一級，於是即把這種學生升入上一級去了。

『有時候因為學校規定的學期已滿，放假在即，教師對於升級有全權的，即便把學生的功課宣告完結，不管學生是不是已經做完了他們的功課。

『有時候竟主張低能的學生的身體比較別的同等能力的學生的身體大些，如果坐同樣大小的坐位，很不方便，所以應把這種低能的學生插入備有坐位較大的上一級去。』

○見 Charles H. Judd 所著 教育之科學的研究 (Introduction to the Scientific Study of Education) 一〇九至一一〇面。

心理年齡和學校作業的關係 心理年齡（或稱智齡）和學校作業間的密切的關係，在現在是不再有人懷疑了。任何學生，倘若在學校中做的功課，沒有滿意的進步，第一步的診斷手續就在測度他的心理年齡。教師對於學生的作業所給與的分數，也可以作為診斷學生心理年齡的主要方法，因為教師所給的分數和學生的心理年齡間表示有密切關係。分數高者，心理年齡也高；分數低者，心理年齡也低。但是比較更為可靠的，要推客觀的智力測驗。據推孟的報告，一年級學生的

心理年齡和學校作業的相關度是百分之七十二(72%)。據麥柯爾的報告，^①第六年級的相關度是百分之七十八。據迪克遜 (Dickson) 的研究，^②五班一年級學生的結果，三十三個留級學生中只有兩個是有正常的智力。

^① 參考麥柯爾著心理與教育測驗之相關 (Wm. A. McCall, *Correlation of Some Psychological & Educational Measurements*)

^② 參考推孟著學校兒童之智力 (The Intelligence of School Children) 第六十四面。

實在年齡的大小，身體的大小以及別種標準都不是分級的準則。可以作為分級的準則的，是學生的作業能力。推孟、迪克遜、韋必爾 (Whipple) 和別的許多人的研究，都顯示心理年齡是學生的作業能力的一種極好的指數。據韋必爾的研究，智力測驗比教師的判斷或教師所給的分數，於分級上更為準確。所以他主張用心理年齡來分級。推孟則更進一層，主張對於所有的學生通通加以智力測驗，務必使每個學生所插入的班級都和他的心理年齡相當。

心理年齡和年級地位的關係 就心理年齡說，無論那一級的學生，其中百分之二十五是屬

較高級，還有百分之二十五是屬較低級。統括說起來，越級 (acceleration) 和留級 (retardation) 雙方都是太多。此地所謂越級是指愚笨的學生說，所謂留級是指聰明的學生說。據推孟的研究，幾乎每級學生的心理年齡的分佈範圍都有從八歲到十四歲的；而當我們把學生分級的時候，對於聰明的學生所做的錯誤比對於愚笨的學生更屬重大。換句話說，就心理年齡而論，一年級中可以有一二級程度的學生，但是習慣上的分班，往往把年紀大而智力低的學生插得高些，年紀小而智力高的學生插得低些。現在國內對於學校兒童的實在年齡和年級的統計，以及兩者相互的比較都無從得到可靠的數目，不過倘能用智力測驗來重行分級，則須要改正的地方一定是很多。

用智齡或智力商數來分級和用教育年齡或教育商數來分級，手續上是相類似的，所以前者和後者可以聯帶討論。

二 由教育測驗分別班級

測驗時應注意之點 用教育測驗分班或重新分班，通常要注意下列各點：

第一、使用的測驗要各級一致。有些測驗對於一二年級是用某一種格式，對於三四年級或五六年級又用另外一種格式。除非這兩類測驗所得的分數是彼此可以比較，則往往發生許多困難。

第二、使用的測驗應當產生一種單獨的分數。有些測驗產生速率分數和準確分數。這種測驗固然於診斷上大有用處，但不熟練測驗的人倘用他來分班也容易發生困難。

第三、使用的測驗應當能量度學校工作之重要方面。除非我們要依照某種科目來分班，我們應當盡量用各種測驗來量度各種科目。

測驗的步驟 測驗的種類和所須測驗的科目既經決定之後，便須注意下列步驟：

第一、在怎樣選擇和怎樣訓練擔任測驗的人。使測驗者事前於專家指導之下先行練習測驗的方法，這是一種最簡捷的辦法。

第二、依照標準手續，施行測驗。倘使被測驗的人數很多，地方又不在一處，則關於時間，測驗者之人數事前須有個計畫。

第三、批閱各種測驗卷，記算學生分數。有須均衡的，必須加以均衡。

第四、把各學生的分數列爲各種表格。

第五、計算各級各種測驗之中數 (median) 或平均數。

第六、把各級各種測驗之常模 (norms) 列爲表格。

分數之合併 把兩種以上的測驗施用於學生以測度他們的總成績，我們便要遇到一種問題。就是怎樣把所得的分數合併起來。一種標準的教育測驗所產生的分數的大小，看測驗中所有問題的多少，規定的時間和記分的方法等等而定。所以各種不同的測驗所產生的分數，大小也不一致。閱讀測驗所產生的速力分數通常每分鐘在百字以上（此指美國的測驗而言），而同樣測驗所產生的瞭解力分數往往可以在每分鐘十五字以下。就克的斯標準算術研究測驗乙組 (Courtis Standard Research Test in Arithmetic, Series B) 說，最高的速力分數是24。所以同一大小的分數，譬如十七分，在某種測驗爲一種意義，在另一種測驗又是一種意義。這種情形的由來，是因爲缺乏一種共同的單位和共同的零點。除偶然的巧合外，沒有兩種測驗所產生的分數是同一單位或由同一零點出發的。

這樣，要把各種不同的測驗所產生的分數合併起來是很困難的，除非他們的單位和零點差不多相同。但是要行合併，總須先有一個共同的基礎。把某種量表分數歸併成別一種量表分數的最普通的方法，就是把各分數看做是對於平均數之差並以標準差爲一單位。⊖這個方法，手續太繁，除非爲精密的研究，不常使用。而比較簡單的方法，就是當零點差不多相等時，把分數依照各測驗的常模的比率做一個反比例的均衡。例如，在測驗甲，第五年級的常模是25，在測驗乙是75。倘使我們把三來乘測驗甲所產生的分數，則測驗甲和測驗乙兩者所各自產生的分數，便可以相互比較。不過，這個方法是先有一個假定，就是對於這兩組分數都給以同等的勢力。還有一個方法，就是依照各測驗的分數的變化率 (variability of scores)：這是麥柯爾所主張的。⊕孟羅的普通調查用算術量表 (Monroe General Survey Scale in Arithmetic) 和斯丹福成績測驗 (Stanford Achievement Test) 都是應用上述簡單的均衡方法。爲實用起見，與其用極精確的方法，不如用這種簡單方法爲是。但是當兩組分數的零點不很相等時候，我們必須從一組分數中加上或減去一種適當的量。

① 參考孟羅著教育測驗理論 (Monroe, Walter S., Introduction to the Theory of Educational Measurements) 二二一面以下。

② 參考麥柯爾著教育測驗 (How to Measure in Education) 三〇面。

各種引伸分數的意義 引伸分數 (derived scores) 之由同一單位和同一零點出發的，則於記錄上和解釋上都很便利。關於這種引伸分數的怎樣推算出來，已經有好幾種提議。而現在使用最廣的，是年齡分數 (age score)。大多數智力測驗中所產生的點分數 (point scores) 都是轉變為心理年齡。這是看各年齡的兒童所得的平均數或中數而定。例如十歲的兒童所得分數的平均數是七十二，則任何兒童得到七十二分者都算作是有心理年齡十歲。大多數成業測驗 (achievement test) 中分數的計算，也依照這個手續。就是把年齡常模 (age norms) 作為轉變點分數為成業年齡 (achievement subject) 的基礎。

③ 所謂點分數，指由測驗直接產生的分數說。例如在測驗中所做對的題數，試做的題數或所達難度之階段，都是屬於點分數。

爲便利起見，各種特殊科目中的成業年齡，漢語中已用許多縮寫的記號來代表。例如 S. A. 就

是科目年齡 (subject age) A. A. 就是算術年齡 (arithmetic age)，成業年齡有時亦用 A. A. 代表。一個學生在一組科目中的平均年齡，就是教育年齡 (educational age)，他的縮寫就是 E. A.

用學生的實際年齡 (chronological age or C. A.) 去除他的教育年齡，所得的商數就是他的教育商數 (educational quotient or E. Q.)。各科目也有商數，叫做科目商數 (subject quotient or S. Q.)。例如 R. Q. 就是讀法商數 (reading quotient) A. Q. 就是算術商數。

再進一步，用學生的智力商數 (intelligence quotient or I. Q.) 去除他的教育商數或科目商數，結果所得的商數叫做成業商數 (achievement quotient or accomplishment quotient or A. Q.)，有些測驗的編造者爲避免重複起見則稱此種商數爲成業率 (achievement ratio or A. R.)。倘把心理年齡去除教育年齡或科目年齡，所得的結果和用智力商數去除教育商數或科目商數是一樣的。因爲這個方法計算比較的便利，所以事實上使用的人也較多。

爲使分數的單位相同，零點相同起見，麥柯爾曾發明過一種轉變點分數爲 T 分數的方法。這

個T分數也是一種引伸分數。他的單位就是十二歲兒童的能力變化率的量數。先把大多數十二歲兒童測驗之後，看這十二歲兒童的分數的分配情形怎樣，就可以得到轉變點分數爲T分數的基礎。詳細的計算方法，請參考孟羅的教育測驗理論，麥柯爾的教育測驗或拙編教育測驗綱要。就容易使人明瞭而論，則T分數似乎不如年齡分數。

教育年齡對心理年齡或教育商數對智力商數 上文說過，分班的第一個目標，在把教育地位同等的學生集合在一起。就這點說，教育年齡——由一組適當的彙選測驗所求得的教育年齡——比心理年齡來得好。怎樣是來得好呢？因爲只有教育年齡能直接顯示何種學生是於教育上處同等的地位。教育年齡能直接測度學生的教育地位；心理年齡只能間接的測度教育地位之是否同等。一個學生的心理年齡和他的教育地位固然有密切的關係，但是因爲還有別種活動勢力存在，所以兩者間不能保持一種完全的相關。申說一句，一個學生的教育地位，除受他的心理年齡的影響外，他的身體的強弱，出席的持久，對於學校功課的態度和勤勉不勤勉等等都是有影響於他的教育地位的。教育年齡對於學生的心理年齡和其他有影響於他的工作的一切動因，都能顧

到。而普通所測得的心理年齡對於這些動因的影響雖然多少也能顯示一些，但是比較教育年齡範圍狹得多了。

然而，就目前所有的教育測驗和特殊能力的診斷測驗說，心理年齡比教育年齡於學生的分班上較爲通用。無論初小學生，即中學生，大學初年級生都可以把心理年齡作爲分班的基礎，雖然有些學校習慣上是根據教育年齡來分班。

分班的第二個基本目標，在把進步速率同等的學生集合在一起。要預斷學生的進步速率將來是怎樣，最好是測度他已往的速率。而學生的心理年齡和教育年齡如果離開了他的實足年齡，幾乎並沒有一點意義，並不能指示學生的進步速率在過去是怎樣。就這點而論，教育商數和智力商數就有極大的功用，因爲兩者都可以顯示學生的進步速率。倘使一個學生的智力，在過去已經是發展得很快，則他的智力商數將依照比例是高的——在一百以上。倘使他的智力在過去已經是發展得很慢，則他的智力商數將依照比例是低的——在一百以下。同樣，倘使他的教育能力在過去已經是發展得很快，則他的教育商數將依比例在一百以上。反之，教育商數也將在一百以下。

重行分班的幾個原則 規模較小的學校在既經分班授課之後，爲求學生能力和工作更加相稱起見，可以利用教育測驗來重行分班，不過須注意下列幾個原則：

一、不應當把學生降級或剝奪他的常規的升級，除非他的教育年齡在打算把他插入的一級的上四分點 \ominus 之下。這樣就不致使愚笨的學生失望，同時對於學生實足年齡的影響（如一班中年齡過大或過小的學生）也可以同時顧及。

\ominus 或譯上二十五分點即 \odot ，或譯下二十五分點即 \odot ，其計算法參考教育大辭書位置數量條，或俞子夷譯教育統計學。

二、不應當把學生升過一級，除非他的教育年齡是超過了預備把他升上的那一級的下四分點。如果覺得這種規定，還不十分穩當，則最好是規定學生的教育年齡須超過了預備把他升上的那一級的中數，方始把他升上去。

三、不應當把教育商數在全校模範的 (typical) 教育商數以下的學生升過一級或一級以上，除非他的教育年齡是在預備把他升入的那一級的中數以上。倘使他的教育年齡是超過了預

備把他升入的那一級的下四分點，同時他的教育商數是在全校模範的教育商數之上，則到學年終了的時候他在二級中所處相對的地位將更屬高一些。但是，倘使他的教育商數是低的（即低於全校模範的教育商數），則到學年終了的時候，他將處於一級的下層，甚至不能及格。

除利用教育年齡和教育商數外，分班時對於學生在各種測驗上所有的成績也宜分別檢查。一個學生的總分數可以很高，但是對於某種測驗的分數容或可以是很低。所以對於算術和讀法等基本科目的測驗，尤宜分別慎重檢查。倘使一個學生在各種測驗中成績都好，惟有算術的成績特別來得壞，則對於這個學生雖然允許他插高一級或二級，同時卻必須對他特別欠缺的能力給以特別的補助。

若在規模較大的學校，學生人數極多，教師也不少，則可以完全利用教育年齡和教育商數來分別班次。就是先把學生依照教育年齡分成若干級。一級之中，再依照教育商數來分為若干組。例如教育商數在一百十以上者為一組，九十至一百十者為一組，九十以下者又為一組。或者照教育商數的大小，把最高的三分之一為一組，中間的三分之一為一組，最低的三分之一又為一組。這樣，

不但可以達到分級的第一個目標——教育地位的同等——並且可以達到了第二個目標——進步速率的同等。聰敏的學生不致因功課容易養成虛驕懶惰等習慣，愚笨的不致因常不及格而自暴自棄。個個學生在一級或一組中都有同等能力的競爭者；個個學生都受到他所需要的教育。一級或一組的地位於無形之中自然提高。

倘使應用智力測驗來分別班級，以上所述的三種原則和其他手續是仍舊可以適用的。只須改教育年齡為智力年齡，改教育商數為智力商數。

三 由教師的判斷分別班級

教師評判之不適當 當學生從家庭或別的學校初進一個學校時，或從同一學校的別個班級中轉入另一個班級中時，倘用教師的判斷作為分別班級的根據，那是不適當的。因為教師對於某學生有了充分的經驗之後，還不敢相信自己判斷之可靠，還不能驟然加以評判，而況是對於一個不相識的學生。做教師的，固然不是全屬盲目的。他們對於最聰敏的學生和最愚笨的學生能準確的分別出來。但是要知道各級學生的能力是延疊的 (overlapping)。所以一個教師可以很明

瞭他自己班中的學生，卻往往不能知道他的班中的最聰敏的學生是不是相等於，較優於，或竟不及別班中的普通學生，或最聰敏的學生。我們如果用測驗來代替教師的評判，則對於學生可以不需要一種事前的相識，可以把異級中的學生相互比較，如同把同級中的學生比較一樣容易。

教師評判之不準確 教師對於學生能力的評判，是一種主觀的測驗。主觀的測驗很容易使教師陷於各種常差（constant errors）。這種常差之最普通的，就是把學生的行為和他的成績混在一起。例如對於有禮貌的學生，或注意教師說話的學生，或常帶笑容使人引起快感的學生，做教師的往往就以爲是能力較高的學生，或在給分數時就多給一點。其實，學生的行為（或意志）於分別班級時，固然應當要加以考慮，但是不應當讓學生的行為來蒙蔽教師對於學生能力的估計。

更有進者，據推孟的研究，做教師的對於學生智力的評判，也常犯一種錯誤，因爲他們忽略了各個學生的年齡的差異和情緒的差異。一個年齡過大（對年級而說）的或情緒快樂的學生，他的智力往往是被估得太高。一個年齡小而插班高的，或性情冷淡的學生，他的智力往往是被估得

太低。又據韋必爾 (Whipple) 對於優秀兒童的深切研究，知道教師雖然不致於把愚笨的兒童看作是普通的或優秀的兒童，但是往往要把優秀的看作和普通的一樣。所以韋氏最終的結論，主張對於優秀兒的甄別選擇和對於低能兒的甄別選擇是同樣須要心理測驗的專家，因為心理測驗的專家所使用的測驗工具是不致於陷入以上所說的常差。

教師評判之重要 話雖這樣說，教師的評判仍是處於重要的地位。因為一個學生的全組織中有幾種要素不是現在的測驗所能測度的。遇到特殊的事例，這種非測驗所能測度的要素，於學生的分級上，反比測驗所測得的部分更屬重要。就測度學生的意志 (purposes) 而說，教師的評判似乎是最直接的。更進一層，做教師的對於學生的體格、情緒以及社交等等特性的推測，似乎也非測驗所能及。概括說來，倘使這些原素是教育的目標，是學校教育所能改進的，則於學生的分級上便應加以注意。所以決定一個學生應當插入那一級時，教師的評判，仍處於一種重要的地位，一直要到這些不易捉摸的能力是證明了並不存在，或是已經包含在客觀測驗的分數的計算中，或是證明了這些能力不是學校教育所能改進的時候為止。

教授年齡的計算 由智力測驗所斷定的是學生的心理年齡；由教育測驗所斷定的是學生的教育年齡；由教師所斷定的，爲便利起見，可以稱做是學生的教授年齡 (pedagogical age)。但是怎樣把教授年齡和心理年齡及教育年齡聯合起來，以決定學生的分級呢？這是一個問題，已經有許多心理學家和統計專家加以研究。據麥柯爾博士的主張，是分爲以下幾個步驟：

一、先用一種或幾種客觀的測驗，把全校的學生測驗一下。能多用幾種測驗——無論是智力測驗或教育測驗——最好，倘使只用一種測驗，而所選的又是屬教育測驗，則最好是用測驗學生的瞭解力的閱讀測驗。

二、核算每個學生的心理年齡或教育年齡（或閱讀年齡）。

三、把每班中的學生，依照他們的心理年齡或教育年齡的大小，順序排列起來。

四、由教師依照他自己的判斷將他班中的每個學生怎樣升級，評定一下。倘使有兩個以上的教師對同一學生做了評判，則他們評定的等次可以平均起來得一平均的等次。

五、由教師評定的一班中等次最高的學生，就給以一班中任何學生對客觀測驗所得最高的

心理年齡或教育年齡。等次較低的，就給以較低的心理年齡或教育年齡。這樣，學生所得心理年齡或教育年齡就是他的教授年齡。

六、把一個學生的心理年齡（或教育年齡）和教授年齡平均一下，就得一個學生的升級年齡（promotion age）。或單用教授年齡作為升級年齡，也可以。倘使一個學生有了心理年齡又有教育年齡，則把三種分數（心理年齡，教育年齡和教授年齡）平均起來。

七、計算每個學生的升級商數（promotion quotient）。舉例如下：

學生名稱	實足年齡	教育年齡	教師評定等次	教授年齡	升級年齡	升級商數
張	140(月)	144	2	138	141	100.7
李	150	138	3	133	135.5	90.3
王	110	133	1	144	138.5	125.9
趙	121	130	4	130	130	107.4

八、根據升級年齡，把學生分級。如屬可能，再根據升級商數於級中再分組。

九、因爲一個學生的升級商數可以表示他的進步速率，所以稍加計算就可以推知他受教育一年之後，他的升級年齡是多少。在下次升級的時候，這個預料的升級年齡，就可以用來替代測驗，這樣就免了每年必須舉行測驗。但是，雖有預估的升級年齡作根據，教師對於學生的評判仍舊是重要的。

用測驗分級之反對論 本章行將結束，順便把反對用教育測驗或智力測驗來分級或升級的論調說一說：

一、強迫年輕的學生和心理上比較成熟的學生競爭——這種論調是代表一種不適當的同情。所以要發生這種論調，是因爲懷抱了一種舊見解：就是以爲一切學生生來是平等的，學生的心理年齡是和實際年齡相緊接或相協調的。換句話說，以爲年輕的學生因爲實際年齡小，心理年齡自然也要低些。其實，據各方面的討究，年輕的學生通常都是一班之冠，而年長的學生往往要十分努力方得相競。依據教育年齡或教育商數來分級或分班正足以補救這種弊病。

二、年輕的學生於社交的適應上發生困難 把年紀輕的學生和年紀大的學生集在一起的

時候，與其說年輕的學生於社交的適應上有困難，不如說年長的有困難。因為年紀較大的學生的活動多半不是年紀較小的學生所能參加，所以年紀大的學生往往厭惡和年紀較小的學生共同作業。這種妬忌是值得我們的研究。現在我們用教育年齡和教育商數來分級，實足年齡雖有大小，智力上不相上下，這種反對論自不成立。

但是有種天才生，年齡極輕就進中學讀書，我們對他將怎樣措置呢？這裏有好幾種意見。有種主張這種學生須延緩他進中學的時期，而使他在小學裏受範圍較廣的教育。但是小學的課程編製是為大多數普通學生着想的，使天才生學習同樣的功課，則他的大部分的時間將消費於不很重要的材料上。所以另有一種主張發現，就是從小學一年級起以至大學，為天才生專開特殊班，這樣，使天才生和天才生集在一起受同樣的教育。倘使他（天才生）在大學畢業的時候，年齡仍舊是太小，可以使他做專門的研究，或從事於別種社會的事業，一直到他身體方面認為充分成熟可以在普通社交團體中立足為止。

至於規模較小的小學，對於天才生之教育，簡直是沒有辦法。惟有聯合好幾個簡陋的小學形

成了一個聯合學校，所謂 *consolidated school*。這樣，個性的差異上或者可以多照顧到一點。

三、忽視身體的健康。有些人以爲天才兒童，精神上雖比較優秀，身體方面常是脆弱。如用測驗分班之後，歸入較高的班級裏頭，功課自然加重，身體的健康定要大受影響。但是據推孟的研究，天才兒童的身體並不比普通兒童的身體脆弱。他們和普通兒童一樣，有些是脆弱的有些是非常強壯的。如果功課的輕重確能影響於身體的健康，則對於身體不甚健康的天才兒童自然於功課方面應當給他比較輕易一些的。

四、偏重了智育方面的特性同時忽視了品性方面的特性。反對用測驗來分級的人往往說，用測驗分級是依照能力來分級，不是依照各學生的意志 (*Purposes*) 或行爲來分級。但是我們要知道學生的善良的意志或行爲可以在一級中養成，也可以在別一級中養成。各種意志間的相關決不如各種能力間相關的密切。更有進者，大多數依照智育方面的成績而升級的學生，也可以依照他們的混合的品性方面的地位來升級。推孟教授曾經研究 ⊖ 優秀兒於智育的和個人的特性上所佔有的範圍。他所研究的特性共有二十四種，就是名譽心，繼續注意力，堅忍，創發性，準確，意

志力，天良，社交適應，領袖才，個人的容貌，歡愉，合作，身體方面的自制力，勤懇，勇敢，可靠，語言的自我表現，智育方面的謙遜，服從，孚衆望，氣質的溫和，情緒的自制，不自私，速率。據他研究的結果，這些特性都是和智力積極的相關，換句話說，就是和能力極有關係。就第一種特性——名譽心——說，於優秀兒中有百分之五十八的相關度。最後一種特性——速率——也有百分之二十八的相關度。其餘各種特性和能力或智力的相關度就在這兩極端之間，就是依照上面所列舉的序次逐漸變動。推氏並說，他能根據這二十四種特性的平均數來預斷一個學生的智力商數。這樣，用測驗來分級並不會忽視學生的品性方面的特性。

⊙推孟著學校兒童的智力(The Intelligence of School Children)第五十八面。

討論或練習：

- (1) 普通分級或分班的弊病在那裏？
- (2) 在經濟比較的不充分或規模較小的學校中怎樣適應低能兒的教學？
- (3) 試用智力測驗把你班中的學生測驗一下，看他們所插的班級是否適當？

(4) 你班中有沒有優秀兒童？如果發現，請把他們身心方面的特點說出來。

第三章 測驗在教育診斷上之功用

一 開始情形之診斷

一個校長或督學對於一所新設的學校，或一個教師對於新設的班級負了責任之後，便須舉行兩種診斷：一種是對於開始情形的普通診斷；一種是對於各級或各學生的特殊缺陷之比較詳細的診斷。

爲討論之具體化起見，本章和以下二章將把讀法教學做例子。讀者苟能把讀法教學的診斷手續領會了，其他各科的診斷手續自然不難類推。以下就是逐步的診斷手續，爲明顯起見逐步的手續用數目字來表明。

一、決定各學生的實足年齡 表一第一行就是表明五個選擇的四年級學生的實足年齡。（這五個學生在本章和以下各章中將看做是一個整個的四年級。）

二、決定學生的開始的讀法能力 桑代克、麥柯爾合著的讀法量表可以測定學生的開始的

閱讀能力（中華教育改進社所編各種測驗也可應用）現在把這測驗的一部分和怎樣施行測驗，怎樣記分等等引述於下：

表一 顯示讀法教學指導上所須要之知識

學生	A	B	C	D	E	平均
1. 實足年齡	121	123	124	118	100	
2. 開始讀法分數	40	43	30	31	38	36.4
3. 開始讀法年齡	121	130	93	96	116	
4. 開始讀法商數	100	106	75	81	116	
5. 開始智力年齡	121	150	83	100	111	
6. 智力商數	100	122	67	85	111	97
7. 開始成業商數	100	87	112	96	105	100

8. 估計最後讀法年齡	131	142	100	105	127	
9. 估計最後智力年齡	131	162	90	109	122	
10. 讀法分數的目標	43	47	33	34	42	39.8
11. 最後讀法分數	43	50	34	36	43	41.2
12. 最後讀法年齡	130	150	104	110	130	
13. 最後成業商數	99	93	116	101	107	103
14. 最後 A. Q. 減去開始 A. Q.	-	-	-	-	-	+3

先把這個讀一下，然後再寫答案。倘使係須要再讀，不妨就再讀一下：

納爾的母親到華達街百貨公司買了十鎊糖，一打蛋，一袋鹽。她總共付了一塊錢。納爾和久胡是和她同去的。他們回家的時候，在伯陽街上看見一輛用三匹馬拖的救火車。

(1) 鹽是在箱子裏，袋子裏，罐頭裏，還是在碟子裏頭？……

.....
.....
.....
(2) 她買了多少蛋?
.....
.....
.....

.....
.....
.....
(3) 他們在伯陽街上看見些什麼?
.....
.....
.....

.....
.....
.....
(4) 百貨公司是在什麼街上?
.....
.....
.....

桑代客麥柯爾讀法量表第二類說明書

怎樣施行測驗 把測驗卷分發給各學生。卷子正面向上。對學生說未喊口令以前，不得把卷子展開。叫學生把卷子第一面上的空白先行填好。主試者先朗讀一段文字，同時使學生閱讀。叫學生口頭回答後，再寫書面答案。在說出『把卷子展開，做』以後三十分鐘，停止學生再做。不得給學生任何幫助。

怎樣閱卷 主試者最好對卷子上第一面的問題先做答案，然後對測驗中所附錄的標準答案研究一下。批閱學生卷子的時候，把全級學生對於第一面上的問題所做的答案先行批閱，然後再批閱第二面，第二面完結，再看第三面。這樣於時間上比較的經濟一些。標準答案紙上的答案，是對的和不對的並列的。但是所有對的答案，並不是就只限於這幾個。所以學生答案的對與不對是可以依照下列幾個原則決定：

- (1) 凡答案之與標準答案紙上所列最壞的答案價值相等的，作為對的答案。
- (2) 完全的句子是不必要的。
- (3) 一個對的答案但另加上一個不對的或沒關係的答案，便作為錯誤的答案。
- (4) 一個對的答案另附了一個適當的添註，仍作為對的答案。
- (5) 凡位置失常，省略或看不清楚的，都作為不對的答案。
- (6) 凡修改，或使用縮寫或句讀，拼法，字體大小有錯誤的答案，仍作為對的答案，除非他們表示了實際的誤解。

(7) 概括言之，一個學生的答案必須要表示他已經適當領會了試卷的文字和所有的問題並能對於這兩者寫些使人可以瞭解的反應文字。

怎樣計算分數 先把學生做對的問題求一個總數，然後照第二表第二行查相對的T分數。這T分數纔是學生的分數，應當照第五表第三縱行的樣子把他登記起來。

表 二

做對的問題	T 分數	做對的問題	T 分數
0	23	19	45
1	25	20	47
2	26	21	49
3	26.5	22	51
4	27	23	53

5	28	24	56
6	29	25	58
7	31	26	60
8	32	27	63
9	33	28	67
10	34	29	71
11	35	30	76
12	36	31	79
13	37	32	82
14	38	33	86
15	39	34	92
16	40	35	96

17	42
18	43

怎樣計算年級分數 年級分數就是一級中學生所得的平均數，參考表五第三行。

怎樣把年級分數和年級常模比較 表三是顯示各年級常模的一個表。就在這個表中找尋

表 三 (顯示二年級 A 至十二年級 B 之年級常模)

年級	2A	2B	3A	3B	4A	4B	5A	5B	6A	6B	7A	7B	8A	8B	9A	9B	10A	10B	11A	11B	12A	12B	標準分數
年級常模	26	30	33.7	37.3	39.6	41.8	44.9	48.0	50.9	53.7	56.0	58.3	59.6	60.9	61.5	62.1	62.9	63.6	64.5	65.4	66.3	68.1	72
受測學生 分數(以千 數計)	02	03	3	6	5	10	5	10	5	10	5	10	5	10	估計	1	估計	1	估計	1	估計	1	03

相當的年級常模，然後把他寫在年級分數之下。如第五表所示的常模是 41.8，因為他是四年級

B 的學生的常模。

怎樣把學生和年齡常模比較 主試者先從第四表第一行中看學生的 T 分數，然後在第二行中尋相當的讀法年齡。把學生的實足年齡除這讀法年齡，就得讀法商數，如第五表所顯示的。凡

表 四 (讀法年齡常模)

T 分數	讀法年齡	T 分數	讀法年齡	T 分數	讀法年齡	T 分數	讀法年齡
21	67	41	124	61	181	81	238
22	70	42	127	62	184	82	240
23	73	43	130	63	186	83	243
24	76	44	133	64	189	84	246
25	79	45	135	65	192	85	249
26	82	46	138	66	195	86	252

27	84	47	141	67	198	87	255
28	87	48	144	68	201	88	257
29	90	49	147	69	203	89	260
30	93	50	150	70	206	90	263
31	96	51	152	71	209	91	266
32	99	52	155	72	212	92	269
33	101	53	158	73	215	93	272
34	104	54	161	74	218	94	275
35	107	55	164	75	220	95	278
36	110	56	167	76	223	96	281
37	113	57	169	77	226	97	284
38	116	58	172	78	229	98	287

39	118	59	175	79	232	99	290
40	121	60	178	80	235	100	293

學生的讀法商數是一〇〇，就是表示他的讀法能力是普通的或中常的，一〇〇以上或不到一〇〇就是表示他高於或低於中常的讀法能力。所有商數均用百乘，以避免小數點。

怎樣解釋 T 分數 測驗學生讀法能力的單位不以做對的題數為準，而以 T 分數為準，這是有道理的。因為做對的題數不是測驗讀法能力的最好的單位，各題的難易不同，原來 33 和 34 分的相差不一定與 43 和 44 分的相差一樣。但是 T 分數 33 和 34 的相差常等於 43 和 44 或 44 和 45 的相差。

T 分數的功用不但這樣，任何學生得到 50 T 分數就可以說他的讀法能力和十二歲兒童的平均讀法能力相等。任何學生得到 70 T 分數，就可以說他的讀法能力高於十二歲兒童的平均讀法能力 20 T 分數或 2 個標準差。得到 35 T 分數，就可以說他的讀法能力低於十二歲兒童的平均讀法能力 15 T 分數或 1.5 標準差。

表 五
分數單式樣

學校名稱——四年級 B, 教師姓名——		年 月 日		
實足年齡 (月數)	學生姓名	T ¹ 分數	讀法年齡	讀法商數
124	丁 某	27	84	68
120	王 某	52	155	129
148	顧 某	40	121	82
134	華 某	46	138	103
平均數.....		41.3	平均數.....95.5	
年級常模.....		41.8	年齡常模 100	

T¹分數再可以解釋如下：

T 分數	為十二歲兒童所 超過的百分數	T 分數	為十二歲兒童所 超過的百分數
25	99	55	31
30	98	60	16
35	93	65	7
40	84	70	2
45	69	75	1
50	50	80	0.1

怎樣解釋讀法商數 怎樣解釋一個學生的讀法商數或一級學生的平均讀法商數，請參考下表第三行。第一行是測驗五千個學生後所得的讀法商數；第二行是這五千個學生的所占的百分數。

讀法商數	學生百分數	解釋
55 以下	0.1	極端下等
55—65	2.4	下下等
65—75	8.0	下等
75—85	13.5	下中等
85—95	18.0	中等
95—105	19.6	上中等
105—115	17.0	上等
115—125	10.5	上上等
125—135	5.8	極端上等
135—145	3.5	
145 以上	1.5	

要表示一個或一級學生的讀法能力至少有兩種有效的方法。一個學生的讀法能力究竟怎樣，第一可以把他的讀法分數和他的年級常模比較。但僅僅使用這個方法，可以鼓勵學校無意或

有意中於實足年齡上把學生留級以便增高學校表面上的效率。因為倘把年級做比較的標準，年齡過大的學生的百分數占得多了，學校的效率表面上就要高些；反之，年齡過小的學生的百分數占得多了，學校的效率表面上就要低些。

第二，可以把他的讀法分數和他的年齡常模比較：這就是讀法商數所表示的。因為讀法商數既把學生和一定的年齡比較，不是和一定的年級比較，他的價值就勝於年級常模，因為學校不能夠造成一種不適當的實足年齡上的留級以便讀法商數之增高，也不能夠造成一種不適當的實足年齡上的升級以便把讀法商數減低。

時間的節省 倘使主試者是擅長統計的手續的，並且他的研究目標只在於年級分數或年齡分數，那末他便無須把每個學生的做對的題數轉變為T分數。他只須做一個分配表，來顯示每種做對的題數的學生數（如做對二十題的有幾人，做對十一題的有幾人，做對十二題的又有幾人）。這樣，分配表的第一行便可轉變為T分數，由許多的T分數得到一個平均的T分數。

再主試者的目標倘僅在學生的閱讀年齡和閱讀商數，則把上面第二表的第一行（即做對

的題數)錄入第四表中,時間上也大可以節省不少。因為這樣把做對的題數轉變為讀法年齡時省節了一種手續。

常模之準確度 桑代客的這一類測驗的年級常模中所包含的學生數,實沒有確切的統計。一種估計的數目,是見第三表中,至於第四表所列的年齡常模,是根據下列的學生數做成一種曲線再修改而成的:

年齡	7-8	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17	成人
學生	128	635	1275	1462	1561	1833	1662	1112	430	65	300

三、轉變開始的閱讀分數為閱讀年齡 上表所述桑代客、麥柯爾的讀法量表的說明書,已經把轉變讀法分數為讀法年齡的方法指示出來。A R等五個學生的讀法年齡是見第一表第三行中。

這種開始情形的發見有兩種重要的功用。第一就是可以避免重行教學。第二就是可以避免

先時的教學 (premature teaching)。有人研究過美國學校中有百分之三十三的留級學生，爲重行教學這些留級生，每年要消耗四千萬金元。這是對於民衆的大損失。但是對於留級生自身的損失恐怕還不止這樣。知道了學生的開始的情形（譬如讀法能力以及別的科目的能力）把他插入相當的年級中受相當的教學，則對於公衆，對於學生使都不致於受無謂的損失。怎樣叫做先時的教學呢？就是課程中間有一部分非常之難的，而強迫學生去學他，或對於某種課程之基本學識還沒有受過相當的訓練，而就把某種課程直接教學學生，這都是先時的教學。知道了開始的情形，不但給學生以適當的指導，就是做教師的也能够措置得當。因爲知道學生在什麼地方有缺陷，便知道什麼地方須要最大的幫助。富脫 (Forte) 曾經說過：『當學生和教師知道他們自己是在什麼地方並向什麼地方走去的時候，他們行程的目標是可以到達的；否則，是可疑的。』開始情形的診斷就是指示給學生和教師『他們是在什麼地方。』

四、決定讀法商數 關於怎樣計算和解釋讀法商數，桑代克、麥柯爾的讀法量表的說明書中已經說過，此地自無須贅述。他如算術商數，拼法商數等等的計算和解釋也是同樣的。

五、決定開始的智力年齡 要決定各個學生的智力年齡，應用團體智力測驗，不如應用個人智力測驗。因為個人智力測驗費時較多，而結果也較為準確。我國現在出版的團體智力測驗只有寥世承等所編製的一種，個人智力測驗只有陸志韋等所編製的比納西蒙智力測驗。在美國則以推孟所編的訂正比納西蒙智力測驗為最通行。他如國民智力測驗 (National Intelligence Test) 及品脫奴班脫生的成業量表 (Pinner-Paterson Scale of Performance) 應用得也很多。

學生A的開始的智力年齡是121。聯同其他四個學生的智力年齡都見第一表第五行。

六、決定智力商數 一個學生的智力商數就是用他的實在年齡去除他的智力年齡所得的商數。例如第一表中學生A的智力商數100，就是用121去除121的結果。學生B的智力商數122，就是123除150的結果。其餘三個學生的智力商數請參觀第一表第六行。

一個學生的智力商數，對於擔任任何學科的教師都是很有價值的，因為一個學生的智力商數的大小就是他的普通智慧的指數。推孟教授曾經說過，一個學生的品性是最最重要，次於品性

的，就是他的智力商數。各種智力商數的意義，如下。

一四〇以上……天才或類似天才

一二〇——一四〇……極優秀智力

一一〇——一二〇……優秀智力

九〇——一一〇……正常或普通智力

八〇——九〇……愚笨

七〇——八〇……類似智力缺陷，有時低能

七〇以下……確定的低能

像這些讀法年齡，讀法商數，心理年齡以及智力商數等等的決定，不但對於教授上供給有價值的指導，並且對於教育指導（研究學生受普通教育的能力和專攻特殊科目的能力後所施的指導）也準備了一種基礎。

普通學習能力 探求教育目標，是教育上的一個問題。探求某種能夠達到這些教育目標的

人（即可以加以教育的人）那是教育上的另一問題。豬牛羊犬馬，都有他們的不同的學習能力，然而只有人類纔看做是有充分的學習能力。換句話說只有人類纔看做能在系統的長時間的教育中獲得相當利益。

不過，學習能力的診斷，並不是把人類和一般動物分別開來就算完了。人類中各人的學習能力是不同的，比犬類中各犬的學習能力的差異更大。更進一層說，人類中有一部分人的學習能力比不上最聰敏的犬，貓或猴的學習能力。

簡括的說，診斷學習能力的手續，是在測驗各個人所能够學習的，或他所已經學得的。就第一種方法說（即能够學習的說），只要使兒童生活於新的環境之中，然後記錄他怎樣學習。就第二種方法說，乃是確定兒童從出生以來，已經生活於一種學習環境之中，我們只要測驗他已經學得的是有多少。

第一種學習能力的測驗，是父母或鄰居對於兒童所加的一種極簡單的沒標準的觀察。這些測驗一定是不準確的，因為有許多恒錯（constant errors）像父母之愛或鄰居的妬忌心，缺少公

平的不變的標準等等。這些主觀的測驗在過去比在現在或者比較的是準確，因為現在有許多獨創的人已經發見了比較更穩妥的測驗標準。

父母對於兒童的測驗雖屬主觀，但也有他的選擇作用，就是極端愚笨的兒童往往因此留在家裏，比較聰敏的就被送進學校。兒童進了學校之後，教育的選擇作用更大。下幼稚園，上至大學，中間無論經過那一階段，都有選擇的作用在裏頭。愚笨的不得升級，特別聰敏的，有特殊班。有些學校規定須有某種學習能力纔得入學。有些國家簡直認天才兒童為一國之資產，與其認為是他的父母所有，不如說是社會所公有。因而選擇了這些天才兒童施以特殊的教育。

話雖這樣說，向來學校中的選擇作用也是偏於主觀的，不合科學的，是因人而異的。要有合於科學，不因人而異的教育的选择，只有施行客觀的標準測驗。標準測驗在目前雖不能算是十分準確，但是已公認為教育選擇中最良的方法。歐戰中美國曾用陸軍智力測驗來選擇各種兵士，成效大著，從此以後測驗的選擇作用更為一般人所知道。科崙比亞大學用桑代客的大學入學智力測驗來選取投考的學生。別的大學和別的較為低級的學校也使用標準測驗來決定學生的去取。至

於低能兒學校之收取學生早已使用比納西蒙的智力測驗。

以前所創製的測驗，屬於個別的（即個人測驗）多，屬於集團的（即團體測驗）少，這於採用測驗來決定學習能力上很不方便。現在各種標準的團體測驗已經發明了不少，測驗的專家於某種準確程度以內能够告訴我們某個兒童富有某種智力商數，某種智力商數可以使兒童做某種學校或學校中某級的工作。推孟教授根據他自己的經驗曾經說過：智力商數八十的學生將不能做超過第三或第四級以上的功課。智力商數七十的學生將不能做超過第五或第六級以上的功課。智力商數八十的學生將以第七級的功課爲限制。智力商數九十的學生，要堅忍耐勞方得畢業於高等小學。教育商數六十，七十，八十，或九十的學生，如果他們的教育機會是正常的，則他們的意義是和上述智力商數可作同樣解釋。

學習特殊科目的能力 斷定學習特殊科目的能力比斷定普通能力要艱難些。目前而論，是使用三種方法來解決的：

（1）測驗學生在一科目（有繼續性的科目）中已往的成績和進步的速率。這個方法是測

驗學生的能力同時包含他的意志 (purpose) 在內。讀法，書法，算術，拼法，作文等都是小學校內多有繼續性的科目。在這些科目中學生的能力怎樣，自第三級起至第八級止都可用團體測驗來測定一下。根據了他們在這些科目中的讀法商數拼法商數等等便可對於他們的將來加以忠告。至於高小學生，可以指定一種工作，要他自己探求，經過一種短期間後，對他施行相當測驗。他將來對於這科目的學習能力怎樣，便可看他最近的進步而加以估計。

(2) 對於專攻科目的成就上所必須的心理能力加以測驗。這個方法是間接的包含學生的意志。就高小的算術一科說，要在這科中有成就須有某種特殊的心理能力。洛哥斯博士^①曾經分析數學能力中所含的心理的要素並編製一種測驗來測驗這些原素。倘使這個分析是準確的，倘使所編製的測驗確能測出這些所發現的心理能力，又倘使證實被測驗的學生確有這些必須要的能力，則我們對於這個學生在數學（例如幾何學）中將來是怎樣，便可作一準確的預斷。

(3) 測驗普通智力。這種預斷法並不要求對於必須要的能力加以嚴格的分析。學業上的成就，無論是在小學的，高小的，大學的，大家都知道是和普通的心理能力有極密切的相關。所以知道

了一個學生的普通智力，便可使我們做一種平穩的準確的預斷。

①洛哥斯著數學能力測驗及其預斷的價值 (Agnes L. Rogers, *Experimental Tests of Mathematical*

Ability and Their Prognostic Value)。科崙比亞大學師範院出版部。一九一八年。

七、決定開始的成業商數。通常每把一級學生的分數和年級常模比較，以定這一級已往效率的高下。但是用這種方法，不如用成業商數來得妥當，因為用年級分數和年級常模比較完全沒有顧到學生智力的高下。而成業商數乃是用心理年齡去除讀法年齡（或其他科目年齡）所得到的。試參觀第一表學生 A 讀法年齡是一二一，他的心理年齡是一二一，所以他的讀法的瞭解力（或讀法的成業商數）是一二一除一二一得一〇〇。其餘學生的成業商數見表中第七行。

成業商數是現今對於研究學習效率，教授效率，視察效率的最確當的一種量數；他是做報告給學生的父母或評判學生的唯一根據；他是一種最好的指數可以指示我們何種學生須要我們的注意或鼓勵，何種學生須要我們的監督，何種學生須要「聽其自然」。

我們每把普通智力較低的學生和實際年齡相似的學生集在一個班級之中，要求他們做平

等的競爭，如果他們成績不佳，便不許他們升級，或竟公然加以訓斥。這是不公平的。有了成業商數，便可對於這種普通智力較低的學生有個公平的待遇，因成業商數是要求這種學生依照他天賦的心理能力所能達的速度而進步的。

還有一種情形發生。就是我們每每督促一個學生竭力向上，但是如果就他身體的健康論，是不應當加以督促的；反之，有時候我們每每抑制一個學生叫他不要操作過勞，而實際上這個學生卻是一個應加強烈的刺激的學生。學生甲在一級中成績最低，照學校的普通習慣說，他就看做是最無效率，甚至全校的好壞都要他來負責。學生乙工作優良，因而得分獨多，全校都稱譽他。可是我們如果對於事實略加考察，則學生甲容或是已經盡量利用他自己的能力，而學生乙倒是一個最沒效率的（即未曾盡其所能）。有了成業商數，我們便可依照他來施行相當的措置。

以上就學生而論，若就教師說，成業商數也有一種功用，例如一班學生都是愚笨的，而要求擔任這班學生教學的教師能產生標準的結果，這是何等的不公平。有了成業商數，便不致於有這種盲目的要求。

有時候做教師的擔任了一班通常智力的學生，而這班學生以前幾年已受了不適當的教學。若是要要求擔任這班教學的教師在一年以內來糾正幾年來的忽略，也是不公平的。第一表中學生的平均成業商數，是表示這一班學生在以前是進步的，並對於教師和督學表示這一班學生在開始時候並沒有受到阻礙。

任何學生或班級，如果有了成業商數一〇〇，便是表示有滿意的進步。做教師的，除須注意學生身體的健康和別種能力的發展外，務必盡量使讀法的成業商數高出於一〇〇。通常，一個教師的首要任務在扶助年輕的優秀兒童。這種優秀兒童在分班時候，插入的班次往往低於他們的能，力，又因為他們於別方面又受了種種忽視，所以他們的成業商數往往比別的學生的低些。

二 診斷的方法

當一個學生於開始的時候就被發覺於解決算術問題（或他種問題）上有困難，我們去決定他所以困難的原因，由近因推至遠因，由遠因推至更遠的因，一直推究到教學的單原（*teaching unit*）。簡單說一句要用標準的測驗來問根究底。測驗之後，知道他對於解決算術問題的困難，是

由於讀法能力之有缺陷，而讀法能力之有缺陷，是由於他對於各種字的意義的認識不精當。凡此種種推究潛在原因的方法就是診斷的方法。診斷既定之後，指示出種種校正缺陷的方略，這就是診斷的機能。沒有診斷的指導，雖施行種種校正學生缺陷的教學，他的成功是全靠機會的。

學校裏頭每天也有他的許多診斷，可是他比不上用標準測驗的診斷，因為他是極浮泛的，有些簡直是極不準確的。如果增進他的準確度，就是增進了校正缺陷的成功機會。有人說要增進診斷的準確度，須要相當的時間。但是要知道一個教師教了一年的算術，他的學生中竟然有許多未曾得到一點進步，苟能行準確的診斷，豈不就是節省了許多時間。

診斷法一 學生自己的內省 這個方法是極顯明的，是常常用的，用不着多大的討論和解釋。大多數學生不但知道他們自己的困難的所在，並且知道所以困難的原因。如果學生能診斷他自己的困難點，則做教師的可以省了不少時間。就是當學生不完全明瞭他自己的困難時，和他談話一二次之後，有經驗的教師也可以得到不少的資料，作為診斷的根據。

診斷法二 正常工作的觀察 教師對於學生的觀察，也是獲得診斷根據的一種方法。學生

甲於讀法上有困難。教師注意他，知道他的眼睛對於每行的字並沒有適當的注視，有時看下去，有時又回上去，全是不規則的。有了這種觀察，便可以對所有困難加以診斷。學生乙對於指定的歷史功課能做完了的很少。教師於觀察之後，知道他於讀歷史的時候，忽然搖頭，忽然動唇，忽然弄髮，沒有集中的注意。這也是對於有經驗的教師的一種診斷的根據。學生丙做算術極慢。有經驗的教師可以做一種試驗的診斷，⊖就是注意他當做加法的時候是不是用手指，足指或舌頭來計數，或是口中低聲自語『七加六成十三，十三加八成二十一。』學生丁不會做命分的除法。觀察之後，知道他的困難之原是由他忘記了把分母倒轉。凡此種種對於學生正常工作的順序上的觀察，無論是簡單的或是極複雜的，都是獲得診斷根據並規定校正方略的資料。

⊙關於算術能力的診斷，最好參考克的新算術練習測驗說明書 (S. A. Courtis, *Teacher's Manual for the*

Standard Practice Tests; World Book Co. 1915.)

診斷法三 工作進程之口頭陳述 有些工作上的困難，無論是由學生自己的內省，或由教師從旁的觀察，都不能明瞭他們潛在的原因。而診斷的目的在指示學生採取一種顯然的行動，務

使可見的缺陷之不可見的原因除掉。如果學生的普通行動和他的書面工作都不能暗示我們怎樣去診斷，則使學生口頭陳述他工作的全歷程，這也是一種診斷法。如使學生走到講壇上去口頭陳述他做加法的歷程，他如果於加法上有任何困難，做教師的便可以明瞭。烏爾氏曾用口頭陳述法發見學生於加法和減法中的心理過程。⊖據他說，有個四年級的學生做 91715 的加法極遲而又極難。經過質問之後，知道他是將較大的數目分拆為小數目然後再加起來。就是當他做這個题目的時候，他對自己說 $9 + 2 + 2 + 2 + 1 = 16$ ，再加 5 成 21 。其他笨重而費時的計數法，烏爾氏舉出的很多。如果不用口頭的陳述法，做教師的很難知道，結果學生的缺陷將永遠存在。

⊙參考烏爾算術中診斷學生工作方法之標準材料 (W. L. Uhl, the Use of Standardized Materials in

Arithmetic for Diagnosing Pupils' Method of Work; 見一九一七年十一月 Elementary School Journal)

診斷法四 測驗結果之分析 有許多測驗的編製不但可作通常意義的測驗用，並且可作

幫助診斷用。孟羅的算術診斷測驗 (Monroe's Diagnostic Test in Arithmetic) 就是診斷

測驗的一種。就實際而論，任何標準測驗多少總有些診斷的價值。

桑代客曾用他自己編製的讀法量表第二類 (Reading Scale Alpha 2) 來測驗學生以求明瞭學生於讀法所以發生不完全的瞭解的原因。根據測驗，他對於誦讀能力分析得很詳。◎依照他的意見，『誦讀是一種極精微的手續。他包含對於一句句子中各個要素之衡量，各要素間彼此適當關係之組合，選擇各要素之某種含義而拒卻別種含義，結合各種勢力以決定最後的反應等等。事實上就是回答一個簡單的節段中的簡單的問題，也包含模範的推理作用之特有的性質。』

『適當的誦讀中(1)每一個字(或詞)產生一個適當的意義，(2)每一個適當的意義都和別的意義比較之後，給以適當的衡量或價值，(3)再對所產生的觀念加以考察或證驗，看他是不是滿足所期望的意義或滿足所以要誦讀之意志。錯誤的或不準確的誦讀是(1)因為各個字(或詞)之錯誤結合，(2)因為對於各個意義看得太重或太輕，(3)不把誦讀中所產生的觀念看做是臨時的，以致當觀念呈現的時候便加以監視或起而發出迎拒的作用。

『尤須注意者，關係詞如代名詞，聯續詞，介系詞等等是含有多少種準確程度的意義。這些詞所

發生的勢力，也看各個人的見解而異。一個學生可以極準確的知道「雖然」二字的意義，但當誦讀含有雖然二字的句子時他容或把雖然二字看做是和「和」或「或」或「倘使」同樣解釋。「對於句中各要素之適當的衡量是很重要的，可是學生於這方面的鑑別力極低。所以大多數答案的錯誤，是由於過分看重某種要素或過分看輕某種要素。」……

◎參考一九一七年六月教育心理學雜誌(*Journal of Educational Psychology*)中 E. L. Thorndike 所著 *Reading as Reasoning: A Study of Mistakes in Paragraph Reading*。

診斷法五 發育史 教育診斷上採用發育史，是和醫學診斷或機械診斷上採用發育史同樣的有效。一個人身體方面有了暗疾，去請教醫生，醫生一定要病人將過去情形完全說出來。你的汽車壞了，汽車修理人一定要問你怎樣把他弄壞的。同樣，要請心理學家來治療一個心理上有所缺陷的兒童，他就要追溯這個兒童的歷史，希望在他的過去事實中得到一點可供診斷用的暗示。發育史不但是追究兒童出生以前的環境，並且追究到兒童的父母和祖父母。有許多根本的教育上的缺陷，都不是突然發生，乃是積漸而來的。當他們最初發生的時候，並沒有注意到他們的根由是

深深潛在於過去之中。所以要診斷成功，必須追究這些根由的起源。

診斷法六 對比 一個教師不能於診斷上成功，往往僅因為他不知道能力缺陷之通常的

原因。例如，有個學生不能表現滿意的進步，因為他的工作方法是無效率的。但是沒有經驗的教師自己也不知道什麼方法是有效或無效，所以他不曾於診斷上成功。遇到這種情形，就可利用對比法。所謂對比可以選擇同級中最好的和最壞的兩三個學生，或這一級的兩三個學生和較低或較高一級的兩三個學生來對比。就是當兩三個最有成就的學生工作的時候，教師從旁加以觀察，觀察了這一組之後隨即就觀察最沒成就的學生工作的情形；或使這兩組學生各個都口頭的陳述他的工作程序；或把兩組加以測驗，並分析測驗結果；或於同期間內對兩組施行任何別的診斷法。對比的診斷，使有能力的和沒能力的學生的不同點顯現出來，這樣便可對沒能力的學生的缺陷容易加以診斷。

診斷法七 能力的全分析 對於某種能力中所含有的感覺的、心理的、運動的、作用加以完全的、澈底的分析，是診斷專家最後的方法。所以說他是最後的方法，因為他是最費時間，並因為診

斷專家使用這個方法來診斷時，如果不能成功，便不能再進一步繼續診斷。一種完全的分析，通常須把上面所述的種種診斷法聯合使用。他是利用上述種種診斷法所得的資料的。

診斷上所必須的技能 診斷專家之成功，第一須對於各種能力缺陷之普通原因有相當的知識。第二，須時時於練習中或經驗中留心去解釋精微的行爲，看這些行爲如同已知原因發生作用之證據。第三，須有使不易見的暗示明顯表示出來的技術。第四，於診斷確定之後須有規定怎樣去校正或治療的知識。

以下列舉幾種能力缺陷之基本原因。時疫之來，通常都可追究到幾個發源地；能力的缺陷，不問他是屬於那一種能力，也可以歸結到幾個基本原因，而當所診斷的缺陷是使人討厭或是屬於頑固性的，尤其是這樣。所以無論何人，於開始診斷之前，應具有能力缺陷之普通的基本原因之知識。

不充分的練習 某種能力於某種學生完全不發生功用，僅因為他們從沒有研究怎樣去發展這種能力。有時候某種能力是發現機能的，但是並不完美，因為缺乏充分的研究和練習。這種情

形是容易解決的，就是對未練習的加以練習，練習不充分的，加以充分的練習。我們所須考慮的問題，是研究和練習充分之後，能力仍不表現他的機能，或表現起來仍舊不完美。

不適當的工作法 工作的方法是很多的，裏頭容或有個最好的方法。學生的範類和氣質，各個不同，容或須要各種不同的方法，或者有一個最好的工作法可以適用於全體學生。但是無論如何，學生工作在標準以下的，總是因為他們使用沒效率的工作法。

基本技術的欠缺 所謂欠缺，或指沒有充分的技術，或是對於新的環境不會應用已得的技術。心理作用，不到潛在的技能變為自動的和可以移轉的時候，不會於文學上獲得欣賞，於歷史、地理、科學、數學問題之大原理或其他任何較高深的原理上起什麼反應。一個畢生於讀法上有困難的，於歷史上也不會有迅速的進步。補救的方法，在使他的基本技能（誦讀）變為自動的。

缺乏興趣 能力之發展，須要充分的興趣或意志，這是不必多說的。世上不知有多少的失敗是由於缺乏對某項工作的興趣所致。

身體的缺陷 任何對於學生能力之診斷，都不能省略對於他的感覺機關的診斷，尤其是對

於眼和耳兩種。因爲一切知識和經驗之獲得，都必須經過感覺機關。如果學生所聽到的僅占對他所說的百分之五十，或所看見的僅占他所看的百分之五十，則教學的效率也不能超過百分之五十以上。

任何診斷，也應考慮學生之反應的機制。凡從感覺機關收進去的，一定要從反應機關施出，教育的圈子 (educative cycle) 纔算完全。學生於圖畫，書法，手工和其他形象藝術上的進步，由於逐漸校正他的肌肉協調上的缺陷，比由於直接校正圖畫或書法能力來得多。

正常以下的智力，低等的先天的智力是能力缺陷之最主要的原因。因爲智力自身是普遍的 (ubiquity)，並且不能加以改良的 (unimprovability)，所以這種缺陷原因是有特殊的意義。但是有許多膚淺的診斷家每每並沒有注意到這智力的重要性。學生於歷史上不能有滿意的工作或者是由於他讀法能力的不健全。但是讀法能力爲什麼不健全？就大多數的事例論，讀法能力不健全，是由於低等的先天智力。所以當學生留級或比較難於加以校正教育的時候，最好施行一種智力測驗。

討論或練習：

- (1) 試用陳鶴琴初小默讀測驗來決定，你班上學生的開始的默讀能力。
- (2) 何謂年級常模？何謂年齡常模？兩者的功用那一個大？
- (3) T分數的意義是怎樣？（請參考商務印書館各種測驗說明書。）
- (4) 何謂學習的普通能力？怎樣診斷？
- (5) 何謂學習的特殊能力？怎樣診斷？
- (6) 成業商數是怎樣求得的？他的用處怎樣？

第四章 測驗在教學上之功用

一 練習測驗之應用

練習測驗之內容 在開始的測驗舉行以後，教師便應依照最好的教育學的手續來施行讀法教學。但這並不是說在這時期中，便可廢除測驗了，因為教學和測驗應當時常繼續的交互在一起的。就基本的技術而論，最好完全在測驗中施行教學。有了算術練習測驗，圖畫練習測驗等等，這兩種教學便可在測驗中施行。讀法的練習測驗，現在還沒有發現，將來苟能編製成功，則學生對於規定的材料，誦讀後瞭解到何種程度，也不難於測驗中明白的表示出來。有效率的教師是常常要對學生的瞭解程度加以測驗的。

基本的學習心理學之知識，是必要的。但是單有這種知識，不一定就可以得到有效率的教學。因為把原理變為實踐往往是極難的。現在把教育測驗在聯結原理和實踐方面之三個重要的貢獻略說一說——練習測驗，非正式的測驗和標準的量表。

雖然練習測驗在初小和高小中都是用的，但是把初小用的練習測驗——克的斯標準算術練習測驗——說一說也已經夠了。這套測驗有四十八張卡片，就是四十八課。除第十三、三十、三十一和四十四課是供測驗用，第四十五、四十六、四十七和四十八課供學習用外，每課含有一種範圍的題目。每課由簡單的題目逐漸變為較難的題目，就是每增加一課，題目的困難就增加一步。學生習完這四十課，就可以說他對於加減乘除中所有的困難已有支配的能力。（俞子夷先生會仿照克的斯的算術練習測驗編了一種算術練習測驗，由商務印書館發行，讀者可先購說明書一讀。）使用這種練習測驗的主要步驟，大概如下：

(1) 用第十三課——這課包含第一至第十三課的各種困難——於規定的時間內測驗全級學生。

(2) 測驗完了之後，教師說出各題的答案，學生交換測驗卷加以訂正。

(3) 凡於本課及格的學生免習第一至第十三課。凡免習第一至第十三課的學生，可以學習別種功課，等到成績不良的學生也能及格之後，再一同學習第十三課以下各課，或竟參加第二步

測驗直接學習下去。

推。
(4) 凡於第十三課不及格的，從第一課起學習。第一課分數及格之後再學習第二課。如是類

(5) 凡於第一課學習不及格，仍須繼續學習第一課，至能及格爲止。

(6) 當學生學完第十二課時，再用第十三課來測驗他，意思就是證明他對於以前各課不是都是學會了。如果他於第十三課及格，他可以學習別的功課，等待別人，或竟單獨前進。

(7) 當全級百分之九十的學生（包含最初免習第一至第十三課的學生在內）在第十三課中都已測驗及格，就用第三十課來測驗。凡於第三十課測驗及格的，免習第十四到第三十課。不及格的從第十四課學起，方法同前。

(3) 教師每天把各個學生的作業記錄下來。時時注意有沒有欺騙的行爲。時時對於學生診斷，遇學生須要的時候應用校正的手段。對於全體學生鼓勵善良的工作。注意學生自己所做的記錄是不是準確。遇必要時，教師須把他們的卷子重行看過一遍，以提高他們記分的標準。

練習測驗使教學個性化 集團教學是很不效率的，尤其當教學技能科目時是這樣。因為集團教學對個個人都注意，結果各個人都沒有獲得益處。

阿米巴對於刺激的反應有三種。第一種積極的刺激如滿意的食料等等，阿米巴遇到這種刺激便起接近的反應。教師也用積極的刺激（如獎勵等等）來誘導學生趨向於工作的善良習慣。第二種是消極的刺激，阿米巴遇到這種刺激就起一種後退的反應。教師也用消極的刺激（如責罰等等）使學生不為工作的惡劣習慣所染。第三種刺激是中性的，阿米巴對於中性的刺激產生中性的反應，實際上等於沒有刺激，也沒有反應。做教師的極想他自己的一言一動都變為積極的或消極的刺激。但是在集團教學中，大多數刺激都是中性的刺激，因為他的刺激對於大多數學生不起任何反應，對於有幾個學生是消極的，對於更少數的學生纔是積極的。各個學生的個性是極不相同的，所以在可能範圍中，最好把個別教學來替代集團教育。而練習測驗就是使教學個性化的一種方法。

練習測驗能自動的把工作適應於各個學生。既以免掉重行教學。也可以免掉先時的教學。

例如用上述克的斯算術練習測驗第十三課測驗以後，凡及格的學生可以無須再學習第一課至第十三課，而不及格的必須從第一課學起。這就是練習測驗於個別教學中的特長。練習測驗還有一種特長，就是他允許各個學生依照他自己的工作方法和自己的工作速率工作，並幫助他去尋求最好的工作方法。

練習測驗使進步的意念加強。練習測驗中，有一定的目標並有測度進步速率的方法。每個學生多把自己每天的作業記錄下來並畫成種種表格以顯示自己的進步。這樣全體學生於學習過程中多有到達目標，增加速率的意念。彼此競爭的結果，進取之心自然加強，普通教學中雖也有分數以爲優劣之標準，但沒有確定的目標，學生於學習中自己也不知道學習到何種地步，有何種速率，所以一級中容或有競爭，但是屬一時的不確定的，所以學生求進之心不强。練習測驗使學生個個對自己負責，個個有超出自己過去記錄的決心，這又是練習測驗的長處。

練習測驗使練習的機會增多。照桑代客的學習的三大基本原則說，學生的意志堅強，能利用結果並有充分的練習機會時，學習的進步一定是很快的。講到練習的機會，普通教學中使學生

練習的機會不是沒有，如於課室中教師口問，由學生輪流口答，或指定一學生至講台上解決算術問題等等都是練習，可是於一時間內只有少數學生得到練習的機會。練習測驗對於各個學生備有全套的練習片（如克的斯算術練習測驗），各個學生都能依照他自己的能力，工作的速度來練習，這於時間的利用上比普通教學經濟得多了。

練習測驗使輔導和診斷便利。練習測驗對於須要扶助和診斷的學生給以極迅速的扶助和診斷，當不須要的時候，也不致給以無謂的徒耗時間的扶助和診斷。因為硬使學生努力而得不到進步，是使學生失望。失望到了某種時期，努力便完全停止，練習測驗中規定學生須自己記錄分數並把分數畫成進步曲線，教師看了進步曲線如果停止，就知道學生正當須要扶助和診斷的時候。

二 新式考試之應用

真假測驗 考試一向是占重要地位的，並且於教育上將永遠是占重要的地位。不過，舊式考試的內容形式和記分方法等等都不很健全，有改良之必要，這是教育家所公認的。教育測驗中的

真假測驗和標準量表都是舊式考試的改良方式的結果。雖然編製測驗的技術目前還不能說已經是極精，但是在許多方面如校閱便利，避免主觀，記分準確，節省時間，都不是舊式考試所能及。現在先舉一個真假測驗的例子於下，然後再討論怎樣計算分數，怎樣編製這種測驗，怎樣應用這種測驗等。

美國地理考試（本國已有地理測驗，為便利計，暫用麥柯爾的現成例子）

以下二十個句子有些是對的，有些是不對的。倘使是對的，你就在「對」的下面劃一劃；倘使是不對的，你就在「不對」的下面劃一劃。對於每句子都須打一記號。倘使你不知道，猜猜看。

1. 普通說來，美國的山脈是東西橫行的.....對的 不對✓
2. 大多數的河流是向北流行.....對的 不對✓
3. 密基山是密雪雪必河以東的最高點.....對的 不對×
4. 華盛頓山是比密基山高.....對的 不對×
5. 卡得斯基山是在梅因 (Maine)對的 不對✓

6. 卡斯開特山比洛基山近於太平洋.....對的 不對×
7. 洛基山比阿派萊張山近於太平洋.....對的 不對√
8. Blue Ridge 是在洛基山脈中間.....對的 不對√
9. 西部比東部多活火山.....對的 不對√
10. “Old Faithful” 是旋風的名稱,他是從 Texas 吹向 Oklahoma 的.....對的 不對×
11. “Grand Canyon” 是在 Cumberland 平原被 Subquehanna 河所割斷.....對的 不對√
12. Pike’s Peak 是在洛基山脈中的.....對的 不對√
13. 米雪雪必河是流入 Great Lakes 中的.....對的 不對√
14. Arkansas, Missouri, Ohio 都是米雪雪必河的支流.....對的 不對√
15. Big Sandy 是美國最大的河.....對的 不對×
16. 大西洋在東, 太平洋在四.....對的 不對○
17. 卡那大在南, 墨西哥灣在北.....對的 不對√

18. 大湖一總有五個……………對的 不對√

19. 在米雪雪必東部大湖中游泳比在米雪雪必西部大湖中游泳是容易沈溺……………對的 不對√

20. 美國中部是比東部或西部較平……………對的 不對√

怎樣計算真假測驗的分數 以上學生對對的是14個，劃得不對或做錯的是5個，沒有做的是1個。再照公式（甲）來計算：

公式（甲） 學生的分數 \parallel 做對的 $-$ 做錯的。所以學生的分數 $\parallel 14 - 5 \parallel 9$ 。

學生的分數爲什麼是等於做對的題數減去做錯的題數？我們現在分幾層來解釋。假使一個學生對於上述美國地理完全不知道，而參加這個測驗，並對每句子都打上一個記號，則依照機會說起來，他劃對的是十個劃得不對的也是十個。換句話說，他的猜對或猜錯的機會是五十對五十或一對一。所以他於上述測驗中的分數是：

$$\text{分數} = 10 - 10 = 0$$

學生的知識是零，計算他的分數的方法也使他得到零的分數。又假如一個學生知道句子中的十

句，而對於其餘的十句亂猜。亂猜的十句中，依照機會起說來，將有五句猜對，五句猜錯。這樣共有十五個對的，五個錯的。但是依照公式甲算起來，他所得的分數仍舊只能相當於他所確實知道的句子數：

$$\text{分數} = 15 - 5 = 10$$

任何學生，對於每句句字都劃對了，則得滿分，就是：

$$\text{分數} = 20 - 0 = 20$$

以上的計算法，是只管學生做對的和做得不對的題數，對於學生省略未做的題數是不問的。這種計算分數的方法是適用於測驗的時間極短，測驗的性質多少含有速率的意義。倘使測驗的時間規定得極長，就是最慢的學生也可以做完，又倘使規定了學生對於每句句字必須個個作答，則依照下列公式計算分數較為便利：

$$\text{公式(乙)} \quad \text{分數} \parallel \text{句數或問題數} - 2 \text{ (做錯的題數或句數)}$$

倘使測驗問題是二十個，又倘使做錯的是五個，則

$$\text{分數} = 20 - 2(5) = 10$$

依照公式甲計算，所得分數也是一樣的：

$$\text{分數} = 15 - 5 = 10$$

只要學生對於測驗中的句子或問題沒有略去不做，公式甲和公式乙所產生的結果是相同的，看下列的推算便可證明：

T = 總題數或總句數 R = 做對題數或句數 W = 做錯題數或句數

則分數 = $R - W$ (公式甲)

$$R + W = T$$

$$R = T - W$$

代入公式甲

$$\text{分數} = T - W - W = T - 2W \text{ (公式乙)}$$

真假測驗之編製 這種測驗的編製頗容易，無論何種科目都可應用。問題或句子的數目以

多爲宜。不過須注意下列各點：

(1) 對和不對的問題或句子的數目須差不多相同。對和不對的問題或句子的先後順序須隨機遇，不得呆板。倘使呆板，則學生容易猜中。

(2) 避免含糊的或意義不明顯的問題或句子。

(3) 避免有暗示性的問題或句子。

(4) 避免不重要的問題或句子（倘使用無關重要的問題來測度，學生於學習上容易養成不正當習慣。）

(5) 避免消極的字句。

(6) 務使問題或句子簡短揭要。

(7) 不要使下一問題或句子就是上一問題或句子的答案。

真假測驗之施行法 這種測驗的實施，最好是把測驗問題印刷或用複印機印成許多卷子，使每個學生都能有一本卷子。如果辦不到印刷或用油印機來印行則以下幾種方法，可以任意採

用一種。第一是使學生把測驗問題逐字照抄，然後作答。這是最勞苦最費時的方法。第二教師把測驗問題寫在黑板上面，每問題之前，註明數目字如1 2 3 4之類。叫每個學生用一空白紙，紙上也寫1 2 3 4等數目相當於問題的數目，問題的本身不必抄寫。學生做的時候只須在相當題目下寫「對」或「不對」。這個方法的不便利之點，就在於問題極多時不能完全寫在黑板上面。第三，叫學生依照問題的數目將1 2 3 4等寫在空白紙上。教師口說問題時，學生就打記號。例如教師口說第三問題時，學生就在3之下打記號，對的打一√記號，不對的打一×記號。這個方法比較上是便利的，不過也有缺點，就是有些學生不容易了解口頭說出的問題，當問題長而繁複的時候更難了解，並且當教師說過一遍或兩遍之後，即繼續說出別個問題，學生不能像對於普通試卷上的問題有充分思考的時間。

真假測驗之校閱 怎樣把這種測驗打分數，也有幾種方法。倘使學生用的是印現成的測驗卷子，則教師可用一空白或未曾用過的測驗卷，把準確的答案填上去，然後把準確答案的一行和學生答案的一行平行的擺在一起，兩相比較，自可很迅速的看出學生的答案是否準確，再於學生

的答案旁做相當的記號。倘使各學生所用的不是印成的測驗卷子，乃是於各問題數目字之後填「對」或「不對」，或打√或×記號，則教師可用和學生所用同大小的紙一張，再在紙上寫1 2 3 4等數目字，各數目字的間隔和學生所填的相同。這樣弄好以後，再在各數目字下填寫準確答案，把這行準確答案擺在學生的答案旁邊，兩相比較，這樣校閱起來也定可快些。不過這種校閱法是假定學生用的紙是有格線的。倘使教師沒有充分時間，則使學生自己校閱或彼此交換校閱，教師只須口說準確答案。當卷子完全校完之後，須把各問題或句子統計一下，看各問題或句子做對的有幾人，做不對的有幾人，省略的有幾人。這步工夫做了之後，便可對學生做種種指導。以後如再舉行測驗，則可用來比較，看各學生是不是進步。所以這是雖然較為費時，卻是必須做的。

真假測驗之優點 我們究竟為什麼要使用這種測驗？這種測驗比普通考試方法的優點是在什麼地方？第一，真假測驗所含的問題較多，學生學習範圍以內的事物，差不多可以完全代表出來，所以對於他的量度是比較的公平。而普通考試所選的題目太少，學生於考試之前往往對於教師將出些何種題目起一種疑慮。有時候，學生且以為自己分數的多寡和平時對於教師的感情有

密切的關係。真假測驗中，學生的成績值一分的就得一分，值二分的就得二分，學生就有一種公平的信仰心。又因為題目多，表示能力的機會也多，學生對於這種方式的考試樂於參加。

就教師而論，他對於這種測驗，尤為樂意。因為校閱卷子起來容易，如果由學生自己校閱，更快。在通常的考試中，學生不能幫助教師校閱卷子。由教師校閱四十本冗長的試卷，真是費時而又苦悶。一則要集中注意來辨別試卷的真好壞，再則要留心給分的不要過寬或過嚴，情緒免不掉要緊張。

真假測驗更含有教育的意味。學生於測驗之後，興趣非常之濃。人人都想知道自己在什麼地方是省略了，在什麼部分是做對了，在什麼部分是做錯了。不瞭解的自然逐漸會瞭解，錯誤的自然會得更正。全部的學習上自可有長足的進步。在教師方面，當測驗之後，也可以從此知道自己的教學效率是怎樣。對於各個學生的能力和困難的地方都可以了然。例如測驗的題目共有二十個，而全體學生的平均做對的題數是十個，則教師的效率可說是只達到百分之五十。知道了這種情形，教師便能自動的改正自己於教學上欠缺的地方。

這種真假測驗對於讀法教學特別有用。試舉一例於下，以明怎樣應用：

把第四級學生（表一）分爲兩組，每組的讀法能力是同等的。要使這兩組能力同等，就於開始的時候舉行讀法測驗（用標準的讀法測驗）。測驗之後，依照學生所得分數之多少列爲一表，最多的在前，最少的在後。把第一名歸入第一組，第二名歸入第二組，第三名歸入第二組，第四名歸入第一組，第五名歸入第一組，第六名歸入第二組。如是類推下去，便可形成讀法能力實際上差不多同等的兩組。

這樣組織成功之後，教師對兩組施行讀法教學而在每星期之末舉行讀法測驗。測驗問題的內容可以從正式讀本中選擇，有時候也可以從他種歷史教科書或地理教科書中選擇。不過這種測驗須含有使兩組競爭的性質。讀法之速率與瞭解兩方面能雙方兼顧最好。爲求明顯起見，再舉一種省時而又含有教育意味的測驗的編造法於下。

第一步，從讀本中選擇兩面文字，這兩面文字是有代表性的（可以代表同一讀本上之別面文字的）。假定選擇讀本上第十至第十一面。第二步就根據這兩面文字，作成二十句「真假」或

「是否」的問題。問題作成之後，就用以下的說明對學生說：

把鉛筆和紙拿出來。把你的姓名寫在紙的上部。把你的讀本拿出來，看第九面。我朗讀第九面的最後一段。你們跟了我讀。我讀完最後一個字的時候，你們就繼續默讀下去。要儘你們的速率讀，同時注意書中的意思。待你們讀完之後，我要把幾個問題問你們，看你們對於剛纔讀過的記得多少。繼續讀下去的是第十和第十一面。讀完第十一面最後一個字的時候，把書合起來。看黑板上，把黑板上的數目字抄在你的紙上。

要測驗學生讀時的速率，最好用碼錶一只。於讀完第九面末段最後一字時，把碼錶開動。學生繼續讀下去時就隔十秒鐘把10寫在黑板上。再隔十秒鐘，把10擦去，寫20。再隔十秒，把20擦去，寫30。這樣，一直到全體學生通通讀完為止。

全體讀完，把書本合上之後，對他們說：

把1到20從上而下抄在你的紙上的左邊。關於你們剛在所讀過的，我要問你們幾個問題。每個問題用「是」或「否」來回答。倘使你不知道或答不出，猜猜看。倘使對於第一問的答案是「是」，你就在1的旁邊寫一「是」字，倘使是「否」，就在1的旁邊寫一「否」字。其餘的問題，都是這樣回答。

倘使學生不能書寫，用記號也可以。如用 \surd 以表示「是」，用 \times 以表示「否」。對於年紀過輕的學生，不妨先做一次練習。

問題要慢慢的讀出來，要響而又清楚，使得每個學生有充分的時間來作答案。

學生由記憶上把測驗問題做完之後，叫他們把答案紙翻過來，在上面再寫1到20等數目字。寫好之後，叫他們把書本展開到第十到第十一面。教師把測驗問題再逐個誦讀一下，看學生當書本展開的時候能夠對答多少問題。

教師把準確的答案讀出來，使學生自己校閱或彼此交換校閱測驗卷，並核算各人分數。校閱的時候只須把做錯的打個記號。省略不做的作為做錯的算。各人的分數等於總題數中減去2乘做錯的題數。例如學生甲在第一部分測驗中做錯六題，在第二部分測驗中做錯三題，他的瞭解分數就是：

$$\text{記憶上瞭解分數} = 20 - 2(6) = 8$$

$$\text{視覺上瞭解分數} = 20 - 2(3) = 14$$

把兩面書上的總字數告訴學生，使他們各自計算每分數讀了多少字。這就是學生的閱讀速率分數。

使各學生保留每星期自己的分數的記錄，以便使他們自己知道於閱讀速率上有沒有進步。這種速率的測驗必然是極粗率的，因為每星期所用的測驗材料，於難度上多少總有些變動。

把記憶上瞭解分數和視覺上瞭解分數，分別算出一平均數。再把閱讀速率分數核算一平均數。教師應當保留這三種分數，作為觀察學生於他的教學之下有沒有進步的參考。

一個第四級學生對於這種非正式的閱讀測驗的結果，是見第六表。

結果之利用 第六表的分數，顯示以下有用的結論：

(I) 學生 A, E 和 H 於讀法是極壞。學生 E 簡直沒有成績可說。這三個學生於讀法的各方面都是有缺陷，容或是由於訓練不充分，容或是由於訓練不適當，容或由於天賦智力低以及別種原因。至於究竟是不是由於訓練不充分，只要加以訓練之後看他有沒有進步，或調查他以前受教育的情形，便可定奪。是不是由於訓練不適當，可以從給與適當訓練之後，測驗他有沒有進步來決定。

六 表

第四級學生對於非正式的閱讀驗分數

學生姓名	記憶瞭解	視覺瞭解	速 率
A	2	6	125
B	5	8	127
C	7	12	150
D	5	16	121
E	0	0
F	4	18	200
G	7	10	150
H	0	2	100
I	13	16	131
J	6	12	146
K	8	10	150

L	18	20	180
M	15	16	172
N	10	12	146
O	9	12	150
P	4	8	104
Q	10	14	140
R	7	12	140
S	12	14	166
T	6	10	127
平均	7.4	11.4	136.3

是不是由於天賦智力的不足，則最好應用標準智力測驗來決定。

(2) 學生D的記憶的瞭解很不好，但是他的視覺的瞭解程度還高。他的讀法的速率表示他記憶瞭解上成績的不好不是由於讀來過快。

(3) 學生 F 的記憶的瞭解極低，而視覺的瞭解則極高。這大概不是由於天賦的記憶力薄弱，乃是由於讀來太快。就他所得速率分數論，他是全級中讀得最快的。做教師的，應當把怎樣支配速率的方法來指示他。

(4) 就記憶的瞭解和視覺的瞭解論，學生 I 是個極好的誦讀者。但是他的速率並不特別高，他於這方面須加訓練。

(5) 學生 L 和 M 於讀法的各方面，成績都極好。他們讀來很快。他們有極強的記憶。他們的視覺的記憶——這是讀法中最重要的一——也是同等優良。

(6) 除第六表所載的記錄外，別的記錄也是有用的。例如將每個問題，做錯的有幾人，做對的有幾人，沒有做的有幾人列為統計表，這樣於教學上也有許多用處。

三 標準測驗之使用

決定最後讀法年齡和心理年齡 學年終了時的學生的最後的讀法年齡和心理年齡，應當預先估計好。智力商數就是預先估計每個學生於一學年末（假定一學年為十個月，即為受十個

月教學之後）應當有怎樣的讀法年齡或心理年齡的最好的幫助。因為智力商數是一個學生於讀法上進步能量的準確的指數。如果沒有智力商數，用讀法商數也可以，不過不如智力商數來得準確。我們對於智力商數九十的學生，在十個月中的進步，只能希望他達到智力商數一百的學生的百分之九十。所以估計一個學生最後的讀法年齡方法是極簡單的，只要把十個月的百分之九十（倘使一學年作十個月算）又該生之智力商數是九十，加到他的開始的讀法年齡上去。倘使一學年作八個月算，學生的智力商數是九十，只要把八個月之百分之九十加到他的開始的讀法年齡上去。學生A的估計的最後的讀法年齡是121加十個月之百分之百，就是131。試參觀表一第八行。同樣，他的估計的最後的心理年齡是121加十個月之百分之百，就是131。試參觀表一第九行。學生B的估計的最後的讀法年齡是130加十個月之百分之122，就是142。他的估計的最後的心理年齡是150加十個月之百分之122，就是162。不過這些計算是假定一學年是十個月的。

對於像閱讀一類含有複雜的機能關係的科目，用智力商數來做預估進步的基礎，是比較可

靠的。但是，對於書法等機能關係極狹隘的科目，用智力商數來估計進步，則不甚可靠。最好是用書法量表求書法商數。

決定年終目標 每個學生的讀法目標應當就是他的估計的最後的讀法年齡，或是他的估計的最後的心理年齡。通常說來，用估計的最後讀法年齡作爲最小限度的目標，比用估計的最後心理年齡來得公平滿意——無論是對於學生或教師。因爲用估計的最後讀法年齡做目標是要求每個學生依照他的能力產生常態的進步。倘使他在過去時期中已經表現超過常態的進步，則以後將永遠保持這種進步。如用估計的最後心理年齡作爲目標，則這種進步是不會保持的。更有進者，用估計的最後心理年齡作爲目標，將使成業商數在一百以下的學生（例如表一學生B）在一年以內做極艱難的工作以彌滿他在成業商數上的缺點。最後，用估計的最後心理年齡作爲目標將使學年短的（即一學年在十個月以下者）學校加重工作。所以讀法之最小限度的目標，以用各學生的估計的最後的讀法年齡爲妥。

用估計的最後讀法年齡作爲目標之後，應當依照讀法量表轉變成一種分數。表一學生A的

估計的最後讀法年齡是13.1。依照表四，得T43。所以他的最小限度的目標是4.3。同樣，學生B的最小限度的目標是4.7。其餘學生的最小限度目標，見表一第十行。

一級的最小限度的目標，就是各學生分數的目標的平均數，例如39.8（參觀表一）。

通常，學校中各科是沒有明顯的目標的；或表面上有確定的目標，而實際上是一種不能到達的理想。這兩者的結果都足使學生對於進步表示「淡漠」的態度。有了這種一定意義的目標，學生纔知道自己走的是什麼路，要達到的是什麼地方，自然進取的心理格外強。

把最近和最後的目標劃為格線圖。各個學生不但須有一年的目標，並且須有逐月的目標。

最好使學生把逐月的目標於格線圖 (Graph) 中表示出來。桑代克、麥柯爾的讀法量表是有逐月的目標的，所以各學生可以把全年目標作十等分（假定一學年是十個月）或五等分（假定一學年是五個月）。表一學生A的開始讀法分數是4.0，他的估計的最後目標是4.3。他一定要在十個月中獲得3分，每個月便須得一分之十分之三。就一個人而論可以畫這種格線圖，就一級或全校各級而論，也可以用一較大的紙畫出這種格線圖，使各人或各級對於自己的目標都可一目

了然。例如表一四年級生，他的開始讀法分數是36·4，估計最後目標是39·8，則畫起逐月目標的格線圖來應表示36·4，36·7，37·1，37·4，37·8，38·1，38·4，38·8，39·1，39·5和39·8。一級目標的格線圖，可以貼在一級的教室之中；全校目標的格線圖，宜貼在比較人人可以注目的地方。

用標準讀法量表在每月末舉行測驗，教師應當在每月末舉行標準讀法測驗。應當鼓勵學生看得每月的測驗是一種重要事情，個個人要努力去到達或超過下月的目標。各測驗卷子記分之後，各學生或各級的結果應當用圖表示出來。

倘使一個學生或一級學生，測驗之後，表示失敗，則教師和學生雙方都須想法尋出失敗的原因。據麥柯爾說上面所述的逐步計畫會施行於一個小學，結果三年級生於第一月之末超過了年終目標。四年級到達了年終目標四分之三。五年級沒有任何進步。六年級和七年級，進步極好。八年級和五年級一樣，沒有任何進步。或由分數之單位太不精確，或因測驗本身之不可靠，各學生，以個人而論，往往表現沒有進步，但就全級而論，往往不是沒有進步的。

討論或練習：

- (1) 你學校中有算術練習測驗片麼？如果有的，你怎樣應用應用的結果怎樣？
- (2) 你對於真假測驗記分的公式有沒有疑問？
- (3) 你班中定有各科的目標麼？是怎樣定的？

第五章 測驗和教學效率的估量

決定最後成業商數 把各個學生在最後標準測驗中所得的分數，轉變為一種讀法年齡。再把預估的最後的心理年齡來除這實際的最後的讀法年齡，所得的商數就是最後的成業商數。把開始的成業商數和這最後的成業商數相減，結果就是一年工作效率的最後估計。

表一學生A的最後讀法分數是43（見表一第十一行）相當於最後讀法年齡130（見表一第十二行）。同表第九行，表示他的計估的最後心理年齡是131個月。把這心理年齡去除他的最後讀法年齡，他就得成業商數990。這樣，學生A實質上已經到達了目標。學生B的成業商數，在這一年之中，增進得很顯著，足見他已超過他的最小限度的目標。然而，他的成業仍在應當到達的程度以下。學生C不但保持了他的高度的開始成業商數112，並且又繼續增高了6點。

就表一這級學生全體而論，全級開始的讀法分數之平均數是36.4。全級讀法分數之目標是39.8。全級實際的最後的讀法分數是41.2。這樣說來，這一級已很超過他的最少限

度的目標。又因爲這級開始的成業商數之平均數是100，而最後的成業商數之平均數是103，所以在這一級中，無論是教師或是學生都知道他們的成績記錄是一種可稱羨的記錄。

教授效率的估計 應用標準測驗於一級學生之後，所得的結果能不能表示這級教師的效率？這些結果是能够表示教師的效率的，不過須有一定的條件。而這些條件在一種實驗的控制之下是可以得到的。我們也可以說這些結果不能表示教師的效率，因爲對於教師效率之公平的估計上之必要條件，於不加控制的測驗情形之下是得不到的。

應用測驗來指導並診斷學生比用測驗來評量教師來得重要。所以在某時期內，最好用測驗來測度學生，不是測度教師，除非做教師的願意用測驗來測度他們自己的效率，或於測度他們的效率時自願誠心合作。當測驗發展到確能公平的估計教師效率的時候，真實有效率的教師自會要求用測驗來衡量他們的效率，而不願意用別種不準確的或含有主觀性太重的方法來衡量。

據麥柯爾的見解，評量教師或視學的效率，並把他們升調時須有以下幾種基本的假定和步驟。

假定：

(1) 學生是教育的系統中的重心或太陽。教師是這太陽的衛星。視學是衛星中的月亮。

(2) 教育上一切附屬品或點綴品之存在都是爲了一個目的：使學生能表現我們所希望於他們的變化。

(3) 這些附屬品或點綴品的價值的高下，能用一個方法來決定：看他們使學生產生所希望的變化能有多少。

(4) 所以選擇或升調教師時的公平基礎是學生所起的變化。

(5) 科學的測驗之本身如果要是公平的，必須要測驗某特定教師下學生所起變化之量。而對於變化之量度，須要兩種測驗：開始的測驗和最後的測驗。

(6) 科學的測驗如果要是公平的，必須要測驗於標準時間內學生所產生的變化之量。

(7) 科學的測驗如果要是公平的，必須要測驗標準學生所產生的變化之量，成業商數就是把學生——不管他的智力是怎樣——轉變爲標準學生的一種方法。

(8) 科學的測驗如果要是公平的，必須是完全的。絕對的完全，不但須測驗兒童於能力上所起變化之量，並且還要測驗他們於意志上所起變化之量。絕對的完全，自然是不可能的，事實上也是不必要的；因為對學生所起的變化中取其一種標樣 (a chance sampling of the changes) 已足夠代表，又因為教師教授一種科目的技能和教授他種科目的技能是積極相關的，例如於教讀法時能使學生起所希望的變化，於教算術時也能使學生起所希望的變化。

滿足了以上的假定或條件，進而估計教師效率（以讀法教學的效率為例）的步驟是：

- (1) 決定某教師所任全級學生開始的讀法分數（表一第二行。）
- (2) 把開始讀法分數轉變為讀法年齡（表一第三行。）
- (3) 於開始的讀法測驗舉行時，對於各學生的心理年齡也決定一下（表一第五行。）
- (4) 用實際年齡去除心理年齡得智力商數（表一第六行。）
- (5) 用心理年齡去除讀法年齡得讀法的成業商數（表一第七行。）
- (6) 估計學年終了時的最後的心理年齡（表一第九行。）

(7) 決定學年終了時的最後的讀法分數(表一第十一行)。

(8) 把最後的讀法分數轉變為最後的讀法年齡(表一第十二行)。

(9) 用估計的最後的心理年齡去除最後的讀法年齡得最後的讀法的成業商數(表一第十三行)。

(10) 從最後的讀法成業商數之平均數中減去開始的讀法成業商數之平均數(表一第十四行)。

倘使相減的結果，差數是零，這個教師就是有效率的。倘使差數是在零以下，那末他的效率就在平均或普通以下。倘使平均的差數在零以上，那末就是表示他的讀法教學的效率高。

(11) 把 1, 2, 5, 7, 8, 9 和 10 各步驟應用於算術科，所得的平均成業商數之差數便可以表示算術教學的效率。至於教授別種科目的效率，也可以同樣決定。

(12) 把各種科目的平均的成業商數之差數再求一平均數，便可以決定一個教師之最後效率。所以教師的效率可以用一公式來表示：

$$\text{教師的效率} = \frac{\text{(讀法成業商數之差數)} + \text{(算術成業商數之差數)} + \text{其他}}{\text{科目數 (或測驗數)}}$$

就表一所示的例子論，教師的效率是：

$$\text{教師的效率} = \frac{103 - 100}{1} = +3$$

這個公式不但可以應用於教師之分等，對於試教半年或一年後之教師加以選擇時，也可以應用。這種選擇教師的方法仍然是粗率的，但是比較通常的選擇法是公平得多了。

學習效率之估計 學生甲是個比較愚笨的兒童，但是他的父親並不知道。教師也不知道他是愚笨，只以為他是懶惰。兩年以前，學生甲是和同等年齡的學生同在一級中學習。因為他的智力低，常處於一級中劣等生的地位，最後由教師把他降級。他的父親接到了這個降級報告之後，對於他常取一種嚴厲管教的態度。

降級之後，學生甲剛巧和年齡更輕，智力更低的學生在同一級中。到了年終，學生甲的父親接得成績優良的報告。學生甲的教師覺得快樂，因為他以為他的教法比學生甲的前任教師是好得

多了，所以能成功。學生甲的前任教師也覺得快樂，因為他以為他把學生甲降級，學生甲纔走上了德性改革的路途。學生甲的父親也覺得快樂，因為他以為他的兒子的學習動機是由於他的適當刺激所激起的。學生甲自己也覺得快樂，因為他不像前一年的那樣不快樂。

今年學生甲是和智力比較優秀的兒童在一起，所以競爭的結果時常失敗。他早已知道同級的學生成績比他好得多，他自己的及格不及格早已置之度外，所唯一憂心的是身體方面的受鞭撻。這樣每隔一年，他是免不掉受身體方面的責罰，一直到了他十分成長之後。

學生甲不過是一個例子。學生中和學生甲有同樣情形的真不知有多少。這些學生每年受不公平的處置，都因為學校所做報告的不妥當。

至於智力上比較優秀的學生，因不公平的處置而所受的損失，恐怕比愚笨的學生所受的損失更要利害，不過這種損失另屬一種範類。大多數優秀的學生，都是在他們最好的學習效率水平線之下學習，換句話說都沒有達到他們的最高的效率。做教師的對於這些優秀兒童的、真真的學習或工作的可能性並不加上一點考慮。所以這些優秀兒童時常毫無困難的獲得優良的成績，得

到種種的獎品。但是除此以外，並沒有一點動機來提高他們的進步的速率。

怎樣的報告，纔是一種公平的報告呢？成業商數不但可以表示全校的效率，並且可以表示一個學生依他自己進步的可能性所已到達的進步範圍。所以成業商數是成績報告中一種必要的項目，應當把他列入報告中送達於學生的父母。

再這種學習效率的量數，對於校長是很有用處。因為做校長的於全校各個學生的學習狀況詳細知道之後，於各個學生的教學的視導以及全校的訓練問題等自能措置裕如。倘和學生的家長會談時，也能把公平的事實表示出來，使吹毛求疵的家長可以誠心悅服，使願意合作的家長不致得不到一點有用的暗示而去。

教學方法和材料之實驗的選擇 學生是教育世界的重心，一切努力都是為學生而開始為學生而終結的。所以教育上一切理論，一切主張，一切設施以及一切從事於教育的人，都要看他們能不能使學生表現善良的變化而定價值的有無，例如教師咧，校長咧，督學咧，教學方法咧，教學材料咧，師範學校咧，教育哲學家咧，諸如此類，有沒有價值，都要看他們能不能使學生的神經紐起變

化。如果他們都不能使學生產生合意的變化，則他們於教育上可稱是沒價值的。教育測驗是很顯著的，因為他必須表示各種事物之教育的效率，最後又必須表示自身有沒有價值。所以他含有自我破壞的性質，他的價值和別種事物的價值一樣，完全要看他使學生產生變化的量和價值而定。

但是，我們假定以上種種教育設施的教育價值在零點以上，或假定他們總是有些教育價值的，則第二個問題又來了：就是兩種教學方法，究竟那一種比較有效率？兩個教師或兩種教科書究竟那個或那種有效率？兩種組織究竟那一種有效率？如前所述，關於這些問題之公平的解答，必須要一種措置得極鄭重的科學的實驗。

以下舉示一種最簡單的教育實驗——同等團體實驗——上的理想的條件。

(1) 舉行絕對準確的開始測驗，選擇兩組絕對同等的學生。這兩組學生不但平均起來能力上是同等，並且要各個比較起來也是同等。就是甲組裏有這樣一個學生，乙組裏也必有這樣一個學生。各組的人數愈多，則兩組間同等的程度愈高。就理論上說，開始的測驗一定要是絕對透澈的測驗，實際上不過是要準確的測度某種教學方法，教學材料，或其他實驗動因，使學生能力上所起

的變化。

(2) 在實驗的全時期中，除實驗的動因外，其他一切實驗時的情形兩組間必須絕對相同。例如實驗的目的在決定兩種教學默讀的方法那一種是比較的有效率，則除甲組用方法甲教學，乙組用方法乙教學外，其餘兩組間的各種情形必須保持相同。

(3) 對於各個實驗動因必須準確的保持所希望的強度。

(4) 絕對準確的測驗各組學生之最後能力。

(5) 對於甲組學生所起的變化和乙組學生所起的變化，做一個澈底的公平的總價值的估計。

(6) 參酌兩組被實驗學生究屬何種範類，並考慮兩種實驗動因的強度之後，再下最後的結論。

(7) 倘使對於兩組實驗時的一切條件以爲是未能絕對相同，必須把結論之統計的可靠度報告出來。事實上，任何實驗，既得結論之後，必須報告結論在統計方面之可靠度。

學校和學區的效率測驗 以上所說是估計效率時比較精密的方法，是適用於小範圍的研究。至於學校或學區全部效率的調查，不得不用比較簡單或比較迅速的方法。下列第七表就是顯示這種效率測驗之比較簡單的方法。

本表的意義是這樣 一九一九年用桑代客的讀法量表第二類測驗 X 學校第三年級，得中數 5.5。而這級的常模是 8.0。一九一九年之中數和常模之差是負 2.5。這級的年級單位是負 0.6。換句話說，這級是在常模以下 2.5，或在常模以下一級之 0.6。第四年級是在常模以下一級。第五、六、七、八各級是各在常模以下一級之 0.5, 0.2, 0.8 或 0.9。若把各級平均計算，則在常模以下一級之 0.7。別種測驗分數的解釋，類推。至於各級在各種測驗中的總表七 顯示用各種測驗測驗各級之後 X 學校的效率和常態效率之比較

桑代客讀法量表第二類

	III	IV	V	VI	VII	VIII	各級平均	六至八級平均
1919中數.....	5.5	10.8	18.0	23.0	24.6	25.9		

常模.....	8.0	15.0	20.0	24.0	28.0	30.0			
差數.....	-2.5	-4.2	-2.0	-1.0	-3.4	-4.1			
年級單位.....	-0.6	-1.0	-0.5	-0.2	-0.8	-0.9	-0.7		-0.6

屈拉別開來填字測驗

1919中數.....	16.5	29.0	38.5	42.5	48.3	55.5			
常模.....	18.5	25.0	30.5	35.0	40.0	43.5			
差數.....	-2.0	4.0	8.0	7.5	8.3	12.0			
年級單位.....	-0.4	0.8	1.6	1.5	1.7	2.4	2.3		1.9

桑代客字彙量表

1919中數.....	21.5	51.5	81.5	106.5	108.5	109.5			
常模.....	30.0	65.0	83.0	95.0	108.0	117.0			
差數.....	-8.5	-13.5	-1.5	11.5	0.5	-7.5			

年級單位……—0.5 — 0.8 — 0.1 0.7 0.0 — 0.4 —0.2 0.1

衰亞拼法測驗

1919中數……13.5 21.5 30.5 47.5 48.3 55.5
 常模 ……………19.6 30.4 37.8 47.7 50.3 54.4
 差數……………—6.1 — 8.9 — 7.3 — 0.2 — 2.0 1.1
 年級單位……—0.9 — 1.3 = 1.0 0.0 — 0.3 0.2 —0.6 0.0

華狄加法量表第二組

1919中數……7.1 12.6 12.0 14.0 15.5 15.8
 常模……………9.0 11.0 14.0 16.0 18.0 18.5
 差數……………—1.9 1.6 — 2.0 — 2.5 — 2.7
 年級單位……—1.0 0.8 — 1.1 — 1.1 — 1.3 — 1.4 —0.9 —1.3

華狄減法量表第二組

1919中數.....	0.7	7.3	8.8	10.0	12.2	12.8		
常模.....	6.0	8.0	10.0	12.0	13.0	14.5		
差數.....	-5.3	-0.7	-1.2	-2.0	-0.8	-1.7		
年級單位.....	-3.1	-0.4	-0.7	-1.2	-0.5	-1.0	-1.2	-0.9

華狄乘法量表第二組

1919中數.....	3.9	7.5	8.7	10.7	14.1	12.5		
常模.....	3.5	7.0	11.0	15.0	17.0	18.0		
差數.....	0.4	0.5	-2.3	-4.3	-2.9	-5.5		
年級單位.....	0.1	0.2	-0.8	-1.5	-1.0	-1.9	-0.8	-1.5

華狄除法量表第二組

1919中數.....	3.4	5.4	4.9	8.0	10.1	9.8		
常模.....	3.0	5.0	7.0	10.0	13.0	14.0		

差數.....	0.4	0.4	-2.1	-2.0	-2.9	-4.2			
差級單位.....	0.2	0.2	-1.0	-0.9	-1.3	-1.9	-0.8	-1.4	

南沙那作文量表

1919中數.....	2.5	3.0	4.1	4.7	5.0	6.2		
常模.....	2.1	2.6	3.0	3.6	4.1	4.8		
差數.....	0.4	0.4	1.1	1.1	0.9	1.4		
年級單位.....	0.7	<u>0.7</u>	<u>2.0</u>	<u>2.0</u>	<u>1.7</u>	<u>2.6</u>	<u>1.6</u>	<u>2.1</u>
平均年級單位	-0.6	-0.1	-0.2	-0.1	-0.2	-0.3	-0.2	-0.2

平均是在常模以下一級之0.2。就六,七,八這三級而說,在各種測驗中的總平均也在常模以下一級之0.2。

至於何以要用年級單位和年級單位的計算方法,也須加上一點解釋。我們知道一九一九年中數和常模之差,在各個測驗間不能彼此比較。例如第八級在作文量表中所得的中數和常模之

差是 $1 \cdot 4$ ，等於 $2 \cdot 6$ 級；而第五級在字彙量表中所得的中數和常模之差是 $1 \cdot 5$ ，只等於一級之 $0 \cdot 1$ 。這樣，在作文量表上差 $1 \cdot 4$ ，是很重要的，因為從第三級到第八級全部常模之增進不過是 $2 \cdot 7$ （ $4 \cdot 8 - 2 \cdot 1$ ）。在字彙量表上差 $1 \cdot 5$ ，是不重要的，因為從第三級到第八級全部常模之增進是 $8 \cdot 7$ （ $11 \cdot 7 - 3 \cdot 0$ ）。所以在把各種測驗所得的差數結合之前，必須先把他們變為可以相互比較的差數。把一切差數變為年級單位的差數，便可以相互比較了。

所謂年級單位的計算是這樣的。把各測驗之各級常模增進之平均數去除中數和常模之差。例如桑代客讀法量表第二類規定第三級的常模是 $8 \cdot 0$ ，第八級的常模是 $30 \cdot 0$ ， 8 到 30 是增進 322 ，平均就是每級增進 $4 \cdot 4$ （ $222 \div 5$ ）。所以就平均說起來，一個代表的學生每升一級要增進 $4 \cdot 4$ 分。這個意思就是說，任何學校，任何年級，經過這個量表測驗之後，如果在常模以下 $4 \cdot 4$ ：就是低於常模一級或在常模以下一個年級單位。表七三年級的中數和常模之差是負 $2 \cdot 5$ ，用 $4 \cdot 4$ 來除，一個年級單位之負 $0 \cdot 6$ （即在常模以下一個年級單位之十分之六或低於常模一級之十分之六）。四年級的中數和常模之差是負 $4 \cdot 2$ ，用 $4 \cdot 4$ 來除，結

果差不多是負 1.0，就是差不多在常模以下一級，或差不多低於常模一個年級單位。其餘各級的年級單位的計算，類推。填字測驗之各級常模增進之平均數是 5，所以這個測驗都是用 5 來除各級中數和常模之差數。

效率測驗之解釋和機能 有了第七表的材料，我們怎樣來推究 X 學校的效率呢？我們可以從四方面來評量 X 學校的成績：(a) 各級在各個測驗中的成績，(b) 各級在全體測驗中的成績，(c) 全校各級在任何測驗中的成績，(d) 全校各級在全體測驗中的成績。

就各級在各個測驗中的成績說，未免太繁瑣，所以略而不說。各級在全體測驗中的成績，是見表七的底部。就是從第三級到八級，每一級，毫沒例外，都是在常模以下一級之 0.1 到 0.6。

這三年級的教師是否比別的教師比較的不效率，那是一個問題，需要特殊的說明。這三年級的學生容或是還沒有開始學習測驗中所測的某種能力。就算術的四則說，這級學生對於減法幾乎完全不知道，但是於除法卻離常模不遠。事實上四種基本方法都是已經教過，除法學會了而減法不懂，足見擔任這級算術的教師平時於教育上着重之點有不妥，須要用一種標準的算術測驗

來診斷各個學生於算術上何處須加練習，何處須加扶助。這一級的學生容或是智力太低，所以各種成績都不好，但是要適當解釋這一級在全部測驗中的結果，所以這樣壞，除根據測驗成績以外，還須對這一級學生的日常行動有種更密切的觀察。這個第三年級的教師容或是全校教師中最有效率的一個教師。

全校各級在各測驗中的成績是見表七最後的兩縱行。從這兩縱行的數目字，可以發見不少有意味的事實。就填字測驗說，全校各級差不多在標準或常模以上一級至二級。這種成績倘使和作文測驗（作文測驗所測的能力和填字測驗所測的能力差不多相同的）的成績不相調和，則我們可以說填字測驗的標準或常模定得太低。就讀法量表，字彙量表和拼法量表說，全校各級是離常模不遠或比常模低了一些。在算術四則測驗中，全校各級，毫沒例外，是在常模以下差不多一級。如果把測驗中所定的常模作為教學的目標，則X學校於下學年對於讀法和算術四則兩種科目須特別注意。

全校各級在全部測驗中的成績，也可以說是全校的效率，是見表七最後第二縱行的底部，就

是負0.2和負0.2（前一種是指三級至八級說，後一種是指六級至八級說。）全表的材料這樣總結之後，我們可以看出X學校各科的成績是比代表的或普通的學校低了一級之十分之二。二。我們對於X學校的其餘一切工作，如果也能加以測驗而平均起來，則X學校的效率也可以說是比普通學校的效率低了一級之十分之二。

教育是爲公衆服務的。他的本身的地位或價值必須從公衆的見地來評定。一切存在的機關須加以測驗的時機已經到了。負社會基金分配責任的人，對於他的代理機關關於金錢上的請求，一定詳加檢查，看他所請求的錢是不是用於值得使用的地方。學校是爲公衆服務的機關之一，自然也逃不出檢查或測驗範圍之外。而一個學校最後的價值之有無，要看學生是否起所希望的變化。如果學生表現所希望的變化之量來得多，學校的價值就來得多，如果沒有善良的變化，則這個學校的價值也就少。要量度學生的變化並作可以證實的結論，目前最好的工具，是標準的教育測驗。標準的教育測驗，是公衆和教師間雙方面的良友。因爲他估計服務於學校的人的效率或價值時，是不帶政治的色彩，是合於科學的方法；對於有效率的，保護他，沒效率的，放棄他。

話雖這樣說，使用教育測驗來估計學校效率的時候，如果不加一點普通的辨別力，也是容易發生弊病的。一個學校的成績可以和他自己定的目標相比較，如果他自己定的目標是很妥當的。但是如果要和別的學校相比較則必須參酌某幾種可能的動因後再下結論。以下幾種有勢力的動因，是評判學校效率時必須注意的：

第一種有力的動因是學校人口的持續性。若干年以前，美國有幾個大都市中出租的房屋多而要租屋居住的人少，房東為招來房客租賃房屋起見，每每貼補搬場費或免除房金一個月。有了這種情形以及別種原因，結果某幾個學校的人口年年要更動到百分之七十五以上。校中高年級的學生差不多全是從別校轉學而來，只有少數是從各該校逐漸升上去的。對於這種學校的學生施行測驗，差不多就是估計附近的學校的效率。概括說一句，一個學校的人口如果不固定，結果可以使學校的效率表示太低或太高或剛好。上述X學校的人口是屬比較的固定的。

第二種有力的動因是學生的智力。在某範圍中，學生智力的高下，可以從他們所處的普通社會地位來決定。就全體而論，在都市中不興盛或貧窮區域產生的兒童，平均起來，智力總比富裕區

域所產生的兒童要低些。但是用智力測驗來決定兒童智力的高下，是比用社會地位來得準確。依照社會地位的估計，X學校的學生的智力是在普通智力之下。這個結論是否準確，須用測驗來證實。我們如果應用成業商數那個公式，則智力高下的影響就可以免掉。

和第二種動因有聯帶關係的就是家庭環境之教育的價值——這是第三種有力的動因。在某種社會中，做父母的教育他們的兒女差不多和在學校裏一樣，而在別的社會中，家庭實際上阻礙了兒童的學習。

第四種有力的動因，是實際年齡上的停滯或超進 (chronological retardation or acceleration)。怎樣來措置這個動因使不發生影響，請參考桑代客、麥柯爾讀法量表的說明書。此地不再多說。

第五種有力的動因，是因成熟而起之能力增進。這個動因是不重要的，倘使我們測驗的目的只在測度比較的效率。倘使目的在決定何種進步是由學校的訓練而來，何種進步是由身心的自然成熟而來，則我們必須於總進步量中減去由成熟而起的進步量，這樣纔可以獲得由訓練而來

之絕對進步量。事實上這兩種進步從未有人把他分開。如果要把他們分別計算，必須有種極精密的實驗。

第六種有力的動因是能力之移轉。這於測驗只擔任一種或兩種科目的教師的效率時尤其要注意。因為學校中的普通訓練，或別個教師的特殊訓練，都能移轉到所測驗的能力上去。有許多特殊科目的教師，恐怕還不知自己所擔任的學生的逐年的進步，多少是受同事加於學生的普通訓練或特殊訓練的影響。

第七種有力的動因是教學上着重點的不同。一個學校，如果他所收受的學生有種種不同的社會的和職業的分配，則他選擇某幾種科目特別加以注意，於某幾種科目則不甚注意，這是合法的。我們對於這種學校選擇測驗來估計他的效率之前，最好是詳細參觀他的教學時間表以研究他的教學上的着重點，或和校長教師等面談，請他們陳述何種科目是他們理想中成績最好的科目，何種科目次之，何種科目又次之。這樣調查明確之後，再選定若干種測驗實地施行，庶不致使這個學校特別占便宜或特別吃了虧。

如果我們所選的測驗不祇一種，是有好幾種，這好幾種測驗又都是有代表性的，則測驗之後所下的結論，便不致於大錯。例如我們對於上述X學校效率的估計，如果只選用算術四則測驗一種，則我們所下的結論未免太嚴酷一點。因為這個學校是很顯明的着重於作文及和作文類似的科目，而不注重於算術四則和讀法等科目。一個學校如果於某種能力表示欠缺，於某種能力表示有特長，則我們對於這個學校，最好再另選幾種測驗，以測度他的平均成績，究竟能否到達規定的常模。但是讀法成績如果在常模以下時，則這個學校平均成績大概總在常模以下。因為讀法能力比作文能力更重要，他是許多種能力之鑰匙。從這幾種情形觀來，我們解釋測驗結果的時候，必須參酌或考慮到一個學校於教學上的着重點。

此外還有許多別的有力的動因，也是不應忽略的。例如一個學校的學年的長短。學年短，教學時間少，當然不能和學年長，教學時間多的學校相比。又如教師的訓練和學校的設備等問題，都是估量一個學校的效率時所必須考慮的。

杜威曾經提議過設置一種教育思想的交換局。標準測驗就是教育思想和教育實踐的一種

特別有價值的交通工具。一個學校表示他所正在進行的，是一件事；表示他所已經成就的，又是一件事。把一地方許多學校的成績和別一地方許多學校的成績兩相比較之後，便可以討論到何種能力是教育上最有價值的，或何種能力中有多少是教育上最有價值的。這樣就可以促進教育目標之科學的確定。

再學校和學校間的效率比較，可以促進教育之國家的概念。一國教育之通盤計劃之形成應當參酌或考慮到國內最不進步的學校的現狀或正在做些什麼事情。全國內那幾區於教育上確是停滯，停滯的程度又是怎樣。用教育測驗，於一定時期內對全國的教育情形舉行大規模的調查，可以使一國內最須要扶助的學校得到國家的扶助。教育測驗是全國教育調查最好的工具，一國教育標準之最低限度也必由教育測驗來確定。

討論或練習：

(1) 何謂實驗動因。

(2) 試提出一種或二種實驗動因，並用測驗來做最後結果之評量。

(3) 何謂學習效率？怎樣決定？

(4) 何謂教師效率？怎樣決定？

(5) 倘使你是教育局長或督學，請參考本章所述學區調查法舉行一學區的調查。調查之後，把發見的事實用統計法表示出來。再和各學校的校長和教員討論改良的地方。

