

中央政治學校
附設

新疆學校圖書館

分類號 217.1911

7538

登錄號 0436

書叢學大

驗測育教

大會書委會員

員 委

丁燮林君 李聖五君 竺可楨君 唐 鈺君 傅斯年君
王世杰君 李權時君 胡 適君 郭任遠君 傅運森君
王雲五君 余青松君 胡庶華君 陶孟和君 鄒 魯君
任鴻雋君 何炳松君 姜立夫君 陳裕光君 鄭貞文君
朱經農君 辛樹幟君 翁之龍君 曹惠羣君 鄭振鐸君
朱家驊君 吳澤霖君 翁文灝君 張伯苓君 劉秉麟君
李四光君 吳經熊君 馬君武君 梅貽琦君 劉湛恩君
李建勛君 周 仁君 馬寅初君 程天放君 黎照寰君
李書華君 周昌壽君 孫貴定君 程演生君 蔡元培君
李書田君 乘 志君 徐誦明君 馮友蘭君 蔣夢麟君
顧頤剛君 羅家倫君 顏福慶君 顏任光君 歐元懷君

mt
61449
18

大 學 叢 書
教 育 測 驗
陳 選 善 著

原 大 夏 大 學 叢 書
商 務 印 書 館 發 行



3 1795 9427 4

自序

標準測驗已成爲教學上一種重要的工具。要使測驗在教學上發生最大的效果，第一，每個教師須有運用測驗的智識及技能，然後測驗才能與兒童的學習發生切實關係；第二，測驗須有診斷的價值，然後各個兒童的困難缺陷才能明白顯示出來，教學才有所依據。這兩點可說是本書的中心概念。

本書之編，取材於前人之著作頗多，不及一一誌謝，良用歉疚。吾妻德音在盛暑之下將原稿抄寫一遍，費去時光不少，我也應該向她謝謝。

民國二十二年十月十日陳選善序於上海。

目錄

第一編 測驗之一般的理論

第一章 測驗在教育上的地位

一 測驗與教育的關係

二 教育測量的可能

三 教育測量的困難

四 教育測量的需要

第二章 測量的要素

一 單位

二 參照點

三 量表

四 測驗

目錄

三

第三章 測驗應具的條件.....一四

一 正確性.....一四

二 可靠性.....一七

三 鑒別力.....一九

四 常模.....二〇

五 施行及校閱的便利.....二一

六 診斷的價值.....二二

七 覆份.....二三

八 經濟.....二三

第四章 智力測驗發展的歷史.....二六

一 初期的嘗試.....二六

二 比納的工作及其影響.....三一

三 比納智力量表成功以後的發展.....三四

第五章 教育測驗發展的歷史.....四三

一 美國體育測量運動的發展……………三四

二 我國測量運動發展的情形……………四九

第六章 測驗的種類……………五三

一 以測量對象論……………五三

二 以試驗方法論……………五六

三 以內容論……………五六

四 以功用論……………五七

第二編 各科測驗與新式考試……………六一

第七章 智力測驗……………六一

一 我國現有的智力測驗……………六一

二 一個智力及智力測驗的界說……………七〇

第八章 國文測驗……………七六

一 默讀測驗……………七六

二 默字測驗.....	七九
三 識字測驗.....	八四
四 作法測驗.....	八五
五 文法測驗.....	九八
六 書法測驗.....	一〇一

第九章 算術測驗

算術測驗的新趨向

一 算術測驗的新趨向.....	一〇八
二 四則測驗.....	一一一
三 應用題測驗.....	一一四
四 練習測驗.....	一一六
五 混合數學測驗.....	一二〇

第十章 英文測驗

英文測驗

一 美國方面的英文測驗舉例.....	一二二
二 我國已編就的英文測驗.....	一二五

第十一章 自然史地科及各科混合測驗·····	一三九
一 自然科測驗·····	一三九
二 史地測驗·····	一三九
三 各科混合測驗·····	一四一
四 其他測驗·····	一四二
第十二章 舊式考試與新式考試的比較·····	一四五
一 引論·····	一四五
二 學校考試的功用·····	一四六
三 舊式考試的缺點·····	一四八
四 新式考試的優點及限制·····	一五三
第十三章 新式考試的編造法·····	一五六
一 新式考試的種類或方式·····	一五六
二 編造規則·····	一六四
三 編造新式考試的步驟·····	一六八

第二編 測驗的實施.....一七一

第十四章 施行測驗的步驟及手續.....一七一

一 舉行測驗的步驟.....一七一

二 施行測驗的詳細手續.....一七六

第十五章 測驗結果的整理(上).....一七九

一 次數分配.....一七九

二 集中趨勢的量數.....一八二

三 離中趨勢或差異度量數.....一九四

第十六章 測驗結果的整理(下).....二一一

四 相關量數.....二一一

五 量數的可靠性.....二二三

第十七章 測驗結果的解釋：年齡量表中各量數的意義.....二三六

一 年齡分數.....二三六

二	智力商數與教育商數·····	二三八
三	成業商數·····	二四四
第十八章	測驗結果的解釋：IQ 分數的意義·····	二五〇
一	IQ 分數·····	二五〇
二	IQ 分數·····	二五五
三	O 分數·····	二六一
四	IQ 分數·····	二六五
五	總結·····	二六六
第十九章	測驗結果的應用·····	二七〇
一	在教育行政及輔導上的應用·····	二七〇
二	在教學上的應用·····	二七九
三	在教育研究上的應用·····	二八〇
第四編	編造測驗的方法·····	二八五

第二十章 編造測驗的普通步驟.....二八五

一 測驗材料的選擇.....二八五

二 問題的編製.....二九一

三 初次的試用.....二九二

四 根據試用的結果加以刷修.....二九三

五 將測驗分為數個相等的卷份.....二九六

六 決定時間的限度.....二九八

七 求信賴.....二九九

八 決定測驗的可靠度.....三〇七

九 說明書的編訂.....三一〇

第二十一章 量表的編製.....三一四

一 量表的編製.....三一四

二 量表的編製.....三一〇

三 量表的編製.....三二七

第二十二章 品質量的表編製.....	三三〇
一 品質量表的性質.....	三三〇
二 編製品質量表的方法.....	三三〇
三 編製品質量表的另一方法.....	三三六
第五編 總結.....	三四三
第二十三章 測驗運動總評價.....	三四三
一 測驗運動的現狀.....	三四四
二 測驗運動對於教育的貢獻.....	三四七
三 測驗運動的流弊及限制.....	三四九
附錄一 漢英名詞對照表.....	三五五
附錄二 漢英人名對照表.....	三五九

表的目錄

一 學業成績與各種測驗成績的相關	二八
二 陸氏訂正比納西門智力測驗預備測驗的日分數	六四
三 陸氏訂正比納西門智力測驗全部測驗的日分數	六六
四 陳氏語體文應用字彙材料表	八〇
五 表示作文測量的主觀性	九三
六 表示計算評判錯誤之數量與方向之方法	九五
七 表示作文成績之變異	九七
八 <i>log</i> 位評判員對於同一篇字所定的分數	一〇一
九 俞氏書法測驗的年級標準	一〇二
十 祁司氏文法與語文測驗的可靠度	一二五
十一 祁司氏字彙測驗的標準	一二七
十二 安氏混合英文測驗的標準與當模：日分數	一三四

十三	安氏混合英文測驗各部分的標準與常模：口分數	一三五
十四	芝加哥大學附屬中學各科分數的分配	一四八
十五	九十一位教員對於一個地理問題的三個答案所給的分數	一五〇
十六	教材內容表：測驗應具的條件	一六八
十七	五十四個學生在一個智力測驗上的分數	一八一
十八	表示中數的求法	一八五
十九	表示算術平均數的求法	一九〇
二十	表示求算術平均數的簡捷法	一九二
二十一	表示求 Q, AD 的方法	一九五
二十二	表示用簡捷法求平均差的方法	二〇〇
二十三	表示標準差的求法	二〇三
二十四	表示求標準差的簡捷法	二〇六
二十五	表示相關係數的求法	二一四
二十六	表示等級相關法的步驟	二二〇

二十七	從 S 化為 Y 的應用表	二二二
二十八	智力商數的固定性	二四二
二十九	根據美國國家智力測驗量表乙與吳特麥柯爾算術測驗所得的算術成業商數的分配	二四六
三十	Y 分數的一種解釋	二五二
三十一	廖氏團體智力測驗量表甲的 Y 分數	二五四
三十二	年齡換算表	二五七
三十三	廖氏團體智力測驗各年齡 Y 校正數表	二六〇
三十四	廖氏團體智力測驗 Q 量表	二六三
三十五	○校正數與距開學時月數對照表	二六四
三十六	○校正數與測量時期對照表	二六四
三十七	Y 量表與年齡量表中各種量數對照表	二六六
三十八	四種量表的比較	二六七
三十九	表示求混合分數的方法	二七五
四十	表示求教學 Y 調級 Y 的方法	二七六

四十一	各年級做對各題目的人數的百分比	二八九
四十二	日價值與百分比對照表	二九三
四十三	表示百分數的求法	三〇二
四十四	表示將原來分數化成日分數的方法	三〇六
四十五	表示求日量表的方法	三一六
四十六	表示擴充日量表之距離的方法	三一八
四十七	表示求各年齡日校正數的方法	三二三
四十八	實足年齡與日校正數對照表	三二五
四十九	各年級分數的次數分配表	三二七
五十	分數與日分數對照表	三二八
五十一	常態曲線下面積的分布	三三二
五十二	表示將百分比的差別別爲日差別的方法	三三四

圖的目錄

一 智力發育曲線.....	九
二 149 位教員對於同一本英文考卷所給分數的分配.....	一四九
三 116 位中學數學教員對於同一本幾何考卷所給分數的分配.....	一五〇
四 表示廣的取樣.....	一五二
五 表示深的取樣.....	一五二
六 表示 σ 和 σ' 的意義.....	一九五
七 表示相關係數的分布圖求法.....	二一六
八 505 個從五歲到十四歲兒童的智力商數的分配.....	二三九
九 表示 \bar{X} 單位的意義.....	二五一



(續)

教育測驗

第一編 測驗之一般的理論

第一章 測驗在教育上的地位

一 測驗與教育的關係

從心理學的立場而論，教育的目的，乃在變更人類的行爲。不論教育的對象是兒童或成人，不論所應用的工具是什麼——學校，書報，戲劇，電影，圖書館，講演會——我們共同的目的，乃在人們的行爲上造成或引起一種變化，使他在已受教育以後和未受教育以前在行爲上有所不同。譬如一個兒童起初對於「 $\infty \times \infty$ 」這個符號的反應是「 ∞ 」或「 ∞ 」或「 ∞ 」或「我不曉得」，經過相當教育以後，我們希望將這個兒童這些不適當的反應變成一個適當的反應——「 ∞ 」。又如一個兒童起初對於「友人在路上落掉手帕一塊」這個情境的反應是「拾起」



(南)

收爲己有，「我們希望這個兒童因爲受了教育的原故能將原有的反應變爲「拾起，奉還物主。」

因此教育專業中有三個大問題：（一）我們在兒童方面究竟應該造成何種變化？何種行爲應該養成？何種行爲應該革除？（二）怎樣去造成這種變化？（三）這種變化究竟已否養成？

第一個問題是教育目標的問題，乃教育哲學，教育社會學的問題。第二個問題是教育方法的問題，乃教育心理學的問題。第三個問題是教育結果考查的問題，乃教育測量學本身的問題。所以測驗在教育上所占的地位是極重要的。我們要研究一種教育的結果，效率，必須要考查這種教育所製造的變化，測量變化的方向及數量。

二 教育測量的可能

物質方面的測量，由來已久，其可能性已爲人所共認。長度，重量，時間都可以十分準確地測量。但是心理測量的可能，截止最近，尙爲多數人所懷疑。品格，興趣，欣賞，態度等等很抽象的對象是否可以用科學的方法去測量，用數目字去計算花草的美麗，風景的幽雅，詩詞的清逸，文章的豪邁，是否可以用科學的方法去估計，用統計的方法去度量？下面兩條重要的原則是教育測量學在理論上的基礎，並且可以解釋上述的懷疑。

（一）凡物之存在必有其數量。這條原則是美國心理學者桑戴克（E. L. Thorndike）在一九一八年所提出的。我們大家都承認人們智慧有高低，品格有優劣，花草有美惡之別，詩詞有雅俗之分。這就是表示我們承

認世界上一切現象，一切特性都有程度的不同。所謂程度的不同，就是數量的有異。

(二) 凡有數量的東西都可以測量。這一條原則是美國測量學者麥柯爾 (W. A. McCull) 在一九二二年提出的。「既然有數量，就可以測量。」這一條原則在理論上是毫無問題的。在實際上，在目前能否做到，這當然又是一個問題。教育的結果——智識、技能、品格、興趣、欣賞、理想、態度——我們現在尚不能一一測量，這是無容否認的。但是我們應該顧到教育測量學的歷史是很短的，在這短促的時期中有這樣的成績，一方面可以表示學者已往的努力，一方面可以指示將來發展的無量。同時在最近的數年中，教育測量的範圍已大大的擴充，更足表示前途的光明。

所以測量精神特性的可能，就理論上講，實在無可否認的。在方法上，精神特性的測量與物質的測量有不盡同者。我們不能把一管尺或一個天平插入學生的腦袋，直接把他們的智力或算術能力加以測量。但是我們可以用間接的方法，去測量這些所謂精神的特性。智力或算術能力都可以在動作中，行為上表現出來的。(其實除掉一人的動作，行為，則其智力，算術能力等就無由表現。)要是某學生盡力做一件運用智力或算術能力的動作，那末，他的動作就是他的智力或算術能力所引起的結果。我們測量了他的動作，便可間接計算他的智力或算術能力的數量。

這種間接的測量方法，不特在教育結果的測量上採用，即在物質測量上，也時常採用的。例如，要測量空氣溫

度的高低，我們祇須觀察寒暑表上水銀的升降。因為水銀的升降乃空氣溫度所引起的結果，正如同學生的行為是能力所引起的結果一樣。

三 教育測量的困難

教育上所用的測量方法與物理學上所用的測量的方法，固然沒有截然的區別，但是測量一個人的能力，究竟比測量一塊石頭的輕重和一幅布的長短，更加困難。一塊石頭，今天重一斤二兩，過了幾天或幾個月，仍舊是重一斤二兩；一幅布，今天長一丈二尺，過了幾天或幾個月，仍舊長一丈二尺。但是一個學生，今天能在一小時間做十個算術題目；到了明天，在一小時內，也許祇能做九個同樣困難的題目。因為一個人的能力或作業成績極易受發育、休息、睡眠、疾病、光線、溫度和心理態度等等的影響，所以是常常變動的。能力的表現既然隨環境而變動，所以更加以正確的測量，實在是十分的困難。

但是教育上測量的不正確，不過是比較的說法，物質的測量亦不是絕對正確的。一個人的能力固然常常變動，但是各種無生命的物件，也時起變化。例如鐵路上的鋼軌，遇熱即脹，遇冷即縮，何嘗不起變化呢？所不同者，物體上所起的變化較為微小，而能力上所發生的變動較為巨大，能引起物體的變化的因子較少，能引起能力的變動的因子較多。所謂測量的正確與否，不是絕對的區別，不過程度上的差別而已。

四 教育測量的需要

教育事業中有三種基本要素：（一）學生，（二）教育目標，（三）教法與教材。這三個要素與測驗都有密切的關係，茲申述於后。

（一）若學生起初的能力或能量不能測量，則無從明瞭其個性。學生之於學校有如原料之於工廠。製造器具的人對於原料的特性必須深切明瞭；設施教育的人，對於兒童的個性，尤須徹底認識。然後才可以因材施教，教育才可以適合兒童的需要。如何可以明瞭兒童的個性，則不得不有賴於測驗的方法。

（二）教育目標若空泛無定，即無價值可言。我們對於任何教育目標至少應該能夠回答三個問題：（1）這目標有什麼價值？（2）這目標的所在地是什麼？（3）學生是否向着目標進行？要解決上面三個問題，都不得有賴於測驗。試以四則為例。學習四則究有何價值？我們或者可以說學習四則可以增進我們解決日常交易問題的能力。但是解決日常交易問題的能力，究竟是否因學習四則而增加，不得不加以考查。考查必須用測驗的方法。學習四則應達到何種程度，須能計算何等難度的題目，須能在規定時間內做幾個題目？這些問題的解答，亦必有賴於測驗。最後，學生已否達到規定的目標，與規定的目標相距多少？沒有測驗，這個問題亦無從解答。

（三）若教法與教材的效果，不能測量，即無法得知其價值。教法與教材都是教育中的工具，用以達到某項

教育目標，用以引起某種變化。所以要明瞭某種教法某項教材的價值，必須考查是種教法是項教材所引起的變化，測量其在兒童行為方面所發生的效果。是欲考查教法及教材的價值，又非測驗不可了。

問題

- ✓ 1. 教育結果的測量與物體的測量有什麼不同之處？
2. 教育結果的測量有什麼特別困難？
3. 教育測量與教育目標有什麼關係？
4. 一種教材或教法的價值怎樣才考查得出來？

參考書報

1. McCall, W. A. How to Measure in Education, Ch. I, Macmillan.
2. Thorndike, E. L. "Measurement in Education", The Twenty-first Yearbook of the National Society for the Study of Education, Part I, Ch. I.
3. 周詞陽，教育測量法概論，第二章，中華。

第二章 測量的要素

每一種測量，不論是物質方面的測量，心理方面的測量，必具有下列幾個要素。茲分述於下。

一 單位

單位即計算數量的單位。測量距離的長短，以尺，寸，丈，里為單位；測量物件的重量，以兩，斤，磅為單位；測量時間的長短，以秒，分，時，日，月，年，世紀為單位。沒有單位，則數量的多少，大小無由表示。

好的單位須適合兩個條件：（一）有確定的意義，（二）有相等的價值。何謂有確定的意義？即同一單位，在許多人的心理中，有同一的意義。譬如一公尺，一公斤，其意義十分確定，絕不致引起不相同的解釋。何謂有相等的價值？就是說第一單位與第二單位間的距離，等於第二單位與第三單位間的距離。例如有甲乙丙丁四個人，甲身長五尺四寸，乙五尺六寸，丙四尺三寸，丁四尺一寸。甲乙相差兩寸，丙丁相差亦兩寸。甲乙之間的差別等於丙丁之間的差別。這是表示尺，寸單位有相等的價值。在前段中我們所述及測量距離，重量，時間所用的單位是最自然，最完善的單位，對於本節所提出的兩個條件，是完全能適合的。

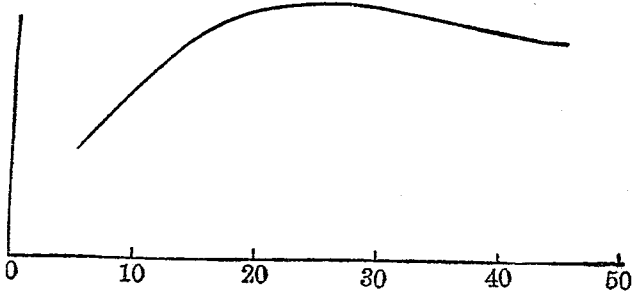
那麼我們在教育測量中，所用的單位，究竟是什麼？智力的高低，品格的善惡，學業的優劣，究竟用什麼單位

去計算？我們可以說尋求一個適當的單位是教育測量學中最重要，同時亦是最困難的工作。教育測驗中所應用的單位，約有下列數種：

(一) 以百分度為單位 百分度的意義及編造百分度量表的方法，將在第二十章中詳細討論。概括言之，一個兒童的百分數可以表示他在某團體中——年齡團體，年級團體——所占的地位。例如某兒童在某種算術測驗上得百分數四十，這就是表示與他同年齡（或同年級）兒童中，有百分之四十的算術程度不如他，有百分之六十的算術程度超過他。百分度單位的缺點就在單位價值的不等。據我們現在所知，在量表的兩端，兩個百分單位間的距離大，在量表的中段，兩個百分單位間的距離小。

(二) 以年月為單位 麥柯爾稱之為生長的單位。兒童教育程度，智慧程度的高低，均以年月為單位表示之。我們常常聽見「教育年齡」、「智力年齡」等名詞，即是以年齡為計算的單位。年齡單位的解釋較為簡單。例如某生的智力年齡為十歲，這就是說他的智力程度與十歲兒童平均智力相等。

年齡單位的缺點亦在單位價值的不等。這是因為人類各種特性的發育曲線或生長曲線成一弧形而非一直線。換句話說，逐年生長的數量是不相等的，發育的速率前後是不一致的。身長的發育如此，智力的發育亦如此。圖一表示智力發育的曲線。據我們現在所知，智力年齡四歲與智力年齡五歲之間的差別大，智力年齡十四歲與智力年齡十五歲之間的差別小。



從桑戴克(克)智力發育曲線

(三)以作業差異度為單位 我國所採用的T分數就是作業差異度單位之一種。T等於十二歲兒童某種特性的分配的均方差的十分之一。T分數的意義及編造T量表的方法將在十八、二十一兩章內詳細討論，現在暫不申述。

(四)以評判差異度為單位 品質量表如作文量表，書法量表等都以評判的差異度為單位。品質量表的編製在後面第二十二章裏有詳細之討論。

二 參照點

參照點即計算的起點。參照點不同，則測量的結果就沒有互相比較的可能。測量溫度以結冰點為參照點(或以結冰點以上三十二度為參照點)；測量陸地的高度，以海洋面為參照點；測量某一個地方的位置，則以英國 Greenwich 地方及赤道為參照點。

參照點有兩種：(一)絕對的零點；(二)人定的參照點。測量輕重，測

量長短，都以零點爲參照點，即以「恰恰沒有一點重量」「恰恰沒有一點長度」爲計算的起點。如以海洋面爲測量陸地高度的起點，以冰點爲測量溫度的起點，這些都是人定的參照點。

教育測量中所應用的參照點，可以說都屬於第二類，即人定的參照點，因爲智力、算術能力、理解能力等的零點甚難決定。現行教育測驗所應用的參照點，極不一致。有以一年級或一年齡兒童作業的平均點爲參照點者，有以編造測驗者所推測的零點爲參照點者，有以某團體作業的平均點以下三個或五個均方差爲參照點者，各種測驗所應用的參照點不同，則所得結果不能直接比較。要改進這種紛歧的情形，須大家共同決定一種參照點，以後大家編造測驗時，即一律以此點爲計算的起點。麥柯爾曾提議以十二歲兒童平均作業以下五個均方差爲參照點，頗得一般測量學者的贊許，但尙未能一律採用。

人定的參照點有一種極大的限制：就是從人定參照點計算起分數不能以「倍數」的方式解釋之。有兩條繩子，一條長四尺，一條長二尺，我們可以說第一條繩子的長度，倍於第二條繩子的長度。有甲乙兩個兒童，甲的智力年齡十二歲，乙的智力年齡六歲，我們却不可以說甲的智力程度倍於乙的智力程度。有甲乙兩個兒童，甲的分數超過同班兒童的平均分數十分，乙的分數超過同班兒童的平均分數五分，我們絕對不能說甲的程度倍於乙的程度。此理甚明，但常爲解釋測驗分數者所忽略。

總結以上的討論：（一）最完善的參照點，是絕對於零點；（二）如用人定的參照點時，則此參照點愈接近

零點愈好；（三）編造測驗者，在可能範圍內，應該應用同一參照點，俾由各種測驗所得的分數，有比較的可能。

三 量表

量表即測量的工具。量表係用等值的單位製成，且定有標準點或參照點。應用量表以測量事物時，即視此事物在量表上距參照點的遠近如何。用權衡以測量輕重，用尺度以測量長短，這些都是測量物質方面事物的量表。在教育測量或心理測量方面，測量量表乃一種測驗結果或分數的排列，以顯明的（最好的）單位表示一種能力的階梯，從零度能力（或者某低度能力）起，到最高能力（或者某高度能力）止。這量表是一種固定的標準，凡個人或團體的能力，都須依此以度量之。

教育測驗所用的量表約有四種：（一）百分量表，（二）年齡量表，（三）作業差異度量表，（四）評判差異度量表。我國測驗所用的 T 量表就是作業差異度量表之一種。

四 測驗

物質方面的測量，有單位，參照點，量表就可以進行。心理方面的測量還須有第四種要素，就是標準的測驗。原來物質是可以直接測量，而心理方面的特性祇能間接測量的。智力，興趣，閱讀能力等，乃抽象的名詞，不是具體的

物件。我們所能觀察到的，所能記載的，祇是人的行為，動作，反應。觀察人的行為，動作，反應，以估計其智力的高低，學業的深淺，品格的優劣，興趣的趨向，這是心理測量根本的方法。所以必須有行為可以觀察，然後測量才可以進行。測驗即引起行為的工具。因此我們可以下一個測驗的定義：測驗是一個或一羣標準的刺激，用以引起人們的行為，根據此行為，以估計其智力，品格，興趣，學業等。我們必須注意到上述定義內「標準」兩個字。施行測驗的時間，說明，評定結果的方法等都有嚴格的規定，舉行測驗者必須遵守預定的標準。

問題

1. 年齡單位有什麼缺陷？
2. 編造測驗者所用的參照點不對，有什麼不便之處？
3. 人定的參照點有什麼限制？
4. 「量表」與「測驗」有什麼區別？
5. 下一個測驗的定義並解釋之。
6. 物質方面的測量有單位，參照點，量表三者就可以進行，何以心理方面的測量還須有第四種要素——標準測驗？

參考文獻

- 1 McCall, W. A. How to Measure in Education, Ch. 9, 10. Macmillan.
- 2 Thorndike, E. L. Mental and Social Measurement, Ch. 2, Teachers college, Columbia University.
- 3 Trause, M. R. Measuring Results in Education, Ch. 1, American Book Co.
- 4 Wood, B. D. Measurement in Higher Education, Ch. V, I, World Book Co.

第三章 測驗應具的條件

一個完善的測驗應具的條件，依測驗的性質，種類而異。下列大綱，是以教育測驗為主要的對象的。

- 一、正確性 (Validity)
 - 二、可靠性 (Reliability)
 - 三、鑒別力 (Discriminative ability)
 - 四、常模 (Norm)
 - 五、施行及校閱的便利
 - 六、診斷的價值
 - 七、覆份
 - 八、經濟
- 茲分別論之：

一 正確性

如一個測驗，真能確實的測量他所欲測量的東西，這個測驗可謂之正確的測驗。正確性是測驗最基本的條件。測驗如不正確，即無價值可言。或者我們可以說，一種測驗的內容，須與其名稱相合。智力測驗所測量的，應該純粹是智力；算術測驗所測量的應該純粹是算術能力；閱讀測驗所測量的，應該純粹是閱讀能力。所以要考查一種測驗是否正確，必須將測驗的結果與一種正確的標準相比較。但是如何求得這正確的標準却是一個不易解決的問題。

茲先以商店售貨員測驗為例。商店售貨員測驗所欲測量的是售貨的能力。我們要考察這個測驗是否正確，必須將一班售貨員在這個測驗上所得的分數，與他們真正售貨的能力——即正確的標準——相比較。在此處這個正確的標準，比較容易求得。售貨員在規定時期以內——一月或半年——售出貨物的價值，可以代表他售貨的能力，即正確的標準。我們祇須求一班售貨員在這個測驗上所得的分數與他們在規定時期內售出貨物的價值的關係，就可以知道這個測驗的正確度（專門名詞稱為正確係數）。相關度愈高，則正確度亦愈高。相關度的意義與計算方法，在第十六章中將有詳細的解釋。

再以智力測驗為例。要考查一個智力測驗是否正確，我們必須將一班兒童在這個測驗上所得的分數與他們真正的智力——即正確的標準——相比較，求二者的相關度。但是在此處，這個正確的標準，却不易求得。我們正因為不曉得這班兒童的智力程度，所以才編造這個智力測驗。假使我們已經有了其他正確的標準，則這個智

力測驗根本就無須編造了。這種困難，正確標準的難求，在考查其他測驗——如品格測驗，教育測驗——的正確性時亦有之。在此種情形之下，我們祇能用其他間接的方法，以考查測驗的是否正確。例如推孟 (L. M. Terman) 在訂正比納西門智力量表時曾用三種標準以考查每個測驗的是否正確。第一，各年齡兒童能做對某個測驗的人數的百分比，是否逐年增加？第二，每一個測驗與全部量表是否相關？第三，測驗的結果與教師的評判，學業的成績，是否相合？

再以教育測驗為例，教育測驗的是否正確，大半要看測驗材料是怎樣選擇來的。例如編造默字測驗時，所用的字必須是常用的字；編造算術測驗時，所用的題目必須包括各種重要的算術過程；編造歷史測驗時，所出的問題必須是歷史中重要的部分。測驗材料的選擇實為編造教育測驗過程中最重要的工作。

總結上面關於正確性的討論，要得到一種正確的測驗，必須顧到下列各點：

- (一) 測驗的內容須經過審慎的選擇。選擇材料的標準將在第二十章中詳細討論之。
- (二) 測驗中應盡力將與所測量能力不相關的分子除去。
- (三) 如其他情形相等，測驗愈長，其正確度愈高。
- (四) 測驗中的題目須有適當的難度。如某題目，全體學生都能做對或都不能做對，則此題目即無價值可言。

二 可靠性

測驗的第二個基本條件，就是可靠。如將一個測驗測量同一個兒童兩次（假設這個兒童在第一次與第二次測量時並無若何之變更），如兩次所得的結果相等，這個測驗可謂之可靠。換一個方法說：如將一個測驗測量同一班兒童兩次，求兩次結果的相關度，相關度愈高，則測驗愈可靠，相關度愈低，則測驗愈不可靠。這個相關係數，專門名詞稱為可靠係數。

測驗的可靠性是很重要的。不可靠的測驗，沒有多大的用處。假使用一把尺來測量一塊布，第一次量之為五尺，第二次量之為四尺，或者甲來測量結果得五尺，乙來測量結果得四尺，這把尺還有什麼用處？假使用一種智力測驗來測量某兒童，今天得智力年齡八歲，明天得智力年齡十歲，這個智力測驗亦沒有什麼功用可言了。

能影響可靠性的因素，約有三種。

第一、客觀 要增加測驗的客觀性，第一，施行的手續必須有劃一的標準。例如測驗中的說明及時間必須嚴格規定。第二，校閱試卷及計算分數必須有客觀的方法，使校閱者不易插入一己的偏見。

第二、適當的取樣 所謂「取樣」，就是測驗材料的選擇。我們可以分兩點來討論：（一）測驗問題的多寡。如其他情形相等，測驗愈長，其可靠性愈高。其關係可用下列公式表示之。

$$r_{nn} = \frac{n \cdot r}{1 + (n-1)r}$$

在公式內， r 代表這個測驗的可靠係數， n 代表測驗增長的倍數， r_{nn} 代表測驗增長後的可靠係數。譬如有一測驗，原來的可靠係數等於 $.8$ ，現在將這個測驗內問題的數目增加一倍，則其可靠係數為：

$$r_{nn} = \frac{2 \times .80}{1 + (2-1) \cdot 80} = \frac{1.60}{1.80} = .89$$

(二) 測驗內問題的難度的支配。如支配適當，則其可靠度可以增高。例如測驗中不應有為全體都能答對或都不能答對的題目。

第三、被試者的反應。被試者對於測驗有無興趣，能否努力，對於主試者能否合作，被試者當時身體上有無不適，心理上無不安。

以上三種因素，如能一一顧到，則不難得一可靠的測驗。

但是「可靠」係比較的名詞，究竟如何才能算為可靠，是很不容易回答的問題。羅雪 (G. M. French) 及施道突 (G. D. Stoddard) 曾定有下列標準：

可靠係數

解釋

0.95—0.99

可靠度甚高；現有的測驗中所不常見的。

請看上面的圖。圖中每一個 \times 代表一個題目，一個 \times 在難度量表上的位置，表示這個题目的難度。第一個圖內二十個題目祇有四等難度，所以祇能鑒別四等程度。第二個圖內二十個題目難度的分配比較均勻，可以鑒別二十等程度。同是二十個題目，但是因為難度分配的不同，鑒別力就顯然有異了。

四 常模

常模普通有兩種，就是年齡常模，年級常模。年齡常模就是每個年齡兒童在某種測驗上所得的平均分數；年級常模就是每個年級兒童在某種測驗上所得的平均分數。所以求常模的手續是很簡單的。我們祇須將測驗去測量多數兒童，求每年齡或年級兒童所得的平均分數，這個平均分數就是某年齡或某年級常模。

常模的可靠與否却是很重要的問題。關於這問題最重要的一點，就是常模所根據的兒童，是怎樣選擇出來的，選擇的方法有無偏倚。譬如我們要為某種測驗求十二歲兒童的常模，最可靠的方法當然是將所有十二歲兒童個個測量，然後求其平均。但是在實際上因為時間，人力，經濟的限制常常不能做到。我們祇能測量十二歲兒童的一部分，以一部分的結果去代表全體。因此發生能否代表的問題。我們要使得小部分足以代表全體，在選擇兒童時必須顧到「隨機取樣」的原則。第二，在「隨機取樣」原則之下，如其他情形相等，常模的可靠度與常模所根據的兒童人數的多寡成正比。測量的人數愈多，則常模愈可靠。

常模的功用，乃在解釋測驗的結果。一個兒童或一班兒童的分數可與常模來比較。常模供給我們一個比較的標準。

五 施行及校閱的便利

施行測驗及校閱試卷的手續須有詳明的規定。說明須清楚簡單，使學生方面至不發生誤會。在必要時，可在正式測驗前加入一、二個例子，或練習題目。測驗必須預備有標準答案，使校閱試卷時有所根據，不致參入主觀的見解。麥柯爾對於測驗的說明有下列的規定：

- (一) 測驗說明務須簡單，同時須清楚，完備。
- (二) 最好用幾個例子或練習測驗，以示測驗的做法。
- (三) 說明須對於被試的人們相適合。
- (四) 說明的次序須是做測驗的次序。
- (五) 說明須分段落。
- (六) 說明須能引起兒童最大的努力。
- (七) 對於兒童的說明以外，還須備有對於主試者的說明。

六 診斷的價值

測驗依其功用可分為兩種：普通測驗與診斷測驗。普通測驗的結果僅能表示一人程度的高低，診斷測驗的結果可以表示一人缺點之所在，特點之所在，在那幾方面或那幾個步驟上已經練習純熟，在那幾方面或那幾個步驟上還未能深切了解，嫻熟。所以普通測驗僅可以為調查或團體比較之用，而診斷測驗却為教學上不可少的工具。下列幾個問題可以幫助我們去考查一個測驗的診斷的價值：

- (一) 這個測驗是普通測驗或是診斷測驗？
- (二) 如其是診斷測驗，這個測驗的編造根據什麼原則？
- (三) 這個測驗能分析或測量幾種功能？
- (四) 這種分析是否合於教學的需要？
- (五) 診斷是個人的還是團體的？
- (六) 測驗說明書中有否述及補救教學的具體方法？

七 覆份

爲增加實際的效用起見，一種測驗至少須有兩份，兩份「相等」的測驗。當然份數愈多，應用上愈便利。例如我們要用測驗來考查一班兒童在一學期中的進步，我們必須測量兩次，一次在學期之初，一次在學期之終。這兩次結果的差別可以代表一學期中程度增加的數量。假說測驗祇有一份，那麼同一份測驗用兩次，就難免「練習」的影響。兩次測驗結果的差別，一部分是受練習的影響，不能完全代表進步的數量。要是這個測驗有好幾份可以替換應用，就可以免掉這種困難。

前面我們說測驗的各份須「相等」，何謂「相等」，尚須解釋。兩份測驗，如能適合下列幾條條件，可以稱爲「相等」：

- (一) 各份測驗須測量同一種能力。
- (二) 各份測驗中所用材料不應有重複的地方。
- (三) 各份測驗的難度須相等。
- (四) 各份測驗分數的差異度須大致相等。
- (五) 每份測驗須是可靠的。

八 經濟

經濟可分時間上的經濟與金錢上的經濟。所謂時間上的經濟，就是指測驗不應太長。太長則費時間太多，應用上就難以推廣了。但是一個測驗究竟應多少長，不能下一個普遍的規定。我們前面說，測驗的長度與其可靠度，有相當的關係。過分注意到時間上的經濟而不顧到測驗的可靠度，是得不償失的。

所謂金錢上的經濟，就是說測驗的價格不應太貴。但是同時我們亦須顧到測驗的美觀等等。陳鶴琴氏所編的小學默讀測驗，測驗卷子以外，另備答案紙條學生的答案完全寫在答案紙條上，所以測驗試卷可以應用多次，這是一個極經濟的辦法。

問題

1. 可靠的測驗是不是一定正確？
2. 考查一個測驗的正確性有什麼困難？
3. 何謂「正確係數」與「可靠係數」？
4. 測驗的長度與其可靠度有什麼關係？
5. 何謂「取樣」、「隨機取樣」？
6. 常模有什麼功用？

7. 參閱測驗與評量測驗有什麼區別？
8. 測驗的預份有什麼功用？

參考書目

1. Kuch, G. M. *The Objective or New Type Examination*, Ch. 2, Scott, Foresman co.
2. Kuch, G. M. and Stordard, G. D. *Tests and Measurements in High School Instruction*, Ch. 4, World Book Co.
3. Synonds P. M. *Measurement in Secondary Education* Ch. 13,14.
4. Wilson, G. M. and Hoke, K. J. *How to Measure (Revised Edition)* Ch. 23, Macmillan.
5. 國定標準國民小學課程標準 二十三章 第十一節

第四章 智力測驗發展的歷史

我們對於社會上任何運動要有正確的認識，必須明瞭其發展的歷史。本章及第五章就智力測驗及教育測驗發展的歷史作一個簡括的敘述。

智力測驗發展的歷史，可以分三個段落來討論。第一，初期的嘗試；第二，比納的工作及其影響；第三，比納智力量表成功以後的發展。

一 初期的嘗試

心理學最初的研究目的在發現人類行爲的普通法則。但是對於同一刺激各人的反應常常不同。此種個人間的差異，初以爲由於試驗時手續上的錯誤，經過長時間的實驗，心理學者才認識此種反應上的差異並非由於偶然的錯誤，乃基於個人間能力上的真正差別。因此個別的研究漸成爲心理學上重要的問題。要研究個別的現象，必須有測量個別的工具，所以個別的研究直接引起測量的運動。

對於個別的研究最有貢獻者爲喀泰爾 (J. Mark Cattell)。喀泰爾爲美國人，留學於德，從馮德 (W. Wundt) 學，後又與英優生學者高爾登 (F. Galton) 相過從，受二氏的影響甚深。一八九〇年喀泰爾在英國雜誌 心理上

發表一篇重要文字，題目「心理測驗與測量。」在這篇文字裏他說：「心理學若不立根基於實驗與測量上，決不能夠有自然科學的準確。」又說：「如其我們規定一個一律的手續，使在異時異地所得的結果可以比較，綜合，則測驗的科學的和實用的價值都可以增加。」這些都是測驗編造方法上很重要的觀念。

在同文中喀泰爾提出下列十種測驗。這幾種測驗頗可以代表當時一般心理學者所應用，所試驗的測驗的性質。

- (一) 握力的測量。
- (二) 動的速率的測量。
- (三) 觸覺的二點閾的測量。
- (四) 引起痛覺的最低壓力的測量。
- (五) 辨別重量的能力的測量。
- (六) 反應時間的測量。
- (七) 說出顏色名稱的速率的測量。
- (八) 將長五十公分的一條線平分爲二，測量其準確度。
- (九) 令被試者憶測十秒鐘的時間，測量其準確度。

(十) 機械記憶能力的測量。

喀泰爾在一八九四年起首用各種測驗測量哥倫比亞大學學生。他在哥倫比亞大學所用的測驗雖然同他在一八九〇年所發表者不盡相同，但是大致相類，測量的對象都屬於感覺能力與動作過程。在一九〇一年威斯拉 (C. Wikler) 報告哥倫比亞大學應用測驗的結果。其中有一部分結果表示測驗成績與學業成績的關係，而以相關係數表示之。相關係數的意義以後我們還要詳細解釋，現在祇能簡單說明。相關係數乃是一種數量，表示兩種事件的關係的深淺。相關係數的數量，從「-」經過「0」到「+」。「+」表示完全正相關，「0」表示零度相關，「-」表示完全負相關。威斯拉所得結果如下：

表一 學業成績與各種測驗成績的相關

學業與反應時間.....	.02
學業與劃去「a」字測驗.....	.09
學業與聯念時間.....	.08
學業與說出顏色名稱.....	.02
學業與理性的記憶.....	.19
學業與聽覺記憶.....	.16

從表一可見各種測驗成績與學業的相關極低。當時威斯拉對於這種結果頗表示失望。其實從現在的眼光看來，這是必然的結果。學業的優劣與智力的高低有深切的關係，而智力的高低，却不是感覺能力與動作過程的測驗所能測量。換句話說，當時喀泰爾與威斯拉所採用的測驗都不是測量智力的好測驗。

除喀泰爾外當時其他心理學者亦有應用心理測驗的嘗試。波爾頓 (T. L. Bolton) 在一八九一年用記憶數目測驗測量了許多兒童。他又請教師評判這些兒童的普通智力，然後求二者——即測驗成績與教師評判的關係。

在一九〇〇年勃格萊 (W. O. Bagley) 用許多動力測驗測量兒童，然後求測驗成績與學業的關係。

同年克伯霍克 (E. A. Kirkpatrick) 用朗數、分牌等等測驗考試了五百個兒童。他把各種測驗的結果與兒童在學校內所得的成績相比較，以決定心理測驗是否能預示兒童在校的成績。他當時極感覺到測驗標準的缺乏，並開始求各年齡的標準。

在一九〇三年克萊 (R. L. Kelly) 發表一篇極可注意的論文，題為「平常的和異常的兒童之心理的與身體的測驗。」克萊的目的在尋求一理簡單的方法，來鑒別平常的與異常的兒童。他所用的測驗大都是體方方面的測驗。

同時比納 (A. Binet) 在法國亦正在多方試驗測量智力的方法。他在這初期的工作，亦如美國心理學者的

工作一樣，並未得有若何的結果。他的工作我們在後面將有詳細的陳述。

綜觀以上各心理學者在這時期中的工作，我們可以得到幾條結論：

- (一) 在這時期心理學者對於測驗的興趣偏於學理的探究，尚未及於實際的應用。
- (二) 這時期所用的測驗，都是單獨的測驗。像後來多種測驗綜合而成的量表——如比納西門智力量表——尚無人編造。

(三) 這時期的測驗在編製方面，施行方面都缺乏嚴密的標準。因為相關方法尚未通行，測驗的功用，可靠度都無從決定。

(四) 這時期所用的測驗都是感覺能力、動作過程的測驗，與「智力」實在不甚相干。雖然愛屏好斯 (H. Ebbinghaus) 的填字測驗及有幾個比納所提出的測驗都屬於所謂高層心理過程，但是這些究竟是例外。

在此處我們萬不可忘記了英國優生學者對於測驗運動的貢獻。高爾登在一八六九年出版了他最重要的著作遺傳的天才。這本書開端的第一句就說：「我在這本書裏要證明人類的能力是由遺傳而來的，無異生物的身體構造之遺傳。」要研究能力的遺傳，沒有測量能力的工具，是無法進行的，高爾登對於能力的測量雖然表示很大的興趣，但是始終並未編造具體的測驗。

高爾登對於測驗學的基本科學統計學，有極大的貢獻。他對於統計學的貢獻有二：(一) 擴充顧特來 (Quo-

Galob) 的百分法，(二)創造相關法。他的同事潘爾遜(K. Pearson)繼續其事業，對於統計方法多所發明，成爲今日之唯一統計專家。

二 比納的工作及其影響

在一九〇五年比納和西門(E. Simon)編造成功第一個智力量表。他們的功績是不可淹沒的。但是我們不要以爲比納的成功是驟然的或偶然的。比納對於智力測驗有長時期的興趣，經過了許多的嘗試，誤走了許多的歧路，才得到最後的成功。

我們閱讀比納在雜誌上發表的論文，不能不驚佩其興趣的廣博，著作能力的偉大。在一八九八年他在科學雜誌上發表一篇極可注意的文章，題爲「個人心理學中的測量。」在此文中，他提及許多測驗，如畫方形，比較線的長短，記憶數目，詞句重組，回答含有道德裁判的問題，了解抽象文章的意義，摺紙等等，他後來都採用於其智力量表中。在此文中，他討論心理測量的原理與方法，極見精確，他後來的成功，在這個時候已種下根基了。

次年，比納又發表了一篇論文，題爲注意與適應，爲比納發表智力量表前之最重要著作。他的目的在研究自動的注意與智力的關係。他的結論是聰明的兒童能有較速的，較完善的適應。他說所謂注意能力就是對於新環境的適應的能力。這個解釋同後來德心理學者斯旦恩(W. Stern)的智力定義幾乎一致了。

但是比納的成功是經過許多嘗試，走了許多歧路而後得到的。在一九〇〇年，比納化了很多時間去測量兒童的頭部。他的目的在比較聰明與愚笨兒童的頭部的測量。他的結果表示大體上聰明兒童稍佔優勢，但是差別很小，互掩很大。為個人診斷起見，頭部的測量可說是完全無用的。

在一九〇四年，比納發表了幾篇研究書法的文字。他問是否能從一人書法而決定其性別，年齡，智力。他自己的結論以為書法可以做一個粗魯的智力指南。

(一)一九〇五年的智力量表——在一九〇四年，法國教育部長委派許多醫學家，教育家與科學家，組織一個委員會，研究公立學校中特別班的管理法；比納亦委員之一。這個委員會的工作把比納對於智力測量問題的興趣從理論方面轉移到實際應用方面。次年他同西門發表一篇論文，題為「診斷異常兒童的智力之新方法。」在此文中他們發表所謂一九〇五年量表。

這個量表雖極簡單，但是有幾點值得我們的注意。第一，這個量表共包含三十個測驗，同以前心理學者所試驗的單獨測驗性質迥然有別。這是比納成功原因之一。原來智力的表現是多方面的。我們對於一人的行為，反應有多方面的觀察，然後對於他的智力的高低才能有較公允的評判。我們現在又知道如其他情形相等測驗題數愈多，則其正確度，可靠度愈高，所以單獨的測驗很難得正確，可靠。

第二，這三十個測驗有難有易，依照其難度而排列，所以成爲一個難度的量表，使同一個量表可以測量高低

不同的智力。這時候比納雖然還沒有用年齡的單位以度量智力的高低，雖然還沒有把測驗歸入年齡的組別，但是讀他這篇論文，可以曉得這種觀念確已粗具形態。譬如比納在這篇文章裏指明三歲，五歲，七歲，九歲，十一歲兒童在他的量表上可以做對什麼測驗。

(二) 一九〇八年的量表——在一九〇八年比納又發表一篇極重要的著作，題為兒童智力之發展。在此文中他詳細說明他第二次的量表，即世所謂一九〇八年的量表。這個量表的最重要點，就是所有測驗——共五十九個——都依其難度歸納入相當年齡組別，組別從三歲組到十三歲組。兒童考試的結果，亦以智力年齡表示之。計算智力年齡的方法，比納亦有大致的規定。這是第一個年齡量表。

(三) 一九一一年訂正量表——一九一一年，即比納去世之年，他發表一九〇八年量表的訂正本，即世所謂一九一一年量表。這個量表與一九〇八年的根本上沒有什麼不同，不過根據應用的結果，添進了許多新測驗，廢棄了許多舊測驗，還有測驗的位置亦有所變動。每個年齡的測驗的數目一律改為五個，如此在計算智力年齡時就便當多了。

到如今，比納西門的智力量表已經過許多的修訂，在方法上，手續上已有許多的改進，他們原來的量表已漸漸失去其功用了。但是後人的工作都是建築於比納所留下的基礎之上，我們現在所有的智力測驗的基本觀念仍舊是比納所賜。我們很贊同賓特納 (E. Pintner) 的話：「在心理學史上，假使我們稱馮德為實驗心理學的真

祖，我們不得不稱比納為智力測量的鼻祖。」

三 比納智力量表成功以後的發展

比納智力量表對於心理學教育學所發生的影響至深且廣，因篇幅所限，我們不能一一詳述。在心理學方面，智力測驗已成為心理研究，心理診斷不可少的工具；在教育方面，智力的測量引起教育家對於個別的通知，以及適應個別的種種設施。茲僅就比納智力量表成功以後智力測驗發展的情形約略述之。

比納智力測量對於英美法德各國的測驗事業都發生重大的影響，各國都有比納西門智力量表的訂正本。英國修訂比納量表者有約翰斯頓 (Katherine Johnson) 女士與溫區 (W. H. Wirth) 及勃推 (O. Burt) 等；美國則有高達德 (H. H. Goddard) 推孟 庫爾門 (F. Kuhlmann) 葉格斯 (R. M. Yerkes) 黑林 (J. P. Hering) 等人；德有斯旦恩 鮑勃太格 (Bobertag) 等；意有特立夫 (Trovati) 和沙飛亞替 (F. U. Saffioti) ；日本、中國、瑞典、俄國、土耳其等亦都有比納量表的訂正本。美國為測驗事業最發達的國家，茲僅就美國智力測驗發展的情形再作比較詳細的申述。

(一) 比納量表之輸入 第一個人介紹比納量表於美國者為高達德氏。高達德為美國著名的診斷心理學家，在一九〇六年任凡因蘭 (Vineland) 地方低能兒童訓練學校心理實驗室主任。一九〇八年他起始應用比

納量表於低能兒童，結果使他深信比納量表的價值，同時使他感覺得比納量表中許多測驗及其排列不盡適用於美國兒童。一九一〇年他發表他的修訂的量表，頗引起一般心理學者的注意。惟因歷史的關係，當時力量表的應用，僅限於低能兒童，未嘗用之於普通兒童，因此社會人士對於智力測驗發生了一種偏見，以為凡被測量者都有低能的嫌疑，所以一般人都願其子女應試。這種偏見對於智力測驗的推廣，頗發生些阻礙。到現在智力測驗已成爲學校行政及教學上不可少的工具，這類的偏見亦已漸漸消滅了。

(二) 斯旦福訂正本 比納西門智力量表最完善的訂正本爲斯旦福大學的訂正量表。主持其事者爲斯旦福大學心理學教授推孟氏。推孟費了五年的功夫，經過嚴密的訂正，在一九一六年才發表他修訂的量表。這個新量表共有九十個測驗，其中五十四個是比納量表中本來有的，三十六個是新採入的。對於測驗的位置亦有許多的變動。測驗的數目從三歲組到十三歲組每組有六個測驗，十一歲從缺，十二歲組有八個測驗，十三歲從缺，十四歲組有六個測驗，以上還有兩個階段，推孟稱之爲普通成人階段與優異成人階段，各有六個測驗。對於施行測驗的手續及記分的方法推孟有極詳明的規定，載氏所著智力測量一書。

此外推孟還有一個很大的貢獻，就是智力商數的引用。前面我們提到比納量表中的分數以智力年齡表示之，但是智力年齡僅表示一個兒童的智力的絕對程度，不能表示其聰明程度的高低。要求一個兒童的聰明程度，必須將智力年齡與實足年齡比較。這兩種年齡的關係的意義，比納亦曾見到，但是沒有規定具體的方法，他祇說

一個兒童要是智力年齡比實足年齡低一歲或兩歲，他就是比較愚笨的兒童。但是用智力年齡與實足年齡的差別以表示一人之聰明程度，乃是極不準確的方法，因為同一數量的差別，因在生長曲線上地位不同，其意義亦不相同。譬如某甲實足年齡三歲，智力年齡二歲，差別是負一歲，乙實足年齡十五歲，智力年齡十四歲，差別亦是負一歲，但是甲乙二人的聰明程度實不相等。據我們現在所知，乙的聰明程度實較高於甲。

最早提出一種計算的方法以避免上述的困難者為德心理學者斯且恩。在他一九〇四年所出版的測量智力的心理的方法一書中，他即提議採用「心理商數」一數量以表示聰明程度。心理商數的求法，是將智力年齡被實足年齡除，其商數即為心理商數。所以一個兒童的智力年齡與實足年齡相等，則其心理商數為1；年齡十二歲的兒童，智力年齡僅有十歲者，得心理商數0.83；反之，十歲兒童有十二歲的智齡者，得心理商數1.20。

推孟自己所搜集的測驗統計證明這種數量的正確。但他稱之謂智力商數，並且為避免小數起見，將商數乘一百，其公式如下：

$$\text{智力商數} = \frac{\text{智力年齡}}{\text{實足年齡}} \times 100$$

智力商數代表聰明程度，100代表中等程度，100以上代表中等以上，100以下代表中等以下。

(三) 其他證正本 除高達德推孟的訂正比納量表以外，尚有幾位心理學者亦從事同樣的工作。

庫爾門曾經發表過兩個訂正本，一次在一九一二年，一次在一九二二年。庫爾門最大的貢獻是把量表擴充到三歲以下的兒童，計有三月組，六月組，十二月組，十八月組，兩歲組，各有測驗五個，自三歲組以上，每組測驗均增加至八個，全量表共有測驗一百二十九個。測驗的位置比較原量表頗有變動，施行測驗的手續亦有詳細的規定。

在一九一五年，葉該斯，勃雷極司(J. W. Bridges)及哈特會克(R. S. Hartwick)發表他們的分點量表。該量表共有二十個測驗，其中有十九個是從比納原來量表中採取來的。測驗實在的數目却不止二十，因為每個測驗包含着幾個部分。分點量表的特點在廢棄比納量表的年齡分組法以及通過與否的記分法。如順背的數目測驗，在比納原量表中(一九〇八年量表)順背兩個數目字是三歲組測驗之一，順背三個數目字為四歲組測驗之一，順背五個數目字為七歲組測驗之一。在分點量表中，這些性質相類而難易不同的測驗都歸於一個測驗之下，完全放棄了比納以年齡分組的方法。又如倒背數目自二十至一這個測驗，在比納原來量表中為八歲組的測驗，通過的條件為「在二十秒鐘內數完由二十到一，並且最多只許漏落或顛倒一個數目。」在分點量表中，能倒數自二十至一者得四分，自十五至一者得三分，自十至一者得二分，自五至一者得一分，放棄了比納的通過與否的記分方法。

在分點量表中，做對某個測驗即可得若干分數，一個兒童在量表上的總成績以在各測驗上所得分數的總和表示之，與比納量表之直接以智力年齡表示測驗成績者在方法上有所不同。但分點量表與年齡量表的區別

不過是手續上的不同，並非原則上的有異，因為一個兒童在分點量表上所得的分數仍須有年齡常模以解釋之。下表為葉該斯所供給的年齡常模表。

年 齡	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
平均分數	15	22	28	35	41	50	58	64	70	74	79	81	84	86	88

例如某兒童在分點量表上得總分五十分，則其智力年齡為九歲，得七十分者，其智力年齡為十二歲。

總而言之，分點量表與比納量表在原則上雖然沒有若何之區別，但從便利一點上着想，似以分點量表為較勝。分點量表的組織較有彈性，分點量表的常模亦較易求得及修改。

在葉該斯的分點量表上求聰明程度的方法，亦與比納量表上求智力商數的方法不同。在分點量表上兒童的聰明程度以他所得分數與同年齡兒童所得的平均分數的比例表示之。這個數量，葉該斯稱之為智力係數 (Coefficient of Intelligence)。同智力商數一樣，智力係數 1.00 代表中等程度，1 以上代表中等以上，1 以下代表中等以下的聰明程度。智力商數與智力係數兩種數量之間關係如何，尚難確定，至少我們可以說他們的意義並不完全相同的。譬如，我們不能說智力商數 120 與智力係數 1.20 代表相等的聰明程度。

黑林的訂正比納-西門智力量表是在一九二二年發表的。他的量表亦是一個分點量表，但在組織上與葉該斯的分點量表不同。黑林的訂正量表共有三十八個測驗，分為五組，第一組應用完畢，即可得一結果。如欲求比較

可靠的結果，則可繼續應用第二組測驗，如再繼續應用第三組，第四組，第五組，則結果的可靠性亦可繼續增加。總之任何一組應用完畢時，即可得一結果。此外黑林還有一個節省時間的方法，就是在應用第二，第三，第四，第五各組時，不必個個測驗都應用，有的可以省去。看兒童在前一組或前數組中所得的分數，決定在本組中什麼測驗可以省去。譬如一個兒童在第一組中所得分數甚高，則在應用第二組時，有幾個容易的測驗就可省去不用，但仍給以分數。反之，若兒童在第一組中所得分數甚低，則第二組中有幾個很難的測驗就可以略去了。

(四) 非文字智力量表 比納式的智力量表都是以文字語言為內容的，施行測驗須用語言文字，被試者回答時亦用語言文字。在理論方面這類量表有一個很重要的限制，就是智力的表現不僅限於文字的或語言的行為，完全用文字的測驗去測量智力，恐祇能測量到智力的一方面，而不能得到智力的真諦。在實際方面，文字智力測驗對於有文字語言困難的兒童及聾啞的兒童都不能適用。因為這種理論上的缺陷和實際上的需要，所以有非文字智力量表的編造。非文字量表中以賓特納與派忒孫 (D. G. Pakenston) 所編造的作業量表應用最廣。

(五) 團體測驗的編造 比納式的智力量表都是個人測驗，每次祇能測量一個人。個人測驗在時間上的不經濟是顯明的事實。學校裏用智力測驗以為分班分組的根據，要在短時間內測量多數的兒童，個人測驗就很不適用。團體測驗同時可以測量多數人在時間上就節省多了。團體測驗的方法，由來已久，學校裏的考試都是團體測量，實驗心理學者亦多有用之者，但是促進團體智力測驗的成功者則為歐戰時美國軍隊中的測量工作。當

美國加入協約國出兵歐戰時，要在短時間內組織成功一個大軍隊，而軍隊中完善的組織要在人當其事，事事得人，要做到這一層，必須依個人的能力以選擇新兵，分配工作。當時美國軍隊就請了許多心理學者主持其事。考查個人能力的方法莫善於智力測量，但是要在短時期內測量一百多萬兵士，個人測驗當然不適用，因此就引用團體測量的方法。當時渥鐵斯 (A. S. Otis) 與推孟已有團體智力測驗的編造，軍隊心理學家就以此為根據編造成第一個團體智力測驗，就是所謂陸軍智力測驗 (Army Alpha Test)。

最初心理學者對於團體測驗多所懷疑，以為團體測驗的結果必不可靠，待團體智力測驗在軍隊中應用成功以後，團體測驗的價值乃為大家所公認，不復如先前的懷疑了。

問 題

1. 個別的研究對於測量運動有什麼影響？
2. 塔爾他們所應用的測驗為什麼不是好的智力測驗？
3. 高爾登對於測量運動有什麼貢獻？
4. 第一個年齡量表是什麼人編造的？
5. 某生實足年齡八歲四個月，智力年齡五歲十個月，試求其智力商數。

教 育 測 驗

9. 王書林, 智力測驗之發達史, 教育雜誌第十九卷三號。
10. 王書林, 比納西摩章量表發表後之智力測驗, 教育雜誌第十九卷第十一號。
11. 張耀翔, 智力測驗錄述, 心理一卷一號。
12. 張耀翔, 科學爲智力測驗, 北京晨報十五年十二月十六日副刊。
13. 廖世承, 智力測驗的歷史, 心理一卷一號。
14. 左任依, 法國心理測驗略史, 測驗副刊號。
15. 黃龍先, 薛登曾, 智力測驗史略, 測驗副刊號。

第五章 教育測驗發展的歷史

教育測驗與智力測驗所測量的對象雖然各異，但是所根據的原則及方法，却是相同的。所以一方面的發展自然而然的影響到他方面。我們將教育測驗與智力測驗發展的歷史分章討論，不過為敘述便利起見，我們不可不明瞭這二者的發展有交互的影響，並非各不相涉的。

一 美國教育測量運動的發展

(一) 菲奢的量表集——標準教育測驗的編造雖然為教育中較近的發展，但是標準測量的觀念可說由來已久。在一八六四年，英國 Greenwold Hospital School 裏有一位教師，名叫菲奢 (George F. Fisher) 曾搜集許多學生成績樣本，編成量表集 (Scale Book) 一書，用以為度量教育成績，學生作品的標準。書中備有各科學生作品的樣本，並為每一樣本評定一種分數，以示優劣。評判某學生某科之作品時，可將其作品與量表集中的各樣本互相比較，以求得與某生作品優劣相等的樣本，此樣本的分數即為某生應得的分數。如此則評定分數時有標準可循，不致漫無依據。菲奢編集量表集時僅憑個人的主觀以評定樣本的分數，在手續上當然不客觀，不精密，但是他所採用的方法與後來的書法量表，作文量表編造的方法大體上是相同的。可惜菲奢的量表集沒有得到當時教

育者的注意，因此對於教育測驗的發展，並沒有若何的影響。

(二) 邁恩關於考試方法的討論——美國大教育家邁恩(Horace Mann)在一八四五年討論口試與筆試的利弊，所論極見精透，與現在的標準測驗所依據的原則，觀念頗多吻合之處。邁恩提出許多理由以示劃一的筆試與個別的口試的利弊。

1. 劃一的筆試比較個别的口試公平些，因為與試兒童可以受同樣的試題，不致有難易的不均。他說：「劃一的筆試不但可以使同一學校內與試的學生受相同的試題，並且可以使各學校的學生受相同的試題。因此同一學校內的學生可以得着一個公允的比較；各學校可以受着同一標準的測量。」

2. 劃一的筆試比較個别的口試時間上經濟些。用一點鐘的時間去考試三十個兒童，若用口試的方法，每個兒童祇受到兩分鐘的試驗；用筆試的方法，每個兒童可以有六十分鐘的試驗。

3. 劃一的筆試比較個别的口試可靠些，因為試驗時間愈長，試題愈多，則考試的結果愈可靠。邁恩說：「對於一個學生所發問題愈多，則測量他的程度的機會愈好。假使祇有一個問題，全校中最好的學生亦許回答不出，雖然後面二十個問題他都能夠回答得一點不錯；或者學校中最壞的學生亦許答對了，雖然對於其他二十個問題是一定答不出的。每個問題是一個不完全的試驗，所以問題愈多，則試驗愈趨於完善。擲一次骰子，那一面向上是很難預測的；擲的次數愈多，則各面向上的次數愈趨於相等。」

4. 口試容易引起臨場的慌亂，因此減少考試的可靠性。

邁恩的討論對於測驗運動的發展雖然沒有發生若何的影響，但是從上述各點看來，可見邁恩早已感覺到劃一的試驗的需要，並富有測驗的思想了。

(三) 來斯的貢獻——在十九世紀的末年，能力心理學尚為一般心理學者所深信，而美國教育界人士尚株守於所謂學科的訓練的價值的學說。學校教員尤為守舊。讀法，算法，拼法，書法等科目差不多占學校課程之全部，而教授的方法又專注重練習與背誦。當時教育界中思想新穎者對於此種現象極為不滿，時倡改革之論。同時學校以外人士對於學校的株守成墨不知改進，亦多所攻讐。因此「新的」科目如手工，縫紉等實用的學科亦漸漸的加入了小學的課程。一般守舊的小學教員及教育學者極力反對實用學科的加入，他們說新的功課一加入，學生就沒有工夫去學習舊有的基本的科目了。雙方爭論甚烈。

在這雙方爭論之中，來斯(J. M. Rice)想到一個解決這個爭論的客觀的方法。他主張用劃一的試驗去考查教授拼法時間多少不同的學校，然後比較各校學生拼法的成績，以決定教授時間的多少與學生的成績有無關係。因此他選定了五十個字，到各學校裏去試驗，同時調查各校每週教授拼法的時間。他的調查的結果表示教授時間的多少與成績的優劣沒有什麼關係。八年之中每天用十五分鐘學習拼法的學生，其成績並不壞於每天用四十分鐘學習拼法的學生！

來斯的結果發表以後，教育界大為震動，許多人說來斯的結果是不可能的，大多數人對來斯的方法及結果都表示深切的懷疑。但是來斯的工作已經引起了少數有思想的教育家的注意與贊同，他對於測驗運動的貢獻是永遠不會磨滅的。來斯後來又編造算術和語言科的測驗。

(四) 桑戴克的貢獻——來斯並沒有編造過現在所謂的標準測驗，他最大的貢獻就是在用客觀的方法去解決教育上的爭論。美國教育測驗運動中的中心人物應推桑戴克。愛里斯 (H. P. Ayres) 曾說過：「我們既稱來斯為教育測驗的創始者，則對於桑戴克應稱之為教育測驗運動的鼻祖。」

一九〇四年，桑氏出版心理的與社會的測量一書，介紹統計的方法以及編造測驗的基本原理。在一九一三年，該書復經一次的修正，至今仍為測量學統計學中重要書籍。

一九〇九年冬，美國科學促進會在波士頓開會，桑氏即在教育組提出其所編製的書法量表，次年三月又發表於哥倫比亞大學師範學院月刊。這個量表為第一種用科學方法造成的教育測量工具，為測量運動中極重要的事件。桑氏在編造書法量表時引用喀泰爾之「等距原理」，後來的作文量表圖畫量表等都是根據這個基本原理而編造的。

(五) 其他的標準測驗——一九〇八年，斯東 (C. W. Stone) 編造一個算術推理測驗，要算是一種最早的標準測驗。桑氏書法量表發表後兩年，希來格司 (M. B. Hillegas) 編成一種作文量表，不久白根漢 (B. F.

Buckingham) 又編成一種拼法量表。這兩種量表的編製，都是受桑氏的指導的。後來，愛里斯亦編成一種書法量表，其方法與桑氏所用者微有不同。愛里斯後又編造一個拼字量表，比以前所編造的更見進步。克底斯 (S. A. Counts) 的算術測驗亦為早期標準測驗之一。他曾用於紐約市學校調查，很引起當時教育界之注意。自一九一五年以後，量表和測驗的編造，日見衆多，但直接或間接莫不受桑戴克的影響。

(六) 學校調查和教育測驗——美國學校調查運動起於一九〇七年的彼斯堡調查，但是到一九一一年紐約市調查，才引用標準的教育測驗。當時克底斯為調查委員之一，用自己所編造的算術測驗，測量了三萬多個兒童。自此以後從事學校調查者都以標準測驗為必需的工具。較大城市如紐約 (一九一三)、鮑爾鐵末 (Baltimore 一九一二)、波士頓 (一九一四) 等先後設立測驗研究部，聘請測驗專家，以進行大規模的教育測量。所以學校調查運動對於測驗的發展和應用的推廣是有很大的影響的。

(七) 較近的發展——以上所述都是測驗運動早期——大概在一九一八年以前——發展的情形。從一九一八年到現在，測驗運動的進展真有一日千里之概。其間重要的事實，約有數端：

1. 「測驗彙集」的編造 一九一八年以前所編造的教育測驗，都是單獨學科的測驗，如算術測驗，默讀測驗，拼字測驗等。後來乃有所謂「測驗彙集」的編造。所謂「測驗彙集」就是由數種學科測驗組織而成的測驗，有時亦包含智力測驗在內。這種測驗彙集可以為普通教育調查之用，在選擇測驗，施行測驗，綜合分數各方面

都比較單獨學科測驗為便利。測驗彙集之最早者為賓特納在一九二〇年發表之教育調查測驗。一九二二年出版之斯旦福大學學業測驗應用最廣。

2. 成業商數的引用 智力年齡表示個人的智力程度，教育年齡表示個人的教育程度；智力商數表示個人與同年齡兒童的平均智力的比較的程度，教育商數表示個人與同年齡兒童的平均學力的比較的程度。這幾種數量都是極重要的。但是有時我們要考查某生的學力是否與其智力相稱，以推斷學生是否努力或教授效率的高低。這個問題却不是以上各數量所能回答的。我們必須直接比較一人的智力與學力，才可以回答智力與學力是否相稱的問題。賓特納在一九二一年提議以「智力指數」與「教育指數」的差別來表示這種關係。在一九二〇年白根漢與孟祿(W. S. Monroe)和弗倫純(R. Fraumeni)先後發見成業商數的數量。成業商數是教育年齡與智力年齡的比例，其計算的公式如下：

$$\text{成業商數} = \frac{\text{教育年齡}}{\text{智力年齡}} \times 100$$

3. 中學各科測驗的編造 初期的教育測驗差不多完全限於小學課程內各學科，測驗的應用亦限於小學，中學及大學還沒有受着測驗運動的影響。一九一八年以來，中學各科的測驗乃逐漸發展，測驗運動的影響亦推廣到中等教育和高等教育階段了。

4. TROFF 制的發明 在一九一八年以後還有一件事足記述者，就是麥柯爾的 TROFF 制的發明。TROFF 制的意義將在第十八章內詳細解釋，暫時且不討論。

以上敘述教育測驗的起源和發展，僅限於美國方面，當然不免遺漏的地方，但是教育測驗以美國為最發達，所以從上面簡單的敘述，讀者亦可以知道一些關於教育測驗發展的重要史實。到現在中小學各種科目莫不有相當的測驗可以應用，教育測驗已成爲學校行政上教學上不可少的工具了。據孟祿在一九二八年的估計，美國的標準測驗總數當在一千三百左右，其中大約有一百五十種是智力測驗。教育測驗中以算術、讀法、英文及歷史各科目的測驗爲最多。美國每年所用測驗的份數沒有準確的統計，據孟祿的估計，其數當不在三千萬之下，其中大概四分之一屬於智力測驗，四分之三屬於教育測驗。美國測驗事業的發達亦可見一斑了。

二 我國測量運動發展的情形

民國七年，俞子夷曾仿造一種小學國文毛筆書法量表，可算爲我國最早的測驗。民國八九年間各大學教育科有測驗科目的設立，各雜誌中關於測驗的文字亦時有發表。至十一年夏，中華教育改進社聘請美國教育測量專家麥柯爾博士來華，主持測驗編造事宜，當時協助進行者，有北京師範大學、北京大學、燕京大學、北京女子高等師範、東南大學等校教授和學生，計共編成測驗四十餘種。

我國編造的測驗，其最重要的特色，就是一致採用麥柯爾所倡的 EBOD 制，使由各測驗所得的分數可以互相比較，應用時亦較為便利。麥柯爾曾有一段話，評論中國各種測驗的價值：

「現在讓我將上列各種測驗，下一個價值的評判。倘若這價值的評判是一種比較的，那末我們應該將各種測驗先與美國最相彷彿的測驗相比較。因為每種心理測驗，都在美國達到最高的發展，而遠駕乎世界各國之上。我們如用這個標準來評判我們在中國所造的測驗，我相信可以下這個結論：我們所造的各種測驗，至少都與美國的標準相等，有許多種竟比美國的為優。然這也不足為榮，因為對於一種繼承物，個個人應該感激，個個人應該想法子去改良。」

但是在十年前所編造的測驗，到現在有許多地方已不適用。課程標準十年來已有很大的變更，舊有的教育測驗似有根本訂正的必要；以前求得的常模，亦未必合於現在學生的程度，亦有重新訂定的必要。總之中國的測驗事業以前的根基下得很好！將來的發展，完全要看我們能否繼續的在方法方面加以改進，在內容方面加以充實，在應用方面加以推廣。這些都是我們從事教育者的責任了。

我國測驗運動中還有一件值得注意的事情，就是民國十二年冬所舉行的全國小學調查。該調查是由中華教育改進社聘請查良劍、德爾滿（E. L. Townan）二氏主持的，共費時三月，到了二十二個城市和十一個鄉鎮，測量了九萬二千個兒童。這個大規模的調查頗引起教育界對於測驗的注意。

二十年六月二十一日，國內研究測驗學者組織之中國測驗學會正式成立，設會所於首都，其宗旨為研究測驗學術，推行測驗方法。該會復有測驗雜誌之刊行。

問 題

1. 比較評審的、量表與現行的品質量表。
2. 劃一的筆試何以比較個別的口試為公平可靠？
3. 試述成業商數的意義。
4. 某生教育年齡12歲4月，智力年齡13歲8月，試求其成業商數並解釋之。
5. 為什麼十年前所造的測驗不很適用於現在的學校？

參考書報

1. Kunda, I. I. (Editor): Twenty-Five Years of American Education, Ch. 5.
2. McCall, W. A. "Scientific Measurement and Related Studies in Chinese Education", Journal of Educational Research, Vol. XI, nos. 2, 3, Feb. and Mar., 1925.

- 3^o Monroe, W. S. Theory of Educational Measurements, Ch. I. Houghton Mifflin.
- 4^o Monroe, W. S. and others, Ten Years of Educational Research, 1913—1927. Part I, Ch. IV. Bureau of Educational Research, University of Illinois, 1928.
- 5^o Tennant, E. I. The Efficiency of Elementary Schools in China. Commercial Press, 1924.
- 6^o 臺灣教育研究會編輯會第一卷第一期
- 7^o 臺灣教育研究會編輯會第二卷第一期
- 8^o 臺灣教育研究會編輯會第三卷第一期

第六章 測驗的種類

一 以測量對象論

以測量對象論，測驗可別為四大類：

(一) 智力測驗 智力測驗所測量的究竟是什麼？智力的性質如何？定義如何？這些問題我們將在第七章內比較詳細地討論，現在暫且不述。智力最普遍的定義是：智力乃學習的能力，智力乃適應環境的能力。

桑戴克曾把「智力」分為三種：1. 抽象智力，即運用抽象觀念的能力；2. 機械智力，即運用實在物件的能力；3. 社交智力，即與人相處的能力。桑戴克這樣的分類不過根據普通的觀察，並沒有客觀的證據的。我們承認運用抽象觀念，運用實在物件，和與人相處的能力是不同的，但是我們並不一定要假設這三種智力之存在。

普通的智力測驗所測量的大概是屬於第一類的能力，即運用抽象觀念的能力。斯丹窺斯脫 (J. I. Stenquist) 的機械能量測驗，顧名思義，是測量機械能力的。社交能力測驗據作者所知，祇有美國的喬治華盛頓大學社交智力測驗。

(二) 教育測驗 教育測驗所測量的乃學業的成績。教育測驗又可分為測驗彙集與各科測驗二類。測驗

彙集在同一測驗之內包含許多學科的測驗，如 *Stanford Achievement Test*, *Pressey Scale of Attainment* 是各科測驗又可依其科目分為若干種，如識字測驗，讀法測驗，默字測驗，文法測驗，算術測驗，代數測驗，幾何測驗，歷史測驗，地理測驗等等。

(三) 品格測驗 普通的觀察可以使我們相信，除了智力以外，個人間品格上的差別亦是很大的。一個人的品格，一部分基於天賦，是與生俱生的；一部分由於環境，是學習而來的。品格的陶冶乃教育中重要目標之一，所以我們一方面對於兒童天賦的品格上的趨向必需有正確的認識，一方面對於兒童現有的品格上的行為尤不得不隨時加以考查。由此說來，品格測驗乃施行品格教育極重要的工具了。

品格一名詞的涵義，所包括的範圍，心理學者並無一致的，嚴格的規定，我們用以代表智力以外其他各種心理的特性。品格測驗，依費里門 (F. N. Freeman) 的意見，可分為下列四類：

1. 意志性情測驗 (Tests of will temperament.)
 2. 情緒和興趣測驗 (Tests of emotional temperament and interest.)
 3. 德性測驗 (Tests of moral disposition.)
 4. 美感測驗 (Tests of aesthetic sensibility.)
- (四) 特殊能力測驗 特殊能力的存在，大多心理學家都承認的。吳偉士 (R. S. Woodworth) 相信除了

天賦的感覺、情緒及反應以外，天賦的稟性又包括特殊的能力。普通的觀察表示人們各有所長，各有所短，假使訓練的機會相等，則所長方面可以表示特殊能力之所在，所短方面可以表示特殊缺陷之所在。特殊能力的分析與測量，現在尚在萌芽時代。據現在所知，有兩種能力與普通智力及其他能力相關極低，可以稱為特殊能力。這兩種能力就是音樂能力與圖畫能力。西旭 (O. H. Serfaty) 對於音樂能力會有極精密的分析並已編成數種可靠的測驗。

二 以試驗方法論

以試驗方法論，測驗可別為個人測驗與團體測驗兩大類。茲將個人測驗與團體測驗的優劣簡括討論如下：

(一) 個人測驗

1. 個人測驗最大的利益就是主試者對於被試者在測量時的行為有真切觀察與控制的機會。主試者在個人測量的情形之下，有充分的機會得到被試者的合作並引起其最大的努力。有時，個人測驗的結果不但可以使主試者斷定被試者智力的高低，並且對於他的性情、言語習慣、情緒的反應等亦可以有相當的診斷。
2. 個人測驗最大的限制就是時間上的不經濟，因為每次祇能測量一個人。
3. 施行個人測驗比較施行團體測驗來得困難。因此，用個人測驗時，主試者必須是訓練有素者方能勝任。

(二)團體測驗

1. 團體測驗最大的利益就是時間上的經濟。
2. 用團體測驗時，主試者對於被試者的行為不能有真切的觀察與控制，因此所得結果的可靠度比較低。各種測驗結果的解釋都依據同一的假說，就是被試者已盡其力之所能去回答測驗上的問題。但是在舉行團體測驗時有許多的原因可以使上述的假說不能成立，如被試者的鉛筆斷了，被試者身體不安適，被試者不肯合作等等。所以為個人的精密的診斷起見，個人測驗是比較妥當。

三 以內容論

以測驗的內容或材料論，測驗可分為文字的與非文字的兩類。

大多數智力測驗都是文字的測驗。其原因約有兩點：第一，第一個智力測驗——即比納西門智力量表——的內容大部分是文字的，以後所編造的測驗都以此為範，因此所用材料亦相似。第二，編造測驗的人大都是教育界人士，因此測驗中所用材料多偏於文字的一方面。

這種偏於文字方面的智力測驗在理論上受許多人的攻擊，說牠所測量的範圍過於狹小；在實施上又有許多的限制，如對於未受文字訓練的人這種測驗就不適用。因此就有非文字智力測驗的編造。至於教育測驗，當然

偏重文字。

四 以功用論

以功用論，測驗有許多可能的分類法，如預測測驗與作業測驗之別，難度測驗與速度測驗之別，普通測驗與診斷測驗之別。茲依次討論之。

(一) 預測測驗與作業測驗 預測測驗的功用，在推測某人在某方面將來的成功，作業測驗的功用，在考查某人在某方面現在的成績。這兩種測驗的功用既異，所用材料亦因之而不同。譬如編造打字的作業測驗，我們祇須測量被試者在規定時間內能打多少字，有多少錯誤，或測量被試者打一篇標準的文字須多少時間，有多少錯誤，就可以評定其打字的能力。編造一個預測的打字測驗，就不能如此了，因為我們的目的，乃是在他們未曾學習打字以前推測其將來的成功，所以我們祇能測量其獲得打字能力的可能性，而不能測量其打字的實際能力。我們必需先對於打字能力做一番分析的功夫，將打字這種作業所包含的基本能量一一分析出來，如手指動作的速率，動作的準確度等。然後根據分析的結果編造一種測驗以測量這些基本的能量，這種測驗才可以稱為預測的測驗。由此看來，預測測驗的編造實在比較作業測驗困難多了。

預測測驗與作業測驗的區別既如上所述，但是有時作業測驗的結果亦可用以比較天賦的能量。假設二人的

訓練相等，則實際能力的差別就可以間接表示天賦能量的差別了。

智力測驗實在是一種預測測驗，因為一人的智力商數可以預示他將來智力的發展。普通的教育測驗大多屬於作業測驗一類。如前面所述斯丹窺斯脫的機械能量測驗，西旭的音樂能量測驗都是預測測驗的例子。教育測驗中的預測測驗有羅傑斯的數學能量測驗 (Rogers' Test of Mathematical Abilities) 及 Iowa Placement Examination, Aptitude Series 等。

(二) 難度測驗與速度測驗 難度測驗測量程度的高低，速度測驗測量作業的速率。以測量算術能力為例，難度測驗的目的在測量某人能回答何等難度的題目，速度測驗的目的乃在測量某人對於某一等難度的題目，在規定時間內能回答幾個。以跳欄為例，難度測驗好像一行高度不一的欄，由低到高排列起來，某人所能跳過的最高的高度就是他的分數；速度測驗好像一行高度相等的欄，某人在規定時間內所跳過的欄的數目就是他的分數。

所以在編造速度測驗時，所有的題目難度須相等；時間的限制必須從嚴，使做得最快的人亦不能在規定時間內做完。在編造難度測驗時，所用的題目須難易不等並依其難度而排列；時間的限制須從寬，在理論上講，最好沒有時間的限制。

速度測驗的記分有兩種方法，一種是以時間為限制，以工作為單位；一種是以作業為限制，以時間為單位。譬

如測量看書的速度，用第一種方法，我們可以計算在規定時間內能看完幾行，行數或字數是計算的單位；用第二種方法，我們可以計算看完一篇規定的文字須多少時間，分鐘或秒鐘是我們計算的單位。這兩種記分方法都可以應用，不過在舉行團體測量時，祇能用第一種，第二種就不大適用了。

(三)普通測驗與診斷測驗 普通測驗或調查測驗的功用，在考查一人在某方面大概的程度；診斷測驗的目的是進一步去診斷一人在某種能力上的特殊優點與缺點之所在。所以從教學的立場而論，診斷測驗的功用比較普通測驗大多了。舉行診斷測驗的第二步就是施行補救教學。所以編造診斷測驗者同時必須預備練習的材料，並指示施行補救教學的步驟。原來測驗本身不能改進教學，測驗的功用乃在表示學生現在的程度並其缺點之所在，教師必須根據測驗所顯示的情形以圖改進，才能使測驗在教學上發生相當的影響。

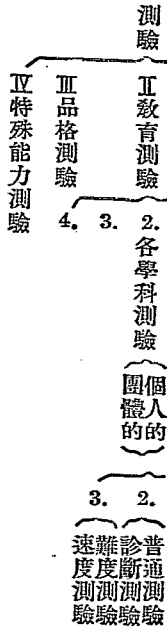
編造普通測驗在選擇材料時，可用取樣的方法。編造診斷測驗，必須先對於所測量的能力做一番詳細的分析，然後將各種步驟，過程一一編入測驗之內，手續上就比較繁重了。

上面所述測驗的種類，可歸納成下表：

I 智力測驗
個人的
團體的
文字的
非文字的
非文字的

1. 預測測驗
作業測驗

第一編 測驗之一般的理論



問題

1. 批評桑戴克對於智力的分類。
2. 比較個人測驗與團體測驗的利弊。
3. 預測測驗的編造何以較作業測驗為困難？
4. 速度測驗的計分法有那幾種？
5. 以作業為限制，以時間為單位的計分法，為什麼不適用於團體測量？
6. 普通測驗與診斷測驗在材料的選擇上有什麼不同之處？

參考書籍

1. Greave, H. A. and Jorgensen, A. N., The Use and Interpretation of Educational Tests, Ch. II, Longmans.
2. Symonds, P. M., Measurement in Secondary Education, Ch. XVII, Macmillan.

第二編 各科測驗與新式考試

第七章 智力測驗

智力測驗可分爲個人的與團體的，文字的與非文字的，已如第六章所述。在本章裏我們先列舉我國已有的各種智力測驗，然後對於智力一概念作一簡括的討論。

一 我國現有的智力測驗

(一) 陸氏訂正比納西門智力測驗 這是一個個人測驗，由陸志韋先生在民國十一年主持編訂。編訂時所測量的學生計約一千四百人，男女各半，年齡從三歲起至二十歲爲止，學級從幼稚園起至高小爲止，且有少數中學生。

1. 內容 本測驗共包含六十五個測驗，有些是創始者向來有的，有些是從各家修正的量表裏採來的，其餘是新加入的。測驗的性質大概可說是仍前人之舊，一大部分依舊用文字，且用問答體。這六十五個測驗分爲兩部，

預備測驗十一，正測驗五十四。茲將這六十五個測驗列后：

預備測驗

- (1) 解說圖畫
- (2) 說出物名
- (3) 解說物名
- (4) 對答問句
- (5) 說錢幣名
- (6) 方形分析
- (7) 順背數目
- (8) 倒背數目
- (9) 方塊叩擊
- (10) 有意義的記憶
- (11) 詞句重組

正測驗

- (1) 指出身上各部
- (2) 比較線的長短
- (3) 數銅元四枚
- (4) 說出自己的姓
- (5) 辨別形式
- (6) 說出自己的年齡
- (7) 數銅元十三枚
- (8) 摹畫方形
- (9) 說明性別
- (10) 三種指使
- (11) 問手指數
- (12) 早上與下午
- (13) 倒背數目
- (14) 拼三角形
- (15) 美惡的比較
- (16) 簡單的迷津
- (17) 心算
- (18) 打紙包
- (19) 指出圖中缺點
- (20) 分明左右
- (21) 說出日期

- | | | |
|---------------|--------------|--------------|
| (22) 摹畫菱形 | (23) 說明字義 | (24) 找尋失物 |
| (25) 數字巧算(簡單) | (26) 正確答案 | (27) 找尋圖樣 |
| (28) 找尋數目 | (29) 歸納測驗 | (30) 說出二物的分別 |
| (31) 一筆畫 | (32) 對較 | (33) 指出謬誤 |
| (34) 插畫圖樣 | (35) 剪紙 | (36) 填數目指使 |
| (37) 說出二物的同點 | (38) 匣子計算 | (39) 複雜的迷津 |
| (40) 皇帝與總統 | (41) 比喻 | (42) 填字(簡易) |
| (43) 數學巧算(複雜) | (44) 說出三物的同點 | (45) 時辰計算 |
| (46) 心算 | (47) 解決問題 | (48) 解說抽象字 |
| (49) 填字(複雜) | (50) 分別抽象字 | (51) 幾何形分析 |
| (52) 巧算 | (53) 理解問題 | (54) 數立方體 |

2. 測驗的次序

(1) 先用預備測驗,後用正測驗。預備測驗從第一個試起至第十一個為止,得一總分數,然後從表二查出相當H分數。例如分數為35,則H分數為12。

表一 陸氏訂正比納西門智力測驗預備測驗的百分數

對	對	對	對	對	對	對	對	對	對	對	對	對
1—30	16	—10	31	5	46	27	61	50	76	72		
2—27	17	—10	32	6	47	30	62	51	77	73		
3—25	18	—9	33	7	48	33	63	53	78	75		
4—23	19	—8	34	9	49	35	64	54	79	76		
5—22	20	—6	35	12	50	36	65	56	80	79		
6—21	21	—5	36	13	51	37	66	58	81	82		
7—19	22	—4	37	14	52	38	67	59	82	83		
8—18	23	—3	38	15	53	39	68	60	83	87		
9—17	24	—2	39	16	54	40	69	61	84	89		
10—16	25	—1	40	18	55	42	70	62	85	92		
11—15	26	—1	41	19	56	44	71	64	86	95		

12	—14	27	0	42	20	57	45	72	66	87	100
13	—13	28	2	43	21	58	47	73	68		
14	—12	29	3	44	24	59	48	74	69		
15	—11	30	4	45	26	60	49	75	70		

(2) 倘使時間短促，那末用了預備測驗，不用正測驗也可。但因測驗數目太少，所得結果難以正確可靠，所以除非時間逼促，千萬不可就此停止。

(3) 預備測驗試驗完畢，求得H分數後，接用正測驗。根據從預備測驗所得H分數，決定從那一個正測驗試驗起。決定的手續如下：

a. 照預備測驗H分數退下十分。如預備測驗H分數為12，退下10分，得2分。

b. 翻到正測驗標明H II 3的那一個測驗（正測驗十三），即從該測驗試起。倘沒有測驗恰巧等於此H分數，再退下一個。（各測驗所標明的H分數表示測驗的難度。）

(4) 以前的正測驗不必試驗，也算通過。

(5) 正測驗連續五個不通過，不必再往下試。

3. 記錄的手續

(1) 在測驗成績紙上先記錄被試者姓名、年齡等項目。

(2) 預備測驗每個記分數，正測驗每個記通過與否。通過的在成績紙上相當括弧內記「十」號，不通過的記「一」號。

(3) 兒童的答案須逐字逐句詳細記錄在成績紙上。在幾個測驗上，兒童的答案很難臨時評定，必先將答案記錄下來，再與標準答案相比較，以定其通過與否。（標準答案見說明書。）

(4) 在成績紙的末頁右下角記測驗開始與終了的時間。

4. 計算總分數的方法

(1) 把預備測驗的分數除三，祇算整數，如有餘數，四捨五進。這就是預備測驗應得的分數。

(2) 正測驗每通過一個算一分。未試驗而作為通過者也每個算一分。

(3) 前二項相加，得一總數。

(4) 從表三查出全部測驗的日分數。

表三 陸氏訂正比納西門智力測驗全部測驗的日分數

對	T	對	T	對	T	對	T	對	T	對	T	對	T
1	-30	15	-11	29	10	43	32	57	42	71	71	71	71
2	-28	16	-11	30	11	44	33	58	53	72	72	73	73
3	-27	17	-10	31	12	45	35	59	54	73	73	74	74
4	-26	18	-8	32	13	46	37	60	55	74	74	75	75
5	-24	19	-6	33	14	47	38	61	56	75	75	76	76
6	-23	20	-4	34	16	48	41	62	57	76	76	77	77
7	-21	21	-2	35	18	49	42	63	60	77	77	78	78
8	-19	22	-0	36	19	50	42	64	62	78	78	79	79
9	-17	23	-2	37	20	51	43	65	64	79	79	80	80
10	-16	24	-3	38	23	52	45	66	65	80	80	81	81
11	-15	25	-4	39	25	53	47	67	66	81	81	84	84
12	-14	26	-6	40	26	54	48	68	67	82	82	92	92

表11 關於 各政黨選舉票的分配

表中

13	—13	27	7	41	28	55	50	69	68	83	100
14	—12	28	9	42	31	56	51	70	70		

本測驗自編訂以來到現在已經十年多了，有許多地方已不適合現在的情形，聽說陸志章先生正從事修訂，我們深盼他修訂的結果能早日公布。

(二) 廖氏團體智力測驗 本測驗是廖世承氏參照美國「國家智力測驗」編造的，分甲乙兩個量表，每個量表包含五種測驗。每種量表測驗的時間，大致在四十分鐘左右。自小學初級三年級起至初級中學二年級止，這個測驗都可適用。

量表甲包含下列五種測驗：1. 算術理解題，共有15題，由淺而深。2. 填字，共有21句，亦由淺而深。3. 理解的選擇，共有8題。4. 同——異測驗，共有8題。5. 形——數測驗，共140個圖形。

量表乙亦包含五種測驗：1. 算術演習題，共有8題。2. 常識測驗，共有8題。3. 字彙測驗，共有8題。4. 比喻測驗，共有8題。5. 校對，共有8題。

關於本測驗的詳細情形請閱說明書。

(三) 劉氏中學智慧測驗 劉廷芳先生曾編有中學智慧測驗兩類。標準已求得，惟尙未發表。兩類測驗的

種類完全相同，每類包含下列十種測驗：1. 見聞，2. 最好的答案，3. 字義，4. 推理的選擇，5. 算術，6. 句義，7. 類推，8. 雜句，9. 類別，10. 數系。

(四) 陳氏圖形智力測驗 陳鶴琴氏編有小學初級和高級用的圖形智力測驗兩類。這是一個非文字測驗，完全以圖畫為內容。

第一類備小學初級之用，包含下列五種測驗：1. 謬誤，叫兒童將圖中謬誤之處劃出；2. 填圖，叫兒童填圖中缺乏的東西；3. 劃圈，叫兒童將圖中多餘之圈劃出；4. 拼圖，每圖中有三個零件，其中兩件可以拼成一樣東西，其餘一件無關係，須劃去。以上四種測驗，每種均有 5 題。5. 較圖，這是一個同異測驗，共 5 題。

第二類備小學高級之用，亦包含五種測驗：1. 填圖，與前同，惟比較難些，共 15 題。2. 分類，每圖中有五樣東西，其中有四樣東西，各自成雙，祇有一樣，不歸入類，須劃去，共 5 題。3. 拼圖，兒童須在五個不完全的圖裏邊，選擇四個拼成一個完全的圖，把其餘一個劃去，共 5 題。4. 辨圖，叫兒童將每個方格中多餘一個圖劃去，共 30 題。5. 交替，共 20 題。

(五) 德爾滿非文字智力測驗 本測驗為燕京大學教授美人德爾滿 (E. L. Terman) 所編造。全測驗共 50 題，每題有五個圖，其中有一個圖與其他四個圖不同類且無關係，被試者須在此圖上畫一 X 號以標記之。時間限定三十分鐘。從初小三年級起至初中二年級止，本測驗均適用。

(六) 德爾滿機械的智力測驗 本測驗係仿美國 *Stenquist Mechanical Ability Test* 編成。全測驗共有二十欄，每欄分兩格，每格內有圖形五個，被試者須將右邊格中之圖形與左邊格中之圖形一一配合。測驗時間限定三十分鐘。從小學五年級起至高級中學三年級止均適用。

(七) 鮑德氏迷津測驗 鮑德氏 (S. D. Porteus) 的迷津測驗是一個年齡量表，由三歲至十四歲，每歲有一個測驗。據陶氏說，迷津測驗可以測量一人的敏捷、謹慎、持重、預計等心理原素，所以大部分為一種性格測驗，足以補比納式測驗之不足而有其特殊的價值。上海市教育局正從事於本測驗的訂正，使適用於我國兒童。

二 一個智力及智力測驗的界說

人類在智力上和天賦的差別這一個概念在智力測驗未發明以前久已成立了。孔子說「惟上智與下愚不移」已明白承認人類在智慧上的差別。柏萊圖 (Plato) 在他的名著共和國裏面亦明白表示人類能力之差異係基於天賦，並提出種種鑒別之標準。人有智愚之別既為古今中外所公認之事實，則智慧之高低當有方法可以測量之。

關於智力之性質及定義中外學者意見紛歧，我們在此處不能詳為討論，讀者可參閱章末所列參考書報。現在讓我們提出一個極平庸但亦是極基本的智力測驗的界說。

(一) 天賦的智力乃一種發展的可能性，必須賴有相當的環境，才能從行爲上表現出來。「智力」乃一個抽象的概念，不能直接觀察，直接測量的。

(二) 我們所能觀察，所能估計，所能測量的祇有人的反應，行爲。根據行爲的觀察以推斷智力的高低，這是常識的方法，亦是智力測量最基本的方法。

在智力測驗沒有通行以前，做父母的，做教師的，對於兒女學生的智力，不常常加以判斷，加以估計嗎？甲兒比乙兒聰明，甲生比乙生愚笨，這類的比較是我們所常常聽見的。試問父母教師根據什麼以下此評斷？我們可以肯定地說，他們所根據的乃是行爲的觀察。這種比較雖然遺漏錯誤的地方很多，但其基本方法——觀察行爲以推斷智力——是不錯的。現行的智力測驗亦採用這個方法，不過手續上較為精密些，審慎些而已。

(三) 要得到一個公平的智力評判，我們的評判應根據於一人全部的或多方面的行爲的觀察或根據於一人行爲的公允的取樣。觀察愈周密，愈完全，觀察的方面愈多，則根據此種觀察以推斷智力，結果愈可靠。評判智力如此，評斷他種能力亦是如此。這一點與測驗內容或材料的選擇極有關係。

(四) 智力的高低乃比較的說法。要比較兩人的智力，必須觀察他們在同樣情境中的反應或行爲，才算公平。測驗的作用乃在供給許多標準的刺激或情境，然後觀察人們對於這些刺激或情境的反應，估計其價值，以評斷其智力之高低。所以測驗不過是引起行爲的工具，以爲估價的對象，評斷的根據。因此我們下一個智力測驗的

定義：智力測驗乃是一羣標準的刺激，用以引起人們的反應或行為的。

(五) 一種行為或反應的價值，由心理學者依主觀而評定之。評斷所依據的標準，雖由主觀而定，但為劃一的，共同的，所以通常稱之為「客觀」的標準。譬如問兒童「倘使你打碎了別人的東西，你應當怎樣？」甲兒回答「賠償」，乙兒回答「不知道」或「哭」，我們說甲兒的反應比較聰明些，這是主觀的評斷。但是以後估計其他兒童的反應時，一律依此標準評斷之，主觀的評斷便變為客觀的標準了。所以在智力測驗方法中反應的估價，與普通評判一樣，基於主觀的見解，所不同者，前者的評斷的標準是劃一的，而後者的評斷的標準則不但各人間不同，即同一人異時所用的標準亦復不一。

(六) 智力測驗所測量的既然是人們的行為，而人們的行為都是獲得的，都是學習來的，都受環境的影響的。我們要從人們獲得行為上的差別來推斷他們天賦上的差別，祇有均等他們所受訓練，均等他們的環境的唯一方法。均等環境以後在行為上所表現的差別，我們可以指為天賦上的差別。所以智力測驗的一個根本應說：就是被試的人，都有同等機會，同等興趣去學習測驗中所包含的材料。

這個應說當然不能完全滿足，但是相對的滿足是可能的。我們編造智力測驗時所用的材料，應為人們日常經驗中所遇見的事件，或者採用多方面的材料，以均衡補償學習機會的參差。例如比納西門智力量表內的材料大都為人們日常生活中所遇見的事件。

(七) 因此我們不能有一個普遍的正確的智力測驗。一種測驗施用於某團體為一個極好的智力測驗，施用於他團體亦許完全不正確。例如文字智力測驗不能應用於不識字的人，美國的智力測驗不適用於英國幼童，城市中的智力測驗，或者不適用於鄉間兒童。應用智力測驗的人，應該根據被試的人們，選擇適當的測驗。

以上陳述對於智力測驗的一個很平庸的界說。

最後一點，就是從教育立場我們對於智力測驗究竟應該採用什麼態度？我們的見解亦是很平庸的。我們以為實際上應用智力測驗的人，對於智力的概念，定義，界說，智力係天賦抑獲得諸問題，有暇研究固好，無暇亦儘可不必深究。應用測驗的人是拿測驗來解決教學上，管理上的實際問題，他們對於智力測驗的態度，應以智力測驗能否幫助他們解決這些實際問題為準衡。例如根據智力測驗來分班分組，教學的效率是否增高？根據智力測驗來推測學生將來的成就，來指導他們求學的途徑，推測是否正確，指導是否有效？我們固不可盲目地過信智力測驗的價值，亦不宜一概抹殺智力測驗的效用。

問 題

1. 陸氏訂正比納西門智力測驗中的預備測驗有什麼功用？
2. 陸氏訂正比納西門智力測驗有那些地方已不適合於現在的情形？
3. 鮑德氏遠津測驗有什麼特殊的價值？

4. 父母與教師對於男女學生的智力的評判何以不能確不可疑
5. 「測驗乃引起反應的工具」這一句話的含義與關係
6. 如何才可以從複雜的行為以推斷天賦的差別

參考書目

1. Colvin, S. S.: Principles Underlying the Construction and Use of Intelligence Tests, in the Twenty-First Yearbook of the National Society for the Study of Education, 1922.
2. Freeman, F. N. Mental Tests, Ch. 18. Macmillan.
3. Finkner, R. Intelligence Testing, Ch. 4. Henry Hoet.
4. Finkner, R. The Individual in School: General Ability, in Chapter 17, The Foundations of Experimental Psychology, edited by Carl Murchison, Clark University Press.
5. Spearman O. Abilities of Man. Macmillan.
6. Terman, L. M. and others, Intelligence and its Measurement, a symposium. Journal of Educational Psychology, Vol. 12, 1921.
7. Thorndike, Ed, L.: Measurement of Intelligence Teachers College, Columbia University.

8. 張耀翔：智慧之定義及範圍，心理一卷二號。
9. 陳選善：我們對於智力測驗應該採取何種態度，教育建設第三集。
10. 孟憲承：智力測驗之爭論與教育學說，教育雜誌第十五卷一號。
11. 陸志革：訂正比納西門智力測驗說明書（商務）。
12. 廖世承：團體智力測驗說明書（商務）。
13. 陳鶴琴：小學圖形智力測驗說明書（商務）。
14. 德爾福：機械智力測驗說明書（商務）。
15. 上海市教育局第三科編：鮑德氏迷津智力測驗說明書。
16. 吳天敏：重訂比納西門智力測驗之經過，測驗第三期。

第八章 國文測驗

國文教學的目標不外發表能力之訓練與閱讀能力之培植兩端。故國文程度之考查亦可分為兩方面，即發表能力（即作文）與閱讀能力（即讀文）。至於字的正當寫法（默字及書法）、字義的了解（識字）、文字的正當應用（文法）乃閱讀能力與發表能力中主要的部分，亦應在考查之列。茲就默讀測驗、默字測驗、識字測驗、作文測驗、文法測驗、書法測驗數種分別討論之。

一 默讀測驗

（一）默讀與朗讀 閱讀可分為朗讀與默讀兩種。這兩種讀法以默讀的應用較大，因為吾人日常一切看報、看書、看信、看佈告等等都用默讀，朗讀的機會實在很少。據實驗結果，默讀的速度又較朗讀為快，所以在時間上默讀又較朗讀為經濟。因此近來學校裏讀法的教學注重默讀，朗讀已不如從前那樣着重了。

但是朗讀仍有他的地位，不應完全忽略。朗讀在初年級裏較為重要。當兒童入學時，已經熟悉了許多字音及其意義，朗讀之功用就在以字音為媒介使字義與字形發生聯結。發音的準確本身當然亦是一種很有價值的訓練。為增進欣賞美術文的能力，朗讀亦有相當功用。

(二) 默讀測驗 讀法測驗可別爲朗讀測驗與默讀測驗兩類。我國尙無朗讀測驗之編造，默讀測驗則已有數種。默讀能力的測量又可分爲了解與速率兩方面。

1. 陳氏初小默讀測驗 陳鶴琴氏所編初小默讀測驗共有兩類，每類包含四個單獨測驗。

測驗一，是測量兒童的辨字能力的。本測驗共 $\textcircled{6}$ 行，每行 $\textcircled{6}$ 字，其中祇有一個是字，餘者皆似是而非字，叫兒童將一個是字的圈出。時間限定四分鐘。

測驗二，是測量兒童用字的能力的。本測驗共有 $\textcircled{6}$ 句子，每句多一個字，叫兒童把那一個多的字圈起，使成爲有意義的詞或句。時限爲五分鐘。

測驗三，是測量兒童了解字義或句義的能力的。本測驗計有 $\textcircled{6}$ 行，每行的左邊有一個字或一句子，右邊有四個圖，其中有一個是指點那個字或句子的，叫兒童將這個圖圈出。時限爲四分鐘。

測驗四，是測量兒童的造句能力的。本測驗也有三十行，每行左邊有一句話，裏面缺少一個字，右邊有四個字，祇有一個是左邊句子裏所缺少的字，叫兒童將這一個字圈出。時限爲五分鐘。

2. 陳氏小學默讀測驗 陳鶴琴氏所編之小學默讀測驗共有五類，可以替換使用，每類都有十篇文章，由短而長，由淺入深，每篇文章下面有三四句問句，每句問句後面有四個答案，其中祇有一個是對的。被試者須將對的答案選出，將對的答案的數目，寫在另外一張答案紙上相當括弧裏面。本測驗是一個難度測驗，了解與速率並顧。

時限爲20分鐘。從初小三年級起至高小三年級止本測驗都適用。

據編者自己所述，本測驗有五大特點：

「(1)方法簡便，只要將對的答案的數目填出，不必將答案寫出。因爲一要寫字便有寫得快慢的關係，不是純粹的看書能力，而且校閱不便當，計分不正確。

(2)機遇少，因爲有四個答案。如果只有兩個答案或用 Yes No (打正負號)的方法，那就機遇多了。

(3)錯的答案的主要字，皆是文章上所有的，這樣可以免去兒童的亂猜。

(4)校閱便當，用標準和做的紙條比較即得，計分也是很正確的。

(5)用紙條寫答案的方法是創作的的方法，試卷可以不止用一次，校閱又非常的便利，實在是最經濟的方法。」

3. 陳氏中學默讀測驗 本測驗共有兩類，每類有十篇文章。其編製方法與格式與陳氏小學默讀測驗同，惟文章篇幅較長，內容較深耳。

4. 艾氏國文理解能力測驗 艾偉氏曾編有文言和白話默讀測驗，方法與格式與陳氏默讀測驗相類。

5. 華氏新學制國語教科書閱讀測驗 本測驗共有兩類方式與陳氏初小默讀測驗完全相同。編者華超氏。

二 默字測驗

(一) 字彙研究 編造默字測驗第一個問題是「用什麼字去測量學生？」根據什麼標準去選擇？我們的答案是默字測驗裏所用的字須是常用的字，常用字的選擇須以應用式發現次數的多寡為標準。

英文中的常用字彙已經有許多人研究過。愛里斯曾根據書信報章文學小說，調查了 383,000 個字，從中得到最常用字 1,000 個作為英文基本字彙。瓊斯 (W. F. Jones) 搜集了一千零五十位小學學生的作文 75,000 篇，得單字 4,532 個。安特生 (W. N. Anderson) 分析了 3,723 封在三十五種不同的職業裏的成人所寫的信，共得字數 361,184 個，單字 9,233 個。桑戴克搜集材料四十餘種，用字彙統計法從 4,565,000 字中，分析得 10,000 個單字，並統計各字發現的次數。此外如蓋茲 (A. I. Gates)，霍翁 (Horn) 亦有常用字彙的研究。

中文常用字彙亦有好幾個人做過研究。這些字彙對於編製教材，教科書，測驗都有極大的功用，茲一一簡述於后：

1. 陳鶴琴氏的研究 我國人研究通用字彙最早者要推陳鶴琴先生。他調查各種語體文書中的字彙，從 354,478 字中分析得單字 4,261 個，又統計各單字在 554,478 字中發現的次數。他的字彙所根據的材料都是語

體文，共分六大類：(1)兒童用書，(2)報紙，(3)雜誌，(4)小學生課外作品，(5)古今小說，(6)雜類。各類材料的字數與百分比見表四。

表四 陳氏語體文應用字彙材料表

材料類別	字 數	百 分 比
兒童用書類	127,203	22.98
報刊類	153,344	27.65
雜誌類	90,142	16.25
小學生課外著作類	51,807	9.34
古今小說類	71,287	12.85
雜類	60,625	10.93
總計	554,478	100.00

2. 敖引德氏的統計 敖引德氏繼續陳鶴琴氏的研究，多能單字78個，合陳氏所得4,261字共4,339個字。
3. 王文新氏的小學字彙研究 王文新氏就好幾種小學國語教科書及各省小學兒童的作文卷子，加以分

析和統計，結果作文方面，從 307, 248 字的材料中，得單字 2, 955 個，教科書方面，從 303, 941 字的材料中，得單字 4, 579 個。合計陳氏語體文應用字彙的 4233 字，共得不同的單字 5, 304 個。王氏又從這 5, 304 字中將「重字」及「在需要上不關重要的字」淘汰，共淘汰 1, 565 字，留 3, 739 字，定為六年小學應識的字。然後再根據六項標準，規定各級識字限度。六項標準如下：

(1) 字的等第 王氏曾將各字依其發現次數的多寡分為十五等。等第愈高的字，其需要愈大，故列在較低年級。等第低的字，列入較高年級。

(2) 在作文上最初發現的年級 某字在某級最先發現的，即能排在某級學習。

(3) 在作文上次數最多的年級 某字在某年級發現次數最多，則其在某級中所占需要比較迫切，分級時即可分在某級。

(4) 在教科書上發現的年級 某字在三部或兩部教科書上均在某級發現，即能列在某級學習。

(5) 字形的難易 筆畫較多或大部分由斜綫曲綫組成的為難，反之為易。難者稍後，易者稍前。

(6) 字義的深淺 對相當於某級年齡的兒童，不待解釋或解釋而容易明瞭者，即列入某年級學習。

根據以上標準，斟酌應用，結果分出一年級識字限度 541 字，二年級 658 字，三年級 654 字，四年級 793 字，五年級 923 字，六年級 931 字。

4. 莊澤宣氏的基本字彙 莊氏就六種字彙之中，求出常用字 2,327 個，備用字 1,241 個，罕用字 1,194 個，共計 5,262 個單字。他所根據的六種字彙如下：

(1) 陳啟字彙計四千三百三十九字，(2) 華文打字機所用字計二千三百九十九字，(3) 商務出版的平民字典所採字計四百五十五餘字，(4) 德人 P. Kraus 所編的常用四千字錄，(5) 平民袖珍字典稿本所採的計三千八百八十二字，(6) 王文瀾氏的小學各級字彙。

5. 杜蔭氏的兒童及成人常用字彙研究 最近杜佐周氏將成堃氏又有兒童及成人常用字彙的研究。他們所根據的材料有下列各項：

- (1) 成人的各種普通讀物：如日報，雜誌，小說，傳單，廣告，及布告等。
 - (2) 兒童的各種普通讀物：如課本，報紙，小說，寓言，童話，及參攷書等。
 - (3) 成人的各種文字作品：如信札，賬目，契約，報告，公文，及紀錄等。
 - (4) 兒童的各種文字作品：如作文，信札，日記，短評，筆記及報告等。
- 結果被調查之字的總數，共為 221,354 個，而不同的單字，共得 4,117 個。

(二) 測驗方式 默字測驗的方式可有單字式與句子式兩種。句子式的默字測驗是將測驗字編在句子裏邊，把整個句子讀給學生聽，叫學生在規定時間內一句一句默寫出來。單字式的默字測驗是由主試者將測驗

字讀出來，解釋一遍，由學生一個一個寫下來就行了。

這兩種方式各有利弊。單字式比較簡單省時，句子式則因與日常生活中書寫的情境較為相似，故結果或較為正確。普通人在日常生活中寫字總是句子式的，如寫信作文，很少機會是單字式的。因此我們要考查被試者寫字能否正確無誤，在自然情境之下，去舉行測量似乎較妥當些。

美國的試驗表示單字式的默字測驗比較句子式的容易些。這一點很容易解釋的。在舉行單字式的默字測驗時，被試者的注意完全集中在幾個單字上，所以寫錯的機會少些；在舉行句子式的默字測驗時，被試者須顧到整個句子，須注意到句子的意義，不能將注意完全集中在字的寫法上，因此寫錯的機會多些。還有，舉行句子式的默字測驗時，每句句字都有時間的限制，而這些時間的限制是參照普通人平時寫字的速率而規定的，所以被試者須照平時速率默寫，對於句子內個別的字，的寫法就沒有充裕時間去考慮了。

(三) 陳氏小學默字測驗 陳鶴琴氏所編小學默字測驗是採用單字式的。這個測驗共分兩類，每類有五十個字，適用於小學三、四、五、六年級。

這兩類測驗的一百個單字是根據陳氏語體文應用字彙用隨機選擇的方法得來的，就是把字彙裏次數最多的（即常用的）2,000字，每隔6個字取兩個字，兩個字中第一個字編入第一類測驗，第二個字編入第二類測驗，每類五十個字，共一百個字。這一百個字是能代表這2,000個常用字的。

據編者自己說，這個測驗的特點有三：

1. 根據字彙的次數，又用隨機選擇的方法選出，是很能代表常用的字。
2. 解釋的方法劃一，皆是每字後用兩句簡單清楚的解釋，不致紊亂兒童的思想。
3. 試驗的手續，極其簡單，只要主試的有一份說明即得。

三 識字測驗

(一) 識字的意義 一個字有形，聲，義三方面，所以必須見形而知其聲（如何讀法）義（如何解釋）聞聲而知其形（如何寫法）義，才算真正完全認識了那個字。在默讀時，必須能見形而知其義，才能了解所讀之內容。在朗讀時，除見形而知其義外，還須能見其形而知其聲。在聽講或談話時，必須能聞音而知義，才能了解對方的講演或談話。聽講時，如須做筆記，則除聞音而知其義外，還須知其形。

(二) 識字測驗的重要 見形知義為閱讀中一個主要的原素。閱讀能力低落當然有許多可能的原因，而了解或誤解字的意義乃其中主要因子之一。假使兒童對於常用字的意義不能了解，則其閱讀能力當然不能提高。

(三) 張氏識字測驗 張耀翔氏曾編有識字測驗，但尚未求得標準。本測驗的功用，在考查被試者讀音與

釋義的能力。測驗中所用的字係從商務印書館出版之增訂放大實用學生字典中本機會之法則，選擇而來，共一百字。

本測驗用爲個人測驗或團體測驗都可以。用爲個人測驗時，其施行之手續爲：主試者將此一百字一一指向被試者問道：「這是一個什麼字？」或「這字怎樣讀法？」然後再問：「這字怎樣講法？」主試者將答案一一記下。用爲團體測驗時，其施行之手續爲：先將試卷分給被試者，令其將試卷上一百字之音與義均注解出來。注音用反切，同音之字，注音符號，或英文拼字均可。解義用訓話或一辭句而含有該字者均可。編者備有答案標準，使校閱時有劃一標準可循。

四 作文測驗

(一) 測量作文能力之困難 作文能力的測量較諸測量默讀能力、算術能力等等爲困難，同時主觀的成分亦較大。一篇作文的品質係多種因素綜合的結果，至少可以分爲內容與形式兩方面。內容一方面又可以分析爲思想、組織、用字等因素。形式一方面又可以分析爲標點、錯字、造句、文法等因素。因素既多而各因素在總成績中應占之分量又無明確的規定，則評判者雖有標準量表可循，仍不得不依主觀的判斷以定各因素的比較的輕重。結果幾位評判員對於同一篇作文所定的分數常不能一致，且往往相殊甚遠。如何減少作文測量的主觀之成分

而增加其精確度可靠度乃作文能力測量中重要之問題。

(二)作文測驗術 周學章氏曾依照桑戴克的方法，編成一種作文測驗術。最近作者復將此術之 FBO 量表造出，以便與其他各測驗相比較。今將該術上之各篇標準文與其價值列後。

周學章氏作文測驗術

燕子 32 0.0

我自己所愛之功課

我很喜歡國文，也喜歡，我也喜歡算術，好買起筆來很用以，所以算術要用心，我也喜歡寫字，好寫起字來子好的，綴法好贖讀中好，我也喜歡鄉土，我也喜子術，我也喜欣。

燕子 27 1.1

雨具

下雨地上必甚濕，姊與我皆張傘穿皮鞋，入學校去，水大地上如何，見同學一位謂我曰，汝勿去，今日明日後日放學三天，我聽其曰，日畢即歸。

燕子 93 1.9

述衛生之道

天下之人衛生的少不衛生者多何也潔靜之人所住之房屋潔所服之衣又潔若人不潔靜皆生疾病學生皆當潔淨可也。

燕子 21 3.2

與方君之談話

余在家與方君閒談，方君曰，汝國民科畢業後至何校求學，余曰，至中法學校求學，余父母之意，欲仍在本校商業科，畢業以後，出外習商，方君曰，余欲至西區學校求學，余父欲出外習商，余曰西區學校不如本校好，方君曰，何不佳乎，余曰嘗聞同學曰，余之同學在西區求學嘗曰，西區學校不及汝學校，余又曰，汝仍本校商業科求學。

燕子 21 4.2

勸用國貨啓

人生無不愛美觀之貨者，中外貨相較，外貨形式，比中國貨美觀，故中人願用之，願美觀之貨，不能經久，是以中國人之錢，皆令外人取去，若國人能用國貨，則國貨自能發達，國家亦能富庶，所以我勸國人，皆買國貨，即我最大之希望，此啓。

燕子 35 5.0

說窶狀

第二編 各科測驗與新式考試

余校每月必發一褒狀，褒狀分五種，凡操行學業出席作業運動，若有最優者，必給褒狀，褒狀以紙爲之，有褒狀二字，印以鉛記，寫明每種爲最優，并寫校長姓名及主任姓名，下印圖章，是爲褒狀，余級現得兩張，一爲出席褒狀，一爲學業褒狀，吾級甚有光榮也。

發子 51 6.0

說同學當如何親愛

同學者，同學校，同師長，而同一讀書之人也，校中百餘人，同居以處，互相切磋，以求學問之進步，智識之漸高，而日進於高明者也，嗚呼，同學與我之關係如此，可不親愛乎，親愛之道如何，少於我者，則愛之如胞弟，衣食讀書，則輔之教之，而導之以前進，長於我者，則親之如師長，尊之敬之，仿之效之，而乞其能教於我，益於我，如此卽爲之親愛，願我同學輩，共勉之可焉。

發子 94 7.0

暑假中見聞一則

暑假既放，予負笈返家，長日如年，悶悶無聊，輒與大兄濟炎作楸枰戲，一日，楚漢相爭，方運箸畫籌間，突聞隔鄰某姓爭聲紛拏，勢若牛吼，使僕探之，蓋兄弟二人爲金錢軀迫所使然者，噫，兄弟手足也，骨肉相關，世間最難得者，固猶枝葉之相扶然，庸何致爭，更何致爲金錢而爭，何人獨不一思，夫所爭者特金錢耳，非有生命相切，若兄弟之親者，

苟兄弟而可爭，則世間可欣可悅，足以奪人者，莫不可爭，安有人倫，奚論手足耶，語未畢，余兄笑曰，爲金錢而骨肉爭，子知之矣，彼爲勢力機會潮流作無形鬼域之爭者，固不謬也，豈止兄弟鬥於牆行見有父子之爭矣，若夫意氣猜度，含沙射影，卽近觀余二人之小遊戲可矣，遂相與大笑，因記之。

雜下 10 8 1

牛女說

世謂牽牛織女，七夕相逢，自古相傳，莫究明其真僞，亦異矣。吾聞極天下之妙者，莫若鬼神，往來風雲，行動如雷，無高能阻，無險能隔，翩然而遇，颯然而合，固非山河之所得阻也，而必待七夕相逢者何哉，噫，我知之矣，蓋上天垂象，以示天下，聖人立教，以化萬民，無非欲致天下於昇平，羣民各得其所而已矣。故設爲牛女之說，以示男子當以耕稼爲事，女子當以紡織爲務，會合自有其時，不可苟且從事，如此，然後男耕女織，各安其業，而天下平萬民安矣。噫，先聖之用心亦盡矣，惜乎後世無有能申而明之者。

雜下 15 9.0

記女子復仇事

女子陳氏，錦縣人，襁褓失怙，母以戚族凋零，來日大難，因懷女而投於江，母死，女以體輕浮沈數十里，卒不死，爲近郭清水菴老尼悟果所得，因以證果名之，女故陶姓，云陳氏者，從尼姓也，稍長，穎悟過人，喜佛學，寡言笑，佛家諸典，

過目了然，尼鍾愛之，撫育顧復，一如生母焉。菴故依山臨水，風景絕佳，顯官魏某，亦錦縣人，愛菴之風光明麗，欲購以爲別墅，尼固執不應，雖多金勿肯也，宦怒置之獄，匝月庚死，菴遂爲魏所有，鳩工擇材，方擬大興土木，突於竣工之前數日，爲一無名女子所刺死，女子亦死，有識之者曰：是卽清水菴證果也，蓋魏姬某嘗遊是菴，愛女慧，極優容之，女卽於尼入獄之日，往依姬爲婢曲意奉侍，益得姬歡，乘某醉而殺之者也。

下： 偷子夷氏亦編有小學綴法量表，可以從小學二年級用起，到初中三年級爲止；平常從小學三年級到初中二年級應用最適宜。偷氏量表有作文樣本 〇 篇，其 〇 價值從 〇 分至 〇 〇 分不等。茲錄偷氏量表中作文樣本數篇於下：

20

王素貞好朋友頂好有一天我到朋友公園遊戲見花開好看我最好看

21

我有一個好朋友他 我是很好的我也 他好的我的朋友的名字叫阿南有一天他同我到外面去 看見一隻猴子在上面做喜看了一 還到蟲完字裏去完還看出喜法看了一看我們還走 走到公花園去完還去坐 坐 坐了一 就回來回到家裏我就到家他就到家裏去了。

20

我的好朋友是曹秉琛哥哥。他天天同我一同上學，在學校裏同我遊戲散學的時候同我一同回家。他對於人很愛護好。象家庭之兄弟，沒有二樣。我有時有過他勸我有危險他教我不要去。做這是我的好朋友。

325

我的最好的朋友，是姓張，叫國英。什麼是我最好的朋友呢？他有一天，我同他到外邊去遊玩，走到一個河邊，看見有許多小兒釣魚，很快活。國英同我遂去看他們。走他這裏，看了一會，忽然釣了一條大魚，一輩小兒們都拍手大笑起來。我亦在旁大笑，不了一失足，跌到河裏去了。羣兒大驚，皆大奔，但是國英不逃，跳下來救我。舉然救了我的命。我非常感激他。不但是我感激，並且鄉村上的人都稱他勇兒。否則那兒都逃，我的生命就不保了！故吾同他是最親愛，和我的兄弟一輩。

30

我是最快樂的人，因為我有很多的好朋友，同我做伴侶。同我講話，同我遊戲，而且非常親暱和善，真所謂一刻不能分離的。倘使你們要問我那許多的小朋友是什麼，我不妨現在把每天的經過情形略略地訴說一下，就可以知道了。

當天初曉的時候，淡淡的曉光，啞啞的雞聲，就把我從夢中吹醒過來，隨即爬起牀來，洗面、整被等完畢，後立刻去尋我的書房裏底課桌上的書本。他是我的好朋友，他就同我講笑話，講有趣的故事，以及種種的事情，總使得我

愉快還有筆墨嘔紙嘔都陪伴就做我的功課，所例做的事情這樣地過了許多時還要到操場上去尋那軒轅浪橋等等。那許多也是我底好朋友他們也。照常的要遊戲尋快樂使我身體康強而不生委倦。還要走到校裏去會面那些美麗的花木和啾啾的小鳥他們也是我底好朋友。因為可以和我接吻和我唱歌使得我樂而忘返總之我自至今一直過這學校生活的時候這些書本呀筆墨紙觀呀浪橋和軒轅嘔都成為我的好朋友他們是能夠每天簡直是天天的陪伴我歡喜我給好處我我於是便做他們唯一底好朋友！

90

我的朋友很多，但是好的却很少，這是什麼緣故？連我自己也不知道，我只覺得我的所謂好朋友，就是一個天天和我遊玩的大自然，牠的面貌服裝，每季必換上一身。牠是沉默的，偉大的，而我因牠沉默，方能得到一種寂寞的安慰，因牠偉大，方能感到人生的價值，——我在課餘的富兒，常到森林下，聽牠幽婉的歌聲，夜深的時候，萬籟俱寂，她能把一個明珠來戲引我，這好像牠是不能不給我的，而於我卻好像不能不接受他的，牠又好像不要我報答的，而我卻是一飯不忘的必要把未來之花果去報答他呵！這真是我的好朋友！

(△是錯字，錯的句讀，脫落等，△是意思不通等。)

(三) 如何增加作文測量之可靠度 作文能力測量之可靠基於兩種因素。第一種因素即評判時主觀的成分太多。表五、為嚴菊生女士所得之結果，表中二十篇作文經一百位評判員依周氏作文測驗衡評判，此一百

位評判員對各篇作文所定分數的差異度見表中第三行。評判之不準確不可靠已極明顯。

表五 表示作文測量的主觀性

作文號數	評 判 人 數	分 數 全 距
1030	100	1.5—9.5
1031	100	1.0—9.5
1032	100	1.0—9.5
1033	100	1.0—9.5
1034	100	0.0—8.0
1035	100	0.0—7.5
1036	100	0.0—8.0
1037	100	0.0—9.4
1038	100	0.0—7.0
1039	100	0.0—8.5
1040	100	1.0—9.5
1041	100	1.5—9.5
1042	100	0.0—9.0
1043	100	0.0—7.5

第二編 各科測驗與新式考試

1044	100	1.5—9.0
1045	100	2.0—9.5
1046	100	0.5—9.0
1047	100	0.0—9.0
1048	100	0.0—7.0
1049	100	0.0—8.0

欲增加評判之準確度，則評判員之審慎選擇與嚴格訓練似為唯一之方法。

第二種因素為學生作文成績自身之變化，換言之，一次的作文不能正確地代表真正的作文能力。學生之作文成績既有變異，縱有訓練完善之評判員，所得結果仍難以正確。

1. 評判員之選擇與訓練 訓練之能否收良好效果與受訓練者之態度與能力均有重要之關係。周學章氏本其豐富的經驗定有四條標準。

- (1) 評判員以對於測驗信任且發生興趣者為合格。
- (2) 評判員以對於評判學生作文有一二年經驗者為最善。
- (3) 評判作文時，評判者必時時刻刻用其堅強之注意力而後可，否則易致錯誤。故選擇評判員時，應注重其忍耐力，凡心思細膩而能持久者為當選之重要條件。

(4) 據試驗結果，作文能力與判文能力為相關的。換言之，作文能力高者，其判文的準確度亦較高。準此，凡作文能力薄弱者，不可請其為評判員。

訓練評判員時，須先備有價值已知之作文多篇，然後令評判員依測驗銜定其價值，再與標準價值比較，以發現錯誤之數量與方向。如是繼續練習，以至各人評判之平均差減至最低限度(0.4)且有減無增時為止。周學章氏定有作文二百六十篇，各文之價值，均依周氏測驗銜經多人評判結果而定者，可用以為練習之材料。(註一)

表六 表示計算評判錯誤之數量與方向之方法

作文號數	標準分數	評判員所定分數	差 (不計符號)	正差數	負差數
1	6.5	6.5	.0		
2	2.6	3.0	.4	+ .4	
3	7.9	7.0	.9		- .9
4	3.2	4.5	1.3	+ 1.3	
5	0.8	2.0	1.2	+ 1.2	
6	4.5	4.0	.5		- .5
7	1.25	1.7	.45	+ .45	

表七 表示作文成績之變異（從周章氏）

作 文 題 目	r	P _α
「 <u>遊蟻</u> 」與「 <u>國難中人民應持之態度</u> 」	.51	.05
「 <u>記災民慘狀</u> 」與「 <u>勸用國貨啓</u> 」	.41	.05
「 <u>振興工業爲強國之本</u> 」與「 <u>青年的出路</u> 」	.38	.03
「 <u>交友宜信說</u> 」與「 <u>國難中人民應持之態度</u> 」	.09	.07
「 <u>國難中人民應持之態度</u> 」與「 <u>寒假見聞一則</u> 」	.78	.03
平 均	.43	

表中 r 乃同一班兒童兩篇作文成績之相關度，r 愈小表示作文成績之變異度愈大，即一次作文成績之愈不可靠。各相關係數均由測驗初級中學一年級學生一百四十二人之結果而得者。五個係數之平均爲 .43，可見一次作文成績之可靠度極低而差異度極大。表中各文均爲曾受精密訓練之一位評判員所評判，其評文之可靠係數爲 .90，P_α 爲 .007，所以表七中兩篇作文成績相關度之低落乃作文成績自身變化所致，不能歸咎於評判員評判之不可靠。

欲減低作文成績之變化度而增加評判之可靠度，最直接之方法即增加每人作文次數，以數篇作文之平均

成績代表其能力。兩次作文之平均成績自較一次為可靠，三次較兩次為更可靠，次數愈多則評判愈可靠。周氏根據表七中之平均可靠係數 $\cdot 65$ ，假設可靠係數 $\cdot 80$ 即可認為滿意，用白朗氏預斷公式計算之，作文五次即可滿足。為得較 $\cdot 80$ 更高的可靠度，作文次數當然須相當增加。

五 文法測驗

(一) 文法測驗的功用 我國國文的教學向來不講求文法，文法上之錯誤概以「不通」二字概括之，僅指出其錯誤而不說明其錯誤之性質，所以學者往往對於同樣的錯誤重複多次而仍不知改正。這是何等地不經濟。有效的教學的第一任務在發現學生缺點之所在，錯誤之所在，然後對症下藥予以糾正，使學生不但曉得自己有無錯誤，並知其錯誤之所在，錯誤之性質以及改正之方法。必有這樣的診斷的教學總能收效。文法測驗之功用即在發現學生在文字上或言語上所犯之錯誤，以為施行補救教學之根據。

(二) 陳氏文法測驗 我國文法測驗已編就者共有兩種。一為陳氏文法測驗，適用於後期小學。一為廖氏文法測驗，適用於初中及高中。陳氏文法測驗共有五十句子，每句裏面有一錯字，這個錯誤是屬於文法方面的，叫被試者改正。茲舉例如下：

1. 皮鞋是牛皮做得。

2. 那個地方我從外沒有走過。
3. 這件事我覺得可非常奇怪。
4. 先生的話我沒好一句不明白。
23. 他雖然是一個平常的人，你又沒有什麼了不得。
24. 這個東西好又好，但是沒有十分用處。
25. 今天下雨不於在家玩罷！
26. 他特可看重你。
47. 昨天受了寒，他的病因諸加重了。
48. 這個婦人是靠得針線生活度日的。
49. 並非明天不下雨我來看你。
50. 這個壞東西又不如不要罷。

每句句子裏面只有一個錯字而校正字也只能用一個唯一的規定字，但這一層不易做到，所以有時候一句句子的適當字可以有兩個或兩個以上。

上面所舉的例，第一句的「得」字應改為「的」字，第二句的「外」字應改為「來」字。第二十六句的「

可」字應改為「別」字，但改為「地」字亦可，所以這句的適當字有「別」、「地」兩個字。

(三) 廖氏文法測驗 這個測驗有兩類，每類有三十個句子。每句子裏邊有二三段虛綫，每段虛綫上應填一個適當的文法字，使那句子成爲一句有意義的句子。句子的排列是由易而難。時限爲十分鐘。茲舉數列如下：

1. ...上海到南京有...遠？
2. 他每天起來練習早操，...他的身體很強健。
3. 我的計劃決定了，前途無論怎樣危險，我...去走一遭。
12. 倘使他們有錢可以養家，也...結婚了。
13. 雨聲...地住了。
14. 你的身體幼弱，...噢這樣的辛苦。
28. 戰爭同老虎一樣，...要吃人的。
29. 小園的花有意...開了。
30. 你...曉得他是無心之失，...還去這樣的責備他！

六 書法測驗

(一) 書法教學的目標 書法是一種工具的學科而同時又為一種藝術。在學校裏書法的教學對於藝術的價值固然不應抹殺，而工具的或實用的價值則尤為重要。書法教學的目標似應以工具的價值為主，以藝術的價值為副。

從形式方面講，書法有正書，行書，草書及大字，中字，小字之別。小字的用處多於大字；行書的用處多於正書草書。故小字及行書的訓練，應認為書法教學的主要的對象。

書法的測量有兩方面，一為速率，一為品質。學校裏書法的教學向來祇注重品質而忽略速率，但從書法的實際的功用看來，速率亦極為重要。書法速率的測量較為簡捷，可於規定時間內，令被試者寫一定的材料，視其所寫字數之多寡以斷定其速率。品質的測量則有賴於書法量表，以為評判之標準。

(二) 測量書法能力的困難 測量書法能力與測量作文能力有同樣的困難，就是主觀成分太多。威爾遜曾報告 N.C.T. 位評判員用愛里斯書法量表評判同一篇字，所得分數自 30 至 100 不等。全部分配如下：

表八 N.C.T. 位評判員對於同一篇字所定的分數

分數	數	次	數
30	40	1	28
40	50	56	50
50	60	71	37
60	70	80	2
70	80		
80	90		
90			
總平均數	245	61.47	12.96
標準差			

欲減少書法測量的主觀的成分而增加其可靠度，評判員之訓練似為最有效的方法。

(三)書法測驗 俞子夷氏於民國七年已編成了四種書法量表，一種正書中字，一種正書小字，一種行書中字，一種行書小字。這幾種量表係做桑戴克的書法量表編造的，未採用工的單位。俞氏求有年級標準如下：

表九 俞氏書法測驗的年級標準

年級	書法			
	正書中學	正書小學	行書中學	行書小學
一年末	4.0	7.5		
二年末	5.5	8.2		
三年末	6.5	8.8	7.0	
四年末	7.5	9.5	8.5	9.0
五年末	9.0	10.0	10.5	10.0
六年末	10.0	10.5	11.5	11.0
七年末	11.0			

最近俞氏又編了一種小學書法測驗，有正書小字和行書小字兩類，速率與品質都可測量。測驗裏所用的字爲「四隻小鳥他們在園中飛好像一個人字」十六個字。這十六個字大概把中國各種的筆劃及常用的間架結構都包括在內，且成爲一有意義的句子，使被試者易於記憶。這種測驗從小學二年級起至初中二年級止均可適用，平常從三年級到六七年級最爲適宜。行書平常教得較遲，可從小學四五年級用起。測驗時間正書限十分鐘，行書限2.5分鐘。

成績分速率和品質兩種。計算速率時祇要數被試者寫的字數，然後從說明書對照表中查出日分數即得。計算品質時須用書法量表（評判正書小字的，用正書小字量表；評判行書小字的，用行書小字量表），其手續如下：

把量表掛在壁上或平放在大的桌子上，把各學生成績逐一沿了量表移動，決定成績和量表那一段最像。決定後再向左右各移過一二段，如果左右各段都不是最像，那麼剛才看最像的那一段上的數目字就代表某學生成績的品質。量表上的數目就是日分數，不必再另查表了。

速率和品質的分數分列比較有意義些。如欲合併計算，可將速率和品質兩個分數相乘，開方就得，或者兩數相加折半也可以。

問題

1. 比較默讀和朗讀的重要。
2. 陳氏中小學默讀測驗有什麼特點？
3. 陳鶴琴氏語體文應用字彙所根據的材料，有什麼可以批評之處？
4. 莊澤宣氏漢本字彙中的常用字，備用字，罕用字，用什麼標準決定？
5. 比較句子式與單字式默字測驗的利弊。
6. 如何才算完全認識了一個字？
7. 評列張氏識字測驗。

8. 作文能力測量之不可靠基於那幾種因素？
9. 要增加作文能力測量之準確度有什麼方法？
10. 書法教學應注意那幾方面？

參考書報

1. Monroe, W. S. *Measuring Results in Education* Ch. 2, 3, 7, 8, 9. Houghton Mifflin.
 2. Monroe, W. S., De Vos, J. C., Kelly, F. J. *Educational Tests and Measurements*, Ch. 3, 4, 5, 6. Houghton Mifflin.
 3. Wilson, G. M. and Hoke, K. J. *How to Measure*, Ch. 2, 3, 5, 6, 7. Macmillan.
 4. 陳勉琴：語體文應用字彙（商務）。
 5. 王文新：小學分級字彙的研究（民智）。
 6. 王文新：小學分級詞彙的研究，*教育研究*二十八、三十期。
 7. 在譯言基本字彙（民智）。
 8. 杜佐周：繙成語；兒童及成人常用字彙研究，*教育與民衆*第四卷第八期。
- 第二編 各科測驗與新式考試

9. 包種顯：小學作文錯字統計與分析，教育研究廿一廿二兩期。
10. 戴新三：小學兒童別字的研究，兒童教育五卷二期。
11. 張維穎：識字測驗，心理雜誌還存（中華）。
12. 傅傑霖：漢字基本字研究的初步教育與民衆二卷二期。
13. 教育研究廿五期，讀法專號。
14. 陳鶴琴：初小默讀測驗編造程序，教育雜誌八卷三期。
15. 陳鶴琴：中小學默讀測驗編造程序，教育雜誌十六卷五號。
16. 周學章：測驗作文能力應知之事項，教育研究四十四期。
17. 俞子夷：編造小學書法測驗方法的概要，新教育六卷四期。
18. 俞子夷：小學校毛筆書法三成績的算法，心理二卷二號。
19. 俞子夷：小學國文毛筆書法，初等教育研究雜誌第十號。
20. 陳鶴琴：初小默讀測驗兩類及說明書（商務）。
21. 陳鶴琴：小學默讀測驗及說明書（商務）。
22. 陳鶴琴：中學默讀測驗及說明書（商務）。

- 23. 陳鶴琴：小學識字測驗說明書（商務）。
- 24. 俞子夷：小學綴法測驗量表及說明書（商務）。
- 25. 陳鶴琴：小學文法測驗及說明書（商務）。
- 26. 廖世承：中學文法測驗及說明書（商務）。
- 27. 俞子夷：小學書法測驗量表及說明書（商務）。

第九章 算術測驗

一 算術測驗的新趨向

算術能力的測量有兩種新趨向。這兩種新趨向都是算術的心理研究的結果。第一種新趨向就是診斷測驗的注重。原來測驗在教學上最重要的應用就是診斷困難，以爲施行補救教學的根據。譬如醫生看病必須將病源之所在診斷清楚然後對症下藥，才可以收成效。要是未把病源捉住，胡亂下些藥，則不但不能把病治好而且有殺人之虞。教學亦然，必先將兒童困難之點診斷清楚，才能規定補救教學的方案。盲目的教學與盲目的下藥一樣地無效，一樣地危險。普通測驗僅能表示程度之高低而不能指出其特殊困難之所在，其功效不過等於一個熱度表。醫生用熱度表測量一個病人的熱度，發現熱度反常，醫生並不能就據此而斷定其病因。一個兒童在一個普通測驗上成績低落，教員亦不能就據此而斷定其困難之所在，由此看來診斷測驗的效用較普通測驗大多了。

診斷測驗在算術一學科裏發達最完備，一則因爲算術科教學的目標已十分確定，二則因爲心理學者對於算術中基本的步驟已有詳盡的分析。一學科的目標不確定，內容的分析不詳盡，則診斷測驗之編造便無所根據。茲將克底斯對於整數四則的分析錄下，以示一斑。

(一) 加法裏的階級:

1. 各種單位數相加: 例如, $1+2, 1+3, 2+3, 6+7, 6+6, 7+8$ 等等。
2. 單位三數相加: 例如, $5+7+6$ 〓
3. 帶上「十」數的: 例如, $38+7$ 〓
4. 七個數目的直行相加: 例如, $79 \quad 11 \quad 37 \quad 84 \quad 75 \quad 42 \quad 98$
5. 進位相加: 例如, $\begin{array}{r} 478 \\ +965 \\ \hline \end{array}$
6. 十三個數目的直行相加: 例如, $339 \quad 128 \quad 657 \quad 812 \quad 179 \quad 892 \quad 973 \quad 451 \quad 734 \quad 971 \quad 284 \quad 817 \quad 252$
7. 長短不等的數目相加。

(二) 減法裏的階級:

1. 各種單位數相減: 例如, $7-5, 2-1, 8-6, 7-4, 9-7$ 等等。
2. 從兩位數減去9或比9小的數目, 無須乎借數: 例如, $19-9, 39-7, 27-5$
3. 有「〇」的直行相減: 例如, $\begin{array}{r} 30 \\ \underline{5} \\ 405 \\ \underline{14} \end{array}$
4. 借上行一個數目的減法: 例如, $\begin{array}{r} 275 \\ \underline{148} \end{array}$

5. 借上行二三個數目的減法：例如，

$$\begin{array}{r} 3463 \\ 1785 \\ \hline 703 \\ 427 \end{array}$$

(三) 乘法裏的階級：

1. 各種單位數相乘：例如， $4 \times 5, 9 \times 8$ 等等。

2. 單位乘數與雙位被乘數相乘而不須進位的：例如，

$$\begin{array}{r} 498 \\ \quad 8 \\ \hline 576 \\ \quad 827 \\ \hline 232 \end{array}$$

3. 與上同，不過要進十的：例如，

$$\begin{array}{r} 31232 \\ \quad 132 \\ \hline 875 \\ \quad 97 \\ \hline 560 \\ \quad 47 \\ \hline 807 \\ \quad 59 \\ \hline 617 \\ \quad 508 \\ \hline 753 \\ \quad 60 \\ \hline \end{array}$$

4. 複位數相乘而不須進十的：例如，

$$\begin{array}{r} 29704 \\ \quad 675 \\ \hline 875 \\ \quad 97 \\ \hline \end{array}$$

(四) 除法裏的階級：

1. 各種單位數相除：例如， $4 \div 2, 9 \div 3, 8 \div 2$

2. 簡單除法而無須帶下的：例如， $48 \div 2, 903 \div 3$

3. 同上，但須帶下的：例如， $2 \overline{) 962} \quad 3 \overline{) 813}$

4. 複位數相除而無須帶下的：例如， $183 \div 61, 148 \div 72$

5. | 6. 有一〇」數的困難的，有兩種方式：

$$\begin{array}{r} 690 \\ 71 \overline{)48990} \end{array}$$

$$\begin{array}{r} 302 \\ 31 \overline{)9362} \end{array}$$

7. | 10. 各種複位數相除而有帶下的，共有四種方式：

$$\begin{array}{r} 72 \\ 83 \overline{)4536} \end{array}$$

$$\begin{array}{r} 68 \\ 49 \overline{)3087} \end{array}$$

$$\begin{array}{r} 89 \\ 63 \overline{)5607} \end{array}$$

$$\begin{array}{r} 79 \\ 36 \overline{)2844} \end{array}$$

有了這樣的詳盡的分析，編造診斷測驗便易於着手了。

算術測驗的第二個新趨勢就是測驗中問題的選擇應以生活的需求或社會的應用為標準。桑戴克威爾遜諸人的研究很明顯地暴露傳統的算術課程的不切於實際，兒童在學校裏所學習的算術有一部分為日常生活中所極不需要的。桑戴克分析了好幾種算術教材發現其中有許多在實際生活中無用的方法和事實，同時許多極有價值的方法和事實反被忽略。這是教學上極大的浪費。這些研究的結果促成了近年來算術課程的改訂。測驗既然是改進教學的重要的工具，測驗的內容自須與教學目標，教材內容相符合。社會的應用為編訂算術教材，同時亦為編製算術測驗的重要的標準。

算術測驗依其性質可分為四則測驗，應用題測驗，練習測驗三類，茲分述如下：

二 四則測驗

四則測驗就是測量加減乘除四種基本算法的。四則測驗又可別為普通調查測驗與診斷測驗兩類。普通測驗的材料的選擇採用取樣的方法；診斷測驗的材料的選擇則採取包攬無遺的原則，將四則中各基本步驟或階級，一一編入測驗之內。四則能力的測量又可分為速率與正確兩方面。

我國四則測驗有下列數種：

(一) 德氏算學四則測驗 本測驗本係凌柯爾所編，由德爾滿求得標準。這個測驗包含速率、正確、診斷三種性質，在我國四則測驗中要算首屈一指了。本測驗共有兩類，難易相仿，可以替換應用。初級小學高級小學均適用。

1. 測驗的內容 本測驗共有八十個題目，加減乘除各二十題，除整數四則外，兼有小數四則。題目的選擇與排列，均以難易階級為根據，每兩個題目代表一種算術階級或算術上一種原則。茲將德氏測驗的前面一部分列下：

	(1)	(2)	(3)
加	$\begin{array}{r} 3 \\ +4 \\ \hline \end{array}$	$\begin{array}{r} 6 \\ +2 \\ \hline \end{array}$	$\begin{array}{r} 7 \\ +5 \\ \hline \end{array}$
減	$\begin{array}{r} (5) \\ 6 \\ -3 \\ \hline \end{array}$	$\begin{array}{r} (6) \\ 8 \\ -4 \\ \hline \end{array}$	$\begin{array}{r} (7) \\ 9 \\ -5 \\ \hline \end{array}$
			(4)
			$\begin{array}{r} 7 \\ +9 \\ \hline \end{array}$
			(8)
			$\begin{array}{r} 8 \\ -0 \\ \hline \end{array}$
			減

加	(9)	(10)	(11)	(12)	加
	3	3	24	50	
	1	0	+4	+6	
	+7	+5	+4	+6	
	-----	-----	-----	-----	

上面(1)(2)兩題代表最簡單的加法：單位數相加，不須進「十」。(3)(4)兩題代表加法另一步驟：單位數相加而須進「十」。(9)(10)兩題是三數相加，比二數相加又難一層了。(11)(12)兩題是兩位數加單位數，比三個單位數相加更難了。以下加法題目都是一步難一步。減法乘法除法的題目亦是照此原則而編擬排列的。

2. 施行的手續 這個測驗的規定時間為十二分鐘，但做到八分鐘的時候，主試者應說「停」並且叫被試者把那時候未了做的那一個題目畫一個大圈，等到大家都圈好了，再叫他們把畫圈的那個題目做完，然後一直做下去。到十二分鐘，主試者再發令停止。到了八分鐘叫停，為的是要計算被試者算法的速率。

3. 計分的方法 從這個測驗可以得到三種分數：曰算，曰速算，曰速對算。曰算就是在十二分鐘內所做對的正確分數；曰速算就是在八分鐘內所做的速率分數；曰速對算就是在八分鐘內所做對的正確分數。各種曰分數的查法，詳見測驗說明書。

此外又有三種診斷計分法：(1)將被試者的速度分數與正確度分數比較，以定被試者的作業為慢而正

確的，或慢而不正確的，或快而正確的，或快而不正確的。根據此種情形，施以適宜的教法。(2) 測驗中加減乘除各有20題，計算時將四種分數分別計算，以發現被試者特殊缺陷之所在。(3) 將被試者已做及未做的題目，列成一表，以發現被試者的困難在那種階段或那種原則上。

(二) 俞氏小學算術混合四則測驗 本測驗原為吳特麥柯爾所編，由俞子夷修改試用並求得中國學校的常模。測驗共分兩類，每類三十四題，內容包括日常需用的各種四則，有整數四則，小數四則，分數四則及諸等等。從小學二年級起至初中三年級止均可應用，平常從小學三年級起至初中一年級止，最為適用。時限為 $\frac{2}{3}$ 分鐘。

這個測驗的編製完全以統計上的難易為標準而對於算術各難易階級則未曾顧到。所以這個測驗不能作為診斷之用。

(三) 崑山算術四則測驗 本測驗亦是俞子夷所編的，因測驗標準係根據崑山各學校成績求出的，故以崑山名。本測驗適用於初級小學。

三 應用題測驗

應用題測驗的編造有兩個應注意之點：第一，測驗中題目須與實際生活中所遇到的問題相近。桑戴克分析算術教科書發現其中有不少極離奇荒謬的問題，為日常生活中所從不遇到者。茲舉數列如下：

某學校有 15 個教室，每個教室平均有 6 個兒童。假如每個兒童在他的石板的兩面各劃了 500 個 \times ，問全體一共劃了幾個 \times ？

我分給一羣兒童每人二角五分，一共分去了二元七角五分，問該羣兒童共有幾人？
編製測驗時應極力避免這些荒誕不經的問題。

第二，應用題目的文字宜力求簡易通俗，以免被試者因文字上的困難而將題目做錯或做不出，使被試者真正的算術能力不能測量出來。

我國的算術應用題測驗有下列數種：

(一) 俞氏小學算術應用題測驗 本測驗可用以考查學生解答應用問題的能力和計算日常算法的能力。測驗內容有整數四則，小數四則，分數，百分，折扣，面積，體積，容量等等，差不多把日常須用的應用題解決法完全包括在內。本測驗共有四類，四類的難易大概彷彿，可以替換應用。從小學二年級起到初中三年級止本測驗都適用，平常從三年級到七年級最適宜。時限為 30 分鐘。

計算分數以答案對不對為標準，而不問算法的合不合。這一點似乎值得討論的，理解題測驗之計分，不但要顧到答案對與不對，似乎還應顧到算法的原則合與不合。美國孟祿的算術理解題就分原則分數與答案分數兩種。

(二)崑山算術應用題測驗 本測驗的性質與上述俞氏小學算術應用題測驗相同，惟內容較為淺易，適用於小學一二年級。編者為俞子夷氏，因測驗標準係根據崑山各學校成績求出的，故以崑山名。

小學一二年級學生識字看書的能力十分薄弱，故本測驗不用印的題目而由主試者將題目口述，以免去文字上的困難。

四 練習測驗

測驗最大的功用乃在診斷學生的困難以指示教學的途徑，所以測量與教學須相輔而行才能取切實的效。練習測驗乃將測驗工作與教學工作打成一片的最有效的工具。練習測驗的編製以算術一學科為最早，到現在讀法及其他學科亦已有相當練習測驗可用了。

(一)俞子夷——麥柯爾算術練習測驗 美國克底斯的算術標準練習測驗為練習測驗中之最早者。後來施都特培克 (J. W. Studebaker) 亦編製了一種算術練習測驗。本測驗係採用克氏測驗的內容和施氏測驗的格式由俞子夷與麥柯爾彙編而成。

本測驗有兩類，兩類題目不同而難易及分量均相等。每類有練習片五十八張，每張一課共五十八課，最後四課（第35, 36, 37, 38課）係學習用的，又有四課（第13, 30, 41, 51課）係測驗用的，其餘五十課為練習的材料。

每課中的題目都是同一類的並依照難度而排列；每一課代表算術中一種階級，舉凡整數四則及小數四則各階級都包括在內。練習片上每個題目下面有一空洞，以備填寫答案之用。

施用本測驗時，每個兒童須有一份練習片，成績摺及空白紙各一份。練習時，叫兒童把空白紙放在練習片下面，把答案寫在練習片空洞下面的空白紙上。這樣，練習片可以應用多次而不致損壞。練習片後面有標準答案，可以叫兒童根據標準答案校對自己的成績或互相交換校對亦可。在成績摺上兒童可將每日的成績記載上去，以顯示進步的狀況。施用本測驗的詳細手續備載說明書，不多述了。

(二) 馬靜軒兒童算術練習測驗片 馮氏算術練習測驗片，專供小學初年級（一二三年級）練習基本四則用的。用完這套片子，可以接用偷氏算術練習測驗。

這套算術練習測驗片，共九十課，其次序依基本四則學習心理的難易而排列，內中十四課為測驗用的（第6, 12, 18, 24, 30, 37, 44, 51, 58, 65, 71, 77, 83, 90等課），其餘七十六課為練習用的。這七十六課練習可分為十四組：第1課到第5課，第7課到第11課，第13課到第17課，第19課到第23課，第25課到第29課，第31課到第35課，第38課到第42課，第45課到第50課，第52課到第57課，第59課到第64課，第66課到第70課，第72課到第76課，第78課到第82課，第84課到第89課。每一組有一測驗，教師可用以考查兒童的程度，決定從那一組練習起。凡第6課測驗及格者，1到5各課就可無須練習了，不及格者應當從第1課起練習。第10課測驗不及格的，應當從第7課起練習。第

18. 課測驗不及格的，應當從第18課起練習……照此下推，第20課測驗不及格的，應當從第20課起練習。

練習時，兒童須各備算術練習測驗片一套，練習簿一本，成績記載摺一份；教師須備練習測驗片一套，教學法一冊。練習時，將練習片印有練習題的一面向上，放在練習簿半透明紙下，使兒童可以看出題目，即在半透明紙上寫答數。練習的時間各課宜一律，編者所規定時間如下：

一下：六分鐘

二上：五分鐘

二下：四分鐘

三上：四分鐘

三下：三分半

練習好了，反轉練習片，將印有答數的一面向上，仍舊放在半透明紙下，將自己算出的答數，和片上的答數對照。每天的成績，指導兒童記入成績記載摺內。

據編者自述，本算術練習測驗片有下列數特點：1. 數目過多過大的題目沒有。2. 題目上不加「+」、「×」、「-」、「÷」符號，以免學校學的情形和社會實際情形，相距太遠。3. 題目反面有答數，兒童做好以後，可以自己檢答，很宜自習。4. 洋紙堅厚，印刷合宜，極便兒童的使用。

(三) 練習測驗的功效 練習測驗的功效約有下列數端：

1. 促進個別教學 班級教學的不經濟是很顯著的事實。班級教學以團體為對象，以適合全班學生的需要為目的，但結果往往一個學生的需要亦不能適合。練習測驗可以使個個兒童按照自己的程度能力速度而進行，不受別人的牽制，可以避免重複的教學與未成熟的教學。

2. 增進學習的動機 練習測驗可以從三方面增進學習的動機。第一，學習的目標確定。學習者對於學習的目標知道得愈清楚，則學習愈努力，進步愈快。盲目的學習是最不經濟的。第二，競爭心的利用。在競爭之下，學習特別努力，成績特別優異，這是最淺近的心理的事實。競爭又可分為與人競爭及與自己以往成績競爭。練習測驗對於這兩種競爭心都已充分利用了。第三，學習者對於自己以往的進步狀況及現在的成績曉得愈清楚，則學習的效率愈高。應用練習測驗時，學生須將自己的成績隨時登載在成績摺上；有了成績摺則對於自己進步的狀況便可一目了然，學習的效率亦因此提高了。

3. 使學生有充分的練習 桑戴克的學習律的第二條便是練習律。練習為學習的必要條件。班級制度下個別問答的方法常常使全班兒童中同時祇有一個兒童在運用他的腦筋，祇有一個兒童得着了練習。施用練習測驗時，個個兒童都得着了練習，時間的經濟，效率的提高是當然的結果了。

4. 幫助診斷工作的進行 練習測驗可以幫助教員診斷學生的困難，以為補救教學的根據。

五 混合數學測驗

此外廖世承等編有一種混合數學測驗，共兩類，每類六十題，包括四則、代數、幾何及三角四類問題。適用於初中及高中。

問題

1. 算術能力的測量有什麼新趨勢？
2. 診斷測驗何以在算術一科中發達最完備？
3. 練習測驗有什麼功效？
4. 批評俞氏小學算術混合四則測驗。
5. 批評俞氏小學算術應用題測驗。
6. 馬靜軒兒童算術練習測驗片有那些特點？

參考書報

1. 德爾滿：算術四則測驗及說明書（商務）。
2. 俞于夷：小學算術混合四則測驗及說明書（商務）。

3. 俞子夷：崑山算術四則測驗及說明書（商務）。
4. 俞子夷：小學算術應用題測驗及說明書（商務）。
5. 俞子夷：崑山算術應用題測驗及說明書（商務）。
6. 俞子夷—麥柯爾：算術練習測驗及說明書（商務）。
7. 馬靜軒：兒童算術練習測驗片（兒童）。
8. 馬靜軒：兒童算術練習測驗片教學法（兒童）。

第十章 英文測驗

英文一學科在我國中小學課程裏占相當重要的地位。考語言文字的作用不外兩端：一為接受，藉語言文字從外界環境中收受意義，一為發表，藉語言文字將意義達至外界或他人。接受意義的途徑不外眼看與耳聽，表達意義的方法不外口述與筆述。根據上面的分析，則英文教學的目標有如下列：

接受方面：

閱讀的能力——閱讀文字以收受意義

耳聽的能力——聽人語言以收受意義

發表方面：

會話的能力——用語言表達意義

作文的能力——用文字表達意義

依此目標，則英文程度的深淺亦可由上列四方面考查之。

一 美國方面的英文測驗舉例

美國已編就之英文測驗不下數十種，雖內容未必盡適合我國學校，但足資借用或參考之處甚多。茲就閱讀測驗、拼法測驗、書法測驗、作文測驗、文法測驗各舉數例以爲介紹。詳細內容請閱章末所列各參考書。

(一) 閱讀測驗 閱讀可分爲默讀測驗與朗讀測驗兩種。默讀測驗中之採用較廣者，有下列數種：

1. *Thornike-McCall Reading Scale* 本測驗共有十類，格式與我國陳氏小學默讀測驗相仿。從二年級至十二年級均適用，從四年級到八年級最爲適宜。測驗的成績以百分數計算之。

2. *Burgess Scale* 本測驗共有四類，難易相等，可以替換應用。適用於三年級到八年級。

3. *Hagerly Reading Examination* 本測驗共有三類，第一類適用於一、二、三年級，第二類適用於小學高年級，第三類適用於初中及高中。

4. *Gates Reading Tests* 本測驗共有兩份。第一份包含三種測驗，適用於小學一、二年級；第二份包含四種測驗，適用於三年級至八年級。本測驗富有診斷的價值。

朗讀測驗中採用最廣者爲 *Gray Oral Reading Test*。本測驗的計分分速度與正確兩種。所謂正確即指錯誤之多寡。錯誤共分下列六種：(1) 完全讀錯，(2) 部分的讀錯，(3) 漏讀，(4) 代替，(5) 插入，(6) 重複。

(二) 拼法測驗 拼法測驗有單字式的與句子式的兩類。單字式的拼法測驗通用者有下列數種：

1. Ayres Spelling Scale

2. Iowa Spelling Scale 係愛普鮑 (E. J. Ashbaugh) 根據安特生的通用字彙而編造的。

3. Morrison-McCall Spelling Scale 共八類，難度大致相等，可以替換應用。

句子式的拼法測驗有 Mourne Timed Sentence Test，凡測驗字都編在句子裏邊，由主試者口讀，叫學生在規定時間內一句一句寫出來。

(三) 書法測驗 桑戴克的書法量表是美國最早的品質量表，以評判的差異度為計算成績的單位。愛里斯的書法量表則以「易於認識」為優劣之標準，與桑戴克所用方法頗有不同。費里門的書法診斷量表與 C. T. Gray 的記分片可作為診斷之用。

(四) 作文測驗 美國通用之作文量表有下列數種：

1. Hillegas Composition Scale 為美國最早的作文量表。

2. Nossan County Supplement to the Hillegas Scale 備有四年級至十二年級的常模。

3. Thorndike Extension of the Hillegas Scale 從四年級至十二年級均適用。

4. Hudelson Composition Scale 從四年級至十二年級均適用。

5. Willing Scale 從四年級至八年級適用。成績的評定分故事分數與形式分數兩項。

Spessey Diagnostic Tests in English Composition 包含下列四種測驗：大寫、標點、文法、造句。

(五) 文法測驗 通用者有 Wilson Language Error Test 及 Charters Language Diagnostic Test。

二 我國已編就的英文測驗

(一) 郝司氏文法與語文測驗 (Keys Grammar & Idiom Tests) 本測驗共有 F, G, H, I, J, K, L 七類，每類分兩部分，一部分採填字式，一部分採校正錯誤式。測驗中的題目係根據編者對於中國學生在英文文法上所犯錯誤之分析而編製，故深合中國學生的需要。七類難度不等：『F』、『G』兩類適合於二年至六年英文程度的學生，『H』、『I』、『J』三類適合於四年至八年英文程度的學生，『K』、『L』兩類適合於六年至十年英文程度的學生。成績的計算均以百分數為單位，百分數係根據十六歲兒童求得。

本測驗的可靠度從郝司氏在嶺南大學試用所得結果可以看出。結果如下表：

表十 郝司氏文法與語文測驗的可靠度

測 驗 類 數	F 與 G	I 與 J	H 與 K	L
年 級	中一上	中二下	中三下	六 一

學生人數	30	20	30	40
相關係數(r)	.89	.88	.85	.88
相關係數的均方錯誤(SD r)	.037	.05	.05	.026
兩類相差平均T分數	4.0	3.1	2.2	2.7

(二) 郝司氏字彙測驗 (Keys' Vocabulary Opposite Test) 本測驗共有 A B C 三類。A B 兩類各有單字 55 個，難度相等，可以替換應用，適合於有二年至八年英文程度的學生。C 類是從 A B 兩類改造的，有單字 35 個，適合於有二年至四年英文程度的學生。A B 兩類的相關度，據郝司氏試用所得結果在 .90 至 .94 之間。成績的計算均以 H 分數為單位，H 分數係根據十六歲兒童求得。測驗時限為 30 分鐘。

本測驗採用「求對」的方式，被試者須將測驗字相對的那一字（即意義與測驗字相反者）寫出，如測驗字為 "sod"，被試者應答 "pad"。被試者不但須了解測驗字的意義並且須能想起一個意義相反的字，所以測驗字雖祇有 35 個，實際上測量到的字却兩倍於此。

本測驗所用的 110 個字係從 200 餘字中選擇出來的。選擇時下列各字類均被淘汰。

1. 偏於測量智慧的字，如 both, day, gave, more, few。
2. 相反字太多的字，如 scatter 有下列數相反字：gather, collect, unite, concentrate, cluster, meet,

assemble, combine 等等。相反字太多，則計分的手續太繁。

3. 專門字，如 *conserve, vertical*。

4. 答案易於猜測的字如 *kind, subjective*。

5. 高年級答錯的百分比高於低年級的字，如 *asleep, obey, backward*。

6. 易於引起錯誤反應的字如 *normal* (許多人答 *primary, kindergarten, concrete* (許多人答 *mind, brick, stone*))

計分的方法頗為繁複，編者備有詳細標準答案及計分規則，詳見說明書。不贅述了。編者並根據六百餘廣東學生的成績規定各年級標準如表十一。

表十一 耶司氏字彙測驗的標準

聽過英文的單位	學 級	年 級	測 驗 時 期	百 分 之 二 十 點	中 數	百 分 之 八 十 點
2.5	小 學	第 五 年	年 年	26	30	34
5	小 學	第 六 年	年 年	30.5	35	39.5
7.5	中 學	第 一 年	年 年	34.5	39	43.5
10	中 學	第 一 年	年 年	38	42.5	47.5

13	中學	學	第二年	年	中	40.5	45	50
20	中學	學	第二年	年	終	43	47.5	52.5
25	中學	學	第三年	年	中	45	50	55
30	中學	學	第三年	年	終	47.5	52.5	58
35	中學	學	第四年	年	中	50	55	60.5
40	中學	學	第四年	年	終	52	57	62.5
45	中學	學	第五年	年	中	54	59	64.5
50	中學	學	第五年	年	終	55.5	61	66
55	中學	學	第六年	年	中	57.5	63	68
60	中學	學	第六年	年	終	59	65	70
65	大學	學	第一年	第一學期	終	60.5	66.5	71.5

大學入學的暫定標準—58

表十一的標準，係以讀過英文單位為根據，年級一行不過備考之參用，各學校英文科的分量有重有輕，以同一年級標準用於各校自非公平之辦法。讀過英文單位的計算，編者規定如下：

1. 一學年每星期25至30分鐘的英文以一個單位計。一學年每星期25至30分鐘的英文以1/2單位計。一學年每星期25分鐘以下的英文以半個單位計。

2. 其他學科如歷史、數學等，以英語教授者照英文科折半計算。所謂英語教授指班上討論用英語而言，僅採用英文教本不能作為以英語教授。

3. 如測驗在第一學期終舉行的，則那一學年的英文單位照1. 2. 規定數折半計算。如在第一學期中間舉行的，則照1. 2. 規定數之 $\frac{2}{3}$ 計算。如在第二學期中間舉行的，則照1. 2. 規定數之 $\frac{1}{3}$ 計算。

(三) 郝司氏聽意測驗 (Keys Auditory Comprehension Tests) 前面我們說，接受意義的途徑有閱讀與耳聽兩方面。默讀測驗所測量的是第一方面，即閱讀文字以收受意義的能力。郝司氏聽意測驗所測量的是第二方面，即聽人語言以收受意義的能力，換言之，即了解英語的能力。

本測驗共有「O」「P」兩類，難易相等，可以替換應用。每類有六十句問句，由主試者讀給被試者聽。每句問句有四個答案（答案是印在測驗紙上的），其中祇有一個答案是對的，被試者須將對的那一個答案劃去。現在舉數例如下：

郝司氏聽意測驗「O」類

1. 問句：

(1) How many is four and two?

(2) I am a teacher. What are you?

- (3) What do you say when your friend is going away?
- (4) Where do people go walking?
- (5) What do you do with your mouth?

2. 答案:

- (1) two four six eight
- (2) write student ro in no
- (3) glad good-bye thank you no sit
- (4) never house street often
- (5) walk hear speak look

本測驗對於各種英文程度的人都適用。若測驗一年至三年程度的，可用測驗中前面四十句問句；若測驗三年程度以上的，其餘二十句也須應用。

(四) 按氏混合英文測驗 (Anderson's Comprehensive English Test) 本測驗的目的在測量閱讀、聽話、與應用英文的能力。適用於中學各年級。內包括填字、字彙 (二種)、文法、看意、聽意、綴法七個測驗，茲分述如下：

測驗一為填字測驗 (Completion)，共有二十句子，每句空去一二字，令被試者在每一虛線上填一字，使成

為有意義的句子。時限為10分鐘。

測驗二為字彙測驗 (Vocabulary-translation) 共有五十個單字，令被試者將各字的意義用中文註在英文字旁。時限為6分鐘。

測驗三亦為字彙測驗 (Vocabulary-reading) 係根據桑戴克字彙測驗編成。共有一百個字，有屬於戰爭的，有屬於家族的，有屬於宗教的，有屬於數量的，有屬於地球的，令被試者在各字前面寫一相當字母，以示類別。時限為8分鐘。舉例如下：

Part III. Vocabulary (Reading)

W-War or fighting

F-Family or relatives

R-Religion or church

N-Number or quantity

E-Earth or its surface

priest	
altar	

nine	
ammunition	

double	
catechism	

numeral	
caneen	

cartridge	
munk	

第一欄 各字測驗或係宗教

	pair
	camp
	marriage

	more
	entrenchment
	singular

	chaplain
	duplicate
	two

	eternal
	nephew
	baptism

	psalm
	brigade
	gulf

測驗四為看音測驗 (Comprehension-visual) 共有二十五個問題，令被試者閱讀後用中文回答，時限為13分鐘。舉例如下：

Part IV-Comprehension (visual)

1. what is the thing to do when you feel hungry?

ans.

2. what is the thing to do when you feel hot?

ans.

3. what is the thing to do when you think your face is dirty?

ans.

測驗五為聽意測驗 (Comprehension-auditory) 共有四段文字，各自獨立，由主試者一段一段地誦讀，讀完

「一段被檢者聽筆答測驗紙上的問句，答案用中文第一段文字知。」

Part V-Comprehension (auditory)

A. Time for reading-thirty-five seconds

Time for writing-two minutes

“Long after the sun had set, Wang was still waiting for Tu and Pan to come. If they do not come before nine o'clock, he said to himself, I will go on to Shanghai alone. At half past eight they came, bringing two other boys with them. Wang was very glad to see them and gave each of them one of the apples he had kept. They ate these and he ate one, too. They all went on down the road.”

「試讀」以後，被檢者聽回答下面的問句。

First Paragraph

1. when did Tu and Pan come?
2. what did they do after eating the apple?
3. who else came besides Tu and Pan?
4. How long did Wang say he would wait for them?

第一段 試聽筆答測驗紙上的問句

測驗六爲文法測驗 (grammar)，共有二十個句子，其中或缺少了字或用錯了字，令被試者填入所缺的字或改正錯字。時限爲10分鐘。

測驗七爲綴法測驗 (composition)，題目是 "The Most Interesting Sight near School." 時限爲10分鐘。

關於校閱與計分的方法編者備有標準答案與計分規則。詳見說明書，不贅述了。編者並根據二十五個教會學校二千個學生的成績規定暫行標準與常模，見表十二、表十三。

表十二 按氏混合英文測驗的標準與常模：口分數

	小學第三年		中學第一年		中學第二年		中學第三年		中學第四年	
	上學期終	下學期終	上學期終	下學期終	上學期終	下學期終	上學期終	下學期終	上學期終	下學期終
Q ₀ 或標準	47	48	50	53	56	59	62	64	67	70
平均或常模	43	45	47	49	52	54	57	58	60	62
Q ₁	35	37	40	43	47	49	52	53	55	57

表十三 按氏混合英文測驗各部分的標準與常模：口分數

測驗一 填字口分數

	小學第三年		中學第一年		中學第二年		中學第三年		中學第四年	
	上學期終	下學期終	上學期終	下學期終	上學期終	下學期終	上學期終	下學期終	上學期終	下學期終
Q _s	46	48	51	54	58	60	62	64	66	68
平均	41	43	46	48	51	53	55	56	58	60
Q ₁	36	38	41	43	46	48	51	53	55	57

測驗二 翻譯口分數

Q _s	48	49	51	53	56	57	59	60	62	64
平均	43	45	47	49	51	53	55	56	57	58
Q ₁	37	40	43	45	47	49	51	53	55	57

測驗三 字彙口分數

Q ₀	46	48	51	53	55	58	62	63	65	67
平均	43	45	48	50	53	54	56	57	58	59
Q ₁	39	40	41	42	46	49	53	54	56	58

測驗四 看意思分數

Q ₀	46	47	51	53	56	58	60	62	64	68
平均	43	44	46	48	51	52	53	54	56	58
Q ₁	35	38	41	44	47	48	50	50	50	50

測驗五 聽意思分數

Q ₀	47	50	53	55	57	59	61	63	65	67
平均	47	48	49	49	50	53	56	58	60	62
Q ₁	40	40	40	43	47	48	50	52	55	58

測驗六 文法分數

Q ₂	47	49	51	53	56	58	60	62	64	66
平均	48	45	47	50	53	54	56	57	59	61
Q ₁	32	36	41	43	46	48	51	52	54	56

測驗七 綴法F分數

Q ₂	50	51	53	55	58	59	61	62	63	64
平均	47	48	50	51	53	55	58	59	60	61
Q ₁	35	38	42	44	47	49	51	52	53	54

表中各數目均為F分數，係根據十六歲兒童求得。上列標準與常模既然是根據教會學校學生的成績而規定，似不能作為全國的標準與常模，各學校英文科的分量有輕有重，繩以同一年級常模更非公平之辦法。

問 題

1. 我國中學英文教學的目標應如何規定？
2. 批評耶司氏的字彙測驗。

② 英語教學法 吳大猷編 中華書局

參考文獻

- 1 Monroe, W. S., DeVoss J. C., Kelly, F. J.: *Educational Tests and Measurements*, Ch 3-6 Houghton Mifflin.
- 2 Wilson, G. M. and Hoke, K. J.: *How to Measure*, Ch. 2, 3, 5, 6, 7, Macmillan.
- 3 Anderson, E. J.: *Comprehensive English Tests*, *Teacher's Manual*, Commercial Press
- 4 Keys, N.: *English Mastery Tests*, *Teacher's Manual*, Edward Evans and Sons.

第十一章 自然史地科及各科混合測驗

前幾章所討論到的國文、算術、英文都是工具的學科，目標較為確定，且着重在技能的訓練與智識的獲得，因此進行測量亦較為簡易。本章所討論的自然科與史地科，一則目標較為籠統，二則着重之點不在技能與事實而在態度、理想、解決問題的能力這些方面，因此四測驗方法的應用就困難得多了。茲就我國已有的測驗分述如下：

一 自然科測驗

廖世承等所編之中學混合理科測驗為測量中學學生理科常識之用，內容包括物理、化學、生物、生理衛生等科目。測驗共分兩類，每類有一百個題目，可以替換應用。題目採用四個答案的彙選式。時限為30分鐘。自初中一年級至高中三年級均適用。

本測驗備有百分數常模與年級常模，百分數係根據十六歲兒童求得，以下各中學測驗均同此。

二 史地測驗

(一) 徐氏中學本國史測驗 本測驗為歷史學家徐則陵氏所編，只有一類而常模亦付缺如。測驗中題目

係根據下列六條原則而擬定的：

1. 史事與其發生時期之關係
2. 辨別史事發生之先後
3. 史事與其發生地點之關係
4. 史事與其人物之關係
5. 史事因果之關係
6. 辨別史事之輕重

本測驗共六十題，分六段，每段十個題目，代表一種原則。題目採用五個答案的彙選式。

(二) 東大附中廖蔭二氏混合歷史測驗 本測驗的內容包括本國史與外國史。題目的範圍很廣，包含中外的典章、文物、制度、名人、學說、發明、宗教、都會、戰爭、黨禍、會議、條約、交通、商埠、紀元等等，共一百題。題目採用四個答案的彙選式。自初中一年級起至高中三年級止均適用。時限為25分鐘。

(三) 楊氏本國地理測驗 本測驗共有兩類，每類有五十二個題目。題目係採用四個答案的彙選式。自小學三年級至初中二年級均適用。測驗中的題目是根據中小學普通地理科本編的，大概包含下列幾方面：

1. 關於普通地理智識的，包括各地方的氣候、山脈、河流、都會、省區的位置。

2. 關於推理的：這類問題須用思想方能解答，不是專靠記憶的。

3. 關於交通的。

4. 關於了解地理上名詞的。

5. 關於各地物產的。

(四) 東大附中廖璋二氏混合地理測驗 本測驗的內容包括本國地理和世界地理。題目的範圍很廣，包含中外的各地的氣溫，火山，地震，潮汐，颶風，經緯度，瀑布，沙漠，風景，河域，水上都會，都城，面積，通商口岸，運河，鐵道，農產，礦產，人口，商業，國債，政策，宗教，人物，建築等等。測驗共兩類，每類有六十六題，係採用四個答案的彙選式。自初中一年級起至高中三年級止均適用。時限為七分鐘。

三 各科混合測驗

(一) 俞子夷小學社會自然測驗 本測驗包括公民，歷史，地理，自然，衛生各種科目。測驗共分二類，每類有二十題，題目係採用四個答案的彙選式。自小學一年級起至初中一年級均適用。時限為七分鐘。

(二) 查良釗學校調查用教育測驗 本測驗專為學校調查之用，共有兩類，每類有一百四十個題目。一百個題目係採用四個答案的彙選式的，包括 1. 歷史，2. 地理，3. 自然，4. 默讀，5. 應用算術五科目，各科目二十題。題目

的排列以難易為標準，各科題目相混，並不分別。末了四十個題目係算術四則題，加減乘除各十題。全部測驗時限為60分鐘，前一百題6分鐘，後四十題15分鐘。從小學三年級至初中二年級均適用。本測驗的日分數係根據十二歲兒童求得的。

本測驗的材料係根據四條原則選擇來的。第一，在實際生活上的重要性。第二，普遍性，凡帶地方色彩太深的試題均刪去。第三，對於小學三年級至初中二年級學生所受教育的適合性。第四，指示以上各年級教學的方針。

(三) 德爾滿學校自用普通教育測驗 本測驗共有三類，第一類為初級中學之用，第二類為高級中學之用，第三類為大學文科之用。據編者說，本測驗有三種特點：1. 內容實用而合時，2. 編製科學的劃一的，3. 施行簡單而經濟。本測驗的功用，約有五端：1. 可作為入學考試之用，2. 可為全校重新分級之用，3. 可為建立全國各種學校各年級的標準之用，4. 可使各校各班與全國的或區域的常模比較，5. 可作為給予轉學生學分之根據。

四 其他測驗

(一) 陳鶴琴小學常識測驗 本測驗共分兩類，每類一百題，採用四個答案的彙選式，內容包括兒童常見常聞或應當知道的各種事物，如動物、植物、物產、天然物、氣候、時節、度量、時間、顏色、生理、物品、地理、圖形等。時限為30分鐘。從小學三年級至初中一年級止均適用。

(二) 俞氏圖畫量表 俞子夷曾編有一種圖畫量表，惟尙未付印。

問 題

1. 歷史教學的目標何在？
2. 批評徐氏本國史測驗取材所根據之原則。
3. 常識測驗有什麼功用？

參考書報

1. Monroe, W. S., DeVoss, J. O., Kelly, F. J. Educational Tests and Measurements, Ch. VII. Houghton Mifflin.
 2. Wilson, G. M. and Hoke, K. J., How to Measure, Ch. 12, 13 Macmillan.
 3. 陳傑夫：化學測驗，中等教育第二期。
 4. 陳傑夫：物理測驗報告，中等教育第二期。
 5. 陳容光：中學科學測驗，心理三卷四號。
- 第二編 各科測驗與新式考試

教 育 測 驗

6. 陳鶴琴：小學常識測驗，心理三卷一號。
7. 廖世承：陳傑夫：混合理科測驗說明書（商務）。
8. 徐則陸：中學本國史測驗一類（商務）。
9. 廖世承：蘇毓芬：混合歷史測驗說明書（商務）。
10. 楊國荃：本國地理測驗說明書（商務）。
11. 廖世承：辜潤琦：混合地理測驗說明書（商務）。
12. 俞子夷：小學社會自然測驗及說明書（商務）。
13. 查良釗：學校調查用教育測驗及說明書（商務）。
14. 德爾福：學校自用普通教育測驗及說明書（商務）。
15. 陳鶴琴：小學常識測驗及說明書（商務）。

第十二章 舊式考試與新式考試的比較

一 引論

從有教育以來就有考試的制度。所謂舊式的考試是讀者所熟悉的；截止最近為學校中惟一的考試的方式。近來舊式考試的缺點一一暴露，舊式考試的不可靠，已有許多實驗的結果可以證實，因此引起大家對於考試方法的研究。前面所討論的標準測驗就是應此種需要而起的。但是標準測驗在應用方面亦有許多的限制。第一，標準測驗編造的手續是極繁重的，不是普通學校教師所能勝任。第二，已有的標準測驗，其內容，目的，常常不能與教員當時教學情形相符合，不合於教員當時的需要。例如教師授完鴉片之戰一段的歷史，要舉行一種考試以考查學生對於鴉片之戰一段歷史的智識。在坊間雖然有許多歷史的標準測驗，但是沒有專門關係於這一段歷史的。教員必須自己編造一種可靠的考試。我們在本章及下章內所討論的新式考試就是應這種實際的需要而起的。一方面可以避免舊式考試的種種缺點，一方面可以補標準測驗的不足。

編造新式考試的基本原則與標準測驗是一樣的。我們的目的就在得着一種正確的，可靠的，客觀的考試。不過在編造手續上比較簡單，而且是教員自己可以隨時編造的。在此處我們必須申明：舊式考試雖然有許多缺點，

但是亦有其特點，我們並不主張把他完全廢棄不用。舊式考試的特點有二：第一，應用舊式考試時，學生有機會自由發表意見，不像新式考試的那樣機械、呆板。第二，舊式考試可以訓練學生用文字發表思想的能力。這雖然不是考試主要的目的，但亦是一種很有價值的訓練。

總結以上的討論，學校中可以應用的考試有三種方式：（一）舊式考試，（二）新式考試，（三）標準測驗。以上三者，各有各的特長，各有各的限制。

二 學校考試的功用

學校中考試制度，由來已久，在教學上，行政上，占有重要的地位。近來有人鑒於考試的流弊，遂有取消考試的主張。其實考試的流弊，是因為考試方式的不良及應用的不當，並非考試制度本身的弊端。我們以為比較穩健的態度，在改良考試，使流弊無由發生，而不在于根本取消考試的制度。

學校考試的功用，綜而言之，可有四端：（一）鼓勵學生的學業。（二）保持學校的標準程度。（三）學業的考查、核算。（四）考查學生的缺點、弱點，以為改進教學的根據。

薩賓司(P. M. Symonds)曾列舉學校考試的功用十五項，頗為詳盡，茲譯錄於下。

（一）使學生知道自己的成績。

- (二) 鼓勵學業。
- (三) 增進競爭 1. 團體間的競爭；2. 個人間的競爭；3. 各人與已往成績的競爭。
- (四) 決定升降。
- (五) 診斷學生的缺點，弱點。
- (六) 決定教學的優劣。
- (七) 決定中學投考生的取捨。
- (八) 決定新生應插入的年級。
- (九) 決定大學投考生的取捨。
- (十) 將學生成績報告家長。
- (十一) 決定獎狀的給與。
- (十二) 以爲教育指導與職業指導的根據。
- (十三) 評判教員的效率。
- (十四) 推測學生將來的成就。
- (十五) 考查學校的效率。

三 舊式考試的缺點

舊式考試的缺點，舉其大者，約有兩端：（一）定分的主觀，（二）取樣的不適當。茲分別論之。

（一）舊式考試定分的主觀，是大家所熟悉的。關於這一點，我們有許多實驗的結果，可以證明。

1. 約翰生 (F. W. Johnson) 的調查 表十四表示芝加哥大學附屬中學各種功課分數的分配。教員的給分，極不一致，是很顯明的。學生在德文班上得 A 等的機會，差不多兩倍於在法文班上得 A 的機會。英文班上不及格的學生竟三倍於家政班上不及格的學生。

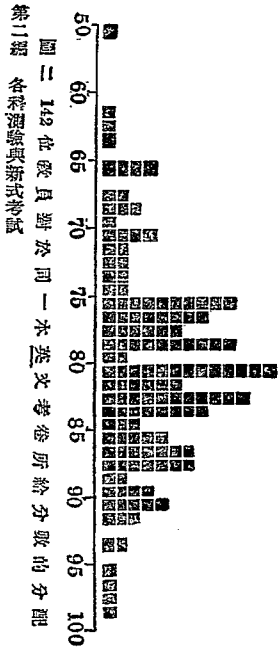
表十四 芝加哥大學附屬中學各科

分數的分配 (從 F. W. Johnson)

科目	分總數	F %	D %	C %	B %	A %
希臘文及拉丁文	868	10.6	16.1	31.8	23.5	17.9
德文	416	8.4	19.5	26.4	23.0	17.1
法文	475	10.9	18.7	33.0	23.0	9.3
英文	1514	15.5	21.7	32.8	23.4	6.5

數	學	1466	14.5	25.2	27.6	21.1	11.5
歷	史	825	8.1	15.9	31.2	30.0	14.7
科	學	672	8.3	16.8	27.7	32.6	14.6
家	政	176	5.7	2.3	27.3	51.7	13.1
平	均	(7297)	11.5	18.9	30.6	27.0	12.0

2. 斯達奇 (D. Starch) 及伊立歐忒 (E. C. Elliott) 的研究 斯達奇及伊立歐忒將一本英文考卷請一百四十二位英文教員評閱給分, 圖二表示這一百四十二個分數的差異的情形。他們又將一本幾何考卷請一百四十五位中學數學教員評閱給分, 結果如圖三。



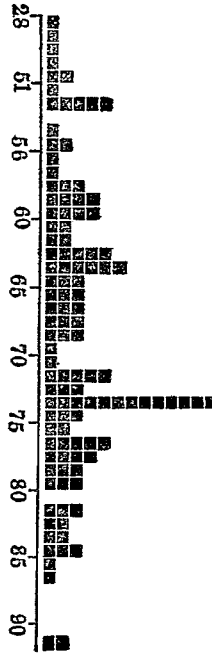
第二節 各種測驗與評卷

表十五 九十一位教員對於一個地理問題的三個答案所給的分數 (從 Ruch)

分數	答案一	答案二	答案三
20	1		9
19	0		3
18	1		21
17	1		12
16	2		17
15	10		15
14	2		3
13	8		1
12	14		3
11	5		1
10	24		3
9	5		1
8	7		1
7	3		0
6	4		0
5	1		0
4	0		0
3	2		1
2	1	3	0
1	0	1	0
0	0	87	0
次數	91	91	91
平均	10.9	0.1	16.1
均方差	3.2	0.4	2.9

3. 羅雪的研究
 羅雪會重複斯達奇及伊立歐忒的研究。他將對於一個地理問題的三個答案請九十一位教員評閱給分，給分的標準從0分起到二十分止。這九十一位教員所給的分數如表十五。

圖三 116位中學教員對於同一本幾何考卷所給分數的分配



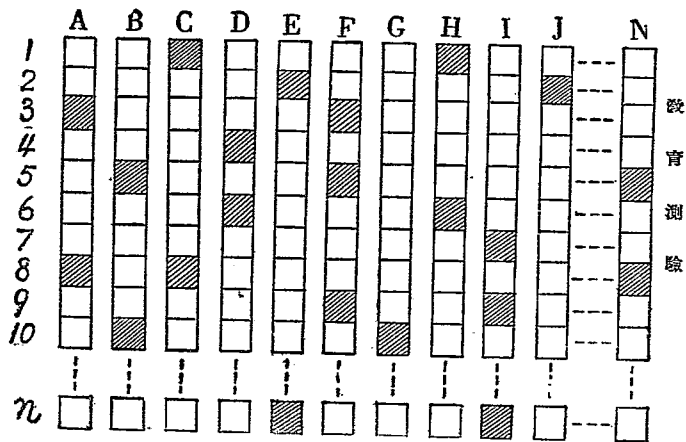
以上各種研究的結果，都表示同樣的趨勢。羅雪曾說：「與其從事於舊式考試的改進，不如變換考試的方式。人們的評判究竟是不可靠的。舊式考試計分的方法是祇能在極小範圍內改良的。」

吳德 (B. D. Wood) 曾述及一件很滑稽，但是真確的事：一九二〇年的夏季，有許多大學歷史教授在評閱歷史的試卷。其中有一位教授，為評閱便利起見，自己做了一本卷子，以為模範。不曉得怎樣，這本模範卷子同其他的考卷錯混在一起。另一位教授照例的拿起這本卷子來評閱，對於這本模範卷子，竟給了一個不及格的分數。為審慎起見，其他教授對於這本不及格的卷子，重複評閱。結果這本模範卷子所得的分數從四十分起，到九十分止！

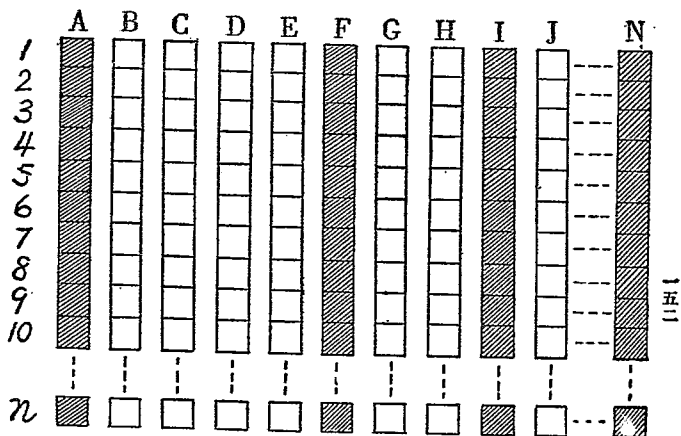
(11) 取樣的不適當 舊式考試第二個大缺點，就是取樣不適當。換一句話說，考試的內容不能代表教材的全部。

我們應該認清考試無論如何總不能包括教材的全部，總是一種有限制的取樣。取樣可有兩種方法；我們可以稱之為深的取樣及廣的取樣。這兩種取樣的方法的不同，可以由圖四、圖五表示之。

圖中 A, B, C, D, …, N 各直行代表教材中的各大段落，各大部分。圖中 1, 2, 3, …, n 代表各大段落中的各小項目，各單零事件。



圖四 表示廣的取樣，圖中各黑方塊代表考試已包括的項目（從Ruch）



圖五 表示深的取樣，圖中各黑方塊代表考試已包括的項目（從Ruch）

舊式考試中題目不過五個六個，但是對於每一個題目却希望學生回答得很詳盡。這是採用深的取樣。新式考試中題目少則五十，多則一百，二百不等，但是每個題目的範圍極狹窄。這是採用廣的取樣。所以舊式考試與新式考試所採用的取樣的方法是根本不同的。從理論上講，廣的取樣較為適當。

四 新式考試的優點及限制

我們可以直捷了當的說，新式考試的優點，就是舊式考試的缺點；新式考試的限制，就是舊式考試的特點。新式考試的利益，有下列數端：

- (一) 定分的客觀。
- (二) 採用廣的取樣。新式考試的內容比較的可以代表教材的全部。
- (三) 在相等時間內由新式考試所得分數較由舊式考試所得分數為可靠。關於這點，我們有吳德羅，杜際司 (H. A. Toops)，孟廠諸氏的實驗結果為根據。
- (四) 校閱的便利，校閱時間的節省。
- (五) 新式考試中每個問題的範圍是確定的，所以對於學生的答案可有較切實的控制。學生不易以假智識來相混淆。

(六)新式考試的教育影響較大。新式考試的校閱既然有客觀的標準，所以教員可利用學生自己校閱自己的考卷。學生考試完畢的時候，正是最好的教學的機會。在這個時候，學生興趣正濃，極願意曉得他那部分做對的，那部分做錯的。教員如能利用這個機會叫學生自己校閱自己的卷子，對於學生答錯的問題，隨時提出討論，必能收極大的效果。

(七)教員及學生都比較的歡迎新式考試。

新式考試的限制，可有下列數端：

(一)新式考試不能予學生以盡量發表思想，見解的機會。

(二)新式考試不能訓練學生以文字發表思想的能力。

(三)新式考試，如應用不當，易流於機械化，使學生專着重強記而忽略理解。

問題

1. 標準測驗在應用上有什麼限制？
2. 主張廢棄學校考試的有什麼理由？這些理由是否充足？
3. 「一次考試無論如何總是一種有限制的取樣」這句話如何解釋？

4. 何以廢的取樣總數的取樣為適宜？
5. 新式考試與舊式測驗的區別何在？
6. 每科考試有什麼特殊的功用？

參考書目

1. McCall, W. A.: How to Measure in Education, P. 123—133, Macmillan.
2. Odell, C. W.: Traditional Examinations and New Type Tests, Ch. 2, 7, Century.
3. Ruch, G. M.: The Objective or New-Type Examination, Ch. 3, 4, Scott, Foresman.
4. Ruch, G. M. and Stoddard, G. D.: Tests and Measurements in High School Instruction, Part III, World Book Co.
5. Russell, C.: Classroom Tests, Ch. II, Ginn.
6. Synonds, P. M.: Measurement in Secondary Education, Ch. III, Macmillan.
7. Wood, B. D.: Measurement in Higher Education, World Book Co.
8. 心理測驗法總論 (四編)
9. 心理測驗法總論 (四編)
10. 測驗法 I 與測驗法 II 參考書目

第十三章 新式考試的編造法

上章我們泛論新式考試與舊式考試的優劣，現在進一步討論編造新式考試的方法，步驟。我們分三部分來討論。

一 新式考試的種類或方式

新式考試的方式繁多，茲就其通用者舉例以說明之。

(一) 回憶式

例一 (地理)

下列各省中最大的城市是什麼？

江蘇

安徽

浙江

河北

例二 (歷史)

將下列問題的答案寫在下面空行上。

1. 中華民國第一任大總統是誰？
.....
2. 五權憲法是何人創始的？
.....
3. 鴉片之戰發生在何年？
.....

例三 (文學)

1. 原道篇是何人做的？
.....
2. 曹雪芹最著名的著作是什麼？
.....

(二) 填字式

例一 (測驗學)

在下面空格裏將相當的字填入，每一空格填一字。

1. 如其他情形相等，測驗愈——其可靠度愈——。
2. 教育的目的乃在——人類的——。

例二 (英文)

1. His coat is more——two fee' long.
2. They did not succeed——catching the thief.

(三) 正誤式

例一 (歷史)

下列的句子，如其你以為對的，請在句子前面空行裏作「十」號；如其你以為錯的，請在空行裏作「C」號。

十1. 胡適之提倡白話文字。

○2. 戊戌政變的主人翁是孫中山。

例二 (測驗學)

○1. 第一個智力測驗是 McCall 所編造的。

十2. TROBE 量表是 McCall 發明的。

十3. 陸志章先生曾訂正比納西門智力測驗。

(四) 是否式

例一 (地理)

1. 中國最長的河流是否長江？ 是 否

2. 河北省是不是中國最大的省分？ 是 否

例二 (測驗學)

1. 中國的測驗是否以百分數為單位的？ 是 否

2. 桑戴克氏是否到過中國，曾幫助中國編造許多測驗？ 是 否

(五) 異同式

例一

倘使下面虛線上下的字句是相同的，在虛線上寫個「○」號，若是不同的，寫個「×」號。

1. 冷×熱。

2. 完全的○整個的。

例二

1. 革命的……演進的 同 異

2. 特殊的……例外的 同 異

3. 積極的……消極的 同 異

(六) 彙選式

例一 (兩種反應)

1. The children (sung, sang) several song.

2. She sang (beautiful, beautifully).

例二 (三種反應)

1. 一點鐘分：(1) 60分, (2) 50分, (3) 45分。……………(1)
2. 寒暑表的用處是：(1) 看時候, (2) 測溫度, (3) 記風的速度。……………(2)

例三 (四種反應)

說明：下面每一個問題後面有四個答案，其中有一個是對的。請你把對的答案的數目寫在問題前括弧內。

(4) 1. 一英尺有：(1) 六寸, (2) 八寸, (3) 十寸, (4) 十二寸。

(3) 2. 旅行沙漠最合宜的獸是那一種：(1) 牛, (2) 馬, (3) 駱駝, (4) 象。

(七) 最好答案式

例一

中國近年來實業衰落是因爲：

1. 帝國主義的侵略。
2. 連年內戰不息, 內政不修, 地方不靖。

3. 人民缺乏毅力不肯向內地投資。

(八) 配合式

例一 (完全配合)

著者	答案	書名
1. 胡適	3	中西文化及其哲學
2. 梁啟超	1	中國哲學史大綱
3. 梁漱溟	2	中國歷史研究法

例二 (非完全配合)

將下列人名及事件配合起來

人名	事件
1. 蔡謬	9 二次革命時被暗殺
2. 梁任公	7 謀復辟不久失敗
3. 胡適	1 雲南起義
4. 胡漢民	10 馬廠督師

第二編 各科測驗與新式考試

5. 袁世凱 文學革命

6. 宋教仁 帝制失敗

7. 張勳

8. 蔣中正

9. 徐世昌

10. 段祺瑞

(九) 比喩式

1. 中國 亞洲 法國

2. MM MM 0

3. 上海 在中國 有如 之在美國

(十) 排列式

將下列事件依其發生的先後排列起來。

() 馬廠誓師

() 五四運動

() 直院戰爭

() 北京改名北平

() 孫中山先生逝世

(十一) 分類式

例一

下列五個人名中，有一個與其他不相類的，在那人名旁劃一直：

康有爲

胡適

章太炎

段祺瑞

梁啓超

例二

下面每組內有五個名詞，這五個名詞中有一個與其他四個不同類的，你把這個不同類的名詞的數目寫在各組後面空括弧裏。

1. (1) 元 (2) 哉 (3) 初 (4) 末 (5) 首……………(2)
2. (1) 殘 (2) 嫉 (3) 媚 (4) 毒 (5) 妬……………(3)
3. (1) 紅 (2) 白 (3) 藍 (4) 黃 (5) 高……………(5)

二 編造規則

有許多編造的規則是各種方式相同的；有的是個別的。爲便利起見，仍分別討論之。

(一) 正誤式

1. 避免無關重要的問題。
2. 避免一部分對一部分錯的句子。
3. 避免故意難人的問題。
4. 避免太長的句子。
5. 避免富有暗示性的詞，如「絕對」、「完全」、「不可能」等。
6. 用普通日常所用的字，避免不常見的，艱難的字。
7. 不要使考試有相互關係的句子，使回答其一就可以回答其他。
8. 考試中對的和錯的句子的數目應大致相等。
9. 考試中對的和錯的句子的排列不應有一定的次序。
10. 正誤式考試中句子數目不應太少。

11. 考試中的問題不要僅僅根據常識或普通智力就可以回答的。

12. 計算正誤式考試的分數應用下列公式，以校正機遇。 $\frac{N - 1}{N} \times \frac{S - 1}{S}$

例如某生在正誤式考試上做對30題，做錯10題，其應得之分數為30減10等於20分。

13. 正誤式考試的格式如下列：

—— (1)
—— (2)

(二) 回憶式

1. 問題須是一種極簡單的反應可以回答的。

2. 寫答案的空格最好排在一直行上，使便於校閱，如下列格式：

(1) _____ ?

(2) _____ ?

(3) _____ ?

(三) 填字式

1. 一句子不應有過多的空格，把一句中幾個最有關係的字空出就可以了。

2. 句子最好不是從教科書中直抄下來的。

(四) 彙選式

1. 彙選式考試中，答案愈多，則猜對的機會愈少，其可靠度則愈高。在可能範圍內，應該有四個或五個答案。

2. 同一考試中，每問題後面答案的數目須一律。

3. 錯的答案不應太明顯，都須與問題有相當的關係例如：

中國第一任大總統是：(1) 孫中山 (2) 袁世凱 (3) 徐世昌 (4) 梅蘭芳

其中第四個答案的錯誤太明顯，如易以「黎元洪」則較妥。

4. 對的答案的位置，不應有一定的排列式。

5. 第一，第二，第三，第四答案是對的答案，次數應大致相等。

6. 說明中應警告學生不要亂猜。

7. 校正機遇的公式如下：

$$S = R \cdot \frac{P}{n-1}$$

其中

S = 分數

R = 做對的題數

W = 做錯的題數

N = 答案的數目

例如某生在四個答案的彙選式考試中，六十個問題中僅知道二十個問題。這二十個問題他當然可以做對。在其餘的四十個問題中，他憑機遇，可以猜對四分之一，或十個問題。因此他在這個考試中做對的題數是三十，做錯的題數亦是三十。照公式他應得的分數為：

$$S = 30 - \frac{30}{4} = 20$$

這個分數恰恰代表他真正的程度。

8. 彙選式的格式如下：

1. ———— (1) ———— (2) ———— (3) ———— (4) ————
.....
2. ———— (1) ———— (2) ———— (3) ———— (4) ————。

(五) 配合式

1. 配合式問題中，須配合的對數應在十與二十之間。對數太少，則機遇太大；對數太多，則做起來不容易，太耗

費時間。

2. 如對數在十以下，最好應用非完全配合式，以增加其可靠度。

三 編造新式考試的步驟

新式考試的編造，應有下列各步驟：

(一) 製造教材內容表，並決定內容各部分的輕重。上面我們一再申說增加一種考試的正確度可靠度的一個重要條件就是取樣的適當，使考試內容足以公允的代表教材的全部。所以在編製問題時，我們必須顧到教材的各部分，並對於各部分之輕重亦予以公允的估計。下表可以表示教材內容表的内容及其功用。

表十六 教材內容表：測驗應具的條件（根據本書第三章）

項 目	百分比
I. 正確性	20%
1. 何謂正確性	
2. 如何考查	
II. 可靠性	20%
1. 何謂可靠性	
2. 如何考查	
3. 影響可靠性的因素	
(1) 客觀性	
(2) 適當的取樣	
(3) 被試者的反應	
III. 鑒別力	10%
1. 何謂鑒別力	
2. 影響鑒別力的因素	
IV. 常模	15%
1. 常模的功用	
2. 求常模的困難	
V. 應用的便利	10%
VI. 診斷的價值	10%
VII. 覆份	10%
1. 何謂覆份	
2. 覆份的價值	
VIII. 經濟	5%
總 計	100%

(二) 草擬題目 根據教材內容表，草擬題目，使教材各部分，得到相當的注重。在草擬題目時，不妨多擬幾個，以備後來的選擇。

(三) 決定題目的數目或考試的長度 決定題目的數目應根據於兩個標準。第一，教材內容的多少。教材豐富，當然可以多出幾個有價值的題目。教材缺少，沒有多少有價值的題目可以出，當然應該少出幾個。第二，考試時間的長短。時間長，題目不妨多出些；時間短，當然祇能少出些。

(四) 修正並選擇題目 在字句上，文字上不妥善的，應加以修正，或者竟加以淘汰。

(五) 評判題目的難度，依其難度排列起來。評判題目的難度，因時間上的限制，祇能根據教員個人的意見。然後依其難度，排列起來，容易的題目在先，難的題目在後。

(六) 規定說明 說明須簡單，須清楚。

(七) 規定記分的規則。

問 題

1. 試就填字式，選擇式，排列式，配合式各舉一例。

2. 選擇式與最好答案式有什麼區別？

第二編 各科測驗與新式考試

3. 照校正權過的公式計算分數，有什麼不公平的地方？
4. 在四個答案的甄選式考試中，某生做對了8題，做錯了9題，照校正權過的公式，該生應得幾分？
6. 比較各種新式考試方式的利弊。

參考書報

1. O'Neil, C. W.: Traditional Examinations and New-Type Tests, Ch. 9—16, Century.
2. Ruch, G. M.: The Objective or New-Type Examination, Ch. Part II, III, IV, Scott Foresman.
3. 測驗第二期，新法考試專號。
4. 艾偉：測驗學上對減錯計算法之研究，測驗期刊號。

第三編 測驗的實施

第十四章 施行測驗的步驟及手續

在測驗運動的初期，測驗的舉行大多為普通調查之用，施行測驗的責任大多由教師以外的人員負擔，所以測驗與教學並不能發生若何密切的關係。近來一般從事教育者才覺悟，要使測驗對於學生的學業發生最大的效果，施行測驗的任務大部分必須由教師負擔起來，因為教育機關內與學生接觸最密，關係最切者莫過於教室內的教師。所以實施測驗的技能已成爲現代教師的專業訓練的一重要部分了。

一 舉行測驗的步驟

舉行測驗的步驟大致如下：

- (一) 規定問題。
- (二) 選擇適當的測驗。

第三編 測驗的實施

- (三) 施行測驗以前應有之準備。
 - (四) 施行測驗。
 - (五) 校閱測驗卷子，求各種分數。
 - (六) 分析並解釋結果。
 - (七) 改進測驗結果所表示的不良現象。
- 茲分條討論之。

(一) 規定問題 應用測驗者必須認清測驗乃行政上教學上一種重要的工具，其功用乃在顯示學生的程度，教學的效果。測驗本身並不能改進教學，舉行測驗更不能解決問題。無目的的施行測驗乃是對於測驗的意義根本誤解的徵象，亦是教育事業中很大的浪費。所以舉行測驗的第一步驟就是明確規定問題，其解決須有賴於測驗者。

教育中須賴測驗解決的問題甚多，大致可別為教師的問題與教育行政人員的問題。茲各舉數例如下：

1. 教師的問題

- (1) 考查學生在開始教學時的程度。
- (2) 決定學生在一月內或一學期內的進步。

(3) 學生成績與常模的比較。

(4) 一班學生各學科成績的比較。

(5) 診斷學生在各學科中缺點之所在。

(6) 考查學生努力的程度。

2. 學校行政人員的問題

(1) 能力分組。

(2) 選擇學生以組織特別班——低能班或高才班。

(3) 學生成績與常模的比較和與其他學校的比較，以決定學校的效率。

(4) 考查學校對於各學科的注重是否得其均衡。

(5) 各種教學方法的比較，以決定其優劣。

(6) 考查一學校，一級的智力及教育程度。

(7) 決定新生應插入的年級。

(二) 選擇適當的測驗 選擇測驗時我們可以利用第三章內所列測驗應具條件以爲選擇的標準。茲爲讀者便利起見，將各選擇的標準重新列下：

1. 測驗是否正確?
2. 測驗是否可靠?
3. 測驗是否客觀?
4. 測驗所測量的是否代表教育中重要的目標?
5. 測驗有無適當的常模，以資比較?
6. 施行及校閱便利否?
7. 測驗的性質與須待解決的問題適合否？譬如我們的問題是考查一學校的教育程度，那麼應該用調查測驗或普通測驗；假使我們的問題是個人的診斷，那麼非用診斷測驗不可了。
8. 測驗有無覆份？
9. 測驗的價格多少？

(三) 施行測驗以前應有之預備 施行測驗以前，主試者必須了解並熟習施行測驗的方法。大致說來，團體測驗比較個人測驗容易施行。個人測驗如比納西門智力量表，其施行的手續甚為繁複，必須經過長期訓練的人方能勝任。

施行測驗前的準備包括：

1. 熟讀測驗的說明書。

2. 在必要時，最好事前舉行小規模的實習，以免臨時的忙亂錯誤。

(四) 施行測驗 施行測驗時，有幾條普通的原則必須遵守的。

1. 主試者必須得到被試者的信仰與合作，引起其最大的努力，戰勝其懦怯的心理。不要造成過分緊張的空氣，以免被試者的慌張；不要有不自然的態度，使被試者跼促不安。

2. 施行測驗時須絕對依照標準的手續。但是有意外的情形發生時，主試者須能隨時隨機應變。不遵照標準的手續與死守成規不知應變都不是正常的辦法。隨機應變這一點，在舉行個人測驗時尤為重要。

3. 測量的環境須安靜，自然。舉行測驗時最好不要有人參觀。

4. 如測量數班學生，而所得結果須互相比較時，則各班測量的情境務求一律，如主試人員，舉行測驗的時間等。

(五) 校閱試卷 校閱試卷時務須依照說明書中所規定的方法。計算分數時務求正確。在可能範圍以內，每本卷子須經過兩人或兩次的校閱。

(六) 分析並解釋結果 將個人或一班的分數與常模比較，然後參照個人或一班他方面的事實，對於所得結果加以一種解釋。

(七) 解決問題 舉行測驗最普遍的弊端，就是測量完畢，校閱完畢，計算完畢，什麼事都了了，而對於如何利用所得結果一點就不顧問。這實在是極大的錯誤，極大的浪費。舉行測驗並不是為裝門面，是因為有問題要解決才舉行的。所以應用測驗時，我們應該每次自問「測驗後怎樣？」

二 施行測驗的詳細手續

廖世承陳鶴琴會列舉施行測驗的手續十三條，茲引述於後：

- (一) 始終保持和悅的態度，不要使學生害怕或刺激過甚。
- (二) 學生坐次須適宜，務避去外來的擾亂。
- (三) 學生桌上不要有什麼東西。未分卷子以前，主試應叫被試把各人桌上所有的東西，一律放在抽屜裏邊，以免妨礙動作。
- (四) 發鉛筆：請每行第一人散給後面的人。
- (五) 發試卷：依每行人數的多寡，發卷子與每行第一人，叫他向後傳。若被試者年齡太小，主試者須一個一個的發給他們，以免錯誤。卷子發後，再問一聲是否每人都有卷子，以免遺漏。
- (六) 填卷面空白。

(七) 倘使遲鈍的學生沒有填寫好卷子面的空白，主試不要開始說明測驗的做法。說明做法時，務引起各學生的注意，喊「做」以後，有必要時，主試可在室內迅速巡行一周，看各學生有沒有翻差頁數或找不到測驗開始的地方的。

(八) 試驗時須絕對依照測驗的說明。如不能背誦，最好看了說明書讀。在必要時，亦可以通融，對於標準的說明予以增減。但這是例外的。更改標準說明的目的乃在使學生了解進行的方法，並不是幫助他去回答測驗內的問題。

(九) 主試講話須清楚，快慢適當，聲音可以使全室的人聽見。對於說明中應注意之點，語氣須格外加重，使聽者容易明瞭。主試講話的態度須鎮靜，使學生能立刻遵照他的話去做。

(十) 禁止偷看或抄襲。

(十一) 計時要準確。最好用碼錶，至少須用有鈔針的錶。測驗開始以後，立刻把開始的時間寫在黑板上或空白紙上，再加上測驗所須要的時間，就得到停止測驗的時間了。例如：

測驗開始時間	11點	18分	20秒
測驗所須時間		5	
停止測驗時間	11	23	20

第三編 測驗的實施

- (十二) 收卷子：叫每行學生將卷子向前面傳遞到第一個人，然後由主試向各行收齊。
- (十三) 收鉛筆。

問題

1. 舉行測驗前何以須規定問題？
2. 選擇測驗須根據什麼標準？
3. 施行測驗何以須依照標準手續？
4. 要戰勝兒童的協怯，獲得其信仰與合作，有什麼方法可用？

參考書報

1. Ternan, T. M.: *Measurement of Intelligence*, Ch. 8, Houghton Mifflin
2. 廖世承——陳鶴琴：測驗概要第十三章（商啓）
3. 潘之英：教育測驗與統計第十二章（世界）
4. 各測驗說明書。

第十五章 測驗結果的整理(上)

測驗的結果必須加以整理，其意義才能明顯。結果的整理必須應用統計的方法。本章將統計方法中最基本，最淺近又最切於實用者詳為介紹，俾學者對統計中通用的量數的意義有明確的了解，對於通用量數的計算有熟習的技能。最重要的統計方法可有五大部分：(一)次數分配，(二)集中趨勢的量數，(三)離趨勢或差異度的量數，(四)相關的量數，(五)可靠度量數。茲分別討論之。

一 次數分配

表十七第一部分代表五十四個學生在一個智力測驗上所得的分數。現在要將這些雜亂無章的分數加以整理，第一種工作當然是將這些分數歸入相當的組別。分組的方法可有四個步驟：

(一) 決定全距。最大量數與最小量數間的距離就是全距。表十七所載五十四個分數中，最大者為 201，最小者為 128，201 減去 128，得 73，這就是全距。

(二) 決定組數與組距。五十四個單獨的分數計算起來，很費時間；為簡便起見可以把這些分數歸入少數的組別。組數與組距有連帶的關係，所以應同時決定，因為組數少則組距大，組數多則組距小。決定組數沒有一

定的標準，大概不應少於十組，不應多於二十組；組數太多則計算仍極麻煩，有失歸組之本意；組數太少則因歸組而發生的錯誤太大，亦非所宜。

全距與大約的組數決定後，就可以決定組距的大小。前例中全距為 1.0 ，假設我們決定大約的組數為 10 ，則以 $1.0 \div 10$ 得 $.5$ ，就是組距。

(三) 決定組別的限度 分配內的組別當然應該包括最大的與最小的分數在內。前例中最小的分數為 1.26 ，所以分配中第一組可規定為自 1.26 起到 1.31 止，第二組為 1.31 起到 1.36 ，以下類推。這樣本無不可，但是為計算的便利起見，組別的限度可規定如下：第一組 $1.25—1.30$ ，第二組 $1.30—1.35$ ，以下類推。這裏有一點應注意者，就是 $1.25—1.30$ 寫得詳細一點，應為 $1.25—1.29.9999\dots$ ； $1.30—1.35$ 實在是 $1.30—1.34.9999\dots$ 。所以分數 1.30 應歸入 $1.30—1.35$ 組內，不應歸入 $1.25—1.30$ 組內。

(四) 將各分數歸入相當組內 表十七第二部分表示歸組的方法。歸組有一個基本假設，就是歸組以後，一組中的各分數都與該組中點相等。譬如表十七內有兩個分數在 $1.25—1.30$ 組內，我們假設這兩個分數的價值都等於該組的中點， $1.27.5$ 。求組距的中點的方法如下：

$$\text{組距中點} = \text{組距} \times \frac{\text{下限度} + (\text{上限度} - \text{下限度})}{2}$$

這個假設雖然是最公允的假設，但未必與事實相符。譬如表十七中 $1.25—1.30$ 組中的兩個分數，我們都假設

爲 127.5，其實這兩個分數原來是 126 和 127，平均爲 126.5，與假設中之 127.5 相差 1.00。這個差別就是因歸組而發生的錯誤。大致而論，分組分得愈細密——就是組數多而組距小——則歸組的錯誤愈小；分組分得愈粗魯——就是組數少而組距大——則歸組的錯誤愈大。我們在前面說，分組的組數不應太少，就是這個緣故。以上敘述將各個分數組織成次數分配的方法。

表十七 五十四個學生在一個智力測驗上的分數

「原來的分數（未歸組）」

185 174 127 183 168 126 177 154 157 189 201 158 160 179 181 155 137 177 164 198 188
 197 151 188 188 169 195 165 185 188 195 176 185 185 179 146 182 153 158 109 176 188
 185 155 178 151 144 191 170 157 172 176 164 191

最大分數 = 201

最小分數 = 126

II. 次數分配表（材料同前）

分數	劃記	次數
200—205	—	1
195—200		4

第三編 測驗的實施

變 異 測 驗

190-195		2
185-190		10
180-185		3
175-180		8
170-175		3
165-170		3
160-165		4
155-160		6
150-155		4
145-150		1
140-145		1
135-140		2
130-135		0
125-130		2
N = 54		

二 集中趨勢的數量

將各個分數已經列成次數分配表以後，下面一步工作普通就是求一個集中趨勢的量數。集中趨勢的量數有兩層價值：第一，我們常常要拿一個簡單的數目去形容，代表全部的分配，集中趨勢的量數，就可以供給這種需要。第二，我們常常要比較兩班或兩組的分數，最簡單的方法，就是比較他們的集中趨勢的量數。通用的集中趨勢的量數共有三種：（一）衆數，（二）中數，（三）算術平均數，依次申述如下。

（一）衆數 衆數的意義是極簡單的。在許多量數中，其次數之最多者，謂之衆數。譬如十一個學生的分數為10, 11, 11, 12, 12, 13, 13, 13, 14, 14, 15，那麼13就是衆數，因為牠發現的次數最多。在已歸組的分配中，次數最多的那一組的中點，就是衆數。例如表十七中的分配，以185—190組的次數最多，這一組的中點，187.5就是這分配的衆數。

衆數的功用在大約指示集中趨勢之所在。衆數不是一個可靠的，固定的量數；在同一分配中其價值常因分組之不同而變易。中數與算術平均數都比較衆數為可靠。

（二）中數 中數即一個分配的中點，在這一點的上下，各有量數的一半。譬如一百個兒童在一個測驗上所得分數的分配的中點為65分，就表示此一百個兒童中，有五十個兒童的分數超過65分，有五十個兒童的分數不及65分。茲將求中數的方法列下：

1. 量數未歸組而數目為奇數者 設有60, 65, 65, 70, 75, 75, 75, 80, 80, 90, 95十一個分數。其中數即為最

中間的量數，即為從任何一端數起的第六個量數，就是75。求中數的公式如下：

$$\text{Median 之位置} = \frac{N+1}{2} \dots\dots\dots (1a)$$

公式中 $\text{Median} = \text{Median 中數}$

$N = \text{次數之總和}$

2. 量數未歸組而數目為偶數者 設有75, 75, 76, 77, 78, 79, 83, 85, 87, 87, 88, 89 十二個量數，因為數目為偶數，所以沒有最中間的量數。最公允的辦法就是以最居中二數之平均為中數。這十二個數目中最居中的二數為79和83，其平均為81，這就是中數。

應用公式1去求上列十二個量數的中數：中數的位置為 $\frac{N+1}{2} = \frac{12+1}{2} = 6.5$ 從任何一端數起，第6.5個量數，其價值恰在79至83之中間，即81。

3. 量數已歸組者 表十八表示求歸組的次數分配的中數的方法。其步驟如下：

- (1) 求 $\frac{N}{2}$ 以2除24得12，所以中數應為分配上的一點，在其上下各有12個量數。
- (2) 從分配之下端，將次數向上遞加（從上端向下遞加亦可），直至含有中數之前一組為止。從125—130組起，到170—175組止，其間共次數26。
- (3) 從 $\frac{N}{2}$ （即27）減去26，得1。我們所需要數足的次數為27，到170—175組為止，已有次數26，所以還

須從上面一組即175—180組中補數一個量數就夠了。

(4) 以中數所在組的次數除需要補足的次數，而以組距乘之。在175—180組內共有8個量數，而我們祇須數一個量數已足；假設一組中各量數是均勻地分配的，那麼這一個量數在175—180組中所占的距離應等於該組組距的 $\frac{1}{8}$ 。該組組距為5，所以實占的距離為 $5 \times \frac{1}{8}$ 或.625。

(5) 以.625加於含有中數組的下限度175上，得175.625就是這分配的中數。

表十八 表示中數的求法（材料同表十七）

分	數	次	數
200—205		1	
195—200		4	
190—195		2	
185—190		10	
180—185		3	
175—180		8	
170—175		3	
165—170		3	

<p>從分配中量數小的一端數起：</p> $N = \frac{54}{2} = 27$ $27 = 2+2+1+1+4+6+4+3+3+(175-180\text{組內之}8\text{分之}1)$ $\text{Mdn} = 175 + \frac{1}{8} \times 5$ $= 175.625$
--

160—165	4	<p>從分配中量數大的一端數起：</p> $N = \frac{54}{2} = 27$ $27 = 1 + 4 + 2 + 1 + 3 + (175 - 1$ <p>80組內之8分之7)</p> $Md_n = 180 - \frac{7}{8} \times 5$ $= 175.625$
155—160	6	
150—155	4	
145—150	1	
140—145	1	
135—140	2	
130—135	0	
125—130	2	
N = 54		

求中數的時候從分配中量數小的一端數起或從量數大的一端數起都可以，用這兩種算法所得的答案應該相同。上面所述的求法是從分配中量數小的一端數起的。如從量數大的一端數起，則算法如下。

$$(1) \frac{N}{2} = \frac{54}{2} = 27$$

(2) 從分配的上端，將次數向下遞加，直至含有中數之前一組為止。從200—205組起至180—185組止，共
次數²⁰

$$(3) \text{從 } \frac{N}{2} \text{ (即 } 27 \text{) 減去 } 20 \text{ 得 } 7$$

(4) 以 $175 - 180$ 組之次數 8 除 7 ，而以組距 5 乘之。 $7 \times 5 = 4.375$

(5) 從含有中數組的上限度 180 減去 4.375 得 175.625 就是這分配的中數。

以上法次，以公式表示之，如下：

從分配中量數小的一端數起，用下列公式：

$$\text{Mdn} = l. l. + \frac{\frac{N}{2} - F}{f} \dots \dots \dots i \dots \dots \dots (1b)$$

公式內：

$l. l.$ (Lower limit) = 中數所在組的下限度

N = 次數的總和

F = 中數所在組以下各組次數之和

f = 中數所在組的次數

i (step interval) = 組距

若以表十八之材料代入公式，則

$$\text{Mdn 之價值} = 175 + \frac{\frac{54}{2} - 26}{8} \times 5$$

$$= 175 + \frac{1}{8} \times 5$$

$$= 175.625$$

從分配中量數大的一端數起,用以下公式:

$$Mdn = u.l. - \frac{\frac{N}{2} - \Sigma f_i}{f} \dots\dots\dots (10)$$

公式內:

u. l. (upper limit) = 中數所在組的上限度

f = 中數所在組以上各組次數之和

若以表十八之材料代入公式,則

$$Mdn \text{ 之價值} = 180 - \frac{\frac{N}{2} - 20}{8} \times 5$$

$$= 180 - \frac{7}{8} \times 5$$

$$= 175.625$$

(三) 算術平均數 以次數除量數的總和，得算術平均數。這就是人人共曉之平均數。茲將求算術平均數的方法列下：

1. 量數未歸組者 求未歸組的量數的算術平均數時，其公式如下：

$$M = \frac{\sum X}{N} \dots\dots\dots (2a)$$

在此公式內：

M = 算術平均數

N = 次數的總和

X = 量數

2. 和號 (讀作 summation)

2. 量數已歸組者 求已歸組的量數的算術平均數時，其公式如下：

$$M = \frac{\sum fX}{N} \dots\dots\dots (2b)$$

在此公式內：

M, N, f 之意義同前

X = 各組的中點

『二各組的次數

其算法詳表十九。

表十九 表示算術平均數的求法

組	距	組距中點	次 數	次數×組距中點
210—205		202.5	1	202.50
195—200		197.5	4	790.00
190—195		192.5	2	385.00
185—190		187.5	10	1875.00
180—185		182.5	3	547.50
175—180		177.5	8	1420.00
170—175		172.5	3	517.50
165—170		167.5	3	502.50
160—165		162.5	4	650.00
155—160		157.5	6	945.00
50—155		152.5	4	610.00
145—150		147.5	1	147.50

140—145	142.5	1	142.50
135—140	137.5	2	275.00
130—135	132.5	0
125—130	127.5	2	255.00
		N=54	9265.00

$$M = \frac{\sum fX^2}{N} = \frac{9265}{54} = 171.57$$

3. 簡捷法 照上述方法，數目很大，計算費時，既不經濟而且又易於錯誤，為便利起見，宜用簡捷法，其公式如

$$M = A + \frac{1}{N} (\sum fX^1) i \dots\dots\dots (20)$$

在此公式中

A = 假設平均數（以任何一組的中點為假設平均數）

X¹ = 各組中點與假設平均數的差（以組距為單位表示之）

i = 組距

$\frac{1}{N} (\sum fX^1) i$ = 校正數

用簡捷法以求算術平均數時，我們先假設一個平均數，然後再校正之。其法次如下：（參看表二十。）

(1) 以任何一組的中點為假設平均數。為計算的便利起見，最好以次數最多或在分配的最中間的一組的中點為假設平均數。現在假定以 $165 - 170$ 組的中點，即 167.5 為假設平均數。

(2) 求各組中點與假設平均數之差，此差別以組距為單位表示之。如 $170 - 175$ 組的中點 172.5 與假設平均數 167.5 之差為 $+5$ ，以組距為單位表示之為 $+1$ ，即正一組； $163 - 165$ 組的中點 163.5 與假設平均數 167.5 之差為 -5 ，以組距為單位表示之為 -1 ，即負一組。餘類推。

(3) 以差數乘次數。如 $170 - 175$ 組的差數為 $+1$ ，次數為 3 ，以差數乘次數得 $+3$ 。

(4) 求「差數乘次數」一行的總和。最簡便的方法是將負號的「差數乘次數」及正號的「差數乘次數」分別相加，然後相抵消。表二十中正號的「差數乘次數」的和為 109 ，負號的「差數乘次數」的和為 65 ，正負相消，尚餘 $+44$ 。

(5) 以次數的總和 55 除「差數乘次數」一行的總和 $+44$ ，再以組距乘之，得 4.07 ，即校正數。

(6) 將校正數 4.07 加於假設平均數 167.5 之上，得 171.57 ，即真正算術平均數。

表二十 表示求算術平均數的簡捷法

組別	次數 f	差數 X'	次數乘差數的簡捷法
200—205	1	7	7
195—200	4	6	24
190—195	2	5	10
185—190	10	4	40
180—185	3	3	9
175—180	8	2	16
170—175	3	1	3(+109)
165—170	3	0	
160—165	4	-1	-4
155—160	6	-2	-12
150—155	4	-3	-12
145—150	1	-4	-4
140—145	1	-5	-5
135—140	2	-6	-12
130—135	0	-7	0
125—130	2	-8	-16(-65)
$N=54$			$\sum f X' = 44$

$$A = 167.50$$

$$Q = \frac{44}{54} \times 5 = 4.07$$

$$M = 167.50 + 4.07 = 171.57$$

三 離中趨勢或差異度的量數

集中趨勢的度量，其方法已如上述。現在我們進一步來討論度量分配中分數的差異度的方法。離中趨勢或差異度的量數的功用，我們可以舉一個簡單例子來說明。用一個智力測驗測量甲乙兩班，甲班的平均智力年齡為八歲四個月，乙班的平均智力年齡為八歲五個月，所以從平均數看來，甲乙兩班的智力程度，並無差別可言。假設我們再考查甲乙兩班中各生的分數，發現甲班中智力年齡最低者為六歲六個月，最高者為十歲六個月；乙班中智力年齡最低者為七歲八個月，最高者為九歲二個月。這些事實表示乙班中學生的程度比較甲班整齊多了。換言之，甲班學生程度的差異度較大，乙班學生程度的差異度較小。所以從差異度看來，甲乙兩班的程度實有重要的差別。估計差異度或離中趨勢的量數共有四種，茲分別說明如下：

(一) 全距 分配中最大量數與最小量數間的距離或差別，就是全距。全距這個量數完全被分配中兩個絕端的分數所決定。我們知道分配中兩絕端的分數是不可靠的，所以全距亦不是一個可靠的量數。

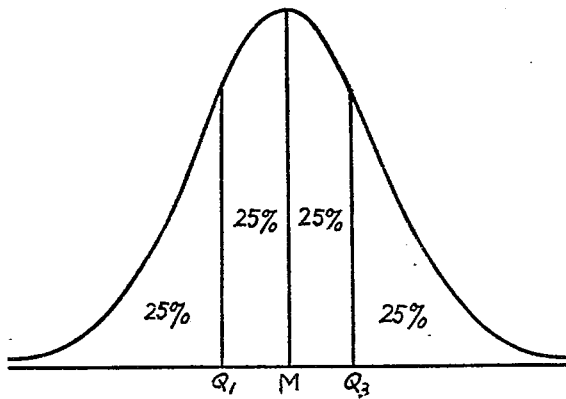
(二)二十五分差 二十五分差等於分配上百分之七十五點與百分之二十五點間的距離的一半。百分之二十五點或稱下二十五分點(Q_1)即分配上的一點，在這一點之下有量數的百分之二十五或四分之一，在這一點之上有量數的百分之七十五或四分之三。百分之七十五點或稱上二十五分點(Q_3)即分配上的一點，在這一點之下有量數的百分之七十五或四分之三，在這一點之上有量數的百分之二十五或四分之一。所以二十五分點與上二十五分點之間有量數的百分之五十或二分之一，如圖六。二十五分差的公式如下：

$$Q = \frac{Q_3 - Q_1}{2} \dots \dots \dots (3)$$

求二十五分差，必須先求上下二十五分點。求上下二十五分點的方法與求中數（即百分之五十點）的方法相似。其法次見表二十一。

表二十一 表示求 Q_1 、 Q_3 的方法

第三編 測驗的實施



圖六 表示 Q_1 和 Q_3 的意義

組 距	中 點	次 數	差 數	次數乘差數
200-205	202.50	1	26.875	26.875
195-200	197.50	4	21.875	87.500
190-195	192.50	2	16.875	33.750
185-190	187.50	10	11.875	118.750
180-185	182.50	3	6.875	20.625
175-180	177.50	8	1.875	15.000
170-175	172.50	3	-3.125	-9.375
165-170	167.50	3	-8.125	-24.375
160-165	162.50	4	-13.125	-52.500
155-160	157.50	6	-18.125	-108.750
150-155	152.50	4	-23.125	-92.500
145-150	147.50	1	-28.125	-28.125
140-145	142.50	1	-33.125	-33.125
135-140	137.40	2	-38.125	-76.250
130-135	132.50	0	-43.125	
125-130	127.50	2	-48.125	-96.250
		N=54		823.750

中數 = 175.625 (從表十八)

$$\frac{N}{4} = 13.5$$

$$Q_1 = 155 + \frac{3.5}{6} \times 5 = 157.92$$

$$\frac{3N}{4} = 40.5$$
$$Q_3 = 185 + \frac{3.5}{10} \times 5 = 186.75$$

$$Q = \frac{Q_2 - Q_1}{2} = \frac{186.75 - 157.92}{2} = 14.42$$

$$AD = \frac{24 \times 11}{N} = \frac{828 \cdot 75}{54} = 15.25$$

以公式表示之如下:

從分配的下端數起,求 Q_1 用以下公式:

$$Q_1 = 1.1 + \frac{\frac{N}{2} - F_1}{f} \cdot i \dots\dots\dots (4a)$$

公式:

1. $L_1 = \text{下} = \text{十五分點所在組的下限度}$

$F = \text{下} = \text{十五分點所在組以下各組次數之和}$

$i = \text{下} = \text{十五分點所在組的次數}$

第十三章 測驗的處理

從分配的上端數起，求 Q_0 ，用以下公式：

$$Q_1 = n \cdot 1 - \frac{\frac{3N}{4} - F^1}{f} \dots\dots\dots (4b)$$

公式內：

$n \cdot 1$ = 下 = 十五分點所在組的上限度

F^1 = 下 = 十五分點所在組以上各組次數之和

從分配的下端數起，求 Q_0 ，用以下公式：

$$Q_0 = 1 \cdot 1 + \frac{\frac{3N}{4} - F}{f} \dots\dots\dots (5a)$$

從分配的上端數起，求 Q_0 ，用以下公式：

$$Q_0 = n \cdot 1 - \frac{N - F^1}{f} \dots\dots\dots (5b)$$

(三) 平均差 平均差即各量數與中數之差的平均，求差數之總和時，不計符號的正負。其公式如下：

$$A.D. = \frac{\sum |X|}{N} \dots\dots\dots (6a)$$

公式內：

A D = 平均差 (Average Deviation的縮寫)

X = 各異數與中數的差別

∑ = 代表“不計符號的正負”

設有6, 8, 10, 12, 14五個分數，其中數為10。這五個分數與中數的差為-4, -2, 0, +2, +4。不計正負符號將這五個差數加起來，然後以次數5除之，得2.4，即平均差。

若分數已歸組而列成次數分配，則平均差的求法與前稍有不同，方法詳見表二十一其公式如下：

$$A D = \frac{\sum fX_i}{N} \dots\dots\dots (5b)$$

公式內：

X = 各組中點與中數的差別

f = 各組的次數

以上方法計算極繁且易於錯誤。為便利起見，可用簡捷法，其公式如下：

$$A D = \left[\frac{\sum fX_i + c(N_b - N_a)}{N} \right] \dots\dots\dots (5c)$$

在此公式內：

X' = 量數與假設中數的差數，差數以組距為單位表示之。

$$= \text{校正數, 以組距為單位表示之}$$

$$= \frac{\text{(真正中數)} - \text{(假設中數)}}{\text{組距}}$$

N_a = 在真正中數以上的量數的次數

N_b = 在真正中數以下的量數的次數

表二十二 表示用簡捷法求平均數的方法

組	距	次	數	差	數	次	數	\times	差	數
200-205		1		5				5		
195-200		4		4				16		
190-195		2		3				20		
185-190		10	$N_a = 28$	2				3(+50)		
180-185		3		1						
175-189		8		0				-3		
170-175		3		-1				-6		
165-170		3		-2				-12		

160-165	4	-3	-12
155-160	6	-4	-24
150-155	4	-5	-20
145-150	1	-6	-6
140-145	1	-7	-7
135-140	2	-8	-16
130-135	0	-9	0
125-130	2	-10	-20 (-114)
	$N = 54$		$\sum 1f \cdot Xi = 164$

真正中數 = 175.625

$N_a = 28$

假設中數 = 177.50

$N_b = 26$

$$\text{校正數} = \frac{175.625 - 177.50}{5} = -.375 \quad \sum 1f \cdot Xi = 164$$

$$\therefore A D = \frac{164 + [-.375(26 - 28)]}{54} \times 5$$

$$= 15.25$$

用簡捷法求平均差的步驟如下 (參看表二十二)

1. 求真正中數。真正中數為 175.625

2. 假設中數。以真正中數所在組的中點為假設中數，得 177.50

3. 求校正數 (C)。以組距除真正中數與假設中數的差別，得校正數， $\frac{175.625 - 177.50}{5} = -1.375$

4. 求在真正中數以上的量數的次數 (N_a) $1 + 4 + 9 + 10 + 3 + 8 = 28$ 。在 175—180 組內八個量數亦歸入 N_a ，因為 175.180 組的中點為 177.50，較真正中數 175.625 為大。

5. 求在真正中數以下的量數的次數 (N_b)。較簡單的方法是將 N_a 從 N 減去，即得 N_b 。54 - 28 = 26

6. 求各組的差數 (K)。以各組中數減去假設中數，再以組距除之，即得。

7. 以各組差數乘其次數，求 (KK_1)。

8. 不計正負符號求「次數乘差數」的總和 ($\sum |KK_1|$) 得 164

9. 將以上各數代入公式 (80)

在常態分配中，從中數起上下各一平均差包含最中間之量數之百分之五十七·五。所以平均差較二十五分差為稍大。

(四) 均方差或標準差 標準差為差異度量數中之最可靠者。標準差與平均差有兩點不同。第一，計算平均差時，不計差數符號的正負；計算標準差時，則將各差數自乘，以避免正負符號的困難。第二，求平均差時，各差數

都從中數算起；求標準差時，各差數都從算術平均數算起。通常代表標準差的符號為希臘字母 σ （讀作 sigma）或英文字母 SD （Standard Deviation 之簡寫）。

標準差之定義為「差數之方之平均數之平方根」。試再以 6, 8, 10, 12, 14 五個分數為例。其平均數為 10。這五個分數與平均數的差為 -4, -2, 0, 2, 4。將差數自乘，得 16, 4, 0, 4, 16。其總和為 40，再以次數 5 除之，得 8。即差數之方之平均數。再求其平方根，得 2.828，就是這五個分數的標準差。標準差的公式如下：

$$\sigma = \sqrt{\frac{\sum fX^2}{N}} \dots\dots\dots (7a)$$

如量數已組成次數分配，則標準差的求法稍有不同。其方法詳見表二十三。公式如下：

$$\sigma = \sqrt{\frac{\sum fX^2}{N}} \dots\dots\dots (7b)$$

在此公式中：

- X = 各組中點與算術平均數之差
- f = 各組的次數

表二十三 表示標準差的求法

組	距	中點	次數 (f)	差 (X)	次數 × 差數 (fX)	次數 × 差數方 (fX ²)
200 - 205		202.50	1	30.93	30.93	956.66

總 帳 簿

195-200	197.50	4	25.93	103.72	2689.45
190-195	192.50	2	20.93	41.68	876.13
185-190	187.50	10	15.93	159.30	2687.65
180-185	182.50	3	10.93	32.79	358.39
175-180	177.50	8	5.93	47.44	281.32
170-175	172.50	3	93	2.79	2.79
165-170	167.50	3	-4.07	-12.21	49.69
160-165	162.50	4	-9.07	-33.28	329.06
155-160	157.50	6	-14.07	-84.43	1187.79
150-155	152.50	4	-19.07	-76.28	1454.66
145-150	147.50	1	-24.07	-24.07	679.36
140-145	142.50	1	-29.07	-29.07	845.06
135-140	137.50	2	-34.07	-68.14	2321.53
130-135	132.50	0	-39.07	
125-130	127.50	2	-44.07	-88.14	3834.33
		N=54			18353.88

110頁

$$6 = \sqrt{\frac{\sum P X^2}{N}} = \sqrt{\frac{18353.88}{54}} = \sqrt{339.887} = 18.44$$

表二十三所列標準差的算法所包含的數目很大而且帶有小數，計算時不甚便利。欲節省時間，減少錯誤，用簡捷法。其公式如下：

$$\sigma = \sqrt{\frac{\sum f(X_i)^2}{N} - O^2 \times i} ; o = \frac{\sum fX_i}{N}$$

$$\therefore \sigma = \sqrt{\frac{\sum f(X_i)^2}{N} - \left(\frac{\sum fX_i}{N}\right)^2 \times i \dots \dots \dots (70)}$$

在公式中：

O = 校正數

X_i = 各組中點與假設平均數之差（以組距為單位表示之）

以簡捷法求標準差的法次如下（參看表二十四）

1. 以任何一組——最好以次數最多或在分配的最中間的一組——的中點為假設平均數。現在假定以 16
- 5—170 組的中點，即 167.50 為假設平均數。
2. 求各組中點與假設平均數之差，此差別以組距為單位表示之。
3. 以次數乘差數，求 $\sum fX_i$ 一行。
4. 求「次數乘差數」一行的代數的總和，而以次數總和除之，得校正數。 $O = \frac{44}{34} = .8148$

5. 以「次數乘差數」一行再乘「差數」一行，求 $f(X_i)^2$ 一行。
6. 求「次數乘差數方」一行的總和，得 770。
7. 將以上各數量代入公式 (σ_s) 求標準差，得 18.44。

表二十四 表示求標準差的簡捷法

分 數	次 數	差 數	次 數	差 數	次 數	次 數 × 差 數 方
X	f	X_i	$f \times X_i$	X_i	$f(X_i)^2$	$f(X_i)^2$
205-205	1	7	7			49
195-200	4	6	24			144
190-195	2	5	10			50
185-190	10	4	40			160
180-185	3	3	9			27
175-180	8	2	16			32
170-175	3	1	3(+109)			3
165-170	3	-1	-4			4
160-165	4	-2	-12			24
155-160	6	-3	-18			36
150-155	4	-4	-16			36

145—150	1	-4	-4	16
140—145	1	-5	-5	25
135—140	2	-6	-12	72
130—135	0	-7	0	
125—130	2	-8	-16(-65)	128
	$N=54$		$\sum f X_i = +44$	$\sum f (X_i)^2 = 770$

$$Q = \frac{44}{54} = .8148$$

$$Q^2 = .6639$$

$$\begin{aligned}
 S D &= \sqrt{\frac{\sum f (X_i)^2}{N} - Q^2} \times 1 \\
 &= \sqrt{\frac{770}{54} - .6639} \times 5 \\
 &= 3.687 \times 5 \\
 &= 18.44
 \end{aligned}$$

在常態分配中，從算術平均點起上下各一標準差，包含最中間之量數之百分之六八·二六。所以標準差較大於平均差，平均差又較大於二十五分差。

(五) 差異係數或比較之差數 以上所述四種量數——全距，二十五分差，平均差，標準差——都是表示

一種分配的絕對的差異度的。但是有時要比較一班在兩種測驗上的程度的差異度或兩班在同一種測驗上的程度的差異度時，除非兩種測驗的單位相同或兩班的平均數（或其他集中量數）大致相等，則這些量數就不能應用了。為應這種需要起見，潘爾遜有差異係數的發明。其公式如下：

$$V = \frac{100 \delta}{M} \dots\dots\dots (8a)$$

公式內：

V = 差異係數

δ = 均方差

M = 算術平均數

從上列公式我們可以知道差異係數不過是差異度量數與集中量數的比例。所以下列公式有需要時亦可以通用的：

$$V = \frac{100 Q}{Mdn} \dots\dots\dots (8b)$$

$$V = \frac{100 AD}{Mdn} \dots\dots\dots (8c)$$

問題

1. 乘數何以不可靠?

2. 爲什麼要歸組歸組有什麼利弊?

3. 比較中數與算術平均數的利弊。

4. 差異係數有什麼功用?

5. 全距何以不可靠?

6. 在一個分配中,上下二十五分點之間包含數數之百分之幾?

7. 根據下列分配求甲乙兩班之

(1) 中數

(2) 算術平均數

(3) 二十五分差

(4) 標準差

(5) 平均差

參考書報

第三編 測驗的實施

組 別	甲班	乙班
65-70	1	
60-66	2	
55-60	3	
50-55	2	1
45-50	8	2
40-45	6	2
35-40	4	3
30-35	7	4
25-30	4	5
20-25	5	6
15-20	5	8
10-15		4
5-10		3
	47	38

1. Garrett, H. B. Statistics in Psychology and Education. Longmans. (中文譯本有劉遵敏：實用統計學，南京書店出版。)

2. 朱君毅：教育統計學，(商務)

第十六章 測驗結果的整理(下)

四 相關量數

以上所述，都是關於表示一個人或一團體在某一種測驗上的成績的統計方法。但是有時候我們還需考查一個人或一團體在兩種測驗上的成績的關係。例如，智力較高的兒童，學業成績是否較優？國文程度較深者，英文程度是否亦較深？這些問題的解答都有賴於相關法。

相關有高有低，有正有負。我們知道一個圓形的周圍常等於其直徑的 3.1416 倍。每次增加或減少一個圓形的直徑，則其周圍亦增加或減少了同數量的 3.1416 倍。換言之，一個圓形的周圍與直徑的關係是固定的，確定的。所以我們說周圍與直徑之間有完全的相關。又如一班兒童在兩種測驗上的成績的等第完全相同，即陳兒在第一種測驗上成績列第一，在第二種測驗上成績亦列第一；張兒在第一種測驗上成績列第二，在第二種測驗上成績亦列第二，等等，則這兩種測驗之間又呈完全相關的關係。

用一種智力測驗，一種圖畫測驗測量一百個兒童，然後將這一百個兒童依其在智力測驗上所得分數的高低分為上中下三組，再計算每組在圖畫測驗上的平均分數。假設這三組的平均圖畫分數大致相等，則智力與圖

畫能力之間顯然無甚關係。換言之，智力較高者，圖畫能力未必較優；智力較低者，圖畫能力未必較差，二者之間的相關度為零。

統計學中表示相關度的量數為相關係數，以英文字母 r 代表之。 $r = 1.00$ 代表完全相關； $r = 0.00$ 代表零度相關。在完全相關與零度相關之間當然還有各級的相關度如：.30, .60, .80, 等等。

相關又可為負的。假如前例中智力上中下三組兒童的圖畫分數以上組為最低，中組較高，下組最高，則智力與圖畫能力之間顯然呈負相關的關係。完全負相關以 $r = -1.00$ 表示之。所以相關係數的價值是從 -1.00 起經過 0 到 $+1.00$ 止。

相關係數的解釋是一個很重要的問題，但因篇幅有限，不能詳論。下述數點在解釋相關係數時必須顧到。第一，相關不一定表示因果的關係。相關係僅僅表示兩種事實關係之有無，關係之方向和關係之深淺。至於因果關係之存在與否，必須考察整個情境，用推理方法以決定之。第二，相關係數的數量依所測量的團體的差異度而變。易所測量的團體的差異度愈大，如其他情形相等，則所得相關愈高；差異度愈小，則相關愈低。所以要解釋一個相關係數，我們必須知道這相關係數是從什麼團體得來的，這團體的差異度如何。第三，測量的錯誤可以增加測量結果的差異度，因此減低相關係數的數量。這種現象，稱為相關減弱，可以用統計學中校正相關減弱的公式以校正之。第四，相關的「高」「低」並無一定的標準可循，要視當時的情境，需要，功用以爲斷。茲錄麥柯爾所定標準

以資參考：

○ 至 ±.4 表示相關之低者；

±.4 至 ±.7 表示相關之切實者；

±.7 至 ±1.00 表示相關之高者。

計算相關度的方法，最通用者有二，詳述於下：

(一) 均方相關法 均方相關係數的公式如下：

$$r = \frac{\sum XY}{\sqrt{6X \sum Y}} \dots\dots\dots (9a)$$

在此公式中：

r = 相關係數

X = 各 X 量數與 X 的平均數的差數

Y = 各 Y 量數與 Y 的平均數的差數

$\sum X$ = X 的均方差

$\sum Y$ = Y 的均方差

公式 (9a) 又可用上式表示：

$$r = \frac{\sum XY}{\sqrt{\sum X^2} \sqrt{\sum Y^2}} \dots\dots\dots (9b)$$

表二十五表示公式(9b)的應用:

表二十五 表示相關係數的求法

學生	在測驗甲 的分數(X)	在測驗乙 的分數(Y)	X	Y	X ²	Y ²	XY
A	50	22	-12	-8.4	144	70.56	100.8
B	53	25	-9	-5.4	81	29.16	48.6
C	56	34	-6	3.6	36	12.96	-21.6
D	58	28	-4	-2.4	16	5.76	9.6
E	60	26	-2	-4.4	4	19.36	8.8
F	61	30	-1	.4	1	.16	.4
G	61	32	-1	1.6	1	2.56	-1.6
H	64	30	2	-.4	4	.16	-.8
I	67	28	5	-2.4	25	5.76	-12.0
J	70	34	8	3.6	64	12.96	28.8
K	71	36	9	5.6	81	31.36	50.4
L	73	40	11	9.6	121	92.16	105.6

平均數	62	30.4		578	282.92	317.0
-----	----	------	--	-----	--------	-------

$$r = \frac{317}{\sqrt{578} \cdot \sqrt{282.92}} = .78$$

用上述方法以求相關係數，常常數目很大，且有小數，計算時甚不便利，且易錯誤。如對偶量數加多，用上述方法尤為不便。欲減輕計算工作，減少錯誤，可用分布圖法。用分布圖法時所用之簡捷公式如下：

$$r = \frac{\sum X'Y' - O_x O_y}{\sqrt{\sum X'Y'^2} - \dots \dots \dots (90)}$$

寫得詳盡些，

$$r = \frac{\frac{\sum X'Y'}{N} - \left(\frac{\sum X'}{N}\right)\left(\frac{\sum Y'}{N}\right)}{\sqrt{\frac{\sum (X')^2}{N} - \left(\frac{\sum X'}{N}\right)^2} \sqrt{\frac{\sum (Y')^2}{N} - \left(\frac{\sum Y'}{N}\right)^2}} \dots \dots \dots (9d)$$

圖七表示公式(9d)的應用。其步驟如后：

變 異 總 和

111 4

圖七 表示相關係數的分布圖求法

f_{xy}	165-160	160-165	165-170	170-175	175-180	180-185	185-190	$f_{y'}$	y'	$f_{y'^2}$	$f(y')^2$	$\sum X^2 y^2$
9												
1	(9)	8	14	7	10	16	24	12	1	3	9	12
6	(6)	2	7	10	16	24	32	16	2	32	64	58
3	(3)	5	7	10	16	24	28	16	2	32	64	58
1	(2)	10	7	10	16	24	28	16	2	32	64	58
2	(1)	7	10	16	24	32	28	16	2	32	64	58
0	(0)	10	16	24	32	28	28	16	2	32	64	58
0	(-1)	7	10	16	24	32	28	16	2	32	64	58
0	(-2)	10	16	24	32	28	28	16	2	32	64	58
0	(-3)	16	24	32	28	28	28	16	2	32	64	58
0	(-4)	24	32	28	28	28	28	16	2	32	64	58
0	(-5)	32	28	28	28	28	28	16	2	32	64	58
0	(-6)	28	28	28	28	28	28	16	2	32	64	58
0	(-7)	28	28	28	28	28	28	16	2	32	64	58
0	(-8)	28	28	28	28	28	28	16	2	32	64	58
0	(-9)	28	28	28	28	28	28	16	2	32	64	58
0	(-10)	28	28	28	28	28	28	16	2	32	64	58
0	(-11)	28	28	28	28	28	28	16	2	32	64	58
0	(-12)	28	28	28	28	28	28	16	2	32	64	58
0	(-13)	28	28	28	28	28	28	16	2	32	64	58
0	(-14)	28	28	28	28	28	28	16	2	32	64	58
0	(-15)	28	28	28	28	28	28	16	2	32	64	58
0	(-16)	28	28	28	28	28	28	16	2	32	64	58
0	(-17)	28	28	28	28	28	28	16	2	32	64	58
0	(-18)	28	28	28	28	28	28	16	2	32	64	58
0	(-19)	28	28	28	28	28	28	16	2	32	64	58
0	(-20)	28	28	28	28	28	28	16	2	32	64	58
0	(-21)	28	28	28	28	28	28	16	2	32	64	58
0	(-22)	28	28	28	28	28	28	16	2	32	64	58
0	(-23)	28	28	28	28	28	28	16	2	32	64	58
0	(-24)	28	28	28	28	28	28	16	2	32	64	58
0	(-25)	28	28	28	28	28	28	16	2	32	64	58
0	(-26)	28	28	28	28	28	28	16	2	32	64	58
0	(-27)	28	28	28	28	28	28	16	2	32	64	58
0	(-28)	28	28	28	28	28	28	16	2	32	64	58
0	(-29)	28	28	28	28	28	28	16	2	32	64	58
0	(-30)	28	28	28	28	28	28	16	2	32	64	58
0	(-31)	28	28	28	28	28	28	16	2	32	64	58
0	(-32)	28	28	28	28	28	28	16	2	32	64	58
0	(-33)	28	28	28	28	28	28	16	2	32	64	58
0	(-34)	28	28	28	28	28	28	16	2	32	64	58
0	(-35)	28	28	28	28	28	28	16	2	32	64	58
0	(-36)	28	28	28	28	28	28	16	2	32	64	58
0	(-37)	28	28	28	28	28	28	16	2	32	64	58
0	(-38)	28	28	28	28	28	28	16	2	32	64	58
0	(-39)	28	28	28	28	28	28	16	2	32	64	58
0	(-40)	28	28	28	28	28	28	16	2	32	64	58
0	(-41)	28	28	28	28	28	28	16	2	32	64	58
0	(-42)	28	28	28	28	28	28	16	2	32	64	58
0	(-43)	28	28	28	28	28	28	16	2	32	64	58
0	(-44)	28	28	28	28	28	28	16	2	32	64	58
0	(-45)	28	28	28	28	28	28	16	2	32	64	58
0	(-46)	28	28	28	28	28	28	16	2	32	64	58
0	(-47)	28	28	28	28	28	28	16	2	32	64	58
0	(-48)	28	28	28	28	28	28	16	2	32	64	58
0	(-49)	28	28	28	28	28	28	16	2	32	64	58
0	(-50)	28	28	28	28	28	28	16	2	32	64	58

$$x^1 - 3 \quad -2 \quad -1 \quad 0 \quad 1 \quad 2 \quad 3 \quad 4 \quad (146)$$

$$f x^2 - 9 \quad -20 \quad -28(-57) \quad 22 \quad 18 \quad 15 \quad 24(79) = 22$$

$$f(x^2)^2 \quad 27 \quad 40 \quad 28 \quad 22 \quad 36 \quad 45 \quad 96 = 294$$

$$C_Y = \frac{2}{120} = .017 \quad C_X = \frac{22}{120} = .183 \quad r = \frac{146}{120} = .017 \times .183$$

$$C^2_Y = .0003 \quad C^2_X = .0334 \quad r = .60$$

$$S_Y = \sqrt{\frac{205}{120} - .0003 \times 5} = S_X = \sqrt{\frac{294}{120} - .0334 \times 5} \quad P.D.T. = \frac{.6745[1 - (.60)^2]}{\sqrt{120}} = .04$$

$$1.31 \times 5 = 6.55 \quad 1.65 \times 5 = 7.75$$

1. 作相關圖，如圖七。以橫坐標自左至右，代表體重，以縱坐標自下而上，代表身長。每一格為一組距。

2. 將各人依其體重與身長歸入圖中相當方格內。設有人體重為 68 公斤（磅），身長為 170 公分（粉），依其身長，此人應歸入相關圖內自上向下數之第三橫行內；依其體重，此人應歸入相關圖內自左向右數之第五縱行內。因此，此人應歸入圖內第三橫行與第五縱行相交處之方格內。

3. 求各橫行與各縱行的次數，分別記載於 T_i 與 X_j 行下。再分別求 T_i 與 X_j 的總和，即次數的總和， T 的總和與的 X 總和應相等。

4. 假設一個 X 的平均數。設以 172.5 即 170—175 組的中點，為身長分配的假設平均數。然後求各身長組中點與假設平均數之差 (Y_1)。此差數以組距為單位表示之。再以次數乘差數，得 Y_1 一行。求 Y_1 行的總和，並以 N 除之得 OY 。再以 Y_1 行乘 Y_1 行，得 $(Y_1)^2$ 一行。將 $(Y_1)^2$ 行的總和以及 OY 代入公式 (70) 求身長分配的均方差 OY^2 。

同樣地求 $M_{FK}, M_{F(X^2)}, \sigma_X$

5. 第四步內所述乃用簡捷法求均方差的手續，前已論及。現在我們須求 $\sum X_i Y_i$ 一行。 $\sum X_i Y_i$ 有正有負，為便利起見，將正負的價值各列一行。

決定各橫行的 $\sum X_i Y_i$ 的手續，可用圖中最上面一橫行為例以說明之。最上面一橫行內祇有一方格內有次數。這方格與體重的假設平均數的差數，即 X_1 為 4，與身長的假設平均數的差數，即 Y_1 為 3。所以這方格的 $X_1 Y_1$ 為 12。將這數目載入該方格的上方右角。這方格內的次數為 1，以 1 乘 12，得 12，即此方格的 $\sum X_i Y_i$ ，將此數載入該方格的下方左角。因為這橫行其他各方格均無次數，故將 12 記入 $\sum X_i Y_i$ 行「+」符號下。

現在我們再計算從上面數第二個橫行的 $\sum X_i Y_i$ ，並從右面第一個方格計算起。第二橫行最右方格的 X_2 為 4， Y_2 為 2，所以 $X_2 Y_2$ 等於 8。將 8 載入該格上方右角。這方格的次數為 3，乘 8 為 24，即該方格的 $\sum X_i Y_i$ ，將 24 載入該方格的下方左角。依此方法，求得第二方格的 $X_2 Y_2$ 為 12， $\sum X_i Y_i$ 為 12，所以 $X_2 Y_2$ 為 24。因這方格的次數為 3，所以該方

格的 $\begin{matrix} X \\ Y \end{matrix}$ 爲 2 乘 6 等於 12 第三方格之 $\begin{matrix} X \\ Y \end{matrix}$ 爲 2 乘 2 等於 4 次數爲 4 所以 $\begin{matrix} X \\ Y \end{matrix}$ 爲 16 第四方格之 $\begin{matrix} X \\ Y \end{matrix}$ 爲 1 乘 2 等於 2 次數爲 3 所以 $\begin{matrix} X \\ Y \end{matrix}$ 爲 6 第五方格之 $\begin{matrix} X \\ Y \end{matrix}$ 爲 0 所以 $\begin{matrix} X \\ Y \end{matrix}$ 當然亦爲 0 第六方格之 $\begin{matrix} X \\ Y \end{matrix}$ 爲 -1 乘 2 所以 $\begin{matrix} X \\ Y \end{matrix}$ 爲 -2 次數爲 1 所以 $\begin{matrix} X \\ Y \end{matrix}$ 仍爲 -2 然後將這橫行內各方格的 $\begin{matrix} X \\ Y \end{matrix}$ 價值正負分別相加得 +58 將這兩個數目載入 $\begin{matrix} X \\ Y \end{matrix}$ 行相當符號下。

用同樣手續求其他各橫行的 $\begin{matrix} X \\ Y \end{matrix}$ 再求 $\begin{matrix} X \\ Y \end{matrix}$ 一行的代數總和得 148

6. 將以上各步所求得的數量代入公式以求相關係數。

(二) 等級相關法 在許多問題中, 尤其在應用心理, 職業心理方面, 能力或性格上的差別祇能以等級表示之, 而不能以分數表示之。根據此種資料以求相關, 則上述的均方相關法就不適用了。有時次數大小, 犯不着用繁複的均方相關法來計算, 亦可用較爲簡單的等級相關法。最通用而較爲可靠的等級相關法爲司墨門 (D. Spearman) 的等級差異法, 其公式如下:

$$P = 1 - \frac{6 \sum D^2}{N(N^2 - 1)} \dots \dots \dots (10)$$

在此公式內:

P (讀作 rho) = 等級相關係數

D = 各人數 X 與 Y 等級的差數

$N = \Sigma X$ 數之總和

等級相關係數的求法見表二十六。表中第一行為售貨員人名，第二行為服務年數，第三行為服務年數的等級。例如C服務年數最久，列第一，O次之，列第二，餘類推。須注意者為B與C應列之等級本為1與2，但因此二人服務年數相同，無法決定何人應列七等，何人應列八等，故採用平均之方法，每人得等級7.5。而服務年數次多之M，則列入等級9。

第四行為各人服務效率的等級，由經理或人事部主任判定之。效率最高者為O，列第一；效率最低者為M，列第十二。第五行為服務年數等級與服務效率等級的差數。第六行為此差數之方。差數方的總和為58。以58代M₂，以12代N，應用公式(10)，得等級相關係數.80。

等級相關係數與均方相關係數是不能比較的，因為兩種方法所根據的基本假說是不同的。根據表二十七可以將D化為r。

表二十六 表示等級相關法的步驟

(1) 售貨員	(2) 服務年數	(3) 服務年數 的等級	(4) 服務效率 的等級	(5) 等級的差 數(D)	(6) 等級的差 數方(D ²)
A	5	7.5	6	1.5	22.5

B	2	11.5	12	.5	.25
C	10	2	1	1.0	1.00
D	8	4	9	5.0	25.00
E	6	6	8	2.0	4.00
F	4	9	5	4.0	16.00
G	12	1	2	1.0	1.00
H	2	11.5	10	1.5	2.25
I	7	5	3	2.0	4.00
J	5	7.5	7	.5	.25
K	9	3	4	1.0	1.00
L	3	10	11	1.0	1.00
N=12				$\sum D$	53.00

$$\rho = 1 - \frac{6 \sum D^2}{N(N^2-1)} = 1 - \frac{6 \times 58}{12(143)} = .80$$

根據表二十七

$$r = .81$$

$$P_{Dr} = \frac{.7063(1-r^2)}{\sqrt{N}} = .07$$

第三編 測驗的實施

表二十七 從 ρ 化爲 r 的應用表

ρ	r	ρ	r	ρ	r	ρ	r
.01	.0105	.26	.2714	.61	.5277	.76	.7750
.02	.0209	.27	.2818	.62	.5378	.77	.7847
.03	.0314	.28	.2922	.63	.5479	.78	.7943
.04	.0419	.29	.3025	.64	.5580	.79	.8039
.05	.0524	.30	.3129	.65	.5680	.80	.8135
.06	.0628	.31	.3232	.66	.5781	.81	.8230
.07	.0733	.32	.3335	.67	.5881	.82	.8325
.08	.0838	.33	.3439	.68	.5981	.83	.8421
.09	.0942	.34	.3542	.69	.6081	.84	.8516
.10	.1047	.35	.3645	.60	.6180	.85	.8610
.11	.1151	.36	.3748	.61	.6280	.86	.8705
.12	.1256	.37	.3850	.62	.6379	.87	.8799
.13	.1360	.38	.3935	.63	.6478	.88	.8893
.14	.1465	.39	.4056	.64	.6577	.89	.8986
.15	.1569	.40	.4158	.65	.6676	.90	.9080

.16	.1674	.41	.4261	.66	.6775	.91	.9173
.17	.1778	.42	.4363	.67	.6873	.92	.9296
.18	.1882	.43	.4465	.68	.6971	.93	.9359
.19	.1986	.44	.4567	.69	.7069	.94	.9451
.20	.2091	.45	.4669	.70	.7167	.95	.9543
.21	.2195	.46	.4771	.71	.7265	.96	.9635
.22	.2299	.47	.4872	.72	.7363	.97	.9727
.23	.2403	.48	.4973	.73	.7460	.98	.9818
.24	.2507	.49	.5075	.74	.7557	.99	.9909
.25	.2611	.50	.5176	.75	.7654	1.00	1.0000

五 量數的可靠度

假設我們要求中國十二歲男子的平均身長(或其他集中量數)及身長的均方差(或其他差異度量數)為得可靠結果起見,最完善的辦法當然是對於個個十二歲男子都加以測量,然後從所得身長的分配求平均數與均方差,這平均數可視為中國十二歲男子身長的真正平均,這均方差可代表中國十二歲男子身長的真正差異度。同樣地,我們去測量中國個個十二歲女子的身長,再求得其真正平均數與真正均方差,既然知道了十二歲

男子與女子的真正平均身長，則男女之間平均身長的真正差別，當然又很容易求得。

但是將一團體中的個人一一加以測量，很不易做到，我們常常祇能測量某團體中的一部分，以一個有限的「樣本」去代表全部的事實。從一個有限的樣本所得的平均數與均方差，因為有限的取樣的原因，常不能與真正平均數或真正均方差相等。換言之，根據有限的樣本所得的平均數與均方差不是完全可靠的。所得平均數或均方差與真正平均數或均方差的差別，專門名詞稱為隨機取樣的錯誤。當然，一量數的隨機取樣的錯誤與其可靠度恰成反比例；錯誤愈大，量數愈不可靠，錯誤愈小，量數愈可靠。在本節內我們將介紹幾個通用的核算因隨機取樣而發生的錯誤的公式。

(一) 平均數的可靠度 表示一量數的可靠度的方法有二，一為標準錯誤或均方錯誤，一為機誤。所謂標準錯誤，即根據各樣本所得量數與真正量數的差別的分配的標準差或均方差。換言之，標準錯誤乃各樣本因隨機取樣而發生的錯誤的均方差。所謂機誤即根據各樣本所得量數與真正量數的差別的分配的二十五分差。(註一) 平均數的均方錯誤的公式如下：

$$\sigma_m = \frac{\sigma(\text{dis})}{\sqrt{N}} \dots \dots \dots (11a)$$

(註一) 在常態分配中，機誤 (P.E.) 等於二十五分差 (S.E.)

在此公式內：

δ_m = 平均數的均方錯誤

$\delta(\text{dis})$ = 全分配的均方差

N = 次數的總和

在常態面積下機誤與均方差有固定的關係，機誤等於 $\cdot 6745$ 均方差，所以平均數的機誤的公式如下：

$$P.E.m = \frac{\cdot 6745 \delta(\text{dis})}{\sqrt{N}} \dots\dots\dots (11b)$$

在此公式內 $P.E.m$ 代表機誤，即 Probable Error 之縮寫。

從上列公式，我們可以發見 σ_B 或 $P.E.m$ 數量的大小，即平均數的可靠度的高低，視兩種因素為斷，即樣本中次數之多少與所得分配的差異度的大小。次數愈多，即 N 的價值愈大，則 σ_B 的價值愈小，即所得平均數愈可靠。所得分配的差異度愈大，即 $\delta(\text{dis})$ 的價值愈大，則 σ_B 的價值亦隨之而增加，即平均數的可靠度愈低。其中理由，熟思之，不難明瞭。

公式 (11a) 與公式 (11b) 的應用與解釋可舉例以明之。在一八八三年，英國人體測量學會求得 68.68 個英國成人男子的平均身長為 67.6 公分，均方差為 6.7 公分。這所得平均數的可靠度如何？這所得結果與英

國成人男子的真正平均身長相差幾何？應用公式 (11a) 平均數的均方錯誤為 $.0277$ 。這結果的解釋如下：真正平均數在所得平均數上下或加減各一均方錯誤之間，即 $67.4614 \pm .0277$ ，即 67.432 與 67.488 之間，的機會為百分之六十八點二六。(註二) 再者，因為在常態分配中，自平均點起上下各三個均方差，差不多包括全體量數在內。(註三) 所以我們可以大致斷定，真正平均數在所得平均數上下或加減各三個均方錯誤之間，即 $67.4614 \pm 3 \times .0277$ ，即 67.377 與 67.543 之間。顯而易見的， σ_m 愈大，則真正平均數所在的可能範圍愈廣，所得平均數當然比較地不可靠。反之， σ_m 愈小，則真正平均數所在的可能範圍愈狹，所得平均數就比較地可靠了。

應用公式 (11b) 平均數的機誤為 $.0187$ 寸。這結果的解釋如下：真正平均數在所得平均數上下或加減一個機誤之間，即 $67.4614 \pm .0187$ ，即 67.44 與 67.48 之間，的機會為百分之五十。(註四) 再者，因為在常態分配中，自平均點起上下各四個機誤，差不多包括全體量數在內。(註五) 所以我們可以大致斷定，真正平均數在所得平均數上下或加減四個機誤之間，即 $67.4614 \pm 4 \times .0187$ ，即 67.39 與 67.53 之間。

(註二) 在常態分配下，由平均點起上下各一均方差，包括全部面積或全體次數之 68.26% 。

(註三) 在常態分配內，由平均點起上下各三個均方差，包括全體量數之 99.73% 。

(註四) 在常態分配內，自平均點起上下各一個機誤，包括全體量數之 50% 。

(註五) 在常態分配內，自平均點起上下各四個機誤，包括全體量數之 99.99% 。

(二) 中數的可靠度 中數的均方錯誤與機誤的公式如下：

$$\sigma(\text{mdn.}) = \frac{5}{4} \cdot \frac{\sigma(\text{dis.})}{\sqrt{2N}} \dots\dots\dots (12a)$$

$$\text{PE}(\text{mdn.}) = \frac{5}{4} \cdot \frac{.6745 \times \sigma(\text{dis.})}{\sqrt{N}} = \frac{.845\sigma(\text{dis.})}{\sqrt{N}} \dots\dots\dots (12b)$$

$$\text{PE}(\text{mdn.}) = \frac{5}{4} \cdot \frac{Q}{\sqrt{N}} \dots\dots\dots (12c)$$

由此可見中數的可靠度較低於平均數。

(三) 均方差的可靠度 均方差的均方錯誤與機誤的公式如下：

$$\sigma_s = \frac{\sigma(\text{dis.})}{\sqrt{2N}} \dots\dots\dots (13a)$$

$$\text{PE}\sigma_s = \frac{.6745\sigma(\text{dis.})}{\sqrt{2N}} \dots\dots\dots (13b)$$

(四) 二十五分差的可靠度 二十五分差的均方錯誤的公式如下：

$$\sigma_q = \frac{1.11 \times \sigma(\text{dis.})}{\sqrt{2N}} \dots\dots\dots (14a)$$

或
$$\delta Q = \frac{1.65 \times Q}{\sqrt{2N}} \dots\dots\dots (14b)$$

由此可見二十五分差的可靠度較低於均方差。

(五) 差數的可靠度 假設我們比較十歲男生與十歲女生在某一個智力測驗上的成績。普通的辦法是去測量許多十歲男生和許多十歲女生，並求男生平均與女生平均的差別。假如男生平均超過女生，我們就斷定男生的平均成績較女生的平均成績為優。但是在下肯定的斷語以前，我們應該問問這差數究竟怎樣可靠，男生的真正平均超過女生的真正平均的機會有多少。計算差數的可靠度的一種公式如下：

$$\delta(\text{diff}) = \sqrt{\delta^2(m_1) + \delta^2(m_2) \dots\dots\dots (15a)}$$

在此公式內：

δm_1 = 第一個所得平均數的均方錯誤

δm_2 = 第二個所得平均數的均方錯誤

$\delta(\text{diff})$ = 差數的均方錯誤

所以欲求兩個平均數的差數的可靠度，必須先求平均數自身的可靠度。

公式 (15a) 的應用與解釋，可舉例以說明之。用測驗測量兩校的兒童，得結果如下：

校 別	學生人數	平均分數	\bar{d} (dis)
甲 校	308	13.88	2.43
乙 校	325	13.69	2.23

甲乙兩校平均成績的差別爲·19。這差數是否可靠？甲校的平均成績是否真正在乙校之上？在下次測量時，乙校平均成績超過甲校的機會有多少？要回答這些問題，我們必須求甲乙兩校平均成績的差數的可靠度。

根據公式 (11a) 甲乙兩校平均成績的均方錯誤如下：

$$\sigma_{m_1} = \frac{2.43}{\sqrt{308}} = .1385$$

$$\sigma_{m_2} = \frac{2.23}{\sqrt{325}} = .1237$$

將這些數量代入公式 (15a) 得

$$\sigma_{diff} = \sqrt{(.1385)^2 + (.1237)^2} = .1857$$

所得差數爲·19，而這差數的均方錯誤爲·1857。

差數的均方錯誤的解釋與其他量數的均方錯誤的解釋相同。我們可以說真正差數在所得差數 (·19) 上下或加減各一個均方錯誤 (·1857) 之間的機會爲百分之六十八。我們可以大致斷定真正差數在所得差數

上下各三個均方錯誤之間，即 $.19 + 3 \times .1857$ 即 $-.37$ 與 $+ .75$ 之間。

我們可以注意到這真正差數所在範圍的下限度為一負數，表示乙校的平均成績亦有若干機會超過甲校的平均成績。所以強然所得平均甲校超過乙校，但是我們還不能十分肯定甲校的平均成績真正在乙校之上。

照通常的標準，差數至少須三倍於差數的均方錯誤，才算完全可靠。上例中甲乙兩校平均的差數 (D) 為 .1，差數的均方錯誤 (σ^2 diff) 為 $\frac{.1857}{D}$ 為 1.02，不到 3，所以這差數不是完全可靠的，我們尙不能十分肯定甲校的真正平均成績在乙校之上。

差數的可靠度又可用機誤來表示，其公式如下：

$$PE_{(diff)} = \sqrt{PE^2(m_1) + PE^2(m_2) + \dots} \quad (15b)$$

照通常的標準，差數至少須四倍於差數的機誤，才算完全可靠。

兩個中數的差數的可靠度，可用下列公式核算之：

$$\sigma^2(diff) = \sqrt{\sigma^2(mdn_1) + \sigma^2(mdn_2) + \dots} \quad (16a)$$

$$PE_{(diff)} = \sqrt{PE^2(mdn_1) + PE^2(mdn_2) + \dots} \quad (16b)$$

(六) 相關係數的可靠度 均方關係數的均方錯誤與機誤的公式如下：

$$\delta r = \frac{(1-r^2)}{\sqrt{N}} \dots\dots\dots (17a)$$

$$PEr = \frac{.6745 \times (1-r^2)}{\sqrt{N}} \dots\dots\dots (17b)$$

從等級相關係數r所得r的可靠度稍低於均方相關係數,其公式如下:

$$PEr = \frac{.7033 \times (1-r^2)}{\sqrt{N}} \dots\dots\dots (18a)$$

在應用以上各量數的可靠度的公式時,我們必須明瞭各量數的均方錯誤或機誤所測量的僅僅為因隨機取樣而引起的錯誤。至於因其他原因而引起的錯誤,均非以上各公式所能測量。

(七) 一個分數的可靠度 一個個人在某一種測驗上所得分數不一定與其「真正分數」相等,換言之,不一定完全可靠。假設恆性錯誤已被排除,則所得分數的可靠度與測驗本身的可靠度成正比例。一個分數的可靠度可用下列公式估計之:

$$\delta (m) = \delta_1 \sqrt{1-r_{12}} \dots\dots\dots (19a)$$

在此公式內

$\delta (m)$ = 一個分數的均方錯誤

σ_1 = 分數分配的均方差

r_{12} = 測驗的可靠係數或自身相關

公式 (19) 的應用與解釋可舉例以說明之。假設一百個兒童在某一種測驗上的平均分數為 150，均方差為 15，測驗的自身相關為 .90。應用公式：

$$\hat{\sigma}_1(m) = 15 \sqrt{1 - .90} = 4.74$$

根據 $\hat{\sigma}_1(m)$ 我們可以說在此一百個人中任何一人的真正分數在所得分數上下或加減一個 $\hat{\sigma}_1(m)$ 之間的機會為百分之六十八。一人的真正分數大概一定在所得分數上下或加減 3×4.74 之間。假定某人得分數 175，則此人的真正分數在 175 ± 4.74 即 170.47 與 179.26 之間的機會為 68%。此人的真正分數大概一定在 175 ± 3 × 4.74 即 160.18 與 189.22 之間。

一個分數的可靠度又可由機誤表示之：

$$P.E.(m) = .6745\hat{\sigma}_1 \sqrt{1 - r_{12}} \dots\dots\dots (19b)$$

問題

1. 就下列各項各舉一具體的例子：

(1) 有相關而無因果關係者。

(2) 有相關而又有因果關係者。

(3) 有相關而因果關係之有無難以確定者。

2. 何謂隨機取樣的錯誤?

3. 所得平均數之均方錯誤 (O.E.) 愈小, 則所得平均數愈可靠, 何故?

4. 均方差的均方錯誤 (O.E.) 如何解釋?

5. 根據下列事實, 求 X 與 Y 的均方相關係數 (r)。

學生	X 分數	Y 分數
A	76	17
B	74	15
C	82	14
D	63	12
E	74	18
F	91	19
G	86	20
H	82	23
I	79	20
J	72	14

6. 根據下表事實, 簡捷法求 r 。

第三編 測驗的實施

	25-30	30-35	35-40	40-45	45-50	f _Y
125-130		1	4	2	6	13
120-125	2	2	6	8	3	21
115-120	3	5	5	1		14
110-115	.1	2	1			4
f _X	6	10	16	11	9	52

Y 國文分數

2. 用線段圖顯示 X 與 Y 的關係。

學生	Y 分數	y 分數
A	30	80
B	32	85
C	24	86
D	20	86
E	24	82
F	35	87
G	37	88
H	24	80
I	33	90
J	39	92

8. 假設所得平均數 = 30.2; $\sigma = 0.00$; $N = 100$;

試求 \bar{Q}_m 並解釋之。

9. 用一種測驗測量兩班, 結果如下:

	M	\bar{Q}_m
甲班	65	1.2
乙班	62	1.4

試求兩班平均的差別的均方錯誤, 並解釋之。

參考書報

同前章。

第十七章 測驗結果的解釋：年齡量表中各量數的意義

標準測驗有一個應具的條件，就是測驗中各分數的意義須確定。一人在一個測驗上所得的分數普通以答對題數為計算的單位，答對十個題目得十分，答對十二個題目得十二分，這種分數可以稱之謂原來分數（*Original scores*）。我們略用思索就可以曉得這種原來分數是完全沒有確定的意義的。他的意義依測驗的長短，難度等而異。所以對於這種意義不確定的分數必須加以解釋，才能曉得各分數所代表的程度。我們在第三章內討論的常模——年齡常模和年級常模——其功用就在解釋原來分數並確定其意義。此外還有許多測驗中通用的量數，亦為解釋原來分數之用。現在我們先討論年齡量表中各種解釋的量數。

一 年齡分數

智力年齡是最普通的解釋的單位。推其原因，約有二端：第一，因為年齡分數易於了解。譬如說，一人的智力年齡是十歲，這就是說此人的智力與十歲兒童的平均智力相等，其意義易為普通人所了解。第二，第一個智力測驗，即比納西門智力量表即以年齡為計算的單位。後來的智力測驗多以比納西門智力量表為模範，故仍襲用其年齡單位。

求智力年齡的方法是很簡單的。如推孟訂正比納量表上自三歲組至十歲組每組有六個測驗，答對每個測驗得智齡兩月。所以比納量表上的分數是直接以智力年齡表示的。團體智力測驗上的分數就是所答對的題數；要將這種分數變成年齡分數，必須先有對照表。我們必須求每個年齡兒童在測驗上所得的平均分數。例如一個十二歲兒童在一個智力測驗上所得分數與十歲兒童的平均分數相等，我們就說他祇有十歲的智力，或者他的智力年齡是十歲。

以智力年齡為計算的單位，亦有幾個缺點：

(一) 單位不等。這點我們在第二章內已經約略討論過了。

(二) 據研究結果，智力增長在一定時期即停止。至於在人生的什麼時期，智力停止生長，尚無一定的說法，大概在二十歲左右。假設智力的生長到十八歲就停止，那麼十八歲以上各年齡的人的平均分數應當與十八歲的平均分數相等，我們就不能有十八歲以上的智力年齡了。假設有一個聰明兒童在智力測驗上所得分數超過了十八歲兒童的平均分數，那麼我們就沒有法子以智力年齡去表示，估計他的智力程度了。所以智力年齡這個數量在應用上受很大的限制。因為這個原因薩孟司和露雪都說智力年齡這個單位不適用於中學的階段。

(三) 求常模之困難。要求可靠的年齡常模，那麼所測量到的各年齡兒童必須是隨機的取樣，這一點是很難做到的，尤其是在教育不普及的社會裏。譬如在中國，我們可以肯定的說，進小學的兒童比較是優秀的份子，

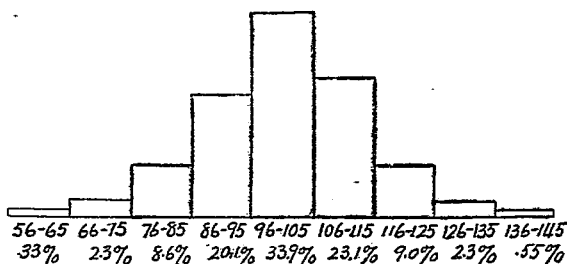
是經過選擇作用的。假如我們根據小學的兒童去求智力的年齡常模，所得結果一定太高，不能代表一般兒童的程度。就是義務教育比較普及的國家，求八歲以下，十二歲以上的兒童的常模，亦就難免選擇作用的困難。八歲以下許多兒童還沒有入學，十二歲以上的兒童有的已入中學，有的已離開學校到社會裏謀生去了，這些人我們求常模時都沒有法子可以測量到，因此所得結果，常不能合於隨機取樣的條件，所求得的年齡常模，就不能可靠了。

教育年齡與科目年齡的解釋與智力年齡同，並且在應用上亦受同樣的限制。學校中許多科目並非逐年都有的；許多學習得來的智識，一經間斷，就免不了遺忘，所以高年齡兒童的成績反不如低年齡兒童的成績，這種現象在事實上是可能的。要是如此，那麼高的科目年齡所代表的程度反不如低的科目年齡所代表的程度，科目年齡變成何等矛盾，何等容易令人誤會的名詞！

二 智力商數與教育商數

智力商數的意義在第四章中已約略述及。智力年齡表示絕對的智力程度，智力商數表示比較的聰明程度；一人的智力程度是逐年增加的，而一人的聰明程度在理論上是固定不變的，智力年齡表示現在的智力程度，智力商數表示已往並且可以預測將來智力發展的速率。

(一) 智力商數的分配 智力商數的分配可以用數學中常態的曲線表示之。這一點我們有許多的測驗



圖八 905 個從五歲到十四歲兒童的智力商數的分配(從 Terman)

結果可以證明。最著名的例子就是推孟所測量的九百零五個從五歲到十四歲兒童的智力商數的分配，如圖八。

智力商數的分配又可以用下列方式表示之：

- 最低百分之一在七十或以下；最高百分之一達到一三〇或以上。
- 最低百分之二在七十三或以下；最高百分之二達到一二八或以上。
- 最低百分之三在七十六或以下；最高百分之三達到一二五或以上。
- 最低百分之五在七十八或以下；最高百分之五達到一二二或以上。
- 最低百分之十在八十五或以下；最高百分之十達到一一六或以上。
- 最低百分之十五在八十八或以下；最高百分之十五達到一一三或以上。
- 最低百分之二十在九十一或以下；最高百分之二十達到一一〇或以上。
- 最低百分之二十五在九十二或以下；最高百分之二十五達到一〇八或以上。
- 最低百分之三十三又三分之一在九十五或以下；最高百分之三十三又三分之一在九十五或以下；最高百分之三十三又三分之一在九十五或以下；最高百分之三十三又三分之一在九十五或以下。

三又三分之一達到一〇六或以上。

如另用一種方式說明之，有如下述：

智力商數達到一一〇者，百人中有二十個人的智力與其相等或超過之。

智力商數到達一一五者，百人中有十個人的智力與其相等或超過之。

智力商數達到一二五者，百人中有三個人的智力與其相等或超過之。

智力商數達到一三〇者，百人中有一個人的智力與其相等或超過之。

反之：

得智力商數九十者，百人中有八十個人的智力超過之或與其相等。

得智力商數八十五者，百人中有九十個人的智力超過之或與其相等。

得智力商數七十五者，百人中有九十七個人的智力超過之或與其相等。

得智力商數七十者，百人中有九十九個人的智力超過之或與其相等。

各個智力商數的診斷的意義，推孟規定如下：

智 商 類 別

140 以上

近似天才或天才

130—140	很優異的智力
110—120	優異的智力
90—110	平常的智力
80—90	愚 笨
70—80	近似缺陷
70以下	低 能

關於上例根據智力商數的分類，我們必需明瞭智力的分配是連續的，其中並無間斷，所以說智商在七十以下為低能，七十至八十為近似缺陷，這些都是人定的類別，其中並無絕對的，自然的界限。低能的診斷，尤非僅僅智力測驗所能做到。

(二) 智力商數的固定性 前面我們說，智力商數代表人們天賦的聰明程度。現在我們要問智力商數能否正確的代表聰明程度，我們有什麼證據？我們假定人們的聰明程度是不變的，所以要回答智力商數能否正確的代表聰明程度一問題，最直接最圓滿的方法是去實地觀察智力商數究竟變動不變動？研究此問題的方法亦很簡單，就是間一年或數年繼續的測量同一羣兒童，再計算各兒童逐年的智力商數，看究竟有無變動或變動的數量。美國心理學者爲了智力商數固定否一問題做了許多實驗，結果指示我們一人逐年的智力商數不能絕對

的固定，但變動的數量很小。所以大致而論，智力商數可以代表一人的天賦智力。從比納西門智力量表上得來的智力商數的固定性，可以由表二十八所載結果表示之。

表二十八 智力商數的固定性（從 Freeman, Mental Tests, 表二十一）

研究人	兒童人數	變動在十分以 上的百分比	中間百分之五十 的變動的限度	平均變動	兩次測驗結果 的相關係數
Terman	435	.15	-3.3到+5.7	4.5	.93
Rugg and Collison	137	.12	-2.3到+5.6	4.7	.84
Garrison	468	.085	-2.0到+4.0 -3.0到+4.0 -3.0到+5.0	5.4	.88
Rugg, L. S.	114		-1.2到+1.9	3.1	.95

表二十八所載事實可以總結如下：

1. 第一次第二次測驗所得智力商數的平均差別是五分。
2. 百分之五十的變動的數量，正的在六分以內，負的在三分以內。
3. 第一次第二次測驗所得的智力商數的相關度大概在 .85 - .95 之間。

費里門以爲智力商數的實在固定性當比表二十八所表示者爲高。因爲兩次測驗所得智力商數的差別一部分亦許是由於測量的錯誤而並非因智力上真正有何變易。

(三) 智力商數的缺點或限制 智力商數的意義及價值既如上述，現在進而討論其困難或限制：

1. 第一個困難即計算高年齡兒童的智力商數時，公式內應用何除數（實足年齡）無一定之標準。這是因爲對於智力生長何時停止一問題尙無一定的結論。推孟用十六歲爲求成人的智商時的除數（所謂成人，指十六歲以上的人），賓特納主張用十四歲爲求成人的智商時的除數（所謂成人，指十四歲以上的人）。

2. 第二個困難 即或者人們智力的生長並不在同一年齡停止。許多實驗結果指示聰明程度的高低與智力生長停止時期的遲早有相當的關係。例如道爾 (E. A. Doll) 研究低能兒童表示智力的生長到十三歲半即停止。第亞彭 (W. F. Dearborn) 根據公立小學及補習學校智力測驗的結果，謂智力生長到十四五歲之間停止。桑戴克根據中學校智力測驗的結果謂中學學生智力的生長到中學完畢時仍未停止。從上面三種研究結果看來，智力生長停止的時期，因人而異。人愈聰明，大概智力生長停止的時期愈遲；人愈愚笨，智力生長停止的時期愈早。要是這個假說是不錯的，那麼我們計算高年齡兒童的智力商數時，似不應以一固定的年齡。其困難情形，可想而知。

教育商數的解釋，亦如智力商數，在應用上亦受同樣的限制。其公式如下：

$$\text{教育商數} = \frac{\text{教育年齡}}{\text{實足年齡}} \times 100$$

三 成業商數

成業商數所代表者乃一人智力與學業的比較數。或者可以說，成業商數可以表示一人的學業與同智力年齡的兒童的學業的比較。所以這個數量一方面可以代表學生的努力程度，一方面可以代表教員的效率。因此有人提議用成業商數為報告學業成績的分數，為考查教員效率的根據。

現在報告學業成績所用的分數，是一種絕對的分數。所以智力低的學生，不論如何用功，如何努力，總不能得着高的分數。一方面智力高的學生，不用努力，不用用功，就可以得到高的分數。因此用絕對的分數表示學生學業的成績，一方面易使智力低落的學生失望灰心，一方面對於智力優異的學生容易養成懶惰的習慣，自傲的心理。這是何等不公平的事。若以成業商數來報告學生成績，則不論學生的智力高低，都有同等機會得高分數，同等機會得低分數。

現在考查教員效率的方法，多以學生學業的成績，來判斷教員的效率，對於學生的智力是不加考慮的。所以教低智力學生的教員，常常顯着效率很低，雖然已盡其力之所能；教高智力學生的教員，雖然不曾努力，亦常顯着效率很高。這是極不公平的方法。用學生成業商數來考查教員的效率就比較公平了。

成業商數的功用既如上述，但是對於成業商數的缺點以及一般人對於這個數量的誤解，我們亦不應有簡單的討論。

第一，成業商數的方法在理論上，假設智力與學業之間呈完全的相關，要是努力和其他因素的影響能夠除去。根據這個臆說，智力愈高學業亦應該愈好，要是每人都能夠盡其力之所能。這個臆說，我們可以說大概是錯誤的。智力與學業之間，大致而論，是有正相關的關係，但是並非完全的相關，並且智力與各學科間的相關大有高低深淺之別，並非相等的。譬如讀法，算術等學科與智力的相關是比較的高；圖畫，手工，音樂等學科與智力的相關就很低了。

據調查的結果，智力商數高的兒童常得低的成業商數，智力商數低的兒童常得高的成業商數。普通的解釋說照現在學校學級編制，愚笨的兒童所在年級常比依照他的智力程度所應在的年級高些，聰明的兒童所在年級常比依照他的智力程度所應在的年級低些。而且，愚笨的兒童常為教員所特別注意，聰明的兒童常被教員所忽略。因此，智力商數低的兒童能盡量應用，發展其天賦的能量，而智力商數高的兒童則反是。這個解釋當然是正確的，但並非完全的解釋。聰明兒童成業商數低，愚笨兒童成業商數高，這個現象一部分是由於上段所述智力與學業並不完全相關的原因。舉一個絕端的例，假設智力與某科作業是毫無相關的。設有五組兒童，年齡相等而智力商數不相等。因為某科作業與智力是不相關的，所以五組兒童的平均作業却是相等的。照此情形，那末智商愈

高，則成業商數愈低；智商與成業商數之間呈負相關的關係，乃無可避免的結果了。以成業商數表示努力的程度，對於聰明的兒童顯然的不公允不正確了。

第二，許多心理學者根據能力不能超過能量的臆說，在理論上成業商數不能超過1.00的數量。這種觀念可以說是一種誤解。智力測驗所測量的能力並不是一種與生俱生不授環境影響的能力，所謂「智力」實在不過是各種能力的平均而已。在某一種測驗上的作業表示某一方面的特殊能力。很自然的，某一方面的特殊能力有超過或低於平均能力的可能，所以成業商數可以超過或低於1.00的數量。

實際測量結果表示成業商數的分配與智力商數的分配差不多，並不限於1.00以下的。表二十九表示根據美國「國家智力測驗」量表乙與吳特（O. Wood），麥柯爾算術測驗所得的成業商數的分配。

表二十九 根據美國國家智力測驗量表乙與吳特，麥柯爾算術測驗所得的算術成業商數的分配

(從Symonds, Measurement in Secondary Education Table 84)

算 術 成 業 商 數	次 數
135—139	2
130—134	2
125—129	12

120—124	9
115—119	16
110—114	30
105—109	39
100—104	34
95—99	21
90—94	18
85—89	17
80—84	3
75—79	2

第三，成業商數的可靠度基於兩種事實。第一種事實就是成業商數所根據的智力測驗與教育測驗的可靠度。假如所根據的智力測驗（或所得智力年齡）與教育測驗（或所得教育年齡）本身就不可靠，那末，從他們所得來的成業商數當然亦不可靠。第二種事實就是智力測驗與教育測驗所測量能力的異同。假如智力測驗與教育測驗所測量的能力是完全不同的，那末，兩種測驗結果的差別（假設兩種測驗本身是可靠的）是真正的，可靠的差別。假如智力測驗與教育測驗所測量的能力大部分或一部分是相同的，那末，兩種測驗結果的差別就大部分或一部分是隨機的，任運的，不可靠的。

據克萊 (T. L. Kelly) 的觀察，現有的智力測驗與綜合的教育測驗所測量的對象有百分之九十是相同的，因此根據二者所得的成業商數是極不可靠的。所以克萊主張將成業商數廢棄不同。智力測驗與單獨學科測驗的內容相同的部分當然要少些，因此各學科的成業商數亦比較可靠些。還有團體的成業商數當然要比個人的成業商數可靠些。

總結以上關於成業商數的討論，成業商數的觀念在教育上是很有用的，但是求成業商數的方法有許多缺陷。我們應該明瞭成業商數的可靠性，正確度不如一般人所想像者之高。我們一方面固然不妨仍舊應用他為施行指導工作的根據，同時亦不可把他當作一種極可靠極正確的數量。

成業商數的低落，除掉上面所述因為智力與學業間並不完全相關，所以對於聰明兒童為不可避免的現象外，還有許多的原因，如學生不努力，教授不良，學級編製之失當，缺乏相當的鼓勵等都為重要的原因。無論如何，低的成業商數乃是學生與學校不相適應的表徵，我們應該進一步究其原因之所在，以圖補救。

問題

1. 從各測驗所得之原來分數何以不能比較？
2. 測量的錯誤何以能增加智力商數的變異？

3. 智力商數這個量數有什麼缺陷或限制？
4. 成業商數與教育商數之區別何在？
5. 據調查結果，智力商數高的兒童常得低的成業商數，智力商數低的兒童常得高的成業商數，這種現象如何解釋？
6. 成業商數何以不是一個可靠的量數？
7. 一個兒童成業商數的低落，有什麼可能的原因？
8. 根據學生的教育程度去考查學生的努力與教員的效率，何以不公平？
9. 智力年齡與智力商數這兩個量數何也不適用於中學階段？

參考書報

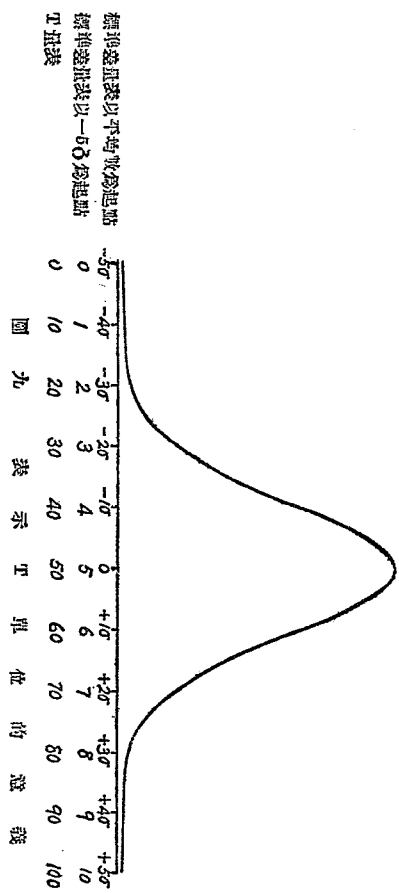
1. Fryeman, F. N.: *Mental Tests*, p. 285—291, Macmillan.
2. Greene, H. A. and Jorgenson, A. N.: *The Use and Interpretation of Educational Tests*, Ch VII, Longmans.
3. Kelley, T. L.: *Interpretation of Educational Measurements*, p. 21—25, World Book Co.
4. Symonds, P. M.: *Measurement in Secondary Education*, Ch. XV Macmillan.
5. Tennant, L. M.: *The Intelligence of School Children*, Ch. IX Houghton Mifflin.

第十八章 測驗結果的解釋：TBOE分數的意義

一 H分數

(一) H分數的意義 H為度量某種特性或能力的單位。此單位等於十二歲兒童（或其他年齡團體）某種特性或能力的分配的標準差的十分之一。其功用和兩寸、秒等等度量單位一樣。H量表的創始者麥柯爾以H名之，含有紀念推孟（Terman）與桑戴克（Thorndike）二氏之意。

每種度量必有一個計算的起點。最完善的起點是零點，如度量時、空、重各種量表都是如此。我們測量智力或算術能力，理論上亦應該從智力或算術能力的零點計算起。但是實際上智力或算術能力的零點無從決定。所以教育上的度量，不得不用人為的參照點。H量表的起點是十二歲兒童（或其他年齡團體）的平均能力以下五個標準差。讀者請參看圖九。



圖九 標準差單位

把零點放在負五個標準差這一點上，可以避免負的分數，而且十二歲兒童（或其他年齡團體）的平均能力恰恰等於 H_0 ，極易於記憶。所以凡個人或全班的 H 分數是 S_0 ，就表示那個人或那一班學生的能力同十二歲兒童（或其他年齡團體）的平均能力恰恰相等。凡得 S_0 的學生，他的能力比十二歲兒童（或其他年齡團體）的平均能力高 $S_0 - H_0$ ；凡得 S_0 的兒童，他的能力比十二歲兒童（或其他年齡團體）的平均能力低 $S_0 - H_0$ 。除類推。

再從十二歲兒童（或其他年齡團體）超過某一個百分數的百分比來計算，我們對於百分數可以得到另一種解釋。參看表三十就可以明瞭。

表三十 百分數的一種解釋

一個百分數	十二歲兒童(或其他年齡團體) 超過那一個百分數的百分比	一個百分數	十二歲兒童(或其他年齡團體) 超過那一個百分數的百分比
25	99	55	31
30	98	60	16
35	93	65	7
40	84	70	2
45	69	75	1
50	50	80	0.1

例如一個兒童得了50分，查表三十就可以知道每一百個十二歲（或其他年齡）兒童中有八十四個的能力比他低，祇有十六個比他高。

編造百分表，根據任何團體都可以。麥柯爾提議以十二歲兒童為標準的團體，他的理由是：第一，十二歲兒童容易測量到，不致有選擇作用的困難。第二，十二歲兒童的能力的差異度極大，而且十二歲兒童能力的分配在全部能力量表上所佔的地位比較最為適中。同時麥柯爾又提及在編製低年級百分表時，可以根據八歲的兒童，在

編製中學量表時，可以十六歲兒童為標準的團體。

IQ分數可以用於各種測驗，所以必須在IQ字之旁註相當符號，以表示IQ分數是從什麼測驗得來的。所以

IQ或IQ_i = 智力IQ分數

IQ_e或IQ_e = 教育IQ分數

IQ_g或IQ_g = 體法IQ分數

IQ_a或IQ_a = 算術IQ分數

餘類推。

(二) IQ分數的求法 求IQ分數有一個很簡單的公式。

$$IQ = 50 + \frac{10(X - M)}{S}$$

假設十二歲兒童在某種智力測驗上的平均分數為163，標準差為10.2，今有某生在此測驗上得179分，其IQ分數應為65.7

$$IQ = 50 + \frac{10(179 - 163)}{10.2} = 65.7$$

在應用標準測驗時，編製者備有現成的對照表，當然可以不必自己去求。祇要把某人在測驗上所得的分數算出，就可以從對照表查出他的IQ分數。表三十一是廖氏團體智力測驗量表甲的IQ量表。設有某生在廖氏團體智力測驗量表甲上得11分，查該表，其IQ分數為88，表示該生的智力恰等於十二歲兒童的平均程度。

教育測驗
表三十一 廖氏團體智力測驗量表甲的T分數

做對的分數	T分數	做對的分數	T分數
0- 1.9	17	62- 63.9	60
2- 3.9	20	64- 65.9	61
4- 5.9	23	66- 67.9	62
6- 7.9	26	68- 69.9	64
8- 9.9	28	70- 71.9	65
10- 11.9	30	72- 73.9	66
12- 13.9	31	74- 75.9	68
14- 15.9	32	76- 77.9	70
16- 17.9	33	78- 79.9	71
18- 19.9	34	80- 81.9	72
20- 21.9	35	82- 83.9	74
22- 23.9	37	84- 85.9	75
24- 25.9	39	86- 87.9	76
26- 27.9	40	88- 89.9	77
28- 29.9	42	90- 91.9	78
30- 31.9	43	92- 93.9	78
32- 33.9	45	94- 95.9	79
34- 35.9	46	96- 97.9	81
36- 37.9	47	98- 99.9	83
38- 39.9	49	100-101.9	85
40- 41.9	50	102-103.9	87
42- 43.9	51	104-105.9	89
44- 45.9	52	106-107.9	91
46- 47.9	53	108-109.9	93
48- 49.9	54	110-111.9	95
50- 51.9	54	112-113.9	97
52- 53.9	55	114-115.9	99
54- 55.9	56	116-117.9	101
56- 57.9	57	118-119.9	103
58- 59.9	58	120-121.9	105
60- 61.9	59		

二 B 分數

(一) B 分數的解釋 B 分數表示聰明程度，其功用與年齡量表中的商數——智力商數與教育商數——同。兩個學生的 B 分數相等，他們的 B 分數未必相同，因為或者他們的年齡大小不一。B 分數代表一人某能力的絕對的數量，B 分數代表一人某能力的相對的程度。知道了某人的 B 分數，我們可以將他的程度同十二歲兒童的程度相比較；知道了某人的 B 分數，我們可以將他的程度同他同年齡兒童的程度相比較。

表三十亦可以拿來解釋 B 分數。B 分數常模總是 50。倘使一個兒童的 B 分數是 50，他的能力就和他同年齡兒童的平均能力相等，所以他的聰明程度是中等。依據表三十，他同年齡兒童中恰恰有百分之五十在他程度之上。倘使一個兒童的 B 分數是 60，他的能力就比他同年齡兒童的平均能力低 10。依據表三十，他同年齡兒童裏面，有百分之八十四比他的能力高。可見此兒童的聰明程度是在中等以下。倘使一個兒童的 B 分數是 70，他的能力比他同年齡兒童的平均能力高 20。依據表三十，他同年齡兒童裏面，祇有百分之十的程度在他之上，百分之九十九的程度在他之下，可見這個兒童的聰明程度是極高了。

各種測驗都可以有 B 分數，所以必須註明，以示區別，所用符號，與 B 分數同。

B 分數本是 (Brightness) 的縮寫，但同時亦含有記念年齡量表的創始者比納 (Binet) 與年級量表創始者白根漢 (Buckingham) 之意。

(二) 百分數的求法。百分數加一個校正數等於百分數。百分數加校正數，各年齡不同。年齡小於十二歲半的兒童，其校正數為正數；年齡大於十二歲半者，其校正數為負數；十二歲半的兒童，其校正數為零。各年齡的校正數如何決定，將在第二十一章內詳細討論。現在我們舉一個簡單例子，以明校正數的意義。在廖氏團體智力測驗量表甲上（參看表三十一），得百分者其百分數為80，得85分者其百分數為85，得90分者其百分數為90。根據測量結果，十一歲半的兒童在廖氏團體智力測驗上的平均分數恰恰為80，十三歲半的兒童的平均分數恰恰為85。設有一個兒童，年齡恰為十一歲六個月，在該測驗上，得了82分，其百分數為82，表示他的程度比十二歲兒童的平均程度低四個百分點。但是這個兒童的程度恰與同年齡兒童的平均程度相等，他的百分數應該是80.149。這就是他的校正數，亦是其他十一歲六個月的兒童的校正數。再假設一個兒童，年齡恰為十三歲六個月，在該測驗上，得了88分，其百分數為88，表示他的程度比十二歲兒童的平均程度高四個百分點。但是他的程度又恰與同年齡兒童的平均程度相等，他的百分數應該是88.149。這就是他的校正數，亦是其他十三歲六個月的兒童的校正數。

各種標準測驗都備有各年齡百分校正數表，所以祇須從表中一查就行了。但是在查表的時候，必須預先知道兒童的實足年齡。我們計算年齡的通俗方法，極不準確，有時與實足年齡相差至兩歲之多。為正確計，必須加以校正。廖世承陳鶴琴二氏曾有一個極簡便的方法，茲介紹如下：

1. 應知兒童年齡是用陽曆或陰曆計算。如用陽曆，可直接計算。

2. 如用陰曆計算的，先求兒童的陰曆年齡與生日，從年齡的數目減去一歲。例如六歲，三月十六日生，暫作五歲計算。

3. 從舉行測驗的日期（陰曆）減去生日（陰曆）。

（1）倘所得的數是正的，就加在減去一歲的年齡上。例如測驗在陰曆五月十四日舉行的，減去三月十六日，得二月二十八日，作為二月。加上五歲，實足年齡是五歲二月。

（2）倘所得的數是負的，就從減去一歲的年齡上減去。例如測驗在陰曆一月二十六日舉行的，減去三月十六日，得負一月二十日，作為二月。從五歲減去，實足年齡為四歲十月。

民國二十年上海市社會局所編印的曆書裏面有一個年齡換算表，較廖陳二氏的方法尤為簡便。讀者請參看表三十二。

表三十二 年齡換算表

出世時 陰曆 月份	換算時 陽曆					
	一 月	二 月	三 月	四 月	五 月	六 月
正 月	十 一 個 月	一 年	一 年 一 個 月	一 年 二 個 月	一 年 三 個 月	一 年 四 個 月

第三編 測驗的實施

正月	陰曆 份曆	十二月	十一月	十月	九月	八月	七月	六月	五月	四月	三月	二月
一年五個月	七 月	○	一個月	二個月	三個月	四個月	五個月	六個月	七個月	八個月	九個月	十個月
一年六個月	八 月	一個月	二個月	三個月	四個月	五個月	六個月	七個月	八個月	九個月	十個月	十一個月
一年七個月	九 月	二個月	三個月	四個月	五個月	六個月	七個月	八個月	九個月	十個月	十一個月	一年
一年八個月	十 月	三個月	四個月	五個月	六個月	七個月	八個月	九個月	十個月	十一個月	一年	一年一個月
一年九個月	十一 月	四個月	五個月	六個月	七個月	八個月	九個月	十個月	十一個月	一年	一年一個月	一年二個月
一年十個月	十二 月	五個月	六個月	七個月	八個月	九個月	十個月	十一個月	一年	一年一個月	一年二個月	一年三個月

十二月	十一月	十月	九月	八月	七月	六月	五月	四月	三月	二月
六個月	七個月	八個月	九個月	十個月	十一個月	一年	一年一個月	一年二個月	一年三個月	一年四個月
七個月	八個月	九個月	十個月	十一個月	一年	一年一個月	一年二個月	一年三個月	一年四個月	一年五個月
八個月	九個月	十個月	十一個月	一年	一年一個月	一年二個月	一年三個月	一年四個月	一年五個月	一年六個月
九個月	十個月	十一個月	一年	一年一個月	一年二個月	一年三個月	一年四個月	一年五個月	一年六個月	一年七個月
十個月	十一個月	一年	一年一個月	一年二個月	一年三個月	一年四個月	一年五個月	一年六個月	一年七個月	一年八個月
十一個月	一年	一年一個月	一年二個月	一年三個月	一年四個月	一年五個月	一年六個月	一年七個月	一年八個月	一年九個月

表三十二的用法如下，先從陰曆年歲減去兩歲，然後增入出世時陰曆月份之直行與換算時陽曆月份之橫行交叉處所載之年月。例如有一人，其陰曆年齡係二十歲，生日係在陰曆十二月。換成實足年齡之法，是先從二十

歲減去兩歲，得十八歲；如換算時（或舉行測驗時）係陽曆十月，則應加之年月，是陰曆十二月之直行與陽曆十月之橫行交叉處的月數，即九個月，所以此人之實足年齡為十八歲另九個月。

實足年齡求得後，即可從測驗說明書內所載各年齡校正數表尋出相當之校正數。然後將該校正數加在H分數之上，即得H分數。茲舉一實例以明之（參看表三十三）。設有一個兒童，實足年齡為七歲十個月，在廖氏團體智力測驗上所得H分數為88。查表三十三，七歲十個月兒童應有之B校正數為21.53 + 27 ÷ 100，就是該生的H分數。

表三十三 廖氏團體智力測驗各年齡
B校正數表

實年	足齡	B校正數	實年	足齡	B校正數	實年	足齡	B校正數
年	月		年	月		年	月	
6	0	50	10	10	8	15	8	-11
6	2	47	11	0	7	15	10	-11
6	4	44	11	2	6	16	0	-12
6	6	42	11	4	5	16	2	-12
6	8	40	11	6	4	16	4	-13
6	10	38	11	8	4	16	6	-13
7	0	36	11	10	3	16	8	-14
7	2	34	12	0	2	16	10	-14
7	4	32	12	2	1	17	0	-14
7	6	31	12	4	1	17	2	-15
7	8	29	12	6	0	17	4	-15
7	10	27	12	8	-	1	6	-16
8	0	26	12	10	-	1	8	-16
8	2	25	13	0	-	2	10	-17
8	4	23	13	2	-	3	18	0
8	6	22	13	4	-	3	18	2
8	8	21	13	6	-	4	18	4
8	10	19	13	8	-	4	18	6
9	0	18	13	10	-	5	18	8
9	2	17	14	0	-	5	18	10
9	4	16	14	2	-	6	19	0
9	6	15	14	4	-	7	19	2
9	8	14	14	6	-	7	19	4
9	10	13	14	8	-	8	19	6
10	0	12	14	10	-	8	19	8
10	2	11	15	0	-	9	19	10
10	4	10	15	2	-	9	20	0
10	6	9	15	4	-	10		
10	8	8	15	6	-	10		

(三) B量表的批評 從上面的解釋，我們可以明瞭B校正數就是各年齡兒童平均B分數與的H₁₀差別。所以要決定各年齡的B校正數，必須先求各年齡的平均B分數，換言之，即各年齡的B分數常模。在第十七章內討論年齡量表時，我們曾提及求常模之困難為年齡量表缺點之一。這種缺點B量表亦不能避免。

三 O 分數

(1) O 分數的解釋 O 分數表示某種能力的年級地位。有了O分數，就可以知道某學生的某種能力相當於那一年級。O本來是(Classification)的縮寫，同時又含有記念克底斯(Courtis)和喀泰爾(Cattell)之意。O分數的解釋如下：

- 1.0 O 代表一年級開始時的兒童的程度。
- 1.1 O 代表讀完一年級一個月的兒童的程度。
- 1.2 O 代表讀完一年級二個月的兒童的程度。
- 1.3 O 代表讀完一年級三個月的兒童的程度。
- 1.4 O 代表讀完一年級四個月兒童的程度。
- 1.5 O 代表讀完一年級上學期的兒童的程度或一年級兒童的平均程度。

1.6.0 代表讀完一年級下學期第一個月的兒童的程度。
餘類推。

從各種測驗所得的 Q 分數，亦應註明，所用符號，與 H, W 分數同。

(二) Q 分數之求法 要求 Q 分數，必須先求 W 分數。W 為 Grade 的縮寫，代表未校正的年級地位，或被試者的程度在測驗時的年級地位。Q 分數可以從 H 分數與 W 分數對照表查出。Q 量表的編造是很簡單的。我們拿測驗去測量各年級兒童，再求各年級兒童的平均 H 分數。譬如三年級的平均 H 分數為 95.0，那麼凡得 H 95.0 分者，其程度與三年級兒童的平均程度相等，其 Q 分數應為 3.5。表三十四為廖氏團體智力測驗 Q 量表。

驗 G 量表

T 分數	G 分數	T 分數	G 分數
72.0	9.7	87.8	12.6
72.6	9.8	88.4	12.7
73.1	9.9	88.9	12.8
73.6	10.0	89.5	12.9
74.2	10.1	90.0	13.0
74.7	10.2	90.6	13.1
75.3	10.3	91.1	13.2
75.8	10.4	91.7	13.3
76.4	10.5	92.2	13.4
76.9	10.6	92.8	13.5
77.5	10.7	93.3	13.6
78.0	10.8	93.9	13.7
78.6	10.9	94.4	13.8
79.1	11.0	95.0	13.9
79.7	11.1	95.6	13.0
80.2	11.2	96.1	14.1
80.8	11.3	96.7	14.2
81.3	11.4	97.2	14.3
81.9	11.5	97.8	14.4
82.4	11.6	98.3	14.5
83.0	11.7	98.9	14.6
83.5	11.8	99.4	14.7
84.1	11.9	100.0	14.8
84.6	12.0	100.6	14.9
85.1	12.1	101.1	15.0
85.6	12.2		
86.2	12.3		
86.8	12.4		
87.3	12.5		

表三十四 廖氏團體智力測

T分數	G分數	T分數	G分數	T分數	G分數
24.4	1.0	40.2	3.9	56.1	6.8
25.0	1.1	40.7	4.0	56.7	6.9
25.5	1.2	41.3	4.1	57.2	7.0
26.1	1.3	41.9	4.2	57.7	7.1
26.6	1.4	42.4	4.3	58.2	7.2
27.2	1.5	43.0	4.4	58.9	7.3
27.7	1.6	43.5	4.5	59.5	7.4
28.3	1.7	44.1	4.6	60.0	7.5
28.9	1.8	44.6	4.7	60.5	7.6
29.4	1.9	45.1	4.8	61.0	7.7
30.0	2.0	45.7	4.9	61.6	7.8
30.5	2.1	46.3	5.0	62.2	7.9
31.1	2.2	46.9	5.1	62.8	8.0
31.6	2.3	47.4	5.2	63.4	8.1
32.2	2.4	48.0	5.3	64.0	8.2
32.7	2.5	48.5	5.4	64.5	8.3
33.3	2.6	49.0	5.5	65.0	8.4
33.8	2.7	49.5	5.6	65.5	8.5
34.4	2.8	50.1	5.7	66.0	8.6
34.9	2.9	50.6	5.8	66.6	8.7
35.5	3.0	51.1	5.9	67.1	8.8
36.0	3.1	51.6	6.0	67.6	8.9
36.5	3.2	52.1	6.1	68.2	9.0
37.0	3.3	52.7	6.2	68.7	9.1
37.5	3.4	53.4	6.3	69.3	9.2
38.0	3.5	54.0	6.4	69.8	9.3
38.5	3.6	54.5	6.5	70.3	9.4
59.1	3.7	55.0	6.6	70.9	9.5
39.6	3.8	55.6	6.7	71.5	9.6

假設兩個兒童或兩班受試的時候不同，而我們比較他們的程度時，用 Ω 分數不是最公平的。假設有甲乙兩個兒童，同在春季始業的三年級肄業。兒童甲是在十月底測量的，得 Ω 分數 ω_1 ，兒童乙是在十二月底測量的，得 Ω 分數 ω_2 ，所以我們可以說兒童乙在十二月底時候的程度比兒童甲在十月底時候的程度高些。但是這不是公平的比較，因為他們被試的時候不同。這個被試時候不同的影響必須加以校正， Ω 分數就是已校正的年級

地位。

表三十五是○校正數與距開學時月數對照表。凡測量時距開學時恰恰五個月者，其○校正數為零；測量時距開學時在五個月以上者，其○校正數為負數；測量時距開學時在五個月以下者，其○校正數為正數。前例中兒童甲是在十月底測量的，假設九月初開學，則測量時距開學時為兩個月，查表三十五應得○校正數。兒童甲的○分數為3.7，所以他的○分數為3.1+3.3=6.4。兒童乙是在十二月底測量的，距開學時為四個月，應得○校正數。兒童乙的○分數為3.2，所以他的○分數為3.2+1=4.2。假如甲乙兩個兒童是同在三年級第五個月的時候測量的，就是在正月底測量的，則兒童乙的成績反不如兒童甲了。

表三十五 ○校正數與距開學時月數對照表

距開學時月數	1	2	3	4	5	6	7	8	9	10
○校正數	+0.4	+0.3	+0.2	+0.1	0	-0.1	-0.2	-0.3	-0.4	-0.5

如用表三十六來求○校正數，則更便易。

表三十六 ○校正數與測量時期對照表

(1) 秋季始業用

陽曆月終	9	10	11	12	1	2	3	4	5	6
O 校正數	+ .4	+ .3	+ .2	+ .1	0	- .1	- .2	- .3	- .4	- .5

(2) 春季始業用

陽曆月終	2	3	4	5	6	9	10	11	12	1
O 校正數	+ .4	+ .3	+ .2	+ .1	0	- .1	- .2	- .3	- .4	- .5

四 F 分數

F 爲 Effort 或 Efficiency 的縮寫，代表學生的效率或一人智力與學業的比較表。F 分數的解釋與年齡量表中的成業商數同，並亦受同樣的限制，茲不重述。F 又含有記念佛倫施 (Franz) 賓特納 (Pitner) 孟祿 (Monroe) 三氏之意。

求 F 的公式是很簡單的：

$$F \text{ 數} = F \text{ 密} + 50 = F$$

加 50 的原因，在避免負數，而使 F 的常模恰等於 50。

五 續 前

以上將 B, C, E 四種分數的意義詳為解釋。我們可以看到 B 量表是極完全的，凡年齡量表中所有各種量數，B 量表中皆有之。麥柯爾曾有一表，比較年齡量表與 B 量表中的各種量數，茲譯錄如下：

表三十七 B 量表與年齡量表中各種量數對照表

年 齡 量 表	B 量 表
C. A. = 實足年齡 (Chronological age)	C. A. = 實足年齡 (Chronological age)
M. A. = 智力年齡 (Mental age)	T. I. = 智力總分數 (Total intelligence)
E. A. = 教育年齡 (Educational age)	T. E. = 教育能力總分數 (Total educational ability)
R. A. = 讀法年齡 (Reading age)	T. R. = 讀法能力總分數 (Total reading ability)
Ar. A. = 算術年齡 (Arithmetic age)	T. A. = 算術能力總分數 (Total arithmetic ability)
其 他	其 他
I. Q. = 智力年齡 = 智力商數 (實足年齡) (Intelligence Quotient)	B. I. = 智力聰明數 (Brightness in intelligence)
E. Q. = 教育年齡 = 教育商數 (實足年齡) (Educational Quotient)	B. E. = 教育聰明數 (Brightness in Education)

R. Q. = 讀法年齡 = 讀法商數 (Reading Quotient) Ar. Q. = 算術年齡 = 算術商數 (Arithmetic Quotient)	Br = 讀法聰明數 (Brightness in Reading) Ba = 算術聰明數 (Brightness in Arithmetic)
共 他	共 他
A. Q. = 教育年齡 = 成業商數 (Accomplishment Quotient) R.A.Q. = 讀法年齡 = 讀法成業商數 (Reading Accomplishment Quotient) Ar.A.Q. = 算術年齡 = 算術成業商數 (Arithmetic Accomplishment Quotient)	F = 教育T分數 = 智力T分數 = 效率分數 (Effort or Efficiency) Fr = 讀法T分數 = 智力T分數 = 讀法效率分數 (Efficiency in Reading) Fa = 算術T分數 = 智力T分數 = 算術效率分數 (Efficiency in Arithmetic)
共 他	共 他

教育測驗所通用的量表共有四種，即年齡量表，百分量表，年級量表及T量表。麥柯爾曾對於這四種量表，根據十五種標準，加以分析的評判。其結果見表三十八。

表三十八 四種量表的比較 最適合標準者〇，尚稱適合者1，不適合者〇。

標	準	T 量表	年級量表	百分量表	年級量表
1. 參照點的確定和便利	2	2	2	1	0

2. 單位的相等	2	0	0	2
3. 量表各距的長度	2	0	2	2
4. 量表分數的可靠性	2	1	1	2
5. 量表的永久性	2	2	2	1
6. 量表單位的習慣性	2	2	2	2
7. 量表分數的通俗解釋	1	2	2	0
8. 單位的世界共同性	2	2	1	0
9. 各量表中分數的比較	2	2	1	1
10. 合併單位的方法	2	2	2	0
11. 計算分數的便易	2	2	2	0
12. 應用商數法的可能	2	2	0	0
13. 編造量表的便易	2	1	2	0
14. 測驗材料全部的應用	2	2	2	1
15. 準備副本量表的便易	2	1	2	0
總 數	29	23	22	11

表三十八所載的結果，雖然是主觀的評判，但是專家的意見是應當重視的。再者，表三十八所列各項標準，其重要性並不相等。如一二兩項極屬重要，一種量表，如對於這兩項標準不能適合，則其效用就受重大的限制。

問題

1. 何以麥柯爾以十二歲的兒童為編製量表的標準團體?
2. 何以麥柯爾以十二歲兒童能力分配的平均點以下一個標準差為計算的起點?
3. 某生在某種算術測驗上得B分數，問十二歲兒童中算術程度在他之上者有百分之幾?
4. 某生在某種算術測驗上得B分數，問同年齡兒童中算術程度在他之上者有百分之幾?
5. 年齡小於十二歲半的兒童其D校正數為正數，年齡大於十二歲半的兒童，其D校正數為負數，何故?
6. 假設兩個兒童是在同時測量的，用Q分數以比較他們的程度，是否公平？何故?
7. 某生在春季始業的四年級肄業，實足年齡為5歲9月，於五月底施以優氏團體智力測驗量表，甲得30分，試求該生的(1)E分數，(2)B分數，(3)G分數，(4)C分數。

參考書報

1. McGall, W. A., How to Measure in Education, Ch. 9, 10 Macmillan.
2. McGall, W. A., How to Experiment in Education, Ch. 5, Macmillan.

第十九章 測驗結果的應用

測驗結果在教育上的應用，可以分三部分來陳述：一、在教育行政及輔導上的應用，二、在教學上的應用，三、在教育研究上的應用。

一 在教育行政及輔導上的應用

測驗在教育行政及輔導上的應用，其重要者有下列數端：

(一) 考查學生的進步 譬如要考查學生在學期中的進步，可在學期之初與學期之終舉行標準測驗，兩次所得分數的差別，就代表進步的數量。

拿個人，一班或一學校的成績與常模比較，更可以得到許多有意義的結論。但是這種比較並不是簡單的，為得到正確的解釋起見，下列各項事實必須顧及：

1. 學生的年齡 對於教育測驗結果的解釋，最容易犯的錯誤就是忽略學生的年齡。我們知道各校中同一年級兒童的年齡並不一致，有時差異極大。下列假設的事實可以明白表示這種錯誤的性質：

學校甲九年級的平均分數……………四〇・一

學校乙九年級的平均分數……………四二・八

九年級的常模……………四〇

九年級常模所根據的兒童的平均年齡……………十五歲六個月

十年級的常模……………四五

十年級常模所根據的兒童的平均年齡……………十六歲六個月

學校甲九年級學生的平均分數……………十五歲七個月

學校乙九年級學生的平均分數……………十六歲五個月

驟然看來，學校乙的程度顯然較學校甲為優。假如我們細細考查這兩級的平均年齡以及常模所根據的兒童的平均年齡，結論便不同了。學校甲九年級學生的平均年齡與九年級常模所根據的學生的平均年齡差不多相等，同時學校甲九年級的平均分數與九年級常模又恰恰相等，這樣看來，學校甲的程度可稱為正常。學校乙九年級學生的平均年齡差不多與十年級常模所根據的兒童的年齡相等（前者為十六歲五個月，後者為十六歲六個月），同時學校乙九年級學生的平均分數遠不及十年級常模（前者為四二・八，後者為四五），這樣看來，學校乙九年級的程度實在在正常以下。

所以在比較兩班或兩校的程度時，或將一班一校的程度與常模比較時，學生的年齡必須顧及。

2. 學生的智力 一班中各學生，各班學生，各校學生的智力程度不等，這是很顯著的事實。所以在比較兩班或兩校的教育程度時，或將一班一校的教育程度與常模比較時，學生的智力程度必須同時顧及。

3. 學生以往的學歷 學生現在的程度當然受以往所受教育的影響。學生現在所在學校，當然不應替以往所在學校領功或受過，所以在考查或比較某一學級或某學校的程度時，對於學生以往的經歷須予以相當的注意。

4. 測驗的內容與學校的課程是否適合，亦是解釋測驗結果時不容忽略的一點。我們必須認清任何教育測驗祇能代表某一學科內容的一部分，或一個有限的取樣，因此不一定與某學校的課程相適合。換句話說，標準測驗是常常不能適合當地情形的。要解決這種困難，祇有在解釋與應用測驗結果時特別審慎些。

(二) 考查教師的效率 如其他條件相等，則根據學生的教育程度以比較教師的效率，是很正當的辦法。但是其他條件是常常不相等的。要公允地比較兩位教師的教學效率，除掉兩班學生當時的教育程度外，還須顧及下列各項事實：

1. 這兩班學生在學年開始時的程度是否相等？
2. 這兩班學生的智力程度是否相等？
3. 這兩班學生的年齡是否相等？

4. 所用測驗的內容與這兩班教材的內容是否同樣地適合？

除了上列四點以外，還有許多其他事項，如學年長短，家庭狀況，學校設備等，在比較教師效率時亦須顧到，不過以上四點比較地重要些。

弗倫純與麥柯爾曾建議用下列公式來度量教師的效率：

$$\text{教師效率} = \frac{\text{（體法成業商數的差別）} + \text{（算術成業商數的差別）} + \dots \dots \dots}{N}$$

在上列公式中，成業商數的差別乃指在學期或學年開始時學生的成業商數與在學期或學年終了時學生的成業商數的差別， N 等於所測量的科目的數目或測驗的數目。

以 \bar{X} 分數代替成業商數，則上列公式可如此寫法：

$$\text{教師效率} = \frac{\text{（末次}\bar{X}\text{—初次}\bar{X}\text{）} + \text{（末次}\bar{X}\text{—初次}\bar{X}\text{）} + \dots \dots \dots}{N}$$

這個公式對於上述1. 2. 兩點均能顧及；雖然不能算十分完善，但是比較現在視察員，督學所採用的方法已經公允客觀多了。

（三）新生的安插及升降的決定 新生的安插，舊生升降的決定，能力分組，特別班的組織等問題都可歸入學生之安插一問題內討論。本節先就新生的安插與升降的決定兩點討論之。

學生的安插有兩條基本的原則：第一，須使學生在插入年級中有達到某種預定標準的可能；第二，須使學生

的程度能力與插入年級的教材的難度相當。

關於決定學生的升降，麥柯爾曾擬定辦法數則，足資參考，茲節錄如下：

1. 任何學生不應留級或退級，除非他的教育程度在留入或退入一級內程度的百分之七十五點以下。
2. 任何學生不應越級，除非他的教育程度在升入一級內程度的百分之二十五點以上。

關於新生安插與舊生升降的決定的根據應着重於標準測驗的結果，但其他事實如教師意見，年齡，健康程度等有時亦應顧及。可用的標準測驗約有數種：

1. 智力測驗 智力與各科學習能力大致呈正相關的關係，所以一人智力的高低大致可以表示他各科學習能力的高低。智力測驗在小學裏尤為適用。

2. 教育測驗 教育測驗可以直接表示學生教育程度的高低，用以決定學生的安插，尤為適宜。麥柯爾相信如有可靠的教育測驗可以應用，則分班時根據教育測驗分數較智力測驗分數尤為相宜。

3. 預測測驗 從理論上講，預測測驗是決定學生安插最完善的根據，因為預測測驗可以直接推測將來的成功。但是這種測驗的發展，尚在幼稚時期，許多學科尚無相當預測測驗的編造。

學生的安插根據於數種測驗的混合分數，當然更為可靠。求混合分數時，除非各種測驗所用的單位相同，不可直接相加。從單位不同的測驗，求混合分數時，一種方法，就是先將由各測驗所得分數，依其差異度的大小，加以

均衡。茲舉例以明之：假設在一個平均數為 50，均方差為 5 的測驗上，甲生得 20 分；在另一種平均數為 120，均方差為 15 的測驗上，甲生得 100 分。現在如將甲生在兩種測驗上所得分數直接相加，得混合分數 120，則第二種測驗在混合分數中所占的重量恰三倍於第一種測驗，因為第二測驗的差異度恰三倍於第一測驗的差異度。為均衡兩測驗在混合分數中所占重量起見，我們必須均衡兩測驗的差異度，其方法為以 3 乘第一種測驗的均方差，或以 3 除第二種測驗的均方差。應用同樣方法於甲生的分數，照第一種手續，甲生的混合分數為 $20 \times 3 + 100 = 220$ ；照第二種手續，甲生的混合分數為 $20 + \frac{100}{3} = 73.33$ 。

表三十九更明顯他表示求均衡分數的方法。

表三十九 表示求混合分數的方法

測 驗	I	II	III	IV	V
平均數	5.60	37.47	35.78	104.71	73.90
均方差	1.76	7.69	4.36	26.79	7.60
均衡乘數	5	1	2	$\frac{1}{3}$	1
新均方差	8.80	7.69	8.72	8.93	7.60
某生分數	5	35	30	100	75
某生的均衡分數 (各測驗重量相等)	25	35	60	34	混合分數 75 = 229

上面我們會說，學生安排的決定的根據，應着重於標準測驗的結果，但其他事實亦不應完全忽略。其中以教師的意見或評斷為最重要。我們都知道教師的評斷是主觀的，不完全可靠的，但同時教師對於學生天天接觸，時時觀察，其意見似有相當的價值。再者，現有的標準測驗尙未臻完善，教師的評斷或可補標準測驗之不足。但如何使教師的評斷與標準測驗的結果有比較及綜合的可能却是一個亟待解決的困難。麥柯爾曾有一極簡便的方法，介紹於下。（參看表四十）

1. 用一種標準測驗（智力測驗教育測驗均可）測量全班學生，求每個學生的H分數。
2. 將學生依其所得H分數之多少，排列起來，H分數最多者列第一。
3. 請教師將學生依程度之高低排列起來，程度最高者列第一。
4. 教師評斷為程度最高的學生，給以一班中最高的H分數；教師評斷為程度列第二的學生，給以一班中次高的H分數，餘類推。這些H分數，可稱之為教學H分數（Pedagogical H Score），簡寫為T_H。
5. 將智力H分數或教育H分數與教學H分數平均，得調級H分數（Promotion H Score），簡寫為H_{pr}。如以為教師的評斷不應與標準測驗的結果相等並重，則在求調級H分數時，可用適當之均衡方法。

表四十 表示求教學H、調級H的方法

學生	智力 T	教師的評斷	教學 T	調級 T
A	25	2	22	22.5
B	22	3	20	21.0
C	20	1	25	22.5
D	18	4	18	18.0

(四) 能力分組 所謂能力分組，就是將同一年級內學生，依其能力之高低，分為數組，使同一班內學生的能力較為整齊，以增進教學的效率。很顯明的，能力分組祇能在人數較多的學校舉行。

依我國學校教育的現狀而論，學級之編製，大半根據學生的年齡及入學的先後，所以一班之中學生能力大相懸殊。在這種情形之下，一方面教師感覺極大的困難，一方面學生亦不能得到充分之裨益。一班之中智力高的學生不能盡量發展和利用他們的智能，漸漸的失了對於學業的興趣而養成了怠惰的習慣；智力低的學生，因為在學業上無論如何努力，結果總歸失敗，亦漸漸的失掉了自信心及好勝心，而對於學業又不知不覺的生了一種厭惡的心理。這種現象不改革，教學的效率是無由提高的。

在前節內我們討論到學生的編入年級應以其智力程度或教育程度為標準，應以智力 T 分數或教育 T 分數為根據。現在討論的能力分組，乃是學生已經依其程度歸入相當年級後，再以其聰明程度或進步速率為標準，

以智力百分數或教育百分數為根據將同一年級內學生歸入相當組別。前者可稱謂豎的分班，後者可稱謂橫的分組。

依照學生能力分入相當組別乃改進教學之第一步，換言之，不過是改進教學的一種必具條件。分組之後，我們還須改換教法，改組教材，使適合兒童之能力，才能收實際的效果。能力分組本身是不能改進教學的。改換教法，改組教材以適應兒童之能力，可有兩種辦法。第一，各組教材相同而升進之速率不等。第二，各組升進的速率相同而教材內容之深淺廣狹則有異。這兩種辦法當然可以兼採並用的。

有許多教育學者，以為要適應各人差異，能力分組仍非最澈底的辦法。他們主張根本推翻班級教學制度，而代以個別教學方法。提倡個別教學最力者要算美國華虛朋氏（C. Washburne）。他的文奈狄卡制（Winnetka System）是名聞教育界的。

（五）特別班的組織 依據兒童智力程度所組織的特別班可有兩種，即高才班與低能班。特別班與能力分組同是適應個別的方法，其不同者，即特別班中的兒童有智力上與通常兒童有更大的差別。應用標準測驗以選擇特別班兒童與應用標準測驗以決定兒童的組別在原則上並無差異，不過對於前者應特別審慎，尤其是關於低能班兒童的選定，所以在可能範圍內特別班兒童的選定應該根據於個人測驗的結果。

（六）教育指導與職業指導的實施 上面所討論關於學生的安插的問題亦可算為教育指導的一種工

作，但教育指導的工作當然不限於此。如選科的指導，升學的指導都在教育指導範圍之內。職業指導的任務在幫助個人獲得對於職業上一個最圓滿的適應，其目的雖與教育指導各異，然二者之間常有密切之關係。

科學的教育指導與職業指導應根據於客觀的事實。實施指導時所根據的事實可有兩種：第一，教育界的情形與職業界的內容；第二，個人的個性，如能力，興趣，性格，體格等。要明瞭教育界的情形與各業的內容當然要從調查研究下手；要明瞭個人的個性，則標準測驗實為最有效的工具。

二 在教學上的應用

測驗在教學上的應用可有下列數端：

(一) 辨別學生的智慧 「因材施教」，「注意個別」，這些都是現代教育上極重要的觀念，因此發現學生的才能，考查學生的個別乃現代教師極重要的任務了。

(二) 考查學生的程度 學校裏的考試普通在學期或學年之終舉行，其實為增進教學的效率起見，尤應在學期或學年之初舉行考試，以考查學生初起的程度。在學習過程的進行中，亦應時時舉行測驗，以示學生進步的狀況。

(三) 診斷學生的困難 測驗在教學上最重要的應用就是診斷困難，以為施行補救教學的根據。在第六

章內我們說教育測驗以功用而論可分為普通測驗與診斷測驗，而二者之中，從教學的立場看來，診斷測驗的效用比較普通測驗大多了。從測量的結果曉得某個兒童的算術程度比較常模為低，這個事實對於教師當然不無相當價值；從診斷測驗的結果曉得某個兒童在加法的某步驟上發生困難，這個事實能切實指示教學的途徑，當然更有價值了。

(四)引起學習之動機 教育心理學中有一條重要的原則，就是學者如能隨時知道自己的成績，則學習的效率可以增加。所以測驗的結果應該使兒童知道，最好能作成圖表，以示進步的狀況並以資比較。練習測驗 (Practice Test) 為教學上極有效的工具，就是這個緣故。

三 在教育研究上的應用

標準測驗差不多已成爲教育上科學的研究所不可缺少的工具。測驗在教育研究上的功用有二：(一)爲等組之用；(二)爲度量試驗因子所產生的效果之用。茲舉例以明之。

我們可以用下述方法——等組法——以比較兩種算術教學法的優劣：

- (一)用一種算術測驗測量一班兒童，以考查其算術程度之高低。如能加用一種智力測驗則尤好。
- (二)根據標準測驗的結果，用配偶法將這班兒童分爲兩個在算術程度及智力上相等的組別。

(三) 用第一種算術教學法於第一組，用第二種算術教學法於第二組。

(四) 經過相當時期，再用標準算術測驗測量兩組的程度。

(五) 應用統計的方法以決定所得差別的是否可靠，以斷定兩種方法的優劣。

標準測驗對於教育中科學的研究的發展實有極大的貢獻。如各科教學法的研究，各科教材的研究，學習過程的研究，性別種別的研究等等無不以標準測驗為研究的利器。

問 題

1. 用六種測驗測量一班兒童，各測驗的差異度（以二十五分差表示之）如下。假定六種測驗在混合分數中所占的分量應相等，試求該生的混合分數。

	算術 測驗	歷史 測驗	地理 測驗	讀法 測驗	數字 測驗	文法 測驗
Q	6	3	5.9	12	6	1
張生分數	20	10	12	15	8	4

2. 求下列各生的教學日分數，測驗日分數。

學生	教育 T 分數	教員評斷的等級	教學 T 分數	關級 T 分數
A	65	3		
B	63	1		
C	61	2		
D	58	5		
E	52	4		
F	49	6		

關級 T = $\frac{2}{3}$ 教育 T + 1 教學 T

3

* 根據上列數據計算各生年終成績

科目	年初的成業商數	年終的成業商數
算術	102	108
讀法	92	100
作文	100	105
作職	105	105

4. 證的分班與橫的分組有什麼分別？
5. 能力分組有什麼利益？能力分組的辦法是否有背於教育機會均等的原則？
6. 科學的分班分組，其目的何在？

參考書報

1. Freeman, F. N.: *Mental Tests*, Ch. XIV Macmillan.
2. Pinnet, R., *Educational Psychology*, Ch. VI Macmillan.
3. Reeb, G. M. and Stoddard, G. D.: *Tests and Measurements in High School Instruction*, Ch. II, III. World Book Co.
4. Synonds, P. M.: *Measurement in Secondary Education*, Ch. XXI, XXII, XXIII, Macmillan.
5. National Society for the Study of Education, 21st Yearbook, Part I, II; 24 th, Yearbook, Part II
6. 上海市教育局：個別智力測驗報告，團體智力測驗報告。

第四編 編造測驗的方法

第二十章 編造測驗的普通步驟

編造測驗的步驟依測驗的性質而異，難有一致的規定。智力測驗的編造在許多方面與教育測驗的編造不同，編造小學各科測驗的規則亦許不盡適用於中學測驗的編造。他如職業測驗，預測測驗，性格測驗的編造，都有各自的特點。在本章內僅就編造測驗最普通的步驟申述之。

一 測驗材料的選擇

一個標準測驗最基本的條件就是正確性。一個測驗的正確度與測驗材料的選擇有密切的關係。所以編造測驗的第一步工作就是根據確定的標準，審慎選擇測驗的材料。選擇材料所依據的標準可有下列數種：

(一) 以實際的用途為標準 這亦是現代課程編製最重要原則之一。某種智識或技能，如在實際生活中極少有應用的機會，則不應採用為標準測驗的材料。反過來說，某種智識或技能，在實際生活中應用次數愈多，則

其價值愈大，或愈重要，編造測驗時應採用之。如桑戴克及霍倫的英文通用字彙的統計，陳鶴琴氏的中文語體文通用字彙的調查，威爾遜 (G. M. Wilson) 的商人應用算術的研究，都為編造測驗極好的依據。

(二) 錯誤的分析 [美人瓊斯曾統計小學學生最易拼錯的字，包雅頓亦有小學作文錯字統計與分析的研究。這種統計與研究的結果可用為編造拼字測驗或默字測驗的材料。美人威爾遜的文法測驗亦係根據於本人對於學生所犯文法上錯誤的研究而編造的。

(三) 課本的分析 根據通用教科書內容的分析，以選擇測驗的材料，亦為編造測驗者常用的方法。這個標準的限制在祇能代表當時實際上所採用的教材而不能代表理想的教材，祇能保守，廣續當時教學的情形而不能有所改進。

(四) 教材的分析 這個標準與第三項大同小異，其利弊亦同前。

(五) 考試題目的分析 考試題目所包括的事實可以代表在教員心目中最重要的教材。

(六) 專家意見的徵集 上述第三、第四、第五、三項選擇材料的方法實際上都可稱為專家意見的徵集。以往的經驗告訴我們，如其他情形相等，多數人的意見較個人的意見為可靠，所以在徵集專家意見時，人數應較多，至少應有三人。同時，選擇專家時標準應嚴格些，寧缺無濫。

(七) 與學業成績的相關 教員所給分數的不可靠，是不容否認的事實，但是大致上總可以表示學生真

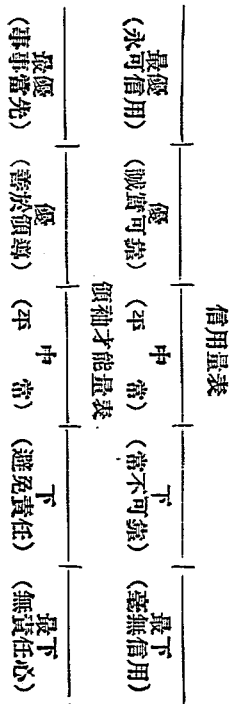
正程度的高低。一種教育測驗的結果與學業的成績之間假設呈零度相關（或相關極低）或負相關，我們對於這個測驗的正確性就應該發生深切的懷疑了。

（八）與客觀標準比較 考查測驗的正確度最直接的方法就是求測驗結果與客觀標準的相關度。在第三章中我們曾以售貨能力測驗為例，說明採用客觀的標準以考查測驗的正確性的方法。

（九）與主觀評判比較 上節內所述客觀標準往往不易求得。在這種情形之下，祇能採用主觀評判為標準了。例如欲考查社交能力測驗的正確性，社交能力的客觀標準不易求得，我們祇能求一班兒童在測驗上所得分數與教師對於他們的社交能力的評判的相關度，以決定該測驗是否正確。

徵集主觀評判時可以採用評判量表。各種評判量表中，以圖示評判法為最便利準確，舉例如下：

說明：在下列各綫相當地位上做一×符號，以表示被評判者的程度。



主觀評判的不可靠乃是心理學者所熟知的事實，所以應用評判法時須極力避免可能的錯誤。評判時最易犯的錯誤有下列數端：

1. 人們的評判有嚴寬的不同。這就是說，各人在評判某性能時在心中的懸的標準高低不同。這種極性的錯誤常使各評判員的評判不能互相比較，綜合應用評判法者必須注意及之。

2. 評判中一種最普遍的錯誤，就是所謂「回音的影響」。所謂回音的影響就是說，評判者對於某人某種特殊性能的評判無形中受他對於這個人的普通好惡的影響。回音的影響的作用使評判的結果不準確。所以應用評判法者必須設法將這種錯誤減少至最低限度。

避免或減少回音的影響可有數種方法：(1) 向評判者指明回音的影響的作用，使其注意。(2) 如要評判一羣人在數種性能上的地位，較好的手續是先評判全體被試者在某一種性能上的程度，再評判他們在另一種性能上的程度，再將其他性能一一加以評判。如先評判某一人在各種性能上的程度，再評判另一人在各種性能上的程度，這種手續就比較容易發生回音的影響。

3. 評判可靠不一定正確。三位評判員評判一班學生的某種性能，假使他們誤用了同一的錯誤的標準，則他們的評判相互間或可有很高的相關（表示可靠），但是同時他們的評判仍然是完全錯誤的。

(十) 與已有測驗比較 將新編測驗的結果與心理學者所公認為正確的測驗的結果相比較，亦為考查

測驗的正確性的一種方法。例如美國的斯丹福大學訂正比納西門智力量表常被採用為考查新編智力測驗的正確性的標準。

(十一) 各年級或各年齡兒童做對某測驗題目人數的百分比，是否依次增加這是一個考查測驗題目的正確性的最有效的標準，牠的用途祇有兩種限制：(一) 有對一種性能並不依年齡而增進，則上述標準當然失去其效用；(二) 有的學科並不是每年級都有的，則上述標準亦不能採用了。

根據這標準以選擇測驗題目的手續大概如下：先用測驗測量許多年齡不一（或年級不同）的兒童，然後計算各年齡（或年級）做對每個測驗題目的人數的百分比，有如表四十一所示者：

表四十一 各年級做對各題目的人數的百分比

題目	做對的人數的百分比									
	年級	3	4	5	6	7	8	9	10	
1	0	1	15	28	48	67	86	98		
2	0	1	1	26	77	100	100	100		
3	0	10	38	58	40	11	24	12		
4	0	0	0	1	20	80	98	100		
5	20	68	95	99	100	100	100	100	100	

做對題目一的人數的百分比，從四年級起到十年級為止逐漸增加。這個題目最適合於七年級，因為七年級兒童做對該題的人數差不多為百分之五十。這個题目的鑒別力並不大，因為曲線的斜度是很平坦的。這個題目適用的年級的範圍甚廣而且做對人數逐級增加，所以仍不失為正確的題目。

題目二比較題目一難些。在六年級以下，八年級以上，這個題目是不適用的。但是在六年級至八年級範圍之內，這個題目有很大的鑒別力。

做對題目三的人數並不依年級而增加。這類題目應該刪去。

題目四與題目二相似，不過比較難些。

題目五很適用於低年級。

依照上述手續評定各题目的價值並適用的年級範圍，以定取捨。

(十二) 將測驗應用於數個能力不等的團體，以考查整個測驗或各個题目的正確度。這個標準與第十項標準是相仿的。譬如要考查一個木匠測驗的正確度，我們先用該測驗測量許多木匠，然後將受測驗者依其資格或現在地位分為師父、徒弟兩組，再計算每組中答對每個题目人數的百分比或每組在整個測驗上所得之平均分數。

這個標準在編造職業測驗，性格測驗時都極有用。

(十三) 心理的分析 如西旭所編音樂才能測驗係根據他個人對於音樂才能的心理的分析而編製的。

二 問題的編製

編製測驗第二步工作就是根據選定的材料編製測驗題目。關於問題的編製，下列數點應加注意。

(一) 問題方式須決定 標準測驗中題目方式的種類大致如第十三章內所列，茲不重述。編造測驗者須參照測驗的可能的用途，材料的性質，以決定採用何種或何數種方式。

(二) 問題的數目至少須比最後所須要的數目加一倍，以備日後的刪除淘汰。

(三) 問題須能包括或代表材料的全部。

(四) 問題的難度須適合於被試者的程度。

(五) 編製各式題目的規則已詳第十三章，茲不重述。

(六) 文字須清楚明白，不引起誤解。

(七) 測驗之格式的規定須審慎從事。廖世承陳鶴琴會規定原則五條，引錄如下：

「1. 所選定的格式必須恰當，使被試者容易明瞭測驗的做法。

2. 格式必須選擇妥當，務使被試者在做測驗的時候，不會弄錯。

3. 格式必須要選得妥當，務使做法簡明省時。
4. 格式須選得妥當，務使校閱時省時省力。
5. 格式要選得妥當，務使測驗所佔的地位很經濟。
- (八) 問題的難度可依主觀的評判大致決定之，並依其難度而排列。

三 初次的試用

測驗的適用與否還須經過實際的試用，才能決定。下列數則可以表示普通的手續：

(一) 用測驗測量各年級兒童，每年級約一百人至二百人。所測量的兒童須能公允地代表各該級的程度。每級中程度最低者與程度最高者都應包括在內。

(二) 在初次試用時，時間限制應寬些，最好使每個兒童都有機會嘗試每個題目。

(三) 有時因時間與經費的限制，不能每年級逐一加以測量，可間級而行之，例如七、九、十一年級或六、十年級。

(四) 如因測驗太長分為數份試用時，則須注意各份應用的次序，以均等練習的影響。假如有三份測驗須試用，則應用的次序應如下列：

1. 每年級三分之一的學生做這三份測驗的次序爲甲乙丙。
2. 每年級三分之一的學生做這三份測驗的次序爲乙丙甲。
3. 每年級三分之一的學生做這三份測驗的次序爲丙甲乙。

四 根據試用的結果加以刪修

根據試用的結果，決定測驗題目的保留或刪除。下列數條表示普通的手續：

- (一) 凡意義含混，易致誤解的題目修正之或刪去之。
- (二) 計算各年級做對各題目的人數的百分比。根據前述標準(十一)刪除不良的題目。
- (三) 將做對各題目人數的百分比根據表四十二化爲標準差或 σ ，以表示其難度。按照標準差或 σ 的價值，重行排列各題目。

(四) 最後保留的題目的數目須以希望所達到的可靠度爲斷。若他種情形相等，測驗愈長，即題目愈多，則測驗愈可靠。

表四十二 σ 價值與百分比對照表

總 標 票 簿

11. 米 四

T	百分比	T	百分比	T	百分比	T	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	99.9899	38	88.49	63	9.68	88	0.007
13.5	99.9887	38.5	87.49	63.5	8.85	88.5	0.0059
14	99.9874	39	86.43	64	8.08	89	0.0048
14.5	99.9861	39.5	85.31	64.5	7.35	89.5	0.0039
15	99.9777	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

英國 國際標準時表

11 表附

級	距	照	等	11米水	0.00011		
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	0.0.05
19.5	99.886	44.5	70.88	69.9	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

五 將測驗分爲數個相等的覆份

在第三章中，我們會說，為增加效用起見，一種測驗至少須有兩個覆份。將最後所決定採用的題目分為兩個或兩個以上的覆份，手續很簡單。假如題目 1, 2, 3, 4, …… 依難度而排列，欲將其分為兩個覆份，可採用下列方式：

甲份： 1 4 5 8 9 12……
 乙份： 2 3 6 7 10 11……

分為三個覆份，可用下列方式：

甲份： 1 6 7 12……
 乙份： 2 5 8 11……
 丙份： 3 4 9 10……

依照上述手續所得之兩個或三個覆份，其平均與均方差可大致相等。關於覆份所需適合的條件，已列第三章，不再申述。

將題目分為兩份或三份以後，應該再試用一次，以決定各份究竟是否相等。在第二次試用時，測量的人數不妨多些，所得結果可用以求該測驗之常模。

六 決定時間的限度

測驗所須的時間限度須用嘗試法（通常可以在第二次試用時）以決定之。在申述決定測驗時間限度的手續以前，對於決定時間限度的原則似乎應該先討論一下。

測驗可大致分為速度測驗與難度測驗，速度測驗測量作業之快慢，難度測驗測量能力之高低或程度之深淺。速度測驗的時間限制的規定當然須嚴格，務使作業最快者在規定時間內亦不能將測驗做完。難度測驗在理論上不應有任何時間限制。在實際上，為避免因少數遲慢學生而過分延長測驗時間起見，仍不得不規定相當的時間限制。但限制應較寬，應使百分之九十或百分之九十五的學生能在規定時間嘗試所有的題目。

決定時間限制可用下述手續。在發給試卷時，同時發給每個兒童三種顏色不同的鉛筆，如黑、紅、藍。從第一次的試用，編造測驗者大致可以曉得這測驗所需要的時間。假設根據第一次試用的經驗，我們估計百分之九十的兒童大概可以在二十五分鐘內做完。我們可以先叫兒童用黑鉛筆做二十分鐘，然後發令叫兒童換紅鉛筆，再過五分鐘，發令叫他們換藍鉛筆。試卷收齊之後再細細考查各生作業的速率，以規定最後之時間限制。

另一種方法是由主試者將過去時間每半分鐘一次寫在黑板上，每一個兒童做完時即將當時黑板上的時間抄在試卷上。試卷收齊之後，將各兒童所需時間列一分配，以作最後之決定。

七 求常模

覆份編好，時間規定之後，就可以求常模了。在這個時候測驗所應用的年級或年齡範圍，當然已經決定了。採用那一種常模，却是尚待決定的問題。可靠的常模必須根據於有代表性的兒童，在第三章內已詳言之。

假設測驗有好幾個覆份，在求常模時，應用的次序必須倒換，以免練習的影響。譬如有兩個覆份，則一半的學生受測驗的次序為覆份一，然後覆份二；另一半學生受測驗的次序應為覆份二，然後覆份一。

常模共有四種：(一)年級常模，(二)年齡常模，(三)百分數，(四)日分數。依次說明之：

(一)年級常模 這是一種以往最通用的常模，但是現在年齡常模已漸漸取年級常模而代之。年級常模在低年級裏比較有用，在中學裏效用較小。

年級常模很容易解釋，很容易求得，但是同時有許多缺點：

1. 各年級學生的平均年齡各校不同；換句話說，「速進」、「在級」及「遲滯」的學生的百分比，各校不同。
2. 年級常模在中學裏並無多大用處。
3. 因為大多數智力測驗是用年齡單位或其他單位的，所以年級常模不能直接與智力測驗結果比較。
4. 如不顧到學生年齡與智力，年級常模很容易引起錯誤的解釋。

麥柯爾的 Ω 分數，就是一種年級常模，不過他將每年級分爲十段，如 4.1 代表四年級上課一個月後的程度，4.5 代表四年級上課五個月後的程度等等。表面上看起來， Ω 分數當然較普通年級常模更爲精確，但是實質上亦未必如此。我們現在的測驗還不能十分可靠，一個測驗分數的均方錯誤或機誤很大，兩個年級間的差別的十分之一恐怕不是現有測驗所能測量出來的。

(二) 年齡常模 年齡常模就是各年齡兒童的平均分數或中數。年齡常模的求法可舉例以明之。各年齡兒童在司丹福學業測驗上的平均分數如下：

年 齡	9	10	11	12	13	14
平均分數	23	33	46	57	66	72

從上表，33 分就是九歲六個月的常模（因爲所謂九歲包括已達九歲而尙未達十歲的兒童，其平均年齡當爲九歲半），66 分就是十歲半的常模，餘類推。兩個年齡間的常模，如九歲七個月，九歲八個月的常模，可用推算法求得之。

年齡常模的優點列舉如下：

1. 年齡單位，不受學級編製的影響。

2. 年齡常模可與智力測驗結果（智力年齡）直接比較。
 3. 年齡標準可以顯示不良的年級編製，如遲滯的學生太多等。
 4. 年齡常模可與實足年齡比較。這種比較為解釋測驗分數必需的步驟。
- 年齡常模的限制列舉如下：

1. 可靠的公允的年齡常模不易求得。求八歲以下，十四歲以上各年齡的常模，難免選擇的影響。
2. 以年齡單位表示學業程度不如年級單位來得自然，習便。
3. 年齡常模在中學裏無多大用處。

（三）百分數 百分數可以表示一人在某一個團體中所居的地位。百分數的求法很簡單，與求中數及上下二十五分點的方法相同。實際上，中數即百分之五十點，上二十五分點即百分之七十五點，下二十五分點即百分之二十五點。求百分點的公式如下：

$$P_p = \frac{PN}{100} \cdot \frac{F}{f} + 1$$

在此公式內：

P_p = 某百分點

第四編 編造測驗的方法

P = 在某百分點以下的次數的百分比

U₁ = 某百分點所在組的下限度

F = 某百分點所在組以下各組次數之和

f = 某百分點所在組的次數

i = 組距

表四十三表示百分數的求法

表四十三 表示百分數的求法(材料同表十七)

分 數	次 數	累積次數	百 分 點	分 數
200—205	1	54	100	201
195—200	4	53	90	194
190—195	2	49	80	188
185—190	10	47	70	185
180—185	3	37	60	179
175—180	8	34	50	175
170—175	3	26	40	167

165-170	3	23	30	180
160-165	4	20	20	156
155-160	6	16	10	147
150-155	4	10	0	126
145-150	1	6		
140-145	1	5		
135-140	2	4		
130-135	0	2		
125-130	2	2		
N=54				

$$P_{10} = 145 + \frac{5.4 - 5}{1} \times 5 = 147$$

$$P_{20} = 155 + \frac{10.8 - 10}{6} \times 5 = 155.67$$

$$P_{30} = 160 + \frac{16.2 - 16}{4} \times 5 = 160.25$$

$$P_{40} = 165 + \frac{21.6 - 20}{3} \times 5 = 167.67$$

$$P_{80} = 175 + \frac{27}{8} - \frac{26}{8} \times 5 = 176.625$$

$$P_{60} = 175 + \frac{32.4 - 26}{8} \times 5 = 179$$

$$P_{40} = 185 + \frac{37.8 - 37}{10} \times 5 = 185.4$$

$$P_{20} = 185 + \frac{43.2 - 37}{10} \times 5 = 188.1$$

$$P_{00} = 190 + \frac{48.6 - 47}{2} \times 5 = 194$$

P。表 P_{100} 的求法似應加以說明。P。為 P_{100} 就是分配中最低與最高的兩個分數。在表十七原來的分數中最低分數為 126，最高分數為 201，所以 P。為 126， P_{100} 為 201。

表四十三中百分點的解釋如下。

全班中有百分之十的學生分數在 194 以上，有百分之九十的學生分數在 194 以下。

全班中有百分之二十的學生分數在 188 以上，有百分之八十的學生分數在 188 以下。

全班中有百分之三十的學生分數在 185 以上，有百分之七十的學生分數在 185 以下。

全班中有百分之九十的學生分數在 147 以上，有百分之十的學生分數在 147 以下。

百分數常模的優點爲：1. 易於解釋，2. 在中學裏，年級常模與年齡常模不很適用，可採用百分數常模。

百分數常模的限制爲：1. 意義不若年級常模，年齡常模之通俗易解；2. 不能與智力測驗結果直接比較；3. 百分數的可靠度比較平均數或百分數爲低。

(四) 百分數 百分數的意義，第十八章已詳述之。百分數爲麥柯爾所命名，但百分數所根據的觀念却早已成立了，如奧偉士、麥戴克與克萊都曾在麥柯爾之先應用過與百分數相似的分數。

百分數均方差的十分之一，較百分數爲可靠。百分數的公式已見第十八章。爲便於討論起見，重列於下：

$$P = 50 + \frac{10(X - M)}{s}$$

在此公式內：

X 代表任何原來分數；

M 代表計算百分數所根據的團體的平均分數；

s 代表計算百分數所根據的團體的均方差；

以10乘(X - M)，爲的是免除小數；

公式中的50，可使平均百分數恰等於50。

百分數的意義可用一個實例以顯示之。將原來分數化作百分數之先，必需求分配的平均數與均方差。表四

十四表示從原來分數化成百分數的步驟。

計算百分數根據任何分配都可以，但為便於比較起見，最好定一個標準團體。小學各科測驗都以十二歲兒童為標準團體。

表四十四 表示將原來分數化成百分數的方法

$$M = 37.19$$

$$Q = 12.24$$

原來分數	$X - M$	$\frac{10(X - M)}{Q}$	$50 + \frac{10(X - M)}{Q}$
80	42.81	35	85
75	37.81	31	81
70	32.81	27	77
65	27.81	23	73
60	22.81	19	69
55	17.81	15	65
50	12.81	10	60
45	7.81	6	56

40	2.81	2	52
37	- 0.19	0	50
35	- 2.19	- 2	48
30	- 7.19	- 6	44
25	- 12.19	- 10	40
20	- 17.19	- 14	36
15	- 22.19	- 18	32
10	- 27.19	- 22	28
5	- 32.19	- 26	24
0	- 37.19	- 30	20

百分數的優點爲：1. 可靠，2. 在中學裏可以應用，3. 如智力測驗亦採百分數制，則可與智力測驗結果直接比較。其限制爲：1. 計算繁複，不易了解；2. 與用年齡量表的智力測驗結果不能直接比較。

八 決定測驗的可靠度

可靠度的意義已詳第十六章，不重述。表示測驗的可靠度的方法有下列數種：

(一) 可靠係數 計算一個測驗的可靠係數，至少有三種方法。

1. 將同一份測驗測量同一羣兒童兩次，兩次測驗相隔期間，不應太短或太久，太短則記憶的影響不能免除，太久則自然發育的影響又發生。同一羣兒童在同一測驗上兩次所得分數的相關度，就表示這測驗的可靠度。

這個方法最不妥當，因為記憶的影響與自然發育的影響很不容易準確地度量而加以控制。除非是單份的測驗，如比納西門智力量表，求測驗的可靠係數時不應用此法。

2. 將測驗分為兩半（普通的手續是將偶數的題目歸為一份，奇數的題目又歸為一份），求兩半之間的相關度，然後應用斯畢門白朗翁 (Spearman) 推測公式以求整個測驗的可靠度。

3. 假如一種測驗有幾個複份，那麼祇要求同一羣兒童在兩個複份上所得分數的相關度，就行了。這個方法最為簡便。

(二) 一個分數的可靠度 上面所討論的可靠係數不是一個表示測驗可靠度最完善的數量，理由如下：
1. 可靠係數祇表示許多對分數的大概關係，至於某一個兒童所得分數的可靠度（這實在是是最重要的問題）却不是可靠係數所能回答的了。

2. 如缺乏其他的資料，可靠係數本身不能告訴我們一個測驗到底如何可靠（或不可靠），因為可靠係數依所測量的團體的差異度而變易。所測量的團體的差異度愈大，則所得可靠係數愈高；差異度愈小，則可靠係數愈低。

假設我們有一種算術測驗，應用於小學第一年級至第四年級兒童，得可靠係數·8，我們却不能說將此測驗應用於第三年級一個年級中，亦能得到同樣的可靠係數，因為這兩個團體的差異度不是相等的。

克萊有一公式，表示可靠係數的高低與差異度的大小的關係，可以使我們從一種測驗在某種團體內的可靠係數，推測同一測驗在他種團體內的可靠係數。其公式如下：

$$\frac{R}{M} = \frac{\sqrt{1-R}}{\sqrt{1-R}}$$

在此公式內：

R = “小”的團體的均方差

M = “大”的團體的均方差

r = 測驗應用於“小”的團體所得的可靠係數

R = 測驗應用於“大”的團體所得的可靠係數

上例公式的應用可舉例以說明之。

主試甲報告某測驗的可靠度為·85，主試乙報告同一測驗的可靠度為·64，這兩個調查所得結果如下：

主試者	可靠係數 r	人數 N	均方差 δ	測量的團體
甲	.95	500	30.3	四年級到十二年級各年級
乙	.64	500	10.1	六年級

將表中事實代入上列公式，求 R 。

$$\frac{\delta \sqrt{1-R}}{M \sqrt{1-r}}; \frac{10.1}{30.3} = \frac{\sqrt{1-R}}{\sqrt{1-.64}}; R = .96$$

表面上看來，主試甲與主試乙所得結果極不一致，但要是我們顧到甲乙所測量的團體的差異度，則甲乙二人所得結果實極相近。

同一測驗在差異度不同的團體中所表示的可靠度迥不相等。所以在報告一種測驗的可靠係數時，必需同時報告這可靠係數所根據的團體的差異度，不然別人就沒有法子去解釋這可靠係數。

因為上面所述兩種理由，所以可靠係數不是一個表示測驗可靠度的最完善的數量。比較有用的一種量數就是一個分數的機誤。一個分數的機誤的公式與解釋已詳第十六章，此處不重述了。

九 說明書的編訂

編製測驗的最後一步，就是編訂說明書。說明書的內容應包括下列各項：

- (一) 本測驗的目的，功用。
- (二) 測驗的正確度，測驗中的材料是根據什麼原則，應用什麼方法選擇得來的？
- (三) 關於如何施行測驗的說明。
- (四) 關於如何校閱試卷的說明，標準答案，或記分標準的規定。
- (五) 常模表，關於如何應用常模以解釋結果的說明。
- (六) 測驗的可靠度。
- (七) 關於如何應用測驗結果的指示。

問 題

1. 根據通用教本內容的分析以選定測驗的材料，有什麼不妥當之處？
2. 何謂回音的影響？如何避免之？
3. 評判可靠不一定正確，何故？
4. 比較年齡，年級，百分數，四種常模的優劣利弊。

5. 可靠係數何以不是表示測驗可靠度最完善的量數?

6. 求下列各題目的信度並將題目分為三個相等的層次。

題 目	做對人數 的百分比	題 目	做對人數 的百分比	題 目	做對人數 的百分比
1	95	11	45	21	58
2	34	12	75	22	52
3	10	13	38	23	56
4	85	14	45	24	82
5	70	15	28	25	19
6	96	16	20	26	16
7	18	17	55	27	34
8	15	18	58	28	46
9	50	19	25	29	85
10	48	20	22	30	78

7. 用一種測驗測量一班學生, 算術平均數為 40, 均方差為 10, 試將下列各生分數化為百分數。

學生	分數
A	40
B	30
C	54
D	28
E	52
F	42
G	31

參考文獻

1. Freeman, F. N. Mental Tests, Ch. 9, 10.
2. Hull, C. L. Aptitude Testing, Part II, Ch. 9, 14.
3. McCull, W. A. How to Measure in Education, Part II Macmillan.
4. Monroe, W. S. An Introduction to the Theory of Educational Measurements, Ch. 4-7. Houghton Mifflin.
5. Ruch, G. M. The New-Type or Objective Examination, Ch. 7. Scott, Foresman.
6. Ruch, G. M. and Stodank, G. D. Tests and Measurements in High School Instruction, Ch. 17-20 World Book Co.
7. 吳中丞 測驗法(第二十一章)(附錄)
8. 吳中丞 教育測驗編法之理論與實際(教育雜誌二十一卷十一、十二號)
9. 吳中丞 論中小學教育測驗之理論與方法(中華教育界第二十卷第八、九、十期)
10. 陳運善 職業性向測驗的編製法(教育與職業第一四四期至一四八期)

第二十一章 T.B.C 量表的編製

一 H 量表的編製

H 分數的意義已詳第十八章，現在我們討論編製 H 量表的方法。

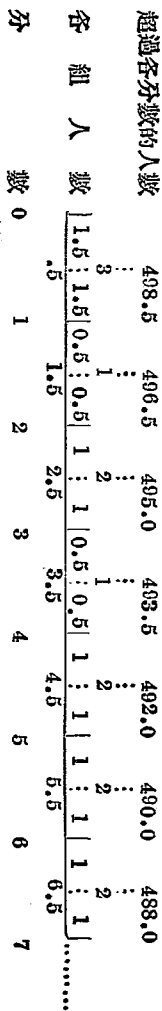
(一) H 分數普通是根據十二歲兒童的編製 H 量表的第一步驟，就是用編就的測驗去測量許多十二歲兒童（十二歲與十三歲之間。）所測驗的人數愈多愈好，取樣須要隨機，不可受選擇作用的影響。

(二) 第二步驟就是計算各人所得分數並將分數列成次數分配，如表四十五第一、二兩行所示者。

(三) 第三步驟就是求表四十五中第三行，即「超過數加」達到數，「說得詳細一點，就是「超過某一個分數的人數加達到該分數人數的一半。」例如表四十五中並無一人超過 $\frac{30}{100}$ 分，得 $\frac{30}{100}$ 分者祇一人，所以超過 $\frac{30}{100}$ 分的人數如達到 $\frac{30}{100}$ 分的人數的一半為在 $0 + \frac{1}{2} = \frac{1}{2}$ ，超過 $\frac{30}{100}$ 分者有一人，得 $\frac{30}{100}$ 分者亦一人，所以超過數加達到數為 $1 + \frac{1}{2} = 1.5$ 。表四十五第三行中其他各數用同樣手續求得之。

「超過數加達到數的一半」的意義似乎須加以解釋。「超過數加」達到數「實際上就等於「超過數。」原來普通測驗上的分數是以完全做對的題數為單位的，完全做對幾個題目，就得幾分，做對題目的一部分是沒

有分數的。假如甲在某測驗上恰恰能做對三個題目，他的實在程度為 $\frac{3}{5}$ ；乙能做對三個題目並第四個題目的一半，他的實在程度應為 $\frac{3.5}{5}$ ；丙能做對三個題目並第四個题目的十分之九，他的實在程度應為 $\frac{3.9}{5}$ 。但是照普通計分的方法，這三個人都得三分。所以照普通計分的方法，「分實在等於 1——1.9999……」；「分實在等於 2——2.9999……」；餘類推。所以要寫得詳細些或準確些，表四十五第一行 0, 1, 2……各數應寫為 0——.9999, 1——1.9999, 2——2.9999……而各組的價值可用各該組的中點代表之，即 .5, 1.5, 2.5……所以求超過 0 分的人數，實際上就是求超過 .5 分的人數。表四十五中有 497 人顯然的在 .5 分以上，有 3 人在 0——.9999 之間，而這 3 人中有幾個人在 .5 分之上却無從決定。最公平的假設就是 3 人中有一半，即 1.5 人，在 .5 分之上，一半在 .5 分之下。所以超過 0 分（實際上是 .5 分）的人數為 $497 + 3 \times \frac{1}{2}$ ，共 498.5 人。上述情形用圖表示之如下：



(四) 第四步就是求表四十五第四行，即「超過數加 1 達到數的百分比」將總人數 500 除第三行各

數再乘 100 即得。例如 $\frac{498.5}{500} \times 100 = 99.7$

表四十五 表示求正確率的方法

做對題數	十二歲兒童數	超過十分數	超過數的百分比	T
0	3	498.5	99.7	23
1	1	496.5	99.3	25
2	2	495.0	99.0	27
3	1	493.5	98.7	28
4	2	492.0	98.4	29
5	2	490.0	98.0	29
6	2	488.0	97.6	30
7	2	486.0	97.2	31
8	4	483.0	96.6	32
9	2	480.0	96.0	32
10	2	478.0	95.6	33
11	10	472.0	94.4	34
12	3	465.5	93.1	35
13	8	480.0	92.0	36

14	8	452.0	90.4	37
15	13	441.5	88.3	38
16	15	427.5	85.5	39
17	18	411.0	82.2	41
18	28	388.0	77.6	42
19	26	361.0	72.2	44
20	34	331.0	66.2	46
21	40	294.0	58.8	48
22	40	254.0	50.8	50
23	41	213.5	42.7	52
24	37	174.5	34.9	54
25	31	140.5	28.1	53
26	35	107.5	21.5	58
27	24	78.0	15.6	60
28	26	53.0	10.6	62
29	21	29.5	5.9	66
30	14	12.0	2.4	70
31	3	3.5	0.7	75

32	1	1.5	0.3	78
33	1	0.5	0.1	81
34	0			85
35	0			90

(五)第五步就是將各百分比根據表四十二化為H分數。例如查表四十二，與百分之99.7相對的H分數為22.5，與百分之99.3(表中與99.3最相近的為99.29)相對的T分數為25.5，所以凡在測驗上得0分的兒童其應得之H分數為22.5或23，得1分的兒童，其應得之H分數為25.5或26，餘類推。

根據十二歲兒童編製的H量表，其距離甚大，足敷應用，但有時仍有擴充之必要。要增加H量表的距離，可照表四十五所示的方法求九歲兒童與十六歲兒童之H量表(或其他年齡之H量表)，然後與H₁₂量表合併，如表四十六所顯示者。

表四十六 表示擴充H量表之距離的方法

做題數	T ₉	T ₁₂	T ₁₆	擴充後之H量表
0	32			22
1	36			26
2	40			30

3	43	33		33
4	46	35		35
5	48	38		38
6	50	40		40
7	52	43		43
8	54	45	34	45
9	58	48	37	48
10	61	50	40	50
11	65	53	42	53
12	70	56	45	56
13		59	47	59
14		63	50	63
15		67	53	67
16		71	56	71
17		75	60	75
18		80	65	80
19			70	85
20			76	91

第四編 稻造試驗的方法

二 量表的編製

各年齡的量表就是各該年齡的量表，所以為計算分數起見，必須先求各年齡的量表。設有一兒童，年齡十歲半，他在量表上的分數就是他的分數；他在量表上的分數就是他的分數。假設他的年齡是十三歲半，則他在量表上的分數就是他的分數；他在量表上的分數，就是他的分數。假設他的年齡是十二歲半，則他在量表上的分數就是他的分數，同是亦是他的分數。

編製各年齡的量表的手續，與前節所述編製量表的手續完全相同。但是這裏發生一種很大的困難，就是對於年齡在十二歲以下或以上的兒童很難得到公允的取樣。十三歲、十四歲、十五歲兒童中較聰明者已進了中學或已離開學校。七歲、八歲、九歲兒童中較愚笨者尚未進小學。這些兒童都難測量到，因此所得結果都不能公允地代表各該年齡的程度。所以編製各年齡的量表，在實際上是很困難的。

麥柯爾曾提出一種簡便可行的方法，詳細陳述於后：

- (一) 用編就測驗去測量各年齡兒童，人數愈多愈好，取樣愈隨機愈好。
- (二) 作一各年齡分數的次數分配表，如表四十七，並將各分數的量表價值寫在表內。
- (三) 求各年齡的總人數。
- (四) 求各年齡總人數的半數，例如七歲兒童的總數為30，半數為15。

(五) 從各次數行的下端向上將次數遞加，直至超過半數的一個次數為止，然後取此次數的一半，加到已得和數上，結果就是「超過與該次數相對的H分數的人數加達到該H分數人數的一半」。例如七歲兒童總人數的半數為175。從七歲的次數分配的下端向上遞加，得 $1+0+3+1+2+0+2+1+4+2=17$ ，若再加上上面的一個次數，就要超過總人數的半數了。將該次數的一半 $2.5=2$ ，加在17之上，得19。查看表四十七，與該次數相對的原來分數為11，分數為34。這就是說，七歲兒童中超過H分數的人數加達到H分數的人數的一半共為19。

(六) 將這「超過數加達到數的一半」用十二歲兒童的總人數除之，求其百分比。例如表四十七中十二歲兒童的總人數為500， $19 \div 500 = 3.8\%$ 。

(七) 從表四十二查出這百分比相對的H分數，得38。這就是說，若所有七歲兒童都測驗到了，並且七歲兒童的H量表亦編造了，那末得原來分數11者，H分數大概是38。

上面(六)(七)兩段所述的方法，基於兩個假說。第一，七歲兒童中有一部分未測量到。假如儘量設法去測量，七歲兒童的人數應與十二歲兒童的人數大致相等。第二，這部分未測量到的兒童是七歲兒童中之較愚笨者。這部分兒童的程度部在已測量到的38個兒童的平均程度以下。這個假設大致是合理的，並與實際相符的。對於八歲，九歲，十歲，十一歲兒童，這個假設更可靠了。

(八) 查表四十七，原來分數11的 T_2 價值為34 T_1 為38 $T_1 - T_2 = 11 + 34$ ，這就是七歲兒童的B校正數。例

如有一個七歲兒童，在測驗上做對了九個題目，他的 B 分數，依照表四十七，為 39。他的 D 分數為 B 分數加 B 校正數， $39 + 34 = 73$ 。因此我們知道這個兒童的程度在同年齡兒童的平均程度之上，依據表四十二，七歲兒童中祇有 5.48% 程度在他之上。

(九) 照上述各步驟，求八歲、九歲、十歲、十一歲的 B 校正數，再以十一歲為例，算法如下：(a) $426 \div 2 = 213$
 (b) $1 + 0 + 6 + 4 + 3 + 13 + 16 + 16 + 22 + 29 + 32 + 40 + (35 + 2) = 199.5$ (c) $199.5 \div 500 = 39.9\%$
 (d) $89.9\% = 52.5 T_{11}$ (e) $52.5 - 48 = 4.5$ 即 B 校正數。十二歲的 B 校正數當然是零。

(十) 十二歲以上各年齡之 B 校正數的求法，與上面所述略有不同。不同之點，就在對於十二歲以上各年齡，我們應顧及表四十七中所遺漏的乃較聰明的兒童。要顧到這一點，可先決定被遺漏的兒童人數，再將該數目加在「超過數加一半達到數」上。以十三歲為例，兒童人數為 452， $500 - 452 = 48$ ，即十三歲兒童中被遺漏的人數。請注意下列算式中如何應用 48 這個數目。(a) $452 \div 2 = 226$ (b) $2 + 1 + 5 + 11 + 19 + 23 + 24 + 39 + 46 + 42 + (24 + 2) = 235$ (c) $235 + 48 = 283$ (d) $283 \div 500 = 56.6\%$ (e) $56.6\% = 48.5 T_{13}$ (f) $48.5 - 52 = -3.5$ 即 B 校正數。

(十一) 用同樣方法，求十四、十五、十六、十七各年齡的 B 校正數。十六及十七兩年齡的 B 校正數，因為人數太少，當然不很可靠的。

(十二) 最後一步是求八歲與九歲、九歲與十歲等之間年齡的B校正數，可用推算法求得之。假使七歲半或九十個月的B校正數為 35.8 ，八歲半或一百零二個月的B校正數為 33.9 ，則七歲半與八歲半之間各年齡的B校正數可用下述推算方法求得之。八歲半與七歲半的B校正數之差別為 $34 - 33.9 = 10.5$ ，那末年齡相差一個月，B校正數應相差 $10.5 \div 12 = 0.875$ ，所以七歲七個月或九十一個月的B校正數為 $34 - 0.875 = 33.125$ ，為簡便起見，等於 33.7 歲八個月或九十二個月的B校正數為 $33.125 - 0.875 = 32.25$ ，為簡便起見，等於 32.3 。照此法求得其他各年齡的B校正數，得表四十八，即實足年齡與B校正數對照表。求任何兒童的B分數祇須從表四十八查出與該兒童實足年齡相對的B校正數，然後將該校正數加在他的B分數上就行了。

表四十七 表示求各年齡B校正數的方法

相對 題數	11	12	13	14	15	16	17
0	23	1	3	1	2	1	
1	25	2	3	3	4	1	
2	27	2	3	2	1	1	
3	28	3	0	6	3	1	
4	29	0	5	5	5	1	
5	29	2	5	9	6	1	

號	紅	標	號	號	號	號	號	號	號	號	號	號	號
6	30	2	6	6	5	1	2	1	0	0	0	2	
7	31	0	10	6	3	5	2	0	0	0	0	0	
8	32	1	8	9	6	4	0	1	0	0	0	0	
9	32	2	10	5	5	2	1	0	0	0	0	0	
10	33	2	6	15	8	6	3	2	0	0	0	0	
11	34	2	11	20	5	4	1	0	1	1	0	0	
12	35	2	9	21	12	3	6	2	1	1	0	0	
13	36	4	14	25	12	4	3	1	1	1	0	0	
14	37	1	12	23	17	12	4	1	1	3	0	0	
15	38	2	13	21	25	15	12	5	2	2	0	0	
16	39	0	17	25	23	22	6	4	0	3	0	0	
17	41	2	17	34	24	31	14	4	4	4	0	0	
18	42	1	5	20	25	20	19	11	5	5	1	1	
19	44	3	3	20	27	32	26	21	3	3	0	0	
20	46	0	4	22	33	42	26	19	5	5	1	1	
21	48	1	4	18	25	35	32	28	10	10	2	2	
22	50		2	6	30	40	35	25	6	6	1	1	
23	52		2	6	27	32	42	24	9	9			

川11四

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		8	13	26	25	23	5	1	
29	66				7	3	21	19	12	5	0	1
30	70				2	4	14	11	7	2	2	
31	75				1	6	3	5	4	1		
32	78					0	1	1	3			
33	81					1	1	2				
34	85					1	1					
35	90											
學生總人數	35	173	347	399	426	500	452	303	113	16	2	
B量表分數	68	59.5	53.5	58	52.5	50	48.5	44	38	28	21	
F量表分數	34	36.0	38.0	44	48	50	52.0	52	54	52	58	
B校正數	34	28.5	15.5	9	4.5	0	-3.5	8	-16	-24	-37	

表四十八 實足年齡與B校正數對照表

第四編 編造測驗的方法

總 經 營 部

冊三第

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

III G 量表的編製

編製 G 量表的手續如下：

- (一) 用編就測驗去測量各年級兒童，人數愈多愈好，取樣愈隨機愈好。
- (二) 作一各年級分數的次數分配表，如表四十九。
- (三) 求各年級的平均 H 分數，這就是各年級的 H 常模。
- (四) 假使表四十九中各年級是秋季始業的而舉行測驗的時間在一月底，則 41.1 為五年級中間程度，相當於 G 分數 5.5；47.1 為六年級中間程度，相當於 G 分數 6.5；兩級相差數為 6 個 T。根據此以推算 5.5 而 G 5.5 之間各 G 分數的相對的 H 分數，得表五十，這就是 G 分數與 H 分數對照表。其他兩年級之間各 G 分數的相對的 H 分數，亦照上述手續求得之。

表四十九 各年級分數的次數分配表

年級	V		VI		VII		T ₅₀
	次數	H 分數	次數	H 分數	次數	H 分數	
0	2	60	1	30			30
1	3	99	2	66			33
2	4	148	0	0	1	37	37

第四編 編造測驗的方法

3	3	123	2	82	2	82	41
4	4	180	3	135	2	90	45
5	2	100	4	200	3	150	50
6	1	55	2	110	4	220	55
7	1	58	2	116	2	116	58
8			1	63	1	63	63
9							66
10							70
總 計	20	835	17	802	15	768	
平 均		41.1		47.1		50.5	

表五十二 G分數與T分數對照表

T分數	G分數
41.1	5.5
41.7	5.6
42.3	5.7
42.9	5.8
43.5	5.9
44.1	6.0
44.7	6.1
45.3	6.2
45.9	6.3
46.5	6.4
47.1	6.5
47.5	6.6
47.8	6.7
48.1	6.8
48.5	6.9
48.8	7.0
49.1	7.1
49.5	7.2
49.8	7.3
50.1	7.4
50.5	7.5

問題

1. 「總數」如何解釋？
1. 總數 2. 總數
2. 年齡十二歲半的兒童，其百分數即等於百分之，何故？
3. 求十二歲以下各年齡之校正數時，何以用總數的總人數去除「總數」？
4. 十二歲以上各年齡之校正數的求法，何以與求總數以下各年齡之校正數的方法不同？

參考書報

1. McCull, W. A. How to Measure in Education, Ch. 10 Macmillan.
2. McCull, W. A. How to Experiment in Education, Ch. 5, Macmillan.

第二十二章 品質量表的編製

一 品質量表的性質

作文能力、書法能力、圖畫能力等的測量，在方法上與讀法能力、算術能力等的測量頗有不同。譬如要進行算術能力的測量，最重要，亦是最繁重的工作就是標準測驗的編造；要進行作文能力的測量，最重要而繁重的工作却不在測驗的編造而在記分標準或評判量表的擬訂。測量算術能力時，以在規定時間內作業的多少——如做對題數——而定程度的高低；測量作文能力時，却不能以作業的多少而祇能以作品的品質的優劣以定程度的高低。品質的評定若僅依主觀的判斷而無客觀的標準以為依據，則結果必不可靠，而各人的評判又不能互相比較。要避免這種困難，須先擬訂記分標準或評判量表，使進行評判時有劃一的標準可循。這種記分標準或評判量表專門名詞稱為品質量表。在第五章我們曾述及英人菲奢所收集的量表集，實為一種粗魯的品質量表。到現在，品質量表的編製在方法上已有極大的改進了。

二 編製品質量表的方法

品質量表通常以評判的差異度為單位。茲參照麥柯爾在 *How to Measure in Education* 書中所述，以

作文量表爲例，將編製品質量表的步驟敘述如下：

(一) 決定一個或幾個作文題目，以及時間的限制。

(二) 決定量表的應用的範圍。

(三) 用所決定的題目測量許多兒童，程度最劣的與最好的，都要包括在內。

(四) 請許多評判員將各篇作文，依其品質的優劣順序排列，並決定應得零分的一篇。

(五) 根據第四步所列之順序，計算評判員中以A篇較優於B篇者有百分之幾，以B篇較優於C篇者有百分之幾等等。依此方法進行，以求得評判員中以各篇較優於其他各篇的百分比。評判員中以最劣的一篇文章較優於最好的一篇文章的百分比，大概是零；以最好的一篇文章較優於最劣的一篇文章的百分比，大概爲100。其餘的百分比，都在零與100之間。

(六) 從上面所求得之百分比減去50%。要從各篇較優於其他各篇的百分比減去50%，理由是由這樣的。假如三十位評判員中，有十五位（即50%）把A篇排在B篇之上，當然還有十五位（50%）把A篇排在B篇之下，所以A、B兩篇作文的程度在三十位評判員的眼光裏是恰相等的。假如三十位評判員中有二十四位（即80%）把A篇排在B篇之上，顯見得在全體評判員心中A篇是較B篇爲優。從80%減去50%，得30%，代表A篇超

過日篇的數量。

(七)根據表五十一將各篇間品質的差別,由百分比化爲機誤(P_H)表五十二表示求P_H差別的方法。
表五十一 常態曲線下面積的分布(以P_H爲單位)

$\frac{X}{PE}$	$\frac{X}{PE}$		$\frac{X}{PE}$	$\frac{X}{PE}$	
	.00	.05		.00	.05
0	0000	0135	3.0	4735	4802
.1	0269	0403	3.1	4817	4831
.2	0536	0670	3.2	4845	4858
.3	0802	0933	3.3	4870	4881
.4	1063	1193	3.4	4891	4900
.5	1321	1447	3.5	4909	4917
.6	1571	1695	3.6	4924	4931
.7	1861	1935	3.7	4937	4943
.8	2053	2168	3.8	4948	4953
.9	2291	2392	3.9	4957	4961
1.0	2500	2606	4.0	4965	4968
1.1	2709	2810	4.1	4971	4974

1.2	2908	3004	4.2	4977	4979
1.3	3097	3188	4.3	4981	4983
1.4	3275	3360	4.4	4985	4987
1.5	3441	3521	4.5	4988	4989
1.6	3597	3671	4.6	4990	4991
1.7	3742	3811	4.7	4992	4993
1.8	3896	3939	4.8	4994	4995
1.9	4000	4057	4.9	4995	4996
2.0	4113	4166	5.0	4996	4997
2.1	4217	4265	5.1	4997.1	4997.4
2.2	4311	4353	5.2	4997.7	4998
2.3	4396	4435	5.3	4998.2	4998.4
2.4	4472	4508	5.4	4998.6	4998.8
2.5	4541	4573	5.5	4999	4999.1
2.6	4602	4631	5.6	4999.2	4999.3
2.7	4657	4682	5.7	4999.4	4999.5
2.8	4705	4727	5.8	4999.55	4999.6
2.9	4743	4767	5.9	4999.65	4999.7

第四卷 乘除開驗的方法

三三三

表五十二 表示將百分比的差別化為 P_{BE} 差別的方法

成 績 樣 本	A > T	N > A	B > N	K > B	E > K	L > E
百分比	50	75	75	84.41	64.47	01.13
百分比減去 50	00	25	25	34.41	14.47	41.13
P _{BE} 的差別	00	1.0	1.0	1.50	.55	2.00

(八)除了這樣直接計算 A, B, A, C, A, D 等 B, C, B, D, B, E 等的 P_{BE} 差別以外,再用各種間接的方法決定各樣本間的差別。例如表五十三中 N, A 間的距離等於 E, N 減去 T, A, B, N 間的距離等於 A, B 減去 A, N, L 間的距離等於 (E, L) - (T, A + A, N + N, B + B, K + K, E)。另外還有其他決定各樣本間距離的間接方法。

(九)從各種直接與間接的方法求得的 P_{BE} 距離的平均為真正的距離。計算的方法愈間接,在求平均距離時所得之分量愈少。

(十)根據求得之平均的 P_{BE} 距離,將各樣本順序排列,並標明各樣本與前一個樣本的 P_{BE} 距離,如下:

樣 本	T	A	N	B	K	E	L
P _{BE} 距離		0	1.0	1.0	1.5	.55	2.0

(十一)根據原來材料記錄各評判員定為零度的樣本。譬如有的幾位評判員以 H 篇,有的以 B 篇,有的以 G

篇爲零度的樣本。

(十二) 計算最中間的零度樣本，假設是 A 篇。

(十三) 計算各樣本與零度樣本間的距離，這就是各樣本的價值，如下：

樣本	A	B	C	D	E	F	G	H	I	J
價值	0	1.0	2.0	3.5	4.05	6.05				

(十四) 從零度樣本起，選擇其他的樣本，各樣本間的距離最好爲一個單位，成爲一個單位相等的量表。較小的距離，爲普通應用計，不大適宜，因爲單位太小，普通人不易覺察其間之差異。以上所述爲編製作文量表的各步驟。作文量表編就後，仍可進而求 HBG 量表。

(十五) 向學校收集兒童的作文，約二千份，各年齡，各年級都包括在內。作文題目用量表中各樣本的題目或相類的題目均可。

(十六) 至少請三位評判員，用製就的作文量表，評定這二千篇作文的價值。以三位評判員的平均評判爲各篇作文的價值。

(十七) 將各年級各年齡兒童的作文分數列成分配。

(十八) 根據十二歲兒童的分配，求 H 量表。根據八歲兒童或十六歲兒童亦可。求 H 量表之方法見第二十

一章。

- (十九) 如有必需,可照第二十一章所述之方法擴充日量表的範圍。
- (二十) 根據其他各年齡兒童的分配,求日量表,方法見第二十一章。
- (二十一) 根據各年級的分配,求日量表,方法見第二十一章。
- (二十二) 編製說明書。

三 編製品質量表的另一方法

麥柯爾在教育實驗法一書中曾敘述一種較簡易的方法,茲介紹如後:

- (一) 決定所要測量的心理特性,並嚴密的下一解說。
- (二) 決定測驗的方式和內容。
- (三) 決定量表應用的範圍。
- (四) 參考前人已編就的測驗,規定測驗中的說明,測驗的形式,以及時間的限制。
- (五) 在測驗卷子上留出相當地位以備填寫下列各項:
 - 1. 姓名
 - 2. 性別
 - 3. 年齡
 - 4. 生日
 - 5. 所在學校
 - 6. 所在年級
 - 7. 測驗日期

(六) 將編就的測驗施用於一百十個學生左右，程度好的與劣的都要包括在內。在規定時限達到前幾分鐘，須先給學生一個警告。

(七) 再將測驗施用於同一組學生，以便對於同一心理特性，如作文能力，得到兩次的測量。

(八) 將此兩組成績，各選留 100 份，每組中最優及最劣的成績不可刪去。

(九) 請三位精細的評判員評定第一組中各樣本的等第，最劣者予以 1 分，次最劣者予以 2 分，依此至 100 分。然後將三位評判員給每樣本的分數平均一下，得每樣本最後的等第。

(十) 用同樣手續評定第二組各樣本的等第。

(十一) 求兩組成績的相關，如相關太低，可增加測驗時限，或修改測驗的內容。

(十二) 本試用時所得之經驗，修正測驗的組織及應用時的說明，然後正式付印。印好後，再將測驗施用於大約二千個兒童。選擇學校，須遵照隨機取樣的原則。在所選之各校中，應將各相當年級中之學生全數測量。

(十三) 看所編測驗的應用的範圍，將八至九，或十二至十三，或十六至十七歲兒童的卷子檢出（如測驗為初小用者，選取八至九歲學生的卷子；為高小用者，選十二至十三歲學生的卷子；為中學用者，選十六至十七歲學生的卷子。）通常，十二歲至十三歲對於各級程度都可適用。不顧品質之優劣，在每一試卷之上註以號數。

(十四) 將選取的卷子，依照品質的優劣，分為十組——A（最劣），B（次最劣），C, D, E, F, G, H,

I, J (最優)

(十五) 再將 A 組中試卷，依照品質之優劣，分為五小組——a (最劣), b, c, d, e (最優)。

(十六) 同樣將其他九組各分為五小組。

(十七) 將 A a 小組中的卷子，按其品質，順序排列。

(十八) 同樣將其他四十九小組 (Ab, Ac, Ad, Ae, Ba ... Je) 中卷子，各按其品質，順序排列。

(十九) 把 A a 組裏幾本最優的卷子與 A b 組裏幾本最劣的卷子仔細比較。倘排定的等第不準確，則兩小組連接點各卷之次第可重行排列。同樣比較其他四十八個連接點各卷之品質，而更正其等第。

(二十) 在記錄紙上，按品質之次第將各卷的號數順序寫出。在最劣卷之號數後寫「分」，在次最劣卷之號數後寫「分」餘類推。

(二十一) 以上第(十四)、(十五)、(十六)、(十七)、(十八)、(十九)、(二十)各步驟，至少請三位評判員去做。做的時候要各人不知道他人所打的記號。

(二十二) 將三位評判員給每本卷子的分數加以平均。照平均分數的大小，再將各卷子順序排列。

(二十三) 從第二十二步所得之順序分配計算超過每一樣本(或卷子)的次數加達到該樣本的次數的一半的百分比。將百分比最近於 99.99% 的樣本檢出，依照表四十二，該樣本的品質為 0。將百分比最近於 99.

33的樣本檢出，依照表四十二，該樣本的品質為25。其他樣本之須檢出者，其百分比見下列第一行，各該樣本之H價值見第二行。如最後量表中的樣本有下列表中之一半就夠了，則可用20, 30, 40, 50, 60, 70, 80各H價值之百分比。若最後量表中所需的樣本數較下列表中所載為多，則可按表四十二查出所須之各百分比。

百分比	99.865	99.38	99.72	93.32	84.13	69.15	50.00
T 價值	20	25	30	35	40	45	50
百分比	30.85	15.87	6.68	2.28	•62	•13	
T 價值	55	60	65	70	75	80	

(二十四) 將上列十三本樣本檢出後，再將在品質次第中緊在每樣本前五本卷子與緊在每樣本後五本卷子檢出。這樣可得十三組——N, O, P, Q, R, S, T, U, V, W, X, Y, Z, ——每組有十一本卷子。將每組中卷子的次序混亂。

(二十五) 請多位評判員將各組中各卷，按品質的優劣順序排列，最劣者得1分，最優者得11分。前面三位評判員所評定的分數在此處亦可採用。

(二十六) 同樣，請評判員將其他十二組中的卷子各順序排列並給予相當分數。

(二十七) 計算每本卷子所得分數之平均數。

(二十八) 本此平均數，由 Σ 組選出最居中的一卷，此卷即為最可代表 \bar{H} 分數 \bar{S}_0 之卷。同樣，從 O, P, Q 等組中各選出最居中的試卷，予以 $\bar{S}_1, \bar{S}_2, \bar{S}_3, \bar{S}_4$ 等 \bar{H} 分數。這十三本樣本與其 \bar{H} 價值即成爲一個品質量表。要評定任何學生的作品（書法，作文，圖畫等）祇須將該作品與量表中各樣本比較，與該作品在品質上最相近的樣本的 \bar{H} 價值即爲該生應得之 \bar{H} 分數。

(二十九) 請至少三位精細的評判員，將原來所收集之二千份卷子與量表中樣本比較，以評定其價值。三位評判員所給分數之平均，即爲該卷之 \bar{H} 分數。

(三十) 將各年級與各年齡兒童（十二歲至十三歲兒童不必再計算）的分數列成分配。

(三十一) 根據各年齡兒童（十二歲至十三歲兒童除外）的分配，求 \bar{H} 量表。

(三十二) 根據各年級兒童的分配，求 \bar{H} 量表。

(三十三) 編製說明書。

問 題

1. 作文能力，書法能力等的測量與算術能力，證法能力等的測量有什麼不同？

2. 百分之八十八的評判員以爲圖畫樣本 \bar{H} 較優於樣本 \bar{O} ，百分之七十五的評判員以爲樣本 \bar{H} 較優於樣本 \bar{Q} 。假設有百分之九十九

由該書中選取之實驗材料，已在本書中詳加介紹。

參考書

1. McCall, W. A.: How To Measure in Education, P. 263—271. Macmillan.
2. McCall, W. A.: How To Experiment in Education, P. 130—139. Macmillan.

第五編 總結

第二十三章 測驗運動總評價

在以上二十二章中，我們將測驗的理論與實際約略討論過了。教育中測驗運動的歷史為時雖不過數十年，但其成績卓著影響鉅大，已無人可否認之。客觀的測驗已取主觀的評斷而代之，標準測驗已成為教育行政上教學上研究上不可缺少的工具。從小學至大學，各級學校，各種教育機關莫不受測驗運動的影響。從機械的技術如縫紉，木工，至高尚的欣賞，如音樂，文學，美術，莫不有相當的測量工具可以應用。現在測量學者仍在努力進行，在方法方面謀改進，在應用方面謀推廣，在範圍方面謀擴充，務將教育的各種結果都納於測驗範圍之內。在這種工作之中，有重大的貢獻，有昭彰的錯誤，有正當的和不正當的熱忱。大概而論，測驗運動截止現在所得的成就對於教育已有極大的裨益；將來的貢獻或可更為偉大。現在讓我們簡單的敘述測驗運動的現狀，將來發展的趨向，以及其可能的不良影響。

一 測驗運動的現狀

教育的功用，在變更人們的行為或反應。這些行為或反應，也就是教育的結果，歸納之可有下列各類：

- (一) 智識的反應；
- (二) 理智的反應或技術，如思想、理論、判斷；
- (三) 技能的反應；
- (四) 行為的反應，道德的、意志的、及其他行為；
- (五) 情感的反應，欣賞的經驗。

上列分類，不過為討論便利計，各類反應之間，初無嚴格的區別，其間互掩部分亦甚大。茲從上列各方面觀察測驗運動已有的成就。

(一) 智識的測量 智識比較容易測量，因此測量運動在這方面的成功最為顯著。我們已經有客觀的工具，可以使教師迅速的、經濟的、公正的決定學生在小學內各學科的成績。關於在中學裏所獲得的智識，亦已有相當工具可以應用。就是在採用新方法最遲疑的大學及研究院裏，所謂「新式」的考試亦已採用。我們可以說，各種智識，無論深淺，差不多沒有不可以測量的。我們又可以說，客觀的測驗及量表的正當的應用，對於智識的教學沒

有不可以改進的。

(二) 思維能力, 判斷能力, 理論能力等的測量 思想, 判斷, 理解等能力的健全的作用, 一部分有賴於豐富的智識, 所以測量智識同時亦測量了思維, 判斷, 理解等能力的一部分。這些名詞所代表的理智的技術都可以測量的, 並且大部分已有相當的工具可以實地測量了。閱讀了解能力測驗, 算術及代數理解題測驗, 管理機械器具能力的測驗, 面部表情評斷能力的測驗等, 雖然不十分精確, 都是各種理智作用的測量工具的實例。如華特生 (E. Watson) 所編的「公正心理」測驗, 可用以決定個人對於宗教的, 政治的, 種族的, 及其他問題的反應, 藉以測量個人思想及判斷的偏執性。所以不僅僅智識的測量, 就是比較複雜的理智的反應的測量, 亦已經證實為可能的。

(三) 動作技能的測量 書法及圖畫的測驗及量表乃客觀測量工具中之發達最早者。其他關於學術方面, 藝術方面, 體育方面, 機械方面, 職業方面各種動作技術的測驗, 亦已經先後造成。在這一方面, 準確的測量並不見得有若何重大的困難。即對於比較難以測量的音樂才能, 天賦能量以及實際能力的測驗都已有顯著的成功。

(四) 品行的測量 品行的測驗發展得比較遲, 成就亦比較小; 大致說來品行測驗現在尚在試驗時期, 還講不到大規模的應用。最近五六年以來, 在美國品格測驗的發展至為迅速, 在一九二七年一年中關於品格測量的出版物竟有二百餘種之多, 所及範圍亦至為廣博。許多心理學家以極大的努力, 致力於測量方法的改進與測量

工具的完成，內中以品格教育研究會梅 (M. A. May) 與霍桑 (H. Harshbarger) 及華特生諸人的貢獻為最大。十年前心理學者認為非測量法所能制御的品格，已逐漸被納入客觀測量範圍之內。品格測量之可能心理學者已不再懷疑之。

品格測量的基本方法與智力學力測量初無二致。品格必賴行為而表現，故品格的測量必須從觀察行為入手。不過品格測量較諸智力及學力測量，有幾點特殊的困難。第一，情境對於品格的表現較諸情境對於能力的表現，影響尤大。情境變易，則所謂「品格」亦因之而變易。例如誠實，有對於銀錢的誠實，對於學問的誠實，對於言語的誠實，而各種的誠實並不完全相關或相關極低。或者，對於親友十分誠實的人，對於外人就不一定誠實了。所以我們不能籠統的說某人何等誠實，我們祇能說某人在某種情境之下何等誠實。誠實如此，其他品格恐怕亦是如此。智力與學力的測量亦有同樣情形，同樣困難，不過在品格測量上這種情形特別顯著，困難特別加重罷了。第二，品格測量的進行不能使被試者覺察。被試者覺察有人在觀察或測量其品格，則其行為就因之而變易。這一點當然增加了品格測量的困難，但是那層困難並非不可消除的，心理學者已經有許多方法可以克服這層困難。

(五)興趣，欣賞，態度，理想等的測量 興趣，感情，情緒，欣賞，態度，傾向，理想等名詞包括許多類不同的反應。常常有人說這種難以捉摸的反應是不能測量的。雖然這方面的測量較其他方面的測量落後，但是說興趣，情緒，欣賞等是不能測量的，乃是重大的錯誤。已有的成功就是最有力的反證。雖然所用的方法還是很粗魯，很費時間，很

不經濟，但是憤怒及奮興的傾向；鑒賞文學，詩詞，音樂，圖畫的能力；消遣及職業的興趣；和愛或殘暴，好學或不好學，爽直或陰深的傾向等等的測量，都已有相當的成功。現在我們還不能或者竟是永遠不能用「紙與筆」的測驗去測量人類的各種反應及反應的傾向，有如醫學裏不能用團體的方面很便宜的迅速的去診斷生理的行為一樣。我們未能編造五分錢一份的測驗去測量此種複雜而難以捉摸的特性，並不能證明測量的不可能。其實，人類任何特性，凡是存在的，都可以應用已有的方法去得到較客觀，較準確的度量。

總結起來，我們可以說雖然現在已有的，價值便宜而施行便易的測驗大多僅能測量教育中較狹隘的結果，如智識及技能，但是許多專家已在努力進行於教育結果其他方面的測驗的編造，並且已有成就表示這種工作的成功是可以預期的。

二 測驗運動對於教育的貢獻

測驗運動的主旨在減少主觀的偏見，增加教育結果的度量的客觀性，準確性，完備性，及精密性。測驗在教育上的功用，前面已有專章討論，現在再將測驗運動對於教育的貢獻簡單陳述於后。

(一) 促進科學的教育 各種科學的最終目的在推測與控制，推測愈準確，則控制亦愈切實。要推測未來，必須根據已往及現在。有了標準測驗，我們得以可靠的度量某人已往學習的成績以及現在的程度，即增進了我們

對於某人未來的成就的推測的準確度。根據此推測以規定設施教育的計劃與方針，則教育上種種設施才不致如無的放矢，而教育的效率亦因此提高了。

分析言之，標準測驗可以促進教育目標的確定，教育工具的改進，與學生學力的診斷。三者俱備，才能將主觀的，非科學的教育推動而進於客觀的，科學的境地。從許多方面，精密的，客觀的測量可以幫助決定具體的教育目標，顯示何種事實何種技能為學生所已習得者，何種為學生所尚未學得者，某種事實或技能須學習至何等程度等等。例如測驗的結果已顯示從前學校裏所教授的算術，字彙，有一部分在實際生活中極少有應用的機會，不值得費許多時間，金錢去學習牠們。學生在學校裏所學習的智識，技能，習慣，態度等，一部分他們早已俱備，這種重複的學習乃教育中的浪費。有時，學生所極需要的學習反為學校所忽視。總之，測驗的結果足以指示我們教育所應努力的目標。客觀的測驗一方面幫助教育目標的建立，一方面使目標更為確定。

各種教育工具的效率必須能互相比較，才有改進的可能。而各種教育機關，教學方法，教材組織，時間支配，鼓勵方法等等不能有公平的，準確的，精密的，換一句話說，科學的比較，除非各種教育工具所產生的結果可以客觀地測量。教育中科學的研究與測量工具的發展的同時並進並不是偶然的事，實勢之必然。沒有測量的工具，則各種教育工具的價值便不能確定，便不能有所改進。

為診斷學生的困難——學業方面的，社交方面的，行為方面的——作業與能力的測驗又為必需的工具。教

學的方案，要有成效，必須詳密規定，對於不良的習慣加以糾正，對於良好的習慣加以充實。教育及「再施教育」應該依照測驗的結果，利用能力之所長，校正能力之所短。這樣的教學才是有效的，科學的，而不是盲目的。

(二) 促進個別的適應 注重兒童的個性為新教育主要的特徵。個別的現象在測驗運動發生以前雖已為教育者所注意，但教育者對於個別現象的認識却因測量工具的發展而愈切實，愈明確。人們的能力不但高低不同而方向亦有異。易言之，人們不但在普通智力上天賦有厚薄，而且各人有其特長，有其特短。依照個人智力的高低以規定成績的標準，根據個人能力之所在以設施相當的訓練，則教育才能適合個性，發展個性。這種教育上的設施，如能力分組，教育指導，公允的競爭，努力的獎勵等，都因測驗的應用而易於實現。

三 測驗運動的流弊及限制

雖然測驗運動在方法上是正確的，在實施上是有效的，但仍不能擔保誤解，流弊的不發生。我們必須審慎於測驗的誤用及誤解。其主要的流弊及限制如下：

(一) 以常模為理想的標準 我們需了解測驗中的常模乃一種統計的量數，僅代表各年級或各年齡的平均作業，而不是某年級或某年齡的理想標準。除非不諳最簡單的統計方法的人，不應將平均的成績與理想的標準相混而自滿於平庸的作業。這個困難的根源乃在應用測驗者智識的缺乏而不在測驗的本身。

(二)以測驗材料及方法作為模範 第二種流弊是教師以測驗中所採用的文章、圖畫、書法樣本及其他材料作為教學的模範，叫學生去模倣，使以後測驗時可以得到較好的成績。雖然測驗中有許多材料可以當作模範用，但這並非測驗主要的功用。多數測驗是嚴格的診斷工具，不應誤用為學習者的模範，有如照愛克孫光以前醫生為病人所規定的食料不能認為合乎衛生的食料模範一樣。這種流弊的防止不在廢棄測驗乃在開導教師。

(三)偏重較具體的教育結果 前面我們已經講過，現在發展最完善而施行最便易的測驗祇限於智識及技能的測量一方面。其他同等重要或更為重要的教育結果，如理想、態度、興趣等，我們還不能便易地圓滿地去測量。假如我們專以學生在這種測驗上的成績去估計教學的效果，則智識及技能實有被過分重視的危險；他方面較複雜而難以測量，但或者更為重要的教育結果實有被忽視的可虞。教師一方面固然不應輕視這種較機械較具體的教育結果的價值，但同時尤需努力於較廣博的教育目標的實現。克伯屈 (W. H. Kilpatrick) 教授曾說：

「現有教育測驗所測量的大多限於學習結果中較機械的方面。一位教育行政當局舉行了一次拼法、算術、或讀法的測驗。教師遲早會得悉他們學生的成績。除非這位行政當局是有見識的，這些教師教學的效率將根據他們學生測驗的成績而被考核。如果這位行政當局能同樣圓滿地測量教師在培植理想與態度方面的成功，使所有的教育結果得以測量，則情形當然不同。但照現在情形，這位行政當局實有使教師忽略於現在尚不能測量

的教育結果，如理想，態度，道德習慣的危險而使教師只注意於教育結果中最機械的一部分。」

上面所說的流弊雖屬真切，但不過屬於暫時的。將來客觀測驗的範圍擴大，則這種流弊自然無形消滅。所以積極的辦法不在取消或減少現有的測驗，乃在努力於他種教育結果的測驗的編造。

(四)客觀測驗與優美生活的衝突的恐懼 或者有人說僅僅教育中較卑低的部分是可以計算的，測量的，而關於精神方面的教育結果將因測量而耗失，測量愈精則耗失愈大。學校將成爲一個智識的製造廠，造成許多律師醫生，個個能令顧主滿意，但個個一點修養也沒有。這種觀念代表一種很普通的心理，恐怕科學與測量，一旦應用於人事方面——家庭，國家，教育，宗教——就會消滅生活的美質，將生活的高尙理想變爲卑陋的物質主義。這種恐懼係根據於一種絕對錯誤的心理學，是完全不足憑的。凡事物之存在者皆有其數量。去測量牠不過是去尋求他的數量。人類現在並不因定量的植物學的成立而減少了對於花草的鑑賞。勇敢性與持久力的測量以及牠們在生理上的根據的考查並不致減少人類的勇敢性與持久力。假使一種品性是值得尋求的，我們對於這品性愈了解，愈能測量，則我們一定愈加努力的去尋求。把人當作神祕的東西並不見得提高了人的尊嚴。將科學與測量應用於教育，我們可以很安全的享受直接的，實際的利益而決沒有侵害理想主義的危險。

總結起來說，應用測驗的人應該辨別測驗中想像的與真正的危險。假使誤認平均成績爲理想的標準，假使將測驗材料與教學材料相混淆，假使不能區別診斷的方法與教學的過程，假使教師除掉通行測驗中所表現的

教育目標以外不承認其他目標的存在，那麼，測驗的影響真有莫大的貽害。但是這些測驗的誤用是極愚笨的，可以完全避免的。這種誤用不應該蒙蔽測量方法根本的正確與測驗應用實際的價值。糾正這種錯誤最好的方法，一方面由科學家擴充測驗的範圍，提高測驗的品質；一方面由教師多方從事於測驗的研究與應用。當然，測驗與教師一樣，亦須與新方法新材料的發展同時並進。測驗，又與教師一樣，對於方法的改進，教育目標的確立，可以有重大的影響。

問 題

1. 品格測量的基本方法與智力學力測量有無區別？
2. 品格測量有什麼特殊的困難？
3. 舉幾個例子表示測驗的結果可以幫助教育目標的確定。
4. 論考模 (Norm) 與標準 (Standards) 的區別。
5. 各個兒童學業的標準應否別？各個兒童的學業的標準應如何決定？
6. 總起來說，測驗運動對於教育的影響，是利多於弊，還是弊多於利？

參考文獻

- 1 Kiepatrick, W. H.: The Foundations of Method, P. 107—109, Macmillan.
2 Thorndike, E. L. and Gates, A. I.: Elementary Principles of Education, Ch. 13, Macmillan.

附錄 1

漢英名詞對照表

Accomplishment quotient	成業商數
Arithmetic mean	算術平均數
Average deviation	平均差
Class interval	組距
Coefficient of correlation	相關係數
Coefficient of intelligence	智力係數
Coefficient of variability	差異係數
Diagnosis	診斷
Discriminative ability	鑒別力

Duplicate form	覆份
Educational age	教育年齡
Frequency distribution	次數分配
Intelligence quotient	智力商數
Lower quartile point	下二十五分點或下四分點
Median	中數
Mental age	智力年齡
Mental quotient	心理商數
Mode	衆數
Norm	常模
Age norm	年齡常模
Grade norm	年級常模
Objectivity	客觀性
Pedagogical T score	教學T分數

Probable error	標誤
Promotion T score	調級T分數
Quality scale	品質量表
Quartile deviation	二十五分差
Range	全距
Rank correlation	等級相關
Reference point	參照點
Reliability	可靠性
Reliability coefficient	可靠係數
Sampling	取樣
Extensive sampling	廣的取樣
Intensive sampling	深的取樣
Random sampling	隨機取樣
Scale	量表
Likert	Likert

Standard deviation

標準差或均方差

Standard error

標準錯誤或均方錯誤

Test battery

測驗彙集

Trait

單位

Upper quartile point

上二十五分點或上四分點

Validity

正確度

Validity coefficient

正確係數

附錄一

漢英人名對照表

Anderson, W. N.	安特生	May, M. A.	梅
Ayres, L. P.	愛里斯	McCall, W. A.	麥柯爾
Bagley, W. O.	勃格萊	Monroe, W. S.	孟祿
Burt, C.	勃淮	Obis, A. S.	渥鐵斯
Binet, A.	比納	Patterson, D. G.	派忒孫
Boberga	鮑勃太格	Pearson, K.	潘爾遜
Botton, T. L.	波爾頓	Pinker, R.	賓特納
Bridges, J. W.	勃雷頓司	Quetelet	闊特來
Buckingham, E. R.	白根漢	Rice, J. M.	萊斯

Cattell, J. Mck.	喀察爾	Ruoh, G. M.	羅魯
Chapman, J. O.	賈波門	Safiotá, F. U.	沙飛亞替
Courts, S. A.	克底斯	Seashore, O. E.	西旭
Dearborn, W. T.	第亞彭	Simon, T.	西門
Doll, E. A.	道爾	Spearman, C.	司畢門
Ebbinghaus, H.	愛屏好斯	Starch, D.	斯達奇
Elliot, E. O.	伊立歐忒	Stenquist, J. L.	斯丹魏斯騰
Fisher, George	非奢	Stern, W.	斯旦恩
Franzen, R.	弗倫純	Stoddard, G. D.	施道突
Freeman, F. N.	費里門	Stone, O. W.	斯東
Galton, F.	高爾登	Studobaker, J. W.	斯都特培克
Gates, A. I.	蓋茲	Symonds, P. M.	薩曼司
Goddard, H. A.	高達德	Terman, E. L.	德爾備
Hardwick, R. S.	哈特會克	Terman, L. M.	推孟

Hartshorne, H.	霍桑	Thorndike, E. L.	桑戴克
Herring, J. P.	黑林	Toops, H. A.	杜波司
Hillegas, M. B.	希米格司	Trèves,	特立夫
Horn, E.	霍翁	Washburne, C.	華盛朋
Johnson, F. W.	約翰生	Watson, G. B.	華特生
Johnston, Katherine	約翰斯頓	Wilson, G. M.	威爾遜
Jones, N. F.	瓊斯	Winch, W. H.	溫屈
Kalley, T. L.	克萊	Wisler, O.	威士拉
Kelly, R. L.	克萊	Wood, B. D.	吳德
Kilpatrick, W. H.	克伯屈	Woodworth, R. S.	吳偉士
Kirkpatrick, E. A.	克伯崔克	Woody, C.	吳特
Kuhlmann, F.	庫爾門	Wundt, William	馮德
Mann, Horace	邁恩	Yerkes, R. M.	賽格赫斯

0486			
375.21 教育測驗			
7538 陳選善			
郵寄	12.26		
抄本	12.26		

中央政治學校附設蒙藏學校

圖書館

借閱者注意

- (一) 加意愛護勿失原有形狀
- (二) 損壞或遺失應照原價加倍賠償
- (三) 借閱以一星期為限期滿欲續借者須持書至館聲明但本館於必要收同時須即繳還
- (四) 逾期不歸還者應照章納金

中華民國二十三年四月初版
再版

(340821精)

大學叢書
(教本) 教育測驗一册

每册定價大洋貳元肆角

外埠酌加運費匯費

著者 陳選善

發行人 王雲五
上海河南路

印刷所 商務印書館
上海河南路

發行所 商務印書館
上海及各埠

版權所
翻印必究

(本書校對者喻飛生)

精五五五C

752738

U 3