

麥柯爾教育測量法撮要

杜佐周編譯

1927

上海民智書局發行

麥柯爾教育測量法撮要

杜佐周編譯

上海民智書局發行

1927

序

近數年來，研究教育與實施教育者，頗具實驗的興味，講求具體的結果，教育已漸成爲一種科學矣。此誠教育前途很可慶賀的事！但既欲把教育當作科學研究，則必須先具有研究的條件。此種條件，大約可分爲三項：一、科學的知識，二、研究的工具，三、研究者的技能和『機械觀』。三者若缺其一，研究的結果，就不能真確；科學就不能成立。教育測量學不特是一種科學的知識，而且是一種研究的工具。研究者若有相當的技能和『機械觀』的態度，則應用之以解決教育的問題，比較教育的成績，及增進教育的效率，其功益當可與自然科學的實驗法相左右。我國中華教育改進社明乎這種科學之重要，數年前曾延聘美國麥柯爾博士（Dr. Wm. A. Mc Call）來華主任編制教育測驗事宜。當時氣象爲之一新；其成績亦頗有足觀者。至今教育界幾乎無不了解教育測量的意義，不可謂非昔日提倡的功勞。

可是遍索國內著譯界，尙無充分相當的書籍，足爲研究此種科學者的參考。此誠不無遺憾！即幸偶得一二種，但都略而不詳；或詳而不盡。非遍於理論，不切實用；即繁於事實，而缺理論。惟有麥柯爾博士所著的教育測量法（How to Measure in Education）一書，既詳且盡；富於理論，又切實用。舉例完備，解釋周詳。不特應用方面多所討論；且於編制方面詳有載述。實可謂教育測量叢書中的傑作。惜乎分量過多，文字繁雜；若譯時無相當方法損益而救濟之，則用爲教科的課本，不甚適宜！編者有鑒於此，前在武昌大學教授教育測量年餘，雖以此書爲課本，但平時每摘譯其重要部分，而遺其非要者，以爲學生的參

考；學生實稱便不少。因此且可減去分量大半；若每週三小時，則一年課程，適可完其全部。其中頗有增添修改處，由編者所見，原為必要。不知麥氏勿至怪他太武斷或多事否？至於舉例，如有國內事實可以應用，而易於明瞭者，亦無不添換之。故無論應用方面及編制方面，均可更切實用，易於倣效。如此，且可免除『不適國情』之弊。

此書倘若已有完全直譯者，則編者此作當為彼所歡迎。編者自信此作將原書的優點未嘗稍遺；同時缺點已多所補救。一方面可為直譯全書的大綱；另一方面亦可為了解全書的幫助。他若此書出版後，願有將全書直譯者，則又當為編者所歡迎；因為初從事於研究教育測量者，應用此作，必較為容易；但若欲為進一步的研究，則參考完全譯本，亦有他的功用。

此書共分為十一章。前六章討論應用教育測驗的方法；後五章解釋編制教育測驗的步驟。讀者固可因其興味與需要之不同，自為斟酌其研究的次序。至於統計部分，因國內已有相當的書籍，故略而不譯。然與本書內容有關係者，仍詳於附錄中。他若 B, C, F 分數的求法，為此書所未載，亦於附錄中補述之。惟這幾部分的材料，借用朱君毅寥世承及錢乃希三位先生的著作頗多；故特在此申明，且表致編者感謝他們的誠意。此書初稿成後，經學生趙演君詳細校讀，得其批評和幫助頗多；編者亦很感謝他。

夫教育經濟的問題，不僅當指金錢而言；且當在受教育者的本身着眼。其應利用科學的方法，實施教育，以免無形中莫大的損失，誠為當今實用教育家所急應提倡了；故編者特敢將此書介紹於

-
- 一、教育行政者，
 - 二、從事教學者，
 - 三、研究教育者，及
 - 四、指導選擇職業者。

倘讀者能應用本書所介紹的方法，為教育前途的改進，且多編制良好的教育測驗及量表，以供教育界的要求，則深為編者所期望的。

中華民國十六年，一月，三十日，杜佐周序於武昌。

麥柯爾教育測量法撮要目次

第 一 章

測量在教育學上的地位.....1

第 二 章

學生分班時應用的測量.....5

概論——分班的目的——用智力測量分班——用教育測量分班——變換年級標準分數爲年齡標準分數的方法——計算教育年及教育商數的方法——用教育年分班與用智力年分班的比較——教育商數與智力商數的比較——分班法與指定班次的表——重行分班的準則——升級及留級的分量——重行分班後的另外手續——較大學校內重行分班的手續——指定新生班次的方法——根據教者的評斷分班——計算教學年的方法——反對用科學測驗分班的理由之討論

第 三 章

爲診斷目的應用的測量.....32

診斷的計畫——診斷的方法——診斷上必要的知識及技能

第 四 章

教學上應用的測量.....42

練習測驗——非形式的測驗(普通考試)——正誤測驗的計算法——正誤測驗的編制法——應用正誤測驗的方法——校對正誤測驗成績的方法——正誤測驗的用處——正誤測驗的缺點——標準量表——教育目標在教育上的意義

第 五 章

評斷教學效率的測量.....66

計算年終成就商數，決定一年教學的效率——評定學習的效率——學校成績的測量

第 六 章

職業指導所用的測量.....76

職業指導的功能——職業指導的方法——智力與職業的關係——品性及體質與職業的關係——個人趨向與職業的關係——已有職業技能與職業指導的關係——興味與職業指導的關係——優秀兒童的職業指導——低能兒童的職業指導

第 七 章

測驗材料的準備及提高確度的方法.....86

理論的和實驗的提高測驗確度的方法——提高職業測驗確度的方法——提高智力測驗確度的方法——智力的準則——編制智力測驗的方法——普通測量智

力的方法

第 八 章

測驗材料的組織及測驗說明的預備.....104

對於測驗排列有影響的事項——測驗的形式——
機械的記錄成績點數的方法——團體測驗與個獨測驗
——做法說明的主要原理

第 九 章

等分測驗.....113

等分測驗的理由——百分表——編制百分表的方法——解釋百分表的方法——年齡表——編制年齡表的方法——年齡分數的解解——年級表——年級表的產生——年級表上P.E.值的恆度——成績比較表——愛禮斯的書法量表——海利葛斯的作文量表——評斷單位的確度及恆度——成績比較表的特性

第 十 章

T 量表.....131

量表編制的方法——測驗材料之預備——表列個獨的問題——正式測驗的選擇及排列——正式測驗的應用及記分——等分答對問題的總數——決定年齡標準分數——規定年級標準分數——量表的特別擴充——增加真確度的方法——量表的印刷——均衡法對於

麥柯爾教育測量法撮要

編制量表的功用——編制量表的簡便法——等分普通
考試成績的方法——量表編制方法的討論——參照點
——測量的單位——拼合記分單位的方法

第十 一 章

信度,客觀性,及標準分數的規定.....157

一個測驗缺乏信度的原因——增加信度的方法——
客觀性的重要——決定及增加客觀性的方法——影
響標準分數的價值之事項

附 錄 一

求平均數,中數,四分點差,均方差,及相關度的方法.....167

附 錄 二

求B, C, F分數的方法.....183

附 錄 三

比較速率及真確度成績的更正公式.....193

附 錄 四

核算實在年齡的方法.....197

附 錄 五

中英人名對照表.....198

附 錄 六

參考書報.....199

麥柯爾教育測量法撮要

第一章

測量在教育學上的地位

教育測量有一種特別的意義。欲知測量之可能，及其在教育學上的地位，我們須先討論下列諸論題：

論題一——凡存在的東西，必有分量。——此題之理，甚為淺明。現對於教育方面，則可略易其辭為：因為教育的功用，學生的程度有變化；這種變化，亦必有分量存在。即無變化，亦非這個論題之不能成立；實因教學不得其法，未得相當效率而已。反對測量者，每謂教育當注重質，不應專注重量。但天下沒有無量的質；故對這個論題是不能反對的。

論題二——凡有分量的東西，必可測量。——近二十年來，歐美各校，凡書法，讀法，作文，甚至智力及品性等，均已用標準測驗測量了。倘科學進步，即讀詩文及聽音樂等的欣賞情感，亦無不可測量。至於測量能如何真確，則是方法的研究；並非測量自身可能與否的問題。總之，凡教育的成績，均可測量。現在所有各種測驗，雖不能十分真確；但比較從前以浮泛的方法，斷定學生程度的優劣，必好得多。至於求測量方法的進步，乃是教育家公共的責任。

論題三——普通一般的教育測量，如天然物的測量一樣。——自然界的東西，用尺衡等器具測量，社會中習用之已久。今用測驗測量學

生的成績，其方法原相類似。所不同者，不過真確程度的關係而已。

論題四——即天然物的測量，亦非完全真確。——天然物測量的真確度雖比現在教育測量的真確度高；但亦非完全真確。不過其差甚微，常人不及見之耳。至若真確的測量如寒暑表等，實均是近代所發明的東西。今若謂教育的標準測驗是完全真確，固然不能；可是我們誠可積極改良之。現在往往有一種批評：每謂標準測驗所測量的範圍甚窄。然天然物的測量亦如此。惟因其如此，故於教學上有很大的用處。例如測驗國文，測其普通程度，雖是必要；測其部分能力，亦非教學上所可缺少的。

論題五——教育測量爲求教育學進步一種不可缺少的工具。——自然科學完全因有真確的測量，故有今日良好的成績。從前教育之不能成爲一種科學，實因沒有真確的測量。柏拉圖（Plato）說：『凡一種藝術，若無衡量爲之助，就沒有什麼的價值可言。』（“When weights and measures are left out, little remains of any art.”）故科學的進步，實有賴於真確的測量。現在教育當注重事實，不當僅注重理論；從前那種猜測武斷的態度，已經用不着了。教育哲學家可以提出關於教育的各種問題，同時教育測量家當用各種方法去解決這些問題。

論題六——教育測量的範圍，比較教育試驗的範圍廣。——依據事實而言，估計教學的品質，有人僅調查圖書館，實驗室，及教室等的物質設備；有人僅調查教者的資格及經驗；有人僅直接觀察教學的方法，學生的行爲；有人僅用普通的考試測量學生的成績。但我們對於各種測量的原理，均當討論；不僅指試驗一項而已。不過主要部分，

仍爲試驗；因此種測量甚有得到真確成績的希望。

論題七——教育除測量外，尚有他種應宜注意的事項。——教育有三種要素：學生，教材和方法，及目標是也。教者當知學生的心理，教學的目標，及應用的教材和方法。教育測量對此種種，均有相當的關係。

論題八——若是學生的能力不能測量，則欲確知他的個性，實不可能。——教者必須先詳細明瞭學生的心理，然後他可選擇相當的教材及方法，以求達到相當的目標。倘若僅知學生的普通能力，而不知其詳細情形，則亦無益；他必不能斷定這個學生究能學習何種學科，進入何種學校。故教者不能以僅知學生的普通能力爲能事；且當知其有某種能力的實在程度如何。愈真確，愈佳。教育測量乃是得到這種心理知識的必要工具。

論題九——若是教育的目標不明瞭，則教育就無意義。——我們論到教育的目標，須先答應三個問題：1、目標的價值是什麼？2、目標的地位在何處？3、學生是否向此目標進行？教育測量乃是解決這三個問題的必要方法。例如教學習字的目標，是使學生有寫字的能力。但這目標的價值，必須先測量寫字的能力對於社會上各種活動有何功用而後能定。他若目標無一定的地位，如寫字應當如何快，如何好等，則這目標的價值就不能定。欲救此弊，必須應用測量。至於欲知學生是否向此目標進行，則尤非測量不可。

論題十——若無測驗測量結果，則教材及方法的價值，亦無從得知。——教材及方法，實欲幫助學生向目標進行。至於是否有此功用，則

必須經測量而後能知。他若教材的選擇，方法的比較，亦非測量不可。故測量實是教育學上最重要的問題。教育雖有別種應注意的事；但非測量每不能顯示其價值。

論題十一——測量學生的程度，當先於教學。——因惟如此，教者可以觀察學生的成績，知其長短；然後酌量情形，採用適當的教學方法。如醫者必須先知病原，而後始能對証投藥，不至於亂。此實教學方面最經濟的方法。

論題十二——教育測量並非是新近的運動。——這種方法，很早已經應用了。薩拉門(Solomon)的格言，差不多都是智慧的測驗。中國從前科舉的方法，亦是測驗的一種。羅馬時代，為父者評定其子的文學程度，每以其在公會所中，能否誦讀碑上的羅馬法律為標準。原其用意，亦莫非是測驗。不過彼時的方法，完全是形式的，毫沒有科學的精神。至1894年，萊士(Rice)始發明用比較測驗(Comparative test)，測量各學校教學的成績，求得標準的分數(Norms)。科學的教育測量至此時始露其萌芽。及桑戴克(Thorndike)利用克德爾福來登的等距定理(Catell-Fulleton equal-distance theorem)，發明教育測量的量表單位，則就更為進步。此後施東的算學測驗(Stone's Arithmetic Tests)，桑戴克的書法量表(Thorndike's Handwriting Scale)，海利葛斯的作文量表(Hillegas' Composition Scale)，阜京海姆的拼法量表(Buckingham's Spelling Scale)，屈來阜的語言量表(Trabue's Language Scale)，胡德的算學基本量表(Woody's Fundamentals of Arithmetic Scale)，葛第斯的算學測驗(Courtis' Arithmetic Tests)，及愛禮

斯的拼法測驗(Ayres' Spelling Test)和書法量表(Ayres' Handwriting Scale)等,遂相繼出世。同時桑氏的學生相與研究編制教育測驗的統計方法,力求完善。至今這種運動,幾乎遍於全球。教學方面及教務調查方面,鮮有不應用之者。各國大學且設有特別專科或研究所,研究這種問題,及養成這種專門的人才,其發達可知。

論題十三——測驗沒有使教育或教育者成爲機械的危險。——倘若應用得宜,測驗實是評定教學方法優劣的最良好最公平的工具。不特可以不使教育成爲機械的東西,且可免除成爲機械的危險。至於教育者自己若不爲測驗所用,則亦必不至有這種弊病。

論題十四——教育測量並無呆板的弊病。——有人每謂測驗妨礙學生的個性;實則不然。無論何種作文,書法,或圖畫等量表,不過是測量的工具。學生可就其能力或程度,而與之比較,以求達到相當的目標;但非必欲倣效之。其實,這種方法很是利用個性;學生用這種方法學習,往往有很大的興味。

第二章

學生分班時應用的測量

一、概論 學生分班時應用的測量,有主要及次要的兩種:

1、主要的,又可分爲三種:如智力測量(Intelligence measurements),教育測量(Educational measurements),及教學測量(Pedagogical measurements)或考試的分數(Teachers' marks)。

2、次要的,亦有數種:如醫藥測量(medical measurements),及年齡

測量(Chronological measurements)等是。

我們所欲討論的，爲主要的三種。概用各科的平均數爲標準；非單就一科的成績而言。至於用各科測量而分班的方法，讀者如了解下面將所討論的問題，自可知之，必無困難地方。

分班的目的：

1、將學科程度相等的學生合爲一班。——若同程度的學生而合爲一班，進步大致可以一致；教者差不多可以用教一人的方法而教全班；因爲學生的需要必是大概相同的。如是，教者在一班之內，用多種教材及各樣方法那些困難，就可免除大半了。他若程度不齊的學生合爲一班，則教者所感受的困難必定很多。

2、將進步速率一樣的學生合爲一班。——欲達到這個目的，學校內須時常重行分班。若從前分班時只注重程度，而不注意進步的速率，則此種手續尤爲必要；蓋分班時，兩方面的事實均宜注意。今欲知學生進步之能否一致，則學生的興味，努力，體康等，亦均宜顧及之。

每科分班，似乎比較用各科的平均數分班好。但好處亦有限度，因小學的課程，每科又是分爲許多段落，這種段落，就學習心理方面言之，未必是十分聯絡的。例如歷史與地理間的差別，未必大於地理本身各段落間的差別，算加法的能力與算減法的能力的相關度(Correlation)，亦未必大於加法與文法間的相關度。

以上所說分班的目的，固甚合理；但此外尚有多種不合理的目的，我們亦宜注意。傑得(Judd)曾論之頗詳。茲照譯之，亦備參考。

1、有時學生的父母，在學校內頗有勢力。爲顧全他們面子起見，學

生的成績雖然不好，亦隨便任他升級了。

2、『有時低年級的人數過多，教者為年輕學生得有位置起見，雖有成績不良的學生，亦勉強讓他升級了。

3、『有時教者對於天資愚鈍或不守規則的學生，急於脫卸教學的責任，故任其升級，讓別位教師去教學。

4、『有時學校因時間的關係，如學期快要終了時，學生所習的功課，必須告一結束；故不管成績怎樣，就任他升級。

5、『有時教者以為愚鈍的學生，身軀每比別的學生大；故任他升級，讓他坐大些的椅子。』

二、用智力測量分班

1、智力年 (Mental age) 與成績品質的關係。——智力與學科的進步，有很大關係。若是一個學生的進步遲慢，欲知其原因，第一須測量他的智力。因為經過種種測驗，智力與成績實有很大的相關度。換言之，學科好的學生，大約智力必強；智力強的學生，大約學科必好。推孟 (Terman) 找到智力與小學一年級學科程度的相關度為 .725；麥柯爾 (Mc Call) 找到智力與小學六年級學科程度的相關度為 .78；笛根生 (Dickson) 測量五班一年級的學生，他找到三十三個退級的學生中，有三十一個的智力是**不及**普通一般同年的學生。由此可知用智力測量分班，實比用年齡，體格，及其他各種方法分班為尤要。

2、智力年與現在年級地位的關係。——一班內大約有四分之一學生的智力居上；四分之一學生的智力居下；另四分之二學生的智力居中。但有些是速進的 (Accelerated)，有些是留退的 (Retarded)。速進的

大抵反是低能的學生；留退的反是優秀的學生！雖因種種關係，如遲入學校，缺乏興味，不肯努力，或教學不得其法等，均可使優秀的學生不能顯其才能，而不能依其智力升級；但尚有他種原因在焉。茲分述如次。

(a) 因為教者有惰性，學生應升級或退級與否，均置之不問！

(b) 因為教者不願他的班內好學生離其班次！

(c) 教者恐怕智力發達特別早的學生，若任其升級，將來對於他的身體，或將有妨害！

上面第一第二兩種情形，是不應當有；第三種意思，亦是不能成立。一個國家最大的希望，全在上智的兒童得好為培養，使其成材，為國家用。今有許多優秀的學生，雖或已超進一級或兩級，然就其智力而言，尚是留退，殊為可惜！故分班時，必須兼用智力測量。

至於用智力測量分班的手續，大約與用教育測量分班的手續相同，故可在下節一起討論。

三、用教育測量分班

1、為重行分班用，選擇測驗，當參照下列數種要旨：

(a) 各級所用的標準測驗，必當一致。有些測驗對於三，四，五年級有一個形式；對於六，七，八年級又另有一個形式。若是這兩種形式的分數不能互相比較，則必將發生許多困難。

(b) 每種測驗當測量學科的一種單獨能力；因為分班的目的與診斷的目的不同。如讀書的速率當與理解力分別測量；否則，初從事於測量者亦必覺有困難。

(c) 所用測驗，當測量校內主要的功課。

2、選擇施行測量的人員；並宜先訓練之。至於訓練的方法，當在專家觀察之下，實際應用測驗，測量學生。

3、施行測量時，當與原所規定的方法符合。若恐全校同時測量時，測量人員的工作有衝突，則可用下面這個方法支配。

表 一

年 級		三	四	五	六	七	八
讀 法	測驗一	1	2	3	4	5	6
填 字	測驗二	2	3	4	5	6	7
加減法	測驗三	3	4	5	6	7	1
作 文	測驗四	4	5	6	7	1	2
乘除法	測驗五	5	6	7	1	2	3
生 字	測驗六	6	7	1	2	3	4
拼 法	測驗七	7	1	2	3	4	5

註：照上表，同時六個測量人員可用 1, 2, 3, 4, 5, 6 或 2, 3, 4, 5, 6, 7 等六種不同的測驗，測量三年級，四年級，五年級，六年級，七年級及八年級六班學生。依次輪轉，毫無衝突之慮，至於測驗類別多少，則可不計也。

4、校對成績，並計算學生的分數。

5、表列學生的分數。

6、計算每一年級的每個測驗的成績中數。(median, 方法見附錄一。)

7、表列每一年級的每個測驗的標準分數。有些測驗只有年中的標

準分數。如是若在年底時施行測驗，必須先改換年中的標準分數為年底的標準分數，然後該次測量所得的成績可與標準分數比較。下表就是一個改換標準分數的簡便方法。

表 二

年 級	三	四	五	六	七	八
年中標準分數	10	14	18	21	23	24
年底標準分數	12	16	19.5	22	23.5	24.5

8、計算每個學生的混合分數(Composite scores)。——例如在表三上第一個學生A的九種測驗成績，為 3, 35, 58, 31, 11, 0, 3, 3, 2.8。若用簡單方法相加，應得 146.8 分。可是用此書所介紹的方法計算，只能得91分。因為每種測驗的記分標準，彼此不同；如僅照原數計算，則上表拼法分數對於混合分數的影響，可一倍於讀法分數；生字分數對於混合分數的影響，幾三十倍於作文分數。(見下表所示的 Q；這就是四分點差，求法詳附錄一。) 我們都知道讀法的重要，並不次於拼法；生字的重要，必不大於作文三十倍。可惜普通計算混合分數時，往往

表 三

某學校各級學生用九種測驗測量的成績表。表內並載有學生的年齡，所有測驗的混合分數，教育年，教育商數，及重行分班後應在的年級。此外尚載有每年級的各測驗及所有測驗的中數及標準分數。表末且附有每個測驗的 Q_1 ， Q_3 ， Q ，及勻和乘數。

實年級 (以月計算)	年級	讀法	填字	生字	拼法	加	減	乘	除	作文	混合分數	教育年	教育商數	應在年級
109	A	3	35	58	31	11	0	3	3	2.8	91	122	112	5
111	B	5	13	21	13	5	0	4	5		54	109	98	4
131	C	12	25	26	14	7	0	0	3	3.8	76	117	89	4
92	D	9	30	56	27	7	0	7	5	2.8	92	123	134	5
103	E	2	13	20	8	9	1	3	3	1.1	44	105	102	4
112	F	10	24	27	13	9	0	4	3	2.8	73	116	104	4
130	G	7	17	29	10	6	0	1	0	2.8	54	109	84	4
139	H	9	18	12	5	11	0	4	6	3.8	71	115	83	4
139	I	0	6	14	2	6	0	0	3	1.9	28	99	71	3
154	J	3	5	13	1	10	6	4	3	1.1	39	103	67	4
131	K	11	16	19	8	6	6	4	1	1.9	60	111	85	4
122	L	10	15	29	41	7	0	3	5	3.8	79	118	97	4
168	M	3	8	39	0	6	6	5	2	1.1	44	105	63	4
149	N	4	20	19	15	6	4	5	3	2.8	65	113	76	4
104	O	4	11	20	2	5	0	3	0	1.9	38	103	99	4
133	P	3	13	19	13	7	0	0	1	1.1	38	103	77	4
104	Q	3	6	7	2	5	0	1	3	0	20	96	92	3
119	R	13	33	81	31	10	0	3	4	3.8	108	128	108	5
147	S	19	27	97	31	3	10	7	2	3.8	116	132	90	5
中數(1919)		5.5	16.5	21.5	13.5	7.10	7.3	9.3	42.5	58	111	90		
標準分數		8.0	18.5	30.0	19.6	9.06	0.3	5.3	0.2	1	71	115	100	
實年級 (以月計算)	四													
.....													
(其餘各級學生的成績以此類推不及備載)														
Q ₁		10.7	27.0	47.5	17.8	10.3	5.2	5.1	3.9	2.9				
Q ₃		24.5	48.6	109.7	49.2	15.2	11.4	12.5	9.1	5.3				
Q		6.9	10.6	31.1	15.7	2.5	3.1	3.7	2.6	1.2				
勻和乘數		1	1	$\frac{1}{5}$	$\frac{1}{3}$	1	1	1	1	5				

不能注意及此。

救濟這種弊病，必先將各種標準不同的測驗分數勻和，然後計算混合分數。如是，才可互相比較。但勻和的方法不當依照分數的大小，而當依照分數的變化率 (Variability)。其方法可用下列說明。

今有學生四人，受兩種測驗，他們的分數，表列如下：

表 四

學生	測驗一	測驗二	混合分數(非科學的方法)
甲	6	403	409
乙	10	404	414
丙	2	403	405
丁	1	402	403

第一種測驗，對於混合分數的影響，比第二種測驗大；因為他的變化率大於第二種測驗的變化率。第一種測驗的分數，從 1 至 10；第二種測驗的分數，僅自 402 至 404，就是等於從 2 至 4。倘使我們用 5 乘第二種測驗的分數，則其限域 (Range) 必將增加，可自 2010 至 2020。如是，亦為 10 點。他若我們用 5 除第一種測驗的分數，則其限域必將縮減，亦為 2 點。今將實在計算表三混合分數的方法分條述之如下：

- (a) 全校學生的每個測驗分數，用次數分配表 (Frequency distribution) 排列之。
- (b) 計算每個測驗的上四分點 (Upper quartile, 或簡稱 Q_3)，下四分點 (Lower quartile, 或簡稱 Q_1)，及四分點差 (Quartile deviation, 或簡稱 Q)。至於求上四分點及下四分點 (方法見附錄一)，此地

並沒有另外的目的，不過為求四分點差而已。所謂四分點差者，就是一種變化率的測量法。

(c) 每個測驗的四分點差，當用上面所述的勻和成績方法，而使之輕重相類。例如下表：

表 五

測 驗	讀法	填字	生字	拼法	加	減	乘	除	作文
Q	6.9	10.6	31.1	15.7	2.5	3.1	3.7	2.6	1.2
勻和乘數	1	1	$\frac{1}{5}$	$\frac{1}{3}$	1	1	1	1	5
新 Q	6.9	10.6	6.2	5.2	2.5	3.1	3.7	2.6	6

上表填字測驗的新Q最大；這並不是因為填字比讀法重要。不過因為如此，Q的數量不必更動，可以省了許多計算的工作。再者填字測驗，每為智力測驗的一種；既可表示智力的高低，故即較有勢力，亦無不妥。讀法測驗的新Q第二大，生字測驗的新Q略比較小，這是應該的。作文測驗的新Q與讀法及生字的新Q差不多。拼法是作文的一分子，故他的新Q小於作文的新Q。至於加減乘除四則測驗的新Q，都比較小；這並不是因為算學不重要，不過因為加減乘除四則有四個測驗。即使拼合計算，亦不過為算學全體的一部分。總之，每個測驗的新Q，若是相差不大，即稍有上下亦無妨。且有時因為時間經濟起見，不得不然。如讀法等，因其為學校內重要的學科，誠不妨使之較有勢力。

(d) 全校學生的分數，每一年級的分數，及每個測驗的標準分數，均

當以上法所求得的該個測驗的勻和乘數乘之。例如讀法測驗，填字測驗，及加減乘除測驗，均當用 1 乘，生字測驗當用 5 除，拼字測驗當用 3 除，及作文測驗，當用 5 乘是。可是這些商數，上面表三並沒有寫出來。原來計算時，這些數目均寫在表的兩格中間，用紅墨水表示，以便校對。

(e) 所有每個測驗的新分數相加，即得實在的混合分數。如此所得每一學生的混合分數，每一年級的混合分數，及每一年級的標準分數，均宜列入表內。下面一個表，就是表明上面表三第一個學生 A 的成績，用這第四及第五兩個步驟計算實在的混合分數的方法。

表 六

測 驗	讀法	填字	生字	拼法	加 減	乘 除	作文	實在的混合分數
學生 A	3	35	58	31	11 0	3 3	2.8	
勻和乘數	1	1	$\frac{1}{5}$	$\frac{1}{3}$	1 1	1 1	5	
新 分 數	3	35	12	10	11 0	3 3	14	91

變換年級標準分數為年齡標準分數的方法。——表三不特表示每個學生的教育年 (Educational age) 及教育商數 (Educational quotient, 或簡稱 E. Q.); 而且表示每個年級的中數和標準分數的教育年及教育商數。但一個測驗的年齡標準分數 (age norm) 未決定以先不能計算一個學生對於該測驗的教育年, 更不能計算他的教育商數。普通教育測驗只報告年級標準分數 (Grade norm), 而不報告年齡標準分數,

誠欠妥當。故有時有變換年級標準分數爲年齡標準分數的必要。表三所表示拼法測驗的第三年級的標準分數是 19.6。假定該級學生的平均實年是 9 歲，則由年級標準分數變換爲年齡標準分數，就是很容易。我們就可說九歲兒童的拼法標準分數是 19.6。用同樣方法，其餘各級的年級標準分數，均可變換爲年齡標準分數。

雖現在每級學生的平均年齡，尙無充分的研究，但亦已有相當可靠的成績。愛禮斯調查 13,868 學生進入第一年級的年齡，找到中數是 80 個月。此外尙有三個研究，報告學生讀完一級功課的平均時間。愛禮斯找到平均包括假期是 12.8 個月；推孟找到平均每一年級的時間，若根據智力年說法，則爲 12.6 個月。他若開萊 (Kelley) 則找到自一年級進至另一年級，平均需時 13.2 個月。所以普通規定每一年級所需的時間，大約爲 13 個月。

已經知道學生平均進入第一年級的年齡，及平均由一級升入他級的時間，則就可編列下表，以求學生的教育年。下表的第二及第三兩行，無論何種測驗，均可應用；至於第一行，則當因測驗不同而異。

試讀下表自右至左，第一年級的學生至第二年五月時，平均年齡當有 89 個月 ($80+9$)。第一年級的標準混合分數爲零，是意定的。至於第二年級的平均年齡則爲 102 個月 ($89+13$)；意定的標準混合分數爲 35 ($71\div 2=35+$)。第三年級的平均年齡則爲 115 個月 ($102+13$)；標準混合分數爲 71，但非意定的。其餘各年級的平均年齡，均照此法求得。除第八年級以上的標準混合分數是意定外，餘均是已經知道的 (見表三)。

表 七

化混合分數爲教育年的表

標準混合分數	平均年齡 (以月計算)	年 級
0 (意定的)	89	一
35 (意定的)	102	二
71	115	三
107	128	四
137	141	五
165	154	六
188	167	七
199	180	八
216 (意定的)	193	九
230 (意定的)	206	十
246 (意定的)	219	十一

第一年級的平均年齡取89,而不取80者,因為表三所表示各測驗的標準分數,都是以五月為根據。意定這級的標準混合分數為零者,因為平均第一年級的學生,對於這些測驗,不能得到任何的分數。至於意定第二年級的標準混合分數為35者,因為該級居第一年級的零分和第三年級的71分之間。他若決定第九年級的標準混合分數,則先平均第七年級超過第六年級及第八年級超過第七年級的分數:然後將此得數(17)加入第八年級的標準混合分數(199)。求第十年級的標準

混合分數的方法，亦是一樣。先平均第八年級超過第七年級及第九年級超過第八年級的分數；然後將此得數(14)加入第九年級的標準混合分數(216)。餘以類推。至於上表必須擴充至第三年級以下及第八年級以上者，因為表三所載學生的教育年，有些是在三年級以下，有些是在八年級以上。

編列上表的目的，原為化每個學生的混合分數為教育年。可是利用此表，亦可分別化每個學生的每個測驗的分數為學科年(Subject age)。如是，必須用該測驗的每級標準分數代替標準混合分數。至求得學科年後，自然就可計算該學科的學科商數(Subject quotient)了。其方法參照下節，就可知道。

計算教育年及教育商數的方法。——例如表三第一個學生就是A，他的混合分數為91。照表七若是他的混合分數不過是71，則他的教育年將為115個月；因為他的學業程度可與普通115個月大的兒童的學業程度相等。但若是他的混合分數是107，則他的教育年將為128個月。今他的混合分數實是91，故他的教育年當在115個月與128個月中間。計算時，須照下面這個公式：

$$115 + \left[\frac{128 - 115}{107 - 71} \times (91 - 71) \right] = 115 + \left(\frac{13}{36} \times 20 \right) = 122;$$

如是，學生A的教育年為122個月。他的實在年齡(Chronological age)，如表三所載，為109個月。因為求教育商數的公式，是

$$\text{教育商數 (E.Q.)} = \frac{\text{教育年 (E.A.)}}{\text{實在年齡 (C.A.)}},$$

故他的教育商數當為 $\frac{122}{109} = 112$ 。其餘學生的教育年及教育商數，亦

都照此方法計算。表三已經一一載明了。

爲便利計，計算學生的教育年，須應用一個比較詳細的表。下表就是爲這個目的而排列的一個例。每一個月都有他相當的混合分數。如此，若是我們知道學生的混合分數，立刻可以找到他的教育年；並無另外計算的必要。

表 八

標準混合分數	平均年齡 (以月計算)	年級
71	115	三
73.8	116	,,
76.6	117	,,
79.3	118	,,
82.1	119	,,
84.9	120	,,
87.6	121	,,
90.4	122	,,
93.2	123	,,
95.9	124	,,
98.7	125	,,
101.5	126	,,
104.2	127	,,
107.0	128	四

平均教育年齡的間距(age interval)爲13個月(理由見前)。今三年級與四年級的混合分數相差爲36點；若是13個月等於36點，則1個月必等於2.77點。如是，自71點起，每月加2.77點，至107點爲止，就得三年級每月的分數。上表就是照此方法計算的。其餘各級均可依此類推。

用教育年分班與用智力年分班的比較。——教育年與智力年對於分班的功用，頗有不同的地方。茲分述之如下：

1、若教育年是用精良的教育測驗規定，則用教育年分班，實比用智力年分班好；因爲教育年是直接表示學生的學科程度。將學科程度相等的學生合爲一班教學，乃是分班的第一要義。智力年雖能測量學生的學科程度，但是間接的。智力年與學業成績雖有很高的相關度，但因另外各種關係，這種相關度尚不能完全。蓋學生的成績，並不是智力一種的結果；他若體康，勤勉，向學態度等，均有關係。至於教育年所表示的成績，乃是這些原因的總結果。故分班時，他的功用比較智力年大。

2、再者，用教育年分班，學生可以沒有超越課程緊要部分而不學習的危險。普通課程的後面一部分，往往比起前一部分重要。若用智力年分班，必不免有這種超越的弊病。

3、還有一層，用教育年分班，學生可不缺乏基本知識的預備。小學校的功課，大概是彼此聯絡而有關係的。若用智力年分班，則升級特別快的學生，必有一部分的基本學科未曾學習；故不如用教育年分班爲宜。

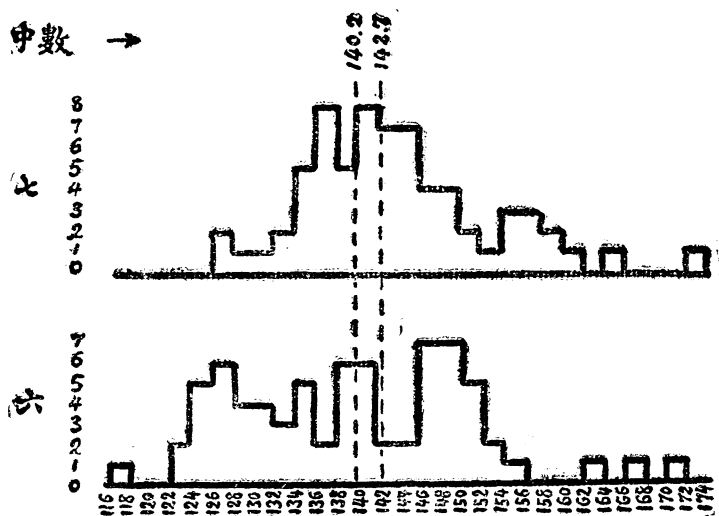
4、但是初年級的兒童，則用智力年分班，較為適宜。且現在所有的教育測驗尚未十分完全，用智力年分班，實亦是一種不可缺少的手續。

教育商數(E.Q.)與智力商數(Intelligence quotient,或簡稱I.Q.)的比較。——分班的第二個目的，是『將進步速率一樣的學生合為一班。』預料將來進步速率的方法，最好是參考學生從前的進步速率。教育年與智力年若離開實在年齡說話，就不能表示學生的從前進步狀況。進步速率大都用E.Q.及I.Q.表示。若一個學生的智力發達特別快，則他的I.Q.必成正比例的高(在100以上)。若發達特別遲，則他的I.Q.必成正比例的低(在100以下)。他若E.Q.，亦是一樣。若是學習能力發達特別快，則學生的E.Q.必成正比例的高於100；若是發達特別遲，必成正比例的低於100。E.Q.分配的形式，往往同I.Q.分配的形式相類似。

現在學校內的學生，是否照教育年而分班？——依據教師的觀察，以及應用教育測驗和智力測驗所得的結果，都可證明現在學生並非照能力及程度分班。下面這個圖形，就是表示兩級程度交叉(Overlapping)的情形。若以教育年計算，每級兩端的相差，差不多有50個月，可以等於平常四個年級。他若兩級間的相差，不過2.5個月而已。換言之，一級程度的限域，幾乎二十倍於相連兩級的相差。

各年級程度交叉的情形，既如上述。但有時每因測量之欠完全及真確，以致『言過其實。』開萊介紹下面這個公式，決定所得的變化率，究竟比實在的變化率大多少？如是，就可知道所得的交叉情形比實在的交叉情形大多少。

圖形一
表示六年級與七年級學生的教育年交叉的情形



實在的S.D.=所得的S.D. $\sqrt{\text{自身相關度的係數}}$ 。

註：S.D.普通譯為均方差 (Standard deviation)，是一種變化率的測量。(求法詳見附錄一。)

照上式，若所有測驗分數的自身相關度係數 (Self-correlation Coefficient, 求法詳後。)是.25，則實在的S.D.，當如下法計算：

實在的S.D.=所得的S.D. $\sqrt{.25}$ = .5所得的S.D.。

這就是說：倘使自身相關度是.25則，實在的交叉情形，只有所得的交叉情形之半。如是，兩級間的相差，必是比所得的兩級間相差大一倍。
分班法及指定班次的表。——表三最後的直行是表示每個學生應

入的年級。關於此層，第一步須決定年級的間距(Grade intervals)。三年級教育年的中數為111個月，八年級教育年的中數為178個月。如是，有67個月(178-111)，須分為五等分，以為五個年級的間距， $67 \div 5 = 13.4$ ，這就是每個年級所應增加的教育年。

因為分班不應完全疎忽 E.Q.，故第二步須決定所有學生的 E.Q.中數。假定所有六個年級的六個 E.Q.中數如下所舉，則其中數很可滿足這個目的的要求。(表三因四，五，六，七及八諸年級從略，故未完全載出。)

年級	三	四	五	六	七	八	中數
E.Q.中數	90	97	105	102	97	103	100—

我們已經知道三年級教育年的中數是111，年級的間距是13.4，及全校的 E.Q.中數是100，我們就可編造指定班次的表(Placement table)及重行分班的表(Reclassification table)。在下表，每個新年級的中數必比前年級的中數大，均以13.4為標準。為分班便利起見，每級中數上下的四分點(quarter point)亦均載入。因施行測驗在學年之末，故表三所載三年級四年級等的學生，均將即成為四年級五年級等學生。因此所以指定的年級，是為將來學年開始時，並非為施行測驗時。下表以111個月為四年級教育年的中數，而不為三年級教育年的中數，就因此故。

表 九

重行分班（為現在的學生）及指定班次
（為將來的學生）的表

	教育年	E. Q.	年級
中數.....	97.6 (意定的)	100	三
上四分點.....	101.1		
下四分點.....	107.5		
中數.....	111.0	100	四
上四分點.....	114.5		
下四分點.....	120.9		
中數.....	124.4	100	五
上四分點.....	127.9		
下四分點.....	134.3		
中數.....	137.8	100	六
上四分點.....	141.3		
下四分點.....	147.7		
中數.....	151.2	100	七
上四分點.....	154.7		
下四分點.....	161.1		
中數.....	164.6	100	八
上四分點.....	168.1		
下四分點.....	174.5		
中數.....	178.0	100	九
上四分點.....	181.5		
下四分點.....	187.9		
中數.....	191.4 (意定的)	100	十
上四分點.....	194.9		
下四分點.....	201.3		
中數.....	204.8 (意定的)	100	十一

重行分班的準則。——用上面方法，重行分班，尚須顧及下面三條準則：

1、任何學生不應留級或退級，除非他的教育年，在所留退的一級內百分之七十五以下。

2、任何學生不應超升一級，除非他的教育年在所超升的一級內百分之二十五以上。

3、倘使學生的E.Q.在全校E.Q.中數以下，不應超升一級或兩級，除非他的教育年在所超升一級內的中數以上。

重行分班。——表三所有的學生，都依據表九所示的標準，並參照上面三條準則，重行分班。表三最後的直行，就是表示將來應在的班級。例如三年級第一個學生A，可以超過一級而進入五年級，因為他的教育年是122個月，在五年級內百分之二十五以上；及他的E.Q.是112，在全校E.Q.中數(100)以上。第二個學生B，照常進入四年級。他的教育年是超過該級內百分之二十五；但他的E.Q.是低於全校E.Q.的中數。第九個學生I，仍舊留在三年級；因為他的教育年低於該級內百分之七十五，而且他的E.Q.只有71。倘若任他進入四年級，則必將失敗；且將要求教學的時間甚多；對於其他學生，亦必將有妨礙。此外學生編級的情形，讀者參看表三及表九，自可知之，此地從略。

但若分班須根據普通一般學校的成績，而不根據自校的過去成績時，則分班表(Classification table)當依據表七編制，不當依據自校各級的中數成績。教育年在 $\frac{89+102}{2}$ 下的，原是一年級學生，應編入

二年級。教育年在 $\frac{89+102}{2}$ 與 $\frac{102+115}{2}$ 中間的，原是二年級學生應編入三年級。教育年在 $\frac{102+115}{2}$ 與 $\frac{115+128}{2}$ 中間的學生，應編入四年級；教育年在 $\frac{115+128}{2}$ 與 $\frac{128+141}{2}$ 中間的學生，應編入五年級。餘以類推。若有猶疑不能決定的學生，則應參考 E.Q. 的大小。如升級以半年為單位，則教育年在 $\frac{115+128}{2}$ 與 128 中間的學生，須編入五年級的上半年；教育年在 128 與 $\frac{128+141}{2}$ 中間的學生，須編入五年級的下半年。其餘各級，亦以此類推。

升級及留級的分量。——如表三所示，三年級學生留一級者兩人，照常進級者十三人，超升一級者四人。麥柯爾參考各方面材料，關於此層，有四條結論：

- 1、普通差不多有百分之四十四學生編在錯誤的年級。
- 2、內差不多有百分之三十四學生與實在程度相差一級。
- 3、差不多有百分之十學生與實在程度相差兩級。
- 4、差不多只有百分之八學生，所在年級是超過他們實在的程度。另一方面，差不多有百分之三十六學生，所在年級是低於他們實在的程度。

重行分班後的另外手續。——用教育測驗分班的方法，上面已經說過。但照此分班，可實行否，還須俟四星期試行後，經教者評斷，而後能定。試行方法如下：

- 1、照教育測驗的成績試行分班，並使學生明瞭試行的意義。
- 2、四星期後，攷試在此試行期內所學習的成績。

3、根據四星期學習的成績排列等級。

4、會議決定學生究竟應該編入何級。

但普通教者的評斷與教育測驗所得的結果，常有不能一致的地方，這是不可不注意的。茲述數例如下：

1、對於特別可以升級的學生，教者的評斷往往低於測驗所得的結果。他若對於特別應當退級的學生，教者的評斷往往高於測驗所得的結果。

2、上段的情形，對於特別可以升級的幼稚兒童，或特別應當退級的年大兒童為尤甚。故他們的升級或退級的程度，必須超過教者的疑團，而後可望實行。

3、學生的教育年及E.Q.，必須超過進入年級的中數；否則，教者或將疑其能力不足而退回。

4、幼稚學生進步速，教者每疑因與愚鈍學生同班的緣故；而每不承認他的實在能力。

較大學校內重行分班的手續。——上面所介紹重行分班的方法，是適宜於較小的學校。內有一種根本的缺點，因為用上法分班，只能知道學生相等的學業成績，而不能知道相等的進步速率。所謂能預告將來進步情形的E.Q.，不過偶然應用而已。

同時應用教育年(或智力年)及E.Q.(或I.Q.)，必須有充分的學生及教師，可以分一級為數班而後可。如是，較優的學生可合為一班；他們升級可以特別加速，或可以另習較深的功課。同時且可集合年齡相若，學年相等，學力相彷彿，社交智識及職業需要相類似的學生，而

爲一班。這樣分班的優點，可有下面數種：

- 1、沒有越級涉略的弊病。
- 2、同班程度相若，有公平熱烈的競爭心。
- 3、沒有智者自滿，愚者灰心的危險。
- 4、沒有因爲缺乏興味及相當難度的工作，而生懶惰性的缺點。
- 5、沒有年齡，智力，體質，及社交智識等相差甚巨的害處。

茲將用教育年及E. Q. 爲標準而分班的步驟，分述如下：

1、單獨用教育年分級。最穩妥的方法，學生不得任意升級，若非他的成績超過該級的百分之二十五；或任意退級，若非他的成績低於該級的百分之七十五。

2、每級再用E. Q. 分班。若學生多時，每級可分爲三班。E. Q. 超過110的爲一班；E. Q. 在90至110間的爲一班；E. Q. 低於90的爲一班。但這種段落的選擇，並不是固定的。一方面當參照E. Q. 排列的情形，一方面當以每班應有若干學生爲標準。有時優生班與劣生班的人數可以較少，中等生班可以較多；有時每班可有相等的人數；均以E. Q. 的大小爲準則。

上面所敘述的方法，無論用智力測驗或教育測驗，均是一樣。以智力年代教育年，以I. Q. 代E. Q. 就可了。若只分級，而每級中不再分班，則教育年或E. Q. 並不是絕對必要。僅用學生的混合分數爲標準，已可得到相當真確的結果。在這種情形下，重行分班表上的教育年，當用混合分數代替之。

指定新生班次的方法。——1、用從前所用的教育測驗，測量學科的

程度。

- 2、以原有的勻和乘數，乘各測驗的成績。
- 3、計算混合分數。
- 4、計算教育年。
- 5、計算 E. Q.。
- 6、指定班次。

四、根據教者的評斷分班

1、根據教者的評斷，為分班的標準，頗有不充分的地方：

- (a) 對於新生不能有充分的評斷。
- (b) 對於自他班加入的學生，不能有充分的評斷。
- (c) 對於自己班內的學生，亦不能有充分的評斷。

教者固可知道最智最愚的學生，但程度介於這兩者間者，實不能明確分別之。分班乃是各級公共的事，並非單獨一班的工作。教者雖能熟悉他自己學生的程度，但不能知道他的學生程度，究竟是好於，次於，或等於別班學生程度的中數。故分班時，有應用科學測驗，比較各班程度的必要。

2、根據教者評斷為分班的標準，頗有欠真確的地方：

- (a) 教者評斷往往混合行為及學業一起。
- (b) 教者評斷往往受學生的容貌，態度，及性情等的影響。例如多血質的兒童，每誤以為高材生；憂鬱質的兒童，每誤以為低能生。
- (c) 教者評斷往往受學生年齡的影響。年大者每評斷為太高，年幼者每評斷為太低。

3、但教者的評斷，對於分班，亦有相當的功用。其重要處，可有下列數條：

- (a) 科學測驗亦須有教者評斷的輔助(理由見前)。
- (b) 各種測驗現尚未十分完備。
- (c) 教者評斷比較測驗簡易，便利，及經濟。
- (d) 教者可注意學生的體質，性情，社交智識等情形，為評斷的參考；但科學測驗則不能。

計算教學年 (Pedagogical age) 的方法。——用智力測驗，決定智力年；用教育測驗，決定教育年；用教者的評斷，決定學生的等級，為便利計，就可稱謂教學年。但如何這樣求得的教學年，可與智力年或教育年相比較，相互用，而定學生的班級呢？今若欲應用教者所記的分數，則這分數必須能與其他教者所記其他年級的分數，或同級異班的分數，互相比較而後可。但班次有大小，智愚有不同，各班程度的變化率又不一致，欲達到這個可以互相比較的目的，亦非常困難。故教者所記的分數，必先使之可與科學測驗的分數相比較；並且規定每級及每班能力的永久標準，而後可以救濟此病。麥柯爾介紹下面一個方法：

- 1、用一種或多種科學測驗，測量全體學生。若無智力測驗及各種教育測驗，而只能用一種教育測驗時，則最好用靜讀測驗 (Silent reading test)。
- 2、計算每個學生的智力年，教育年，或讀書年 (Reading age)。
- 3、各班學生依照此年排列等級。

4、教者亦用評斷方法，排列學生的等級。若有數位教師排列一樣學生，則當平均他們所評斷的等級。

5、給教者所評斷最高的學生以那一班或那一級的最高智力年，教育年或讀書年；給次高的學生以那一班或那一級的次高智力年，教育年或讀書年。此年就為教學年。其餘類推。這是幫助教者所排列的學生等級，可與他班或他級學生比較的唯一方法。否則，就沒有另外確定數目，可以幫助計算教學年了。

6、平均教學年及智力年，教育年或讀書年，則得進步年 (Promotion age)。有時教學年亦可代替進步年。但若學生有智力年，教育年，及教學年，最好平均之；並宜參考測驗的數目及性質，斟酌各年的勢力。

7、以實在年齡除進步年，則得進步商數 (Promotion quotient)。公式為：

$$\frac{\text{進步年}}{\text{實在年齡}} = \text{進步商數} \quad \text{詳細方法，可參考下表。}$$

表 十

學生	實在年齡	教育年	教者評斷的等級	教學年	進步年	進步商數
A	140	144	2	138	141	100.7
B	150	138	3	133	135.5	90.3
C	110	133	1	144	138.5	125.9
D	121	130	4	130	130	107.4

8、分級當根據進步年；同級中分班，當根據進步商數。

9、進步商數表示學生進步的速率；由此可以預知學生來年的進步

年。將來能升級否，可以根據這未來的進步年而定。如是，不必每年用測驗測量了。但來年升級時，仍當參照教者的評斷。

照表三分班，尙未利用這種方法。可是這種方法，比較更加進步了。

討論一般反對用科學測驗分班的理由。—— 1、年幼的學生被逼而與年大的及智力比較成熟的學生競爭。這原是古人的觀念，以為智力發達必與年齡增加相平行。其實經過種種測驗，年幼者每冠全班，而年大者每反被逼而追隨於年幼者之後。根據教育年及 E. Q. 分班，實可救濟這種弊病。

2、年幼學生若升級快，則有社交方面的種種困難。其實年大者每以與年幼者同班為可恥；故社交方面的困難，不在年幼者，而實在年大者。蓋年幼者往往富有適應環境的能力。有人每欲故意延擱上智兒童進入中學的時期，而叫他學習範圍較廣的功課。可是這種添加的功課，未必十分有價值。兒童的智力或學力特別強者，誠當速其畢業。如是，他可早日得到有價值的學識，將來可以早日為社會用。國家前途，實賴有此等人才耳。他若學校過小，不能有相當救濟個性不同的機會，則當合數校而為一個規模較大的學校，必可較為有益，如所謂混合學校(Consolidated school)等是。

3、致命學生的教育有涉略間隔的危險。其實教育測驗反可救濟這種弊病。若是一個學生受相當的教育測驗，得到很高的成績，我們就可知道沒有什麼間隔。再者教育測驗不充分時，且可兼用教學年救濟之。

4、疎忽年幼兒童的體康。普通每謂聰明的兒童，身體往往細弱而不

強。其實照推孟調查的結果而論，此說不能成立。聰明的兒童，弱者固不免，強者亦不少。比諸普通兒童的體康，亦未嘗弱也。

5、疎忽年幼兒童品性的發展。推孟對於這層，亦有詳細的測驗。如和善，合羣，服從，堅忍，愉快，勇敢，謙讓，公平，及節制諸美德，每與智力成正比例。(a)因為智者每能明瞭人與人間的關係；(b)因為智者每能預料行為的結果；(c)因為智者每富有自制的力量。雖有時升級特別快的學生，亦沒有良好的成績，但這或因為他從前班內的功課太容易，已經養成一種學而不力的習慣之故。

上面所述的理由，固是反對用科學測驗分班的方法；實則反是說明科學測驗的功用。因為用教育年及 E. Q.，或智力年及 I. Q. 為分班的標準，實可救濟上面大部分的弊病。至若兼用教學年，則幾可救濟全部的弊病了。

第三章

為診斷目的應用的測量

一、診斷的計畫

無論新來的校長或新來的教師，應執行兩種診斷(Diagnosis)：普通診斷及精密診斷。

為謀討論時愈覺具體起見，我們將用讀法教學來說明。

1、決定每個學生的實在年齡。(方法見附錄四。)

表 十 一

讀法教學的指導表

學 生	A	B	C	D	E	平均數
1、 實在年齡	121	123	124	118	100	
2、 年初讀法分數	40	43	30	31	38	36.4
3、 年初讀書年	121	130	93	96	116	
4、 年初讀書商數(R.Q.)	100	106	75	81	116	
5、 年初智力年	121	150	83	100	111	
6、 智力商數(I.Q.)	100	122	67	85	111	97
7、 年初成就商數(A.Q.)	100	87	112	96	105	100
8、 預定年終讀書年	131	142	100	105	127	
9、 預定年終智力年	131	162	90	109	122	
10、 年終讀法分數目標	43	47	33	34	42	39.8
11、 年終讀法分數	43	50	34	36	43	41.2
12、 年終讀書年	130	150	104	110	130	
13、 年終成就商數(A.Q.)	99	93	116	101	107	103
14、 年終 A.Q.減年初 A.Q.	—	—	—	—	—	+3

2、計算每個學生的年初讀法分數。(上表是根據桑戴克—麥柯爾的讀法量表(Thorndike-McCall Reading Scale)所測量的成績。)

3、化每個學生的年初讀法分數為年初讀書年。(方法見桑戴克—麥柯爾的讀法量表用法(Directions for Using Thorndike-McCall Reading Scale。))

4、決定每個學生的讀書商數(Reading quotient, 或簡稱 R.Q.)。其

公式如下：

$$\frac{\text{讀書年}}{\text{實在年齡}} = \text{讀書商數。}$$

5、決定每個學生的智力年。(方法隨測驗而異。)

6、決定每個學生的智力商數。其公式如下：

$$\frac{\text{智力年}}{\text{實在年齡}} = \text{智力商數(I. Q.)}$$

I. Q. 的大小，是表示聰明程度的高低。任何學科的教師，都有知道學生的I. Q. 之必要。茲就其大小，而分述其意義。

I. Q. 在 140 以上者……………類似天才或天才。

I. Q. 在 120 至 140 間者……………最優秀的智力。

I. Q. 在 110 至 120 間者……………優秀智力。

I. Q. 在 90 至 110 間者……………正常或普通智力。

I. Q. 在 80 至 90 間者……………愚笨。

I. Q. 在 70 至 80 間者……………類似低能。

I. Q. 在 70 以下者……………確定的低能。

教育有兩種問題：(a) 決定教育的目標，(b) 決定學生達到目標的能力。診斷這種能力的簡便方法，一方面須測量學生能學什麼東西；另一方面須測量他已學過什麼東西。前一個方法，須置學生於新動境內，然後觀察他所學習的成績；後一個方法，須假定他自生以後，就在學習的動境中，只要測量他已學得多少就可了。不幸因為缺乏相當的標準測驗，這種方法的進步，甚為遲慢！至近年來，心理學家始能確實知道一個學生若有相當的 I. Q.，始能做什麼學校或什麼年級的工作。

推孟根據他自己的經驗，有下面一段結論：

『60 I. Q. 的學生，不能做三四年級以上的工作；70 I. Q. 的學生，不能做五六年級以上的工作；80 I. Q. 的學生，充其量可做七年級的工作；至於90 I. Q. 的學生，則若有充分的努力，或可畢業於中學』。

如求學的機會合於普通的標準，則凡 E. Q. 只有60, 70, 80或90的學生，可與相同的 I. Q. 一樣解釋。倘若智力年或教育年尚未充分發達，則上面所列的標準，亦未必能達到。這就是說明這些兒童實際必是遲進的。

上面所討論的學習能力，是普通的。此外尚有特別學習的能力。診斷這種能力的方法有三種：

- (a) 測量學生從前單獨一種學科的成績，及其進步的速率。——如讀法，書法，算學，拼法，作文等，在小學校內都是繼續教學的。這些學科的能力，自三年級至八年級均可用團體測驗測量。根據學生的讀書商數，拼法商數 (Spelling quotient) 等，就可決定他宜於學習何種學科。在中學校內，可叫學生在短時期內，專心學習一種功課，然後測量他的成績。如是，根據他的進步情形，可以預定他將來的希望。
- (b) 測量對於某種學科最有關係的精神作用。——例如欲望學習數學有良好的成績，必須具有學習這種學科的特別能力。其方法宜先分析關於數學方面的各種精神作用，然後編制測驗以測量之。
- (c) 測量普通智力。——無論在小學，中學，或大學內，學生的學業

成績與普通智力有很高的相關度，所以知道學生的普通智力，亦可預定將來特別學習的成績。

7、決定年初的成就商數(Accomplishment quotient, 或簡稱A. Q.)。——普通測量過去的效率，往往將一班的成績與年級標準分數比較。這種方法遠不及用成就商數為測量的標準。因為成就商數顧及學生學習的能力，而普通方法則否。求成就商數的公式，就是：

$$\frac{\text{讀書年(或其他學科年)}}{\text{智力年}} = \text{成就商數。}$$

例如表十一，A學生的讀書年是121，他的智力年亦是121；所以他的讀書A. Q. 為 $100\left(\frac{121}{121} \times 100\right)$ 。其餘學生的讀書A. Q.，均用一樣方法計算。

A. Q. 差不多是最新最真確的測量學習及教學效率的方法。評斷學生的工作，用此方法，最為公允；因為學生可就其生成的智力，為相當比例的進步。試看表十一，C學生的讀法成績，居全班之末。普通的評斷者，必將謂其效率甚低。B學生的讀法成績，居全班之首。普通的評斷者，必將謂其效率甚高。其實適相反。就生成的學習能力而論，C學生的效率高居全班之首；B學生的效率遠在全班之末。因為前者的I. Q. 只有67；後者的I. Q. 有122。

有時教者或將教學能力較低的學生；若欲必期望達到標準的成績，這是很不公平。今用A. Q. 就可免除這種弊病。再者有時教者或將教學會受不良教學的學生；若欲必期望於一年內補足數年的疎忽，這亦是不公平。今用A. Q.，亦可免除這種弊病。

一個學生或一班學生有 100 A. Q. 者，其進步可謂已經滿足普通標準的要求。但在相當的範圍內，教者當竭力提高 A. Q.。對於年幼聰明的兒童，尤當特別注意；因爲他們所在的班次，往往低於他們智力的要求。

二、診斷的方法

學生的學科成績不良，必有不良的原因。教者若欲施補救的方法，必當先診斷這種原因。例如讀書能力與計算理想算學問題的能力有關係。若是經過考試，兒童對於計算理想算學問題的能力不及格，則診斷的職務是要找到他的原因，及原因的原因。總言之，須自各方面去找原因。如不能計算理想算學問題是因爲讀書能力不充分，讀書能力不充分是因爲生字缺乏的緣故等等，都是診斷時所應找到的。故診斷的目的，就是用各種方法，詳細分析學生的學科能力的缺點；並且探求各種缺點的根源。但若沒有科學的方法，而欲實行此事，則必不免有猜測及武斷的弊病。中時少而誤時每多也！

平日學校內實日有診斷的事實發生。有些的效果比較是很好；有些比較是沒有什麼效果。今若診斷的確度增加，則找到根本原因的次數亦必自然增加。但欲得到較高的真確度，必須有充分時日的研究；同時亦必須有科學診斷的知識。有許多教者，並不了解診斷的方法及其功用，以致教學的成績不良，殊爲可惜！倘將來一旦診斷法果能精密真確，則對於時間及努力的經濟，必有很大的貢獻。

1、學生自省法(Introspection by pupil)。——這種方法，學生自己時常應用。此地毋須詳細討論。學生用這種方法，頗能找到自己困難

的地方，及困難的原因。若學生用這種方法，可以自己診斷，則教者誠不必再用很精密的方法去診斷了。但若用這種方法，尚不能知道自己困難的原因時，則教者可用談話法救濟之，因此亦可為診斷的幫助。

2、平常作業的觀察 (Observation of normal work)。——最普通的診斷方法，是要在被診斷那個學生的作業上得到有些暗示。例如一個學生讀書有困難，則當觀察他讀書時的情形。究竟他的眼動是否有規則？他的口唇是否作微動等？均可幫助教者診斷他的困難原因。他若學生不能完全理解歷史的功課，或對於計算數目的能力特別遲鈍等，亦均可用此方法診斷。

3、學生自己口述所用的方法 (Oral tracing of process)。——有許多困難，若僅用內省法及觀察法，或不易找到實在的原因。遇着這種情形，教者可叫學生口述他所用的方法。例如教統計學，若學生不能懂得計算中數的方法，則可叫他到黑板前面，口述他計算中數的方法。如是，他的困難及困難的原因，立刻可以找出。至於其他學科有困難者，亦均可用此方法診斷之。

4、分析測驗所得的結果 (Analysis of test results)。——有許多測驗，並非單獨測量學生的成績，且亦有益於診斷困難的原因。其實無論何種標準測驗，都有診斷的功用。例如讀法測驗可以分別測量速率及理解力，算學測驗可以分別測量理解問題的能力及計算問題的能力等，均可為診斷困難原因的幫助。

5、調查學生發展的歷史 (Developmental history)。——此層適如醫者詢問病人的過去情形一樣。知道他過去的情形，就可知道他現在疾

病的原因；然後投藥醫治，不至於亂。今診斷學生學科缺點的原因亦如此。這種歷史的調查，不當僅限於學生自己方面，而且應當及諸他的父母及祖先方面的情形；因爲許多能力都受遺傳的影響。他若學生從前的學校歷史，亦宜調查。

6、比較優等生與劣等生所用的方法 (Contrast of opposites)。——教者每因不知道一門學科所有缺點的普通原因，故診斷時往往不能得到良好的結果。例如一個學生不能有充分的進步，實因爲他的學習方法不良；今若教者自己不能知道什麼方法是好，什麼方法是壞則他診斷時必無成功的希望。若欲救濟這種弊病，最好比較成績好的學生及成績次的學生所用的方法；觀察他們的動作；並且叫他們口述自己所用的方法。如是，他們不同的地方可以顯然現露；乃有裨益於診斷必不鮮。這種方法，對於沒有經驗的教師，用之尤爲適宜。

7、完全分析學科的能力 (Complete analysis of ability)。——分析關於學科能力的感覺，動作，及精神各方面情形，以爲診斷的方法，乃是最後一步的計畫。一因爲用這種方法較爲費時；二因爲用這種方法若仍失敗，就沒有另外的方法可以應用了。完全分析學科能力的意義，就是包括上面所討論的各種方法而言。如用學生自省法，平常作業觀察法，學生自己口述所用的方法，分析測驗所得的結果，調查學生發展的歷史，及比較優等生與劣等生所用的方法是。現在我們可舉一種很簡單的學科如拼法爲例來說明。若是一個學生的拼法成績不好，當有下面幾種原因，可以研究。

(a)或因覺官方面有缺點。(如音聽不明白，字看不清楚等。但均可

用測驗測量之。)

- (b)或因智力薄弱。(可用智力測驗測量。)
- (c)或因發音錯誤。(拼錯的字，叫學生讀之，即可知其大概情形。)
- (d)或因注意力不充足。(如只看見半個字那種情形是。)
- (e)或因記憶力薄弱。【須用記憶距(memory span)等測驗測量。】
- (f)或因字的意義不明瞭。(叫學生解釋拼錯字的意義，即可知其情形。)
- (g)或因手指等動作不靈敏。(有時口拼的能力勝過於寫拼的能力，就是這個緣故。)
- (h)或因偶然錯誤。(明知字的拼法，但不知不覺間因不注意而拼錯者。)
- (i)或因惡習慣遷移。(須調查學生過去的學校歷史。)
- (j)或因一種怪習慣的作用。(如有些人於每字之末，任意添加字母“e”等習慣是。)
- (k)或因沒有興味，不肯努力。(關於教學方面，或關於學生自己方面。)

如拼法這樣簡單的學科，若學生的成績不良，倘且有上面種種原因應當注意，可知診斷非是容易的事。現在另有一個問題，當然要發生：就是教者實行診斷後，如已找到困難的原因，是否對於原因必有補救的方法？欲解決這個問題，必須先有充分的實驗而後可。但現在這種知識，尚非常缺乏！有些情形，似乎補救非常困難。不過若是學生的智力尚優，則學校應竭力設法救濟之；不當任其不能，而妨礙他的別種

功課。

三、診斷上必要的知識及技能

診斷之能成功與否，恆視診斷者具有下面四種條件與否爲斷：

- 1、知道各種學科普通缺點的普通原因。
- 2、能觀察學生的動作，爲診斷的幫助；並能用自己的經驗解釋之。
- 3、有發見種種暗示的能力。
- 4、有對於某種診斷，應用何種補救測量的知識。

下面乃是一般學科普通缺點的普通原因。若是教者應用診斷的方法，爲教學便利起見，無論如何，須先有這種知識。

1、練習不充分。——有許多學生，功課不好，不過因爲沒有充分的練習。這種原因，容易補救；故毋庸討論。但既有充分的練習，尙有不滿足的情形時，則教者須特別注意了。

2、學習不得其法。——每門學科有每門學科的學習方法；每個學生有每個學生的學習方法。所以有普通學習的方法全體學生可以應用外，尙須有各科及各人的學習方法。大概學生成績不及普通的程度，若非是他的天資愚笨，必因他所用的學習方法不適宜。例如有些學科每有兩方面可以進步，如速率及品質；並且有相互的關係。有時學生的成績不好，每因他的工作太快，或因他的工作太慢。這就是學習不得其法的例子。

3、根本的智識或技能不充足。——此層可分兩方面討論：(a) 工作本身方面的智識或技能缺乏；(b) 或別種知識或技能遷移到新工作方面來的程度不充足。例如學習文學，歷史，地理，算學，及其他各種學

科，若是根本的知識或技能缺乏，或不缺乏而不能應用，則必沒有良好成績的希望。他若讀書能力不好的學生，去研究別種學科如歷史等，亦將必有困難。救濟的方法，宜先養成有根本的知識及技能；並且有隨時應用的能力。

4、對於學習沒有興味。——興味對於學習的重要，任何人都不能否認。普通學業失敗，大抵因為缺乏興味之故。

5、身體方面有缺點。——無論診斷何種能力，均須仔細測驗身體各方面的情形，尤以耳目兩部分為重要。倘使學生只聽到一半，或看見一半，則教學的效率亦必不能超過於一半。不僅如此，診斷且宜注意學生反應機關那部分的情形。例如學習圖畫，寫字，手工等學科，必須調查學生筋肉的發達及應用的情形。

6、智力薄弱。——低能差不多是功課不好的主要原因。因為這種原因是最普通，且是沒有什麼方法可以救濟使之進步，故覺得特別重要。普通診斷每疎忽此層，故所得的結果每浮泛不切實際。其實，例如歷史成績不好，每因讀書能力不足；讀書能力不足，每因智力薄弱。所以每遇學生成績不好，欲診斷他的原因時，最好先用智力測驗測量他的智力。

第四章

教學上應用的測量

凡教學一種學科，最好當未教學以前，用一種測驗，測量學生已有的程度。如是，教者始可就學生的程度，而施適當的教學方法。但此并

不是謂第一次測量以後，毋須再用測驗。其實測量與教學宜相互而行。至於教學技能方面的功課，尤宜如此。如算學，習字等，測量差不多可佔教學的大部分功夫。其他功課如讀書等，現雖尚未有充分的練習測驗。但若讀的材料，選擇適宜，學生讀時，能表示他的理解及速率的程度，則先後比較成績，獲益亦必不鮮。

教者具有充分的學習心理知識，固是必要；但教學方面，並不能僅以此為能事。蓋有時原理往往不易應用到實際方面去。關於這種問題，此地當然不能討論；但教育測量，對於此層，有三種重要的貢獻，不得不略為述之，即練習測驗(Practice tests)，非形式的測驗(Informal tests)，及標準量表(Standardized scales)三種測量的功用是。茲分別詳細討論之。

一、練習測驗——我們現舉一個小學用的練習測驗，就是葛第斯的標準算學練習測驗(Courtis Standard Practice Tests in Arithmetic)為例來說明。這種測驗共有四十八張硬片子。每張片子，是為一次的功課。第十三，三十，三十一，及四十四四張片子，為測驗用。第四十五，四十六，四十七，及四十八四張片子，為學習用。其餘每片只含有一種算學例子。功課進行，自易至難。每進一課，就加一種算學上的困難。若是一個學生熟習四十課時，他就熟悉加減乘除所有的困難了。練習時，每一個學生須有一組練習測驗。

除這種練習功課以外，每個學生尚須有一本學生的練習記錄簿(Student's Practice Pad)。這種簿子，是透明紙做的。學生練習時，把那張練習片子放在一張透明紙底下。如是，他可看見問題，同時且可

在透明紙上計算。因此這些練習片子，以後還可應用。這誠是最經濟的方法。學生的練習記錄簿上尚有另外的紙，以備學生畫圖表及曲線等，記錄他自己每日進步的成績之用。

此外教者尚須備有一本簿子，叫做教師的記錄簿 (Teacher's Manual)。這本簿子內有關於應用練習功課及學生練習記錄簿的詳細說明。同時教者且可記錄每個學生的每日成績。

至於應用練習測驗的步驟，有如下面所述數條：

1、先用第十三張片子，測量全體學生的程度。所以用這張片子測量的緣故，就是因為這張片子包含自第一張至第十二張片子所有困難的問題。每個學生須將這張片子放在練習記錄簿的第一張透明紙下面；問題向上。教者出令開始後，學生須繼續演算問題，至教者出令停止為止。

2、停止後，學生交換練習記錄簿，當教者報告對的答案時，學生互相校對錯誤。

3、凡沒有錯誤的學生，不必再練習自第一張至第十二張片子的功課。有時用第十三張片子測驗兩次，較為妥當。沒有錯誤的學生，就可學習另外的功課，等到次的學生趕到為止。或者他可繼續受以後較難的測驗，找到他須練習的地方，然後練習。

4、另外學生，若是第一次測驗有錯誤，則須練習第一課，即是第一張片子。若是練習第一課後無錯誤，則練習第二課。餘以類推。

5、若是第一課尚有錯誤，則當繼續練習之，至沒有錯誤為止。

6、當學生完習第十二課時，立刻再用第十三張片子測驗，證明他

究竟已經熟習所有的問題否。若無錯誤，他可做另外的功課；或者受以後較難的測驗，練習以後較難的問題。

7、當學生已有百分之九十個人，已經熟習第十三課以前的功課時，則須用第三十張片子測量他們的程度。沒有錯誤的學生，可以免習自第十四課至第二十九課。若有錯誤，則自第十四課起，繼續練習，至第二十九課止。方法如前。

8、教者須挨日記錄學生的成績；防制他們欺騙的舉動；診斷他們困難的原因，並用適當方法補救之。此外且宜鼓勵學生有良好的成績；指導他們記錄自己的工作；並宜時時審查他們的演題。

現尚有一句話須說明：就是所有片子的背面，都印有問題的答案。所以學生若將片子反一面放在透明紙底下，可以自己校對成績；或者互相校對。如是，教者可以完全注意個人教學那方面了。

應用練習測驗的方法已講完，茲將其功用分述於後：

1、練習測驗是個人教學的一種方法。——學生的興味、程度，及學習的方法，彼此不同，今若用彼此一樣的方法教學，其成績不良，必無疑義。這種情形，尤以技能方面的功課為甚。全體教學的目的在全體；但實際上並沒有幾個人能得其益。今欲救濟這種弊病，必須應用個人教學的方法。練習測驗，乃是個人教學方法最良的一種。沒有這種方法，個人教學差不多是不可能的。其理由如下：

(a)練習測驗可使學習的功課適合學生的程度。他可只在從前沒有學過的或須再學的地方用功夫。這就是未教學以前測驗的功用，葛第斯的標準算學練習測驗的第十三張片子，就是一個例子。

(b)用練習測驗，學生可用自己的方法學習；且可因此找到較好的學習方法。即如加減乘除這樣簡單的功課，學生所用的方法，亦是彼此不同。寇備 (Kirby) 實驗所得，找到一個學生最適宜的方法，對於另外學生未必適宜。有許多學生，若非受環境逼迫，力求進步，往往不知道去找良好的學習方法，及其進步之可能。

(c)用練習測驗，學生可就其能力而求進步。無論何種測驗，都可找到學生進步速率不同的情形。欲求教學有相當的效率，必須注意此層。

2、練習測驗可以鼓勵學生有力求進步的精神。——練習測驗可使學生知其目前及將來的目標，且可測量他的進步及去目標的遠近。如是他學習時，必特別努力，特別有興味。每日學生並可自己畫圖表，表示進步的情形；前後比較成績。這是利用最有效果的競爭心。此外個人的成績可與團體的成績比較，團體的成績又可與另外團體的成績比較。凡此種種，均於教學上有很大的幫助。

練習測驗有自己記錄成績的必要，既如前述。學生因此學習功課，每特別有興味及努力。努力及興味同時存在，乃是一種遊戲的變形。普通教學並無這種記錄；有時雖有成績報告，但都是比較的，並不是成績確實的表示。所有報告且不過說明某某成績好，某某成績次而已；並不說明某某成績確比某某成績好幾多。次的學生因此每以為他的天資愚笨，雖努力亦不能如人；好的學生每以為他的天賦獨厚，可以不努力而勝他人。這種弊病，可以立見。至若用練習測驗，則可鼓勵學生擔負自己應求進步及達到目標的責任。懶惰的人達到目標必遲；

勤勉的人達到目標必快。學生若知道自己的成績，及離開目標之遠近，則即欲懶惰，亦不可能了。

3、練習測驗可使學生有充分的練習。——桑戴克的第二條學習律，就是練習律 (Law of exercise)。若是一個學生學習的目的明確，效果律 (Law of effect) 應用得宜，而且練習充分，則進步迅速的條件，都完備了。

今若有一個教師教學加法，他同學生說：『你們將我說的題目寫下來。』然後叫張生：『你口算第一行。』然後又叫李生：『你口算第二行。』照此方法，繼續做完。則每個學生都等着他自己的次序；每個學生只有一次機會練習。這種呆笨的方法，恐怕現在尚是很普通！倘使一班有四十個人，一個等着一個，每人用心思的時候，不過佔四十分之一。他若用練習測驗，個個學生可以同時練習。如是教學的效率，可以增加四十倍。

4、練習測驗對於診斷一層，非常有益；教者幫助學生，亦非常便利。——用練習測驗，學生需要教者幫助的地方，教者甚易知道。不要幫助的地方，教者亦可毋庸濫費。若是一個學生用一種努力，沒有相當的結果，他必喪心沮氣，而無興味。久之，則所謂努力者，亦將終歸消滅。這種情形，教者往往不知。今用練習測驗，學生有日畫進步曲線之必要。若曲線停止上升，須有教者的幫助時，可以一望而知。至於毋庸教者的幫助時，而教者尚幫助之，則不特徒費精力與光陰，且有害於學生學習的態度。

有效的診斷，教者必須先發見學生困難的地方，然後找到困難的原

因。練習測驗乃是診斷的一種最好方法；因為學生每日所做的功課，往往可以露出他的困難痕跡來。

用葛第斯的標準算學練習測驗，並沒有過於注重機械訓練的危險；因教師的記錄簿上印有相當時間的分配，且力戒任意逾越其時間。教者若加以注意，可無意外的弊病發生。惟教者應用這種練習測驗時，若與普通功課分離，則養成計算的能力，未必能應用於社會生活上去；這是最大的危險。所以這種測驗，須在有生氣的環境下練習才好。譬如用設計方法教學，學生須做種種作業；當他們知道需要這種計算能力，或找到他們的計算能力缺乏時，則用這種練習測驗學習，必比較有興味，有效果。

二、非形式的測驗（普通考試）——考試制度在西洋還是占很重要的地位。例如美國全國大約有 700,000 個小學，中學，及大學的教員。普通計算，每個教員每年有二十次考試。如是，每年差不多有 14,000,000 次考試。連出題目，監視試場，及評閱卷子，每次考試差不多要費三點鐘。按此計算，每年考試費去的時間，幾乎有 42,000,000 點鐘；其重要可知。照我的意見，考試是一種很重要的教育測量。就是將來亦不能廢止的。熱心用標準測驗的人，並不是輕視考試制度；其實標準測驗，亦不過是一種改良的考試而已。再者，現在我們已有種種改良考試的方法。下面所舉一個正誤測驗（True-false test）的例子，就是說明教者如何可以利用這種改良的考試，使考試不僅為考驗學生成績的工具，而且可成爲一種很有興味的工作。

正 誤 測 驗

下面二十句句子，有些是對；有些是錯。若是對的，你把「對」字底下畫一畫；若是錯的，你把「錯」字底下畫一畫。每句句子，都須照此方法做個記號。倘使你不知道，你任意猜測可也。

- 1、只要肯練習，隨便什麼事情都可做成功。 對 錯 ×
- 2、所謂標準測驗者，就是我國平常所用的考試。 對 錯 ∟
- 3、學習律有三：就是傾向律，練習律，及效果律。 對 錯 ∟
- 4、凡存在的東西，未必都有分量可以測量。 對 錯 ∟
- 5、測量學生的程度，當先於教學。 對 錯 ∟
- 6、以實在年齡除智力年，即得智力商數。 對 錯 ∟
- 7、先求得教學年，然後可以計算教育年。 對 錯 ×
- 8、陸志韋先生是北京大學的心理教授。 對 錯 ∟
- 9、How to Think 一本書是 J. Dewey 著的。 對 錯 ∟
- 10、診斷是只測量學生困難的地方。 對 錯 ○
- 11、兒童的天性是善的。 對 錯 ∟
- 12、功用說(Theory of Utility)是根據效果律而成立的。對 錯 ∟
- 13、本能的發展及衰弱，是逐漸的。 對 錯 ×
- 14、女子的發育比男子快。 對 錯 ∟
- 15、男子的智力，普通是比女子高。 對 錯 ×
- 16、求平均數 (mean) 的公式是 $\frac{n+1}{2}$ 。 對 錯 ∟
- 17、抱獨身主義的人，因其天性使然。 對 錯 ∟
- 18、患者易於為惡。 對 錯 ∟
- 19、學生的成績不同，就是教學的效率不同。 對 錯 ∟

20、習慣是第二天性。

對 錯 \angle

上面零零碎碎的考試，就是一個正誤測驗的例子。他的範圍廣窄，可因所測量的知識範圍廣窄而定。我隨意這樣草率寫成這個測驗，原欲表示這種測驗編制之簡便。無論那個小學，中學，或大學的教員，都可照此方法，隨時造成正誤測驗，測量學生的成績。

這種測驗所有對錯的句子，均須印刷清楚。每個學生須有一份。測驗時，學生當在「對」字或「錯」字底下畫一畫，表明他的意見。測驗後，校對成績，可用「 \angle 」及「 \times 」記明學生答案的是非。若有不答的句子，則可用「○」記明之。

計算成績，可用下面(A)這個公式：

(A) 學生分數 = 是的答案 - 錯的答案。

例如一個學生的成績，是的答案有十四，錯的答案有五，沒有答的有一，則照公式計算，當如下：

$$\text{學生分數} = 14 - 5 = 9。$$

我們須先說明學生答案的數目須減答錯的數目的理由。假定一個學生，對於所測量的功課完全不知道，今受這種測驗，每句隨便做個記號，則照碰機會的公例，十句句子或可答對，十句句子或將答錯。換句話說，他猜測的結果，必有百分之五十是對，百分之五十是錯。他的分數當如下式計算：

$$\text{分數} = 10 - 10 = 0。$$

總而言之，若學生的知識等於零，則照此方法計算，亦等於零。現在若他知道十句，猜測十句；則十句猜測之中，必有五句是答對，五句是答

錯。如是，他雖只知道十句，而答對的實有十五句($10+5=15$)；故計算他的實在程度，須用下法：

$$\text{分數} = 15 - 5 = 10。$$

若是一個學生完全答對，則他的分數當然是 $20 - 0 = 20$ 。

此處尚須注意的，就是調查是否有未曾答的句子。若是測驗的時候甚短，則有許多句子必將遺漏不答。遇見這種情形，不答的句子可置之不問。計算成績，就照上面這個公式。但若測驗時間非常充分，即工作最慢的學生亦能完卷；或教者曾申明，如遇不知道的句子，可以任意猜測，則計算成績時，當作錯誤算。如是，用下面(B)這個公式，較為便利。

$$(B) \quad \text{學生分數} = \text{句子總數} - 2(\text{錯的答案})。$$

若是句子的總數是二十，內有五句是答錯，則照上式，得

$$\text{分數} = 20 - 2(5) = 10。$$

若照第一個公式(A)，則得

$$\text{分數} = 15 - 5 = 10。$$

上面兩個公式，當沒有不答的句子時，則答數可以符合。不過若有未曾答的句子時，則必須用公式(A)。至於沒有未曾答的句子時，則用公式(B)，較為簡單。因為他的第一個數目，彼此一樣，第二個數目均為錯的答案之倍數，計算時自然要方便些。

有許多人或疑這種測驗是碰運氣的。其實不然。倘使我們現在用一個銅元擲五萬次，結果大約必有二萬五千次人頭向上，另二萬五千次人頭向下。麥柯爾曾實驗這種碰機會的結果，果然找到二萬五千次人

頭向上，二萬四千九百九十九次人頭向下。他說：『若是數的時候沒有錯誤，兩個數目必可絕對相等。』總之，若是數目多，機會必是平均。故正誤測驗所用的句子，亦愈多愈好。

上面已將正誤測驗的計算方法詳細說明；茲將進而討論編制的方法。其應注意的，大約有如下數點：

1、對錯句子的數目，須大約相等。——若是一個小孩發見對的句子多於錯的句子，或錯的句子多於對的句子，則當他遇着不知道的句子時，他就可利用這種情形，猜測句子之對錯。所得的成績，就可比他應得的高。但是這種情形，聰明兒童始能利用，愚笨者則不能。

2、每次所用測驗的句子，對的和錯的數目，亦不應絕對一樣。——若是常常一樣，則學生回算已經答過的句子，比較對的和錯的數目，若發見已答對的句子多於已答錯的句子，則他遇着不能答的句子時，必以為錯。另一方面的情形亦如一樣。故對於成績，將必不能十分真確。所以編制正誤測驗時，有時對的句子須比較多；有時錯的句子須比較多；有時須彼此一樣。無論何種呆板的方法，均不宜用。

3、不應將對的句子都放在前面；亦不應將錯的句子都放在前面；亦不應將對及錯的句子，互相交叉。最好須用偶然碰機會的方法，決定次序。

4、應沒有不明瞭的句子。——前面那個正誤測驗，如第十一及第十四兩句句子，都是不大明瞭。試先說明第十一句。因為人類的本能並不都是善的，故謂『兒童天性，本是善的』這句話是錯，實有相當的理由。但本能雖非全善，可是有為善的可能性。就此而論，則謂『兒童天

性，本是善的』這句話是對，亦能成立。至若第十四句，亦有兩方面可以解釋。如謂『女子的發育，比男子快』是根據春情發動時期的情形而言，則這句句子是對；他若根據兒童生後幾年的情形而言，則這句句子是錯。倘因為這種句子的緣故，學生的分數發生有不公平的影響，他們必不能滿意。故出題時，務須特別的謹慎。但這種情形，教者自己每不能知道。不過測驗後，學生若有機會自己校對成績，則每可發見之。發見後，這種句子自然不應計算。

5、教者須知道這種測驗究竟是測量什麼東西。句子宜簡單明瞭，不宜用否定辭。所有冗贅文字的弊病，均宜除去。

6、不應有暗示的句子。——例如『長江是中國最長的河流。』這類句子，學生即不知道，亦必以為對的；因有『長』字為之暗示。但反面情形，亦是如此。例如『大西洋是世界最大的海洋。』學生若不知道，亦將以為對的；因有『大』字為之暗示。一個測驗，若是這種句子過多，則難度必即隨之增加。有時測驗結果反是負，就是因為這種相反的暗示句子太多之故。如是，照上面的公式計算，即使學生實際上知道幾句，亦必受此影響而不能得有分數矣。故這種方法，實不適宜。但我們要測量學生知識的確度時，則用之亦未始不可。不過普通測驗最好不用這種句子。合於理想的方法，要使毫不知道的學生，得到零分。

7、後邊的句子，不宜解答前邊句子的意義。

8、免除瑣碎及太普通的句子，恐怕學生因此養成不良的學習習慣。

應用正誤測驗的方法。——這種測驗當印得很整齊，每個學生須有一份，前面已經說過。但有些學校印刷機器未曾置備，教者應用這種

測驗，必須另有方法救濟而後可。下面所述幾種，可任擇其一而用之。

1、教者口述句子，學生照錄之；然後開始測驗。但這種方法，非常費時；且乏興味。

2、將所有句子寫在黑板上；每個學生只照寫句子之數目。然後開始測驗，記「對」或「錯」於數目旁邊；或簡用正負號表示亦可。這種方法，比較方便；可是尚須費黑板上的功夫。

3、叫學生先寫1, 2, 3, 4, 5, 6等數目；然後教者口讀第一句句子，學生記其「對」或「錯」於數目旁邊。其餘以此類推。這種方法是最容易，最簡便。但有一種缺點：就是有些學生不能聽得清楚；長且複雜的句子，尤易錯誤。他若句子可以看見時，則就可沒有這種弊病了。故教者宜先觀察情形，究竟何種方法，較為適宜，然後採用之。

校對成績的方法。——1、若是每個學生有一張題目，則教者校對時，可取一張沒有用過的句子，自己先用「對」或「錯」記明之；然後將此紙放在學生答案的旁邊校對。此法甚為便利。

2、若是學生沒有題目，他們不過寫「對」或「錯」於數目旁邊時，則教者可用一張紙，如學生一樣，寫數目於其上，並自記號之；然後與學生成績校對。此法亦甚便利。

3、最後一個方法，若是學生用格子紙，照教者指導依次排列句子的數目，則學生自己可以校對，或互相校對。其法教者當報告每句句子的答案，令學生校對之。若學生校對時，沒有句子作參考，則教者宜先讀句子，然後報告答案。如是，學生可以知道那句句子，他是答對；那句句子，他是答錯。校對完後，教者當問學生第一句有多少人答錯，叫

他們舉手表示；第二句有多少人答錯，叫他舉手表示。餘以類推。同時教者須記錄每句答錯的人數；然後收集試紙。

但是學生自己校對，每有欺騙的事情。若說沒有人欺騙，這就是欺騙人的話。不過教者若能善為處理，欺騙的事情亦不至發生；可是這亦無補於事。社會中責備極嚴，人有欺騙行為，就沒有人肯原諒；故為學生時，倘有不誠實的習慣，教者誠宜設法發見之，改正之。這是對他們行為上真正的貢獻。凡教者都應有這種態度。

正誤測驗的用處。——1、這種測驗可以包括教材的大部分，測量學生較大限域的能力。所以比較普通考試只出幾個題目，測量學生的程度公平；因為普通考試所包括的範圍每甚狹小，學生往往猜測教員的意旨而預備功課。幸而碰着的人，分數必多；不幸碰不着的人，分數必少。實則他們的程度或是一樣的。

2、正誤測驗能增進教員與學生的感情。他若普通考試，往往使學生與教員的感情疎遠；因為學生往往疑信教員記分不公平，且恐怕他的平日行為或於分數有影響。至於正誤測驗則不然。學生所得的分數，都是他應得的。所有成績，他可自己校對；故絕對沒有不公平的地方。

3、學生都歡喜這種測驗。(a) 因為這種考試的規則是很公平；(b) 因為這種測驗包括範圍甚廣，沒有用文字的地方。

4、教者亦比較歡喜這種測驗。(a) 因為容易計算成績；(b) 因為計算成績時，學生可以代勞；(c) 沒有記分不公平的恐懼。

5、這種測驗比較有教育的意義。其理由如下：

(a) 教學最好的機會，是在考試之後；因為此時學生最欲知其成績，

希望進步之心更切的緣故。

(b)測驗後，討論公之於全體。

(c)考試次數可以增加。考試對於學習有功效，對於進步有幫助，已經實驗證明的。寇備同 葛第斯等均謂：學生測驗時所學的東西，比較學習時所學的還要多，這就是這段話的明証。

(d)包括範圍甚廣，覆習的機會彼此平均。

(e)可以知道學生困難的地方。

6、用這種測驗，教者可以知道全體學生的情形，及教學的效率。若教者記錄答錯每句句子的百分比，他就可知道學生學習及自己教學的情形如何。

7、這種測驗可以訓練學生誠實的品性。有時教者或學生交換重覆校對學生自己已經校對的成績，每可發見有欺騙的行為。發見之後，讓同學公同批評。這種教訓，很有教育的意義。

正誤測驗的缺點。——1、用這種測驗，學生沒有機會可以表示他們組織材料的能力。但有時這種測驗亦可用很複雜的句子；倘使學生能理解而能答對，亦可知道他的組織能力。不過測量發表思想等能力，最好用作文測驗。

2、這種測驗只能測量知識，不能測量技能及做事的能力。但其實不然。如算學的正誤測驗，學生可用算式証明答案之是非；這是測量他的能力及技能。

3、正誤測驗不能告訴我們什麼地方或什麼緣故，學生有錯誤；故這種測驗並不是診斷的。可是教者如欲診斷學生困難的地方及困難的

原因，則可應用較詳的測驗補充之。

正誤測驗的用處，編制的方法，及其優點和缺點，上面已經詳細討論。這種測驗對於讀法教學，尤為適宜。茲將其應用方法，詳述於下。

1、根據學生讀書的程度，分一班為兩組。——先用標準讀法測驗測量學生的程度；然後分學生為兩組。兩組的程度，必須相等。其方法先將測驗的成績依次排列。最多的分數為第一，次多的為第二，餘以類推。然後將第一歸入第一組，第二歸入第二組，第三歸入第二組，第四歸入第一組，第五歸入第一組，第六歸入第二組，第七歸入第二組等等。如是，兩組的程度可以絕對相等。試再用表說明之。

表 十二
依次分組表

甲組	1	4	5	8	9	12	13	16	17	20	21	24	等
乙組	2	3	6	7	10	11	14	15	18	19	22	23	等

照此方法分組後，每組各自練習讀書能力；且本競爭的精神，以謀優勝他組為目的。

2、每個星期終了時，教者用正誤測驗測量他們的程度。其步驟如次。

- (a) 選擇測驗的材料。——普通就可在讀本上選擇適宜的材料，為測驗用。但有時用歷史地理等材料亦可。
- (b) 測驗時，須注意理解及速率兩方面的成績。
- (c) 假定某某讀本之第十及十一兩頁，可以代表其餘各頁的程度，

則這兩頁無論學生曾已讀過與否，均可為測驗用。教者可用上面所討論的編制正誤測驗方法，在這兩頁書上，選擇二十句正誤的句子。測驗時，對學生可用下面一段指導：

『你們各須預備一枝鉛筆及一張白紙。紙的上邊各自寫你們的姓名。然後取出讀本，從第九頁最後一段讀起。起初我同你們一齊朗讀，等到這頁最後一字讀完，你們須自己繼續靜讀第十及十一兩頁。讀時不要出聲。讀得愈快愈好；但須懂得意思。讀完後，我要問你們幾個問題，測量你們能記得多少。當你們讀完第十一頁的最後一個字時，即將書本閉了；眼看黑板上所寫的數目，並將那數目寫在你們的紙上。』

教者計算讀書時間的方法，自第十頁的第一個字起，過十秒鐘，寫 10 於黑板上；過二十秒鐘，將 10 拭去，寫 20 於黑板上。餘以類推；至全體學生讀完為止。前段說每個學生讀完時，須將當時黑板上所寫的數目，記錄紙上，就是這個數目。此後教者須用下面一段指導，測量理解能力。

『你們每人須在紙的左邊，自上至下，寫 1, 2, 3, 4, 5, ……… 20。關於你們現在所讀的材料，我將問幾個問題。每個問題，你們可用「是」或「否」答應。如遇到不知道的句子，任意猜測可也。若是第一句句子是對，寫個「是」字；若是錯，寫個「否」字。餘以類推。』

若是學生尙未能寫字，則可以叫他們用校對的記號「∟」代替「是」字；用「×」代替「否」字，若是被測量的兒童年齡較幼，則測

量前須有一種學習，使他們知道這種測驗的性質，教者讀問句時，聲音須很響亮明白，不宜太快；且宜使每個學生有充分的時間答應問題。

當全體學生答完問題時，叫他們將紙反一面；再在左邊，自上至下，依次寫1至20。然後叫他們重開書本，試看他們參觀書本時能答多少問題。

校對時，教者報告對的答案，學生自己校閱；並且自己計算分數。有時互相校對亦可。如有未曾答的句子，亦作錯誤計算。每個學生的分數，須用錯誤的倍數，去減句子的總數。若是一個學生第一次測量有六個錯誤，第二次測量有三個錯誤，則他的理解分數當如下：

$$\text{記憶的理解分數 (Memory comprehension score)} = 20 - 2(6) = 8。$$

$$\text{對閱的理解分數 (Visual comprehension score)} = 20 - 2(3) = 14。$$

此後，告訴學生第十及十一兩頁書的字數，叫他們以每分鐘為根據，計算所讀的字數，此即為速率分數。

再者，學生須自己記錄每次測驗的成績，先後互相比較，特別注意速率的進步。可是這種比較，有時不甚公平，因為讀的材料時時變換，不能有相等的難度。

同時教者當計算全班或全級的记忆理解分數，對閱理解分數，及速率分數的平均數。前後比較，以觀教學的成績。

用這種非形式的靜讀測驗，測量一班四年級學生的程度，其成績表列如下：

表 十三
一班四年級學生的靜讀成績

學 生	記憶理解分數	對閱理解分數	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
平均數	7.4	11.4	136.3

~~~~~  
上表可給我們下面幾種有價值的貢獻：

1、A, E, 及H三個學生的讀書能力, 非常幼稚。他們每種成績都比平均數低。學生 E 竟至不能讀一句書! 這或因訓練不足, 或因教學不宜, 或因他的智力太薄弱, 或因另外種種緣故。教者對於第一層, 當調查他的教育歷史; 對於第二層, 當用別的教學方法, 看他進步怎樣; 對於第三層, 當用智力測驗測量。

2、學生D的記憶力不好。看他的記憶理解分數, 就可知道。至於他的對閱理解分數, 比較一般程度都要高。他若他的速率成績雖不慢, 亦不快。由此可知他的記憶理解成績不好, 不是因為讀得太快的緣故。

3、學生F的記憶理解成績甚劣, 對閱理解成績甚優。這不是因為他的記憶力薄弱; 實際上, 因為他讀得太快。看他的速率成績, 就可知道。對於這種學生, 教者當限制他的速率, 使他注意記憶方面的成績。

4、依據記憶理解及對閱理解的成績而論, I 是很好的學生; 可是他讀得很慢。故此後他應當特別注意速率方面的成績。

5、L及M兩個學生的讀書程度很好; 每種成績都比平均數高。

6、這種測驗尚有另外的用處。例如校對成績時, 學生如遇有錯誤, 即須舉手表示; 故當時即可討論, 且全體學生可以知道。再者若有一句句錯誤的人數甚多, 則教者就可知道那一部分, 尚須特別教學。

用非形式的測驗, 測量朗讀(Oral reading) 的速率及品質。——葛萊(W. S. Gray) 曾有很精密的實驗, 測量學生朗讀的速率及品質。他的方法, 無論何人, 均可應用。測量時, 須顧及下面數種條件:

1、必須個人單獨測量。

2、須有很清靜的地方。如是，別的學生不能察知測驗的情形，而獲意外的便宜。

3、教者給學生應讀的材料；幷說：『我願意你朗讀這頁書。當我說「開始」，你從第一個字讀起，讀到我說「停止」時爲止。若是你碰到難讀的字，你盡你的能力繼續讀下去，不要叫我幫助。』

4、測驗時，若是學生碰着困難的字，猶疑讀不出，教者可以幫助他；但須當作錯誤計算。倘使讀的一頁之第一個字不是一句句子的起頭，則可等到學生讀完那句後，再記時刻；至他讀完爲止。速率即以每分鐘所讀的字計算。當學生讀時，教者並宜記錄他的錯誤。葛萊分錯誤爲六種；茲列舉如下：

(a) 一個字完全讀錯。

(b) 字的一部分讀錯。

(c) 遺漏。

(d) 代替。

(e) 添加。

(f) 重覆。

但除此以外，照我的意思，尚有兩種錯誤，應當記錄：(a) 交換字的位置，(b) 段落讀錯。

至於計算朗讀品質成績的方法，就是以100乘錯誤的次數，然後以所讀字的總數除之。所得之商愈大，即成績愈壞。

學生平均的成績，無論速率或品質，就是一班或一級的成績。無論



個人或全體的成績，均須繼續保存，以便先後比較，而觀進步。

三、標準量表——說明這種標準量表的用度，須參照前章的討論。前章所載的表十一，就是這節所將討論的材料。

1、預定年終的讀書年及智力年。——每個學生年終時應有的讀書年及智力年，必須預先決定。因為 I. Q. 是學習功課如讀法等成敗的索引，故可用他為根據，以決定學生年終的讀書年或智力年。若是一個學生的 I. Q. 是 90，則他經過一學年（十個月）的學習，如與 I. Q. 是 100 的學生比較，可以希望有百分之九十的進步。所以預算他的年終讀書年或智力年，就是他的年初讀書年或智力年加十個月的百分之九十的進步。若是一學年只有八個月，則當僅加八個月的百分之九十的進步。例如表十一 A 學生的年終讀書年，當是  $121 + \frac{100}{10}$ ，就是 131 個月。他的年終智力年，當是  $121 + \frac{100}{10}$ ，亦是 131 個月。B 學生的年終讀書年，當是  $130 + \frac{122}{10}$ ，就是 142 個月。他的年終智力年，當是  $150 + \frac{122}{10}$ ，就是 162 個月。每一學年都以十個月計算的。

I. Q. 可以做預定一般複雜學科進步的標準。可是對於書法等，並無實效。故如欲預定書法的進步。教者必須另外測量特別能力。如缺乏這種測量，則依照教者的主見，任定一個目標亦可。

2、規定年終時應達到的目標。——每個學生的讀書目標，就是年初時所預定的年終讀書年或智力年。普通以讀書年為目標，比較以智力年為目標好；因為以讀書年為目標，學生可就其能力而謀相當的進步。若是一個學生，從前進步很快，已經超過他的目標，他仍可繼續上

進，謀該學年應有的進步。他若以智力年為目標則不然。再者以智力年為目標，若是一個學生的A.Q.低於100時，例如表十一的學生B，則在一學年內，當做很難的功課，才能趕上他的智力應有的程度。他若學年太短，用智力年為目標亦不適宜；因為時短課重，很難完成責任。故教學讀法，用讀書年為目標，較為適宜。

這種預定的目標，乃是最小的限度，並非最大的限度。他并不包含學生從前欠缺應補的功課。故不管學生的A.Q.怎樣，不當以此為最高目標。這種目標，不過限定學生無論如何，必須及格而已。教者仍宜用種種方法，使學生超過這種目標。但是教者亦不宜有最高限度的目標，因為另外功課或將受此影響，而減少學習時間，致有不良的成績。可是讀法一科，對於其他功課，都有補益，目標稍高亦無妨。

普通教學，並沒有一定可見的目標；所以學生對於進步一層，沒有什麼責任。他們不知道當時的程度怎樣，亦不知道將來的程度應當怎樣。今學年開始時，先測量學生的程度，而預定將來的目標，乃是救濟這種缺點。

普通教學，有時雖有目標；但往往過高，不近事實。這因教者自信力太高，或因他們的希望太奢之故。如作文等的目標，就往往有這種毛病。

普通教育所定的目標，雖或能適合一般的程度；但若全體學生同一目標，則為害仍大。因學生程度不同，進步必不能一致。這種普通的目標，實違背心理學的根本原理。今若根據學生的智力，規定他應當達到的目標，則就可無這種弊病了。

3、用圖形表示目前及年終的目標。——教學不應僅有年終的目標，而且應有每月的目標。這種目標，學生且宜自畫圖形表明之。預定每月目標的方法，就是年終目標預定後，若是一學年是十個月，則以十除之；若是一學年是五個月，則以五除之。例如表十一學生 A，年初時讀法分數是 40，所預定年終的分數是 43，故他這一學年應得 3 點 ( $43 - 40 = 3$ )；每月應得十分之三點 ( $3 \div 10 = .3$ )。教室內應有一張大紙，每個學生可將他如此預定的每月目標，畫在上面，以資觀察。

4、用圖形表示全校各班的目標。——若是一校中各班都有規定的課程，並且預定目前及年終的目標，則學科進步必可加速。這種目標，宜用圖表揭示於公眾共見之地。每班每月的進步，就可在此圖表上表示之。如是，學生可以知道何班先達到最小限度的目標。個人的成績若表示於衆，低能學生未免有羞恥之心；可是團體成績如此表示，就可免除這種情形。

目標在教育上的意義。——用教育量表為教學的方法，一般教育家往往反對。他們受裴絲德拉齊 (Pestalozzi) 的影響，每以為訓練 (drill) 是一種乾燥無味的方法。可是這種態度，表面上雖是可取；實際上並無價值。訓練倘得其法，未必沒有興味。且一種技能之養成，不特自身很有價值，且於將來的教育有很大的關係；訓練實是一種不可缺少的手續。總之，教者若無目標，則必不能有適當的方法，可使學生進步；學生若無目標，則必無適當的努力，去求進步。有明瞭的目標，然後有相當的努力。教學無論何種學科，每月須測量學生的成績，前後比較其進步。倘使經過一個月的教學，而尚沒有相當的進步，則無論教者

或學生，均須探求原因，而謀補救的方法。

## 第五章

### 評斷教學效率的測量

此章所討論的材料，亦是根據第三章所載的表十一；故討論時，須時與上面兩章相聯絡。

一、計算年終成就商數，決定一年教學的效率。其方法如下：

- 1、照每個學生的年終讀法分數，而定年終讀書年。
- 2、以預定的年終智力年除年終讀書年，即得年終成就商數。這就是這一年教學效率的表示。
- 3、這年終成就商數與年初成就商數相減，就是兩年教學效率優劣的比較。

看表十一，學生A的年終讀法分數是43，應得年終讀書年130個月，可是他的預定年終智力年為131個月；以這智力年除年終讀書年，得成就商數99。他是差不多達到預定的目標。學生B這年進步很快，他的年終成就商數比較預定的大得多。但是他的程度，若就他的智力而論，還是落後的。至於學生C，不特保持他原來很高的成就商數(112)，而且增加4點。

這班學生的年初平均讀法分數是36.4（見表十一最末一格），預定的年終平均讀法分數是39.8。可是年終實在所得的平均讀法分數是41.2，超過目標1.4。年初平均成就商數是100，年終平均成就商數是103；故這一年的教學效率比前一年要好百分之三（103—100）。可是

這種比較的結果，是否確實表示教學效率之不同？依據麥柯爾的意思，尚須顧及下面數種條件。

- 1、學生為教學的主體，教師為教學的客體。
- 2、教學惟一的目的，是使學生得到有用的變化(Desirable changes)。
- 3、教學的效率，只有一種方法可以測量：就是測量學生有用變化的數目之增加。
- 4、選擇教師或升進教師，當以學生的有用變化數目之多寡為標準。平常每以教師的學識，人格，容貌，聲音，書法等為根據，實不適宜。現在經過許多測驗，並沒有找到這些情形與學生成績有很大的相關度。
- 5、公允的科學測量，必須當我們測量學生所得有用變化的分量時，始能成立。此種測量必須有先後兩次。其方法，以後再討論。
- 6、公允的科學測量，必須在標準時期(Standard time)測量學生的有用變化之分量。
- 7、公允的科學測量，必須測量標準學生(Standard pupils)的有用變化之分量。前面所討論的成就商數，乃是使學生而為標準學生的一種方法。

滿足上面幾種條件，評定教學的效率，就是前面所討論的表十一之用度。茲仍以讀法一科為例，簡略述之。

- 1、用標準讀法測驗，測量學生原有的程度。
- 2、根據讀法分數，而定讀書年。
- 3、同時用智力測驗，而定學生的智力年。
- 4、用實在年齡除智力年，以求 I. Q.。

- 5、用智力年除讀書年，以求讀法的A. Q.。
- 6、預定年終智力年。
- 7、年終時測量程度，得到年終讀法分數。
- 8、根據年終讀法分數，而定年終讀書年。
- 9、用預定年終智力年除年終讀書年，以求年終的讀法A. Q.。
- 10、以年初的平均A. Q. 去減年終的平均A. Q.，若所得的商是零，則這個教師是有正常的教學效率；若是比零大，則其效率大於正常；若是比零小，則其效率小於正常。
- 11、他若計算算術的A. Q.，則只重覆1, 2, 5, 7, 8, 9, 及10七個步驟可了。先後A. Q. 的差異，就是算術教學效率的表示。其餘學科，亦可用同樣方法，比較教學的效率。

12、計算A. Q. 差數的平均數，就得教師的教學效率之最後評定。弗萊辰 (Franzen) 介紹下面這個公式：

$$\text{教學效率} = \frac{(\text{讀法A. Q. 的差數}) + (\text{算術A. Q. 的差數}) + \dots\dots}{\text{學科數目}}$$

照此公式，表十一所表示的教學效率，當為：

$$\text{教學效率} = \frac{103 - 100}{1} = +3。$$

二、評定學習的效率。——有一個學生，天資很愚笨，他的父母及教師都不知道。他們以為他的成績不好，因為他很懶惰的緣故。二年以前，他與同年天資較優的學生同班，致不及格而退級，他的父母得到報告，當然不快活。退級後，運氣碰得好，他班內同學的天資，比他還要笨；所以年終時，他的成績比較好得多。他的父母得到報告，覺得很快活，以為他們的兒子這一年得到適當的刺激，成績進步了。他的教

師亦很快活，以為他的教學效率比較別人好。他從前的先生亦很快活，以為這個學生這一年的進步，是受他的鼓勵緣故。學生自己亦很快活，因為他的成績不復如從前那樣壞了。但是第二年升級後，又不幸碰得智力較優的同學，他年終的成績又不好了。他的父母及教師，又得到一種不快的感想。

上面一段故事，並不是罕見的。在普通教學情形以下，類似這樣的人很多。這種成績的好壞，並不是教學效率之不同，學生之努力有異；實是不以智力為標準，而評定學生成績之優劣的緣故。

不特劣等的學生如此，就是優等的學生亦常受這種影響。普通有許多天賦很高的學生，因為他的父母及教師不知道，而他的成績每低於他的能力所可有的程度！救濟這種弊病，除求成就商數以外，并無他法；因為成就商數不僅測量教學的效率，而且測量學生個人學習的效率。

三、用實驗的方法，選擇學習方法，學習材料，及教學方法。——什麼學習或教學的方法比較好？什麼課本比較適宜？什麼練習測驗比較有用？什麼組織比較妥當？欲解決這些問題，除用最精密的科學實驗，沒有另外的方法。此地詳細的情形，固不能論及；但其最重要的條件，不得不略述之。

1、宜用很真確很充分的測驗，分學生為兩組。兩組學生的程度，須彼此一樣。所謂一樣的程度者，非僅指平均數而言；實際上，甲組每個學生須有他程度相當的同學在乙組內才好。他若兩組學生的人數加多，亦可增加兩組程度類似的情形。

2、在實驗期內，各種情形須絕對一樣。例如一種實驗比較靜讀教學的方法，則除甲組用A法教學，乙組用B法教學外，其餘情形，必須一樣。

3、實驗要點，始終當有一樣的強度。

4、每組實驗終了時，須有精確充分的測驗，測量成績。

5、比較及評論甲乙兩組的成績。

6、根據實驗所得的結果，論斷方法的優劣。

7、用統計方法，說明結論的信度。

四、學校成績的測量。——測量學校的成績，可以用下表所示的方法說明。

表十四

某校每年級每測驗的成績與標準分數比較的情形

|         |  | 讀法分數比較 |      |      |      |      |      |      |
|---------|--|--------|------|------|------|------|------|------|
| 年級      |  | 三      | 四    | 五    | 六    | 七    | 八    | 總平均  |
| 1919年中數 |  | 5.5    | 10.8 | 18.0 | 23.0 | 24.6 | 25.9 |      |
| 標準分數    |  | 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 |      |
| 年級基位    |  | -0.6   | -1.0 | -0.5 | -0.2 | -0.8 | -0.9 | -0.7 |
|         |  | 填字分數比較 |      |      |      |      |      |      |
| 年級      |  | 三      | 四    | 五    | 六    | 七    | 八    | 總平均  |
| 1919年中數 |  | 16.5   | 29.0 | 38.5 | 42.5 | 48.3 | 55.5 |      |
| 標準分數    |  | 18.5   | 25.0 | 30.5 | 35.0 | 40.0 | 43.5 |      |
| 差數      |  | -2.0   | 4.0  | 8.0  | 7.5  | 8.3  | 11.0 |      |
| 年級基位    |  | -0.4   | 0.8  | 1.6  | 1.5  | 1.7  | 2.4  | 1.3  |



| 年 級    | 其餘各學科分數比較 |      |      |      |      |      | 總平均  |
|--------|-----------|------|------|------|------|------|------|
|        | 三         | 四    | 五    | 六    | 七    | 八    |      |
| ⋮      | ⋮         | ⋮    | ⋮    | ⋮    | ⋮    | ⋮    | ⋮    |
| 平均年級基位 | -0.6      | -0.1 | -0.2 | -0.1 | -0.2 | -0.3 | -0.2 |

上表的讀法如下：在 1919 年，用讀法量表測量各級的讀書能力。三年級的中數是 5.5，這量表的標準分數是 8.0，相差為 -2.5；故該級的讀書程度，比較標準程度低十分之六的一年級，即所謂年級基位(Grade unit) -0.6 是也。他若四年級的中數是 10.8，標準分數是 15.0，相差為 -4.2；故該級的讀書程度，比較標準程度差不多完全低一年級，即所謂年級基位 -1.0 是也。其餘各級，以此類推。各級的年級基位的平均數為 -0.7，這就是表明這個學校各級的讀書程度，比較標準程度低十分之七的一年級。其他各種測驗的每級年級基位相加，而求其平均數，則即為每級各種學科的成績與標準程度比較的情形。若各種測驗的各級的年級基位相加，而求其平均數，則即為各級各種學科的成績與標準程度比較的情形；這就是全校成績的表示。

※上表所介紹的新法，就是先求年級基位；然後比較成績。因為每種測驗的記分標準，彼此不同；故各測驗中數與標準分數比較的差數，不能互相比較。若一個測驗自三年級至八年級分數相差甚大，另一個測驗自三年級至八年級分數相差甚小，則用這兩個測驗測量所得的分數與各測驗的標準分數相減的差數，必不能就其表面價值而互相比較。今用年級基位，就是救濟這種弊病。其計算的方法如下：

以每年級某測驗的標準進步量，除該級中數與標準分數的差數，即得年級基位。例如上表讀法測驗，三年級的標準分數為8.0，八年級的標準分數為30.0。自三年級至八年級的進步量為 $30.0 - 8.0 = 22.0$ 。每級的平均進步量為 $\frac{22.0}{5} = 4.4$ 。故若一個年級的讀法成績低於標準分數4.4時，就是低於標準程度一年級（ $\frac{4.4}{4.4} = 1$ ）。上表三年級讀法分數與標準分數相差是-2.5，應用公式， $\frac{-2.5}{4.4} = -0.6$ ，此即三年級讀書程度低於標準程度的年級基位。上表四年級讀法分數與標準分數相差是-4.2，被4.4除，差不多等於-1；故該級讀書程度比較標準程度差不多要低一年級。其餘類推。至於填字測驗，三年級的標準分數為18.5，八年級的標準分數為43.5。自三年級至八年級的進步量為 $43.5 - 18.5 = 25.0$ 。每級的平均進步量為 $\frac{25.0}{5} = 5.0$ 。故每級填字分數與標準分數的差數當以5除之，然後可得每級的年級基位。

這種年級基位，尚可用他法計算。每校自己各年級進步的平均數可以代替每級的標準進步量。可是照此方法計算的年級基位，各校必不相同；互相比較成績，非常困難。至若用標準進步量計算，則各校的程度，就可根據年級基位，互相比較。

再用相近兩級的標準級距，代替各級一樣的標準進步量，計算年級基位亦可。例如三年級的讀書中數與該級的標準分數的差數，用三年級至四年級兩級標準分數的級距除之，亦可得一種年級基位，表示進步的情形。但有時高一級的標準分數，甚至等於或少於低一級的標準分數。倘使兩級的標準分數是一樣，則進步必等於零。無論何種差

數，若被零除，則所得的商，必是無定；故此法有時不能應用。

再有一種較好的方法，是用鄰近兩級或三級的平均標準進步量，除該級中數與標準分數的差數。用這種方法計算年級基位，可以免除我們上面所用方法的一種困難。普通學科的進步低年級每快於高年級。上面我們所用的方法，是以自三年級至八年級的平均進步量除各級中數與標準分數的差數，實未能顧及此層。可是用鄰近兩級或三級的平均標準進步量計算，甚費時間，不宜採取。

標準分數往往以學年終了時(六月)為標準，今若在十一月時，測量學生的成績，則計算年級基位，當用何法呢？請參觀下表。

表十五

某校讀書成績與標準分數比較的情形

| 年 級        | 四    | 五    | 六    | 七    | 八    | 平均數 |
|------------|------|------|------|------|------|-----|
| 十一月時分數     | 39.6 | 44.9 | 49.0 | 54.9 | 59.5 |     |
| 下年六月時應得的分數 | 43.3 | 47.8 | 53.1 | 58.1 | 61.2 |     |
| 標準分數(六月時)  | 41.8 | 48.0 | 53.7 | 58.3 | 60.9 |     |
| 差 數        |      |      |      |      |      |     |
| 年 級 基 位    |      |      |      |      |      |     |

註：上表所留的差數，年級基位，及平均數諸空白，讀者須照上面所討論的方法，自己計算。

自十一月至下年六月，尚有七個月，占一學年十分之七。上表就是以這七個月為標準，根據十一月時的成績，計算下年六月時應得的分數。其方法如下： $(44.9 - 39.6) \times 7 = 3.7$ 。 $39.6 + 3.7 = 43.3$ 。這就

是四年級六月時應得的分數。 $(49.0 - 44.9) \times .7 = 2.9$ 。 $44.9 + 2.9 = 47.8$ 。這就是五年級六月時應得的分數。其餘以此類推。可是計算八年級的六月時分數，則為另外的問題。其計算方法，須用標準分數。如  $(60.9 - 58.3) \div (58.3 - 48.0) = .25$ 。 $(58.1 - 47.8) \times .25 = 2.6$ 。 $59.5 + (2.6 \times .7) = 61.2$ 這是八年級六月時應得的分數。

參觀前面表十四，我們可分別評定下面幾種成績：

- 1、每年級每種測驗的成績。
- 2、每年級各種測驗的成績。
- 3、各年級每種測驗的成績。
- 4、各年級各種測驗的成績。

但有四點應當注意的：

1、若一個教師在各級內，只教一門功課，則測量他的教學效率時，須計算各級每種測驗的平均年級基位。

2、若一個教師教一級內所有的功課，則測量他的教學效率時，須計算每級各種測驗的平均年級基位。

3、用這種方法評定教學的效率，且可為教學方法改良的指導。

4、評斷學校教學的成績，且可為經費分配等的標準。

但一個學校的成績：不能任意與他校比較。若是下面所舉幾種條件不一致，則比較時必不能公允。

1、學生的數目必須固定。若學生時常轉學，則測量該校學生的程度，如測量他校學生的程度一樣；其結果，難免使該校的成績太低或太高。

2、學生的智力必須類似。用智力測驗可以調查這種情形。如用成就商數比較成績，則就可沒有欠公允的弊病。

3、家庭的情形必須一樣。有些家庭，父母教學的分量，差不多如教師所教學的一樣。如是，與沒有父母教學的學生比較成績，必不公允。

4、實年的速進或留退學生的數目，必須彼此相彷彿。若是一校中，年輕聰明而被留退的學生多於年大愚笨而被速進的學生，與其他情形恰恰相反的學校比較成績，則即教學方法一樣，而前者必勝於後者；故欠公允。

5、受生理方面發達的影響，而才能特別增加的情形，亦須彼此一樣。此層須俟實驗解決。

6、學力遷移的情形，亦須一致。此層若比較教師的教學效率，尤為重要；因為有許多功課的進步，每因受另外功課進步的影響，並非完全是本科教學的結果。

7、注重功課的情形，亦須一致。有些學校因社會環境不同，職業需要有異，有些功課必特別注重；故成績亦必較優。比較教學效率時，必須注意此層。

此外尚有幾種情形，應當注意：

8、學年的長短。

9、教師的學識。

10、學科的多寡。

11、學校的設備等。

各學校的成績互相比較，可以促進全國的教育進步。教育行政官且

可因此知道何處教育應當特別提倡，何種教育應當特別注意。換言之，教育測量可以為診斷全國教育之用；且可因此規定全國教育的標準程度。

## 第六章

### 職業指導所用的測量

兒童在學校內，尚有教師相當的指導。可是一離學校而入複雜的社會，則自沉自浮，再沒有人管理了。職業選擇，完全是碰機會的！幸而碰得好，還有成功的希望；不幸而碰不好，則不特個人一生受害，即社會國家亦受其影響。所以職業指導是現在很重要的問題；無論何校，不容疎忽的。

一、職業指導的功能，可分為二：

1、幫助學生找到一種職業，對他的興味及能力很適宜，對於社會亦很有利益。這種職業，大約分為三類：

(a) 第一類職業，有相當的收入；同時自己學識有繼續進步的機會，位置有繼續升進的希望。

(b) 第二類職業，收入甚微；但與自己意志相合。同時且可受一種特別的練習或訓練。

(c) 第三類職業是暫時的。但在短時期內，可以得到很大的經濟幫助；由是可以繼續上進求學。

2、使學生在求學時代，有一生事業的動機，此即哈佛大學 (Harvard University) 校長伊里 (Eliot) 所謂 “The life career motive” 是也。學

生對於學習自身有興味者，固不乏其人；但此種人實不多見。然若有一個終身事業的動機，則興味必可成立，努力必可增加，成績必可進步。

二、職業指導的方法，須有下面幾種步驟：

1、須先調查各種職業需要人才的情形；同時須研究報酬及工作等狀況。

2、學生自己須研究這種調查的結果。如是，他可選擇相當的職業，而有相當的預備；且可因此得到普通重要的見識。

3、用各種方法，調查或測量學生對於各種職業的興味及能力。

4、選擇職業須得職業指導員的許可及指導。

5、有充分職業教育的預備。

6、預備職業時，須有適當的指導。如選擇的職業不適宜，當更換之。

7、受職業教育後，須有充分的實地練習。

8、介紹入相當職業做事。

但此地只能就其關於測量一部分而研究之；至於職業指導的詳細情形，不能論及。

三、智力與職業的關係。——若是一個小孩來問教師，什麼職業對他最適宜？什麼功課他應當預備？教師答應這個問題以先，應該知道什麼？教育測量能幫助他多少？第一，須用智力測驗測量各種職業代表人物的智力；因此可以知道各種職業所需要智力的限度。推孟曾有這種測驗，找到現在各種職業代表人物的智力，實參次不齊。專門學校學生的智力甚至有比走江湖者尚不如！但是比較智力商數的中數

或下部四分點 ( $Q_1$ )，則專門學校學生的智力實高於走江湖者；上等職業家的智力實高於小夥友，救火夫，警察，及汽車夫等。第二，須測量學生的智力，究竟宜於學習何種職業？教師或職業指導員並宜告訴學生，何種職業對於他的智力最為適宜。學生雖可選擇較高的職業；但職業的要求若超過智力愈遠，則成功的機會愈少。時運通，家庭有勢力，或另外有好的品性，則能勉強敷衍，亦未可知；但希望成功，實不容易。故指導學生時，須注意及此。此外尚宜注意興味。因為同一智力階級的職業，每有多種；學生可擇其興味近者而習之。用這種方法，幫助學生選擇職業，不特對於學生自身很有利益，即對於僱用者亦很有貢獻。有些營業每因用人不當，而致效率不良；每因效率不良，而致失敗！今若學生各有他們相當的職業，則僱用者自然可以間接受其利益了。

決定職業所需智力的限度，及找到學生應習的職業，實是一種很困難的問題。照桑戴克的意見，普通智力有三種：1、抽象智力 (Abstract intelligence)，2、社交智力 (Social intelligence)，3、工藝智力 (Mechanical intelligence)。我們平常所測量的智力，並不限定何種智力的。

在單獨一種智力的範圍內，個人所有的各樣能力，彼此往往有很高的相關度。至於兩種智力的各樣能力，則雖有正的相關度，但其高度終不及一種智力範圍內各樣能力的相關度。富於抽象智力的人，往往強於處理抽象的觀念。學習科學原理，及算學或化學的公式等，最為適宜。此種人物往往宜為律師，科學家，及神學家等。富於社交智力的人，往往強於處理人類相互的關係。若這種人有領袖的才能，則宜為



軍官辦事人物等。至於這種人有服從的性質，則宜為政治家，商業家，或很和順的婦人。他若富於工藝智力的人，往往宜於學習汽車，輪船，飛艇，及其他各種工作。

測量這三種智力，並不能一樣真確，也並不是一樣容易。普通所有的智力測驗，大都是測量抽象智力。然同時亦有小部分測量其他兩種智力。

四、品性及體質與職業的關係。——一個人選擇職業，適宜與否，並不僅以智力為標準；同時亦宜視品性及體質之如何。智力與道德雖有正的相關度，但究竟不是一樣的東西。至於體質與智力或道德的相關度，則更低了。

職業各有道德的限度，適如各有智力的限度一樣。有些職業只要身體強健，道德標準並不必十分高。例如許多下等職業，需要誠實，勤勉，文雅，同情心等，甚為微少。他若高等職業，則大不相同。如大理院的院長，戰爭時的軍官等，則其責任重大，非有高尙的品性不可。

至於體質方面亦如此。如軍隊之須有強壯的身體，及競力職業之宜有相當氣力等，都是普通所熟知的。他若開礦運貨諸工作，亦非強有力者為之不可。幸而這方面的測量，比較容易，而且很真確。如醫藥測驗及體格測量等，都已有很好的成績；惟賴職業指導者善利用之而已。可是容貌醜美等，則仍非客觀的方法可以測量。因其與測量品性的情形類似，故拼在下段討論。

測量品性及容貌醜美，不如測量智力那樣容易。在現在時代，差不多完全用主觀的方法測量。如誠實，文雅，忠信，勤勉，優美等，我們只

好依據個人朋友的評斷及個人自己的意見而定上下。前者當然比較後者真確。倘所有評斷的人都是很能幹的評斷家，而且人數很多，知道被試者的情形甚詳，沒有不公平的態度，則其確度尤可增加。但所困難者，各人沒有一定的標準，可以評斷別人。故有時自己批評，亦是必要。假使個人自知甚明，則當其選擇職業時，或不至於故意自為揚誇或貶抑也。

霍林威士 (Hollingsworth) 在他所著的職業心理學 (Vocational Psychology) 書中，報告自己評判確度的研究。克德爾 (Catell)，勞斯威塞 (Norsworthy)，威爾士 (Wells) 等，亦有同樣的測量。霍氏 在大學三年級百五十個女生中，選擇二十五個互相熟悉的人，互相評判。每種品性底下，如脾氣，智力，優美，鄙俗，文雅，和善等，依次排列各人的地位，評判者自己亦在內。他找到結論，有如下數種：

- 1、自己評判的錯誤，比他人評判的錯誤大一倍。
- 2、假定他人的評判是真確，則自己評判有一種很普通的錯誤：就是對於不好的品性，往往低其評判；對於好的品性，往往高其評判。但據克德爾的研究，科學家自己評判，則無這種普通的錯誤。
- 3、對於好的品性，若一人能很真確評判自己及他人，則他必有這種品性。至若對於不好的品性，則適相反。
- 4、評判自己很真確的人，往往評判他人亦很真確。可是品性不同，則這種情形亦不同。

根據上面的結論，我們可以知道自己評判，只有他人評判的一半確度。但若用男人為被試；所有各種品性有一定的定義；評判時，不用自

己的主觀，而用他人評判自己的想像；或評判時，知與自己選擇職業有關係，則這種自己評判的確度，或可增加，亦未可知。總之，這種方法亦有相當的價值，職業指導誠可利用之。

五、個人趨向與職業指導的關係。——智力與職業的關係，及智力測驗對於職業指導的貢獻，上面已經詳細論及。但智力測驗的功用，亦有他的限度。下面三段話可以說明這個意思：

1、若是要指導學生得到特別適宜的職業，必須應用各種特別的測驗。如算學，圖畫，音樂等能力，非用特別測驗診斷不可。平常所用的智力測驗，實不能測量特別職業所需的要素。所以我們必須用各種類似的，或分析的測驗，測量個人對於特別職業的趨向。如桑戴克的書記測驗(Tests for Clerical workers)，西宜華(Seashore)的音樂測驗(Tests of musical capacity)，萊傑(Roger)的算力診斷測驗(Test for the diagnosis of mathematical ability)等，皆其例也。

2、我們應當知道職業愈低，需用智力愈微；故智力測驗愈不重要。如簡易計算及校對諸工作，智者愚者均可爲之，其效率必無甚差別；因此等工作毋需很高的智力。倘使愚者對於這種工作有充足的趨向，則他所做的工作，或可比智者所做的好。職業愈下，志趨愈低，應用智力的機會愈少。職業愈高，責任愈重，應用智力的機會愈多。職業的心理學家之有貢獻於公司的業務，不僅有時能選擇智慧的工人，並且有時知道可以毋庸選擇智慧的工人。

3、我們應當知道除出特別趨向以外，尚須注意各種職業所需的特別能力。上智的人去做上等的職業，也遇有失敗的時候；這往往是因

特別能力缺乏的緣故。從前芝加哥大學 (University of Chicago) 的研究院有一個學生，性情和善，智力甚高，進一大工廠做事，用滑尺 (Slide rule) 計算百分比，但竟無良好的工作；人頗以為異。後經測驗，找到他是缺乏算學的能力。但改入他部做事，曾不幾日，成績甚著。可見前所失敗者，實因一種特別能力缺乏的緣故。

可是測量這種特別能力缺乏與否，實不容易。有時亦並非必要。且個人不同處頗為微小，普通職業指導直可忽視之。不過遇到極端的情形時，則不得不用特別測驗測量。

六、已有職業技能與職業指導的關係。——有些工廠不僅僱用工人，而且訓練工人。其中一部分人，訓練期滿以前，必將被淘汰。彼等勢必將謀工於他處，這是第二個僱用者的問題，與職業指導員的問題不同。他不僅應當測量他們的學習能力，而且應當測量他們已有的技能。可是學校內若有職業的訓練，則職業指導員的責任，當亦與第二個僱用者一樣。測量這種已有技能的測驗，大約可分為三種：1、口問測驗，2、圖畫測驗，3、動作測驗。彼此均可測量特別技能之存在與否。下面幾句問話，乃是測量鐵匠技能的例子。致平器具 (Flatter) 什麼用？當一塊鐵在火內，有火星飛出，表示什麼？使鋼絲有相當的硬度，你用什麼東西？做鐵的凹線，你用什麼器具？除出口問以外，尚須用圖畫測驗，叫被試者一一說明各種器具的用途。但有些人知道做而不知道說，故除出口問以外，尚須用動作測驗以補充之。

七、興味與職業指導的關係。——職業指導第一步須測量學生的智力，品性，體康，趨向，及職業需要的技能。第二步須表列學生最有成

功希望的職業，使他自己及他的父兄知道。如此，學生可從這些職業中，根據他自己的目的及興味，選擇相當的職業。興味愈強，成功的機會愈多。

但個人的興味與智力有衝突時，當怎樣呢？有時因有這種衝突，竟產生很不幸的結果。有的終身努力，而竟無一成；有的可以成功，但無適當的機會。處理前者，有三要點應當注意。分述之如下：

1、智力不能改變；他若興味；則因成熟關係或用機械方法可以使之遷移。換言之，我們可以創造興味；但不能創造智力。

2、一種工作，初時或無甚興味；但將成功時，則必興味突生。他若金錢的酬獎，有價值的稱譽等，亦可以令人快樂而生興味。

3、若超過一定的程度後，興味增加對於成功的效果，並不如相當智力增加的效果。可是這一條不過是一種臆說，現在尚不知道怎樣興味，怎樣智力，才可有怎樣的成就。

興味的改變，乃是職業指導最重要的問題。一個職業家從前或願為工程師，一個詩家以後或願為醫士；往往不能一定。但若小學或中學教課的內容，是代表職業的內容，則桑戴克研究大學學生的興味，有下面數條結論：1、學生的興味，大概是固定的；2、學生的能力，亦有恒而不變；3、自中學至大學時期的興味及能力的變化，大於自小學至中學時期；4、沒有一個很明瞭的時期，興味及能力成為很固定；5、學生的興味，每可為能力的表示。對於第五條，柏里傑士 (Bridges) 及陶林傑 (Dollinger) 曾再有研究，找到興味與能力的相關度是很低，不過<sup>2</sup>或<sup>3</sup>而已。所以他們的結論謂興味不足為能力的表示；興味及能力必

須各有測驗。至於此層，桑氏自己必亦以為應當如此。

關於職業指導，凱雷爽(Crathorne)尤有切要的調查。他研究自中學至大學，對於一生事業的興味，有否中途改變的問題。他在二千個一年級的大學生中，找到百分之五十七人，進中學時候已定有一生的事實；但未進大學以先，這百分之五十七人中，差不多有百分之五十已經改變宗旨。男女間亦沒有什麼不同。由是可知在中學一年級的時候，只有百分之二十五學生已經確決一生的事業，其餘大部分未入職業以先，實沒有一定的目的。職業選定後，還是要改換的。這種不定的狀況，是有害或是有益，不能一定。學校究竟應否有各種職業的設備，幫助學生從早決定職業，乃是教育家很重要的問題。

欲成就上面種種工作，必須有一個職業指導員。他不僅應有真確的眼光，且宜有科學的訓練。此外尚應能分析各種職業，找到需要的各種特性及其分量。至若編制及選擇各種測驗，測量各種特性，亦為必要。再者，他尚宜能應用測量的結果，使學生得到適宜的職業預備；然後分配學生至各方面去，使各有相當的位置。

八、高材生的職業指導。——社會最大的耗費，就是大材小用。一般僱用者竭力想得聰明能幹的人，以為己用。可是被僱用者能否在此得到充分的機會，發展他們才能；而對於社會，是否可有相當的貢獻，彼等每置之不問。彼等為自利計，不惜以他人為犧牲；但社會實不容有這種情形。社會固宜指導個人不至選擇職業高於他自己的智力；同時亦不宜使上智的人埋沒於下等職業中。故若對於個人的自由沒有很大的妨礙，社會應指導個人得到他的能力或天性最相近的職業。因惟

如是，社會始能興旺發達。

社會不能任上智的人自在混雜的職業界中，尋覓他的工作；否則，一入迷路，爲害無窮，將來回頭，就很難了。柯愛(Coy)女士研究上智兒童的志願，找到大部分人將來所喜歡做的事，均低於他們自己的智力。其中差不多沒有一個人願意做一種事，是他的能力所不能及。所以柯氏主張學校不僅當用客觀的測驗測量學生的才能，並宜養成他們有高尙適宜的志願。

九、低能兒童的職業指導。——有些人恐怕指導低能的兒童而入適宜於低能兒童的職業，與平民主義有衝突。其實這種方法不特沒有違背平民主義，而且是很慈善的事業。不僅對個人有利益，而且對社會有很大的貢獻。其理由有三：

1、低能的人在需要智力甚微的職業中做事，容易成功。他能成功，並不是因爲他的智力高於職業，實因爲他只在這種職業中有相當的能力，有相當的競爭心。至於智力較高的人，自然可以升做較高的職業，而得較大的報酬。

2、低能的人在需要智力甚微的職業中做事容易成功，亦因爲社會願意報酬他在此職業中所做的工作之故。有許多人做個清道夫，尙可得到相當的生活；至若去教書，勢必將餓死。這就是社會願意報酬他的清道工作，不願意報酬他的教書工作。

3、低能的人做用力及機械的事，每很有興味；做用心思的事，每有非常的苦痛。有許多人願意掘溝搗石，但不願意讀書，就是這條的例子。

## 第七章

### 測驗材料的準備及提高確度的方法

#### 一、理論的和實驗的提高測驗確度的方法

何謂真確的測驗?——測驗通性有多種：如確度 (Validity)，信度 (Reliability) 及客觀性 (Objectivity) 等。但以確度為最要。所謂確度者，乃指一種測驗所測量的能力與用他種方法所規定或他種測驗所測得該能力相符合的度數。若一個測驗實在測量他所欲測量的東西，而且測驗的全部確有同樣的作用，則這測驗必是真確的測驗。

我們若問一個有經驗的心理學家：『一個測驗究竟是測量什麼東西？』他必將說：『在相當的環境中，測量做測驗紙上所有材料的能力。』我們若再問：『測驗成績好的人，是否實際工作的成績亦好？』他就不能確定這兩種成績真確相關的程度。

我們若是用二十個算學題目，測量一班學生。倘使 A 生做對十八個題目，是否他在別種地方，遇到一樣的題目，亦能做對十八個？實沒有人能確定。再者，若 B 生或 C 生僅做對十四個題目或十二個題目，是否他們在實際生活中，計算的能力成正比例的次於 A 生？亦沒有人能知道。換一句話說 測驗的結果同實際生活的結果，不是絕對相符合的。不過雖是如此，但亦必有多少關係；因為測驗的情形與實際生活的情形並非大相懸殊。測驗的方法進步，則他的確度亦自然可以隨之增高。

增高確度的方法。——若測驗的方法愈近實際生活的情形 則測驗



的結果愈可與生活的能力比較。例如學生拼字的能力，可用下面三種方法測量：1、考調他的書信及作文的成績；2、默寫包含生字的句子；3、默寫生字。第一種方法，最合實際生活的情形，當然是最好。第二種次之；第三種又次之。至於學生文法的能力：可用下面兩種方法測量：1、分析他所做的文字（筆寫或口述）；2、測量他的文法知識。第一種方法當然要比第二種好。可是欲求測驗更合於自然的情形，則測驗的手續必愈麻煩。

測驗類似實際生活的情形，彼此不同。霍林威士曾討論四種根本的測驗：

1、模型測驗 (Miniature test)。——用一種模型的器具代表工作的全部或主要部分。測量時的情形與實際的動境一樣。例如梅克哥馬絲 (Mc Comas) 所製電話板的模型，測量接電話的工人；工人可以實際做應接的事，且同時有最精密的記時機 (Chronometric attachments)，記錄其速率及真確度。

2、樣本測驗 (Sampling test)。——這種測驗，係測量被試者實際做樣本的能力；並非用另外模型測量。例如用打字機及默寫方法等，測量速記者的技能；用抄錄表冊或在電話簿上找地址等，測量書記的技能是。大凡教育的測驗，均屬於這類。我們測量學生讀書的能力，則叫他讀樣本的材料；測量學生算學的能力，則叫他計算樣本的問題。

3、同理測驗 (Analogy test)。——這種測驗，並非測量職業一樣或類似的工作。其實乃是測量職業預備的能力。例如測量學音樂者的資格，則測量他的辨別音色等能力是。在歐戰時，桑戴克等編制一種測

驗，診斷學習飛行者的能力。他們分析與飛行有關係的身心兩方面能力如平衡的感覺(Sense of balance)；傾斜的知覺(Perception of tilt)；及遇感覺方面突然的震動，有相當的鎮靜(Nerve-resistance to sudden sensory shock)等，亦是同理測驗的一個例子。

4、經驗測驗(Empirical test)。——這種測驗，是依據『試行錯誤法』(Trial and error method)所得的結果編制而成。編制者並不如造模型測驗，樣本測驗，及同理測驗時，有一定的目的。他初時試用許多測驗，凡足為實際測量應用者保存之；否則刪去之。普通職業所需要的根本能力，現在尙未有明瞭確定的分析，可為編制測驗的根據；故一般心理學家每謂這種測驗，往往尙比他種測驗好。

一個測驗若沒有不相關切的分子(Irrelevancies)在內，則必與實際生活更相接近。試以算學為例，倘使沒有與算學不相關切的困難。則所有問題，必可更近於實際。不然，若是測驗指導太複雜，則學生必不易明瞭，他們計算的結果，必非實在的成績。再者，若是問題措辭不宜，則亦必將有文字的困難。所以編制測驗時，這種不相關切的分子，必須仔細除去才好。

他若測驗的形式，亦是一種不相關切的分子。這並不單是材料方面的關係；其實用材料的方法，亦是一種重要的問題。至於測驗的形式，大約有如下數種：1、問話法(Question mould) 2、填補法(Completion mould)，3、分類法(Classification mould)，4、對比法(Matching mould)，5、及其他種種。這些形式往往成為不相關切的困難分子。對於兒童，尤其如此。他若關於速率的測驗，則這種困難，甚有妨碍；因兒童的速

率成績，必多少受他適應測驗方法遲速的影響。

推孟討論美國軍隊所用的智力測量 Alpha，有下面一段話。他說：『這個測驗的問題，答應時，差不多都不用書寫的工作；被試只要畫一畫，或畫一個十字×，或做一個記號就可了。』用這種方法，究竟爲什麼緣故呢？因爲1、測驗時，省時；2、校對成績時，省時；3、更加客觀；4、最要緊的理由，是使這測驗要測量所欲測量的東西。這個測驗的目的，原爲測量智力。若須寫字甚多，則差不多要變成測量寫字的速率了。各人畫一畫，畫一個十字×，及做一個記號等，速率都是差不多。他若寫字的速率，則就必不相同了。

至於測驗太長，亦不適宜。因爲太長的測驗，並不能測驗所欲測量的東西；其實將不過測量疲倦而已。所以一個測驗必須有相當的長短，不至有發生疲倦的危險才好。若是有些兒童比較另外兒童容易疲倦，以及要比較年幼和年長兒童的不同時，尤當注意。

疲倦可用下面四種方法免除之：

1、測驗宜短。

2、若是要測驗有相當的信度非長不可時，則應分爲幾部。每兩部中間，可有相當的休息。

3、若測驗有許多短的測驗，則短的測驗應當難易交叉；測驗的性質，且當互相更易。

4、測驗應有各種不同而且具有興味的形式及材料。前面所說智力測驗 Alpha 所用的材料，大都是很有興味，可爲這段的例子。

上面關於防止不相關切分子的方法，並非一定有實驗的根據；不過

我們因此對於有些問題，知所特別注意而已。故應用時，仍宜小心謹慎。若主試者要測量純粹的算學能力，則如文字等不相關切的分子，自當除去；測量純粹的語言能力，則非語言的不相關切分子，亦自當除去。但若要將社會上生活實際需要的算學問題實行測量，則除去不相關切的分子，亦未必十分有價值。例如兒童在店中購買貨物，其他與計算無關，足以擾亂其心思的事情必多。故他必須有充分的注意力，貫徹其目的，而後可無錯誤。可是這種情形，測量時甚難顧及，且不易有相當的設施；故另外用文字等困難代替之，未始非一妥適的方法。再者，所謂不相關切的分子，亦因各種測驗性質不同而異。若一個測驗故意測量兒童實際工作時的複雜精神作用，則這測驗即有特別的形式，亦可說沒有不相關切的分子。有些測驗且故意用這種方法編制，如吳偉士—威爾士的指導測驗(Woodworth—Wells Direction Tests)是。茲摘錄一段為例，以供參考。

“With your pencil make a dot over any one of these letters F G H I J, and a comma after the longest of these three words: boy mother girl Then, if Christmas comes in March, make a cross right here....., but if not, pass along to the next question, and tell where the sun rises.....Give the correct answer to this question: ‘Does water run uphill?’.....and repeat your answer here.....Do nothing here(5—7=.....), unless you skipped the preceding question; but write the first letter of your first name and the last letter of your last name

at the end of this line:”

若測驗包羅愈廣，則確度愈可增加。一個測驗若採用所測量的能力範圍以內所有的材料，則包羅必可甚廣。同時確度亦可因之增加。這種方法，若主試者只欲測量範圍很窄的能力，或教材的一部分時，則甚易施行。即使要詳細測量範圍很大的教材，亦無不可應用：不過需時當多，學生受測驗時，差不多將與學習時間一樣了；故不經濟。所以要測量範圍較廣的材料或能力時，必須先找到比較經濟的方法。

1、一個測驗若採用所測量能力的各種樣本知識為材料時，可以包羅甚廣。例如測量學生能拼多少字，知道多少字，用字的能力怎樣？固不必把字典裏面個個字都加測量。我們若任意採取多少字測量之，亦就足殼了。做這種樣本測驗，我們必須顧及幾種條件：(a)選擇材料時，須採用『任意採取』的方法；(b)必須有充分的數目。所謂『任意採取』的方法，就是我們可以任取字典裏面每頁第一個字或第兩個字，或每第二十五頁所有的字等為材料。什麼字當被選擇，完全是碰機會的。用這種方法編制測驗，我們且可用類似的材料造成另一種測驗，以備與原測驗比較。但用這種方法編制的測驗，只能告訴我們學生有全部知識的百分之幾的程度；除出測驗所包含的材料外，並不能告訴我們他知道其餘什麼材料。故其用處，亦有限度的。

2、欲救濟上段最後一個缺點，我們可用模樣材料 (Type material) 編制測驗，以謀達到包羅豐富的目的。用這種測驗測量技能，尤為適宜。凡屬技能的功課，必含有主要的單位或方法。若測量這些主要的工作，就可完全知道這門功課的能力。例如加減乘除是算學主要的單

位。計算加法時，進一位是一種模樣的工作；計算除法時，指定小數點的位置，亦是一種模樣的工作。諸如此類，不可勝舉。若一個測驗包含所有的模樣工作，則就可算為這門功課能力的代表了。包老 (Ballou) 同孟祿 (Monroe) 的算學測驗，即是根據這種方法編制的。

3、另外一個選擇測驗材料的方法，是依據社會上應用的價值為標準。這種方法，已甚普通；且比上面所討論『包羅豐富』的原理好。我們若應用這條原理，則測量學生拼字的能力，算學的能力，以及其餘歷史等能力，只當以在社會上有價值的材料為限度。如是，我們所得到學校成績的報告，當可更有價值。最明白的例子，是愛禮斯的拼法測驗。他用社會上常用的一千個字編制。其他學科，亦可依照同樣的方法，造成相當的測驗。

4、一個精密完全的測驗，不應僅測量學生能做多少，或能做如何好；而且應測量他能做如何快。這一層對於診斷學生的能力或習慣，非常要緊。其理亦易明白。學校的主要目的，要養成學生執行社會上各種事業有充分敏捷的能力。這不僅需要普通的能力，而且需要做事時有相當的速率。否則，學生自己亦無大益及充分的快樂。故學生的能力可分為三部分測量：(a)分量或難度(How much or how difficult)；(b)品質或真確度(How well or how accurately)；及(c)速率(How rapidly)。例如測量讀書的能力，這三方面都應測量。

提高職業測驗(Trade test)的確度的方法。——我們如何可以知道一個測驗是測量所欲測量的能力？我們只能用求相關度的方法以決定之。一個兒童受智力測驗所得的成績，是否與學校或外界的批評相

符合？這種問題有兩種求相關度的方法可以解決：

1、用一個測驗，測量許多學生；並保存他們測驗的成績。然後繼之以長時期的精密觀察，看他們實際動作的優劣，是否與測驗的結果成正的相關度；若是相關度甚高，則這測驗必是真測量所欲測量的能力了。如是，用這測驗就可預先知道兒童作事的能力，排列他們的等級。

2、用一個測驗（我們正欲求其確度）測量一班兒童。若是他們的能力我們已經知道，則比較測驗的成績是否與另外方法所規定的優劣相符合，就可知道這個測驗的確度之如何。現引美國 1918 年的心理公報 (Psychological Bulletin) 中一段文字來說明。盧爾 (Ruml) 雷賓孫 (Robinson) 及 赤珀門 (Chapman) 諸人是用這種方法求得測驗的確度。利用這一段文字：且可說明職業測驗編制的步驟。

『若將編制測驗這個問題詳細分析起來，有幾種根本條件，必須顧及。一個好的測驗，第一須能分別技能的等級；第二在不同的地方及性質不同的被試者，亦須有一致的結果；第三只須費少量的時間及精力。

每種職業能力差別的等級甚多。然為便利計，可分為幾大部分。普通每用新手 (Novice)，學徒 (Apprentice)，工人 (Journeyman)，及專家 (Journeyman Expert) 四種。所謂新手者，絕無職業的技能。所謂學徒者，已有職業一部分的知識；但尙未能做重要的工作。至於工人，則已完全能做職業中的工作。他若專家，則不特能做所有的工作；且能有極精巧靈敏的手腕。

有時職業測驗，須能分別上面每種中各人的技能。至於分別各種間

的能力，則尤為測驗重要的條件。普通對於這樣分類，有三種職業測驗：就是口述測驗，圖畫測驗，及動作測驗。

口述的測驗，是最普通；因為在短時間內，可以測量多數被試者。且極經濟，無特別設備的必要。決定職業能力之有無，甚為適宜。有時且甚真確，可無他種測驗的幫助。

編制一種測驗，須有十二個步驟：1、動機 (Priority)，2、指定工作 (Assignment)，3、調查 (Inquiry)，4、搜集材料 (Collection)，5、編選問題 (Compilation)，6、初步編制 (Preliminary sampling)，7、修正 (Revision) 8、決定形式 (Formulation)，9、最後編制 (Final sampling)，10、估量價值 (Evaluation)，11、規定級次 (Calibration)，及12、印行 (Editing)。

上面十二個步驟可以拼合為六個大段討論：

搜集關於職業的報告（包括最初四個步驟）。——例如軍營裏交來軍隊中所應用各種職業的名單一紙，要求各種職業須有相當的測驗（第一步驟）。然後根據這張名單，指定長官為相當的調查（第二步驟）。調查每種職業的情形，有三個應注意的地方（第三步驟）：

1、決定該種職業有測量之必要否？

2、決定該種職業所包含必須測量及可測量的分子：就是從事職業的人，是否可以依據技能的程度，而排列他們的等級？因為有些職業，只須用單純的動作，並沒有甚麼等級可以排列的。

3、決定應用何種測驗。有些職業，如打字等，是一種技能；用動作測驗比較用口述測驗好。有些職業，如管理機器等，則是一種知識；用圖



畫測驗及口述測驗，比較更為適宜。

經過這樣調查，我們就可知道這種職業是否應當測量，可以測量。既決定後，則可進而搜集各種材料(第四步驟)。其方法如下：1、職業專家的意見，2、職業公會中職員的意見，3、關於職業的文字，4、職業學校辦事人員的意見，5、僱傭者的意見，及其餘等等。用這種方法調查，就可知道這種職業含有什麼要素；如何可使從事這種職業者有良好的成績。

編選問題(包括自第五至第八四個步驟)。——材料收集後，須編選許多問題(第五步驟)。普通大約須有四十個或五十個。答應每個問題，須有切實的職業知識。根據經驗所得，好的問題宜具下面五種條件：

- 1、必須用職業中應用的文字。
- 2、必須是一個完全的單位，毋須另外的說明。
- 3、必須免除可碰機會的問題。如是，可無猜測答案的危險。
- 4、問題宜短，且只需要短的答案者才好。
- 5、意義必須明瞭，沒有誤解的危險。

本來所有大宗的問題，可用這五條定例審定之。有不合者刪除之；然後將其所餘編成一個初步的測驗(第六步驟)，測量許多工人。依據他們的答案，而定每個問題的價值，及分別其難易。所有的被試者，當任意招集，以免這種方法只適合一部分人或一個地方的習慣。測量時，至少須有兩個主試。如是，修正時可以互相貢獻意見。

初步測驗可供下面幾種校對的用處(第七步驟)：

- 1、這測驗是否適合職業的情形？
- 2、這測驗是否代表職業良好的練習？
- 3、何部分應當修改，應當補充，或應當刪除？
- 4、這個測驗是否代表該職業自新手至專家全部限域的程度？
- 5、這個測驗是否代表該職業全部各種方法？

依據諸問題的答案，修正測驗之不妥當處；然後再定測驗的形式（第八步驟）。

最後排列（第九步驟）。——用上面所排列的測驗，測量每組（新手，學徒，工人，專家）的代表人物二十人。對於新手，須測量智力較高，常識較富，但無職業能力的成人。同時亦宜測量另外已有職業知識者多人。測量時的情形，當如同在兵營中一樣。

然後將測驗的成績送交統計部。統計專家將必仔細研究這種成績及每個問題的答案。如是，可以決定每個問題的價值，選擇相當材料，使有平均的勢力。

估量測驗的價值（第十步驟）。——若是一個好的測驗，則一個專家受測驗時，當能答對問題的全部或幾乎全部；一個工人當能答對問題的大部分；學徒只能答對問題的小部分；一個新手當不能答對任何問題。然而這並不是說每個問題所有專家都能答對；多數工人能答對；少數學徒能答對；沒有新手能答對。其實，普通很少問題有這種傾向。

可是有些問題可以顯然分別新手學徒等的界限。新手失敗，而學徒，工人及專家能答對。有些問題，學徒失敗，而工人及專家能答對。有些問題，則僅專家能答對。諸如此類的問題，都有相當的價值，可使

成爲良好的測驗。主要的條件，是要這種情形若用曲線表示，他的傾向必須向上才好。若是一個問題，工人答對者多於專家，學徒答對者多於工人，新手答對者多於學徒，則必不適宜，應當刪去。另外問題，均應用此方法選擇，使有相當的勢力。

規定級次(第十一步驟)。——除出上面幾種步驟以外，尚有一個步驟，須詳細說明，就是規定測驗的級次。若是每個問題得四分，則當有多少分才是表示專家的程度，多少分才是表示工人的程度等，是又不可不研究的。決定這種標準的方法，就是計算專家得到的平均分數及工人得到的平均分數等等。普通專家的分數當必多於工人，工人的分數當必多於學徒，學徒的分數當必多於新手。然有時有幾個工人的成績與下級專家的成績一樣高，有時有幾個學徒的成績與下級工人的成績一樣高。兩組之間，往往有交叉的情形。規定級次的目的，是要劃定每兩組間的界限。如是，則交叉的情形就可減少了。

當這樣的界限規定後，就可預備印行(第十二步驟)，分發各兵營，實行測量矣。』

假定我們用算學問題，測量一組學生；同時不使學生疑惑，預備一種遊戲的動境，學生在此生活中，可以遇到一樣計算的問題。若是這種測驗及遊戲動境下計算的成績各是這種能力可靠的表示，則當測驗與這觀察所得成績的相關度只有平常的高度時，是否我們可以斷定這個測驗並非測量真正的能力？倘當測量時，非常謹慎，則這相關度是可告訴我們：這個測驗不過大概並非確實測量遊戲生活中的計算能力。我們批評這個測驗以前，須知遊戲動境下的計算成績，是否

與實際動境下的計算成績有限高的相關度？普通不同動境的相關度，並不比測驗與動境的相關度高。其實，我們所討論的真能力，事實上是罕有的。許多能力彼此總有多少的不同；若是要一個測驗與每個動境均有很高的相關度，這是不可能的事；因為每個動境各有他的特性。

用實驗的方法，除去每種測驗及每種動境的特性，如是能得到完全的相關度，亦未可知。這樣把能力分析，頗具理論的興味；就診斷方面而論，對於補救教學(Remedial instruction)，亦有很大的貢獻。但只為預測將來生活能成就與否等，則一種能力必須在他的特別情形下，才可起一種相當的作用；故我們實不能從事於這種詳細的分析。測驗的目的，在選擇將來可以成功的人才。一個測驗所得的成績，即沒有與實際生活上能力完全相關，亦不能說這個測驗沒有價值。可是當相關度很低時，或低於同類測驗與實際能力的相關度時，則這測驗就沒有存在的可能了。

## 二、提高智力測驗的確度的方法

智力的準則。——如何我們可以決定一個測驗是真正測量智力？前面我們已經講過，只有求相關度的一法。一個測驗所得的結果與用智力的事項有高的相關度時，這就是智力測驗優良的證據；這亦是最後選擇測驗唯一的方法。若是這種情形已滿足，則可無庸他求了。換一句話說，估量測驗的價值，必須依照智力的程度，確實排列學生的等級，然後與測驗的成績比較，而求其相關度。

這種智力的排列，必須有能幹精密的人，觀察學生具體的行為，除

去無關於智力的事項，而後估量智力的程度。這種觀察可以延長至一生；如是，可用一生的事實來證明。用這種方法排列後，則可用『試行錯誤法』編制一個測驗，測量智力是否與觀察所得的結果有相當的相關度？若用這種方法找到一個測驗的確度，則另外測驗亦可照樣編成，而與這個測驗比較，求其相關度。

若是這種觀察的判斷，是智力最後的一個準則，則又何以不完全應用之？我們用測驗測量，不僅為比較的更加經濟，其實亦因為比較更加客觀，可以預測將來。主觀的評論，若不十分小心謹慎，往往是非顛倒，而有偏見。父母的批評，尤有這種弊病。因為這個緣故，所以應用鐵面無私而且比較真確的智力測驗，測量智力，一天普遍一天。另外一個智力測驗的功用，則為預測將來；因為用智力測驗測量智力，不必待有事實，而後始可知其程度。

編制測驗，是否必須依照『試行錯誤法』進行？若我們用心理的方法分析智慧，是否可以給我們有些編制測驗的指導；省去許多時間及手續？這固然是可以的；但仍須用實驗的方法，找到其確度。茲根據神經系統方面的作用，說明分析智慧的內含。

智力的分析。——1、良好感應結（神經聯絡）的數目。——心理學上最重要的智愚標準之一，是精密分析的能力（Piece-meal activity or minute analysis）。這就是研究一種動境中細小分子的能力。對同一的動境，若感應結的數目愈多，則智力愈高。

2、感應結組織的方法。——對於新穎的思想或精巧的工作，欲有相當的效率，必須具有靈敏的精神作用。精神作用能靈敏與否，每視感

應結組織適宜與否為轉移。兩個兒童的感應結，數目雖或相等，但他們的智愚可以絕對不同。這就是因為感應結的組織不同之故。

3、造成或消滅感應結的易度。——普通心理所謂可型性(Plasticity)者，就是造成及消滅感應結的意思。可型性的高低與智愚成正比例。聰明的人，不特造成感應結很快，即去消滅感應結亦很快。愚笨的人，不特造成感應結很慢，即去消滅感應結亦很慢。

4、保持良好感應結的能力。——這層亦與智力有很高的相關度。無論何種學習，若今日學之，明日忘之，則有何用？普通所謂學得快，忘得易，實無心理的根據的。

編制智力測驗的方法。——1、一個智力測驗應當是一種學習測驗。他的方法，測量以前的多，測量將來的少。——上面所討論的四種智力特性，凡是智力測驗都應測量。有兩個方法可以應用：(a)給被試者一個學習的動境，自易至難，自簡單至複雜，測量(1)所學的分量，(2)所學的程度，(3)學習的速率，及(4)保持所學的能力。可是這種測驗，執行時非常困難；普通所應用者亦少。(b)測量根據天然發展所得到的樣本能力。這種測驗是測量以前所學的東西，並非測量以後的。

一個測驗可以測量上面四種智力的特性。例如兩個兒童同一年紀，若是一個兒童的感應結造成或去除特別快，保持能力特別強，則必有較多的感應結。所以比較感應結的多少及其組織的情形，就可知道他們的智力了。

2、一個智力測驗應測量許多特性。——智慧雖是一個混合的東西，每種特性不能獨自分離起作用；但同時亦有他的相當地位。所以愈多

的特性被測量，診斷可以愈安全。否則，學生或因特別緣故，必將發生不公平的結果。

3、一個智力測驗應測量各種不同而可以比較的特性的樣本能力。——理想的方法，是要測量所有的特性；然後就其與智力的關係，而定其價值。如此所得的總數，始可代表智力的全部。可是這種工作，非常困難！所以我們只得注意比較可以分別的特性。如視力聽力等，無論人獸，都沒有什麼差異，我們不能用之以別智愚。我們所應注意的，是可以辨別智愚的特性：如分析複雜的動境；在同一時內，能注意多數的分子；精神作用善於遷移；及能處理抽象事項，識別相互的關係等能力，均當測量。

但各種特性對於智力的關係，並不彼此相同。我們當就其關係之多少，而分別其價值。不特各種特性應當如此；即同一特性，若其分量不同，則其影響於智力，亦必有異。有些特性，若未達到一定的分量，或已超過一定的分量時，對於智力的高低，就不能發生什麼影響。

給各特性以相當的價值時，測驗者必須很小心，只給一種特性一次的價值。例如我們測量被試者的文字及算學能力，而給各種能力以相當的價值。今若算學測驗問題的文字很困難，意義很複雜，則名雖測量算學能力，實則測量文字能力。如是，就可說給文字以兩次的價值了。這種情形，非常普通！救濟的方法，當用統計法計算每個測驗與其他測驗的相關度。若兩個測驗間的相關度很高，則他們名目雖異，性質必同。這是不可不注意的。因為我們的目的要所測量的特性與智力有高的相關度，並非要各測驗自身有很高的相關度。

給各種特性以價值時，尚須顧及該特性的勢力是受他種特性存在與否的影響如何而定。例如對於銀行家，智力是非常要緊；但不誠實則無用。換言之，智力的價值，對於銀行家因誠實而更顯著。他若此類，不勝枚舉。

4、一個智力測驗應測量兒童有均等機會發展的特性。——這就是說測驗所用的材料及方法，當適合一般兒童的環境，沒有偏重偏輕的弊病。無論已經受過學校教育或未曾受過學校教育，無論在鄉間或城內，無論男或女，無論貧或富，均當一樣。因問題稍有偏向，結果就不能正確。

5、一個智力測驗應測量兒童有一樣動機發展的特性。——選擇的方法，最好觀察兒童於本能方面富有興味的各種活動；然後收集測驗的材料。

6、一個智力測驗的正確答案的百分比，須隨兒童的年齡增加。——普通智力成熟，大約在十八歲左右。平常所用智力測驗的根本假定，以為兒童所得到的智識，技能，及能力的總和，(a)是他現在智力的測量；(b)是與他所受遺傳的智力成正比例；(c)可供預知他的將來智力的程度。智力成熟以前，若一個兒童進步快於其他同年的兒童，這就是他的智力高於其他兒童的表徵。

7、一個智力測驗應測量遷移學習的能力。——這就是應用所學於其他動境的意思。這種能力與智慧有密切的關係。智力測驗若不顧及此層，即不能算作完全的測驗。智力測驗不測量學習的成績，亦即因此之故。



普通測量智力的方法。——茲摘譯推孟所著一段文字說明之。

1、實試法。——囑被試者試做各種難度不同的工作，然後就其成績比較其智力。這種方法亦有相當的確度；但須有充分的時間試做及考驗。

2、骨相學。——提倡這種假科學的方法測量智慧的人，就是高爾（Gall）。依據他的意見，以為各種特性，如智慧，喜鬥心，同情心等，觀察腦殼各部凹凸之不同，就可知道。這種學說之不能存立，乃是顯然的事實。其理由如下：

(a)各種特性並非在腦中有一定的位置。

(b)腦蓋的外形，並非腦系發達真確的表示。

3、直觀法。——直接觀察人的行為，言語，及容貌，為評斷智力的根據。這種方法甚為簡便，有時亦有相當的效用。可是錯時多，對時少；往往不能有相當的確度。其原因有下面數種：

(a)判斷沒有一定的標準。

(b)易受別種情形的影響。

(c)易被外觀如容貌等所欺。

(d)言語不能為智慧的表示。

(e)因朋仇的關係，判斷不易公平。

4、智力測驗。——精神方面有各種不同的作用。每種作用需要相當的智力。先知這些作用的品質；然後可知精神的全部，而定智力的高低。

5、彙集各種精神作用的樣本能力，為測量的標準。——如記憶能

力,理解語言能力,生字的數目,識別時間及空間的能力,聽從指導的能力,常識,判斷能力,辨別常物同異的能力,計算的推理,創造的能力,發見悖理的能力,聯念的豐富及敏捷,組織的能力,及抽象推理的能力等是。有這些作用成績的平均數,就可以表示個人普通智力的程度。

## 第 八 章

### 測驗材料的組織及測驗說明的預備

#### 一對於測驗排列有影響的事項

測驗的形式。——編製測驗,第一要緊,須注意記分(記錄點數)的經濟,真確,及客觀的情形。有兩種方法可以解決這個問題。

1、適宜的編製方法。

2、適宜的記分方法。

學生的反應(Response)須有簡易,短略,容易管理(Controlled)之可能。如是,計算成績時,必甚便利。其實,任何繁雜的精神作用,亦可用一個字,一個字母,一個記號,或一個數目表出之。用這種方法作測驗反應的手段時,記分就可很快,且易真確。各人不同的意見,亦可減少。

測驗中每個問題須只有一種反應是對的。如是,可做一個答案校對表,計算必甚方便。否則,評斷必費時間;且不容易真確;主觀的影響,亦容易潛入。

再者,反應須有一定的位置。如用方形,或圓圈,或其他方法表示位

置之所在等是。如是，校對時，亦可便利許多。

茲述桑戴克及麥柯爾編制航空測驗中一個問題逐漸改進的情形爲例來說明。

1、Compare the lines as they were before with what they are now.

這個問題須用很長說明的答案；校對時，很不便利。試看下面改良的方法。

1、Are the lines shorter than they were before, longer than they were before, or the same as they were before?

這個問題可以用較短的答案；且應用何種答案的形式，亦有相當的暗示。但答案尚沒有一定的位置。試看下面一個方法，就覺得更進步了。

1、Are the lines (1) shorter than they were before, (2) longer than they were before, or (3) the same as they were before?.....(2)

假定(2)是這個問題的真確答案，則用這種格式，比較好得多；因爲答案是很簡單，很容易管理，而且有一定的位置。若有成績校對表互相校對，則知道正誤，甚爲便利。

但尚有一個缺點，因爲這個問題正的答案是“longer”。答“shorter”的人，比較答“same”的人，錯誤必還要大些。故前者須受罰兩點，後者只須受罰一點。若用下法排列，則給與相當分數，尤爲便利。

1、Are the lines (1) shorter than they were before, (2) the same as they were before, or (3) longer than they were before?.....(3)

這個問題，正的答案是“longer”，既如上述；故宜寫(3)。若答案的

括弧內寫 2，則當受罰一點；3 與 2 的不同是 1。若答案的括弧內寫 1，則當受罰兩點；3 與 1 的不同是 2。故用這樣排列，正的答案與括弧內的數目的相差，就很容易計算了。

不特測驗應用這種方法排列：就是普通考試亦可利用這種方法增加成績計算的確度，及除去主觀的影響。而且時間方面，亦可比較經濟。茲再引數例說明之：

1、桑戴克的敏捷測驗(Mental Alertness Test)的一段：

Make a cross in the square before the best answer  
to each question.

(1) Why are prunes a good food? Because they

|  |
|--|
|  |
|  |
|  |
|  |

grow in California.

are wholesome and economical.

are served in boarding houses.

make an attractive dish.

2、柏來雪(Pressey)的精神作用測驗(Mental Survey Test)的一段：

girl—woman: boy—man

Examples: Sun—day: moon—.....

good—bad: big—.....

(1) woman—girl: man—..... (11) hill—valley: high:—.....

(2) kitten—cat: punny—..... (12) arm—elbow: leg—.....

(3) sky—blue: grass—..... (13) truth—falsehood: straight line—.....

3、葛林(Green)的組織測驗(Organization Test)的一段：

Write numbers in these spaces

|                                        |  |
|----------------------------------------|--|
| (1) I am, see, how tall.....           |  |
| (2) of the cold, afraid, they were.... |  |
| (3) a dog, a boy, had.....             |  |

機械的記錄成績點數的方法。——有這種方法，記錄點數，必更覺  
 方便。用這種方法之可能與否，全視測驗排列的方法如何而定。故有  
 討論的價值。其方法之普通者，有如下數種：

1、若被試者的答案有一定的位置，如在測驗的右邊，或左邊，則最  
 簡便的方法，可用一張試驗紙，將其答案填好；然後照此答案的樣本  
 放在被試者的試驗紙上，與他的答案雙雙平行而核對之。

這種方法亦可應用於被試者的答案在每行的中間者。其法把一張  
 測驗紙填好答案，將其旁邊部分割去之。如是，將此試驗紙放在被試  
 者的試驗紙上，可於空隙間見被試者的答案，然後核對之。最好用這  
 原樣的試驗紙，因為找到答案相當的地方特別容易。如要常時應用，  
 可用比較堅牢的紙板。

2、有時被試者的答案並不十分明瞭。可是真確與否，對於成績很有  
 關係。例如學生須在相當字下畫一畫，或在相當理由下做一個記號，  
 以及劃去字母等，用上面那個方法核對，雖亦便利；但用透明紙為做  
 核對樣本的材料，尤為適宜。其法先將測驗紙填好答案，然後用透明  
 紙放在上面，做相當的記號。如是，可用此紙放在被試者的測驗紙上，  
 核對他的答案。如智力測驗及算學測驗等均可應用之。

3、尚有一個方法，記錄測驗成績的點數時，只可見其正的答案，不

能見其其餘部分。其法用一張白紙，鑿刻小孔，只使正的答案顯出。例如此類，用透明紙黑塗其餘部分。而僅留正的答案所在地方，亦無不可。如是，校對成績，亦甚方便。

用上面這個方法，且可繼續分別記錄點數。例如胡德麥柯爾的算學基本測驗 (Woody-McCall Fundamentals of Arithmetic Test)，加減乘除諸算題均混雜一起。校對成績時，最好依據算法的不同，前後分別校對。故可用四張透明紙，分別填好各法的答案，繼續分別校對之。

團體測驗與個獨測驗。一、個獨測驗，就是一次只單獨測量一個學生。有些測驗及對於不識字的兒童，必須應用這種方法。至於另外有些測驗及對於年紀較大而且識字者，則可用團體測驗；就是一次測量多人。

兩種方法各有利弊；但善於編製測驗者，往往欲使團體測驗亦宜於測量不識字的人。

可是有時雖可應用團體測驗，但較之個獨測驗結果，必將稍次。若為診斷目的，則用個獨測驗，尤為適宜。其理由如下：

1、測量時，學生的外表及其偶然的反應，均可為知道他的缺點之參攷。用單獨測驗，容易得到這種幫助。

2、單獨測量時，可就被試反應的情形，找到其他適宜的個獨測驗，因為一種測驗的成績，每可為選擇他種測驗的參考。個獨測驗特別富有這種好處。

3、用個獨測驗，測量的情形容易合於標準。

但用團體測驗，時間方面，可以更加經濟。例如用個獨測驗，若每人

須三十分鐘，則先後測量五百人，至少須費三百點鐘。他若用團體測驗，則至多不過費四十分鐘够了。即行分組測量，每組五十人，亦可省下時間不少。

團體測驗有時且可對於被試者更加公平。1、有些實驗，須測量學生在相當時期內所得的進步。今用個獨測驗，若主試須費一星期工夫完其工作，則最後一個被試者或可多得一星期學習的利益，或將多受一星期遺忘的害處。2、有時用個獨測驗，後被試者可得先被試者的利益；因為先被試的學生可以宣布測驗的性質於他人。3、有些測驗，主試的情形，前後很難一致。有些被試者可以特別便宜，有些則否。故用個獨測驗，前後測量，頗難公平。4、團體測驗，時間天氣，均可一致。

團體測驗及個獨測驗各有其功用，已如上述。以軍隊的事實說明，最初用團體測驗。如找出不識字或精神異常的人，則再用團體或個獨測驗診斷之。學校內情形亦可如此。普通可以應用的通則是：

『開始用團體測驗，為大多數的診斷。然後對於少數學生，如照團體測驗成績所示，須特別研究者，則再用團體測驗或個獨測驗。』

## 二、做法說明的主要原理

1、做法說明須簡單明瞭；同時須使學生有充分的了解。——冗長的說明不特費時，而且使心思混亂，不易明瞭。不特對於兒童如此，即對於成人亦有這種弊病。因為冗長的說明需要記憶力甚大，要緊的地方容易忘却。倘有一點錯誤，全部成績均受影響。這種冗長的弊病，往往係由於與測驗不相關切的說話所致。我們所當注意者，說明的目的並非給兒童一種教學。對於測驗的做法，若已有充分的指示，則已够足

了；原不必從事於瑣碎的說明。

再者，冗長的說明易使沒有經驗的主試混亂。因為新從事測驗者若非記得說明所示各種步驟及步驟的次序，往往不能自信應用裕如。若說明太冗長，則所感的困難必多。且他還須維持試場中的秩序，注意時間的經過，顧及試紙的支配，及被試者了解說明的情形。凡此種種，無一不足增加其困難。

但不充分的說明。其為害且甚於冗長的說明。有時且可使測驗的目的完全失敗。欲求說明的適宜，須用實驗方法解決之。緊要地方不宜有一絲遺漏。

2、做法說明可先舉例表示，或用練習測驗表示。——用這種方法，比較用字說明更加經濟，且易明瞭。其理由有三：(a)容易記憶，(b)容易注意，(c)容易知道錯誤。

這種方法之普通者：(a)主試在黑板上表示測驗實際的工作；(b)有些做法表示，係利用測驗的例子，與測驗印在一起。試看下面這個樣式：

『這是一個常識測驗。下面共有十六個問題；每個問題有三個答案。你們仔細看！那個答案是最好？就在他前面方格內做個十字×的記號。如下例：

|    |   |                                              |
|----|---|----------------------------------------------|
| 舉例 | { | 為什麼我們要用火爐？因為                                 |
|    |   | <input type="checkbox"/> 他好看。                |
|    |   | <input checked="" type="checkbox"/> 他可使我們溫暖。 |
|    |   | <input type="checkbox"/> 他是黑的。               |



上面這個例，第二個答案是最好，應在他前面方格內做個十字×的記號。現在你們開始做第一個題。開始後，繼續做去，等到叫「停」為止。』……………

若說明很複雜時，必須有這種舉例說明才好。兒童的個性各不相同，故應適測驗情形的快慢，亦必不能一樣。如舉例說明，則可救濟這種弊病。

說明種類甚多，列舉如下：

- (a) 主試在黑板上舉例說明。
- (b) 叫兒童先做一種容易的模型測驗。
- (c) 印練習測驗及說明在正式測驗的後面，或在前面第一頁。
- (d) 說明及舉例等與正式測驗印在同面。
- (e) 若時間並非緊要的問題，則正式測驗的前面數題可為舉例用。

主試巡行室內，如被試有未知者，可幫助之。可是被幫助的部分，當然不能算為正式的工作。

他若不計費用，則僅將年月日等空白以及說明和練習測驗印在前一面，較為適宜。

- (f) 有些測驗不宜舉例，因為舉例就是教學生作實際的工作。如皮內(Binet)及西門(Simon)所編的智力測驗等，就有這種情形。

3、做法說明必須適宜於被試者的全體，而且宜一致。——測量學校的成績或能力，說明必須簡易明瞭，人人能懂；要沒有一個學生因為不懂得說明，而不能做測驗的工作才好。欲達到這個目的，說明不應比測驗內最容易的材料難。若說明比測驗內最容易的材料難，則學

生雖或可以得到多少分數，有時竟或因此而得到零數。在如此情形下則學生並不受實際的測驗，實受說明的測驗而已。故說明有求適宜於被試者全體的必要。

至於宜求一致這一層，可分兩方面討論。

(a)機械方面的事項，如寫名字，性別，年齡，年級·月日等，與測驗本部無甚關係，故無求一致的必要。

(b)非機械方面的事項，對於測驗的成績，必有多少的關係；故必須一致。對於各年級，當完全相同，鄉村學校的同班學生，年齡大小不同，這層尤為緊要。若說明相同，主試在同一時內，可以測量許多學生。所以一種測驗的說明，不特須求適宜於全體，且宜力求一致。

智力測驗有時以懂得說明的能力，為實際成績的一部分。故所用的說明，尤須絕對一致。

4、說明的次序當於動作的次序相符合。——這層有三個理由：(a)容易記憶，(b)容易應用，(c)容易模倣舉例的說明。

5、說明須以動作為單位而分段落，一步一步做。——這層亦有三個理由：(a)容易記憶；(b)動作時，注意力容易集中；(c)免除種種與測驗不相關切的錯誤。

6、說明對於被試者須有相同濃厚的興味。——其理由如下：(a)興味為努力的主要條件。興味不相同，努力亦不能一致，成績亦必不能公平。興味最濃時，個人的工作易於一致；可為實際能力的表示。故有用相當的方法提高興味之必要。

若與測驗的目的無妨碍，主試者須說明測驗的目的及大約計算成績的方法。若被試者知道測驗的成績是根據於工作的速率及品質，則其興味必可提高。但詳細的說明，亦不適宜。其理由有三：(a) 因為簡略省時；(b) 因為有時被試者不應知道實在記分的方法；(c) 有時欲免除被試者欺騙的行為，且有嚴守秘密之必要。

另外一個提高興味的方法，是利用被試者的競爭心。學校裏差不多沒有另外的刺激比較競爭心更有勢力，更為自然，更為健全。

他若用設計方法，測量作文寫字等，亦為提高興味之一法。

7、對被試者有說明外，對主試者亦應有相當的說明。——對於主試者的說明，須易了解及實行。有兩種方法，可以應用：

(a) 主試者對被試者應說的話，須有明確的規定。

(b) 對於主試者的說明，可插入對於被試者的說明中間；但須用另外方法區別之。

## 第九章

### 等分測驗

為什麼要等分測驗？——測驗的根本目的，是要發見學生真正的差異。欲發見真正的差異，一個測驗不但應有充分的確度；且應具有下面各種性質：

1、每個學生須得到零以上的分數。若每個學生都得到零分，則我們就不能知道誰好誰次。即若只有一個學生得到零分，我們亦不能知道他究竟比他人要次多少。換一句話說，得到零分的學生，我們就不能

確知他的程度。

每班學生程度的限域非常大，要救濟上面兩種弊病，須使低能的學生亦得到相當的分數。欲達到這個目的，測驗之一部分材料，必須在他們的能力範圍以內；同時須有充分的時間，使能做完其測驗。

2、沒有一個學生應得完全的分數。完全的分數不能真確測量學生的程度。其理由與得到零分的人一樣。

3、不應有不能分配的分數。若一個測驗的記分方法只有成 (passed) 及敗 (failed) 兩種，則容易發見這種弊病；因為同是成的人，或同是敗的人，就沒有等級可分了。

記分究應如何精密？則當視成績的用途如何而定。就普通情形而論，一個測驗至少須能分學生的能力為七等；若欲求相關度，則至少須能分為十三等。這是最低的限度，其實，等數愈多則愈精密。如比七等尚少，則成績的用途就鮮；如比十三等尚少，則相關度就不能十分精確。故一個測驗的一部分材料必須很容易；對於被試者全體，當沒有什麼困難。然後逐漸增加難度，至最優秀的人亦覺有困難為止。可是一個測驗的每部分材料，若有精密記分的可能，則即測驗的材料甚少，其所得的結果，亦可與材料甚富的測驗之結果相彷彿。例如施東的算學推理測驗 (Stone's Reasoning Test in Arithmetic)，雖只有十二個問題，但可分學生有十二多種的程度；因為每個問題的各部分均有相當的記分，可以分別學生的程度。

我們若要知道個人確實的成績，則上段所討論的條件，必須顧及。但若只要知道一班的成績，而且用中數或超過測驗一部分的人數百

分比爲標準時，則卽有不能分別真確程度的零數或完全分數，亦無不可。可是用平均數(mean, 求法詳附錄一) 爲標準時，則這種極端的分數，就可影響於全班的平均分數了。

4、若量表(Scale)上兩點的距離彼此相等，換一句話說，若量表有相等難度的級次，逐漸上進，則此量表必比較更爲適用。因惟如此，學生不同的程度始能精確比較。

5、一個測驗應有相當的統計結果。若徒用文字說明，絕不相宜。

6、除非每種測驗所得到的兩種成績可化爲一個分母計算，則實際不同的程度必不能找到。例如葛第斯的加法測驗 (Courtis Addition Test)，若學生甲的速率成績是十題，真確成績是百分之九十，學生乙的速率成績是十二題，真確成績是百分之七十五，則誰的成績好？倘使沒有一個表可使速率成績固定，比較真確的成績；或使真確成績固定，比較速率的成績，我們就無另外良好的方法可以比較了。這種表的意思，就是應當告訴我們：假使學生乙的速率成績是做十題，非做十二題，則他的真確成績當幾何？葛第斯曾創作一個公式，將他算學測驗的速率分數及真確分數化爲一種分數計算。(詳法見附錄三，原文見 1920 年美國哥倫比亞大學的師範院季刊(Courtis and Thorndike: "Correction Formulae for Addition Tests"; Teachers College Record, January, 1920.)。

等分測驗的方法頗多：第一種叫做目的表 (Goal scale)。例如葛第斯的輔導測驗 (Courtis Supervisory Tests)，一個學生若能拚對一個特別拼字測驗中二十個字至二十五個字，則可假定他已達到相當的拼

字程度，能拼對普通一千個字。他若一個學生只能拼對十七個字至二十個字，則可假定他只能拼對五百個字。餘以類推。第二種叫做重見次數表(Frequency-of-occurrence scale)。例如鍾斯的生字測驗(Jones' Vocabulary Test)，學生知道一個字，他應得的分數，以這個字重見於十本普通教科書的次數為標準。同樣有凱得一羅索奴富的自由聯想測驗(Kent-Rosanoff Free Association Test)，測量個人心緒異常的情形，以個人的反應稀罕與否為根據。其他尚有百分表，年齡表，年級表，成績比較表及 T 量表五種。這五種量表比較更為普通，應用亦比較廣大；故將詳細討論。

#### 一、百分表(Percentile Scale)——以百分比為單位(Percentile Unit)。

編制百分表的方法。——表十六第一行表示一個測驗問題的數目，第二行表示十一歲兒童答對每格問題的人數，第三行就是答對某格問題的人數百分比。這第一行與第三行就是這個測驗對於十一歲兒童的百分表。用同樣方法，任何年齡，任何學科，及所有年齡一起的任何百分表，亦可造成。造百分表的方法，同計算四分點(Quartile)的方法一樣。 $Q_1$ 就是百分之二十五； $Q_2$ 就是百分之五十，或稱中數； $Q_3$ 就是百分之七十五。其餘的百分比，亦是用一樣方法計算。若成績依優劣的次序排列，從次的逐漸至好的，則百分之二十五，就是總人數四分之一那個人所得的分數；百分之五十，就是總人數中間那個人所得的分數；百分之七十五，就是總人數四分之三那個人所得的分數。其餘百分之十，百分之二十等，均可用同樣方法計算。最低的分數是零，最高的分數是百。

表 十 六

編造十一歲兒童受某種測驗的成績百分表的說明表

| 答對問題的數目 | 學生的數目 | 百分比 |
|---------|-------|-----|
| 0       | 1     | 0   |
| 1       | 1     |     |
| 2       | 1     |     |
| 3       | 1     |     |
| 4       | 1     |     |
| 5       | 1     |     |
| 6       | 1     |     |
| 7       | 5     |     |
| 8       | 4     |     |
| 9       | 2     |     |
| 10      | 6     |     |
| 11      | 4     |     |
| 12      | 3     |     |
| 13      | 4     |     |
| 14      | 12    | 10  |
| 15      | 15    |     |
| 16      | 22    |     |
| 17      | 31    | 20  |
| 18      | 20    | 30  |
| 19      | 32    |     |
| 20      | 42    | 40  |
| 21      | 35    | 50  |
| 22      | 40    | 60  |
| 23      | 32    | 70  |
| 24      | 29    | 80  |
| 25      | 22    |     |
| 26      | 16    |     |
| 27      | 16    | 90  |
| 28      | 13    |     |
| 29      | 3     |     |
| 30      | 4     |     |
| 31      | 6     |     |
| 32      | 0     |     |
| 33      | 1     | 100 |

解釋百分表的方法。——表十七是說明用百分表的方法，及解釋一

個十一歲兒童所得分數的情形。測驗A是一個讀法測驗，他的百分比是根據表十六來的。測驗B是一個算學測驗；測驗C是一個拈法測驗。一個學生受A測驗，得到成績22，就等於百分比60。受B測驗，得到成績39，就等於百分比55；因為39在38與40之間，故其百分比在50與60之間，即是55。受C測驗，得到成績10，即等於百分比50；因為10的成績有三個百分比(40, 50及60)，故擇取其中間一個，即是50。

表十七

十一歲兒童受三種測驗用百分比方法計算其成績的說明表

| 測驗   | 百 分 表 |    |    |    |    |    |    |    |    |    | 學生成績 | 學生的百分比 |     |
|------|-------|----|----|----|----|----|----|----|----|----|------|--------|-----|
|      | 0     | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 |      |        | 100 |
| 測驗 A | 0     | 14 | 17 | 18 | 20 | 21 | 22 | 23 | 24 | 27 | 33   | 22     | 60  |
| 測驗 B | 1     | 25 | 30 | 34 | 36 | 38 | 40 | 43 | 46 | 50 | 62   | 39     | 55  |
| 測驗 C | 4     | 6  | 7  | 9  | 10 | 10 | 10 | 11 | 13 | 17 | 24   | 10     | 50  |
|      | 中數    |    |    |    |    |    |    |    |    |    |      | 55     |     |

為實用的目的，這個學生受這三種測驗，所得成績百分比的地位，可以中數為標準，就是55。這個百分比，是表示他的成績比較他同年的平均程度稍高。

若一個兒童只有十歲，則當用十歲的百分表計算他的成績。

就理論方面而言，用所有測驗百分比的中數，尚不能表出一個學生真正的程度。據賓德納 (Pintner) 的意思，尚須進一步再求比較精確的百分比。因為測量愈充分，變化率愈可減少；故有求高級百分表 (Super-percentile table) 的必要。所謂高級百分表者，就是許多中數百



百分比的百分表；乃可以真確表示百分比的價值。其編制的方法，如編制一個測驗的百分表一樣。許多同年學生的中數百分比代替表十六第一行答對問題的數目而已。

二、年齡表 (Age Scale)——以長進率為單位 (Growth Unit)。

編制年齡表的方法——編制年齡表，只要決定每歲相當的標準分數 (Age norms)。上面所討論的百分比單位，若以年齡為根據時，則只能告訴我們一個學生比較他同年齡的學生好多少或次多少；若以年級為根據時，則只能告訴我們一個學生比較他同年級的學生好多少或次多少。但我們不說一個學生所得的分數是在同年的程度第百分之二十五，五十，或七十五等，而要說這個學生的程度是等於九歲的平均分數，十歲的平均分數，或十一歲的平均分數等時，則當用何法呢。欲達到這個目的，我們須用長進的單位，計算每歲的標準分數。若已有每歲的標準分數，則任何學生的成績可以立刻化為教育年及教育商數 (教育測驗)，或智力年及智力商數 (智力測驗)。試看下表，

表 十 八

一個十歲兒童受三種教育測驗計算教育年及教育商數的說明表

| 測 驗    | 年 紀 |    |    |    |    |    | 學生的學生的 學生 的 |      |             |
|--------|-----|----|----|----|----|----|-------------|------|-------------|
|        | 8   | 9  | 10 | 11 | 12 | 13 | 分數          | 教育年  | 教育商數 (E.Q.) |
| A 平均分數 | 4   | 8  | 12 | 15 | 18 | 20 | 12          | 10   | 100         |
| B 平均分數 | 20  | 30 | 38 | 45 | 50 | 55 | 46          | 11.2 | 112         |
| C 平均分數 | 85  | 84 | 82 | 80 | 78 | 75 | 83          | 9.5  | 95          |
|        | 中 數 |    |    |    |    |    |             | 10   | 100         |

就可明白。

年齡分數的解釋。——上表測驗 A 的平均分數，八歲的兒童是四；九歲的兒童是八等等。這個十歲的兒童得到十二分，適合十歲兒童的標準分數；故得教育年十歲。今實年與教育年適相等，故他的 E. Q. 是 100  $[(10 \div 10) \times 100]$ 。

100E. Q. 就是表示這個兒童受 A 測驗，他的程度合於一般的標準。他受 B 測驗，得到 46 分，在 45 至 50 距離五分之一地方。45 是十一歲兒童的標準分數，故這個兒童受這個測驗的教育年是十一歲零十分之二。他的 E. Q. 是 112  $[(11.2 \div 10) \times 100]$ 。

測驗 C 用時間的單位計分。費時愈長，程度愈低。這個學生的成績是 83，適在 84 與 82 兩項成績中間。84 是九歲兒童的成績，82 是十歲兒童的成績，故這個兒童的教育年是九歲半。他的 E. Q. 是 95  $[(9.5 \div 10) \times 100]$ 。

他教育年的中數是 10，E. Q. 的中數是 100，故他是一個正常的兒童。倘使這個兒童是十歲半，則他受 A 測驗的 E. Q. 將為 95.2  $[(10 \div 10.5) \times 100]$ ；受 B 測驗的 E. Q. 將為 106.6  $[(11.2 \div 10.5) \times 100]$ ；受 C 測驗的 E. Q. 將為 90.5  $[(9.5 \div 10.5) \times 100]$ 。有時用月計算，比較便利。如以最後一個 E. Q. 為例，就可用下法計算。

$$9.5 \text{ 歲} = 114 \text{ 月} \quad 10.5 \text{ 歲} = 126 \text{ 月}$$

$$(114 \div 126) \times 100 = 90.5$$

其結果都是一樣。

有些年齡表毋須如表十八那樣變換數量。西特弗訂正皮內西門的

智力測驗(Stanford Revision of the Binet-Simon Intelligence Scale)就是一個例子。這個測驗共分許多部分；做對一部分，就得二個月的智力年。故他的成績開始就用月計算的。

三、年級表(Grade Scale)——以年級變化率為單位(Grade Variability Unit)。

年級表的產生。——我們普通用點為計時間的單位，用元為計財產的單位，用磅為計重量的單位，用尺為計長度的單位。至於測量教育的成績，則用『或然的錯誤』(Probable error, 就是P.E.)，均方差(Standard deviation, 就是S. D.)，或其他的變化率，為計分的單位。此地我們只能討論利用這種單位編制量表的方法。詳細情形，當看胡德所著的算學成績的測量(The Measurement of Some Achievements in Arithmetic 在 Bureau of Publication, Teachers College, New York, U. S. A. 出版)。現僅能就其編制之步驟而略述之。

假定主試者要為三年級編制一個加法量表 (Addition Scale)，則當取如下步驟：

1、用自己的判斷，選擇許多加法問題。難易各不相同；最易至最難的都有。

2、用這些問題測量許多三年級的學生。

3、計算學生做對每個問題的百分比。若有百分之九十學生做對一個題目，則這題目當然是很容易；在量表上的地位應當很低。若有百分之五十學生做對一個題目，則這題目的難度當然適中；在量表上的地位必在中間。他若只有百分之五學生做對一個題目，則這題目當然

是很難；在量表上的地位應當很高。所有問題在量表上的地位，都是依據這樣百分比之大小而定。百分比愈大，問題愈容易；在量表上的地位亦愈低。百分比愈小者，其情形適相反。

表 十 九

P.E. 值對數表

| P.E. | .00  | .05  | P.E. | .00  | .05  | P.E. | .00  | .05  | P.E. | .00     | .05    |
|------|------|------|------|------|------|------|------|------|------|---------|--------|
| 0    | 0000 | 0135 | 1.5  | 3441 | 3521 | 3.0  | 4785 | 4802 | 4.5  | 4988    | 4989   |
| .1   | 0269 | 0403 | 1.6  | 3597 | 3671 | 3.1  | 4817 | 4831 | 4.6  | 4990    | 4991   |
| .2   | 0536 | 0670 | 1.7  | 3742 | 3811 | 3.2  | 4845 | 4858 | 4.7  | 4992    | 4993   |
| .3   | 0802 | 0933 | 1.8  | 3896 | 3939 | 3.3  | 4870 | 4881 | 4.8  | 4994    | 4994.6 |
| .4   | 1063 | 1193 | 1.9  | 4000 | 4057 | 3.4  | 4891 | 4900 | 4.9  | 4995.2  | 4995.7 |
| .5   | 1321 | 1447 | 2.0  | 4113 | 4166 | 3.5  | 4909 | 4917 | 5.0  | 4996.2  | 4996.6 |
| .6   | 1571 | 1695 | 2.1  | 4217 | 4265 | 3.6  | 4924 | 4931 | 5.1  | 4997.1  | 4997.4 |
| .7   | 1816 | 1935 | 2.2  | 4311 | 4354 | 3.7  | 4937 | 4943 | 5.2  | 4997.7  | 4998.0 |
| .8   | 2053 | 2168 | 2.3  | 4396 | 4435 | 3.8  | 4948 | 4953 | 5.3  | 4998.2  | 4998.4 |
| .9   | 2291 | 2392 | 2.4  | 4472 | 4508 | 3.9  | 4957 | 4961 | 5.4  | 4998.6  | 4998.8 |
| 1.0  | 2500 | 2606 | 2.5  | 4541 | 4573 | 4.0  | 4965 | 4968 | 5.5  | 4999.0  | 4999.1 |
| 1.1  | 2709 | 2810 | 2.6  | 4602 | 4631 | 4.1  | 4971 | 4974 | 5.6  | 4999.2  | 4999.3 |
| 1.2  | 2908 | 3004 | 2.7  | 4657 | 4682 | 4.2  | 4977 | 4979 | 5.7  | 4999.4  | 4999.5 |
| 1.3  | 3097 | 3188 | 2.8  | 4705 | 4727 | 4.3  | 4981 | 4983 | 5.8  | 4999.55 | 4999.6 |
| 1.4  | 3275 | 3360 | 2.9  | 4748 | 4767 | 4.4  | 4985 | 4987 | 5.9  | 4999.65 | 4999.7 |

4、假定A,B,C,D,E,F,G,諸問題學生做對的百分比如下表,我們就可用上面那個P.E.值對數表,將百分比化爲離開該三年級成績中數上下的P.E.距了。

表 二 十

用P. E. 值對數表化學生做對問題的百分比  
為離開中數的P. E. 距的說明表

| 問 題        | A     | B     | C    | D  | E   | F    | G    |
|------------|-------|-------|------|----|-----|------|------|
| 做對的人數百分比   | 95    | 85    | 60   | 50 | 40  | 12.2 | 2.2  |
| 與 50 百分比相減 | -45   | -35   | -10  | 00 | 10  | 37.8 | 47.8 |
| P. E. 距    | -2.45 | -1.55 | -0.4 | 00 | 0.4 | 17.5 | 3.0  |

A與B兩題難度的不同為(95-80),就是相差百分之十。P. E.的不同為(2.45-1.55),就是.90P. E.。

C與D兩題難度的不同,就百分求而論,亦是百分之十;可是P. E.的不同只有.4,比AB兩題難度相差的一半還弱。

D與E兩題難度的不同,就百分比而論,亦為百分之十。其P. E.的不同與CD兩題一樣,亦只有.4。

F與G兩題難度的不同,就百分比而論,亦為百分之十;可是P. E.的不同為1.25(3.0-1.75)。

觀上面所述的情形,百分比不能表示難度的不同。若要真確表示難度的不同,必須用P. E. 值計算。CD兩題及DE兩題的百分比相等,P. E. 亦相等;這因D題為諸題難度的中數,C與E兩題離開中數距離相等的緣故。F與G兩題的不同最大,這是因為這兩題的百分比離開中數最遠之故。做對A題的百分比最大,在中數的下端最遠,且因上表是表示各題難度的不同,故做對人數的百分比愈大,問題必定是愈容易。

5、現在須計算三年級加法能力的零點 (Zero point) 之 P. E.。決定零點的方法，適與上面所討論的相反。現在可不再問算對某題的人數百分比多少，但僅問三年級學生不能算對一個問題的百分比幾何。假定有百分之四學生不能做對一題，則照表十九所示，當在中數下 2.6 P. E.；這就是零點。

6、計算每個問題在零點以上的 P. E.。如是，年級表就告成功。每個問題在零點以上的 P. E.，看下表就可知道。計算的方法，就是用零點的 P. E. (2.6 P. E.) 去減如表二十所示的每題的 P. E.。

表 二 十 一

| 問 題            | A   | B    | C   | D   | E   | F    | G   |
|----------------|-----|------|-----|-----|-----|------|-----|
| 在零點以上的 P. E. 值 | .15 | 1.05 | 2.2 | 2.6 | 3.0 | 4.35 | 5.6 |

7、有時須取消 P. E. 沒有相等階級的問題。例如胡德的算學量表，組 B (Woody's Arithmetic Scale<sup>8</sup>, Series B)，就是從組 A 裏面選出來；然後有差不多相等的階級。

8、若編制一個量表，不僅為一級，而為小學全校之用，則每一年級當重覆如上所述編制三年級年級表的第二，第三，及第四三個步驟。

9、自一級的中數至相連他級的中數，計算 P. E. 距。其法就是計算一級學生所得分數超過相連他級中數的人數百分比；然後由百分比校對 P. E. 值。這種 P. E. 每落在兩級中數之間。將這樣直接決定兩種及另外間接決定幾種 P. E. 平均，則得更加真確的 P. E. 距。

10、利用這種間隔 (Interval)，將每題離開他自己一級的中數的 P. E. 距化為離開公共零的參照點的 P. E. 距。

11、然後決定每題最後的全校 P. E. 值；並平均每題在各級的 P. E. 值，而決定他在量表上的位置。計算平均數時，有幾級的 P. E. 值比較更有勢力；因為近每級人數分配中間的 P. E. 值每比近兩端的 P. E. 值更為可靠。

12、若只單獨做一個小學的全體量表，則可省了許多手續。開始就計算各級全體學生做對每個問題的人數百分比。再將每題的百分比化為 P. E. 值；然後求得不能做對一題的人數百分比之 P. E. 值為零的參照點，以供規定各題在參照點以上的地位；量表遂成。

13、另外一個簡單方法，以六年級的量表代表小學全體的量表。因其居小學全體的中間（自三年級至八年級）；故可代表全體的程。

年級表上 P. E. 值的恒度 (Constancy)。——我們知道一根鐵的長短，可以歷數百年而不變。可是 P. E. 值則不然。例如胡德的加法量表第五題是 $3+1=?$ ；他的 P. E. 值是 3.26。這種 P. E. 值並非絕對不變的。普通有兩種原因，可以使此類問題有時變更其價值。

1、課程的改良及教法的進步，均有相當的影響。再以上面一題為例，對於一般兒童，其困難在於加法的符號。若另有一題 $7^2$ 加 $2^6$ ，第二數寫在第一數下面，則其難度可比 $3+1$ 還容易。 $3+1$ 這種簡單的題目，對於兒童反這樣困難，以後教學者將必注意及此，使兒童早學這種符號的意義。如是，自然將變更其 P. E. 值了。

2、分班的改良，亦有相當的影響。利用測驗，逐漸使分班可以更加真確；同時可以減少各年級學生能力交叉的情形，而增加各年級相差的距離。如是，程度分配情形，必有變化；P. E. 值自然亦將改變。

再者，年級制現已逐漸失掉他從前的機械形式。有些學校改八級爲七級；有些學校用分科制代替年級制；有些學校根據教育測驗及智力測驗的成績，兩級間添設另外班次，以滿足特別教學的要求。故將來編制測驗，用年級爲根據，不如用年齡爲根據。

四、成績比較表 (Product Scale) ——以成人作業的變化率爲單位 (Variability-of-Adult-Performance Unit)。

愛禮斯的書法量表。——這個量表以成人實試結果的變化率爲單位；與其餘如桑戴克的書法量表等，以成人的評斷爲標準者不同。愛氏根據清楚的程度 (Legibility)，規定量表；桑氏則以普通價值 (General merit) 爲根據。前者注重用度 (Function)；後者注重外表 (Appearance)。以清楚程度爲標準，有兩層理由：1、寫字的目的是供人覽讀。以易讀的程度 (Readability) 爲比較成績的標準，實甚適宜。2、清楚的程度，可用科學的方法測量。在一定時間內，讀相當的分量，比較真確的程度，就可知道書法清楚程度之如何。至於以普通價值爲根據，則不能用這種科學的方法測量。

編制這種量表，以成人誦讀寫字樣本平均時間的長短爲標準。其方法將成人誦讀每個樣本的時間平均，然後再依據時間的長短，將許多樣本前後排列，做個次數面 (Frequency surface)；並在次數面的底線上分爲十個相等的間隔，最低的速率爲0，最高的速率爲100，平均的速率爲50。凡書法樣本的速率適合20, 30, ……100的地位者，則取爲該地位的樣本。如是，量表遂成。

用這樣量表的方法，就是將一個學生的寫字成績放在量表旁邊，逐



漸向上移動，至找到品質甚相類似的樣本爲止。如是，這個學生的成績，就照這個樣本的分數規定。

**五、成績比較表——以評斷的變化率爲單位 (Variability-of-Judgment Unit)。**

如海利葛斯的作文量表，絲大齊 (Starch) 的書法量表，及桑戴克的圖畫量表等，均是這種量表的例子。其編制的方法，都是一樣。茲以作文量表爲例，分述其編制步驟如下：

- 1、收集成績各不相同的許多文章；最劣的與最好的都有。
- 2、請許多能幹的批評家依據成績優劣，排列這些文章的次序；並決定應得零分的一篇。
- 3、從第二步驟所規定的次序表上，計算A篇文章好於B篇文章，C篇文章等等的評斷百分比；次計算B篇文章好於C篇文章，D篇文章等等的評斷百分比；至成一表，計算每篇文章好於另外文章的評斷百分比爲止。排列最劣的文章好於最好的文章的評斷百分比，大概是0；排列最好的文章好於最劣的文章的評斷百分比，大概是100。其餘的百分比，大都在0與100之間。
- 4、上面所有的百分比減去百分之五十(50%)。
- 5、根據百分比的大小，校對P. E.的不同。例如下表：
- 6、下表是直接比較兩種樣本文章的P. E.。此外尚有用許多間接方法，決定這種的不同。例如下表NA的距離等於TN減TA；BN的距離等於AB減AN；LE的距離等於TL減(TA加AN加NB加BK加KE)等等。

表 二 十 二  
表示化評斷百分比為 P.E. 值  
(根據表十九)

| 樣本成績         | A 好於 T | N 好於 A | B 好於 N | K 好於 B | E 好於 K | D 好於 E |
|--------------|--------|--------|--------|--------|--------|--------|
| 百分比          | 50     | 75     | 75     | 54.41  | 64.47  | 91.13  |
| 百分比<br>減50   | 00     | 25     | 25     | 34.41  | 14.47  | 41.13  |
| P.E. 的<br>不同 | 00     | 1.0    | 1.0    | 1.5    | .55    | 2.00   |

7、計算所有直接及間接方法找到 P.E. 不同的平均數。這就是真正的不同(True difference)。方法愈間接，對於計算平均 P.E. 不同的勢力愈弱。

8、依次排列各種樣本的 P.E. 距。例如：

| 樣本成績   | A | T | N   | B   | K   | E   | L   |
|--------|---|---|-----|-----|-----|-----|-----|
| P.E. 距 |   | 0 | 1.0 | 1.0 | 1.5 | .55 | 2.0 |

9、在原來所收集的材料中，計算評斷某某樣本只有零的價值之人數。有的說 K 應得零，有的說 B 應得零，有的說 N 應得零。諸如此類，不勝枚舉。

10、計算中數零的樣本。茲假定中數零的樣本是 A。

11、然後計算每種樣本在零的 P.E. 距。這叫做量表值(Scale value)。因為 A 與 T 有一樣價值，故這量表當如下。

| 樣本成績 | A 或 T | N   | B   | K   | E    | L    |
|------|-------|-----|-----|-----|------|------|
| 量表值  | 0     | 1.0 | 2.0 | 3.5 | 4.05 | 6.05 |

12、從零的樣本起，選擇其餘的樣本；中間大約須有一個P. E. 的間隔。較小的間隔，則為普通應用計，就不適宜；因為一個P. E. 亦只有百分之七十五評斷者能辨別其不同。但為科學研究計，或有專家要比較差別較小的成績時，則用之亦無不可。若兩種樣本有相同的量表值，則同時用兩種，或單取評斷者意見更加接近的一種。

評斷單位的確度及恆度。——編制成績比較表之可能，是根據於克德爾福來登的等距定理，及桑戴克應用這個定理於教育量表編制的方法。這個定理就是：『除非恆見的差異及永所未見的差異外，則屢次所見的差異，若其次數相同，必定相等。』（“Differences Which are equally often noticed are equal unless they are always noticed or never noticed”。）葛第斯曾用實驗方法，欲證明這個定理是否真確？他用圖形做測驗的材料。先請評斷者評斷圖形的大小；然後實際測量。結果，找到人的評斷是依惠栢（Weber）的公例而異其程度。圖形小時，稍有不同，即可辨別；圖形大時，非大有不同，不能辨別。換言之，屢次所見的不同，只當圖形大小差不多一樣時，則次數相等，圖形亦相等。否則，就不相等。故這個定理對於圖形方面，不能成立。若作文，寫字等，適如圖形有絕對的差異，則這個定理恐亦未必真確。

可是即使這是事實，成績比較表仍有他的價值。教育方面所要知道的差異，實有多種。有的是絕對的差異（Absolute differences），就如葛第斯所指出來的是。有的是難度的差異（Difficulty differences），如胡德的算學量表等是。有的是評斷的差異（Judgment differences），如現所討論的成績比較表是。前所討論的百分比表，年齡表，及年級表

上的數值，是根據學生實地測量，能做何種困難的問題而定。這些量表若用成人意見，不用學生的成績為標準，亦可改為成績比較表。可是實際上並沒有人提倡這種方法，因為教育所注重者乃在問題實在的難度，並不在主觀評斷的難度。但對於作文及寫字等，則不然。我們注重實際的價值，並不注重真確的難度。再者，並不注重絕對的價值，而實注重校內外能幹的評斷者之所評斷的價值。

至於恆度，則即如量長短的米突尺，亦非始終絕對不變的。作文的評定，因受評斷者的年齡，性別，曾所受的教育，以及作文的目的等影響，而易其標準，誠為意中之事。現在教育測量者正開始研究這種問題。

成績比較表的特性。——作文，寫文，圖畫等量表，所以與他種量表不同者，因為不是測驗；不過是一種計分的工具而已。所以稱謂成績比較表者，一方面即因此故。至於百分比表，年齡表，及年級表等，則可稱謂工作表；可以用之以測量學生的程度。故實際上是一種測量的工具，並非記分的工具。例如胡德的加法量表是一種測量的工具；他的記分工具，乃是真確的答案。我們收集學生許多作文的成績，則是測驗；至於成績比較表，則是記分的工具。記分的工具及被記分的成績均可製為量表。下表是說明被記分的成績，記分量表，記分的工具，及量表單位的關係。

表 二 十 三

| 被記分的東西 | 記分的量表 | 記分的工具 | 量表的單位 |
|--------|-------|-------|-------|
| 人的高度   | 距線    | 碼尺    | 碼，呎，吋 |

|              |     |       |        |
|--------------|-----|-------|--------|
| 開車時間         | 時   | 錶     | 點,分,秒, |
| 水的熱度         | 溫度  | 寒暑表   | 度      |
| 葛第斯的算學測驗—B組  |     |       |        |
| (a)速率        | 速率  | 無     | 問題     |
| (b)真確度       | 真確度 | 做對的答案 | 問題     |
| 胡德的算學量表—B組   | 難度  | 做對的答案 | P. E.  |
| 桑戴克—麥柯爾的讀法量表 | 難度  | 做對的答案 | T.     |
| 絲大齊的書法量表     | 品質  | 寫字樣本  | P. E.  |
| 作文           | 品質  | 作文樣本  | P. E.  |

## 第 十 章

### T 量表(T Scale)——以年齡變化率為單位

#### (Age Variability Unit)

#### 一、量表編制的方法——以讀法量表為例，其步驟如下：

測驗材料之預備。——1、選擇許多短的，難易逐漸不同的，可以代表校內及校外讀法的材料的，沒有困難科學名詞的，而且城內及鄉間都可適用的散文及詩文。

2、編造許多關於材料內容的問題。其所要求的答案宜短，容易計算；且宜明瞭，沒有文字的困難。難度宜與原文相類似；前後問題亦宜沒有什麼相互的關係。此外尚宜有充分的數目，使測驗有相當的長度。

3、請有經驗的教師答應問題；并請他們排列問題對於兒童的難度。

最容易的放在前面；最難的放在後面。

4、測驗依法排列後，用相當的印刷，并宜附有測驗的指導。竭力使初步的測驗同正式的測驗一樣。

表列個獨的問題。——5、用這初步的測驗，測量數百自三年級至八年級的學生。人數愈多愈好；使有常態的分配。

6、校對學生答案的對誤。

7、刪除不明瞭或不易記分的問題。

8、表列其餘的每個學生的每個問題。其方法如下：

| 學 生 | 第一段問題   | 第二段問題     | 第三段問題 | 其 餘   |
|-----|---------|-----------|-------|-------|
|     | 1 2 3 4 | 1 2 3 4 5 | 1 2 3 | ..... |
| A   | 1 1 1 1 | 0 1 1 0 0 | 0 0 0 | ..... |
| B   | 1 0 1 1 | 1 1 1 1 1 | 0 1 0 | ..... |

9、計算答對每個問題的人數。然後以總被試人數除之，求得答對問題的百分比。

10、依據下表，化答對問題的百分比為 S. D. 這就是每個問題的實在難度。可是用這種方法所找到的難度，亦不過大致可靠。下表原是假定學生能力是照常態分配的。每級學生或每歲學生的能力雖或適合這種假定，但合數級為一組，則能力分配的曲線上面，必將呈平坦的現象；故未能完全合於標準。不過其確度已可滿足普通的應用了。

表 二 十 四

S.D. 值對數表

| S.D. 值 | 百分比       | S.D. 值 | 百分比   | S.D. 值 | 百分比   | S.D. 值 | 百分比    |
|--------|-----------|--------|-------|--------|-------|--------|--------|
| 0      | 99.999971 | 25     | 99.38 | 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   | 75.80 | 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 |

而且因此可以省了許多時間。化百分比為 S. D. 的方法，大致可表示如下：

|          | 第一段問題 |      |       | 第八段問題 |        |       |      |
|----------|-------|------|-------|-------|--------|-------|------|
|          | 1     | 2    | 3     | 1     | 2      | 3     | 4    |
| 答對百分比    | 99.99 | 99.7 | 98.13 | 0.61  | 0.0003 | 0.008 | 0.53 |
| S. D. 價值 | 13.   | 22.  | 29.   | 75.   | 90.    | 88.   | 76.  |

正式測驗的選擇及排列。——11、照實在的難度，重新排列問題。如位置更動頗大者，則當除去。就是已經排好的測驗，亦當重新整理。因為初步的測驗必比正式的測驗大，有些分子且已移動位置；故每分子的難度在正式測驗中，未必與在原來測驗中的難度一樣。

每個問題的難度，一部分固是實際的難度，一部分亦因以前問題對於被試者的精神作用有些影響，一部分亦因離開第一個問題距離的大小而易其難度。一個問題愈放在測驗的前面，可以愈覺容易。再者，



若與另外問題相離，則難度亦將不同。所以位置移動很大的問題，必宜除去。惟因此故，初步測驗須受整理，而後可得正式測驗。

照上面所討論的方法，問題的難度是根據全部學生的成績。普通能顧及此，已經很彀了。可是有時間題的難度，各級不同。這級以為易者，他級或以為難；他級以為易者，這級或以為難。這種問題，叫做不完全的問題(Transient questions)。若欲免除這種問題，則當在每級中重覆上面第九及第十兩個步驟。如是，測量表自可比較進步。不過所費時間甚多，得不償失耳。

12、若量表上有很大的空缺，不能用移動問題的方法救濟時，則可拼合兩個或多個問題以救濟之。用上面第九及第十兩個步驟，計算同時答對兩個問題或多個問題的百分比；然後再由百分比化為S. D. 的價值。

但要注意計算這種百分比，并非平均答對這個問題的百分比及另一個問題的百分比就得。再者，尚當注意兩個問題或多個問題拼合後，必須永遠當做一個問題看待。兩部分或多部分必須完全答得真確，然後可得這個問題的分數。

13、將這樣選擇及排列的材料同測驗指導印在一起，遂成為正式測驗。

正式測驗的應用及記分。——14、為最後測量的目的，選擇許多學校。一起計算，至少須有 500 個自十二歲至十三歲的學生。其餘年齡亦須有相當代表的學生。至於找到很幼及很大的代表學生，自然要比較困難些；因為年幼的有些尚未進入學校，年大的有些已經離開學校

了。

15、用此測驗測量所有三年級至八年級的學生。至於自十二歲至十三歲的學生，則無論在不分班級的學校內或在中學校內，均須測量。

16、每個問題的答案無論是對是錯，均須照原定記分的原則仔細記分，以求一致。普通部分記分(Partial credits)比較是不經濟，且其功用并不一定要比只用對錯兩種方法記分大。可是只用對錯兩種方法記分，亦非絕對必要遵守的。

17、每個問題所有真確的答案，以及最次的答案可認為對者，和最好的答案應認為錯者，均須記錄，用表列之，以為將來記分參考之用。

18、每個學生的測驗本子當歸入他的年齡及年級一類中。下表就是說明這種方法。本國數字代表年級；阿剌伯數字代表年齡（實年）。

表 二 十 五

|       |       |       |       |       |       |
|-------|-------|-------|-------|-------|-------|
| 6—7   | 7—8   | 8—9   | 9—10  | 10—11 | 11—12 |
| 三     | 三     | 三     | 三     | 三     | 三     |
| 四     | 四     | 四     | 四     | 四     | 四     |
| 五     | 五     | 五     | 五     | 五     | 五     |
| 六     | 六     | 六     | 六     | 六     | 六     |
| 七     | 七     | 七     | 七     | 七     | 七     |
| 八     | 八     | 八     | 八     | 八     | 八     |
| 12—13 | 13—14 | 14—15 | 15—16 | 16—17 | 17—18 |
| 三     | 三     | 三     | 三     | 三     | 三     |
| 四     | 四     | 四     | 四     | 四     | 四     |

|   |   |   |   |   |   |
|---|---|---|---|---|---|
| 五 | 五 | 五 | 五 | 五 | 五 |
| 六 | 六 | 六 | 六 | 六 | 六 |
| 七 | 七 | 七 | 七 | 七 | 七 |
| 八 | 八 | 八 | 八 | 八 | 八 |

等分答對問題的總數。——19、計算每個十二歲學生答對問題的總數。表二十六的第一行及第二行，就是表示這種方法。

20、計算所有十二歲學生超過每項成績的人數，加達到該項成績人數之半的百分比。如表二十六有500個十二歲的學生，內497個的成績比零的成績好。得到零分人數之半加497，則得498.5；就是500的百分之99.7。其餘做對一個問題，兩個問題，三個問題等等，均以此類推。如表二十六第三行所表示的是。

21、將百分比化為S. D.的價值，就是量表的分數。如百分之99.7的量表分數當為23；百分之99.3的量表分數當為25。其餘百分比的量表分數，亦均以此類推。這就是所有學生若不能答對一個問題，則在量表上應得23分；所有學生若答對一個問題，則在量表上應得25分等等。

決定年齡的標準分數。——22、表列每歲學生答對各項問題的人數。看表二十七，就可知道。

23、計算每歲學生的總數，每歲學生的總分數，及每歲學生的平均分數。表二十七下面數行，就是表示這種成績。

表 二 十 六

等分答對問題的分數

| 答對問題的總數 | 十二歲學生的數目 | 超過數加達到數之<br>半的百分比 | 量表的分數 |
|---------|----------|-------------------|-------|
| 0       | 3        | 99.7              | 23    |
| 1       | 1        | 99.3              | 25    |
| 2       | 2        | 99.0              | 27    |
| 3       | 1        | 98.7              | 28    |
| 4       | 2        | 98.4              | 29    |
| 5       | 2        | 98.0              | 29    |
| 6       | 2        | 97.6              | 30    |
| 7       | 2        | 97.2              | 31    |
| 8       | 4        | 96.6              | 32    |
| 9       | 2        | 96.0              | 32    |
| 10      | 2        | 95.6              | 33    |
| 11      | 10       | 94.4              | 34    |
| 12      | 3        | 93.1              | 35    |
| 13      | 8        | 92.0              | 36    |
| 14      | 8        | 90.4              | 37    |
| 15      | 13       | 88.3              | 38    |
| 16      | 15       | 85.5              | 39    |
| 17      | 18       | 82.2              | 41    |
| 18      | 28       | 77.6              | 42    |
| 19      | 26       | 72.2              | 44    |
| 20      | 34       | 66.2              | 46    |
| 21      | 40       | 58.8              | 48    |
| 22      | 40       | 50.8              | 50    |
| 23      | 41       | 42.7              | 52    |
| 24      | 37       | 34.9              | 54    |
| 25      | 31       | 28.1              | 56    |
| 26      | 35       | 21.5              | 58    |
| 27      | 24       | 15.6              | 60    |
| 28      | 26       | 10.6              | 62    |
| 29      | 21       | 5.9               | 66    |
| 30      | 14       | 2.4               | 70    |
| 31      | 3        | 0.7               | 75    |
| 32      | 1        | 0.3               | 78    |
| 33      | 1        | 0.1               | 81    |
| 34      | 0        |                   | 85    |
| 35      | 0        |                   | 90    |

表 二 十 七

表示自七歲至十七歲學生答對上表第一行所列問題的人數

| 問題       | 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     | 1     | 2    | 0    | 1    |    |
| 6        | 30   | 2    | 6     | 6     | 5     | 1     | 2     | 2     | 1    | 0    | 0    |    |
| 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     | 20    | 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    | 24    | 17   | 8    | 2    |    |
| 28       | 62   |      |       | 2     | 3     | 13    | 26    | 25    | 23   | 5    | 1    |    |
| 29       | 66   |      |       |       | 7     | 3     | 21    | 19    | 12   | 5    | 0    |    |
| 30       | 70   |      |       |       | 2     | 4     | 14    | 11    | 7    | 2    | 1    |    |
| 31       | 75   |      |       |       | 1     | 6     | 3     | 5     | 4    | 1    |      |    |
| 32       | 78   |      |       |       |       | 0     | 1     | 1     | 3    |      |      |    |
| 33       | 81   |      |       |       |       | 1     | 1     | 2     |      |      |      |    |
| 學生總數     | 35   | 173  | 347   | 399   | 426   | 500   | 452   | 303   | 188  | 16   | 2    |    |
| 量表總分數    | 1190 | 6280 | 13098 | 17673 | 20332 | 25117 | 23516 | 15968 | 6125 | 843  | 116  |    |
| 平均量表分數   | 34   | 36.3 | 39.5  | 44.3  | 47.7  | 50.2  | 52.0  | 52.7  | 51.9 | 52.7 | 58.0 |    |
| 比較準確的平均數 |      |      | 35.   | 41.   | 46.   | 50.2  | 51.   | 60.   |      |      |      |    |

24、計算比較真確的平均分數。三年級以下的年級，沒有什麼測驗可以測量。在自三年級以上的年級中，所被測量的七歲及八歲兒童，大抵都比較聰明；因為比較愚笨的兒童尚在幼稚園，一年級，或二年級中。所以表二十七所表示他們的平均分數，比較實在的成績高。在中校中，除出對於十二歲的學生外，並未施行什麼測驗。十三歲以上的學生，聰明者大概已離開小學；留讀者大都是比較愚笨的。所以表二十七所列這些年紀的平均分數，比較實在的成績低。而且因為人數太少，成績亦不十分可靠。

這種選擇的情形，以及選擇的分量，均是很明瞭的事實。參攷表二十七的每歲學生總數，就可知道。至於十二歲的學生，則全在校內；即有少數已入中校，但亦都受測驗；故人數共有 500 個。離開十二歲愈遠，被測量的人數愈少。照理論方面而言，每歲學生的數目，彼此均當相等的。

求十二歲以上或以下年齡的比較真確的平均分數，是要求每歲的中數成績。我們須有一個假定：就是在學校內或社會中，所有每歲兒童的數目，須與十二歲兒童的數目相等，就是應共有 500 個。如是，我們可以知道年幼未經測量者，是在中數下；年大未經測量者，是在中數上。例如七歲的兒童，只有 35 個被測量。500 個減 35，得 465，離開假定的中數（就是第 250 個七歲兒童）甚遠；故無從計算比較真確的平均分數。八歲的兒童，只有 173 個被測量。500 減 173，得 327，離開假定的中數（就是第 250 個八歲兒童）亦尚遠；故亦無從計算比較真確的平均分數。但九歲的兒童有 347 個被測量。500 減 347，得 153，比 250 少；故可

計算比較真確的平均分數。其方法將 250 減去 153，得 97；即實在中數所在之地。故從上向下數，至第 97 個學生所得到的量表分數 35，就是比較真確的平均分數。用同樣方法，可以找到十歲及十一歲兒童的比較真確的平均分數。至於計算十三歲及十四歲兒童的比較真確的平均分數，則更簡單。因為未曾被測量的學生都在中數上面；所以只要從上向下數至第 250 個學生的地方，就得比較真確的平均分數。他若十五歲，十六歲，及十七歲學生，被測量者沒有超過 250 個，故比較真確的平均分數，就不能計算了。

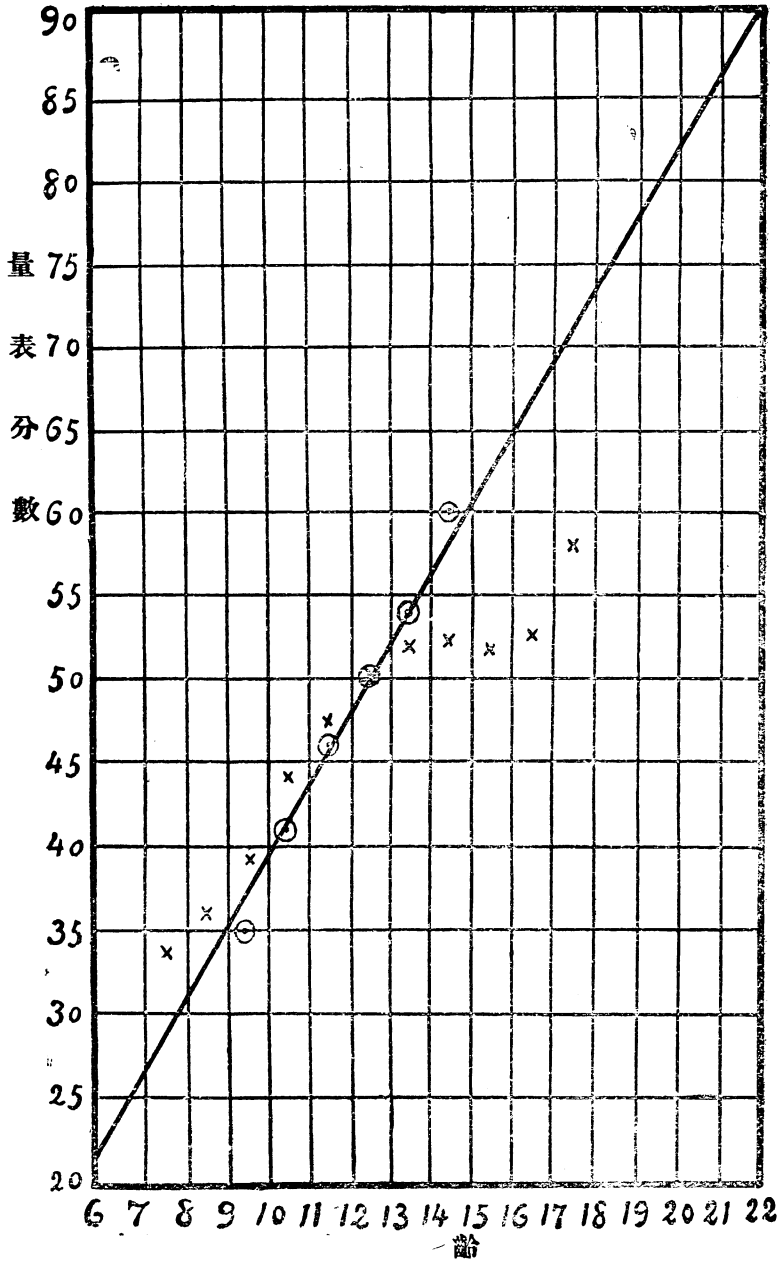
前面已經說過，低年級的平均分數是太高；高年級的平均分數是太低。但比較真確的平均分數，其情形適相反。低年級的是太低；高年級的是太高。計算比較真確的平均分數時，假定未經測量的低年級學生，完全在該歲中數之下；未經測量的高年級學生，完全在該歲中數之上。這種假定，必須普通考試完全真確，升級標準毫無錯誤，才能成為事實。可是實際情形並不一定如此。幸而這種人數比較必是很少；比較真確的平均分數比較原來的平均分數更為真確，並非沒有根據。試看下圖，就可明白。十字×代表原來的平均分數，圓圈代表比較真確的平均分數，所畫的直線與圓圈很相接近，就可幫助我們估定真正平均分數所在的地位。這就是年齡的標準分數。

## 圖形二

註：原來的平均分數 × × ×

比較真確的平均分數 ⊙ ⊙ ⊙

真正的平均分數——





上圖的直線是代表真正的平均分數。雖無實際成績可為根據，但引長之，可知其他年齡的大概情形。且每月的標準分數，亦可從此直線上計算。在表二十七所謂原來的平均分數及比較真確的平均分數者，自然都是根據七歲零六月，八歲零六月兒童等的成績為標準的。

規定年級的標準分數。——25、照編造表二十七的方法，以在三年級上，三年級下，四年級上，四年級下等的學生數目，代替各年齡的學生數目，然後計算每級中每組的平均分數，就得每級中每組的標準分數。

這個量表特別的擴充。——上面這個步驟，已經完成量表及規定年齡和年級的標準分數的工作。但若有時中學學生或小學學生的能力，不能在此量表上規定時，則此量表宜向上或向下擴充。如在表二十六，有一個學生答對問題多於33時，我們除用81+的分數表示他的成績外，就沒有另外方法可以表示；故宜用下法擴充。

26、重覆第十九個步驟，測量中學十六歲的學生。有時即是十四歲或十五歲的學生，亦可應用。

27、假定在中學十六歲的學生有200個，只占本地方十六歲兒童百分之20，則這200個學生大抵要比其餘800個兒童聰明。

28、自次數分配表下面起，計算十六歲學生超過35題的人數加達到35題的一半人數之和。用同樣方法，計算34題及33題的超過人數加一半達到人數之和。

29、計算每題照上法所得人數的百分比。其除數當是1000，並非是200。如是，可略校正其餘未被測量的十六歲兒童之欠缺。

30、化百分比爲S. D. 價值。

31、答對34個問題高於答對33個問題4點，答對35個問題高於答對33個問題9點。原來答對33個問題是81分，則81加4得85，是答對34個問題的分數；81加9得90，是答對35個問題的分數。如是，這個量表就向上擴充。

至於量表向下擴充，實非必要。就遇必要時，照上法測量八歲兒童就得了。但所得分數的不同，須與原來照十二歲兒童所定的最低分數相減；并不如上所述的方法相加。

增加真確度(Accuracy)的方法。——T 量表只有一個缺點：就是量表上端及下端的分數，有不能十分可靠的傾向。這雖不甚要緊；但改正這種缺點，有時亦有相當的價值。其法對於十一歲及十三歲的兒童，重覆表二十六所示的方法。這樣所得到的三種答對各項問題的量表分數之平均數，差不多可與增加十二歲被試者人數三倍的成績一樣。若慮十二歲以上的十三歲兒童不如十二歲以下的十一歲兒童之多，則三種量表分數拼合以前，可用十二歲的量表分數爲標準，使十三歲的及十一歲的量表分數均等。其方法亦極簡單。例如對於十一歲的兒童，答對22個問題，得量表分數51；對於十二歲兒童，當得50；對於十三歲兒童，當得49。如是，所有十一歲的量表分數均加1，所有十三歲的量表分數均減1，則兩組分數就可與十二歲的量表分數比較了。這是假定這三個年齡的變化率，是差不多相等的。如遇必要時，這種應加及應減的分量，若用答對21, 23, 20, 24等數目的問題，幫助只用答對22個問題的決定，則必較爲真確。

量表的印刷。——32、每套測驗須附有一小冊，詳細說明應用測驗及校對成績的方法。其餘如何計算學生的分數及班級的分數，和如何利用年齡的標準分數及年級的標準分數等，均當詳細解釋。有許多很好的測驗，因不注意及此，故就毫無用處，殊為可惜！

現在量表編制的工作，已經完竣。所得的量表，已可應用。但另外尚有一個方法，須待說明：就是利用均衡法 (Calibrator)，編制重覆的量表 (Duplicate Scales)。

均衡法對於編制量表的功用。——因為編制量表時，其數不只一個，故有用均衡法的必要。所謂均衡法者，就是一組在各測驗中相同的問題。均衡法雖非是絕對必要，但若測驗的數目甚多，同樣學生不能同時受許多測驗時；或為別種原因，只用一部分測驗測量這組學生，另一部分測驗測量另一組學生時，則均衡法亦有相當的價值。

若一組學生的能力分配與另一組學生的能力分配一樣，則不必用均衡法。但若兩組的能力不相等，則每個問題在量表上的難度，對於每組學生必不相同；所有 S. D. 的價值，就不能一組與一組比較。如是，當最後排列測驗時，有些問題就不能隨便移動。同時每組問題的量表分數，及根據這種分數規定的年齡標準分數和年級標準分數，亦無互相比較的可能了。至於均衡問題，則為每組學生所共應答，故對於每個單獨問題或所有全體問題，均可利用之，以指定其 S. D. 的價值。

假使一組十二歲的兒童有百分之 49 人對答一個均衡問題，另一組十二歲的兒童有百分之 52 人答對這個均衡問題，校對表二十四，根據

第一組的成績，這個問題應得 50S. D. 價值；根據第二組的成績，應得 49S. D. 價值。平均計算，第二組的能力自然比第一組的好。今若將用均衡問題測量所得兩組 S. D. 價值的差異平均，這就是兩組能力平均差異的表示。

現在假使這平均差異是兩點偏向於第二組，則欲使兩組所有其他問題的 S. D. 價值可以互相比較，必須對於第二組的每個問題加兩點；或對於第一組的每個問題減兩點；或同時對於第二組的每個問題加一點，對於第一組的每個問題減一點。如是得到所有問題的難度價值，可如所有一切問題都被兩組學生全體測量過一樣。用這種方法編制量表，雖實際上所有材料並未都被全體學生測量，可是已得到如全體學生都受測量的利益了。

這種均衡法一方面可使不同的問題有相當的量表價值，可以互相比較；另一方面亦可應用於各許多問題的總和，求其量表分數，使有相當的價值，可以互相比較。

上面兩段所述，亦不過大致是真確。每個特別問題的難度價值，或許多問題總和的量表分數，一半原是根據該組學生的平均能力而定；但一半亦因能力分配的形狀而異。若一組學生就全體而言，是好於或次於他組學生，則用均衡法要平減其差異；蓋均衡法只關涉各組間平均的差異。

但無論用這種方法使各種量表的分數可以互相比較與否，最好須平均所有年齡的標準分數及年級的標準分數，以求較確的年齡及年級的標準分數。但既用均衡法使所有量表的價值可以互相比較，則其

實只要規定一個量表的年齡及年級的標準分數，已足殼了。

編制量表的簡便法。——等分普通考試成績的方法。——這種簡便編制量表的方法，就是省去上面第三，第四，第五，第六，第七，第八，第九，第十，第十一，及第十二諸步驟。有時選擇測量問題的目的，原為診斷之用；故不應為統計及記分方面的緣故，刪去其中的問題。有時排列問題，除出難度外，尚有另外須考慮的地方；故決定問題的次序，毋須經初步測驗之試用。例如編制讀法量表，排列問題時，欲使前後問題的難度逐漸增加，實為不可能之事；因為散文或詩文的問題必須各在一起。再者，測驗問題亦不必一定依難度而進行；若略有難易的先後，已可滿足普通一般測量目的的要求了。

若是一個量表的目的，不在決定年齡及年級的標準分數，則方法實可簡便許多。在這種情形下，所需要的只要測量每個學校的十二歲兒童就殼了。故可減為下面五個步驟：

- 1、直接預備正式測驗。
- 2、測量未經選擇的十二歲兒童。
- 3、校對成績，計算每個學生答對問題的總數。
- 4、計算答對某數問題的超過人數加達到人數之半的百分比。
- 5、化百分比為S. D. 價值；這就是量表分數。如是，量表遂成。

這種方法，如此簡便，教師誠可利用之，等列考試的成績。其方法如下：

- 1、在教室內施行考試。
- 2、用「對」及「錯」的方法，校對每個問題的答案。或用其他適宜的

方法。

- 3、計算每個學生的總分。
- 4、計算超過某種分數的人數，加達到某種分數的人數之半的百分比。
- 5、化百分比為S. D. 價值；這就是量表分數。
- 6、每個學生成績的記分，當根據這種量表分數，不應根據原來的點數。

教師用這種量表分數，保存學生的成績，很為便利。如是，前後考試的分數就可互相比較。再者，這種分數有統計的形式，可用簡便加法計算混合分數。他若用 A, B, C, D, E 等表示成績時，則就不能了。

## 二、量表編制方法的討論

參照點 (Reference Point)。——無論何種測量，記分時必須有相當的起點，這就是參照點。如經度以格林威齊 (Greenwich) 為參照點，緯度以赤道為參照點等是。其他無論測量距離，時間，重量，勇敢，讀法能力，計算技能等，都應有記分的參照點。

若沒有一定的參照點，則科學的進步必受很大的妨礙。精神方面的測量，欲研求一個相當的參照點，為時已經很久。普通的傾向，大致欲研求被測量的能力的絕對零點為參照點。其結果，遂使每個量表各有一個不同的參照點。這種混亂情形，實可阻止心理測量的進步。

不特每個測驗的參照點不同，即規定參照點的方法，亦不一致。1、不分等級的測驗之參照點，每以沒有得到該測驗的分數為根據。2、有些量表的參照點以編制者所猜定的零點為根據。3、有時以評斷者平

均評定零的成績為量表的參照點，如評斷量表 (Judgment Scale) 等所用的是。4、以低年級學生受一種測驗，不能得到任何分數的人數百分比為標準所規定的零點為參照點。5、有時以被試者平均成績下的3S. D. 為參照點。6、有時只以最低的成績為參照點。及其他種種。

因為規定零點的方法不能一致；因為不知每種能力的零點究竟是什麼；因為即使實際找到零點，亦必不易明瞭，結果必致解釋成績時，沒有充分的價值；因為現在所用的各種方法，必定產生許多不同的參照點；且因為有許多不同的參照點，可以發生許多混亂的情形，所以麥柯爾介紹一個普通的參照點。這種參照點，可應用於中小學所有的測驗。他並不是用零點為參照點；但用十二歲兒童的平均成績為參照點。無論何種能力或特性，不管絕對的零點是在什麼地方，均可應用。

上面所敘述編制量表的方法，就是用這十二歲兒童平均的成績為參照點。無論何個學生，若得到50量表分數，他的能力就等於十二歲兒童平均的程度。一半十二歲的兒童在他之上，另一半十二歲的兒童在他之下。若是一個學生只得到40量表分數，則他的能力在十二歲兒童的平均成績下一個S. D. (就是十個單位)。換言之，十二歲的兒童有百分之84.13超過他的程度。他若一個學生得到75量表分數，則他的能力在十二歲兒童的平均成績上兩個半S. D. (就是二十五個單位)。換言之，十二歲的兒童只有百分之0.62超過他的程度。

實際上就算學的意義而言，這種量表的零點是在十二歲兒童的平均數下五個S. D.。這就可說明這種量表的零點，並不是實際的參照點。這種數學的零點在平均數下五個S. D.上，而不在平均數上，有四

個理由，可以說明。

1、免除正負號的麻煩。若以平均數為零點，則學生的分數將必當用正的三個 S. D. 或負的兩個 S. D. 等報告。今用 50 或 75 等報告，則便利多了。

2、這種方法可有自 1 點至 100 點的分數限域。規定 50 點為參照點，且很容易記憶。

3、量表上下兩端有充分的範圍，任何學生的成績，均可表示。至若只用從正的三個 S. D. 起至負的三個 S. D. 止，則就不能了。

4、這種數學的零點，可與讀法，拼字，作文，以及其餘能力的測驗所假定的絕對零點相接近。調查阜京海姆屈來阜胡德等所編制的測驗，找到他們所規定的零點，不過稍高於十二歲兒童的平均成績下五個 S. D. 而已。故用十二歲兒童的平均成績下五個 S. D. 為零點，理由既充足，實際應用亦適宜。

應用何種參照點，亦視測量作何用度而定。普通公認比較學生能力的最良方法，是用一般學生能力的平均數為參照點。另一方面，普通亦公認若用倍數說法，表示學生成績，則最好用能力絕對的零點為參照點。例如我們要說學生 A 的能力兩倍於學生 B 的能力，則用這絕對的零點，是較為適宜。麥柯爾所介紹的參照點，差不多完全可以滿足第一種情形的要求。至於他以十二歲兒童的平均數下五個 S. D. 為數學的零點，亦可滿足第二種情形的要求。因為上段說過，實際參照點下五個 S. D. 差不多與絕對的零點一樣。雖非完全相同，但用此法可以彼此一致。如是，可以免除混亂的困難。



用十二歲兒童的平均分數為參照點，或可發生一種缺點：就是在參照點以上或以下的分數，不能表示一個兒童究竟有多少能力；因為普通十二歲兒童對於一種能力或可以很高，對於另外一種能力或可以很低。若用絕對的零點，就無這種弊病。可是這種絕對的零點，實不容易找到。且教者知道十二歲兒童平均的程度，用之以估量一個學生的能力，其方法必可比較用理論的絕對零點為參照點更為容易。再者，測驗材料自身亦可為估量學生能力的參考，原不必一定要用絕對的零點。

另外一種便利的參照點，則為個人誕生的時候。西特弗訂正皮內西門的智力量表就是用這種參照點。一般智力測驗都應用之。其不甚適宜者，因需用一種不十分滿足的測量單位。這層以後將再討論。

測量的單位 (Unit of Measurement)。——測量須有單位，如測量須有參照點一樣。如測量山的高度，參照點是海平面；測量單位是呎。測量時的久長，參照點是耶穌誕日；測量單位就是世紀，年，月，日，時，分，秒。

關於精神方面的測量，不特有許多不同的參照點，且有許多不同的單位。有些用簡單記分的單位，有些用一級的變化率為單位，有些用所有年級的變化率為單位，有些用評斷的變化率為單位，有些用成人作業的變化率為單位，有些用比較的價值 (Relative Worth) 為單位，以及其餘種種。可是此書則提議只用一種單獨普通測量的單位。凡小學所應用的測驗，則用十二歲兒童的成績變化率；中學所應用的測驗，則用十六歲兒童的成績變化率；小學初年級所用的測驗，則用八

歲兒童的成績變化率。最好，就是用S. D.。

用這種變化率為單位，是桑戴克及其學生發明。這種發明，對於測量方法的進步，有很大的貢獻。可是只用一級的變化率為單位時，則仍不十分適宜；因為各方面情形如學級制度的變更，分班方法的改良等等，必將在很短時期內，變易所用單位的意義及量表的價值。無論何種單位，若根據這種機械的，易變的年級成績，則必缺乏固定的性質。普通只有一種變化率，可具有充分的意義；就是根據相當固定一組的成績變化率為標準。

另外一點，對於量表的編制，頗具有價值：就是皮內及西門所發明，推孟所改進的量表，以生時為參照點，以年月為單位。這種參照點及單位，人人都可懂得；故頗為一般所應用。但有幾種缺點，茲分述之如下：

1、這種單位，雖說是固定的；但在量表上，並不是一定都相等。八歲至九歲的間隔，往往比十四歲至十五歲的間隔大。不特智力方面如此，即其餘能力亦或如此。

2、有許多能力，自十六歲後，就無進步。故所用年齡的單位，就成零點。倘若退步，則實際上比零還要小了。

3、因為有些學生入學甚遲，有些學生離校甚早，故欲編制一個量表為八歲以前及十二歲以後的兒童，就覺困難。換言之，在這種情形下，所編制的量表就不能測量很劣及很好的學生。

但因為解釋容易，計算便利，故已成為很普通的方法。且智力非至老年時，或因其他特別情形，不至突然衰退；故用之實頗相宜。至若為

他種能力測量的單位，則就不甚妥當。

至於用十二歲兒童成績的變化率十分之一的S. D. 為單位，則在量表上，任何一點的難度都相等；故具有桑戴克及其學生所用單位的特性。同時這種單位亦以實年為根據；故亦具有推孟及其先進者所做工作的長處。用這種單位編制量表，上下都有充分的餘地，為實際之應用。即量表上下兩端的擴充及編制量表自身的手續，亦均不十分麻煩。總言之，這種單位實包含兩種逐漸發達的科學心理測量的主要方法：一為桑氏所採用的變化率，一為推氏所採用的實年；蓋這種單位就是固定一歲的變化率。普通均用T名之，叫做T分數。原以紀念桑推二氏(Thorndike 及 Terman) 的貢獻。

從前用年齡的變化率為單位，只可為同年學生比較用。但心理測量中，所需要者，在使某歲學生可與其他各歲學生比較，或某歲學生的平均可與其他各歲學生的平均比較。用T分數，就可滿足這種要求。

上面所提議的單位及編制量表的方法，差不多測量無論何種能力都可應用。如欲編制一個量表，測量一個學生能做何種難度的工作，可說是毫無困難。其他如欲比較價值等，凡有等別分數之可能者，均可應用之。

成績比較表可改為T量表。如是，可使成為工作的量表。我們前面曾經說過，成績比較表如作文，寫字等，不過是一種記分的工具。因為作文，寫字等，須一次記分；故記分者有用這種已經編定記分的工具之必要。可是若進一步，將十二歲兒童用成績比較表所得到的分數，照上面表二十六所示的方法，求得量表分數，則就成為T量表。如是，

所有量表，均可互相比較。

拼合記分單位的方法(Method of Combining Scoring Units)。——普通測量及量表，計算學生的分數，大都將做對的單位用簡便法相加。例如受葛第斯的加法測驗，則學生的分數就是他所做對問題的總數。受胡德的加法測驗，則學生的分數亦是根據他所做對問題的總數；但用P. E. 表示問題的難度。計算P. E. 數目的大小，決定學生程度的高低。他若受桑戴克的生字測驗，學生的分數亦以所超過P. E. 距的總數為標準。至於用納索(Nassau)的作文量表或愛禮斯的寫字量表，則計算學生的分數，就以超過某種成績單位的總數為根據。

用簡便方法計算單位總數，表示學生成績的量表，至少有三種不同的計算方法。

1、成績比較表的計算方法：校對時，將成績逐漸依表向上比較，找到達到相當的地位為止。那以下的成績，必比較要次些。

2、如胡德的加法測驗等，比較超過P. E. 的大小，決定學生能做如何難的問題。可是一個學生能做對四個P. E. 難的問題，未必一定能做對四個P. E. 難度以下的問題。

3、如桑戴克及海格底(Haggerty)的讀法量表用百分之二十錯誤法(20% error method)，如胡德的算學基本量表用百分之五十錯誤法(50% error method)等，在量表上有百分之二十或百分之五十錯誤的一點，決定一個學生的分數或一班學生的分數；並觀察這點以上及以下的情形。

以上三種方法，第一種最為妥當。至於第三種，則宜於難度級距不

等的量表；方法甚為複雜，普通不大應用。

除簡便法計算總分數 (Method of simple total) 外，尚有積重計算法 (method of cumulative total)。茲用孟祿的標準算學推理測驗為例，來說明這兩種方法的不同。試看下表：

表 二 十 八  
簡便法與積重法計算總數的比較

| 問 題           | A | B | C | D | E | F | 最高分數 |
|---------------|---|---|---|---|---|---|------|
| 零點以上的 P. E. 值 | 1 | 2 | 3 | 4 | 5 | 6 |      |
| 簡便法相加         | 1 | 1 | 1 | 1 | 1 | 1 | 6    |
| 積重法相加         | 1 | 2 | 3 | 4 | 5 | 6 | 21   |

這兩種方法不同的情形，可用下面一個比喻表明之，就是計算人的高度。這個比喻，證明積重法實有欠妥的地方。

|       | 腳底 | 踝節  | 膝蓋   | 臀骨   | 肩頭   | 頭頂   | 總高    |
|-------|----|-----|------|------|------|------|-------|
| 量表值   | 0  | 3 吋 | 22 吋 | 41 吋 | 61 吋 | 74 吋 |       |
| 簡便法相加 | 0  | 3,, | 19,, | 19,, | 20,, | 13,, | 74 吋  |
| 積重法相加 | 0  | 3,, | 22,, | 41,, | 61,, | 74,, | 201,, |

但下面另外一個比喻則不然。其實且可證明積重法的長處。這個比喻就是一個小孩轉運木頭；就其轉運磅數而付工錢。在這種情形下，只得用積重法計算。

|       | 第一條  | 第二條  | 第三條  | 第四條  | 總 數  |
|-------|------|------|------|------|------|
| 量表值   | 10 磅 | 15 磅 | 20 磅 | 35 磅 |      |
| 簡便法相加 | 10,, | 5,,  | 5,,  | 15,, | 35 磅 |
| 積重法相加 | 10,, | 15,, | 20,, | 35,, | 80,, |

這兩種方法各自測量不同的東西。簡便法相加，測量能力的高度；積重法相加，測量所做工作的分量，據上例而言，前者測量小孩能運如何重的木頭；後者測量小孩已運多少重的木頭。

兩種方法各有他的用度。但對教育測量，積重法並沒有特別的好處。實際我們估量一個學生的能力，大都是用簡便法相加的結果。比較兩個人的能力，普通同比較兩個人的高度一樣。若欲解釋用積重法相加的成績，使有相當的意義，我們必須記牢一個學生得到四分，另外一個學生得到十分，這第二個學生所做的分量是比第一個學生多二倍半；並不是第二個學生的能力比第一個學生高二倍半。解釋成績時，若能注意及此，則用無論何種方法，沒有什麼不同。因為兩種方法計算的分數，雖非成一定的比例，但彼此必有很密切的相關度。當一個測驗測量學生最高能力時，尤其如此。

用簡便或積重的方法，測量學生能做如何難的材料，不應與測量速率的目的相混。普通速率測驗的每個測驗分子，大都需要同樣時間工作。所以所有的單位，彼此相等；學生的速率可根據在一定時間內所做分子的數目而定。但有時速率測驗的工作單位，需要不同的時間；在這種情形下，必須規定每個工作單位需要幾何時間。換一句話說，就是每個工作單位須有他的時間值 (Time value)。欲測量學生工作的速率，就可將學生做完測驗分子所有的時間值相加。但宜注意每個分子的時間值，並不是從全個測驗開始時算起。

由上而論，可知T量表不僅有較好的參照點及測量的單位，而且有較妥的拼合分數方法。用這種方法，每個學生可得一個純粹的量表分

數;且可僅用簡便法計算;同時且可保存原有的測驗材料。即使沒有受過特別訓練的人,亦可應用之。所有目的,只用等分總分數的方法(Scaling total score),就可達到。從前所討論的種種編制量表方法,除等分測驗分子以外,尚須另有計算學生分數的方法。至於 T 量表所用的方法,則第二個步驟就可不必要了。

再者,用等分總分數的方法拚合記分的單位,尚有另外一種好處。普通預備重覆測驗,手續非常麻煩。若用一種量表測量學生,他的分數根據做對測驗問題的總數,則每個測驗分子必須與重覆測驗上每個測驗分子有絕對相等的難度。這種條件,在 T 量表上,並非必要。即使兩種量表沒有相等的難度,一個量表亦可為其他量表的重覆量表。

附註: T 分數的意義及求法,上面已經詳述。此外尚有 B, C, F 三種分數。普通與 T 分數相聯,稱謂 T, B, C, F 分數,可是這後三種分數的意義及求法,原書未及備載。可是從事於教育者,則不可不知;故在附錄二中補述之。

## 第十一章

### 信度,客觀性,及標準分數的規定

#### 信度 (Reliability)

個測驗缺乏信度的原因。——所謂信度者,是同一主試用同一測驗測量同樣的被試者兩次或多次所得結果相同的程度。欲得完全的信度,必須有一樣的主試,用兩個絕對一樣的測驗,依照一樣的手續,測量一樣的學生。凡此種種,均是得到最高信度的條件;缺乏這些條

件，就是缺乏信度的原因。

一個測驗缺乏信度的第一種原因，是主試因受測驗以外的種種影響，致先後的行為各不相同。這種情形，均可使被試者的分數特別高或特別低。如喧嘩的學生，教師的笑容，或飯後不消化等，均是可能的的主要原因。有時機會可使第一次測驗特別便宜，有時或使第二次測驗特別便宜。兩次測驗的成績既不能相同，故信度亦必減低了。這些致損信度的原因，均非屬於測驗本身的。

一個測驗缺乏信度的第二種原因，是主試因受屬於測驗本身的種種影響，致先後的行為彼此不同。這種原因，或是在測驗的指導方面，或是在記分方面，或是在統計方面。但普通最主要的原因，是因為沒有充分的說明。理想的說明，須確實表示測驗時及記分時等所發生種種情形的處理方法。測驗的指導 (Instructions) 是測驗的一部分，其應確定及一致，當如測驗材料一樣。他若如何記分，如何統計，亦當有一致的指導。如皮內的測驗做法，都有一定的標準，即其例也。

一個測驗缺乏信度的第三種原因，是學生自己時時刻刻的變化。如汽車的鳴聲，家犬的吠聲，遊戲的衝動，久坐的疲乏等，不勝枚舉。無論內部或外部的影響，均可左右他的成績，使之或高或低。兩次測驗的時期既不能相同，成績亦必將有差異。因此，測驗就減小信度。

但用每個學生兩次的成績絕對相同，做決定測驗信度的標準，亦非時常公允。普通每有幾種固定的原因，致令成績發生絕對的不同。所以一個測驗的信度，每比實在的信度低。這些原因，應當除去；或解釋成績時，有相當的說明；然後可以決定實在的信度。如因第一次測驗



所得的經驗而進步,及所測量的能力有自然的發展等,均是其例。蓋學生的能力,因年齡的增長,或經驗的加高,是時時刻刻有變化的。

用什麼方法,可以免除這些固定的原因 有四種方法,可以利用:

- 1、適宜的間隔(Optimum interval);
- 2、用重覆測驗(Duplicate test);
- 3、實驗的核減(Experimental allowance); 及
- 4、自身相關度(Self-correlation)。

前三種方法的目的,乃在發見每個學生兩次成績的絕對相同。至於最後一種方法,則為比較而用。

所謂適宜的間隔者,就是前後兩次測驗有相當的間隔,學生的能力不至發生什麼不同。零的間隔,自然是不可能的。兩個測驗不能在同一時間測量同樣學生,這是很明了的事實。間隔時間愈長,受第一次測量的練習影響愈小;但亦不可太長。否則,因遺忘而減小的能力,或將因常態成熟而進步的能力補充之。

選擇適宜的間隔時,須參攷下面兩種情形:

- 1、能力因成熟長進,往往比因遺忘退步慢。
- 2、有些測驗具有一種特性,學生不易傳達做法於他人;且在短時間內,即將遺忘絕對的做法。

用重覆測驗,可以幫助上法的不足。用這種方法,若是速率測驗(Rate tests),則可免除一部分練習的影響;若是難度測驗(Difficulty tests),則可免除全部分練習的影響。

用實驗核減的方法,就是利用比較組(Comparable group),決定間

隔的影響，而後核減因此影響發生的成績差異。

求自身相關度，是免除各種固定錯誤的最簡便方法。可以單獨應用，亦可與選擇適宜間隔及利用重覆測驗兩種方法同時應用。這種方法，就是用同一測驗或重覆測驗測量同樣學生兩次所得成績的相關度。若相關度等於零，則這個測驗就毫無信度。無論另外各種品質如何好，亦沒有價值了。相關度愈大，信度愈高。

一個學生的成績，雖或因受練習的影響，或成熟的增進，而前後不同；但若別種情形一樣，則無論間隔如何長久，相關度亦不至突然減小。

增加信度的方法。——一個測驗的信度，應當如何高？普通的答應，當然是愈高愈好。若自身相關度是零，則這測驗就無價值；若自身相關度是合於1，則這測驗的信度是完全。普通有五種標準教育測驗的自身相關度：.55, .7, .75, .8, 及 .9。測驗的成績，完全根據學生的分數。一班的分數，往往是學生成績的平均數或中數。若一班有充分的人數，則雖測驗的自身相關度只有.55，而其成績亦已可靠；足以滿足一般應用的目的。但若這種測驗的分數將為診斷單獨學生用，則自身相關度須在.9以上。可是在.9以下，亦有相當的價值；因為一個測驗的自身相關度在零以上，必比較零的自身相關度好。

若是一個測驗的信度，尚未及可以應用的目的，則如何可以增加之？上面已提出各種方法，如何減小主試的變化，使有比較合於標準的測驗手續；如何減小被試者的變化，使成績不受別種影響而生意外的差異等是。

此外尚有三種方法,可以去除缺乏信度的原因:

1、增加測驗的長度。

2、平均一個測驗數次測量的成績,或一個測驗與重覆測驗的測量成績。

3、拼合上面兩種方法。

但若只有一個測驗,則經屢次測量後,學生對於內容及方法,必日益熟悉;欲求有相當的信度,實不容易。可是亦因測驗的性質不同而異。故當有很高信度的必要時,須用重覆測驗。重覆測驗乃是決定測驗價值的最要手續。再者,重覆測驗可以免除教學(測驗本身的內容及方法等)的危險。

至於應用一個測驗或重覆測驗幾次才可得到相當的信度,及應用一個測驗或重覆測驗數次可以得到如何的信度,則可用斯比門(Spearman)的求自身相關度的公式(Self-Correlation formula)。

$$r_x = \frac{Nr_1}{1 + (N-1)r_1}。$$

若一個測驗同一個重覆測驗的自身相關度( $r_1$ )是.8,現要找到應用這個測驗幾次(N),才能得到相關度( $r_x$ )是.9? 代入上面公式,求N的數值。

$$.9 = \frac{N(.8)}{1 + (N-1).8}。 \quad N = 2.25次$$

現若應用這個測驗四次,則可得到如何大的相關度( $r_x$ )? 代入上面公式,求 $r_x$ 的數值。

$$r_x = \frac{4(.8)}{1 + (4-1).8} = .941。$$

## 二、客觀性(Objectivity)

客觀性的重要。——同一主試用同一測驗測量同樣的被試者兩次，若所得到的成績完全符合，則這個測驗就完全可靠。但若是不同的主試用同一測驗測量同樣的被試者兩次，所得到的成績完全一樣，則這個測驗可說是有完全的客觀性。若兩個主試所得的成績不符合，則就完全屬於主觀了。普通一個測驗的客觀性必比信度低，因為加了主試的不同。客觀性的大小，亦視測驗性質的不同而異。如溫度及脈動等測驗，比較人的美貌等測驗，必容易富有客觀性。估量人的高度，原是主觀的；可是利用測量的器具，就可增加客觀性。這種情形，在教育測量方面，亦是一樣。所有的測驗並非絕對是主觀的，亦非絕對是客觀的。客觀性實如信度一樣，是一種程度大小的關係。無論何種測驗，大抵均居於這兩個極端的中間。

客觀性對於測驗之編制，非常重要。普通每謂客觀的標準，是科學的教育測量之母，誠非過言。從前測量教育成績的方法，差不多完全是主觀的。教育測驗的產生，實因為有這種不滿足的情形。各種科學的進步，亦均因這種測量工具的改進，減少這種人差(Personal equation)而已。驗證(Verification)一字，已成為科學中最普通的名詞；教育進步，得成為一種科學者，全恃這種測量工具及方法之發明，使能得到可驗證的成績。

如何可以決定及增加客觀性？——一個測驗的客觀性，可以用下面一個公式決定：

$$\text{客觀性} = \text{信度} - \text{人差}。$$

但用直接的方法計算客觀性,比較更加容易。這種直接法同決定信度的方法一樣,就是比較兩個主試所得成績相同的程度,或計算同樣學生兩組分數的相關度。至於所應預防的固定錯誤,亦與決定信度時一樣。所以求自身相關度的方法,對於決定客觀性,與決定信度一樣便利。

然則如何可以增加客觀性呢?教育的問題與其餘科學的問題並無不同之處。解決這種問題的第一步驟,當如前節所示,利用各種方法,增加測驗的信度。第二步驟,就是決定不同主試的人差情形及分量;並且照量解釋之。將來增高信度,或就可為增加客觀性的最利便良好的方法。

客觀性亦如信度一樣,可因全部的測量方法合於一定的標準而增加。若是兩個主試用不同的方法,施行同一的測驗,則成績不能一致,是可預料的。他若記分方面,亦如此。再者,如統計方面,無一定的方法,則所得到的結果,亦必不能一致。倘若應用測驗及統計方面均有充分說明,則可免除許多變異。至於對於記分方面,則現已逐漸應用比較客觀的方法。

### 三、標準分數(Norms)

影響標準分數的價值之事項。——普通應用的標準分數有兩種,我們必須分別清楚。1、成績的標準分數(Standards of achievement), 2、為應有的成績標準(Standards for achievement)。前者是一歲或一級學生的平均成績;後者是一歲或一級學生所應達到的目標。前者叫做標準分數(Norms),後者叫做成績標準(Standards)。

若是標準分數為最宜於比較的一組學生之代表成績，則必更有價值。倘若只能有一個標準分數，則這個標準分數當為全國學生的標準分數。求得這種標準分數，並不要測量全國所有的兒童。今若選擇各級能力的代表兒童，兒童數目有相當的比例；然後測量之，即已足數。所謂相當比例者，就是某級能力的學生占全體學生百分之十時，則這級能力的學生受測驗者，亦當占受測驗學生的全體百分之十。

若是標準分數是固定不易者，則必較有價值。標準分數之固定與否，原視當時被試人數之多寡及選擇被試者之適宜與否而定。選擇被試者的方法，上段已略言之。至論被試者的數目，則人數愈多，標準分數愈固定。但繼續增加人數，並不一定繼續增加固定性。我們須要知道者，就是必須有若干被試者，然後才能得到固定的標準分數？關於此層，現在尚無絕對真確的方法可以介紹：不過普通均用一種簡便法決定之。就是平均各組被試者數目不同的成績，然後觀察平均數的變動狀況。若增加一百個新被試者，尚未變動以前所定的標準分數時，則這標準分數便是已經固定了。

規定固定標準分數所需被試者的人數，亦視標準分數的類別而異。普通有三種標準分數：1、平均的工作標準分數(The average performance norm)；2、在一個量表上規定一個測驗所在的特別年齡的標準分數(The placement of a test at a specific age on an average scale norm)；3、百分比或變化率的標準分數(Percentile or variability norms)。決定一個測驗分子在年級上的位置，同規定百分比的標準分數相類似。普通量表如皮內西門的智力量表等，在年齡表上決定一個

測驗分子的位置，都是根據百分之二十五兒童失敗及百分之七十五兒童成功的年齡為標準。選擇這些百分比為標準，原是照百分之二十五最劣的學生不能通過這個測驗分子，百分之五十中等學生及百分之二十五最好的學生應當通過這個測驗分子的原理。因為近中數的學生必比近百分之七十五的學生多，因為近百分之七十五學生必比在全體極端地方的學生多，故欲得到固定的百分比的標準分數，需要學生的數目須比決定一個測驗分子在年齡表上的位置多。至於平均的標準分數，則有少數被試者，就可得到相當的固定性。

若求標準分數的方法有明白的說明時，則這標準分數必更為有用。這雖是很明了的事實；但普通往往疎忽此層，並不說明求得的方法！究竟是平均的標準分數，抑是中數的標準分數，抑是七十五百分比的標準分數等，均不得而知之。

若標準分數有充分的報告時，則必較為有用。有一個著作者報告一級各班最高的成績，為測驗的標準分數。這種方法可以鼓勵教師的教學達到最高的標準。可是這種刺激有其利亦有其弊；因為容易過重一種科目。所以比較適宜的方法，當報告最高及最低的分數；或進一步，當報告上部四分點及下部四分點的分數；或更進一步，當報告所有百分比的分數。報告愈加完全，標準分數愈加有用。但無論有幾個標準分數，或只有一個標準分數，最好報告一種簡單的平均分數。

若標準分數兼具有普徧性(Universal)及地方性(Local)者，則必較為有用。一個平均分數，很廣闊的區域可以應用，固然很好；但為許多特別地方，亦應有區別的標準分數。可是普徧性及地方性雖是很重

~~~~~  
要。但若只有良好的標準分數，亦不能使另外條件不良的測驗成爲良好的測驗。

最後，若標準分數同時包含有年齡及年級的意義，則必較爲有用。若只爲年級的，則學制不同，就不能應用。至若兼爲年齡的，則就可免掉這種弊病。再者，利用年齡的標準分數，可以計算讀書年及讀書商數，拼法年及拼法商數，智力年及智力商數等。其餘功用，本書亦已先後提及多次，茲不贅述了。

附 錄 一

求平均數,中數,四分點差,均方差,及相關度的方法

一、平均數——平均數為表示分配的集中趨勢的方法。其求法有多種,茲分別述之。

1、設有二十個學生的國文分數為 45, 80, 45, 65, 50, 88, 73, 65, 84, 63, 40, 50, 75, 50, 40, 45, 75, 74, 74, 60。求其平均數,即將各個分數相加,得一總數1241,然後以次數的總和20除之,得62.05,即為平均數。

2、有時同樣分數,發現若干次數,即可將分數列為一行,次數又列一行,然後以次數乘分數,而得量數的總和;再以此數的總和除之,亦可得平均數。如下表:

表 一

平均數的求法

分數	次數	分數×次數
56	2	112
60	4	240
65	6	390
66	5	330
68	6	408
70	7	490
71	5	355
72	4	288
74	3	222
75	2	150
	44	44)2985(67.84

3、若分數或量數已歸入組距，則平均數的求法，宜照表二。

表 二
另一種平均數的求法

組 距	次 數	量數與假設平均數的差	次數×差數
5—10	3	—4	—12
10—15	4	—3	—12
15—20	7	—2	—14
20—25	7	—1	—7
25—30	9	0	—45
30—35	6	1	6
35—40	5	2	10
40—45	3	3	9
45—50	2	4	8
	<u>46</u>		<u>33</u>
$-45 + 33 = -12, \quad -12 \div 46 = -.26, \quad -.26 \times 5 =$ $-1.30 \text{ 校正數. 假設平均數} = 27.50, \text{ 平均數} = 27.50 + (-1.30) = 26.20.$			

上表的方法程序如下：

- (a) 求次數的總數得46。
- (b) 假定平均數在25—30一組，(依略約觀察，此組的次數最多；且居中間地位。)而以其組距的中點27.50為假設平均數。
- (c) 求各組距與25—30組距的差，而得差數—20，—15，—10，—5及5，10，15，20；再以5(組距單位)除之，得—4，—3，—2，—1及1，2，3，4。寫於上表的第三行上。

(d)以差數乘次數，將其結果寫於上表的第四行上。將負號的總和 45與正號的總和 33 相加，得-12；以次數的總和46除之，得-1.26。再以5乘之（因先以5除），得-1.30，此即為校正數。

(e)真正平均數，即為假設平均數加上校正數，如 $27.50 + (-1.30) = 26.20$ 。

二、中數——中數亦是測量集中趨勢的一法。就是量表上的一點，將量數的數目，分為兩半，左右相等。欲求若干量數的中數，必須經過二步手續。第一步，將量數依大小排列，從一端數起，至量數一半止。此處為分配中心，兩邊量數相等，即為中數。茲將求中數的方法列下：

1、量數數目的總和，為單數者，則中數的求法，當如下表。其公式為：中數 = $\frac{N+1}{2}$ ，這公式內的 N，就是次數的總和。

表 三
中數的求法

量 數	算 法
60	次數的總和 = 11 \therefore 中數的位置 = $\frac{11+1}{2}$ $= \frac{12}{2} = 6$ 即中數的價值 $= 75$
65	
65	
70	
75	
75	
75	
80	
90	
90	
95	
次數 11	

上表的算法，即先將分數依大小排列；然後加其次數得11，為單數。11加1得12，以2除之，得6。此即十一個分數內的中數為任何一端數起之第六個量數；故中數為75。

2、量數數目的總和為雙數者，其中數的求法，則詳下表：

表 四

另一種求中數的方法

量 數	算 法
75	中數的位置 $= \frac{12+1}{2} = \frac{13}{2} = 6.5$
75	
76	
77	
78	
79	∴中數的價值
83	$= \frac{79+83}{2}$ $= \frac{162}{2} = 81$
85	
87	
87	
88	
89	
次數12	

3、若量數或分數，已歸入組距者，則求中數的手續稍繁，見表五。

表 五
再一種求中數的方法

組 距	次 數	算 法
5—10	2	$\frac{64}{2}=32$ 從上數下, $32=2+3+4+$ $6+9+(30-35$ 內的10分 之8). 中數的價值= $30+\frac{8}{10}\times 5=30+4$ $=34.$ 或從下數上,亦可計算.
10—15	3	
15—20	4	
20—25	6	
25—30	9	
30—35	10	
35—40	8	
40—45	7	
45—50	6	
50—55	5	
55—60	4	
	64	

上表的方法程序如下:

- (a) 以2除次數的總和64得32。
- (b) 從次數的上端數下,直至含有中數的組距之前一組(25—30)為止,共24。
- (c) 以32減24得8。
- (d) 以30—35組內的次數10除8,而以組距5乘之得4.0。
- (e) 以4加於含有中數的組距的最低限度30上,得34,即為中數。

三、四分點差——四分點差一名二十五分差,簡稱為Q,由英文Quartile Deviation 簡譯而得。試將量數由小而大,順序排列,分為相等四段。從最小的量數,或最劣的成績數起,第一段終了處的量數為下四分點,或稱下二十五分點(Q_1);第二段終了處為中數;第三段終了

處爲上四分點,或稱上二十五分點(Q_3)。求四分點差的公式如下:

$$Q = \frac{Q_3 - Q_1}{2}$$

茲更舉例以明之。

1、量數分立而未歸入組距者,其求四分點差的方法如下表:

表 六

求四分點差的方法

例 一	例 二	例 三	例 四
12	12	12	12
13	13	13	13
14	14	14	14
15 Q_1	15 Q_1	15 Q_1	15 Q_1
16	16	16	16
17	17	17	17
18	18	18	18
19	19	19	19
20	20	20	20
21 Q_3	21 Q_3	21	21
22	22 Q_3	22 Q_3	22
23	23	23	23 Q_3
	24	24	24
		25	25
			26
4)12(3 $Q_1=14.5$ $Q_3=20.5$ $\therefore Q = \frac{20.5 - 14.5}{2}$ $= \frac{6}{2} = 3$	4)13(3, 25 $Q_1=14.5$ $Q_3=21.5$ $\therefore Q = \frac{21.5 - 14.5}{2}$ $= \frac{7}{2} = 3.5$	4)14(3.5 $Q_1=15$ $Q_3=22$ $\therefore Q = \frac{22 - 15}{2}$ $= \frac{7}{2} = 3.5$	4)15(3, 75 $Q_1=15$ $Q_3=23$ $\therefore Q = \frac{23 - 15}{2}$ $= \frac{8}{2} = 4$

上表所列的算法共四種,各不相同,茲說明之如下:

(a)若次數的總和可以爲4所除盡(例一),或不可除盡,而其小數點

以下之數不及.5者(例二),則 Q_1 與 Q_3 的地位在兩數的中間。

(b)若次數的總和,不能為4除盡,但其小數點下之數為.5(例三);

或較.5大者(例四),則 Q_1 與 Q_3 的地位,適為整數。

2、若量數已歸入組距,則求四分點差的方法如下表:

表 七

另一種求四分點差的方法

組 距	次 數	算 法
0—5	4	
5—10	5	$Q_1 = 15 + \frac{.5}{9} \times 5$
10—15	7	$= 15 + .28$
15—20	9	$= 15.28$
20—25	10	
25—30	12	
30—35	8	$Q_3 = 35 - \frac{5.5}{8} \times 5$
35—40	5	$= 35 - 3.44$
40—45	3	$= 31.56$
45—50	2	
50—55	1	$\therefore Q = \frac{31.56 - 15.28}{2} = 8.14$
	$\frac{4)66}{16.5}$	

上表的求法,甚為明瞭;因求 Q_1 與 Q_3 的方法,與求中數的方法一樣。既得 Q_1 與 Q_3 ,則求 Q 就很容易了。

四分點差的意義,可述之如下:(a)在一常態分配曲線圖上,一個四分點差包含全數百分之二十五。正負各一個四分點差($\pm Q$)等於全數中間百分之五十。(b)在一常態分配曲線圖上,從中數起,前後展

開各百分之二十五量數，共百分之五十量數，其任何一個量數與平均數相差，不能超過四分點差之外。例如上表所載事實的四分點差為8.14，意即這量數的中間百分之五十量數與平均數相差，無一能超過8.14之外者。

四、均方差——表示離中趨勢的又一法為均方差，亦名標準差。英文原名為 Standard Deviation，簡稱爲 S.D.，亦可用希臘字母 σ 表示之。按均方差為表示離中趨勢最完善的方法。其定義如下：『均方差為差數之方的平均數的平方根。』故欲求均方差，必(a)先求各量數與平均數的差數。(b)方之。(c)求各方的總和。(d)以次數的總和除之。(e)再求其平方根。茲將求均方差的方法詳下：

1、若先得真正平均數，而再求差數者，則公式如下：

$$\sigma = \sqrt{\frac{\sum d^2}{n}}$$

在這個公式內，d 為差數，n 為次數的總和， Σ 為總和的符號， σ 為均方差。其算法見下表：

表 八
求均方差的方法

量 數	差數 (d)	差數方(d ²)
65	-11	121
66	-10	100
68	-8	64
72	-4	16
72	-4	16
74	-2	4
74	-2	4

75	-1	1
77	1	1
79	3	9
81	5	25
82	6	36
83	7	49
84	8	64
86	10	100
76 = 平均數 n = 15		610 = $\sum d^2$
$\therefore \sigma = \sqrt{\frac{610}{15}}$ $= \sqrt{40.66} = 6.4$		

2、若量數已歸入組距者，則可用簡捷法，即求量數與假設平均數的相差。其公式如下：

$$\sigma = \sqrt{\frac{\sum fd^2}{n} - C^2} \times i$$

在這個公式內，C為校正數，i為組距，其餘如前。茲舉例以明之。

表 九

再一種求均方差的方法

組 距	次數	與假設平均數的差	次數乘差數	次數乘差數方
0—5	2	-6	-12	72
5—10	3	-5	-15	75
10—15	4	-4	-16	64
15—20	4	-3	-12	36
20—25	5	-2	-10	20
25—30	6	-1	-6	6
30—35	7	0	-71	
35—40	6	1	6	6
40—45	5	2	10	20
45—50	3	3	9	27
50—55	2	4	8	32
55—60	1	5	5	25

$48 = n$	38	393
$i = 5$	$\frac{-71}{-33}$	
$c = -33 \div 48 = -.68$		
$c^2 = -.68^2 = .4624$		
$\Sigma fd^2 = 393$	$\frac{\Sigma fd^2}{n} = \frac{393}{48} = 8.2$	
$8.2 - .4624 = 7.7376 = \text{均方差}^2$		
$\therefore \text{均方差} = \sqrt{7.7376} = 2.78 \text{ (組距)}$		
$2.78 \times 5 = 13.90 \text{ (單位)}$		

均方差的簡捷求法，已如上述。茲將其特性約略述之。(a)在常態分配曲線圖上，一個均方差佔全數之34.13%；正負各一個均方差，共佔全數中間之68.26%。(b)在常態分配曲線圖上，從平均數起，前後各展開34.13%，共合68.26%量數。其中任何一個量數，與平均數的相差，不能超過均方差價值之外。(c)若分配對稱，或近於對稱，則以6乘均方差，可以包括全數百分之九十九有奇。此法可常用之，以驗均方差之有無錯誤。

五、相關度——相關是一種方法，定奪一組人，或一組學校，或其他團體，對於兩種成績的關係。倘使有絕對的關係，並且那個關係是正的，相關係數(r)為+1.0；要是那個關係是負的，相關係數為-1.0。要是沒有關係，則相關係數為0。舉例如下：

相關度可以解答如下列一類的問題：這個智力測驗或教育測驗是否可靠？這兩種測驗是否測量同樣的能力？通常說長於文學的人不善科學這句話是否確實？數學做得快的人，是否計算也正確？教師的評

	測驗1	測驗2	測驗1	測驗3	測驗1	測驗4	測驗1	測驗5
學 生	分數	分數	分數	分數	分數	分數	分數	分數
A	2	6	2	12	2	6	2	12
B	3	8	3	10	3	10	3	10
C	4	10	4	8	4	8	4	8
D	5	12	5	6	5	12	5	6
	r=+1.0		r=-1.0		r=+.8		r=-.8	

判與測驗的等第，是否符合。學業成績與日後的成功有多大關係？這一類的問題，都可用相關方法求出關係的數量。

求相關的方法很多。最通行的，為1.均方法(Product-moment Method)。公式如下：

$$r = \frac{\sum xy}{N\sigma_x\sigma_y}$$

或列成下列的公式：

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \cdot \sum y^2}}$$

這個公式可用下邊的例子來說明。

(a) 依照各個人的號數，把兩種測驗的分數排列起來。比方第一號的分數在測驗 I 為 2，在測驗 II 為 50，就照表上的寫法。

(b) 求兩種分數的平均數。測驗 I 為 7.0，測驗 II 為 57.5。就理論上說，均方相關祇用平均數；但實際上中數有時也可用。

(c) 求測驗 I 分數和平均數的差數x，測驗 II 分數和平均數的差數

例如測驗 I 的平均數為 7.0, 第 1 號的分數為 2, 比較平均數少 5; 所以在 x 行下寫一負 5。又如第 24 號的測驗 I 分數為 12, 比較平均數多 5; 所以在 x 行下寫一正 5。

表 十

核算均方相關的方法

學生	分 數		平均數的差數		差數自乘		xy
	I	II	Ix	IIy	x ²	y ²	
1	2	50	-5	-7.5	25	56.25	37.5
2	3	50	-4	-7.5	16	56.25	30.0
3	4	50	-3	-7.5	9	56.25	22.5
4	4	80	-3	22.5	9	506.25	-67.5
5	5	20	-2	-35.5	4	1406.25	75.0
6	5	60	-2	2.5	4	6.25	-5.0
7	5	40	-2	-17.5	4	306.25	34.0
8	5	50	-2	-7.5	4	56.25	15.0
9	6	70	-1	12.5	1	156.25	-12.5
10	6	40	-1	-17.5	1	306.25	17.5
11	6	70	-1	12.5	1	156.25	-12.5
12	6	50	-1	-7.5	1	56.25	7.0
13	7	50	0	-7.5	0	56.25	0.0
14	7	70	0	12.5	0	156.25	0.0
15	7	40	0	-17.5	0	306.25	0.0
16	7	70	0	12.5	0	156.25	0.0
17	7	60	0	2.5	0	6.25	0.0
18	8	20	1	-37.5	1	1406.25	-37.5
19	8	60	1	2.5	1	6.25	2.5
20	8	90	1	32.5	1	1056.25	32.5
21	9	80	2	22.5	4	506.25	45.0
22	9	60	2	2.5	4	6.25	5.0
23	10	90	3	32.5	9	1056.25	97.5
24	12	60	5	2.5	25	6.25	12.5
平均數	7	57.5	總數或Σ		124	7850.00	434.0
N=24							-135.0
						299.0	

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}} = \frac{299}{\sqrt{(124)(7850)}} = \frac{299}{986.61} = .303$$

$$P.E. = .67449 \frac{1-r^2}{\sqrt{N}} = .67449 \frac{1-.303^2}{\sqrt{24}} = .125 \therefore r = .303 \pm .125$$

(d) 把 x 和 y 的數目自乘。如 -5 自乘為 25 ， -7.5 自乘為 56.25 等是。

(e) 求 x 和 y 的相乘數。如 $-7.5 \times -5 = 37.5$ ， $-4 \times -7.5 = 30.0$ 等是。

(f) 求 x^2 和 y^2 的總數。 $\Sigma x^2 = 124$ ； $\Sigma y^2 = 7850$ 。

(g) 求 xy 的總數。正的 xy 數 $= 434$ ，負的 xy 數 $= 135$ 。兩數的總數 $\Sigma xy = 299$ 。

(h) 把所得的數目代入公式， $r = .303$ 。

上表內的 P. E. 為 Probable Error 的簡稱；譯為『或然錯誤』，或『機誤』。或然錯誤為相關係數可靠與否的量數。大約人數愈多，則或然錯誤愈小；人數愈少，則或然錯誤愈大。故求相關係數，必求其或然錯誤。上表內， $r = .303 \pm .125$ 的意義，可詳述如下。 x 與 y 的相關度為 $.303$ ，然這 $.303$ 之係數，乃由 24 個學生的測量而得。假使將數千萬應有盡有的同類學生的 x 與 y 而測量之，然後求其相關度，則其相關係數，必為最後的真確數目，而必非 $.303$ 無疑。故 $.303$ 之係數，難免因機遇而得的錯誤。今其或然錯誤為 $\pm .125$ ，意即今所得的相關係數 $.303$ 與真確相關係數的相差，不至超過 $+.125$ 或 $-.125$ 。其機遇為一與一之比。換言之，真正 x 與 y 相關的係數，有一半機會，不出 $.303 + .125$ 與 $.303 - .125$ 之外，或不出 $.428$ 與 $.178$ 之外。如是，今所得係數可靠的程度可知了。

2、均方相關比較是很可靠，不過此外尚有等級相關(Rank Correlation)的方法。此法非常利便；可用斯比門的公式(Spearman "Footrule" Formula)。

$$R=1-\frac{6\sum G}{N^2-1}$$

用上列公式得到R以後，尚須化成r。潘阿生 (Pearson) 曾有一個對照表(表十二)。舉例如下：

(a) 先把各個人測驗I的分數，列成比較的等第。如2分列第一或1；3分列2；4分有兩個，平分3, 4等第，所以各列3. 5；5分有四個，把5, 6, 7, 8四個等第平均起來，各得6. 5；餘類推。測驗II的分數也列成等第，最小的分數列在最前。如20分的兩個，平均1, 2兩等第，各得1. 5；40分的有三個，平均3, 4, 5等第，各得4；餘類推。倘使以最大的分數列在第一也可，不過兩種測驗的等第必須要一致。

表 十 一

核算等級相關的方法

學生	分 數		等 第		超過第一次等第 數 Gain (G)
	I	II	I	II	
1	2	50	1	8.5	7.5
2	3	50	2	8.5	6.5
3	4	50	3.5	8.5	5.0
4	4	80	3.5	21.5	18.0
5	5	20	6.5	1.5	
6	5	60	6.5	14.0	7.5

7	5	40	6.5	4.0	
8	5	50	6.5	8.5	2.0
9	6	70	10.5	18.5	8.0
10	6	40	10.5	4.0	
11	6	70	10.5	18.5	8.0
12	6	50	10.5	8.5	
13	7	50	15.0	8.5	
14	7	70	15.0	18.5	3.5
15	7	40	15.0	4.0	
16	7	70	15.0	18.5	3.5
17	7	60	15.0	14.0	
18	8	20	19.0	1.5	
19	8	60	19.0	14.0	
20	8	90	19.0	23.5	4.5
21	9	80	21.5	21.5	
22	9	60	21.5	14.0	
23	10	90	23.0	23.5	.5
24	12	60	24.0	14.0	
N=24				$\Sigma G=74.5$	
$R=1-\frac{6 \Sigma G}{N^2-1}=1-\frac{6(74.5)}{(24)^2-1}=.224$					
參照表十二, $r=.37$					

(b)核算超過第一次的等第數。如 $8.5-1=7.5$; $8.5-2=6.5$; 餘類推。

(c)求超過第一次等第數的總數。 $\Sigma G=74.5$ 。

(d)代入公式, $R=.224$ 。參照表十二,化成 r ; $r=.37$ 。

表 十 二

化R爲r的對數表

$$r = 2 \cos \frac{\pi}{3} (1-R) - 1, \quad R = 1 - \frac{6 \Sigma G}{N^2 - 1}$$

R	r	R	r	R	r	R	r
.00	.000	.26	.429	.51	.742	.76	.937
.01	.018	.27	.444	.52	.753	.77	.942
.02	.036	.28	.458	.53	.763	.78	.947
.03	.054	.29	.472	.54	.772	.79	.952
.04	.071	.30	.486	.55	.782	.80	.956
.05	.089	.31	.500	.56	.791	.81	.961
.06	.107	.32	.514	.57	.801	.82	.965
.07	.124	.33	.528	.58	.812	.83	.968
.08	.141	.34	.541	.59	.818	.84	.972
.09	.158	.35	.554	.60	.827	.85	.975
.10	.176	.36	.567	.61	.836	.86	.979
.11	.192	.37	.580	.62	.844	.87	.981
.12	.209	.38	.593	.63	.852	.88	.984
.13	.226	.39	.606	.64	.860	.89	.987
.14	.242	.40	.618	.65	.867	.90	.989
.15	.259	.41	.630	.66	.875	.91	.991
.16	.275	.42	.642	.67	.882	.92	.993
.17	.291	.43	.654	.68	.889	.93	.996
.18	.307	.44	.666	.69	.896	.94	.996
.19	.323	.45	.677	.70	.902	.95	.997
.20	.338	.46	.689	.71	.908	.96	.998
.21	.354	.47	.700	.72	.915	.97	.999
.22	.369	.48	.711	.73	.921	.98	.9996
.23	.384	.49	.721	.74	.926	.99	.9999
.24	.399	.50	.732	.75	.932	1.00	1.0000
.25	.414						

上邊得到的r爲.30或.37,究竟這個相關度算是高的,還是低的:據各人的經驗,覺得

從0至±.4的相關度算是低的;

從±.4至±.7的相關度算是很有關係;

從 $\pm .7$ 至 ± 1.0 的相關度算是高的。

朱君毅，寥世承，杜佐周。

附 錄 二

求 B C F 分數的方法

一、B量表的編制法——T量表的編制法，已詳於本書第十章，茲不贅述。T分數表示一個人的總能力；年齡愈大，能力愈增進。但B分數則不然，乃並不隨年齡而增進的。使某生年齡為10歲，B為50；及至12歲或13歲，仍為50，以此終其一生。

假使能得各年齡兒童代表之數，即各得500個而成普通分配曲線狀態者，則B量表的編制法與編制T量表的方法一樣。可是吾人實際執行測驗，因欲得代表歲(12—13)的兒童數，其他年齡的兒童往往不能達到代表的數目。13歲以後，聰明者已進中學，所留在小學者為一部份愚笨的兒童。反之，12歲以前，其留在高小者為一部份聰明的兒童；其他部份尚未進入高小。所以不得不另求他法編制B量表。茲分述如下：

1、作一各年齡分數次數分配表，並書12歲的T量表於右方。(下表的T分數是根據於11個12歲兒童的成績求得的。)

2、計算各年齡的總數。如(10—11)歲為6，(11—12)歲為9，(13—14)歲為8。

表 一

各年齡分數次數分配表

分數 \ 年齡	10—11	11—12	13—14	T_{12}
0	1	1		33
1	0	1		39
2	2	1		41
3	2	2	1	44
4	1	2	2	50
5		1	2	56
6		1	2	61
7			1	67
學生總數	6	9	8	
$6 \div 2 = 3$ $1 + 2 + (2 \div 2) = 4$ $\frac{4}{11} \times 100 = 36.3$ $T_{10} = 54.5$ $54.5 - 41 = 13.5 \text{B.C.}$		$9 \div 2 = 4.5$ $1 + 1 + 2 + (2 \div 2) = 5$ $\frac{5}{11} \times 100 = 45.4$ $T_{11} = 51$ $51 - 44 = 7 \text{B.C.}$		$8 \div 2 = 4$ $1 + 2 + (2 \div 2) + (11 - 8) = 7$ $\frac{7}{11} \times 100 = 63.6$ $T_{13} = 46.5$ $46.5 - 56 = -9.5 \text{B.C.}$

3、算出各年齡的半數。

4、自(10—11)歲, (11—12) 歲的下方將各數加上直至超過半數的一數為止, 然後將此數折半加上。如(10—11) 歲行, $1 + 2 + (2 \div 2) = 4$; (11—12) 歲行, $1 + 1 + 2 + (2 \div 2) = 5$ 。此即得某分數超過數加 $\frac{1}{2}$ 達到數之意。如(10—11) 歲得2分者2人, 超過數加 $\frac{1}{2}$ 達到數為4; 同理(11—12) 歲得3分者2人, 超過數加 $\frac{1}{2}$ 達到數為5。

5、然後將所得數用(12—13) 歲兒童的數目(就是11個)除之, 算出百分比。

6、化百分比為T值,得 $T_{10}=54.5$; $T_{11}=51$ 。

7、(12-13)歲得2分的T值為41,與3分的T值為44。自54.5減41得13.5,即10歲的B校正數(或簡稱B.C.);自51減44得7,即11歲的B校正數。

8、在12歲下的B分數算法,既如上述。至於在13歲以上的B分數算法稍有不同。前者失去愚笨的兒童,所失去者在全距的上端;後者失去聰明的兒童,所失去者在全距的下端。此失去的兒童必須加入乃可。如表一,失去的為(11-8),加入得7。然後以(12-13)歲的總數11除之,得百分比63.6;化為T值,得46.5,即13歲得5分的T值。而12歲得5分的T值為56; $46.5-56=-9.5$,為負數。比12歲半小的兒童,將所得T值加校正數,故為正;比12歲半大的兒童,將所得T值減校正數,故為負。

9、既得各年齡的B校正數,然後用插進法求每月的B校正數。若試驗之月為六月,則可如表二。

表 二

年齡與B校正數對照表

年 齡	B 校 正 數
10—6	13
10—8	12
10—10	11
11—0	10
11—2	9
11—4	8

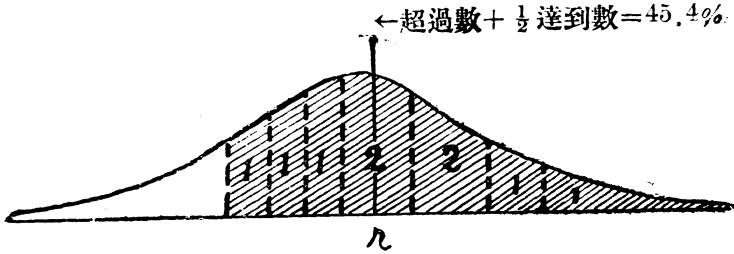
11—6	7
11—8	6
11—10	5
12—0	3
12—2	2
12—4	1
12—6	0.
12—8	—2
12—10	—3
13—0	—5
13—2	—6
13—4	—8
13—6	—9

表一 (11—12) 歲的學生, 僅得該團聰明的部分, 上面已經說過。觀圖形一, 更可明白。被測驗者, 僅占曲線的右部分; 左部空白處就是表示愚笨者尚未進入舉行測驗的最低年級。如表一, (11—12) 歲兒童得 3 分者有 2 人, 故達到數為 2; 2 人以上尚有 4 人為超過數。超過數加 $\frac{1}{2}$ 達到數等於 5, 就是 r 點以上的人數。化為百分比 45.4, 即為超過數加 $\frac{1}{2}$ 達到數的百分比。化為 T 值 51, 即 r 點所在的 S.D. 值。但此 T 值以 (11—12) 歲為標準; 至以 (12—13) 為標準, 則得 3 分者的 T 值為 44。兩者相差為 7, 這就是 (11—12) 歲的 B 校正數。

(13—14) 歲兒童的優秀者亦不入測驗的範圍, 因已出舉行測驗的最高級; 故所留者為曲線的左部分。圖形二, 右部的空白處就是表明所失優秀的兒童, 其數為 (11—8)。如表一, (13—14) 歲之得 5 分者 2 人, 即

圖 形 一

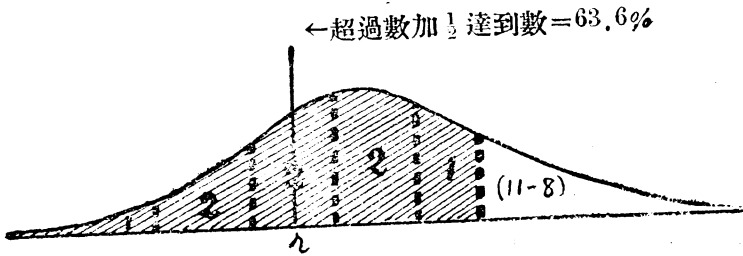
[依表一(11-12)歲的次數分配]



為達到數;其上尚有 $1+2+(11-8)=6$,為超過數。 $\frac{1}{2}$ 達到數加超過數 $=1+6=7$,即 r 點以上的人數。化為百分比 63.6 ,即超過數加 $\frac{1}{2}$ 達到數的百分比;化為 T 值 46.5 ,即 r 點所在的 $S.D.$ 值。(12-13)歲得5分的 T 值為 56 ,相差 -9.5 ,這就是(13-14)歲的 B 校正數。

圖 形 二

[依表一(13-14)歲的次數分配]

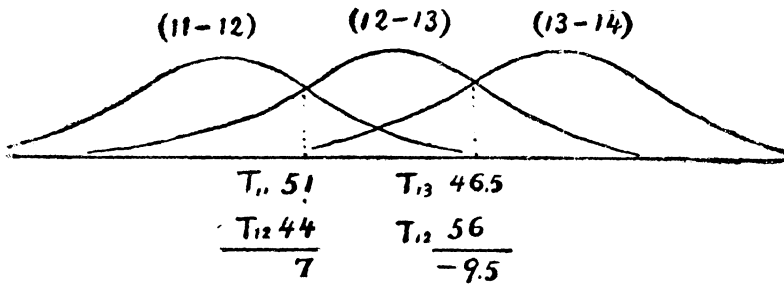


又設三曲線如圖形三:中線代表(12-13)歲的次數分配,左線代表(11-12)歲的次數分配,右線代表(13-14)歲的次數分配。設(11-12)歲與(12-13)歲曲線相交之處為同得3分的兒童;(12-13)歲與

(13—14)歲曲線相交之處為同得5分的兒童。同得3分者以11歲為標準，T值為51；以12歲為標準，T值為44；相差為7，即兩曲線相差之度，亦即11歲與12歲聰明相差之度。同理，12歲與13歲相差為-9.5；其為負數者，因以12歲為標準的。

圖形三

三曲線相關圖



二、C量表的編制法——C量表亦由T量表而來，為分班之用。其編制方法如下：

- 1、作各年級分數次數分配表。如表三。
- 2、將次數乘T值。
- 3、將各級T值平均，而得各該級的標準分數：第五年級為43.7，第六年級為47.9，第七年級為55.8。
- 4、假定秋季始業，而測驗在六月末，則43.7為五年級末的程度，即

表三

年級次數分配表

年 級	五		六		七		T ₁₂
	分 數	次 數	T 值	次 數	T 值	次 數	
0	2	66	1	33			33
1	1	39	1	39			39
2	2	82			1	41	41
3	2	88	4	176	1	44	44
4	1	50	5	250	2	100	50
5	2	112	2	112	1	56	56
6			1	61	3	183	61
7					2	134	67
總 數	10	437	14	671	10	558	
平 均		43.7		47.9		55.8	

爲六年級初的程度。同理,47.9爲七年級初的程度,55.8爲八年級初的程度。然後將各級相差的數分爲每月的數,如下表:

表 四
年級 T 值對照表

T 值	年 級 (G)
43.7	6.0
44.1	6.1
44.5	6.2
44.9	6.3
45.3	6.4
45.8	6.5
46.2	6.6

46.7	6.7
47.1	6.8
47.5	6.9
47.9	7.0
48.7	7.1
49.6	7.2
50.3	7.3
51.0	7.4
51.9	7.5
52.7	7.6
53.4	7.7
54.2	7.8
55.0	7.9
55.8	8.0

5、G值表示由T值而得的年級地位，為絕對之數。若欲表示某生在某級優劣的地位，則觀試驗日期距開校日的月數而定自G加或減的C校正數。看下表，距開校1月，C校正數為0.4，離2月為0.3，餘類推。五年級C的標準為5.5，六年級C的標準為6.5，其餘類推。若某五年級生的C為5.5，則為中等生；大於5.5，為優等生；低於5.5，為劣等生。無論何時測驗，只將C校正數自G加減之，即得該生該級優劣的地位。

表 五

C校正數與距開校月對照表

月 終	1	2	3	4	5	6	7	8	9	10
C校正數	.4	.3	.2	.1	0	-.1	-.2	-.3	-.4	-.5

三 努力分數的計算——上述 BC 兩法及前述 T 法皆可用以編制各學科的量表。如算術的量表，則有 T_a, B_a, C_a ；讀法的量表，則有 T_r, B_r, C_r 。此外又可用以編制智力的量表，如 T_i, B_i, C_i 。努力分數 (F) 者，即各科的 T 值減智力的 T 值。以一學科計算，即為某科的努力分數，如 $F_a = T_a - T_i$ ，為對於算術的努力分數； $F_r = T_r - T_i$ ，為對於讀法的努力分數。若各科合計之，則為教育的努力分數，如 $F_e = T_e - T_i$ 。以一人計，為某人對於某科或教育的努力分數；以全級計，為某級對於某科或教育的努力分數。若用以比較各校，則可計算各校對於某科或教育的努力分數。若智力的 T 值小於教育的 T 值，即 $T_e > T_i$ ，則努力分數為正，表示學生對於功課用心；若 $T_e = T_i$ ，則努力分數為零，表示學生用力平常；若 $T_e < T_i$ ，則努力分數為負，表示學生對於功課怠惰而不用心。若以學校經費與努力分數合算，則知每一元努力的分量。譬如先計算全校每生的教育努力分數，又計算每生平均用費，然後以後者除前者，即可得每元的努力分數。各校相較，則知何校最為經濟或最不經濟。由是而求其原因，以求改良的方法。用努力分數以計算各校的效率，可得如下的利益。

- 1、可減少學校經費於最低的限度，而不損害教學的效率。
- 2、足以鼓勵教師的努力，增進其奮勇的精神。
- 3、此種計算方法自然示高效率的學校應得多量的經費。
- 4、可以將學生對於學科的努力為教師薪俸高低的分辨。
- 5、可使學校減少無益的設備與裝飾。
- 6、使學校不保留畢業生，以增加表面的成績。

- 7、學校以地位的關係而多聰明的學生者，不足以爲效率的表示。
- 8、使學校延長授課時日，減少缺課人數。
- 9、使學校減少留年齡過大的學生於低年級的傾向。

以上九端，皆用努力分數計算學校效率的利益；至用以診斷一市或一縣的教育，則可知與努力相關的各種情形如下：

1、以努力分數與教師薪俸求相關度，則可知薪俸影響於努力分數至如何程度。

2、以努力分數與教師記分(Teacher's score)求相關度，則可知教師之影響於努力分數至如何程度。

3、以努力分數與學校衛生求相關度。

4、以努力分數與每生用費(平均)求相關度。

5、比較各種相關度，則知何者爲增加努力最重要的分子。

關於TBCF分數，寥世承先生有很好一段解釋，茲照錄之。

『T是什麼？T爲量學生某種特性的單位。T分數表示各個人對於某種特性的總數量，或稱爲總能力分(Total ability)。所以要將測驗分數化成T分數的緣故，至少有兩種原因：1、除掉題目難易的影響；2、有一普遍的標準，可以比較各個人的總能力。

B(Brightness)是一種年齡分數，各年齡的B，猶之12歲的T(指以12歲爲根據的T量表而言)。年幼的學生，T分數大概都不甚高；但是B分數，却可以很高。年高的學生，T分數大概都不甚低；但是B分數，却可以很低。因爲個人的T分數是逐年逐月加大的；但是B分數變更很少。所以看了一個人的B，就可以知道他在同年齡學生中所處的地位。

了。

C (Classification) 是一種年級分數。有了 C 分數對照表，可以知道一個人的年級地位，供學校行政人員分班時的參考；並可與全國同等學校比較程度的高下。

F (Effort) 是一種努力分數，比較各個人或全班智力與學力的上下，可藉以度量教育效率』。

錢希乃，杜佐周。

附 錄 三

比較速率及真確度成績的更正公式

任何人類活動的成績，時間必是一種很重要的因素。若是別種條件一樣，時間愈長，工作必愈多；時間愈短，工作必愈少；此誠淺而易見的。假設一個兒童，八分鐘做完十二個加法算題，則若不受別種影響，十分鐘必可做完十五個算題。今若不問兩次測驗時間的長短，而僅比較十二個算題與十五個算題的成績，則所得結論，必不真確。因為測驗的時間既不相等，我們就無從得知十五個算題的成績是比十二個算題的成績好，或次，或相等。

他若測驗的情形或目的不同，成績亦有異。倘忽略此層，而求結論，亦必不能公允。葛第斯及桑戴克測量一組成人的加法能力；前後四次，情形各不相同。第一次，在普通情形下執行測驗。第二次，注重真確度 (Accuracy)；先告訴被試者，對於每個算題，至少須算兩回。如兩個答數不符，則當重算第三回或第四回，至自己相信已經真確為止。

第三及第四次，則注重速率 (Speed or rate)。一次每題只有二十七秒鐘，另一次每題只有二十二秒鐘。每次測驗有三種成績：1、試算的題數；2、算對的題數；3、真確度，就是試算題中算對的百分比。六個被試者的成績，表列如下：

表 一
計算加法算題的成績

測驗	測驗情形	A	B	C
第一試	普通情形下	4—3	15—9	11—10
第二試	注重真確度	4—4	4—4	4—4
第三試	注重速率17.8	0%	25%	5%
測驗	測驗情形	D	E	F
第一試	普通情形下	16—11	16—14	24—8
第二試	注重真確度	6—6	6—6	6—6
第三試	注重速率17.8	45%	65%	55%

上表讀法如下：被試 A，第一試試算四個題，算對三個；第二試試算四個題，算對四個；第三試試算17.8題，但無算對者，故真確度為零。其餘類推。

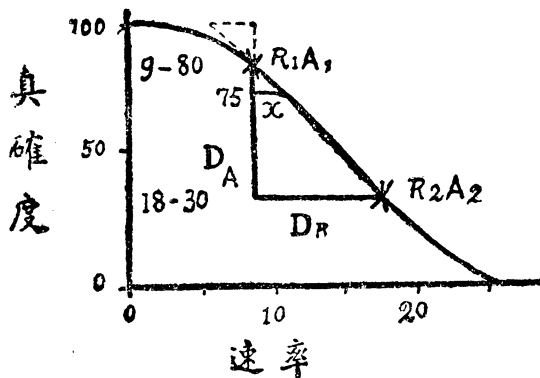
看上表，被試者的第一試成績，彼此相差甚遠。至於第二試成績，彼此幾乎完全一樣。可是他們並未改變其能力，不過更易其工作的方法而已。被試 A 第一試工作，非常過細；每個算題，必行校對一回。至於被試 B 及 F，則第一試時，工作很快，故不十分真確。但 B 及 F 均可採用

A 的方法，試看第二試的成績，就可知道。A 及 B 的成績，完全一樣；F 則不過多做兩個算題。就第一試成績而論，A, B, C 三個被試者的能力，絕不相同；但就第二試成績而論，則他們的能力，完全一樣。由此可知單獨一次測驗的成績，不能決定個人真正的能力。然多次測驗的成績，前後既不能一致，則又如何能使可以互相比較？何種成績，究竟表示真正的能力？欲解決這兩個問題，工作的速率及真確度，必須聯絡一起說話。其法先定有標準的真確度(Standard accuracy)，而後比較速率；或先定有標準的速率(Standard rate)，而後比較真確度。茲仍以比較加法能力為例，葛桑二氏取 75% 為標準的真確度。(因為第八年級受加法測驗的真確度中數是 75%)。然後應用他們的加法測驗的更正公式 (Correction Formula For addition tests)，計算達到標準真

圖 形 一

$$X : D_R :: (A_1 - 75) : D_A \quad R_{75} = R_1 + \frac{D_R(A_1 - 75)}{D_A}$$

$$X = \frac{D_R(A_1 - 75)}{D_A} \quad R_{75} = 9 + \frac{9(80 - 80)}{80 - 30}$$



$$R_{75} = 9 + \frac{45}{50}$$

$$R_{75} = 9 + .9$$

$$R_{75} = 9.9$$

確度的應有速率。如是，就可互相比較。這種更正公式，乃是根據人類工作曲線 (Performance curve) 的普通形式而產生；就是根據時間愈長，真確度愈高；時間愈短，真確度愈低的通例。

上圖曲線上的十字 \times ，表示前後兩次不同情形下測驗的加法成績。上邊一個，代表八分鐘試做9個算題，真確度為80%。(用 R_1 ， A_1 表示之。)下邊一個，代表八分鐘試做18個算題，真確度為30%。(用 R_2 ， A_2 表示之。)從上邊這個十字 \times 起，畫一條直線，與真確度軸線相平行；又從下邊那個十字 \times 起，畫一條直線，與速率軸線相平行。如是，這兩直線相交，可與能力線 (Ability line) 成一個直角三角形。這直角的兩邊的長度，是已經知道的。底線為18—9 ($R_2 - R_1$)，就是兩次測驗的速率相差 (D_R)；縱線為80—30 ($A_1 - A_2$)，就是兩次測驗的真確度相差 (D_A)。若現在從真確度線上的標準真確度點，作一垂直線，則可成一個小的直角三角形，適與第一個三角形相類似。這小三角形的一邊為80—75 ($A_1 - \text{Standard accuracy}$ ，或 $A_1 - A_s$)；另一邊為 X ，就是應加速率的未知數 (The unknown increment of rate)。倘若以100%為標準真確度，則這小三角形當在大三角形以外。(上圖虛線的三角形是。)如是，應加速率的未知數，當為負數 (就是減速率)。由上可知達到標準真確度的速率，是原來速率加速率應加數 (Initial rate + rate increment)。我們若有一個被試者受兩次測驗的速率及真確度成績，且其真確度在15%至85%的範圍以內，則任取何度為標準的真確度，均可照下面一個公式計算達到標準真確度的速率。

$$R_s = R_1 + \frac{D_R(A_1 - A_s)}{D_A} \quad (\text{實例見上圖})$$

在上面公式中, R_s = Standard rate (達到標準真確度的速率, 或簡稱標準速率)

R_1 = lowest observed rate (所得最低的速率)

D_R = difference in the two observed rates (所得兩次速率的相差)

A_1 = accuracy at lowest observed rate (所得最低速率的真確度)

A_s = accuracy adopted as standard (所採取的標準真確度)

D_A = difference of observed accuracies (所得真確度的相差)

在上圖, 有一個被試者的成績, 是完全照這公式計算的。找到他若照75%的標準真確度工作, 他的速率當為9.9題 (根據原來八分鐘的時間); 若是取100%為標準真確度, 則速率當僅為5.4題。由此我們可以推知這個被試者的算題是5.4時, 則真確度為100%; 算題是9.0時, 則真確度為80%; 算題是9.9時, 則真確度為75%; 算題是18.0時, 則真確度為30%。其實他的能力, 都是一樣。今若要將這個被試者的能力與他人的能力比較, 欲求比較公允而且真確, 則必當根據標準的速率, 而比較真確度的分數; 或根據標準的真確度, 而比較速率的分數。

葛第斯, 桑戴克, 及杜佐周

核算實足年齡的方法

先求出被試者的陰歷年齡和生日，從陰歷年齡中減去一歲，即得應有年齡；復從測驗那天的陰歷某月某日減去陰歷生月生日。減後所得的數目若是正的，即加於應有年齡上；若是負的，即從應有年齡中減去此數。例如有一個14歲的學生是陰歷4月生的，試期是在8月。他的應有年齡，就是從14減去1得13歲，再從8月減去4月得4月，這4是正數，加於13歲上，即得實足年齡13歲4個月。又有一個14歲的學生是在陰歷11月生的，試期亦在8月。從14減去1得13，從8減去11得-3，這3是負的，應從13中減去此數；所以他的實足年齡是12歲9個月。

附 錄 五

中英人名對照表

Ayres 愛禮斯	Mccomas 梅克哥馬絲
Ballou 包老	Mouroe 孟祿
Binet 皮內	Nassau 納索
Bridges 柏里傑士	Norsworthy 勞施威塞
Buckingham 阜京海姆	Pearson 潘阿生
Catell 克德爾	Pestalozzi 裴絲德拉齊
Chapman 漆珀門	Pinter 賓德納
Courtis 葛第斯	Plato 柏拉圖
Coy 柯愛	Pressey 柏來雪
Crathorne 凱雷爽	Rice 萊士

Dickson 笛根生	Robinson 雷賓孫
Dollinger 陶林傑	Roger 萊傑
Eliot 伊里	Rosamoff 羅索奴富
Franzen 弗萊辰	Ruml 盧爾
Fulleton 福來登	Seashore 西宣華
Gall 高爾	Simon 西門
Gray 葛萊	Solomon 薩拉門
Green 葛林	Spearman 斯比門
Haggerty 海格底	Starch 絲大齊
Hillegas 海利葛斯	Stone 施東
Hollingworth 霍林威士	Terman 推孟
Jones 鍾斯	Thorndike 桑戴克
Judd 傑得	Trabue 屈來阜
Kelly 開萊	Weber 惠柏
Kent 凱得	Wells 威爾士
Kirby 寇備	Woodworth 吳偉士
McCall 麥柯爾	Woody 胡德

附 錄 六

參 考 書 報

- Buckner, C. A. : Educational Diagnosis of Individual Pupils;
Teachers College, Columbia University, New York, 1919.

-
- Bloomfield, M: *Youth, School, and Vocation*; HoughtonMifflin Co., Boston, 1915.
- Burgess, M. A. : *Measurement of Silent Reading*; Russell Sage Foundtion, New York, 1920.
- Chapman, J. C. : *Trade Tests*, Henry Holt Co., New York, 1921.
- Chapman, J. C. and Rush, G. P. : *Scientific Measurement of Classroom Products*; Silver, Burdett Co., Boston 1917.
- Courtis, S. A. : *The Gary Public Schools : Measurement of Classroom Products*; General Educational Board, New York, 1919.
- Coarts, S. A. and Thorndike, L. T. : *Correction Formulae for Addition Tests*; Teachers College Record, Columbia University, New York, 1920 (first munder).
- Davis, J. B. : *Vocational and Moral Guidance*; Ginn Co., New York, 1914.
- Fretwell, E. K. : *A Study in Educational Diagnosis*; Teachers College, Columbia University, New York, 1919.
- Hillegas, M. B. : *Scale for the Measurement of Quality in English Composition by Young People*; Teachers College, Columbia University, New York, 1912.
- Holligworth, H. L. and L. S. : *Vocational Psychology*; D. Appleton Co., New York, 1916.

-
- Kelly, T. L.** : Educational Guidance: An Experimental Study in the Analysis and Predication of Adility of High School Pupils; Teachers College, Columbia University, New York, 1914.
- Kruse, P.** : The Overlapping of Attainments in Certain Grades, Teachers College, Columbia University, New York, 1918.
- Moore, W. S.** : Measuring the Results of Teaching; Houghton Mifflin Co., Boston, 1918.
- Monroe W. S., De Voss, J. C. and Kelly F. J.** : Educational Tests and Measurements, Houghton Mifflin Co., Boston, 1917.
- Pintner, R. and Paterson, D.** : A Scale of Performance Tests; Warwick and York Co., Baltimore, 1917.
- Rogers, A. L.** : Experimental Tests of Mathematical Ability and Their Prognostic Value; Teachers College, Columbia University, New York, 1918.
- Starb, D.** : Educational Measurements; Macmillan Co., New York, 1917.
- Terman, L. M.** : The Measurement of Intelligence; Houghton Mifflin Co., Boston, 1916.
- Thorndike, E. L.** : Introduction to the Theory of Mental and Social Measurements; Teachers College, Columbia Unive-

- rsity, New York, 1913.
- Trabue, M. R. : Completion Test Language Scales; Teachers College, Columbia University, New York, 1915.
- Van Wageningen, M. J. : Historical Information and Judgment of Elementary School Pupils; Teachers College, Columbia University, New York, 1919.
- Whipple, G. M. : Classes for Gifted Children; Public School Publishing Co., Bloomington, Ill, 1919.
- Wilson, G. M. and Kremer, J. H. : How to Measure; Mac Millan Co., New York, 1921.
- Woody, C. : Measurements of Some Achievements in Arithmetic; Teachers College, Columbia University, New York, 1916.
- Yerkes, R. M. and Others : A Point Scale for Measuring Mental Ability; Warwick and York Co., Baltimore, 1915.
- Yoakum, C. S. and Yerkes, R. M. : Army Mental Tests; Henry Holt Co., New York, 1920.
- 朱君毅：教育統計學；商務印書館，1924。
- 寥世承：教育心理學，第三十六課至第四十課；中華書局，1924。
- 錢希乃：麥柯爾測驗編造的TBCF制，教育雜誌第十五卷，第九號，商務印書館，1923。

美國勞工狀況 邵元冲著

著者要將美國過去和現在勞工運動的經驗及現在勞工狀況，指示我國人以勞工運動的組織方法，並指出應該免除的歧途。因此實地調查兩年，走遍了七八十處地方，得到的材料又加上大半年的整理，方纔成書。對於他所負的使命，真算達到。是我國空前的著作，是勞工界的福音，是勞動運動者的嚮導。

全一册 定價六角

廣 (11)

版權所有翻印必究

中華民國十六年八月初版

麥柯爾教育測量法撮要 (全一册)

每册定價大洋八角

外埠酌加郵費匯費

原著者

William A. Mc Call

編譯者

杜佐周

印刷者

上海塘山路三十一號
民智印刷所

發行者

上海河南路九十九十一號
民智書局

分發行處

漢口湖北街漢潤里十一號
武昌省政府前
民智書局分店

分售處

杭州 永 漢 北 坊
廣州 壽 安 坊
各省各大書坊

總發行所

上海河南路中市
民智書局
九十五至九十一號

簿記與會計

上册

國立暨南商科大學會計學會編纂

上册定價 平裝一元二角
精裝一元五角

- (一) 本書目的乃供商業學校、中學校、師範學校暨工商界作課本或參攷之用。
- (二) 分上下兩册，上册為普通商業簿記與會計，以商業組織種類而分：獨資商店簿記及會計，合夥營業會計，股份有限公司會計等；掌管股份公司會計，合併公司會計等。
- (三) 本書為便利學者計新舊式並列。
- (四) 本書編纂由淺入深，由簡而繁，學者不致發生隔閡。
- (五) 本書以實用為主，關於理論方法，概從略。
- (六) 本書討論之重要問題，皆列入細目，極便查攷。

廣 (47)

記敘文作法講義

孫假工著

定價七角

教學初級中學國文，極感困難，採用是書，能將這種困難打破。編者曾經在東大附中實地施用。全書包含寫景，敘事，遊記，三部分。關於緊要的地方，摘錄作多名家作品做例子，使讀者容易明白。為初中一二年級最適合的國文教材。



80760707

廣 (86)

資本論解說

德國考茨基著

戴季陶譯

胡漢民補譯

精平定價一元三角

馬克斯資本論是有名難讀底書，但是我們讀了考茨基所著的解說，就會識得馬克斯經濟學底大旨，不會覺得難讀了。因為資本論底難讀，不在乎馬克斯用語的艱深特別，而在讀者還未有相當預備底工夫。考茨基是幫我們預備好的。他將馬克斯研究底方法，組織底體系，處理底問題，理論底要點，都給我們解說清楚。所以由資本論出版到現在，歐美各國所有從事於解釋的書，還未有比他好的。戴季陶先生在前幾年譯出大半，最近又由胡漢民先生補譯完全。中間還經過朱執信先生和李漢俊先生校訂，可知此書底價值。

各國革命史略

邵元冲講演

定價三角

本書共分七講。講美國的獨立，法國的大革命，德國革命運動，俄國的革命及革命後之建設事業，以及現代各國革命之趨勢，如土耳其，印度，波斯，愛爾蘭，朝鮮等均經講及。敘述簡要，議論明斷，極

英士先生革命小史

邵元冲編

定價四角

辛亥革命光復上海，陳英士先生自民國五年為袁黨所害後，迄今已有十載。陳先生一生完全盡力於革命運動，與中華民國有甚深的關係，故其歷史最值青年們誦讀。現經邵元冲先生編成，全書十章，對於英士先生一生重要的事業和其奮鬥的精神，均有簡明的敘述。

