

教育講話叢書

陳選善主編

7701

# 教育測驗

陳選善著



世界書局印行

## 教育講話叢書編輯凡例

- 一 本叢書就教育各部門，約請專家，分別編著，定名為教育講話叢書。
- 二 本叢書編輯主旨，在就教育各部門，給予讀者一種綜合的鳥瞰，以為專門研究之基礎。
- 三 本叢書內容着重在歷史背景之敘述，發展趨向之指示，各家學說各派主張之介紹，參考材料之提供，研究方法之指導，俾啟發讀者思想，引起讀者興趣。
- 四 本叢書文字力求流利生動，深入淺出，俾便閱讀而利了解。
- 五 本叢書讀者以師範生，大學教育系學生，中小學教師，一般從事教育工作人員及對於教育有興趣者為對象。

教育講話叢書書目及編著者一覽

教育哲學講話

傳統先編著

心理學講話

張燦翔編著

教育心理學講話

高君珊編著

教育社會學講話

陳科美編著

西洋教育史講話

林漢達編著

中國教育史講話

林漢達編著

比較教育講話

韋 懋編著

教育及學校行政講話

胡祖蔭編著

教育測驗講話

陳選善編著

教育統計學講話

沈有乾編著

教學法講話

董任堅編著

幼稚教育講話

董任堅編著

小學教育講話

孫貴定編著

中學教育講話

黃式金編著

社會教育講話

俞慶棠編著

鄉村教育講話

古 樸編著

職業教育講話

何清儒編著

教育研究法講話

陳選善編著

教 育 講 話 叢 書

陳 選 善 主 編

---

教 育 測 驗 講 話

陳 選 善 編 著

---

世 界 書 局 印 行

## 自序

本叢書編輯的主旨，內容的側重點和文字的體裁已詳編輯凡例，但本書是否能夠達到預期的目的，是頗成問題的。不過我想讀者看完本書以後，對於測驗的性質功用和發展應有的趨向，總可以有一種概括的認識。

本書因限於篇幅，對於編造測驗的技術和新式考試的方法，未及敘述，讀者可選閱附錄一所列參考書目。關於處理測驗結果的方法讀者可閱讀本叢書沈有乾先生編教育統計學講話。本書之編著承陸高誼先生給予鼓勵，內子德音將原稿抄寫一遍，一併致謝。

陳選善 中華民國三十三年五月

# 教育測驗講話目次

第一講 教育測驗的範圍和性質	一
一 測量與測驗	一
二 直接測量與間接測量	二
三 測量的要素	三
四 測驗的種類	五
第二講 測驗發展的歷史	一
一 智力測驗的發展	一
二 教育測驗的發展	一九
三 我國測驗運動的回顧與展望	二五
第三講 測驗的條件	二九
一 信度	二九
二 效度	三三
三 常模	三八

四 施行及校閱的便利

五 複份

六 價格

第四講

智力測驗

一 個人智力測驗

二 團體智力測驗

第五講

國文測驗

一 閱讀測驗

二 默字測驗和識字測驗

三 文法測驗

四 作文測驗

五 書法測驗

六 國文常識測驗

第六講

數學測驗

一 四則測驗

四二

四三

四三

四五

四五

五九

六八

六八

七二

七六

七七

八三

八四

八五

八五

二	應用題測驗	八七
三	練習測驗	八八
四	中學數學測驗	九一
第七講 其他各科測驗		
一	常識科測驗	九四
二	社會科測驗	九五
三	自然科測驗	九八
四	英文測驗	九九
五	各科混合測驗	一〇三
六	其他各科測驗	一〇四
七	教育測驗評價	一〇四
第八講 性格測驗		
一	作業測驗	一〇六
二	心理問卷	一〇八
三	評量表	一三一



第九講 測驗結果的解釋

一 年齡分數	一三七
二 年級分數	一三九
三 百分數	一四〇
四 標準分數	一四〇
五 T分數	一四二
六 智力商數與教育商數	一四四
七 B分數	一四七
八 成績商數	一四九
九 分數的比較性	一五三
第十講 測驗結果的應用	一五五
一 團體調查	一五五
二 個人診斷	一五八
第十一講 學校設備與設施的測量	一七〇
一 學校設備的測量	一七〇

二 學校設施的測量	一七五
第十二講 測驗運動的回顧與前瞻	一八七
一 成就	一八七
二 批評	一八九
三 基本認識	一九〇

# 表 目

表一	蕭孝嶸訂正墨跋智力量表百分數常模表	三九
表二	蕭孝嶸訂正墨跋智力量表標準差常模表	三九
表三	根據艾偉識字測驗推算之各級識字數量(以音義俱知字數為準)	七四
表四	根據艾偉識字測驗推算之各級識字數量(以音義俱知知音不知義或知義不知音字數之半為準)	七四
表五	楊黃李合編識字測驗之年級常模	七五
表六	根據楊黃李合編識字測驗推算之各級識字數量	七六
表七	艾偉修訂初等代數測驗常模	九二
表八	艾偉中學英語測驗量表甲乙之常模	一〇二
表九	艾偉高中英語默讀測驗常模	一〇三
表十	學生性格評量表	一三三
表十一	常態分配下T分數之解釋	一四三
表十二	第二次斯丹福比納量表各年齡之智商差異度	一四五

表十三	表示計算 $G_t$ $G_p$ 的方法	一六〇
表十四	邵爽秋教學效率測量表	一七六

# 第一講 教育測驗的範圍和性質

## 一 測量與測驗

教育中的測量並不限於測驗。「測量」二字相當於英文字 Measurement，「測驗」相當於英文字 Test。測量可以包括測驗，而測驗不過測量的一種工具和方法而已。

譬如我們要考核一個學校的設施，可以有兩種着手的方法。第一，我們可以考核這學校的行政組織是否健全，教師人選是否合格，校舍設備是否妥適，教材教法是否完善。我們要對於上列各方面加以正確的考核，當然不應完全憑考查者個人主觀的判斷，而要事先釐訂客觀的標準，然後根據這些標準，作精密的考核。最後以數量的術語，來表示考核的結果。這種工作雖然用不到什麼測驗，我們可以稱之為測量。因此我們對於教育測量可以下一個定義：凡對於教育任何方面，根據客觀的標準，作精密的考核，而以數量的術語來表示考核的結果者，均可稱之為教育測量。

考核一個學校的設施，另一種着手的方法，就是直接考核這種學校教育，在學生身心方面所引起的變化。換言之，學生在德智體羣各方面的成績。這裏有一個重要的假設，就是學生是學校的中心對象，學校教育的目的，就在促進學生身心各方面健全的發展。學校行政組織、教師、校舍、教具、課程、教材、教法都是為

幫助學生健全的發展而設的。西洋有一句諺語，*The test of pie is in the eating*。我們可以說 *The test of a school is in its students*。學校教育的成敗要看這學校所造就出來的學生如何。所以考核一個學校的設施最直接而有效的方法就是考查學生的成績。考查學生的成績就用得到測驗。

這本書所討論的雖然着重於測驗，但下面所講到的原則亦適用於一般的教育測量。

## 二 直接測量與間接測量

一切測量可分為直接的與間接的兩類。直接測量祇適用於事物或現象，可直接觀察而加以度量者，例如距離的長短，物體的輕重是。凡事物或現象不能直接觀察而加以度量者，必需採用間接的測量。一切心理現象均屬於此。譬如一人的智力、品格、興趣，我們無從直接加以觀察；所能觀察的，祇是人的行為、動作、反應。從行為、動作、反應的觀察，以度量其智力的高下、品格的優劣、興趣的趨向，這便是間接的測量，亦是心理測量根本的方法。所謂「視其所以，觀其所由，察其所安，人焉廋哉，人焉廋哉」，心理現象的測量，還逃不出這種方法呢。

這種間接的測量方法並不限於心理現象，自然科學有時亦採用的。例如度量空氣的溫度，要看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，正如兒童的行為是其智力、學力或品格所引起的結果一樣。

但偶然的，隨意的觀察，不能正確地辨別人們智慧的高下，品格的優劣，興趣的趨向。因為引起行為的情境在各人既不盡相同，對這些不同的情境所發生的行為，當然不能嚴格比較。張某在一個範圍極小，稽核嚴密的機關充會計員，十餘年來絕無弊端。李某在一個範圍龐大，而組織散漫的機關中充收納員，每日銀錢進出以數萬元計，服務不到一年，即發生弊端。我們根據這種事實，似乎不能貿然斷定張某比較李某誠實，因為兩人所處情境不同。假使易地而處，說不定張某不到一月就會作弊呢。要作公允的比較，我們應該安排一些標準的情境，觀察人們在這些標準情境中所發生的反應，以鑑別其智慧的高下，學問的深淺，或品格的優劣。所謂測驗就心理學者所安排的一些標準的刺激，用以引起人們的反應的。因此我們對於測驗可以下一個定義：測驗是一個或一羣標準的刺激，用以引起人們的反應，然後根據這些反應，以度量其智力、品格、興趣、學業等。

測驗的性質和形式，與普通學校考試大致相仿，但測驗自有其別於普通學校考試之特徵在。我們在第三講再為詳述。大致說來，測驗的材料較為精選，取樣較為廣博，記分較為客觀而結果較為可靠。

### 三 測量的要素

要進行一種測量不論是物質方面的，或心理方面的，必需具備下列幾種基本的要素。

第一、計算的單位 測量距離的長短，以寸尺丈里為單位；測量物體的重輕，以兩、斤、磅為單位；測量時

間的久暫，以秒、分、時、日、月、年、世紀爲單位。顯然的，沒有計算的單位，測量就無從進行。

像測量輕重長短所用的單位有兩種重要的特徵。第一，這種單位有普遍的意義，換句話說，這種單位可以離開測量的對象而存在的。尺寸可以用來計算一塊布的長度，亦可以用來度量人的高度。而且其意義十分確定，絕不致引起不同的解釋。第二，這種單位有相等的價值。換句話說，第一個單位與第二個單位間的距離等於第二個單位與第三個單位間的距離。例如有甲乙丙三人，甲身長五尺四寸，乙五尺六寸，丙五尺八寸，甲乙相差二寸，乙丙相差亦二寸。我們可以說甲乙之間的差別，等於乙丙之間的差別。我們可以如此說法，因爲尺寸的單位具有相等的價值。

在心理測量中我們亦必需有計算單位，去表示智力的高下、品格的優劣、學業的程度。我們所用的單位究竟是什麼？心理測量中所用的單位是否有普遍的意義，有相等的價值？這些問題我們在第九講中都要詳細講到。我們現在可以說尋求一個適當的單位，是心理測量中最重要同時亦是最困難的工作。

第二、參照點 參照點就是計算的起點。最好的參照點是零點。身高體重的測量即以零點爲計算的起點。但是有許多事物或現象的測量，並無「絕對的零點」可求。例如要說明某一個地方在地球上的位置，因爲地球是圓的，東西南北都是比較的說法，並無自然計算的起點。因此我們就不得不規定一個地點作爲計算的起點。我們都知道現在國際間公認的參照點是南北以赤道爲計算的起點，東西以英國（Greenwich）地方爲計算的起點。他如測量溫度以結冰點（或冰點以下三十二度）爲參照點，測量陸地的



高度以海洋面爲參照點。這些都可稱爲人定的參照點。

心理測量中所採用的參照點，可以說都屬於人定的參照點一類，因爲智力、閱讀能力、算術能力等的零點甚難決定。同時，現行心理及教育測驗所採用的參照點頗不一致，因此從各測驗所得結果往往不能直接比較。要改進這種紛歧的現象，須由心理學者共同決定一種參照點，以後編造測驗即一律以此點爲計算的起點。

第三、量表 量表就是測量的工具。量表上載有度量的單位，且往往註明度量的起點。應用量表以度量事物，祇須觀察此事物在量表上距參照點若干單位。一把尺便是量表最簡單的例子。溫度計上水銀管旁邊的分度表，便是度量溫度的量表。在心理測量方面，量表乃是作業成績（如書法圖畫作文）或測驗分數的排列，以顯明的單位表示一種能力的階梯。這種量表是一種固定的標準，凡個人或團體的作業，都須依此以度量之。

第四、測驗 物質方面的測量有單位、參照點、量表，就可以進行。心理現象的測量還須有第四種要素，就是標準測驗。前面我們說過心理現象不能直接觀察，祇有從行爲上可以觀察出來，測驗就是引起行爲的一種工具。

#### 四 測驗的種類

測驗的分類可根據幾種不同的觀點：

一、以測量對象為標準，測驗可大別為下列幾類：

(一)智力測驗 人類的智力高下不等，是有目共見的事實。孔子說「惟上智與下愚不移」已明白承認人類在智力上的差異。心理學者對於智力的性質，尚無一致的見解。各人所下定義，亦頗多出入之處，因不在本書討論範圍之內，不能詳為敘述。茲就兩個最普通最簡單的定義，說明智力的性質。

第一、智力是適應環境的能力。環境有簡單有複雜。飢則食，寒則衣，這是最簡單環境的適應。主持一國的政治，應付國際的變化，這是比較複雜環境的適應。個人處任何複雜艱難、變化繁多的環境，而能應付裕如，這人是比較聰明的。一人處簡單單調的環境，而仍不能有圓滿的適應，這人是比較愚笨的。

第二、智慧是學習的能力。一個人學習得愈快而所能學習的事物愈高深複雜，則其智力愈高。有些人既能學些簡單的技能，做些機械的工作，若涉及抽象的理論，便格格不入了。有一位心理學者說，智慧是抽象的思维能力，不失為一個良好的定義。

桑戴克 (E. L. Thorndike) 曾把智力分為抽象智力，即運用抽象概念的能力；機械智力，即運用實在物件的能力；社交智力即應付人事的能力。

(二)教育測驗 教育測驗所測量的乃學業的成績。教育測驗又可依科目分為若干種：如閱讀測驗、算術測驗、史地測驗等等。

有些科目如寫字、作文、圖畫，其測量方法與讀法算術科的測量頗有不同。測量閱讀能力或算術能力主要的工作是標準測驗的編造。測量書法或作文成績主要的工作，是評判量表的擬訂。譬如測量算術能力，以在規定時間內在測驗上做對的題數而定程度的高低。一篇作文，只有好壞的問題，並無正誤的問題，所以只能以品質的優劣，定程度的高低。品質的評定若僅憑主觀的判斷，而無客觀的標準，則評定的結果，必不可靠。因此心理學者就有作文量表書法量表圖畫量表的編造。這種量表實際上是幫助評定分數的一種工具，專門名詞稱為品質量表。所謂品質量表不過是許多作文樣本或書法樣本的排列，從最壞的到最好的，而這些樣本的品質，已由心理學者根據多數人的意見，運用統計學的原理，詳加釐定。在評定學生的作文或書法時，我們祇須同量表上的樣本比較，以定分數。如此評定者有一標準可資遵循，不致全憑主觀了。

(三) 特殊才能測驗 特殊才能的存在，大多數心理學者都承認的。普通的觀察表示人們各有所長，亦各有所短。假定訓練的機會相等，則所長方面，表示特殊才能之所在，所短方面，表示特殊缺陷之所在。特殊能力的分析與測量尚在未成熟的階段。據現在所知，音樂和圖畫的才能比較是獨立的，與智慧幾無關係。此外計算能力，機械能力，和社交能力雖不若音樂和圖畫才能那樣和智慧全無關係，但是相關亦不高。這幾方面已有相當測驗可資應用。

(四) 性格測驗 普通的觀察使我們相信，除了普通智力，特殊才能以外，個人間又有性格上的差別。

性格的涵義和類別，心理學者並無一致的見解，我們不妨以性格一名詞，來代表智能以外其他各種心理的屬性，如情緒、性情、意志、品格、興趣、態度等等。這方面的測量工具可說尚在試驗時期。

(五)職業測驗 職業測驗可分為職業性向測驗、職業能力測驗兩類。前者之功用，在指示一人所適宜的職業，後者是度量一人在某種職業上實際的能力的。

### 二、以測量人數為標準

以測量人數論，測驗可分為個人測驗與團體測驗兩類。個人測驗最大的優點，就是主試者對於被試者的反應，較有親切觀察與嚴密控制的機會，因此所得結果亦較團體測驗正確而可靠。團體測驗的利益，就是時間上的經濟。

### 三、以測驗材料為標準

以測驗所用材料為標準，可分為文字與非文字兩類。大多數智力測驗都是文字的，或語言的測驗。譬如比納(A. Binet)、西蒙(T. Simon)智力量表的用法，大部分靠言語問答。聾啞的兒童當然不能適用，主試與兒童之間，方言上不能互相了解，也發生困難。文字測驗對於未受文字訓練的成人或兒童亦不能適用，所以幼童測驗大都採用圖畫作為材料。另一類智力測驗，着重在做，不專靠文字或言語的對答。最普通的是機巧板。做測驗時，需將各種形式不同的木板安放在適當的地位，以所須時間及錯誤定智力的高下。我國的益智圖頗與此相類。另一種常用的材料是迷津。受試者須尋覓一條最簡捷的出路而不走入迷

路。這類測驗與比納式測驗性質上略有不同，可以補文字語言測驗之不足。至於教育測驗當然偏重文字。

#### 四、以測驗功用為標準

以測驗功用為標準，測驗可有幾種可能的分類法，如預測測驗與成績測驗之別，難度測驗與速度測驗之別，調查測驗與診斷測驗之別。茲依次討論之。

(一) 預測測驗與成績測驗 預測測驗的功用，在推測某人在某方面將來的成功；成績測驗的功用，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。譬如要測量打字成績，祇須看被試者在規定時間內，能打多少字，有多少錯誤，或打一篇標準文字，須多少時間，有多少錯誤。要預測一人打字能力是要在他未學習打字以前推測他未來的成功，所以不能以打字做測驗的材料。預測測驗的編造，大致先依據心理的分析，選擇材料，然後根據試用的結果，確定其效度。預測測驗的編造比成績測驗困難多了。

實際上預測測驗與成績測驗在功用上不能嚴格區別。據研究結果，成績測驗往往亦是最好的預測測驗。換句話說，一人已往的成績，是推測他將來造就最好的根據。

(二) 難度測驗與速度測驗 難度測驗度量程度的高低，速度測驗度量作業的快慢。例如算術的難度測驗目的在度量某人能回答何等難度的題目；速度測驗目的在度量某人對於某一等難度的題目，在規定時間內，能回答幾個。譬如跳欄，難度測驗好像一行高度不一的欄，由低到高排列着，某人所能跳過的

最高的欄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。所以普通運動會裏的跳欄是一種速度測驗，跳高是一種難度測驗。

在編造速度測驗時，題目難度須相等，時間限制須從嚴，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，題目須難易不等，並依難度而排列，時間限制須從寬，在理論上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；一種是以工作為限制，以時間為單位。譬如度量閱讀速度，用第一種方法，可以計算在規定時間內能看完幾行，行數或字數便是計算的單位。用第二種方法可以計算看完一篇規定的文字須多少時間，分秒便是計算的單位。這兩種方法都可採用。不過在舉行團體測量時，祇能用第一種方法，第二種就不適用了。

(三) 調查測驗與診斷測驗 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，在診斷一人缺點之所在，在那幾方面或那幾個步驟上，已經了解透澈或練習純熟。那幾方面或那幾個步驟，尚未明了或嫻熟。所以調查測驗僅可為團體比較之用，而診斷測驗，才是教學的利器。診斷測驗的編造，必須根據教材詳盡的分析。

## 第二講 測驗發展的歷史

歷史是有繼續性的。任何運動的歷史，可以追溯到很遠的時代。但為敘述便利計，我們不妨規定一個敘述的起點。在測驗運動中，普通公認法國心理學者比納為智力測驗的創始者。美國教育學者來斯（M. Rice）為教育測驗的創始者。我們現在敘述測驗發展的歷史，不妨從比納來斯說起。

### 一 智力測驗的發展

一、初期的嘗試 在比納發表第一個智力量表以前，心理學者對於心理測驗的編造，已有許多的嘗試。一八九〇年卡泰爾（J. Mark Cattell）在英國雜誌心理上發表一篇文字，題曰「心理測驗與測量」，提出十種測驗，並力主測驗手續標準化和成立測驗常模的重要。在該論文之末，附有英國優生學者高爾登（F. Galton）的家語，由此可見二人之關係。同時，美國波爾敦（T. L. Bolton），克伯崔克（E. A. Kirkpatrick），克萊（R. L. Kelley），諾斯維塞（N. Norworthy），英國柏推（C. Burt）及比納本人都編造和應用過許多心理測驗。費里門（F. N. Freeman）對於這時期心理學者的工作，有下列幾條結論：

（一）這時期心理學者對於心理測驗偏於理論的探討，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。

(三) 這時期的測驗在編製方面，施行方面，都缺乏嚴密的標準。測驗的效度和信度，少有人注意。

(四) 這時期的測驗大多屬於感覺與動作過程的測驗。

二、比納的貢獻 第一個智力量表是比納和西蒙在一九〇五年提出的。這個量表雖極簡單，但有幾點值得注意的。第一、這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不同。第二、量表中大部分測驗屬於所謂高層心理過程，與以前心理學者所用的感覺和動作過程測驗性質不同。第三、這三十個測驗，依照難度而排列，所以成爲一個難度量表。這時候，比納雖然還沒有用年齡單位表示智力的高低，還沒有把測驗依其難度歸入年齡組別，但是這種概念確已粗具形態。

一九〇八年比納和西蒙發表訂正智力量表。這個量表的最重要點，就是測驗都依難度歸入相當年齡組別，測驗結果亦以智力年齡表示之。計算智力年齡的方法，比納有大致的規定。這是第一個年齡量表。一九一一年即比納去世之年，他又發表一種訂正量表。這量表與一九〇八年的沒有什麼重要的區別，不過根據應用的結果，增加了若干新測驗，廢棄了若干舊測驗，對於測驗的位置亦有所變動。每個年齡的測驗數目，一律改爲五個，如此在計算智力年齡時就便當多了。

到如今，比納西蒙的智力量表已經過許多人的修訂，在方法上、組織上，已有許多改進，他們原來的量表已漸漸失去實用的價值了。但是後人的工作都是建築在前人所留下基礎之上。我們現有的智力測驗



基本的概念仍舊是比納所賜。我們可以引賓特納（R. Pintner）的話「在心理學史上，假使我們稱馮德（Wm. Wundt）為實驗心理學的鼻祖，我們不得不稱比納為智力測量的鼻祖。」

三、比納智力量表的修訂 比納智力量表發表以後，引起各國心理學者極大的注意。英、美、德、意、日本、中國、蘇聯、瑞典、土耳其都有比納智力量表的修訂本。在中國有陸志韋兩次的修訂。在美國有高達德（H. H. Goddard）推孟（L. M. Terman）庫爾門（K. Kuhlmann）葉格斯（R. M. Yerkes）黑林（J. P. Herring）諸人的修訂本。前三人仍採用年齡量表的組織，後二人採用積分量表的方法。年齡量表與積分量表的不同，在量表中各測驗的組織或排列的方法。前者將測驗歸入相當年齡組別，測驗結果直接以年齡分數來表示；後者將量表中各測驗依照難度排列，測驗結果以積分表示，然後再用年齡常模來解釋。從便利一點着想，積分量表似較勝於年齡量表。積分量表的組織較有彈性，測驗標準的修訂，亦較便易。我國的修訂本在第四講有詳細的介紹。美國各種修訂本中，以推孟的司丹福修訂本為最完善而應用最廣，茲簡單介紹於後。

推孟是美國司丹福大學心理學教授，所以他的修訂本稱為司丹福修訂本。推孟費了五年的工夫，經過廣博的試用，在一九一六年才發表他的修訂量表。量表中共有九十個測驗，其中五十四個是比納量表中原來有的，三十六個是新採入的。對於測驗的位置，根據試用的結果，有許多變動。測驗的數目，從三歲組到十歲組，每年有六個測驗，十一歲從缺，十二歲有八個測驗，十三歲從缺，十四歲有六個測驗。以上還有兩

個階段，推孟稱之爲普通成人階段，優秀成人階段，各有六個測驗。

推孟在修訂時所用的方法，有幾點值得提出來一談的。第一，量表標準化時所根據的兒童，力避選擇的作用。第二，每個測驗的效度，均經下列方法決定。將每個年齡的兒童，依其智商分爲三組，智商在九十以下者爲一組，九〇至一〇九者爲一組，一一〇或以上者爲一組，然後求三組兒童通過每種測驗者之百分比。假定這三組兒童通過某一種測驗之百分比，依智商之高低而遞減，那麼這種測驗可認爲有效的智力測驗，不然就須淘汰。第三，測驗的排列和記分的標準，務須使各年齡兒童的平均智商與其年齡相等，不然則測驗的地位須加變動，或記分的標準須加改變。第四，推孟對於主試者的說明及記分的方法有極詳細的規定。

此外推孟還有一個很大的貢獻，就是智力商數的引用。智力商數就是智力年齡與實在年齡的比例。智力年齡的解釋要根據實在年齡，這一點比納亦曾見到，但沒有提出具體的方法，以表示二者之關係。比納祇說，一個兒童的智力年齡比實在年齡低一歲或兩歲，他就是比較愚笨的兒童。但以二者之差數來表示聰明程度，不是準確的方法，因爲同一差數，因在生長曲線上地位之不同，其意義亦隨之不同。例如甲兒智力年齡四歲，實在年齡三歲，智齡超過曆齡一歲，乙兒智力年齡十一歲，實在年齡十歲，智齡超過曆齡亦是一歲，但甲乙二人之聰明程度，實不相等。據我們現在所知，甲兒實較乙兒聰明。智力商數的公式爲：

$$\text{智力商數} = \frac{\text{智力年齡}}{\text{實在年齡}} \times 100$$

則甲乙兩兒之智商應爲

$$\text{甲兒智商} = \frac{4}{3} \times 100 = 133.3$$

$$\text{乙兒智商} = \frac{11}{10} \times 100 = 110.0$$

甲兒比乙兒聰明多了。

最早提出商數之方法者爲德心理學者斯坦恩 (W. Stern) 鮑勃太格 (O. Robertag) 及美心理學者庫爾門。斯坦恩名此商數爲「心理商數」推孟改稱爲「智力商數」。

在一九三七年即第一次斯坦福訂正量表發表後二十一年，推孟又發表第二次修訂本。第二次修訂工作費了推孟十年的工夫。共事者有斯坦福大學心理學教授滿利 (M. A. Merrill) 等。

據二十餘年應用的經驗，第一次訂正本缺點，其重要者有下列數點：(1) 量表兩極端的取樣頗有誤差，尤其是四歲組以下，普通成人組以上。(2) 第一次訂正量表中若有若干測驗效度欠高，計分繁複容易練習。(3) 施行說明及計分指導不夠清楚具體。(4) 缺乏複份量表。第二次修訂的目的就在改進上述種種缺陷。

第二次訂正本的特點：(1) 共有量表兩份，可以替換應用兩份的難度、難度距離、效度、信度均相等。(2) 量表的全距比較第一次訂正本更加擴充，應用的範圍更廣。(3) 第一次訂正本有九十個測驗，第二

次訂正本每份量表有一百二十九個測驗，分配如下：從兩歲到五歲（每半年一組）共七個年齡組別，每組有七個測驗，其中一個係替換測驗；從六歲到十四歲共九個年齡組別，每組有六個測驗，另外普通成人組八個測驗，優秀成人組 I, II, III, 各六個測驗。（4）低年齡增加非文字測驗並富有興趣的材料。（5）施行說明及計分標準更加詳細具體，使所得結果更加客觀可靠。（6）修訂所根據的被試者的選擇務求避免取樣上的誤差，如被試者地理的分布、家庭職業、年齡與學級的關係等等因素無不顧到。（7）除智齡及智商兩種分數仍沿用外，又編製了智商與標準分數（Standard Scores）對照表，以備從事研究工作者的應用。（8）計算高年齡智商的方法與前不同。推孟在發表第一次修訂量表時，規定凡實在年齡在十六歲以上者，求智商時公式中的除數概用 16。第二次修訂本規定計算智商的方法如下：十三歲以下用實在年齡作為除數，十六歲以上概以 15 為除數，年齡在十三歲至十六歲之間的兒童，以 13 加上超過 13 歲的月數的  $\frac{2}{3}$  為除數。譬如十四歲兒童以十三歲八個月為除數，十四歲半的兒童以 14 為除數，十五歲兒童以十四歲四個月為除數。

據推孟自己說這種計算方法對於智力生長逐漸停止的事實較舊法為適合。賽門司（P. M. Symonds）對於推孟第一次修訂量表規定計算高年齡智商的辦法，亦以為智力生長到十六歲驟然停止的假定，是不合乎事實的。（註一）我們以為賽門司的話頗有討論之餘地。推孟計算智商，在十六歲以下，用實在年齡為除數，十六歲以上用十六為除數的辦法似並不包含智力生長驟然停止的假定。智力發育的速

率從十三歲（假定的）逐漸減慢的事實，在選擇十三歲以上各年齡測驗時，應該已經顧到。換言之，假定從十三歲到十四歲一年間智力發育的速率，比十二歲到十三歲一年間智力發育的速率慢，則十四歲測驗與十三歲測驗難度的差別，應該比十二歲測驗與十三歲測驗難度的差別小。如此則智力發育逐漸停止的事實應該已經充分顧到，並不需要其他方法來校正。推孟第二次修訂量表所規定計算高年齡智商的辦法實在是一種「嘗試與錯誤」的方法，並不一定在理論上有何根據。推孟自己說：「用嘗試與錯誤的方法，我們曉得用這種辦法（指計算高年齡智商的辦法）會同最後所採用計分標準，使高年齡兒童的智商分配相似低年齡兒童的智商分配。一個理想的量表似乎是應該如此的。」（註二）這問題關鍵所在，恐怕是量表中高年齡的測驗不能正確代表各該年齡兒童的平均作業。這一點推孟亦是承認的。他說：「我們要注意十三歲以上的智力年齡與十三歲以下的並不具有同樣的意義，因為這些智齡並不能代表各該年齡隨機羣體的平均作業。」（註三）

四、非文字智力測驗的編造 比納式智力測驗都是以文字語言為內容的；測驗說明用語言文字，被試者回答亦用語言文字。在理論方面，這類測驗有一個重要的限制，就是智力的表現，不限於文字的或語言的反應，完全用語言文字的測驗去測量智力，恐不免有偏。在實用方面，文字智力測驗，對於有文字語言困難的兒童，及聽啞的兒童，都不能適用。因為這種理論上的限制和實際上的需要，所以有非文字智力量表的編造。賓特納與派忒孫（D. G. Paterson）所編造的作業量表，鮑德氏（S. D. Porteus）的迷津測

驗都是非文字智力測驗的著名例子。

五、團體智力測驗的發展 比納式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，學校裏的老試都是團體測驗，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第一次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，事得其人。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頗多懷疑，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。

六、智力的因子分析 我們敘述智力測驗發展的歷史，不能不提及智力的因子分析一問題。智力測驗的發展大部分為適應實際上的需要。心理學者對於智力的性質及內容極少有系統的研究與分析。英

國心理學家斯比門 (O. Spearman) 在四十年前就發表他的智力兩因說。照他的意思，種種作業都靠着兩種因子，一種是普遍的，各種作業中多少有幾分；一種是特殊的，祇是一種作業中有，他種作業中沒有的。這是最早的智力因子分析。在普遍因子和特殊因子外，後來又有團體因子的發現。所謂團體因子，乃幾種作業所共有，而又不是一切作業所遍有者。在智力方面已被發現的團體因子，有記憶、語文、計算、知覺速率、動作速率、歸納能力、演繹能力、空間關係等。這種能力的分析，目的在尋求那些不相關的因素，以爲編造測驗的根據。從事這種研究的有斯比門和美國賽斯東 (L. I. Thurstone) 克萊 (T. L. Kelley) 諸人。這種能力分析的研究，現在雖然離完成的階段尙遠，將來對於測驗的影響一定是很巨大的。

## 二 教育測驗的發展

標準教育測驗的編造，雖然爲教育中較近的發展，但是標準測量的觀念可說由來已久。在一八六四年，英國 Greenwich Hospital School 裏有一位教師，名叫菲奢 (George Fisher) 曾搜集學生作業樣本，編成量表集 (Scale Book) 一書，用以爲度量學生作業的標準。書中備有學生各科作業的樣本，每一樣本定有一種分數。評判某學生某科之作業時，可將其作業與量表集中各樣本比較，以尋求與某生作業優劣相等的樣本，此樣本的分數，即爲某生應得的分數。如此則評定分數時有標準可循，不致漫無依據。菲奢這量表集表集時，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。

美國教育家邁恩 (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。

(一) 劃一的筆試比較個別的口試公平些，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。」

(二) 劃一的筆試比較個別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。

(三) 劃一的筆試比較個別的口試可靠些，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。」



(四)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。

一、來斯的貢獻 在十九世紀的末年，能力心理學尙爲一般心理學者所深信，而美國教育界人士尙株守於所謂學科之訓練的價值的學說。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爲不滿，時倡改革之論。同時學校以外人士對於學校株守成墨不知改進，亦多所攻擊。爲適應社會之需要，實用的學科漸漸的加入了小學課程。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科目了。

在這雙方爭論之中，來斯想到一個解決這種爭論的客觀的方法。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，其成績並不低於每天用四十分鐘學習拼法的學生！來斯的結果引起教育界很大的注意。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表示深切的懷疑。但是來斯的工作引起了少數有思想的教育家的注意與贊同，奠定了教育測驗運動的基礎。

二、桑戴克的貢獻。來斯所用的測驗並沒有經過嚴密標準化的手續，所以他對於編造標準測驗的技術並沒有什麼貢獻。他的貢獻乃在用客觀的方法去解決教育上的問題。美國教育測驗運動的中心人物應推桑戴克。愛里斯 (L. P. Ayres) 曾說過：「我們既稱來斯為教育測驗的創始者，則應稱桑戴克為教育測驗運動的鼻祖。」

一九〇四年，桑氏出版心理與社會測量一書，介紹統計的方法以及編造測驗的基本原理。一九〇九年冬，桑氏在美國科學促進會教育組提出其所編的書法量表，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，為測驗運動中極重要的事件。桑氏在編造書法量表時，引用卡泰爾之「等距原理」，後來的作文量表圖畫量表都是根據這個基本原理而編造的。幾種早期的標準測驗如斯東 (C. W. Stone) 的算術推理測驗，希來格司 (M. B. Hillegas) 的作文量表，柏金漢 (B. R. Buckingham) 的拼法量表，都是在桑氏指導之下完成的。此外早期完成的測量工具有愛里斯的書法量表，拼字量表和科底斯 (G. A. Courtis) 的算術測驗。

三、學校調查運動。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教育測驗。當時科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 (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。

四、較近的發展 以上所述都是測驗運動早期發展的情形。近二十年來美國測驗運動重要的趨勢，約有數端可述者。

第一、測驗範圍的擴展 早期的教育測驗完全限於小學階段，以後逐漸擴展到中學。到現在中小學各種科目莫不有相當測驗可以應用。據門羅（W. S. Monroe）在一九二八年估計，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。教育測驗屬於算術、讀法、語文、及歷史各科目者最多。每年所用測驗份數估計不在三千萬之下，其中大概四分之一屬於智力測驗，四分之三屬於教育測驗。據李氏（J. M. Lee）在一九三二年的估計，適用於中學的測驗約有六百餘種。美國測驗事業的發達亦可見一斑了。

第二、診斷測驗的注重 早期的測驗大都為普通調查之用，祇能作團體之比較，不能作個人之診斷。為求測驗對於教學發生切實效果計，測驗須具有診斷的功用。測驗學者對於診斷測驗的注意，代表着一種很健全的發展趨勢。

第三、性格測驗的發展 教育的結果原不限於智識和技能，興趣、態度、理想、德行、情感、欣賞都是教育的重要結果。這些結果比較智識技能來得抽象而難以捉摸，測量方法的發展亦比較落後。近二十年來，經過許多心理學者的研究與試驗，這方面的成就已頗有可觀。

第四、測量方法的擴充 近年來教育者，尤其是所謂「前進」教育者，對於測驗運動的批評，就是測量的範圍過於狹窄，偏重具體的而忽略抽象的教育結果。克伯屈（W. H. Kilpatrick）在二十多年前，曾明白指出這種流弊。

「現有教育測驗所測量的大都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，這些教師的教學效率，將根據他們學生的測驗成績而被考核。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。」（註四）

講到這一個問題，我們不得不承認「紙筆」一類所謂客觀測驗的方法，在應用上是有極大的限制。現在的趨勢是將測量的方法大大擴充。教育中測量工作的任務，是在供給可靠的證據，以顯示教育目標之已否實現。測量的方法並不限於紙筆一類的測驗，凡足以供給證據以顯示教育目標之已否實現者都是有效的測量方法。觀察、面洽、問卷、作品分析等方法，雖不如紙筆測驗之客觀，但確能供給紙筆測驗所不能供給的事實。在美國近年來「測驗」一名詞已被「考核」（Evaluation）起而代之，這並不僅僅是名詞上的變易，實代表測驗運動的趨向重要的轉變。

### 三 我國測驗運動的回顧與展望

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。民國十年廖世承陳鶴琴合編的智力測驗法出版，十一年費培傑把比納一九一一年量表譯成中文。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（W. A. McCall）來華主持測驗編造事宜，當時協進行者有北京師範大學、北京大學、燕京大學、北京女子高等師範學校、東南大學等校教授和學生，計其編成測驗四十餘種。十四年廖陳合編測驗概要一書，就是當時所編的各種測驗的彙編。

我國編造的測驗，最重要的特色，就是一致採用麥柯爾所倡的T.B.C.F.制，使由各測驗所得的分數，可以互相比較，應用時較爲便利。麥柯爾曾有一段話，評論中國各種測驗的價值（註五）

「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。因爲每種心理測驗，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：我們所造各種測驗，至少都與美國的標準相等，有許多種竟比美國的爲優。然這也不足爲榮，因爲對於一種繼承物，個人應該感激，個人應該想法子去改良。」

民國十二年冬舉行全國小學調查。(註六)該調查是由中華教育改進社聘請查良劍德爾滿(E. J. Terman)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，國內研究測驗學者組織之中國測驗學會正式成立，其宗旨為研究測驗學術，推行測驗方法，並刊行測驗雜誌，可惜該會現在已無形停頓了。

我國測驗事業，在民國十一、十二年之際，確樹下了很好的基礎。可惜提倡測驗的祇是幾位大學教授，對於各級學校始終沒有發生什麼影響。因此測驗運動的高潮，隨着麥柯爾的去華也就低落下去了。近十年來蕭孝嶸對於智力測驗性格測驗有許多重要的貢獻，艾偉新編了一套小學教育測驗，總算打破了測驗運動的岑寂。就麥柯爾來華時所編的四十幾種測驗說，祇有陸志韋訂正的個人智力量表，在民國二十五年有第二次的訂正，其餘測驗二十多年來都未曾加以修訂，而且大多數去，經絕版，連買都買不到了。考我國測驗事業「不景氣」的主要原因，就是提倡測驗的人與應用測驗的人中間脫了節。換言之，祇有大學教授在提倡測驗，而一般中小學行政當局及教師，尚不知如何利用測驗，以謀教學之改進。為補救這種情形，我們有下列的建議。

(一)測驗編好了，沒有人利用，也是徒然的。所以怎樣利用測驗以謀教學之改進，應該有人來指導。這種責任我們以為應該由各地方教育行政當局及各區國立師範學院肩負起來。美國測驗事業發展的一

種重要原因，是推廣和指導機關的設立。這種機關大別爲兩種。一種是附屬於地方教育行政機關的，一種是附設於州立大學的。以前江西省曾有智力測驗局的設立，上海市教育局曾有測驗統計科的設置，可惜不久因人事的變遷就取消了。

(二)測驗的編製，在時間上和金錢上，都是很耗費的。而且，一種測驗編好以後，並須隨時修訂。這種工作不是個人力量所及，而應由機關以團體的力量，聘請專家，長期研究。

(三)爲切於教學上之應用計，以後應注重診斷測驗，練習測驗之編造。

(四)從技術方面說，我國早期的測驗都是採用EBCT制的。T分數的最大缺點，就是不容易了解，使應用上發生相當阻力。陸志章第二次修訂比納智力量表，以及近來艾偉蕭孝蟻所編的各種測驗，都已放棄T分數而採用年齡分數或年齡年級的百分數常模。我們認爲祇少小學階段的測驗應該放棄T分數而成立年齡及年級常模。

註一 P. M. Symonds, *Measurement in Secondary Education* P. 320.

註二 L. M. Terman & M. A. Merrill, *Measuring Intelligence* P. 30

註三 同上

註四 W. H. Kilpatrick, *The Foundations of Method*, pp. 107-108.

註五 W. A. McCall, "Scientific Measurement and Related Studies in Chinese Education," *Journal of*

Educational Research, Vol XI, nos. 2, 3. Feb. & Mar., 1925.

註六 E. L. Terman, The Efficiency of Elementary Schools in China, The Commercial Press, 1924.



### 第三講 測驗的條件

一個完善的測驗應具備若干條件。這些條件在編製測驗或選擇測驗時均應顧到。

#### 一 信度

測驗的第一個基本條件，就是信度，或稱可靠度。用一種測量工具度量某種事物或現象兩次，（假定這種事物或現象固定不變）如兩次所得結果相等，這種測量可算是完全可靠的。不論物質的或心理的測量，直接的或間接的測量，都有信度的問題。

在心理測量方面，影響測驗結果信度的因素，可大別為兩類。第一類是屬於測驗本身的。標準測驗與普通考試最明顯的區別，就是標準測驗對於施行的手續和記分的方法，有嚴格的劃一的規定，使應用測驗者，有所遵循，而不致攙入個人的意見。於施行測驗的手續說，如主試者的說明及測驗時間的限制，均應有詳明的規定。主試者對於測驗做法或內容，多解釋一句或少解釋一句，都有影響測驗結果的可能。有時很微細的變易可發生重大的影響。例如美國應用最廣的司丹福比納智力表三歲組有一個問兒童性別的測驗。照規定說明，對於男女兒童的問句略有不同。對男孩應該問「你是男孩還是女孩？」對女孩應該問「你是女孩還是男孩？」這樣規定的緣故，是兒童對於別人的問話，有機械地複述末了一個字的趨

向。假如問一個女孩「你是男孩還是女孩？」回答是「女孩。」主試者就不能斷定她是否已能辨別性別。假如照此問句考驗一羣女孩，則答對的人數一定比較照規定問句多。所以編造測驗者和應用測驗者，都應該認識「說明」乃測驗的一部分，對於施行的手續，主試的說明，編造者應明確規定，應用者應嚴格遵守。

就記分的方法說，更爲舊式考試與標準測驗顯著的區別。舊式考試定分的主觀，已爲彰明較著的事實。斯達奇 (D. Starch) 及伊立歐忒 (E. C. Elliott) 將一本英文考卷請一百四十二位英文教員評閱給分，各人所定分數從五十分至九十九分不等，最低的和最高的分數竟有四十九分之差。他們又將一本幾何考卷請一百四十五位數學教員評閱給分，各人所定分數從二十八分至九十二分不等，最高與最低的分數，相差更巨。

吳德 (R. D. Wood) 曾述及一件很滑稽的事：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。其中有一位教授爲評閱便利起見，自己做了一本卷子，以爲評閱定分的標準。不知怎樣，這本標準卷子同其他考卷錯混在一起。另一位教授照例的拿起這本卷子來評閱，竟給了一個不及格的分數。爲審慎起見，其他教授對於這本不及格的卷子，重加評閱。結果這本標準卷子所得的分數，最低爲四十分，最高爲九十分！

舊式考試定分的主觀，主要的原因，是問題的方式對於被試者的反應不能有嚴密的控制。標準測驗

所採用的問題方式，如是非式、選擇式、排列式、配合式，對於被試者的反應，均有嚴密的限制，記分就比較客觀了。

在司丹福比納智力量表四歲組有一個畫方形的測驗。兒童畫的方形各式各樣，何者可算及格，何者算不及格，不易評定。為增加記分的客觀性起見，編者備有記分標準片，上面畫着許多方形，有的認為「不及格」的，有的認為「及格」的。記分時可拿着記分標準片來比較，記分者個人的主觀就可減少。

影響測驗結果的信度第二類因素是屬於被試者方面的。物體的測量，測量的對象比較固定不變，祇要測量工具可靠，觀察準確，兩次測量的結果就可以一致。一塊石頭今天重三百斤，過了幾天或幾個月，仍舊重三百斤。一幅布，今天長一丈二尺，過了幾個月仍舊長一丈二尺，兩次測量的結果可以相同。但是一個學生今天能在一小時內做對十個算術題目，到了明天，在一小時內也許祇做對九個難度相等的題目。一個人的作業成績極易受疲勞、睡眠、疾病、光線、溫度、嘈聲和心理態度等因素的影響，所以是時時變易的，而測驗結果的信度就因此低落了。

就這一點說，物體的測量和心理現象的測量，亦不過是程度上的差別而已。一個人的作業成績固然常常變動，但是無生命的物體，也受環境的影響而發生變化。例如鐵路上鋼軌遇熱即漲，遇冷即縮，何嘗不起變化呢？所不同者，物體所起的變化較為微小，而心理現象所發生的變化較為巨大，能引起物體變化的因子較少，能引起心理現象變動的因子較多而已。

這一類因素固然與測驗本身無關，但主試者亦可加以相當的控制。如舉行測驗的環境須安靜自然，舉行測驗時要引起兒童的合作，要避免兒童的疲勞。

測驗結果的信度常受上述兩類因素的影響，而第二類因素卻與測驗本身無關，因之測驗本身的信度常略高於測驗結果的信度。

上面我們討論信度的性質和影響信度的因素。

一個測驗的信度怎樣考查呢？在心理測量中，我們所注意的往往是個人之間的比較。用一種測驗測量一班兒童兩次，假使這班兒童兩次測驗成績比較的地位沒有變動，那麼這測驗可算是完全可靠的。變動愈多，則測驗的信度愈低。要考查兩次測驗成績比較的地位有無變動，可用統計學中相關的方法，求兩次測驗成績的相關度。統計學中表示相關度的量數為相關係數，在這個場合，稱為信度係數，以英文字母  $r$  代表之。 $r = 1.00$ ，代表完全相關， $r = 0$  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。相關係數又可為負的，但負的信度係數可說是絕無僅有的。

上面所說求信度係數的方法，即將一種測驗應用於一班兒童兩次，求兩次測驗成績的相關實在不是妥善的方法。因為第一次測驗時所犯的錯誤，在第二次有重複的傾向，使所得信度係數較實在的信度係數價值略高。所以除非是單份測驗，求信度係數時不應用此法。

比較妥善的方法有兩種。假如一種測驗有幾個相等的複份，那麼可以用一班兒童在兩個複份上所

得分數的相關，來表示測驗的信度。另一種方法是將測驗分爲兩半，（普通的手續是將偶數的題目歸入一半，奇數的題目歸入另一半）求兩半之間的相關度，然後用斯比門白朗（Spearman-Brown）推測公式求整個測驗的信度。

信度係數的價值既然有高低，那麼信度係數要怎樣高才可使我們滿足呢？這是要看測驗結果的用途而定的。照克萊（註一）的提議，爲團體比較起見，測驗的信度係數不應低於50，爲個人診斷起見，不應低於94，因爲平均分數的信度比個別分數高。我們還須注意克萊所定標準是指根據一個年級的兒童所求得的信度係數，因爲同一測驗的信度係數依所根據的團體的差異度而異的。所根據團體的差異度愈大，則所得信度係數愈高，差異度愈小，則信度係數愈低。

## 二 效度

測驗的第二個基本條件就是效度。如一種測驗確能測量他所欲測量的現象，這種測驗可謂之有效。我們可以想到直接的測量如身長體重的測量，並無效度的問題，而間接的測量便會發生效度的誤差。所謂間接的測量，就是說以一種事物或現象的度量，替代另一種事物或現象的度量。例如觀察寒暑表上水銀的升降，以度量溫度，就是間接測量了。但這種間接測量卻是完全正確的，換句話說，並無效度的誤差，因爲溫度和水銀高度之間，有近於完全的相關。假如以兒童身長的度量，來代替他們體重的度量，就有效度

的誤差，因為身長與體重之間相關並不完全。

心理現象與教育成績的測量都屬於間接測量一類，所以都有效度的問題。要考查一種測驗的效度，需將測驗結果與一種有效的標準比較，觀察其間之相關，相關之深淺即表示測驗效度之高低。這種相關係數專門名詞稱為效度係數。

但要考查測驗的效度，這種有效的標準卻往往不易求得。譬如要考查一種智力測驗的效度，需先求得智力有效的標準。但我們正因為沒有準確的方法以度量智力，所以才編造智力測驗。假使已經有了其他有效的標準，那就沒有編造測驗的必要了。在這種情形之下，我們祇能用間接的方法以考查測驗的效度。一種常用的方法是看各年齡兒童能做對測驗的人數的百分比是否逐年增加。我們知道在發育期間兒童的智力是逐年增進的，所以一般六歲兒童的智力比五歲兒童高，五歲兒童的智力比四歲兒童高。假使某種測驗真能度量智力，則六歲兒童能做對這測驗的人數的百分比應較高於五歲兒童的百分比，五歲兒童能做對的百分比，應較高於四歲兒童的百分比。換言之，一個有效的智力測驗應能鑑別智力高下不同的兒童。這種從測驗的鑑別力以決定其效度的方法，不惟適用於智力測驗，亦適用於他類測驗。譬如有效的教育測驗需能鑑別學業優良與學業低劣的兒童，有效的情緒測驗須能鑑別情緒穩定與情緒反常的兒童，有效的品格測驗須能鑑別品格優良與品格惡劣的兒童，有效的興趣測驗須能鑑別興趣不同的兒童。鑑別力愈大，則測驗的效度愈高。

考查測驗效度另一種間接的方法，是將新編測驗的結果與心理學者所公認為有效的測驗相比較。例如美國斯丹福比納智力量表常被採用為考查新編智力測驗效度的標準。

有時，這種有效的標準不難直接尋求。譬如一個百貨公司要招請五十名售貨員。我們可以編造一種測驗以為選擇的工具。但編就的測驗究竟能否鑑別售貨能力的高低，換言之，這種測驗的效度如何，一定要試用以後才能知道。我們可以將這五十名售貨員應考時在測驗上所得的分數保留起來。在一月或半年以後計算各人售出貨物的件數或價值，這便是售貨能力的標準。然後求測驗分數與售貨成績的相關。換言之，應考時在測驗上分數較高的人，售貨成績是否較優。這種關係愈深，即證明測驗的效度愈高。總之，測驗的效度不能取決於主觀的評斷而應有事實的證明。

以上就一般測驗說明效度的意義。現在專就教育測驗的效度再加申述。

(一)教育測驗須與教育目標相符合。測驗是一種教育的工具，用以完成教育目標的。所以測驗的範圍須能包括各項重要的教育目標，而不應有偏闕。我們不容諱言，現在發展最完善的測驗是偏於智識技能一方面，而其他同等重要或更為重要的教育結果，如品格、理想、態度、興趣、思考能力、判斷能力、合作能力等，我們還不能便易地、圓滿地去測量。假如我們專以學生在智識測驗技能測驗上的成績，去考核教學的效果，則智識技能實有被過分重視的危險；同時較抽象而難以測量但更為重要的教育結果，實有被忽略的可虞。

就個別學科測驗而論，一種學科測驗須與該學科的教學目標相符合。

(二)測驗內容須能代表教材的全部。四五十分鐘的考試或測驗總不能包括一學科教材的全部，換言之，總是一個有限的樣本。我們所能做到的，是使得測驗內容能公允地代表教材的全部。要做到這一點，就要研究測驗材料取樣的方法。

測驗材料的取樣可有兩種方法。我們可以稱之為深入的取樣，及廣博的取樣。普通學校考試題目不過四五十個，但對於每一個題目卻希望學生回答得很詳盡，這是採用深入的取樣。標準測驗題目少則四五十，多則一二百，每題範圍較窄，但教材的全部都能顧到，這是採用廣博的取樣。從考試或測驗的效度一點講，廣博的取樣實較優於深入的取樣。

另一種取樣的方法，尤其適用於技能測驗的，是採用所謂標型的材料。這種方法假定一種技能的學科包含許多特殊的單元或過程。因此測量一個兒童對於這許多特殊單元的能力，就可以測量到他對於全部學科的能力。譬如算術的四則包含許多特殊的技能單元，如加法的進位，減法的借數。一種取樣周全的四則測驗應包含這些單元在內。

從上面的討論，我們可以想到，如其他情形相等，一個測驗題目愈多，則取樣愈周。普通學校考試的結果往往不能代表學生真正的學力，一部分原因是試題數目太少，材料取樣不周。

總起來說，要使測驗內容取樣周全，應該(1)採用廣博的取樣方法，(2)採用標型的材料，(3)增加



測驗的長度。

(三)測驗題目須有適當的難度。很明顯的，凡全班都能做對或都不能做對的題目，是毫無價值的，因為這種種题目的鑑別力恰等於零。在難度測驗中，題目難度的分配是很重要的。測驗中題目所代表的難度等級愈多，則其鑑別力愈大。

測驗题目的難度，要看一班學生答對人數的多寡而定。但一個题目的難度並無絕對性，常依被試者所處環境所受教育（教材的排列及教學的側重點）而異的。譬如城市兒童能回答「電車怎樣會走的」一定比鄉村兒童多些；反過來說，能回答「大麥和小麥有什麼分別」，鄉村兒童一定比城市兒童多些。

(四)測驗與實際生活情境愈接近愈好。譬如要考查兒童默字的能力，我們可以（1）從他們的作文、日記、信札，考查錯字的數量；（2）從默寫句子考查錯字的數量；（3）從默寫單字考查錯字的數量。第一種辦法最自然，最接近實際生活情境。第三種辦法最不自然，最接近實際生活情境。所不幸的，測驗情境愈是接近實際生活，愈是繁複，而難於控制。

(五)測驗須盡力除去不相干的因子。所謂不相干的因子，就是與測驗所測量對象不相關的因子。例如，算術測驗的文字應極簡易，不要使兒童因文字上的困難而影響測驗的成績。難度測驗的時限，應該從寬，否則就變成速度測驗了。測驗不應過分冗長，使兒童發生疲勞。測驗的做法不應複雜，使兒童不知如

何着手。

在結束效度的討論以前，我們應該提及一個重要的概念，就是一個測驗的效度，是對某一個羣體而說，並無絕對性。對某羣體有效的測驗，對另一羣體未必有效。應用於某種情境下有效的測驗，在他種情境下，不一定同等有效。譬如美國的智力測驗，不加修訂，不能適用於我國，因為兩國兒童環境迥異。二十年前編造的教育測驗不能適用於現在，因為課程標準已經屢次修改。再如對於居住城市的四五歲兒童，「汽車與電車誰走得快」一試題，或者是一個有效的智力測驗，對於鄉村兒童就不適用了。這樣說來，測驗的效度是依被試者的環境背景與經驗而異的。

### 三 常模

效度和信度是測驗主要的條件，此外還有許多次要的條件，常模是次要條件之一。

常模普通有兩種，就是年齡常模和年級常模。年齡常模就是各年齡兒童在測驗上所得的平均分數。年級常模就是各年級兒童在測驗上所得的平均分數。此外還有百分數常模和標準差常模。這兩種常模常以一個年齡或一個年級團體為根據。所以我們可以有年齡團體的百分數常模、年級團體的標準差常模、年級團體的百分數常模。這四種常模告訴我們在測驗上得某分數，相當於某一年齡或某一年級團體中什麼地位。表一表二是蕭孝隣訂正墨跋智力量表百分數常模表和標準



差常模表之一部分。譬如一個兒童曆齡二十五個月得分數9分我們在表一曆齡一直行中先找出這個兒童的曆齡，然後在這橫行中找出他的分數，此分數相當於百分位置數20，這就表示他在同年齡兒童中居百分之二十的位置。換言之，在同年齡兒童中成績不如他的有百分之二十，成績相等或超過他的有百分之八十。又譬如一個兒童曆齡三十個月，分數35，我們在表二「曆齡」一直行中，先找着這兒童的曆齡，然後在這橫行中找出他的分數，此分數相當於標準差值1.0，表示他在同年齡兒童中所居的位置是在平均點以上一個標準差。常模的功用在解釋測驗的結果。一個兒童或一班兒童的分數可與常模比較。常模供給我們一個比較的標準。

就年齡常模與年級常模比較而論，大部分智力測驗及心理測驗採用年齡常模，大部分教育測驗採用年級常模。大致說來，年齡常模較年級常模為穩定，因為年級常模常受學級編製的影響。各校各年級學生的平均年齡不同，將一級成績與年級常模比較時，如不顧到學生的年齡，很容易引起錯誤的解釋。年齡常模還有一點便利，就是可直接與曆齡比較，計算各種商數，如智力商數、教育商數等。年齡常模和年級常模較適用於小學階段，在中學無多大用處。常模的代表性是很重要的問題。在求測驗常模時所測量的人數總是有限的，以一部分代表全體，就發生取樣的問題。要使得一部分可以代表全體，第一，在求測驗常模時必需顧到「隨機取樣」的原則。換言之，取樣宜公允，不應有偏。第二，在隨機取樣原則之下，常模的代表性與常模所根據的兒童人數成正比，人數愈多，常模的代表性愈大。

年齡團體的隨機取樣，實際上是可望而不可即的。測驗的常模大都以入學兒童為根據的。以入學兒童為根據的年齡常模，在低年齡與高年齡兩端，往往缺乏公允的代表性，因為八歲以下的兒童，一部分尚未入學，十二歲以上的兒童，一部分已離開學校。因為這種困難，周學章（註二）曾提議求年齡常模時，索與以在校兒童為限，未入學或已離校的兒童不必加以顧慮。為名實相符計，這種常模可稱之為「在校年齡常模」。

常模因所根據團體的性質，有全國常模、地方常模、種族常模、性別常模等之分。現在國際性的測驗尚無人嘗試，兩性之間據各種測驗應用結果大致上並無顯著的差別，所以種族常模和性別常模，我們可以略而不論。就全國常模與地方常模比較，似以後者較為切用，尤其我國區域如此廣闊，而各地方教育程度又相差懸殊，以某地方某學校成績與全國常模比較似無多大意義。

各種測驗的常模除非根據相似的團體，往往不能嚴格比較。所以供給常模者，應將所根據團體詳加敘述，使應用常模者有斟酌的餘地。

再者，學校課程教材的變動，自然要影響到教育測驗的常模，所以常模有時時修訂的必要。

在結束本節以前，常模與標準的區別應加說明。常模代表實際上平均的成績，標準代表理想中應有的成績。常模固可幫助標準的確定，但二者卻不可混為一談。賽門司（註三）曾提出四項原則，在規定標準時應該顧到的：（1）規定標準應顧到各門學科或各項活動相對的價值。一種學科的標準不應定得太高，

致妨礙其他學科平均的發展。(2)規定一學科的標準時應顧到現在及將來其他學科的需要。譬如算術科的標準應顧到學習代數時所必需的算術能力，代數一科的標準應顧到學習三角、幾何或物理所必需的代數能力。(3)規定標準時應顧到社會的需要。譬如書法就一般人而論，應認爲一種工具的學科，所以書法的標準，祇要能滿足社會的需要就夠了。(4)規定標準時應顧到個人之間能力的差異。

#### 四 施行及校閱的便利

施行測驗及校閱試卷的手續愈簡單，則應用時愈便利，而所得結果亦愈可靠。每種測驗對施行及校閱的手續應有詳明的規定。關於測驗的說明，麥柯爾（註四）曾有下列的規定：

- (一) 測驗說明須簡單而清楚。
- (二) 說明中可利用例子以示測驗的做法。
- (三) 說明須劃一而適合被試者的程度。
- (四) 說明的次序應照着做測驗的次序。
- (五) 說明須分段落。
- (六) 說明須能引起兒童最大的努力。
- (七) 對兒童的說明外還須備有對主試的說明。

校閱的便利大致要看題目方式及排列的形式如何。編造測驗者祇要肯用心機，這一點是不難做到的。黃覺民（註五）曾有考卷校正器的創製已交商務印書館代為製造發行。

## 五 複份

為增加實際效用起見，一種測驗至少須有兩份，兩份「相等」的測驗。當然份數愈多，應用愈便利。例如意要考查一班學生在一學期中的進步，可在學期之初舉行一次測驗，學期之終再舉行一次，兩次測驗分數的差別可以代表一學期的進步。假使測驗是單份的，那麼同一份測驗用兩次，就難免練習的影響。測驗如有好幾份就可以替換應用，就可避免或減少練習的影響。

前面我們說測驗的各份須「相等」。兩份測驗如能適合下列條件就可以稱為「相等」。

(一) 各份測驗須測量同一能力。

(二) 各份測驗所用材料不應重複。

(三) 各份測驗的難度須相等。

(四) 同一班兒童在各份測驗上所得分數的平均及差異度須大致相等。換言之，同一分數在兩份測驗上的意義應相等。兩份測驗的相等性，當然要有事實的證明。

## 六 價格

測驗價格便宜，應用自可推廣。陳鶴琴所編小學默讀測驗，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。

註一 T. L. Kelley, Interpretation of Educational Measurements, p. 29.

註二 周學章，T B C F 制之檢討，測驗二卷二期。

註三 P. M. Symonds, Measurement in Secondary Education, chapter XIII.

註四 W. A. McCall, Measurement, Chapter VI.

註五 黃覺民考卷校正器的創製，教育雜誌第三十卷第七號。



## 第四講 智力測驗

### 一 個人智力測驗

一、陸志韋訂正比納西蒙智力測驗 本測驗由陸志韋在民國十一年主持編訂的。編訂時所測驗的學生約計一千四百人，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。所測驗的學校都在江浙兩省。

本測驗共有六十五個試題，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，其餘的是新加入的。測驗的性質大概可說是仍前人之舊。一大部分仍舊用文字，且用問答體。

本測驗放棄了年齡量表的形式而採取績分量表的組織。先求測驗上應得總分數再從對照表求T分數，B分數。本測驗施行的手續及計算分數的方法，可參閱陸志韋訂正比納西蒙智力測驗說明書。（商務出版）

民國二十五年陸志韋吳天敏發表第二次訂正本。這次修訂的動機，據修訂者自己說，可有三點：（一）舊本裏吩咐兒童的語句有些不能普遍了解的地方，爲打破地方性而使測驗可以普遍應用，這些語句有加以修改的必要。（二）舊本中一部分試題，或受時間性的限制，於現在社會情形不相適合，或帶着地方色

彩太重，應加修改。(3)根據以往應用的經驗，兒童智商的高低常與其年齡成反比。換句話說，年齡愈大的兒童智商愈低，年齡愈小的兒童智商愈高。這種情形大約由於低年齡的試題太容易而高年齡的試題太難了。

在編制方面，第二次訂正本又放棄了T量表而重新採用年齡量表的組織，測驗的結果可以直接用智力年齡來表示。另外又備有績分量表，可以從兒童在測驗上所得總分數求出他的智力年齡。

這次訂正本共有五十四個試題，一部分是舊編裏留下來的一部分是從推孟的修正本裏採選來的。測驗的性質大約可分為語言文字、數目、解圖和機巧四類。五十四個試題之中有能以不同之標準應用於不同之年齡者，故五十四個試題實際上可按七十五個試題應用。

本測驗自三歲至十一歲，每歲有試題六個，每題代表智齡二個月；十二及十三歲各有三個試題，每題代表四個月；十四及十五兩歲共有六個試題，每題代表四個月；十六十七及十八歲（即成人組）共有九個試題，每題代表四個月。

本測驗自六歲至十四歲各年齡試題的選擇和排列，係根據修訂本應用所得結果而定，六歲以下和十四歲以上各年齡試題排列的位置，並非以修訂者自己所得結果為根據，乃參照其在舊本中的位置或在美國比納測驗上的位置而定的。所以本測驗應用的年齡是從六歲（足歲）至十四歲。六歲以下或十四歲以上的兒童，雖然也可以勉強用，但恐所得成績難以準確。

本測驗的成績可以按智慧年齡計算，也可以按積分計算，同時兩者可以互相轉變。求得智慧年齡則可照普通公式求智商。但計算高年齡兒童的智商時，應該用什麼除數，修訂者未加規定，殊使應用者無所適從。

本訂正本所根據的受試者共六百三十三人，其中男的三百二十七人，女的三百〇六人，年齡從六歲到十四歲；除十四歲以外，每一歲受試者最少七十人，其中男女各以三十五人為最少。這六百三十三人完全是北平城內小學校的學生。從這些兒童的籍貫和家境看來，修訂者「認為此次所選擇的兒童是很滿意的，是可以代表中國一般兒童的。」

但是這次修訂本所根據的兒童恐仍難免取樣的誤差，因為六七歲的愚笨兒童尚未入學，而十三十四歲的聰明兒童已經畢業了。又為求得公允的取樣起見，在選擇兒童時，應顧到其年齡與所在年級的關係，因為據已有結果，同年齡的兒童在高年級的，比在低年級的聰明些。修訂者對於這一點已經相當顧到。他們所規定各年級兒童年齡的高限是：一年級兒童不得超過八歲十一個月，二年級不得過九歲十一個月，三年級不得過十一歲三個月，四年級不得過十二歲三個月，五年級不得過十三歲六個月，六年級不得過十四歲六個月。這種限制確實可以去掉一部分「落伍」的兒童，但在高年級顯然限制太寬（十二歲在四年級的，十三歲在五年級的，十四歲在六年級的都未去掉。）而低年級又無年級太高的限制，所以最後據以重訂標準的那些兒童，低年齡的不免是較聰明的，高年齡的不免是較愚笨的，不能算是最公允

的代表。

編訂年齡量表最重要的工作，便是各測驗年齡標準的決定；換言之，各種測驗應該放置在什麼年齡組別裏而使各歲兒童的平均智商平均都在一百。編訂年齡量表最成功的推孟是完全用理論上毫無把握的試誤方法，才收到這種效果的。這次重訂者採取了渥鐵斯的理論，用百分之五十及格為各測驗的年齡標準。渥鐵斯的理論已有人指出其破綻，是不適用的。重訂者採用百分之五十的標準，結果各歲兒童的平均智商除去十二歲以上，都似太高，平均約高半歲。重訂者乃將各年的測驗往下降了半年或一年的標準，以謀補救，但所得結果，似仍未能令人滿意。

茲將第二次訂正本各測驗錄後：

### 三歲組

- (一)比較線的長短
- (二)說出自己的姓
- (三)數銅元四枚
- (四)說出自己的年齡
- (五)指出身上各部
- (六)順背數目(四個字)

### 四歲組

- (一)辨別形式
  - (二)數銅元十三枚
  - (三)說出顏色
  - (四)說出物名
  - (五)三種差遣
  - (六)畫方形
- 五歲組

- (一)問手指數
- (二)上午與下午
- (三)倒數數目二十至一
- (四)三角形拼成長方形
- (五)說明性別

(六) 簡單的迷津

六歲組

- (一) 解說圖畫 (二) 學說語句 (三) 順背數目 (六個) (四) 有意義記憶 (五) 方形分析 (六) 指出謬誤

七歲組

- (一) 作語句 (二) 對較 (三) 指出圖中缺點 (四) 找尋失物 (五) 對答問句 (六) 解決問題

八歲組

- (一) 有意義記憶 (二) 默寫 (三) 找尋數目 (四) 方形分析 (五) 學說語句 (六) 說出日期

九歲組

- (一) 找尋圖樣 (二) 正確答案 (三) 填數目指使 (四) 描畫圖樣 (五) 順背數目 (七個) (六) 心算

十歲組

- (一) 詞句重組 (二) 學說語句 (三) 對答問句 (四) 有意義記憶 (五) 數學巧術 (六) 剪紙

十一歲組

- (一) 時辰計算 (二) 解決問題 (三) 填字 (四) 數立方體 (五) 倒背數目 (五個) (六) 比喻

十二歲組

(一) 方形分析 (二) 複雜迷津 (三) 匣子計算  
十三歲組

(一) 幾何形分析 (二) 默寫 (三) 指出謬誤  
十四歲十五歲組

(一) 歸納測驗 (二) 方形分析 (三) 數立方體 (四) 順背數目 (八個) (五) 倒背數目 (六個)  
(六) 有意義記憶

成人組

(一) 數學巧術 (二) 三物相同之點 (三) 幾何形分析 (四) 詞句重組 (五) 心算 (六) 複雜填字 (七)  
分別抽象字 (八) 理解問題 (九) 數立方體

關於編訂本測驗的經過，施行手續和記分標準，可閱下列二書：陸志青吳天敏第二次訂正比納西蒙測驗說明書（商務），吳天敏第二次訂正中國比納西蒙智力測驗之經過（商務）。沈有乾有以上兩書書評，載中國心理學報第一卷第四期。

二、蕭孝嶸修訂墨跋量表 墨跋量表 (The Merrill-Palmer Scale) 爲美人斯達次門 (R. Sturman) 所編製，專爲測驗幼稚兒童之用，係一種個人測驗，經蕭孝嶸於民國二十五年詳細修訂，並求得各項常模。蕭氏修訂本量表時所測驗兒童共一千五百三十一人，年齡從十八個月起至七十一個月止，男女

約各半。據修訂者估計，受試者所代表的社會階級並無過高或過低之弊。

本量表所包含的測驗可以分爲兩類：（一）全是全否的測驗，此係指受試者能做即能合格的測驗；（二）分數測驗，此係指可有各種時間分數或各種成績分數的測驗。此兩類測驗各有十六種，但因各種分數測驗可有多種不同標準，故十六種分數測驗共有七十二個測驗標準，加上十六個全是全否的測驗，全量表實際上共有八十八個測驗標準。這八十八個測驗依照難度排列而且是根據年齡分組的，從十八個月起至七十一個月止，分成九組，每組相距六個月，每組中測驗從八個到十二個不等。

本量表修訂時所測驗兒童的年齡，雖然是從十八個月到七十一個月，但據蕭氏自己申明，只有二十四個月至六十五個月的兒童，可屬於診斷範圍以內。倘若受試兒童的年齡在二十四個月以下或在六十五個月以上，他在此量表上所得的分數，只可視爲一種參考，並不能可靠地診斷他的智慧。

實施測驗時的手續或次序，是從符合受試者年齡的測驗做起。例如某兒童的年齡爲三十一個月，主試者便應從 30 和 31 個月年齡組中的測驗試起。倘若在這些測驗中有一部分不能通過，便應當嘗試年齡較低一組的測驗。倘若前組中的測驗完全通過便應當嘗試年齡較高一組的測驗，直至某一組測驗完全不能通過爲止。

本量表備有甲乙兩種表格：甲種表格是爲測驗時紀錄結果用的，乙種表格是爲轉錄已經記下的結果用的。記錄時所用的符號有四種：（一）正號「+」表示通過；（二）負號「-」表示失敗；（三）「○」表





(二) 在這個測驗以後共有十個「+」號，將10加入21，得31。這便是第一次總分。

(三) 再從基本測驗（即測驗21）看起，發現測驗27上有「R」符號。這個測驗的號數小於第一次總分（31），所以應該加上一分，得32。這是第二次總分。

(四) 測驗30上亦有「R」符號，其號數較小於第二次總分，所以應當加上一分，得33。這是第三次總分。

(五) 測驗31是一個免試的測驗，其號數較小於第三次總分，所以亦應當加上一分，得34。這是第四次總分。

(六) 測驗34上有「R」符號，其號數與第四次總分相等，所以應當加上一分，得35。這是第五次總分。

(七) 測驗43上亦有「R」符號，但其號數較大於第五次總分，所以不應加入總分內。

(八) 以後的測驗成績就不必再看了，因為以後如再有免試或拒絕的測驗，它們的號數必定較大於第五次總分。這個兒童的總分數為35。

本量表為一種積分量表。從各項測驗所得總分數，須加以解釋。為解釋總分數於量表也見。訂者備有下列幾種表：(一) 分數與智齡對照表。得了智齡，智商不難算出。例如上述兒童得總分數35，從對照表中查出35分相當於智齡35個月。該兒童智齡為40個月，以40除35，得88，即該生之百分。應用分數表示的標準差表。此表的用處在應用標準差以解釋分數的意義。知道了某生的智齡和總分數，便可直接從此表

上查出該生在同年齡兒童中所佔的地位，此地位以該年齡兒童分數分配的標準差表示之。(三)應用智齡表示的標準差表。此表的用處在應用標準差的單位以解釋所得智齡的意義。(四)應用智商表示的標準差表。此表的用處在應用標準差以解釋所得智商的意義。照蕭氏所得結果，同一個智商，對於各年齡兒童似缺乏一致的意義。換言之，智商的意義並無絕對的一致性。譬如一個二十五個月的兒童在本量表上得智商80，另一個六十五個月的兒童亦得智商80，這兩個兒童的聰明程度並不相等；因為同是智商80，但在各該年齡兒童的智商分配上的標準差地位並不相同。(五)分數與百分位置對照表。從此表可以查到一個兒童在同年齡兒童中的百分地位。以上所述各種解釋分數的方法，當然不必同時並用，根據需要採用一種或兩種就行了。

茲將本量表中各種測驗錄下，以示內容之一斑。

(一)全是全否測驗：(1)交足，(2)使木塊行走，(3)一足站立，(4)鏡內認識自己，(5)摺紙，(6)抽繩，(7)使用剪刀，(8)配合顏色，(9)握拳與移動拇指，(10)數二塊木頭，(11)畫圈，(12)拇指與他指相對，(13)仿繪十字圖形，(14)畫星(三試對一)，(15)畫星(全對)，(16)單字重述。

(二)分數測驗：(1)簡單問題，共有三個測驗標準，(2)字組重述，共有三個測驗標準，(3)鄔林(Walsh)樺板測驗A種，共有六個測驗標準，(4)鄔林樺板測驗B種，共有六個測驗標準，(5)相套立方體測驗，共有四個測驗標準，(6)十六塊立方體測驗，共有三個測驗標準，(7)二枚鈕扣，共有七個測驗

標準，(8)四枚鈕扣，共有四個測驗標準。(9)席弦形式板測驗，共有六個測驗標準。(10)小紅塔，共有四個測驗標準。(11)馬圖測驗，共有五個測驗標準。(12)六塊立方體的錐體，共有六個測驗標準。(13)畫迷一，共有六個測驗標準。(14)畫迷二，共有三個測驗標準。(15)人形測驗，共有兩個測驗標準。(16)德氏(Deroly)配偶測驗，共有四個測驗標準。以上共十六種測驗，七十二個測驗標準。

關於本量表修訂經過，施行手續及記分標準，可閱蕭孝嶸修訂墨跋量表（商務，二十六年一月）。

三、蕭宗機械能力測驗 本測驗係蕭孝嶸、亮東根據美國「明內所他機械能力測驗」(Minnesota Tests of Mechanical Ability)選編而成。「明內所他機械能力測驗」是美國許多種機械能力測驗中比較最精確的一種。蕭宗兩氏所選用的材料有下述三種：

(一)空間關係測驗 此項材料共有長方形木板兩塊，及形式不同的小木塊若干。受試者須將這些小木塊插入木板的孔中。每種木板應根據時間與錯誤記分。時間分數就是做完一種木板所需的秒數；錯誤分數就是受試者誤置木塊的次數。本測驗係一種個別測驗。

(二)集合測驗 此項材料共有三箱。箱內各格有機件若干種，都拆散為幾部分，受試者須在規定時間內將分散的部分配合為一種完整的東西。以A箱為例，共有機件九種：擴大螺旋帽、水管夾、紙夾、發條衣釘、鏈環、瓶塞、按鈴、腳踏車鈴、扁鎖。B箱有機件八種，C箱十六種。本測驗的記分係以各件東西集合正確的分數為根據。每件東西集合正確可得十分。倘若某件東西只有一部分集合正確而未能完成，亦可得一部

分分數。本測驗係一種個別測驗。

(三)紙形板測驗 此項材料共有圖形五十六個，每個有兩種形式，在左邊的爲一個完整的圖形，在右邊的爲分散的部分。受試者須按右邊的形式將左邊的圖形劃分爲形式相同的部分。本測驗是一種團體測驗。時間限制爲十五分鐘。記分的方法即是以做對多少爲應得的分數。

蕭宗二君曾用上述三種測驗，測量過初中三年級男生二百人。所得結果約有三點：

(一)根據測驗的結果，在十四歲至十八歲之間，機械能力的高低是與年齡成正比例的，即是年齡愈高成績愈好。

(二)機械能力測驗成績與學校勞作成績相關很低，(最高僅·一七，最低爲·〇八。)

(三)機械能力測驗成績與普通智慧能力(根據蕭孝嶸「同填測驗」)相關亦很低，(最高爲·四〇，最低爲·〇一。)

本測驗的說明及記分手續編者均有詳細規定，可參閱蕭孝嶸亮東機械能力測驗之標準化心理教育實驗專篇第三卷第一期，國立中央大學教育實驗所出版，民國二十五年六月。

四、胡毅木板構成測驗 本測驗係胡毅根據賓特納及派忒孫之作業測驗(Pintner and Paterson Performance Tests)及我國舊有益智圖之原則改編而成。測驗材料計有木板十塊，其中小方形一塊，小三角形兩塊，中三角形一塊，大三角形一塊，梯形二塊，長方形二塊，斜方形一塊。另有測驗圖十八張，上印

黑色圖形。受試者須在規定時間內，用小木板把紙上的黑色圖形蓋起來。每一圖工作時限爲三分鐘。編者將這十八個測驗，依其難度分爲六組，每組三個測驗。爲節省時間起見，測驗時不從頭做起而自第三組做起，然後第四、第五、第六、第二、第一組。進行中若某組三測驗全體能通過，以前各組皆免試作爲通過。若進行中某兩組接連全部未通過，以後各組均免試，作爲不通過。據試用結果，每一次測驗所需時間，連解釋練習在內，最多須四十分鐘左右。本測驗係一種個人測驗，測驗詳細手續尙待規定，常模亦未求得。

編者曾用本測驗測量過三百餘人，包含各種職業在內。根據試測結果，本測驗有下列各優點：

(一) 確能免除文字及教育影響。施行測驗時祇需用最少之語言，有時全不用亦可。受試者無須作語言或文字之反應。學校及社會教育對於本套測驗中工作只有間接影響。對於失學民衆、邊地民族、學齡前兒童，及文字上有缺陷者，其效力與普通學童或成人相等。

(二) 用具簡單。只須測驗圖一本，說明及記錄簿一本，木板一套（十塊），碼錶一只，鉛筆一枝。一切均可裝入普通公事皮包，重量約等於雜誌兩本，極易攜帶。

(三) 作法簡易。所有測驗，除第一係作練習用外，皆用一律之方法。不同點只在圖形之區別及所用木板之差異，但主試者對此二點完全不用費心。

(四) 應用寬廣。難易兩端俱能與最難最易相隔甚近。各年齡各職業之人皆可應用本測驗。所測三百餘人中，無一人全盤作出，亦無一人全不能作出。故用任何方法計分，皆無人得零分及最高分。

關於本測驗之編製及試測經過可參閱胡毅木板構成測驗編製及試測經過載中國教育學會第三屆年會報告，二十五年二月。

五、鮑德氏迷津測驗 迷津測驗為美國鮑德氏所編造，為一年齡量表，共三歲至十二歲迷津各一，十四歲迷津一，成人用迷津二。據編者自己說：迷津測驗的目的乃在測驗各人在具體的新情境下的設計能力、慎審和心理上敏捷的傾向。這種測驗雖然是測驗心理能力的，但是因為衝動、躊躇、暗示性、精神病和憤激性，能影響受試者的成績，所以這測驗大部分亦是測驗被試者氣質上的能力的。

上海市教育局在民國二十二年曾採用迷津測驗材料在京滬一帶測量過七千五百餘人，結果我國兒童在該測驗上的成績較優於美國的標準。曹日昌於二十四年在定縣用迷津測驗測量了一百八十四人，內中男生八十五人，女生九十九人，年齡從三歲到十三歲，所得結果顯示除實在年齡三歲、四歲和五歲的三組以外，其他各組都是測驗年齡比實在年齡落後，平均測驗商數為 85.17，與上海市教育局所得結果恰相反。這兩次試用所得結果的不同，大概由於受試者不同和施行手續參差所致。照這兩次試用經驗，鮑德氏迷津測驗在未加詳細修訂以前，似不能認為適用於我國兒童。

關於迷津測驗在中國應用之結果可閱讀下列各參考書：

迷津智力測驗報告專號（大上海教育第二卷第八期）

曹日昌 試用鮑德氏迷津測驗初步報告（中國心理學報第一卷第三期）

陳選善 關於鮑德氏迷津測驗在中國之應用（測驗三卷一期）

關於迷津測驗之性質可參閱朱鎮蓀譯，迷津測驗之理論與實際（測驗二卷二期）（本文係 S. D. Porteus, Guide to Porteus Maze Test 之譯文）

## 二 團體智力測驗

六、蕭孝麟訂正古氏兒童智慧測驗 本測驗係美人古納夫（F. L. Goodenough）原編，乃一種團體測驗，適用的年齡從四歲到十歲。本測驗並無時間限制，大概年幼兒童多於五分鐘或十分鐘完畢。施行測驗時祇須分發白紙鉛筆，叫兒童自由地畫一個人，手續甚為簡便。但批分方法相當複雜，古氏定有詳細標準以盡量減少個人主觀的攙入。據古納夫本人的研究，本測驗與斯丹福比納智力量表的相關（各年齡兒童分別計算）平均為.76，測驗的信度（各年齡兒童分別計算）在.80至.90之間。古氏原著為 F. L. Goodenough, Measurement of Intelligence by Drawing, 1926.

本測驗經蕭孝麟詳細修訂。向兒童的說明，古氏原來只是叫兒童畫一個人，蕭氏卻詳細規定須畫一個穿短衣的男人，而且要畫一個整個的人，臉應當朝着前面，手應當畫出。這樣的說明，可以減少以後批分的困難。在批分標準方面，因為中西服裝之不同，故不得不有所修改。記分標準力求詳盡，以減少主觀的成分。每一項目的記分方法，經過二種分析手續而後決定。第一種手續是確定通過每一項目者的人數百分

比是否與年齡成正比，而且是否可以區別優常劣三個等級的兒童。這種手續是確定記分標準的效度。第二種手續是使每張畫由四人分別記分。倘在某一項目下各人記分之結果不同，則由各人說明其理由，以確定困難之所在，然後製造一種記分標準，並再予以嘗試，至各人記分之結果完全相符為止。這種手續乃力求記分之客觀，以增加測驗結果的信度。

本測驗記分標準分四十九個項目，每與其一項相符者得一分。茲將最先十個項目錄後：(1) 有頭，(2) 有腿，(3) 有臂，(4A) 有軀幹，(4B) 軀幹之長度較大於其闊度，(4C) 有肩，(5A) 臂與腿之依附處，(5B) 兩腿附於軀幹且兩臂依附之處必須正確，(6A) 有頸，(6B) 頸部外線可與頭部外線或軀幹外線成一連續而無方角之線。

這些項目只是記分標準的綱目，每項目記分方法，另有詳細說明。例如第一第二兩項目，有下列的說明：

有頭 頭與軀幹須有區別。僅有面部各種器官而無表示頭部的外線，則不能通過。但有表示頭部的外線而無面部各種器官，卻可通過。

有腿 正面須有二腿，而側面則可有一腿或二腿。如只有一腿而在此腿上附有二足，亦可通過。此處只須注意於腿的數目而不必注意於腿與軀幹是否接連。

此外修訂者又將記分指導中凡非文字所能詳盡描寫之各點製成圖例，以爲記分之根據。此等圖例



多半代表兩個極限之情形，一爲能通過的圖形之最低極限，一爲不能通過的圖形之最高極限。換一句話說，前者係表示一個圖形至少必須達到某種程度才能通過。後者係表示一個圖形縱然好到某種程度還是不能通過。此種圖例其有益於記分之客觀性當非淺鮮。

根據記分標準所得的分數可用下列各表解釋之：(一)分數與智齡對照表，得了智齡，可照普通公式計算智商。(二)以測驗分數爲單位之標準差值表。(三)以智齡爲單位之標準差值表。(四)以智商爲單位之標準差值表。(五)各年齡的測驗分數與其百分位置對照表。本測驗應用範圍從四歲至十一歲止。

關於本測驗修訂的經過，施行的手續，批分的標準，及分數的解釋，詳蕭孝燦訂正古氏兒童智慧測驗國立中央大學心理教育實驗專篇第四卷第一期，民國二十八年六月。

此外張革對於古氏畫人測驗曾加試用；沈有乾對於本測驗之記分方法曾有一種研究，可參閱下列各文。

張革古納夫畫人智力測驗在中國之應用（中華教育界第二十三卷第五期）

張革古納夫畫人測驗說明（浙江教育第一卷第一期）

沈有乾張革畫人測驗在中國應用之問題（中華教育界第二十三卷第九期）

沈有乾古納夫智力測驗之研究（浙江教育第二卷第二期第五期）

七黃覺民訂正幼童智力圖形測驗 本測驗係美國哥倫比亞大學師範院教授賓特納及克能海原

著，測驗材料全係圖形，做法簡單而有趣，頗合測驗幼童之用，經黃覺民加以訂正並求得我國常模。本測驗係團體測驗，應用的範圍從四足歲到十足歲。

黃氏訂正工作重要的有下列幾點：（一）增加預備測驗藉以減少正測驗中的指導語因為訂正者覺得「在中國語言未統一前，測驗既常用方言指導，那麼不論何種標準測驗，凡是口頭指導語用得愈多的，則所得結果亦愈不可靠，全國各地愈難比較。」（二）原測驗中圖形，不合於中國兒童之用者加以改動。（三）測驗訂正後先在上海試驗四次，後分託全國二十五個商務分館，請有經驗的教界人士分赴城鄉測驗，計收回測驗卷可用的有三千四百二十三卷。「吾國所編智力測驗，曾在這麼多地方測驗過而求出常模的，似乎只有本測驗。」訂正本有T分數常模及年齡常模兩種。

關於本測驗的施行手續及計分方法請閱說明書（商務出版。）

八、蕭孝嶸小學智慧測驗 本測驗係蕭孝嶸所編，於一九三九年在商務印書館出版。本測驗包括下

述四種材料：（1）矯正測驗，（2）辨同測驗，（3）仿繪測驗，（4）辨位測驗。全部材料全係圖形，不用文字。此外主試尚需要下列四種材料：（1）矯正測驗大號圖形一張，（2）辨同測驗大號圖形一張，（3）辨位測驗標準圖形五張，（4）仿繪測驗記分標準圖形一本。前三項為測驗時所須用者，後一項為記分時所須用者。本測驗之用途有二：（1）小學一年級入學考試，可以此種測驗之分數為取錄之標準，（2）本測驗之分數亦可用為一二年級分組教學之根據。

在應用本測驗時每次測驗之人數以十二人至十六人爲限。倘若環境靜寂而主試的手續已極純熟，則一次測驗二十人亦無不可，但至多不得超過此數。

根據八百五十五個一年級生的結果，測驗分數的中數爲 36.18，均數爲 37.92。

據編者言，本測驗具有下列六種優點：(1) 測驗材料最能引起並維持兒童的興趣，因此能充分測量其所具有之能力。(2) 過去環境的差異對於測驗的結果，只能發生很少的影響。(3) 測驗手續完全標準化。(4) 測驗時間極短，不致產生疲勞現象。(5) 測驗易於實施。(6) 測驗的效度頗高。根據八百五十五個小學一年級兒童的結果，測驗分數與第一學期學科成績的相關爲 .64，此爲除去年齡因素後之相關。

關於施行本測驗之手續、時限及常模，可參閱蕭孝嶸小學智慧測驗說明書。(商務。)

九、陳鶴琴圖形智力測驗 陳鶴琴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，完全以圖畫爲內容。

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，叫兒童填圖中缺少的東西；(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，每種均有三十題；(5) 較圖，這是一個同異測驗，共四十五題。

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，共十二題。(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，不歸人類，須劃去，共三十題。(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，共三十題。(4) 辨圖，叫兒童將每個方格中多餘一個圖劃去，共三十題。(5) 交替，共二百題。

十、廖世承團體智力測驗 本測驗是廖世承參照美國「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，本測驗都可適用。

量表甲包含下列五種測驗：(1) 算術理解題，共有十五題，由淺而深。(2) 填字，共有二十一題，亦由淺而深。(3) 理解的選擇，共有三十二題。(4) 同異測驗，共有五十題。(5) 形數測驗，共一百四十個圖形。

量表乙亦包含五種測驗：(1) 算術演習題，共有二十二題。(2) 常識測驗，共有三十五題。(3) 字彙測驗，共有三十五題。(4) 比喻測驗，共有三十二題。(5) 校對測驗，共有五十題。

十一、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所編造，共有二類。每類測驗有九十題，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，被試者須在此圖上劃一×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。

十二、德爾滿機械智力測驗 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。

十三、劉廷芳中學智慧測驗 本測驗有兩類，每類包含下列十種測驗：(一)見聞，(二)最好的答案，(三)字義，(四)推理的選擇，(五)算術，(六)句義，(七)類推，(八)雜句，(九)類別，(十)數系。本測驗雖已編就，似未曾發行。

關於以上各種測驗的內容可參閱廖世承陳鶴琴測驗概要第五章。(商務)

關於施行手續，時限及常模，可閱讀各測驗說明書(商務)。但這幾種測驗因編造較久，當時所求得之常模未必適用於現在。

十四、陳禮江成人用非文字團體智力測驗 本測驗係陳禮江陳友端合編，包含下列六種測驗：(一)迷津共八題，(二)異同識別共三十題，(三)劃去異類，共三十題，(四)立方數配合共十四題，(五)時間順遂共十八題，(六)幾何形分析共三十二題，全部測驗共一百三十二題。各測驗的時限均有規定，總共三十三分。

編者曾求三十五個大學生的測驗成績與學業成績的相關，得相關係數 $.784$ 。根據六十一位文盲及受初等教育者的結果，測驗成績和學業成績的相關為 $.364$ 。本測驗與廖氏團體智力測驗(甲種)的相關，根據會受初等教育及中等教育者四十六人的結果為 $.786$ 。

本測驗常模及T量表尚未求得。本測驗祇有量表甲一種，量表乙尚在編造中。

關於本測驗編造經過可閱讀陳禮江成人用非文字團體智力測驗編造經過及其初步結果教育雜誌第二十六卷第一號。

十五、蕭孝燦大學入學心理測驗 本測驗是蕭孝燦根據賽斯東一九二二年所發表之大學心理測驗改編的，曾於民國二十六年試用於國立中央大學。據試用結果本測驗之效度與信度均相當滿意。

本測驗時限為三十分鐘，但與需時數日之入學試驗卻有七的相關。

本測驗為一種自理測驗。主試的手續異常簡單，只是將測驗卷發給受試者，等他們填好了姓名以後，就說一「做」字，到了三十分鐘，便說一「停」字。這種測驗在主試方面不需要專門的訓練，就是未曾學過心理學的人也可以擔任主試工作。

十六、沈有乾智力測驗 沈有乾在浙江大學曾編智力測驗兩類，試用於投考大學教育系之學生，甲類在民國二十二年試用，乙類在二十三年試用，試用結果顯示測驗信度在.92與.97之間。以測驗分數推測入學後學業成績，較入學考試為正確。甲類包含十種材料，乙類包含九種材料。

關於本測驗內容及試用結果，可閱讀沈有乾大學入學考試中學力測驗之試用（測驗二卷二期）。

十七、郝耀東大學入學智力測驗 郝耀東曾編造抽象智力測驗一種，包含六種材料：（1）比例，（2）

立方計算，（3）類別，（4）紛亂詞句，（5）填數，（6）關係，共一百四十九題。於民國二十三年試用於安徽大學。郝氏又根據美國塞斯東編的巧智測驗（*Ingenuity Tests*）編成智力測驗一種，有甲乙兩份，各有十二

題，於民國二十四年試用於安徽大學。關於這兩種測驗的內容及試用經過，可參閱郝耀東一個大學入學智力測驗的嘗試。（測驗二卷四期）

此外張耀翔曾編製大學生智力測驗一種，為北平師範大學入學考試之用。其內容有（1）指示（2）辨圖，（3）推理，（4）類似，（5）配數，（6）常識六種材料。陸志韋亦編有大學智力測驗為燕京大學入學考試之用，其內容有指示、詞句重組、算術應用、數系、類推五種材料。（見朱君毅編教育測驗與統計四十三頁。）（商務）

以上各種測驗大都是測量所謂普通智慧的。近年來能力分析的研究，使比納式的智力測驗在理論上發生動搖。那種忽而測驗記憶，忽而測驗推理，忽而測驗語言文字，忽而測驗數目計算的辦法，確乎是很難認為合理的。譬如體格的測量，我們決不會把身長、體重、肺量、握力等等結果，雜亂混成一個莫明其妙的籠統價值。不過能力的分析是一個十分困難的問題。最近的研究結果，雖然可以說在方法方面已經略見端倪，但離着可以根據來編造測驗的階段尚遠。所以比納式的智力測驗仍有其功用的。

## 第五講 國文測驗

國文教學的目標，不外發表能力和了解能力之訓練兩端。故國文程度之考查，亦可分為兩方面，即發表能力（即說話與作文）與了解能力（耳聽與閱讀）。至於字的正當寫法（默字及書法）字義的了解，字音的讀法（識字）文字的正當用法（文法），乃了解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 閱讀測驗

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著重了。

但是朗讀仍有其地位，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。

讀法測驗可別為朗讀測驗與默讀測驗兩類。我國尚無朗讀測驗之編造，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。

一、陳鶴琴初小默讀測驗 本測驗共有兩類，每類包含四個單獨測驗。

測驗一、是測量兒童的辨字能力的。本測驗共四十行，每行五字，其中祇有一個是字，餘者皆似字而非字，叫兒童將一個是字的圈出。

測驗二、是測量兒童用字的能力的。本測驗共有三十句句，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。

測驗三、是測量兒童了解字義或句義的能力的。本測驗計有三十行，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。

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。

二、陳鶴琴小學默讀測驗 本測驗共有五類，可以替換使用。每類有十篇文章，由短而長，由淺而深。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。

三、陳鶴琴中學默讀測驗 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。

關於這幾種測驗的內容可參閱廖世承陳鶴琴測驗概要。

四、艾王小學國語默讀測驗 本測驗係艾偉王全桂合編，共有十類，分低中高三組，各組第一類於二十八年已由中華書局印行出版。低組代表二下、二上、三中組代表三下、四上、四下、高組代表五上、五下、六上、六下。每組中各類測驗難度大約相等。測驗材料係從兒童通訊、國語讀本選來改編而成。低中組測驗用字係以教育部編印頒行之小學初級分級暫用字彙為依據。選擇材料時所採取的原則如下：

（1）測驗材料包括故事、時事通訊及各種敘事的文章、詩詞等韻文，因為它另具一種特性，不便夾雜。

（2）每組各篇文字的長短，漸次增加：低組自十餘字至百字左右，中組自七十餘字至二百字左右，高組自二百餘字至三百六十餘字。

（3）每篇文字竭力避免片段的截取，而設法使成爲一篇整段的文字，有頭有尾。

（4）不適宜於小學生閱讀的材料設法避免。

測驗問題的編製採用四答選一法。至於每篇文字後面的問題數，則視文字之長短，內容之難易而定。測驗時限一律爲三十五分鐘。測驗卷的格式，低中兩組採用答案寫在測驗卷上的辦法，高組採用答案紙的辦法，以省紙張並便於保存。艾氏曾於民國二十五年作大規模之試用，求有年級常模。

關於本測驗編造經過以及施行手續，年級常模，可參閱艾偉小學教育測驗說明書（中華）艾偉王

全桂編造小學國語默讀測驗之經過（測驗二卷四期）

五艾偉中學文白理解力測驗 本測驗分量表甲乙兩類。量表甲包括白話文一篇，題曰感化，文言文一篇，顧亭林書吳潘二子事，用以測量初高中六級之理解與速率二者。量表乙包括三篇短篇白話文，美國人的母親日兵體操，和三篇短篇文言文，歸熙甫野鶴軒壁記，左宗棠上賀蔗農書，顧亭林生員論，用以測量初高中六級之閱讀能力。量表甲無規定時限，量表乙則有。甲乙兩量表之相關，其係數為 $.82$ ，比率為 $.83$ ，信度尚高。就甲乙兩量表與學校國文成績求相關，其係數平均為 $.66$ 。艾氏並求有甲乙兩量表之年級常模。

關於本測驗原文及應用結果，可閱讀艾偉中學文白測驗結果之比較研究（國立中央大學心理教育實驗專篇第二卷第一期民國二十四年六月）

朗讀測驗我國尚付缺如。茲介紹美國葛雷（W. S. Gray）朗讀測驗以示方法之一般。本測驗有十段文字，由淺入深。測驗的方法是令學生朗誦，記錄其所須之時間，同時記錄其錯誤。錯誤分為完全讀錯、部分讀錯、漏讀、代替、插入、重複六種。本測驗所測量的乃正確與速率兩方面而不及於了解。本測驗是一種個人測驗。

閱讀能力的考查又可從觀察眼動入手。從閱讀時眼動的情形，可以決定一人之閱讀能力。眼動情形可以從三方面觀察（1）眼停次數，（2）眼停時間，（3）回視次數。大概閱讀能力好的，眼停次數及回視次數少，而眼停時間短。

## 二 默字測驗和識字測驗

一個字有形音義三方面。在默讀時，須能見形而知其義，才能了解所讀之內容。朗讀時，除見形知義外，還須能見形而知其音。在聽講或談話時，須能聞音而知義，才能了解對方的講演或談話。聽講如須做筆記，則除聞音而知其義外，還須知其形。默字測驗是看兒童聽了字音，曉得了字義，能否把牠寫出來。識字測驗是看兒童看了字形，能否知其音和義。

一、陳鶴琴小學默字測驗 默字測驗可有兩種方式，即單字式和句子式。句子式是將測驗字編在句子裏面，讀給兒童聽，叫他們把整個句子寫出來。單字式是由主試者將測驗字讀出來，解釋一遍（以免同音字之誤），叫學生寫出來。這兩種方式以單字式較為簡單。本測驗是採用單字式的，共有兩類，每類有五十個字。這一百個字是根據陳鶴琴語體文應用字彙中最常用的（即次數最多的）二千個字，用隨機選擇的方法得來的。方法是每隔四十個字取兩個字，第一個字編入第一類測驗，第二個字編入第二類測驗。所以這一百個字是可以代表二千個常用字的。

二、艾偉識字測驗 張耀翔曾編有識字測驗一種，共分十組，每組一百字。本測驗之功用，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，選擇而來。

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先將所用之某組測驗材料一百字，依次寫在一張紙上，每字大小約半寸見方。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釋出來。注音符反切，同音之字，注音符號或英文拼字均可。釋義用訓詁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。本測驗的信度甚高。艾偉根據張氏所得結果，求出本測驗各組相互的等級相關，大多數在 .90 以上。

艾偉就張氏所選十個字組中之第四第九兩組，採用團體測驗方式，改編爲識字測驗量表甲第一第二兩類。第一類九十五字，第二類九十八字。測驗方式採用四答選一法，對於字音字義，各擬四種答案，注音符兼用假借字、反切、注音符母三種方法。本測驗適用的年級自小學五年級上學期起，至高三下學期止。時間爲三十五分鐘。

測驗成績分音義俱知，知音不知義，知義不知音，音總及義總五項，其中以音義俱知一項爲最準確，最能分別各組的程度。據試用結果，第二類比第一類難。但兩類之相關甚高，就音義俱知一項成績而言，其相關係數爲 .917。

艾氏於民國二十一年用本測驗在南京杭州兩處測驗三千五百餘人，求有各級常模。

本測驗選字所根據之實用學生字典共有一萬三千四百六十九字或約爲一萬三千五百字。第一類測驗有九十五字，每一字可代表一百四十二字；第二類測驗有九十八字，每一字可代表一百三十八字。根據此種計算，以定各級之識字數量，若以音義俱知成績爲進，則有如下表所列：

表二 根據艾偉識字測驗推算之各級識字數量（以音義俱知字數爲準）

年 級	第一類	第二類
小 五 上	2245	1975
小 五 下	2684	2298
小 六 上	3237	2739
小 六 下	3640	3151
初中一下	4509	4014
初中二下	5161	4612
初中三下	5424	4803
高中一下	5801	5145
高中二下	5970	5293
高中三下	6187	5546

若以知音不知義或知義不知音之字爲半識字，而以兩個半識字等於一個全識字，以此字數加在音義俱知之字數上，則各級之識字數量，有如下表所列：

表四 根據艾偉識字測驗推算之各級識字數量（以音義俱知知音不知義或知義不知音字數之半爲準）

年 級	第一類	第二類
小 五 上	4235	3905
小 五 下	4780	4362
小 六 上	5322	4903
小 六 下	5832	5382
初中一下	6671	6325
初中二下	7258	6788
初中三下	7631	7120
高中一下	7878	7353
高中二下	7930	7407
高中三下	8096	7533

關於本測驗材料之來源及試用經過，可閱讀下列兩種參考材料：

張耀翔識字測驗（心理雜誌選存中華）

艾偉漢字測量（國立中央大學心理教育實驗專篇第一卷第二期）

三、楊黃李合編識字測驗 楊敏祺黃裳李智合編之識字測驗，材料從李智之民族中心小學常用字彙隨機選擇而來，共有兩份，各七十字。每字附有三個音，四個意義，由被試者選擇其一。凡一字音義俱答對者得一分。

編者用本測驗在廣州測驗過二千七百餘人，求得各年級常模如下：

表五 楊黃李合編識字測驗  
常模之年級之驗

年級	測驗甲	測驗乙	平均
六	54.0	52.0	53.0
五	47.4	47.1	48.5
四	36.4	39.2	37.6
三	29.4	29.0	29.2
二	18.0	16.7	17.4

按李智之常用字彙共有五五五二字，是測驗甲或乙中每字約代表李氏字彙中七九字，如是則各年級實在識字總數約如下表：

表六 根據楊黃李合編識字測驗推算之各級識字數量

年級	測驗甲	測驗乙	平均
六	4280	4130	4190
五	3760	3900	3850
四	2900	3110	2980
三	2330	2300	2320
二	1430	1320	1380

本測驗的信度，以各年齡分別計算，平均在·九〇以上。

關於本測驗編造經過可閱楊敏祺黃裳李智識字測驗經過的報告（教育雜誌第二十六卷第十一號）。

### 三 文法測驗

我國國文教學向來不講求文法，文法上之錯誤，概以「不通」二字概括之，僅指出其有錯誤，而不說明錯誤之性質，所以學者往往對於同樣的錯誤重複多次，而仍不知改正。文法測驗之功用，即在發現學生在文字上所犯之錯誤，以為教學之根據。

一、陳鶴琴小學文法測驗 本測驗適用於後期小學。內容有五十旬句子，每旬裏面有一錯字，這個錯



字是屬於文法方面的，叫被試者改正。

二、廖世承中學文法測驗 本測驗有兩類。每類有三十個句子，每句裏面有二三段虛線，每段虛線上應填寫一個適當的文法字。

關於以上兩種測驗的內容，可參廖世承陳鶴琴測驗概要第九章。

#### 四 作文測驗

測量作文能力的工具是評判量表或稱品質量表。在評判量表中，載有許多篇優劣不一的作文樣本，各篇樣本的品質，由編造者依多數人的判斷，用統計的方法評定，定有一種分數。評判時可將一篇作文與量表上樣本比較，然後給分。如此評判者有標準可循，不致全憑主觀了。但評定一篇作文的優劣是相當困難的事。文字有許多體裁，如議論文、敘事文、描寫文。文字體裁不同，評定的標準亦就不同。一篇作文的品質係多數因素綜合的結果，至少可以分為內容與形式兩方面。內容又可以分析為思想、組織、用字等因素；形式又可分析為造句、文法、錯字、標點等因素。因素既多而各因素在總成績中應佔之分量，又無明確的規定，則評判者雖有標準樣本可資比較，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。如何減少評判者主觀的成分乃作文能力測量中重要的問題。

我國作文量表計有兩種：

一、周學章作文量表 本量表載有作文樣本十篇，代表十個等級。這十篇作文的優劣，係根據二百七十四人的評判評定，最劣的一篇分數爲0，最優的一篇分數爲9。茲抄錄三篇樣本於下，以示一般。

雨具 價值1.1

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

說褒狀 價值5.0

余校每月必發一次褒狀，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，上有褒狀二字，印以鈐記，寫明某種爲最優，並寫校長姓名，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。

記女子復仇事 價值9.0

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，過目了了，尼鍾愛之，極有顧復，一如生母焉，菴依山臨水，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，官怒置之獄，匝月庾死，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，是即清水菴證果也，蓋魏姬嘗遊是菴，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。

本量表係應用精密的統計方法始克編成。惟有一缺點就是各篇樣本的題目不一致，應用時不免有困難。

二、俞子夷小學綴法量表 本量表可以從小學二年級用起，到初中三年級爲止。量表上載有樣本十八篇，量表上所定各篇分數就是「分數」。除最初三篇外，其餘各篇題目一律，用的題目是「我的好朋友」。茲錄樣本數篇於後：

我的好朋友 價值 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

我的好朋友 價值 50

我的好朋友是曹兼霖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

我的好朋友 價值 90

我的朋友很多，但是好的卻很少，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。——我在課餘的當兒，常到森林下，聽牠幽婉的歌聲，夜深的時候，萬籟俱寂，她能把我一個明珠來戲引我，這好像牠是不能不給我的，而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）

根據作文量表評定作文的優劣，雖然可以稍稍減少主觀的成分，但各人對於同一篇作文，所評定的

分數仍不能一致，有時相殊甚遠。周學章對於此問題有詳細的研究。據其研究，訓練可以增加評判的準確度。訓練的方法，可將價值已知的作文多篇，令評判員依作文量表評定給分，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，評判的準確度不難增加。周氏集有作文二百六十篇，各文之價值均已評定，可用作練習之材料。要求結果之準確，一篇作文至少要由三位評判員評定，然後求其平均分數。周氏的研究又顯示一次的作文成績，不能代表一人真正的程度。據周氏估計，五篇作文的平均分數可認為相當正確。

周氏的研究見周學章測驗作文能力應知之事項（教育研究第四十四期，民國二十二年五月）。

俞氏周氏的作文量表都是測量作文的普通品質的。另外有一種量表具有診斷的功用，比較更有價值。我們可以舉美國范魏格寧（M. J. Yan Wagener）作文量表為例。本量表分三種，一種是評量議論文用的，一種是評量敘事文用的，一種是評量描寫文用的。每篇作文的品質從三方面來評定，所以有三種分數。第一是思想內容，第二是結構，第三是形式。這三種分數又可依編者所定之公式合為總分數。編者又將評定各種分數時所應顧及之點詳為釐定，使評判更有依據。

照編者規定，評判議論文之思想內容應顧及：

1. 文不離題，
2. 富有興趣，

3. 思想連貫，
4. 見解清晰，
5. 用字適當。

評判敘事文之思想內容應顧及：

1. 情境之充分的說明，
2. 會話自然而適當，
3. 敘事步驟分明，
4. 宕筆之應用，
5. 善於形容，
6. 措辭適當而多變化。

評判描寫文之思想內容應顧及：

1. 觀察點之先後一致，
2. 描寫生動，
3. 感情的反應，
4. 措辭有力而新穎。

評判作文之結構應顧及

1. 和諧，
  2. 聯絡，
  3. 文勢或着重點，
  4. 文句之變化 and 複雜。
- 評判作文之形式應顧及：

1. 拼法，
2. 標點，
3. 大寫字，
4. 文法，
5. 段落。

這一類含有診斷作用的作文量表頗值得吾人之借鏡。

最後我們應該認識作文量表完全是評分的工具。假使要比較兩個或兩班兒童的作文成績，則測驗的情境如題目、時限、說明，必須劃一，比較才能公允。

## 五 書法測驗

書法是一種工具的學科，而同時又爲一種藝術。書法教學對於藝術的價值，固然不應完全抹殺，而工具的或實用的價值則尤爲重要。書法教學的目標應以實用爲主，美觀爲副。所以小學課程標準規定寫字的目標是「養成正確敏捷的書寫能力」，不再提美觀問題了。小學寫字作業，照課程標準的規定，有中小字正書行書的練習，至於草書祇要認識就夠了。

書法的測量有品質和速率兩方面。速率的測量較爲簡便，可於規定時間內，令被試者寫一定的材料，視其所寫字數，以評定書寫速率。品質的測量，則有賴於書法量表，以爲評判之依據。

俞子夷早於民國七年編成四種書法量表，一種正書中字，一種正書小字，一種行書中字，一種行書小字。這幾種量表是做照桑戴克的書法量表編造的。

後來俞氏又編了一種正書小字量表，一種行書小字量表，速率與品質都可測量。計算速率祇需數被試者在規定時間內（正書限四分鐘，行書限二分半鐘）寫的字數，然後從對照表查出T分數。計算品質須將學生作品與量表上樣本比較，找出品質最相近的樣本，該樣本上的數目，就是品質的T分數。

正書小字量表載有十六種樣本，行書小字量表有十三種樣本。樣本所用的字爲「四隻小鳥他們在園中飛好像一個人字」十六字。這十六個字大概把中國字各種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。本量表小學二年級至初中二年級均可適用，平常從小學三年級到初中一最適宜。行書教得較遲，可從小學四五年級用起。

含有診斷性的書法量表我國尙無人編造。我們不妨舉美國費里門所編的書法分析量表爲例。這量表實際上包含五個量表，測量（一）斜度的一致，（二）行列的整齊，（三）筆劃的品質，（四）字母的形式，（五）間隔的適當五方面。

應用書法量表評定書法的優劣，多少含有主觀的成分。因此幾位評判員，對於同一作品，所評定的分數，往往大有參差。要減少主觀的成分，評判員的訓練似爲最有效的辦法。

## 六 國文常識測驗

廖世承編有國文常識測驗一種，適用於中學。內容包括應用文、聲韻、作品、作家、經史、詩詞、新文學等材料。



## 第六講 數學測驗

數學測驗依內容可分為四則測驗、應用題測驗、代數測驗、幾何測驗等；依功用可分為調查測驗、診斷測驗及練習測驗三類；依對象又可分為正確、速率及難度三方面。

### 一 四則測驗

一、德爾滿算術四則測驗 本測驗是麥柯爾在中國時所編，由德爾滿求得常模。本測驗包含速率、正確、診斷三種性質，前後期小學均可適用。

本測驗共有兩類，難易相仿，可以替換應用。每類測驗有八十題，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難度。茲將本測驗前面一部分列下：

(4)	(3)	(2)	(1)	(5)	(9)
7	7	6	3	6	3
+9	+5	+2	+4	-3	1
-----	-----	-----	-----	-----	-----
(8)	(7)	(6)	(5)	(8)	(10)
8	9	8	6	8	8
-0	-5	-4	-3	-4	0
-----	-----	-----	-----	-----	-----
(12)	(11)	(10)	(9)	(8)	(10)
50	24	8	3	8	8
+6	+4	+5	+7	+5	+5
-----	-----	-----	-----	-----	-----

上面(1)(2)兩題是單位數相加不須進十；(3)(4)兩題是單位數相加而須進十；(9)(10)兩題是三數相加；(11)(12)兩題是兩位數加單位數，以下加法題目一步難一步。減法乘法除法的題目亦是照此原則而編排的。

二、俞氏小學算術混合四則測驗 本測驗原爲吳特(C. Woody)麥柯爾合編，由俞子夷修改試用，並求得中國學校常模。測驗有兩類，每類三十四題，內容包括整數四則、小數四則、分數四則及諸等，從小學二年級起至初中三年級均可應用，平常從小學三年級至初中一最爲適用。

本測驗的編製完全以統計上的難易爲標準，對於算術各難易階級則未曾顧到，所以不能作爲診斷之用。

三、崑山算術四則測驗 本測驗亦俞子夷所編，因常模係根據崑山各校成績，故以崑山名。本測驗適用於初級小學。

四、艾郭南氏小學算術測驗 本測驗係艾偉郭祖超於民國二十五年合編，是一種診斷測驗，包括整數四則的基本結合和算術的各種演算能力。全部測驗計有基本結合60題，整數的高級演算68題，小數15題，分數17題，共160題。題目的排列根據二十一年十月部頒課程標準的條文，每一條文下的題目，視該條文所含因子的多寡而定，至少有一題，最多有十七題。測驗時間一二年級三十五分鐘，三年級至六年級五十分鐘。本測驗共有四頁，取活頁制。第一頁測驗一上至二上應養成之能力，第二頁測驗二下至三上應養成

之能力，第三頁測驗三下至四下應養成之能力，第四頁測驗五上至六下應養成之能力。應用時視被試者所在年級而定發給測驗紙的頁數。

關於本測驗編造經過及年級常模，可閱艾偉小學教育測驗說明書（中華）及艾偉郭祖超編造小學算術測驗之經過（測驗二卷四期）。

## 二 應用題測驗

應用題測驗之編造有兩點應該注意。第一、測驗題目須與實際生活中所遇到的問題相近。如龜兔同籠一類荒誕不經的題目應完全避免。第二、文字宜簡易通俗，以免被試者因文字上的困難而做錯或做不出。

我國算術應用題測驗有下列數種：

一、俞氏小學算術應用題測驗 本測驗內容包括整數四則、小數四則、分數、百分、折扣、面積、體積、容量等，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，難易相仿，可以替換應用。從小學二年級至初中三年級本測驗均適用，從小三至初中一最相宜。

本測驗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，似乎還應顧到算法原則合與不合。美國門羅的算術理解題測驗就分

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，因測驗常模根據崑山學校成績，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口述題目，以免去文字上的困難。

### 三 練習測驗

測驗最大的功用，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，才能取切實的效果。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。練習測驗的功效約有下列數端：

(一) 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，可以避免重複的與未成熟的教學。

(二) 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。

(三)使學生有充分的練習。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。

(四)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。

一、愈子夷麥柯爾算術練習測驗。美國科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，由愈子夷與麥柯爾彙編而成，由商務印書館出版。

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，最後四課係學習用的，又有四課（第13, 30, 41, 54課）係測驗用的，其餘五十課為練習的材料。每課中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本測驗說明書，不多述了。

二、馬靜軒兒童算術練習測驗片 馬氏算術練習測驗片，由兒童書局出版專供小學初年級（一二

三年級）練習基本四則用的。用完這套片子，可以接用俞氏算術練習測驗。

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，其餘七十六課為練習用的。這七十六課練習可分為十四組：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。

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。（兒童書局出版。）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。練習的時間均有規定。

三、雷震清算術四則基本練習片 本練習材料適用於小學高年級。本材料以高小全部算術為基本，提出須熟練的基本方法，分析步驟，按部就班，循序漸進，正確和速度，兩方兼顧。本練習材料包括（一）整數四則，包含加減乘除四種的基本結合，進位、退位、多項式、多位式，以及乘法中常用的加法，除法中常用的減法等；（二）分數四則，包含分數化法，分數加減乘除等各種計算方法；（三）小數，包含小數加減乘除；（四）百

分法，包含百分數的寫法和小數分數的關係。

全部材料共九十課，分三部份：第一部分為整數四則，計加法十一課，減法七課，乘法九課，除法十課，共三十七課；第二部分為分數四則，計分數化法六課，加法七課，減法乘法除法各五課，共二十八課；第三部分為小數和百分法，計小數加減七課，小數乘除五課，百分法十三課，共二十五課。

本材料每種包含測驗及練習二項：第一課為測驗（百分法第一二課均為測驗），包含一種中各項進度及計算內容；第二課以後為分步練習。如整數四則中的加法，第一課為各種形式的加法，可以測驗學生看是否都會做；第二課至第十一課為分步練習，每課代表一種進度。學生做第一課時，如有錯誤，先查錯誤的所在，如錯誤在多項式的加法，則練習多項式的加法一課，錯在橫式，則練習橫式一課。總之，發現什麼錯誤，便練習什麼，不必一一去做，以省時間而增興趣。每課練習時間一律為五分鐘。

#### 四 中學數學測驗

一、廖世承混合數學測驗 本測驗共兩類，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。

二、艾偉修訂初等代數診斷測驗 本測驗所用材料係美入道格拉斯 (H. R. Douglas) 所編初等代數標準診斷測驗甲組第一種與乙組第一種。甲組測驗測量初等代數基本運算能力，包括（一）加法

及減法，(二)乘法，(三)除法，(四)簡單方程式四個測驗。乙組測驗測量甲組所測範圍以外之能力，包含(一)分數，(二)劈生，(三)公式及分數方程式，(四)聯立方程式，(五)圖，(六)開平方、指數及根號，(七)二次方程式七個測驗。修訂時對於甲組測驗未有變動，乙組測驗則刪去最後三個測驗。甲組測驗內四個測驗各有十題，乙組測驗內四個測驗各有五題。測驗時限甲組三十四分鐘乙組五十七分鐘。

艾氏會將甲乙兩組測驗京滬杭一帶中學校初中二三年級學生六百四十三人，內中男生三二七人，女生三一六人。被試者以曾習代數一年者為限所得均數如下：

	測驗一	測驗二	測驗三	測驗四	總計
甲組測驗	6.34	5.59	5.85	6.30	24.56
乙組測驗	1.20	2.56	2.18	1.85	8.32

附註：上列常模係五月底之成績。如測驗時期不在五月，則成績當有差別。

按學生個別成績所求甲乙兩種測驗之相關，最高係數為·八四，餘則多在·六五以上。若依各校均數以求相關，則係數為·九四，校正後得·九九。關於本測驗之信度似可滿意。甲乙兩種各測驗間之相互關係，大多在·八〇以上。甲種各測驗間之相關，大體較乙種各測驗間之相關為高。又測驗成績與學校算學成績之相關最高達·七〇。

本測驗之優點約有下列數項：(一)測驗問題均經審慎選擇，使包含各種形式及各種難度。且有充分的測驗時間，使學生得有嘗試全部問題之機會，故學生所具代數能力，得以充分表現。(二)測驗手續簡單，時間經濟，計算分數，完全客觀。(三)本測驗具有診斷之價值。根據本測



驗所得結果，以診斷班級弱點及施行補救教學可有相當信任。至個人診斷則僅供建議及指示之用，蓋測驗中每一過程之下每種形式僅有一二問題者，欲以之診斷個人，恐未必可靠。

關於本測驗修訂經過題目內容，施行手續，及詳細常模可閱艾偉郭祖超初中代數測量（測驗三卷一期）。

## 第七講 其他各科測驗

### 一 常識科測驗

一、陳鶴琴小學常識測驗 我國早期編成的常識測驗有陳鶴琴小學常識測驗及俞子夷小學社會自然測驗兩種。陳編小學常識測驗共有兩類，每類一百題，採用四答選一式，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。

二、俞子夷小學社會自然測驗 本測驗包括公民、歷史、地理、自然、衛生各科材料。方式同陳編小學常識測驗。

關於以上兩類測驗的內容及常模可參閱廖陳合編測驗概要第十二章及各測驗說明書。

三、艾偉朱亞男合編初級小學常識測驗 本測驗係艾偉朱亞男於民國二十五年合編，由中華書局出版，共有三類，一二兩類大部係選取低年級的共同教材，一小部分是中年級教材，程度較淺；第三類純取材於中年級的共同教材，程度較高。本測驗的範圍包括社會、自然、衛生三科。在每類測驗裏，此三科題數的分配，是以教育部規定的教學時數為主要的根據。照部定教學時間，低年級社會自然衛生教材的比率為3:3:2，中年級為2:2:1。每類測驗共八十題，在第一第二兩類測驗裏社會、自然各三十題，衛生二十題；第

三類社會自然各三十二題，衛生十六題。測驗材料的選擇是以下列幾個原則為依據的：（一）分析各種常識教科書的共同教材；（二）採用各校常識科的共同試題；（三）顧到部頒課程標準；（四）搜集與兒童實際生活有關的材料；（五）參考已有的小學常識測驗。測驗年級常模已求得見艾偉小學教育測驗說明書（中華）。關於本測驗編造經過，參閱艾偉朱幼顏編造初級小學常識測驗之經過（測驗二卷四期）。

## 二 社會科測驗

一 徐氏中學本國史測驗 本測驗為徐則陵所編，測驗中題目係根據下列六條原則而擬定的：

- （一）史事與其發生時期之關係。
- （二）辨別史事發生之先後。
- （三）史事與其發生地點之關係。
- （四）史事與其人物之關係。
- （五）史事因果之關係。
- （六）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。題目採用五答選一式。

二 廖世承蘇毓棻中學混合歷史測驗 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。題目採四答選一式。初高中均適用。

三、楊國荃本國地理測驗 本測驗共有兩類，每類有五十二題，四答選一式。自小學三年級至初中二年級均適用。測驗題目是根據中小學普通地理課本編的，大概包含下列幾方面：

- (一) 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。
- (二) 關於推理的：這類問題須用思想方能解答，不是專靠記憶的。
- (三) 關於交通的。
- (四) 關於了解地理上名詞的。
- (五) 關於各地物產的。

四、廖世承章潤珊中學混合地理測驗 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，採用四答選一式。初高中均適用。

以上四種測驗都是早期編就的。測驗內容和常模與現行課程標準及學生程度未必適合。讀者可參閱廖陳合編測驗概要第十二章及各測驗說明書。

五、艾偉孫邦正小學歷史測驗、地理測驗 本測驗於民國二十五年編就，由中華書局出版。歷史測驗和地理測驗各有兩份，一份是五年級用的，另一份是六年級用的。測驗材料係從通用史地教科書（計歷史教科書五部，地理教科書三部）和各校自編社會測驗及教師的試題選擇出來的。選擇材料的原則有下列各項：

（一）在實際生活上的重要性。

（二）材料的普遍性。

（三）適合學生的程度。

（四）注重本國部份（本國部份材料約佔四分之三，外國部份材料約佔四分之一）。

（五）相當的顧到小學社會科課程標準。

（六）關於歷史方面，特別注重近代史料。

（七）關於地理方面，特別注重長江流域一帶事蹟。

本測驗每份有一百個題目，採四答選一式。測驗時限為三十分鐘。年級常模見艾偉小學教育測驗說明書。關於編造經過可參閱艾偉孫邦正編造小學歷史測驗和地理測驗之經過（測驗二卷四期）。

六、艾偉孫邦正小學社會測驗 本測驗有兩類，每類有甲乙兩份，第一類適用於五年級，第二類適用於六年級。試題即採自艾孫合編之歷史測驗和地理測驗。本測驗年級常模見艾偉小學教育測驗說明書。

### 三 自然科測驗

一、廖世承等中等混合理科測驗 本測驗內容包括物理、化學、生物、生理、衛生等材料。測驗分兩類，每類有一百題，可以替換應用。題目採用四答選一式。初高中均適用。

二、艾偉秦湘蓀小學高年級自然科測驗 本測驗共有三類，第一類適用於五年級，第二類適用於五下六上，第三類適用於六下，每類一百題。測驗時限為三十分鐘。本測驗的着重點不在技能而在態度、理想及解決問題的能力等方面。本測驗的內容係根據課程標準，各種教科書，及已有之測驗題而決定的。高級自然教材大約可分為動物、植物、礦物、天文、氣象及應用科學五部份。前三部份五年級較為注意，六年級則後二部份題數較多。選材的標準有下列數端：

- (一) 注重兒童實際生活中所能經驗及生活中重要部份的材料。
- (二) 注重兒童在環境中所能接觸和所能想像的自然現象材料。
- (三) 採取各種自然教本中共同認為必要的材料。
- (四) 參考已有之小學自然測驗，注意許多共同認為主要的材料。
- (五) 相當的顧到自然科課程標準。
- (六) 注重普遍的問題，不取局部的。

(七)所取材料都須合乎小學五六年級學生程度的。

本測驗常模見艾偉小學教育測驗說明書。關於編造經過，可參閱秦滄蓀編造小學高年級自然科測驗之經過（測驗二卷四期）。

三、包靜元小學高級衛生測驗 本測驗有兩類，各一百題，採四答選一的彙選式。本測驗材料經過審慎的選擇，其依據有下列三項：（一）四部比較通行的高級小學衛生教科書；（二）各校衛生測驗卷；（三）部頒小學衛生科課程標準。選材的原則有下列數項：（一）有普遍性的；（二）專家所公認的；（三）合於學生程度的；（四）日常生活中所常用的；（五）環境中所急需的。根據衛生科教科書的分析，衛生教材可分為營養、生理衛生、疾病醫藥、急救、保健、公共衛生六大類，然後統計各類材料所佔分量，作為製題的標準。本測驗常模似未經發表。關於本測驗編造經過，可參閱包靜元編造小學高級衛生測驗的經過（測驗二卷四期）。

#### 四 英文測驗

一、祁司英文測驗 (Keys English Mastery Test) 本測驗由上海伊文思圖書公司出版，共有四種：

(一) 文法與語法測驗 本測驗共有七類，各類難度不盡相等。每類分兩部份，一部分採填字式，一部分採改正錯誤式，測驗中題目係根據編者對於中國學生在英文文法及語法上所犯錯誤之分析而

編製。

(二) 字彙測驗 本測驗有 A B C 三類，A B 兩類難度相等，C 類較易。測驗的方式採「求對」式，被試者須將與測驗字意義相反之字寫出。

(三) 聽意測驗 這是測驗了解英語的能力的。測驗的方式，由主試者將規定的問句讀給被試者聽，每句問句有四個答案（答案是印在測驗紙上的），被試者就四個答案中選擇一個。

(四) 默寫測驗 由被試者將主試者讀出的語句寫出來。

二、安特孫混合英文測驗 (Anderson Comprehensive English Test) 本測驗是測量閱讀、聽話、和應用英文的能力的，包括填字、字彙（翻譯）字彙（歸類）文法、看意、聽意綴法七個測驗。由商務印書館出版。

綴法測驗計分的方法頗為新奇。校閱試卷時，須將錯誤記下，錯誤分拼字、文法、大寫三種。若一字犯了二種或二種以上的錯誤，只算一個錯。假使一句句子無意義時，校閱者將應刪去的字劃去，然後加入最少限度的字，使這句句子合於文法而有意義。每加入一字，算一個錯。編者規定計分的公式如下：

$$\text{分數} = 2 \left( \frac{N^2}{100(M+1)} \right)$$

N = 全篇字數

M = 錯誤總數



本公式在統計學上的根據未蒙編者說明頗屬費解

三、白克萊默讀測驗 (Brinkley Silent Reading Test) 本測驗方式與陳鶴琴默讀測驗相同。由伊文思圖書公司出版。

關於以上三種測驗的內容可參閱王書林心理與教育測量第十六章及各測驗說明書。

四、艾氏中學英文測驗 本測驗係艾偉所編，計量表甲兩類為高中用者，量表乙兩類為初中用者。每類分字彙（五十個字）改字（二十句），填字（二十句），消字（十句）四種，後三種可稱為文法測驗。茲舉例如下：

量表甲

1. 字彙

	1	2	3	4
(A)	(3) tiger	afraid	large	animal
(B)	( ) b by	youth	girl	little child
		cradle		

2. 改字

(A) (boys) There are two boy in the classroom.

(B) ( ) He write a letter to his father.

3. 填字

(A) (3) The students paid much attention \_\_\_\_\_ his lecture.

1 2 3 4  
(on, for, to, of)

(B) ( ) This boy depends too much \_\_\_\_\_ the notes he has taken.

1 2 3 4  
(with, of, in, on)

4. 消字

(A) (was) His father was died last night.

(B) ( ) There were many people went to Shanghai last week.

艾氏並求得各年級常模可資比較(見表八)是表甲乙中第一類與第二類之相關為.98, 信度可謂甚高。就測驗內容而分為字彙文法兩種, 由此而求相關其係數均在 . 九〇以上。

表八 艾氏中學英語測驗表甲乙之常模

	初中一	初中二	初中三	高中一	高中二	高中三
量表乙	第一類	34.41	50.15	61.37	72.63	
	第二類	25.09	42.03	53.49	64.80	
量表甲	第一類			36.15	58.05	69.38
	第二類			41.67	49.66	60.45

五、艾偉高中英語默讀測驗 本測驗共有三類，性質相同，計每類有英語文章八篇，每篇之下有問題三四個或四五個。讀完一篇文章，即可解答其下之問題。如問題之中有不懂者，可將文章重讀一遍，再行解答。八篇文章之中，前五篇非常之淺，後三篇較為艱深。問題總數為五十。閱讀與解答問題時限為四十分鐘。茲將艾氏所求得之常模錄於後：

表九 艾偉高中英語默讀測驗常模

	第一類	第二類	第三類
高 中 一	67.13	52.19	67.64
高 中 二	72.49	59.58	74.02
高 中 三	73.33	63.75	78.70

附註：上表所列常模為數為用二乘過者

## 五 各科混合測驗

一、查良釗學校調查用教育測驗 本測驗專為小學調查之用，共有兩類，每類有一百四十個題目。一百個題目採用四答選一式，包括歷史、地理、自然、默讀、應用算術五種科目。各科目二十題。題目的排列以難易為標準，各科目題目相混，並不分列。末了四十個題目係算術四則題，加減乘除各十題。從小學三年級至初中二年級均適用。

二、德爾滿學校自用普通教育測驗。本測驗共有三類，第一類爲初級中學之用，第二類爲高級中學之用，第三類爲大學文科之用。據編者說，本測驗有三種特點：（一）內容實用而合時，（二）編製係科學的劃一的，（三）施行簡單而經濟。本測驗的功用，約有五端：（一）可作爲入學考試之用，（二）可爲全校重新分級之用，（三）可爲建立全國各種學校各年級的標準之用，（四）可使各校各班與全國的或區域的常模比較，（五）可作爲給予轉學生學分之根據。

各測驗包括國文、英文、自然科學、社會科學、混合數學五科目。

## 六 其他各科測驗

以上就我國已有各種教育測驗作簡單之介紹，此外中小學科目中如公民、圖畫、音樂、體育、手工各科我國尚無測驗可用。王書林在心理與教育測量一書中曾介紹美國在這幾方面的標準測驗，讀者可以參

## 七 教育測驗評價

中小學課程內各種學科大致可別爲工具學科或技能學科，如國文、算學、外國語；內容學科或知識學科，如常識、社會、自然、欣賞學科，如圖畫、音樂；及實踐學科如公民、衛生四大類。這種分類當然不是絕對的。如

國文教學一部分含有欣賞的目標；圖畫音樂的教學，亦應有技能的訓練；行爲的實踐當然要根據正確的智識。不過這種分類方法可以大體上說明各科教學側重點之所在，不無相當價值。大致紙筆一類的測驗對於工具學科比較最爲適宜，內容學科次之，對於欣賞學科和實踐學科，就不甚相宜了。

一種教育測驗的評價，最重要的標準，就是測驗內容和教學目標是否相符。換言之，測驗的應用是否能促進教學目標的實現。這一點我們在第三講已經提到。我們翻閱小學課程標準就可以明瞭常識社會自然各科的教學，不應注重零碎事實的記憶，而應着重於關係和原理的了解，態度理想志趣的培養，以及探求知識和解決問題方法的訓練。美術音樂各科的教學，不僅要相當顧到技能的訓練，而應着重於欣賞能力愛美興趣和創作能力的陶冶。就我國已編就的各科測驗而論，對於各科教學目標是否能充分符適，是一個值得考慮的問題。

## 第八講 性格測驗

性格測驗比智力測驗教育測驗發展較遲，而成就亦較少。大致說來，性格測驗現在尚在試探時期。還講不到大規模實際的應用。許多心理學者正以極大的努力，致力於測量方法的創造和測量工具的編製。性格測量的基本方法，與智力測量學業測量初無二致。一人的性格必賴行為而表現，故品格的測量必須從觀察行為入手。不過性格測量較諸智力及學業測量有幾點特殊的困難。第一、情境對於性格的表現較諸情境對於能力（智力及學力）的表現，影響尤大。這一點對於測驗材料取樣的問題增加了不少困難。第二、性格測量的進行不可使被試者覺察，不然就不易觀察到被試者的真正性格。這一點當然增加了性格測量的困難，但心理學者已經有許多方法可以克服之。

性格測量有三種方法，就是作業測驗，心理問卷，和性格評判。茲分別說明如下。

### 一 作業測驗

所謂作業測驗就是由心理學者安排一種情境，觀察被試者對於此情境的反應，以度量其性格。這類測驗在我國尚無人編造。美國已有許多嘗試，其中規模最大的要推哈之兄和梅（H. Hartshorne and M. A. May）的研究。他們受「社會宗教研究社」（Institute of Social and Religious Research）

的委託，費了五年的工夫，搜集和創造了許多關於誠實(Honesty)服務(Service)自己(Self-control)的測驗。茲略述其方法於後。

「誠實的測驗方法，是給受試者以欺詐的機會，察其反應。用普通智慧測驗的形式，材料分兩種，微有不同處，而受試者不知，兩種材料間位分發，抄襲便可發現。用普通智慧測驗形式，收卷後將答案暗中抄出，幾天之後，將卷子發還受試者，令各自校對批分，若私自將原來錯誤改正，便可查出。用極難的題目，不違法不易答對，例如閉目以鉛筆描線於某種圖畫中，若描得太正確，便有偷看的嫌疑。用兩類相等的測驗，一次監視極嚴，一次不加監視，若兩次成績相差太遠，便有欺詐的嫌疑。用體力測驗，例如握力，先由主試者指導試若干次，暗記其分數，再令自試若干次，報告最高分數，若所報告分數超過指導試行時太多，就是欺詐的表現。用錢幣作遊戲或計算，給受試者偷竊的機會，察其是否誠實。用兩次測驗，初次只問某事知否或能否，第二次則須詳細回答，兩次相比，便可見初次是否虛報。用一張書籍名單，其中一部分書名是假的，令受試者指出讀過的書籍，若所報有虛妄的書籍，又是不誠實的表现。

「所謂服務是指爲他人犧牲自己的利益。測驗的方法，包括以下各種：受試者受某種測驗可以參加個人競賽或團體競賽，各有獎品，受試者若肯放棄個人競賽的權利而參加團體競賽，便表現服務精神；一班受試者有得某種獎金的希望，主試者提出獎金用途，如各人均分，給最出力的人，購紀念品留於班上或校中，捐作慈善事業，令受試者投票表決，可見其服務精神；受試者每人得贈品文具一盒，同時主試者述鄉

村學童窮苦情形，每人發一封袋，婉勸——但不命令——其將各種用品捐贈若干，捐贈物品的質量便是服務精神的表示。」

「自治測驗分兩部：一部測驗受試者的堅持 (Persistence) 精神，給受試者一種極難讀的故事（如一字和一字間不加空白，一字的字母間多加空白，字母中不規則地間着大寫小寫，）觀其繼續努力到何處；給受試者一種不容易解答的謎題（如九連環，）觀其嘗試若干時間；此外如令受試者一脚獨立，或尋找失物，也以時間長短作堅持精神的準則。另一部測驗受試者的抑制能力，觀察受試者能否專心某種乏味工作（如作加減法，數書上每行有若干字母，或受別種長時間的測驗，）不受引誘（如讀有趣的故事，嘗試新奇謎題，看圖書，吃糖果。）」（註一）

哈之兄和梅的研究報告共有三冊書名如下：

*Studies in Deceit.*

*Studies in Service and Self-Control.*

*Studies in the Organization of Character.*

## 二 心理問卷

心理問卷的形式和普通問卷相似，內有許多問題，由被試者回答，從回答中觀察其品格、態度、興趣等。



心理問卷依問題的性質可分為直接的或明顯的和間接的或隱晦的兩種。所謂直接的問題，就是所問的就是實在所要知道的，大多數興趣問卷屬於這一類。所謂間接的問題，就是表面上問的是一樁事情，而所考查的實在是另外一樁事情，大多數品格問卷、情緒問卷屬於這一類。心理問卷我國已有數種，茲介紹於後。

一、蕭氏訂正個人事實表格 鄔馬個人事實表格 (Woodworth-Matthews Personal Data Sheet) 為確定兒童的情緒穩定性的一種工具。根據在美國試用結果，本表格似尙屬有效而可靠，現經蕭孝嶸加以修訂。

蕭氏修訂工作有幾點足述的：(一)各項目的去留。原表格有問題七十五則，修訂者刪去了五則，留七十則。刪去的原因，或是兒童不能了解，或是內容重複，或是試用結果有百分之六十以上人數的答案與鄔馬二氏的病態記分標準相符。在未經選擇的兒童中既有百分之六十在某項目上與標準相符，則此種反應不得視為病態的表徵；換言之，此項目並無區別常態與病態的功效，所以應當取消。(二)記分方法的改換。原表格的記分方法係以病態的反應為根據。凡與每項的病態反應（何者為病態反應由原編者規定）相符者得一分，所以分數愈大，表示愈有病態。蕭氏採用的方法恰與此相反。他是以常態的反應為根據的。凡與每項的常態反應（何者為常態反應由原編者規定）相符者得一分，所以分數愈大愈表示常態，分數愈小，愈表示病態。

蕭氏曾將本表嘗試用於小學四、五、六年級，計四年級男生二九六人，女生二〇一人，五年級男生三五〇人，女生二五三人，六年級男生二五〇人，女生一九〇人，年齡從九歲到十五歲，並求得各年級和各年齡平均分數。從這些結果，我們可以察見三個顯著的事實：（一）受試者的年級或年齡愈高，則其分數愈大；換言之，分數有與年級或年齡成正比的趨勢。此種趨勢在男女兩性中均無例外。這個事實可以表示情緒的穩定性係隨年齡而增進。（二）男性的情緒穩定性較高於女性。此種差異在各年齡中皆為一致。（三）男女兩性的差異度在情緒穩定性上大約相等。

用此結果以診斷病態，蕭氏主張以分配中央百分之五十為常模的範圍。凡在下四分點以下的可認為有病態的表徵。但此種標準係假定的，是否正確尚有待於事實的證明。換言之，這種心理問卷的效度尚待事實證明。

茲將蕭氏訂正個人事實表格錄後。下面表格中識以橫線之答案即為標準答案。受試者所圈消之答案如與一個標準答案相符，便可得一分。

蕭氏訂正個人事實表格

你在答覆下面各項問題的時候，對的應當在「是」字下劃一橫線，不對的應當在「否」字下劃一橫線。一切問題都是必須答覆的。

一、你喜歡獨自一個人玩耍而不喜歡和別的兒童玩耍麼？

是  否

二、別的兒童讓你和他們玩耍麼？

是  否



二十、你常常咬你的指甲使你的指頭痛麼？

二一、你從前是不是一個結巴子？

二二、你現在是不是一個結巴子？

二三、你在靜坐的時候可以心神安定麼？

二四、你從前有沒有搖動你的頭，或搖動你的肩的習慣呢？

二五、你曾經因為發了很大的脾氣，以致眼睛發紅麼？

二六、你因為碰到東西而跌倒的時候比別人多些麼？

二七、有沒有一種食物使你吃後要嘔吐呢？

二八、你平常都是快樂麼？

二九、你曾經覺得沒有一個人愛你麼？

三十、你常常容易生氣弄得哭起來麼？

三一、你容易因小的事情發生惱怒麼？

三二、你喜歡戲弄別人弄得他哭麼？

三三、你能够忍受痛苦和別人一樣麼？

三四、你看見血的時候能够不怕麼？

三五、你曾經覺得在弄傷別人的時候，有一種快樂麼？

三六、你曾經覺得在弄傷動物的時候，有一種快樂麼？

是 是 是 是 是 是 是 是 是 是 是 是 是 是 是 是  
| 否 否 否 否 否 否 否 否 否 否 否 否 否 否 否 否



五四、你曾經昏倒過麼？

五五、你常常昏倒麼？

五六、你是否常常覺得你的家庭對你不好，但是你又知他們的確是對你好呢？

五七、你的教師平常待你好麼？

五八、你曾經覺得一切的東西都是不實在的，而因此心裏不安麼？

五九、你曾經覺得有人跟着你走，而因此心裏不安麼？

六十、你曾經覺得有人謀害你麼？

六一、你在走過大街或露天地方的時候，覺得心裏不安麼？

六二、你坐在小房間裏把門關了的時候，覺得心裏不安麼？

六三、你在要決定一件事情的時候，覺得很難決定麼？

六四、你曾經很想放火去燒一件東西麼？

六五、你曾經很想偷東西麼？

六六、你想你所怕的東西比別人多些麼？

六七、你容易交朋友麼？

六八、你容易討厭別人麼？

六九、你很迷信麼？

七十、你曾經覺得你是一個很壞的人麼？

是 否

是 否

是 否

是 否

是 否

是 否

是 否

是 否

是 否

是 否

是 否

是 否

是 否

是 否

是 否

是 否

是 否

關於本心理問卷修訂的經過，可閱蕭孝嶸馬個人事實表格的修訂及其結果。（測驗第三卷第一期）

二、蕭氏修訂×○測驗 ×○測驗爲美心理學家濮來西 (S. I. Prescay) 所創製，用以研究情緒的發展與反常。此種測驗有A B兩種，A種爲用於成人者，B種爲用於兒童者。這兩種測驗均經蕭孝嶸加以初步的修訂。

B種測驗包含三個測驗，測驗一調查一人所憎惡之事物，測驗二調查一人所憂慮之事物，測驗三調查一人所愛好之事物。每個測驗原有一二五個字，修訂後減爲一百個語句。受試者須將各測驗中所憎惡的，曾經憂慮過的，或喜歡的事物畫消，然後再將各測驗每行（每個測驗有一百個語句，分二十行排列，每行有五個語句）中最憎惡，最憂慮或最喜歡的事物圈去。

蕭氏曾用該測驗在京滬漢三處各學校測驗了一千四百餘人，年齡從十一歲至二十歲，男女約各佔半數。統計之手續係先求得每個年齡中每種反應之次數，此即指每個年齡中畫消每個語句者之人數。此種次數以各該年齡之人數除之，即得各年齡每種反應之百分數。於是分受試者爲二組：一組之年齡爲十一至十五歲，其他一組之年齡爲十六至二十歲。前者爲年齡較幼之組，以Y代表之，後者爲年齡較長之組，以O代表之。於是在此二組中求其對於每個語句的反應之平均百分數而比較之。若某語句之反應此二組之差別至少爲百分之十，則此語句視爲可以區別二組者。此等語句分爲二類：一爲年齡較幼組多於年

齡較長組之語句，一爲年齡較長組多於年齡較幼組之語句。前者以  $\times \vee \circ$  代表之，後者以  $\circ \vee \times$  代表之。根據此種統計結果以計算各人分數，其手續爲確定在每人所畫消之語句中何者屬於  $\times \vee \circ$  類，何者屬於  $\circ \vee \times$  類。以後者之數減前者之數，即得此人之區別分數 (differential score)。此分數愈大表示情緒愈幼稚；分數愈小，表示情緒愈成熟。區別分數顯然是與年齡成反比的。

第二步統計手續爲確定各年齡在各測驗中每行所圈之語句以何者爲最多，其最多者稱爲「衆數」語句。由此可以察見各年齡中所最憎惡，最憂慮和最愛好之事物，並各年齡對於各種事物態度上的變化。根據此種統計結果以計算分數，可將各人所圈消者與「衆數」語句比較，計其不符之數。此種分數稱爲離中分數 (deviation score)。此種分數可以表示一人在所測驗之三種態度上與多數人差異的程度。若差異過大，卽有情緒反常之症候。

A 種測驗係用於成人的，共包含四種測驗：第一種測驗係關於所不喜歡之事物，第二種測驗係關於認爲具有關係之事物，第三種測驗係關於所憎惡之事物，第四種測驗係關於所憂慮或所懼怕之事物。統計手續和計分方法與 B 種測驗同。詳細情形可閱蕭孝嶸修訂  $\times \circ$  測驗之初步 (測驗二卷二期)。

茲將蕭氏修訂  $\times \circ$  測驗 B 種節錄於後：

$\times \circ$  測驗 (B 種)

測驗一



指導：在下面二十行中，倘有一個語句是指一件壞的事物，請你畫消這個語句。畫消多少聽你自便。但你必須把你所認為壞的事物完全畫消。

1. 謙恭 竊笑 小氣 精神衰弱 調戲

2. 輕看 卑鄙 身體殘缺 跳舞 迂緩

3. 吝嗇 玩弄別人 誇口 猶豫不決 女郎

.....

18. 家庭 宗教 說人壞話 虧本 考試不及格

19. 書 不守法律 信神 漂亮 吸煙

20. 橫槓 牌 罷課 胡思亂想 學校

### 測驗二

指導：在下列二十行中，倘有一個語句所指的事物是你曾經憂慮過的，或曾經使你覺得不自然的事物，請你畫消這個語句。畫消多少聽你自便。但你必須把你曾經憂慮過的事物或曾經使你覺得不自然的事物，完全畫去。

1. 污穢 神 找錯處 破壞 憂愁

2. 靈魂 不公平 失敗 刀 偵探

3. 良心 不測之禍 愁悶 失望 癩狂

.....

18. 記心不好 紅臉 父母 惡夢 藥物

- 19. 男子 胡思亂想 睡眠 耳語 警察
- 20. 癡心 思慕 金錢 犯罪 卑鄙

測驗三

在下面二十行中，倘有一個語句是指你所喜歡的事物（或你有興趣的事物），請你畫消這個語句。畫消多少聽你自便，但你必須把你所歡喜的事物完全畫消。

- 1. 算命 搖船 海邊 山 遊藝會
- 2. 露天生活 網球 遊山 小吃 娛樂公園
- 3. 科學 美術 商業 軍事 醫學
- .....

- 18. 祕書 工程師 看護 軍人 教師
- 19. 報 詩 辯論 書籍 雜誌
- 20. 醫生 政治家 大學教授 銀行家 科學家

第二步的指導：細心地看一遍，不要更改你已經畫了的記號；不要注意這些記號。在每個測驗每行中照着下面的規則去圈一個語句。

- 測驗一（在第一面）請你把每一行中一件最壞的事情加上一圈，每行必須有一個圈，不可遺漏一行。倘若你不能確定，請猜一猜。
- 測驗二（在第二面）請你把每一行中你曾經最憂慮的一件事情加上一圈，每行必須有一個圈，不可遺漏一行。倘若你不能確定，請猜一猜。

測驗三（在第三面）請你把每一行中你所最歡喜的一件事情加上一圈，每行必須有一個圈，不可遺漏一行。倘若你不能確定，請猜

不要注意你從前所畫的記號，你所圈的語句多半是已經有了記號的，但不一定是冇記號的。一個語句可以有一種記號，也可以同時有兩種記號。

請記得，在做完這個測驗以後，每一行中應當有一個圈。

全部測驗做完了以後，那就是說，畫消與打圈兩種手續都做完了以後，請你舉手以便主試知道你做完了。

三、朋路透品性測驗 本測驗係美國斯巨福大學教授朋路透 (R. G. Bernreuter) 所編，共一百二十五個問題，茲錄數題於後，以見其大概：

8. 你時常無故感覺不快樂否？
12. 你時常害羞面紅否？
89. 你在決定一件重要事件以前，是否喜歡徵求別人的意見？
103. 你同一素不相識的人談話感覺困難否？
125. 在討論會裏你自動的貢獻意見時，是否覺着怪不好意思的？

受試者祇須對每一問題選擇「是」「否」「？」三種答案之一，每題可照四種方法或標準批分，得以下四種分數：

甲、精神不健全的傾向 這項分數高的人情緒不平穩，分數低的人情緒平穩。

乙、自持性 這項分數高的人不嫌寂寞，難得求人，有不顧他人意見的傾向。分數低的人喜朋伴，時時

需要他人的鼓勵和勸告。

丙、丙外向 這項分數高的人興趣向內，富想像，生活於一己主觀的世界中。分數低的人精神向外，活潑多動，少憂慮，生活於客觀的實在世界中。

丁、自現和自卑 這項分數高的人和人交接佔權勢，善操縱。分數低的人和順而有畏縮的傾向。

例如對第八題回答「是」的，甲種分數七分，乙種分數負一分，丙種分數三分，丁種分數負一分。所以「時常無故感覺不快樂」是精神不健全的重要症候，同時也指示自持性弱，內向和自卑。此外如害羞面紅也是精神不健全和內向的表示，同時些微指示自持性弱和自卑。徵求別人意見的主要表示是自持性弱，和人談話感覺困難的主要表示是自卑。反之，在討論會裏喜歡自動貢獻意見是精神健全，自持性強，外向和自現的表示。四種分數的記分標準各不相同，但並非絕無關係。據編者兩次試用的結果，四種分數的相關係數如下：

甲與乙 負·三九（指示情緒不平穩的人有自持性弱的傾向，但並不顯著。）

甲與丙 負·九三至·九五（指示情緒不平穩的人趨於內向。）

甲與丁 負·八二至·八七（指示情緒不平穩的人趨於自卑。）

乙與丙 負·二八（指示自持性強的趨於外向，自持性低的人趨於內向，但關係並不深切。）

乙與丁 負·五〇至·五五（指示自持性強的趨於自現，自持性弱的趨於自卑。）

丙與丁 負·七三至·七五（指示內向的人趨於自卑，外向的人趨於自現。）

據編者報告，四種分數的信度自·八五至·九二，雖尚不及最好的智力測驗或教育測驗，已可辨別個人間的差異。若用以研究團體間的差異當然更無問題。編者又把本測驗結果和已經證實的同類測驗比較，計算效度係數，也都在·八四以上。

本測驗施行的手續至為簡單。答案沒有準不準的分別，不能作弊。時間也沒有限制，大約不消三十分鐘可以答完。欲保答案確實，必須受測驗者誠意合作，最好令答者自記分數，不經他人之手，祇少允許不將結果公開。若在學校等處施行團體的測驗而測驗的目的在得全體分配情形則可不記姓名；同時如受測驗者願知一己分數，可各用一他人不能查出的號數。

本測驗一百二十五個問題及記分標準已由沈有乾陳選善譯成中文，並由沈氏在杭州試用，求得百分位常模。沈氏求常模所根據的受試者係四〇五個中學生內中男生二〇四人，女生二〇一人。

據沈氏所得常模，甲乙丙三種分數男女相差極微，常模幾可通用，惟丁種分數則男女差異非常顯著。是項常模係根據中學生求得，若受試者非中學生或年齡與中學生不相若者，則上述常模即不適用。我們應當照性別、年齡、教育程度分別編造常模，以利應用。

關於本測驗，可閱沈有乾性情態度興趣的測驗（教育與職業一二八期）；沈有乾劉景琦朋路透氏品性測驗在中國之常模（教育與職業第一六一期）；沈有乾中國民族性之一斑（教育雜誌第二十五

## 卷第一號)

四 主要興趣測驗 本測驗係何清儒就 Allport & Vernon 合編的主要興趣測驗 (A Study of Values) 譯就，並參酌我國情形編成。在這測驗中，人的興趣分爲六種主要的，所有各種細微的興趣都歸納在這六種中。這種分法雖未免概括粗草，但是與普通思想極爲相近。所以在實用上，頗爲便利。這六種所代表的意義，分別簡略說明如下：

(一) 理論 理論的興趣，可說是研究真理的興趣。凡這項興趣濃厚的，對於科學的研究，和哲理的探討，都必熱心。這種人大多數是沉靜，有思想。對於細微的事亦極認真。

(二) 經濟 經濟的興趣，自然包括對於財務的注意。但此外對於一切實利的興趣，亦都在內。在這種興趣上高的人，凡事以實用爲目的。對於理想觀念的事，不甚注意。

(三) 藝術 凡對於藝術有興趣的，必歡喜文學、音樂、美術、工藝一類的事。這類的人對於一切事務祇求滿足美的觀念，而忽略真理和實用。

(四) 社會 有社會興趣的人，對於人的問題，極爲注意。在人與人的關係上，如家庭、親友等，或是對公共問題如衛生、慈善等，這種人都願竭力顧全。

(五) 政治 有領導、統制別人慾望的人，或是喜居高位、追求名譽的人，差不多都有濃厚的政治興趣。凡對行政組織及法律有信仰的，亦都有政治的趨向。這種人對於權利、義務分辨得很清楚。

(六) 宗教 宗教的興趣是無需說明的。凡信有超人的勢力，統管宇宙，或是以爲信仰可以改變人生的，都是宗教興趣的表現。對於儀式、禮節、感覺趣味的，亦可算是有宗教的趣向。

本測驗的功用就在辨別個人對於以上六種興趣的高低。本測驗分兩部分，第一部共三十題，第二部共十五題。茲將各部做法說明錄後並附題數則。

### 第一部

標明：方法：下面的話或問題每項有兩個替換的答案。將你個人的傾向在下面方格中，按下列方法用數目

假若你與(甲)同意，與(乙)不同意，在上格內

寫3，在下格內寫0；

假如你與(乙)同意，與(甲)不同意，在上格內

寫0，在下格內寫3；

假如你對(甲)的趨向比對(乙)的高，在上格

內寫2，在下格內寫1；

假如你對(乙)的趨向比對(甲)的高，在上格

內寫1，在下格內寫2。

1	2	0	3	(甲)
2	1	3	0	(乙)

除這四種數目以外，不可在問題後寫任何數目。這測驗沒有時間限制，但不可在某一項上遲延過久，並且除非真不能決定時，不可將某項遺漏。

(1) 科學研究的主要目的，應該是發現真理，不在實際應用。

(甲)是；(乙)否。

(2) 你想大藝術家專顧自己，而忽略別人的情感，是有理由的麼？(甲)是；(乙)否。

(3) 因為人有競爭和自強的天性，廢止戰爭是空幻的理想。

(甲)是；(乙)否。

(4) 假若你是大學教授，並且有必需的才能，你喜歡教那一樣？(甲)詩學；(乙)物理化學。

第二部

方法：下列問題每項有四個可能的答案。將這四個答案，按照你個人意見，排列先後，在左方格中寫出次序的數目：

最合你意見的寫

1

次要的答案寫

2

				(甲)
				(乙)



再次的寫

3

最不能代表你的興趣的寫<sup>4</sup>

你亦許想到別的與你意見相合的答案。但是請你務必由所提出的選擇，列成次序。如有不決，可以猜定。如實不能猜定，即可付缺。

(1) 你以爲一個好政府應該：

甲、多扶助貧苦，疾病合老弱的人。

乙、發展工商業。

丙、在政治與外交上採取道德手段。

丁、在國際上造成優勝的地位。

(2) 依你的意見，一個人在一星期內工作謀生，星期日最好是：

甲、讀科學的書以求深造。

乙、比賽運動。

丙、赴音樂會。

丁、聽德育演講。

記分方法由原編者規定。上例中第一部第一題甲答案表示理論的興趣，乙答案表示經濟的興趣。第

二題甲答案表示藝術的興趣，乙答案表示社會的興趣。第二部第一題甲答案表示社會的興趣，乙答案表示經濟的興趣，丙答案表示宗教的興趣，丁答案表示政治的興趣。第二題甲答案表示理論的興趣，乙答案表示政治的興趣，丙答案表示藝術的興趣，丁答案表示宗教的興趣。最後由各項興趣總分數之多少以定個人興趣的趨向。

本測驗由何清儒在清華大學試用過。本測驗的效度似有待於事實之證明。可參閱何清儒主要興趣測驗（教育與職業第一四一期）。

五、職業興趣測驗 中華職業教育社編有職業興趣測驗一種。本測驗係仿美國斯旦福大學教授施厥浪（E. K. Strong）所編的職業興趣測驗，共分七節，計一百九十五問題。前五節的問題為對於各種職業、工作、娛樂、學程、人物的好惡。第六節每題舉職業或其他情境兩種，受試者須表明其興趣的先後。第七節甲部舉工作十項，受試者須指出他最願作的及最不願作的各三項；乙部舉擇業標準十項，受試者須指出其最重視者及最輕視者各三項。第一節共五十題，第二、三、四、五、六節各二十五題，第七節甲乙兩部各十條。現在把每部分的說明抄在下面，並各舉數例：

## I

以下各種職業的工作你願做否？酬勞的優劣，社會上地位和進步機會且不管，祇看你喜歡不喜歡做每種職業的工作，若喜歡那種工作，將好字加一圈，若討厭那種工作，將惡字加一圈。若既不喜歡又不討厭，將平字加一圈。每項必須回答，但不必細心考慮，不妨圈得快些，因

爲我們要的是你頃刻間的態度。

建築工程師.....好平惡

機械工程師.....好平惡

土木工程師.....好平惡

電機工程師.....好平惡

軍官.....好平惡

## II

以下各種動作，照上面方法看你喜歡或討厭它，加圈於適當的字。快圈，不可遺漏。

修理鐘表.....好平惡

診裝電綫.....好平惡

種植花木.....好平惡

急救受傷者.....好平惡

佈置房屋內部陳設.....好平惡

## III

以下各種娛樂方法，也照前面辦法將適當的字圈出。快圈，不可遺漏。

釣魚.....好平惡

打獵.....好平惡

- 拍網球 ..... 好平惡
- 公園散步 ..... 好平惡
- 打拳 ..... 好平惡

IV

以下各種學校學程，也請圈出你的好惡。快圈，不可遺漏。

- 文學 ..... 好平惡
- 心理學 ..... 好平惡
- 植物學 ..... 好平惡
- 動物學 ..... 好平惡
- 地理 ..... 好平惡

V

以下各種不同的人，也請圈出你的好惡。快圈，不可遺漏。

- 樂觀的人 ..... 好平惡
- 悲觀的人 ..... 好平惡
- 守舊的人 ..... 好平惡
- 時向人借東西的人 ..... 好平惡

VI

以下各項，每相對的兩項中，請你挑選一項。如喜歡上面一項，請在上面空白中加一鈎。如喜歡下面，在下面加一鈎，如兩項相同，在中間加一鈎。

- |      |       |     |     |     |     |        |
|------|-------|-----|-----|-----|-----|--------|
| 電車司機 | ..... | ( ) | ( ) | ( ) | ( ) | 電車售票   |
| 警察   | ..... | ( ) | ( ) | ( ) | ( ) | 救火員    |
| 車夫   | ..... | ( ) | ( ) | ( ) | ( ) | 廚司     |
| 設計   | ..... | ( ) | ( ) | ( ) | ( ) | 施行已定計畫 |
| 親自辦理 | ..... | ( ) | ( ) | ( ) | ( ) | 請人代辦   |

VII

A. 下列關於某種機器（例如打字機，汽車）的十項工作，請挑選你最願做的三項，和最不願做的三項在括弧內鈎出。

- |       |       |     |     |     |            |
|-------|-------|-----|-----|-----|------------|
| 最願的三項 | ..... | ( ) | ( ) | ( ) | 創造新機器學說    |
|       | ..... | ( ) | ( ) | ( ) | 使用這新機器     |
|       | ..... | ( ) | ( ) | ( ) | 改良這新機器的一部份 |
|       | ..... | ( ) | ( ) | ( ) | 計算使用這機器的費用 |
|       | ..... | ( ) | ( ) | ( ) | 監督製造這機器    |

B. 你選擇職業時，下列各項中，最重視那三項，最輕視那三項，請在括弧內鈎出。

最重三項

最輕三項

( )	.....	( )	.....	薪水
( )	.....	( )	.....	位置穩固
( )	.....	( )	.....	升任的機會
( )	.....	( )	.....	上司以禮相待
( )	.....	( )	.....	能利用所有智識及經驗

本測驗的功用，在考查職業興趣的趨向以爲職業指導及職業選擇之參考。應用以前須將本測驗試用於各職業界中人，根據所得結果，確定每題對於每兩種職業的鑑別力，然後釐訂每種職業的記分標準。此種手續自極繁重。郝雁題謝涵在沈有乾指導之下，根據八十二位大學教育組學生及二百二十八位非教育組學生（內中理工數學系學生二百十五人，英文系學生十三人）在本測驗上所得結果，釐訂教育界記分標準。照此標準記分，教育組與非教育組在本測驗上分數的差別確甚顯著，計教育組中數爲八八·八，而非教育組中數爲負三六·三。若將英文系學生十三人除外，非教育組尙餘二百十五人，可稱爲理工組。理工組的中數是負四〇，教育組中只有一人不及這分數，而教育組的中數八八·八，理工組也只有  
一人能超過，可證興趣測驗可成爲極有用的工具。爲推廣本測驗的效用起見，我們須做照同樣手續，釐訂各職業界記分標準。

本測驗的信度亦經有人調查。陳選善曾在大夏大學將此測驗試教育系學生二十一人，隔一月後用原題再試之，察其兩次答案之異同，可以見測驗之信度。所得結果經沈有乾用統計方法加以分析，證明兩次所作答案與亂答迥不相同，各節測驗的信度約在·九〇左右。後沈有乾又用郝謝兩君求得之教育興趣之記分標準，比較受試者兩次在測驗每節所得分數，得全測驗的相關係數爲·七一。總之根據沈氏所定記分標準用以比較教育組學生教育興趣之高低雖尚嫌不甚可靠，但若用以辨別各種人之教育興趣固可勝任而有餘。關於本測驗有下列四種參考：

沈有乾本社所編興趣測驗之可靠性（教育與職業第一三三四、五期）

沈有乾續談本社興趣測驗之可靠性（教育與職業第一四〇期）

郝雁題謝涵教育組學生與非教育組學生興趣之辨別（教育與職業第一三七期）

沈有乾續談教育組與理工組學生興趣之不同（教育與職業第一三八期）

### 三 評判量表

主觀評判的不可靠乃吾人所熟知之事實，但心理測量根本上事實上都離不了主觀的評判。心理測量根本上離不了主觀的評判，因爲主觀的評判往往是考查「客觀」測驗最後的根據。心理測驗事實上還離不了主觀的評判，因爲性格的考查尚不能完全用客觀的測驗，仍有賴於主觀的判斷。所以評判雖然

有許多困難，仍不失為一種有用的方法。

評判的方式，有排列法、對比法、人與人比較法、指定等級法、直線評判法數種，茲簡單說明於後。

一、排列法 譬如評判一班學生的誠實，可由教師將這班學生依其誠實的程，度排列起來，最誠實的列第一，最不誠實的列最後。這種方法的優點，是將一班學生互相比較，而不是將每個學生同一種抽象的標準比較。如是，可以避免因評判者心中所懸標準之不同（如標準太高或太低）而發生常性的誤差。其缺點是一班人數過多則排列起來非常麻煩。

二、對比法 對比法亦是一種互相比較，而手續比排列法更為繁重。用對比法時，評判者須將一班中每位學生同其他學生個別比較。例如有甲乙丙丁四人，我們先將甲和乙、丙和丁比較，再將丙和丁比較，然後應用統計方法用數字表示四人的程度。

三、三人與人比較法 用這方法時，先由評判者選出三人或五人，代表某項性格或能力的三個或五個等級，然後將要評量的人與這三五人各各比較，決定與何人最相近，即列入此人所代表的等級。這種方法好像利用作文量表評判作文一樣，不過以實在的人代替了作文的樣本。這種方法很具體，但因各人所選定的「標準」人不同，所得評判就不能互相比較。

四、指定等級法 將各評判的項目分為五個或七個等級。評判量表上各級之下，加以註釋，以示各級所代表之意義，由評判者指出被評者所居的等級。表十是這種評判量表的一個實例。



表十 學生性格評量表

受評量者姓名\_\_\_\_\_作評量者姓名\_\_\_\_\_

學校\_\_\_\_\_年級\_\_\_\_\_年\_\_\_\_\_月\_\_\_\_\_日

1. 下列各項性格，每項名目後之註釋代表該項所包含之意義或事項。每評量一項時請專就該項着想。
2. 每項性格分爲五級，每級下之註釋代表該級之意義。受評量者與某級最相近，即請在某級下作一✓號。
3. 如同時評量數人，可將此數人在同一性格上先作比較，然後決定高下。
4. 個人之好惡須力求避免。個人所喜之人未必在各項性格上居高，個人所惡者未必盡無可取之處。
5. 評量結果對受評量者完全保守祕密，故作評量者可秉公判定，不受拘束。

項 目 \ 等 級	最 優	優	中	下	最 下
外表—服裝，態度， 修飾，舉止	(令人敬重)	(整潔大方)	(平常)	(不甚整齊)	(令人生厭)
思想—觀察，分析， 組織，理論	(極有條理)	(清楚不亂)	(平常)	(常不合理)	(紊亂無序)
才力—學習，會通， 運用，創造	(出乎常人)	(頗爲優強)	(平常)	(常顯薄弱)	(遲鈍不敏)
信用—言語，時刻， 交往，事務	(永可信任)	(誠實可靠)	(平常)	(常不可靠)	(毫無信用)
處羣—交友，合作， 公德，合羣	(受人推戴)	(頗能合作)	(平常)	(常有衝突)	(衆怒所歸)
領袖力—指揮，引導， 措置，應付	(事事當先)	(善於領導)	(平常)	(祇受命令)	(事事被動)
責任心—盡職，耐勞， 進取，專心	(不計艱難)	(盡心從事)	(平常)	(避免責任)	(無責任心)
人格總評	(高尚出羣)	(極爲優良)	(平常)	(極多缺點)	(卑鄙惡劣)

五、直線評判法 將各項性格或能力用一直線表示，一端代表最高點，一端代表最低點，中間亦不妨加入幾條界線，代表中段各等程度。兩端或中間界線下面可加註釋以備評判者之參考。評判者可在直線上任何一點作一記號，以示受評者之程度。

直線評判法（或稱圖示評判法）與指定等級法並無多大區別，不過直線評判法並不強制評判者將受評者分列五個或七個等級，可作較為精細的評判。不過人的評判根本並不十分準確，評判量表分得細，評判的準確度實際上並不因此而增加，反而加重了評判的困難。塞門司的研究指示評判量表上的等級大概不必多於七等。

六、寫像配合法 這方法是哈之兄和梅所創始的。他們曾用之以考查兒童的「服務性」(Service)。評判量表之編製及應用和作文量表書法量表差不多，不過以一段一段文字的描寫，代替了普通品質量表上的作品樣本。他們還有一種方法叫做「猜猜看是誰」(Guess Who Test)與寫像配合法是大同小異的。

關於評判法的研究，有許多重要的結果，茲擇其最切要者分條簡述於下，以為採用評判法者之參考。

(一) 評判的項目必須加以明確的解釋。將一種性能分析為具體的項目，可以增加評判的效度，如蕭孝嶸訂正勒氏內外傾品質評定量表，包含四十個項目。蕭氏訂正馬氏通內外傾品質評定量表包含十五個項目。

(二) 評判須根據於被評者以往及現在的行為。評判者對於被評者事前應有長期的觀察。

(三) 評判者與被評者應相識有素，但二人之間關係不可太密切。太密切者易有偏見，有偏見就要減少評判的效度。據實驗結果，如評判者與被評者有很深的交誼，則對於良好性格的評判偏於太高，對於不良性格的評判偏於太低。

(四) 評判的準確度依所評判的性能而異。這就是說有的性能容易評判，有的難於評判。大概一種性能愈客觀，則愈容易評判。所以用評判法度量性格時，最好限於比較客觀的性能。

(五) 人的評判有寬嚴的不同，這就是說各人在評判時心中所懸的標準有高低的不同。這種常性的誤差，如經發覺，可以用統計方法校正之。

(六) 評判有一種普通的誤差就是所謂「回音的影響」。這就是評判者對於某人某種特殊性能的影響，無形中受他對於這人普通好惡的影響。

避免或減少回音的影響，可有數種方法：(1) 向評判者指出「回音的影響」的作用，促其注意。(2) 如要評判一班兒童在幾種性能上的程度，應先評判全體兒童的甲種性能，然後評判他們的乙種性能，不要先評判張生在各種性能上的程度，再評判李生在各種性能上的程度。簡言之，評判的手續應以性能為單位，不要以人為單位。

(七) 衆人評判比較個人評判可靠。所以採用評判法時，應有五個至十個獨立的評判，然後求其平均。

(八) 評判者必須對於評判量表的應用有相當的訓練與經驗。訓練可以增加評判的信度。

註一 引錄沈有乾教育心理一二五——二七頁。

## 第九講 測驗結果的解釋

一人在一種測驗上所得分數普通以答對題數計算，答對幾題就得幾分。這種分數可稱為原來分數 (raw scores)。我們略加思索，就可以知道這種原來分數，並無確定的意義，所以必須加以解釋。在第三講所討論的常模，就是用來解釋原來分數的。此外測驗上還有幾種常用的量數，亦可作為解釋原來分數之用，這種分數可稱為引伸分數或解釋的分數 (derived or interpretative scores)。

### 一 年齡分數

一人的智力、情緒、生理組織、教育程度，都依年齡而逐漸發展，所以一人各方面發展的程度可以用年齡來表示。因此我們可以有智力年齡、情緒年齡、生理年齡、教育年齡等。茲就智力年齡與教育年齡加以申說。

年齡分數的意義極易於了解。譬如說，某人的智力年齡是十歲，這就是說此人的智力相當於普通十歲兒童。某人的教育年齡是八歲，這就是說此人的教育程度相當於普通八歲兒童。

年齡量表，如比納智力量表上的分數，是直接以年齡表示的。團體測驗上的分數，普通以答對題數為單位。要將這種分數轉化年齡分數，必須有對照表。我們須先求各年齡兒童在測驗上所得之平均分數。例

如某人在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，或者說他的智力年齡是十歲。從上面的說明，讀者不難領悟到年齡分數與年齡常模實在是一件事情。

智力年齡有幾種限制，在應用時須加注意。

一、單位不等。這是因為智力的發育曲線大概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。身長的發育如此，智力的發育大概亦如此。據我們現在所知，四歲到五歲一年之間智力發展快，或生長的數量大；十四歲到十五歲一年之間智力發展慢，或生長的數量小。

二、年齡量表不夠廣闊。智力的發展不是永久繼續，全無限制的，總有停止的時期。智力的發展究竟在幾歲停止不是容易解決的問題，而且恐怕不是人人一律的。假定智力的發展到十五歲停止，那麼我們就不能有十五歲以上的智力年齡了。一個聰明兒童在智力測驗上所得分數超過了十五歲兒童的平均分數，我們就沒有法子再以年齡分數去表示他的智力了。這是智力年齡分數在應用上的限制。因此賽門司和露西 (G. M. Ruch) 都說智力年齡不適用於中學階段。

三、求各年齡兒童平均分數（實在就是年齡常模）時取樣的困難。這一點在第三講討論年齡常模時已經說過了。

四、解釋上的限制。兩個人智齡相等，假使實在年齡相差很遠，不能認為智能相同。智齡三歲的低能成人與常態三歲兒童的智能絕對不能比擬。克能海 (K. S. Cunningham) 用桑戴克所編 C A V D 智力

測驗，比較智力年齡相同的常態兒童與低能成人，顯示他們在能力組織上有顯著的差別。對於常態兒童很難的項目，低能成人反以為容易。常態兒童以為容易的題目，低能成人反覺得很難。滿利用比納智力量表比較年齡相同的高才、常態和低能兒童，安特律其（C. G. Aldrich）和道爾（E. A. Doll）用格賽兒（A. Gesell）兒童發育程序單與墨跋量表比較同智力程度的常態兒童與低能兒童，亦得同樣的結果。總之，年齡分數的應用，將一個兒童的作業與他年齡兒童的作業比較，多少總有點勉強而不自然。

教育年齡與學科年齡的限制，大致與智力年齡相同。

## 二 年級分數

兒童年齡和所在年級大致是正相關的，所以兒童的智慧和教育程度，在相當限度內，亦隨年級而遞增。因此我們可以用年級的單位，來表示一人的智力與教育程度。從測驗上所得分數轉化年級分數，必須有對照表。我們須先求得各年級兒童在測驗上之平均分數，換言之，就是年級常模。譬如某人在教育測驗上所得分數與六年級兒童平均分數相等，那麼他的教育年級分數就等於六·五。

年級分數的優點就是易於解釋和便於應用。其缺點為（一）單位不等，（二）易受編級制度的影響，缺乏固定性，（三）有些科目並非各級俱備，因此成績並不隨年級而遞增。麥柯爾的C分數亦是一種年級分數。

### 三 百分數

百分數的計算，或根據年級或根據年齡，其意義在第三講討論常模時已經說過。將原來分數轉化百分數的方法，可查閱任何統計學書籍，此處不敘述了。一種測驗採用百分數的，在說明書上備有現成對照表，可以查用。

百分數的優點爲（一）易於了解；（二）用百分數是將一個兒童的作業與同年齡或同年級兒童的作業比較，較年齡分數年級分數的解釋來得自然；（三）百分數在中學大學階段均可適用。

百分數的缺點爲（一）單位不等。在百分量表的兩端，兩個百分單位間的距離大，在量表中段，兩個百分單位間的距離小。不過這一種缺點往往言過其實，因爲在常態曲線的中段，百分數與標準分數（標準分數的意義後面再講）有近乎直線的相關，除掉在百分之八十五度以上或百分之十五度以下。（二）兩個絕端的百分數往往不甚可靠。（三）百分數不能用數學的方法來處置。

### 四 標準分數

標準分數（Standard Score）的計算或者根據年齡團體，或者根據年級團體。假定分配是常態的，從一人的標準分數可以曉得在同年齡或同年級兒童中有百分之幾分數超過或不如他的。從原來分數轉



化標準分數可以用下列公式

$$z = \frac{X - M}{\sigma}$$

公式內  $z$  代表標準分數， $X$  代表某人的原來分數， $M$  代表某年齡團體或年級團體的平均分數  $\sigma$  代表該年齡團體或年級團體分數分配的標準差。所以標準分數是某兒童的分數與同年齡或同年級兒童的平均分數的差別，此種差別以該團體分配的標準差為單位表示之。假定某甲年齡十歲，在某種測驗上得三十分。要將此原來分數轉化標準分數，先要曉得十歲兒童在該測驗上的平均分數及其分配之標準差。假定平均分數為二十，標準差為五，則某甲之標準分數為

$$z = \frac{30 - 20}{5} = +2$$

換言之，某甲在同年齡兒童中處在該團體平均點以上兩個標準差的地位。假使我們要曉得在十歲兒童中有百分之幾的分數超過他，百分之幾的分數不如他，可查閱「常態曲線下面積計算表」這種表。普通統計學書籍附錄中都可以找到。

標準分數的優點為（一）標準分數，係根據統計量數中最穩定的平均數及標準差求得，所以較百分數為可靠而穩定。（二）標準分數可以用數學方法來處置。（三）單位相等。標準分數的缺點是難於了解不

夠通俗。

## 五 T 分數

從標準分數的公式，我們可以見到標準分數可正可負，而且標準差的單位太大，所以標準分數往往有小數。為避免這兩種麻煩，柯爾提議用標準差十分之一為計算的單位（如此可避免小數）而以分配之平均點下五個標準差為計算的起點（如此可避免負數）。這種分數叫做 T 分數。T 分數實在是一種變相的標準分數。

從原來分數求 T 分數可用下列公式：

$$T = 50 + \frac{10(X - M)}{\sigma}$$

用前面計算標準分數的例子，則某甲的 T 分數為

$$T = 50 + \frac{10(30 - 20)}{5} = 70$$

T 分數與 z 分數的轉化可以用下列公式：

$$T = 50 + 10z$$

T分數的解釋，亦要根據「常態曲線下面積計算表」，不過須稍加推算。麥柯爾 *How To Measure in Education* 書內第二十三表可以用來直接解釋 T 分數，不必再加推算。茲節錄該表之一部分如下：

表十一 常態分配下 T 分數之解釋

T 分數	百分數	T 分數	百分數
10	99.9968	55	30.85
15	99.977	60	15.87
20	99.865	65	6.68
25	99.38	70	2.28
30	97.72	75	0.62
35	93.32	80	0.13
40	84.13	85	0.023
45	69.15	90	0.0032
50	50.00		

這個表告訴我們在常態曲線橫坐標任何一點之上有全體面積百分之幾。假如一種測驗是根據十歲兒童標準化的，某生在測驗上得 T 分數 40，則十歲兒童中分數超過他的約有百分之八十四，不如他的約有百分之十六。讀者所應注意的，就是根據這個表來解釋 T 分數，我們是假定測驗標準化所根據團體的分數分配是合乎常態的。

編造T量表根據任何團體都可以。麥柯爾曾提議一切測驗的標準化以十二歲兒童為根據。他的理由是（1）十二歲兒童容易尋到不致有取樣的困難。（2）十二歲兒童能力或作業的差異度極大，而且十二歲兒童能力或作業的分配，在整個分配上所占的地位比較最為適中。這兩點是否與事實相符似乎頗有討論之餘地。我國在民國十一年所編造的測驗大都是照麥氏的意思，以十二歲兒童為標準化的團體的。

## 六 智力商數與教育商數

智力商數的意義，我們在第二講中已約略述及。智力商數乃智力年齡與實在年齡的比例。或者可以說智力商數是一人之智力程度與同年齡兒童之智力程度的比例。這種分數一方面表示一人以往智力發展的速率一方面可以作為預測將來智力發展的趨向。

不過智力商數在理論上有許多問題，在應用上有許多困難。先討論理論上的問題。

一、智力商數的固定性 一人的智力商數逐年有無變動？假定大致固定而變動很小，那麼根據一人現在的智力商數，就可以推測他將來的智力發展。假定變動很大，那麼這種推測就不可能。這是所謂智商固定性的問題。對於此問題的許多實驗，指示一人的智力商數，逐年多少有些變動，但假定環境上無重大的變易，而所用測驗相當可靠，則逐年有保持其原來比較的地位的趨勢。大概智商的固定性依所用測驗

(團體測驗或個人測驗)兒童年齡及兩次測驗相隔期間的久暫而異。從個人測驗所得智商較團體測驗的固定。六歲以上兒童的智商較六歲以下兒童的固定。兩次測驗相隔期間愈短，則智商變動愈小。晚近的研究又指示兒童所處環境上的變異往往有使智商發生大量變動的可能。

二、智商意義的一致性 所謂智商意義的一致性，即謂同一智商在各年齡有一致的意義。這是引用智商者的一個基本假設。換言之，在各年齡兒童智力分配上占同一點地位的兒童，其智商應相等。譬如在各分配上下四分點的兒童其智商應相等。反過來說，兩個年齡不同而智商相同的兒童在各該年齡分配上所占的相對地位是相同的。一個六歲兒童得智商八〇與一個十二歲兒童得智商八〇，此二人在各該年齡分配上應占相等的地位。簡言之，各年齡智商的差異度應相等。此與前節所討論個人智商的固定性完全是兩個問題。

表十二 第二次斯丹森  
比納量表各年齡智商差異度  
(從 Terman)

年齡	人數	智商的標準差	
		量表L	量表M
2	102	16.7	15.5
2½	102	20.6	20.7
3	99	19.0	18.7
3½	103	17.3	16.3
4	105	16.9	15.6
4½	101	16.2	15.3
5	109	14.2	14.1
5½	110	14.3	14.0
6	203	12.5	13.2
7	202	16.2	15.6
8	203	15.8	15.5
9	204	16.4	16.7
10	201	16.5	15.9
11	204	18.0	17.3
12	202	20.0	19.5
13	204	17.9	17.8
14	202	16.1	16.7
15	107	19.0	19.3
16	102	16.5	17.4
17	109	14.5	14.3
18	101	17.2	16.6

從比納式測驗所得的智商似乎能相當的滿足這一個條件。例如第二次斯丹福比納量表，其各年齡智商的差異度如上表所示。

表十二顯示各年齡智商差異度雖然不能絕對相等，除年齡六歲及十二歲外，相差極微。智商差異度與年齡並無一致的關係。

這一個條件即同一智商在各年齡意義相同，並非所有測驗均能適合。例如蕭孝嶸修訂墨跋量表所得結果對此條件即不能滿足。據蕭氏所得結果，智商七十一在曆齡三十五個月的分配上占  $1.1\%$  的地位，在曆齡四十九個月的分配上占  $1.2\%$  的地位，在曆齡六十八個月的分配上占  $1.2\%$  的地位。（參看蕭孝嶸修訂墨跋量表第十一表）同一個智商，因被試者年齡不同，其意義亦因之而異。

要使各年齡智商有一致的意義，則各年齡測驗分數的差異度與各年齡的平均分數或常模應保持某種確定的關係。這一點費里門（註一）討論得最爲詳盡。假定智力發育的曲線爲一直線（年齡量表就是這樣假設的），則智力年齡的差異必需逐年增加與年齡成比例。假如六歲兒童的智齡分配的標準差是十二個月，那麼到九歲，智齡分配的標準差，應該是十八個月，到十二歲應該是二十四個月，以此類推。假如不能適合上述情形，那麼在六歲得智商一二〇與在其他年齡得智商一二〇，其意義就不能互相比擬。

智力商數在應用上的困難有下列幾點：

一、第一個困難即計算高年齡兒童的智力商數時，公式內應該用什麼除數（實在年齡），沒有一定

的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。例如推孟用十六（最近用十五）爲求成人智商的除數，賓特納主張用十四。

二、許多研究指示人們智力的生長並不在同一年齡停止。大概聰明的人，智力生長停止較遲，愚笨的人智力生長停止較早。假定尋常兒童智力生長十六歲停止，愚笨兒童十三歲就停止，最聰明兒童要延遲到二十歲，計算智商時如以十六歲爲最大的除數，則愚笨兒童十三歲以後的智商要逐漸減少，聰明兒童十六歲以後的智商就要逐漸增加了。

因爲有上述兩種困難，賽門認爲智力商數不適於中學階段。

教育商數表示一人的教育程度與同年齡兒童教育程度的比較。要使教育商數在各年齡有一致的意義，則各年齡教商分配的差異度亦須相等。教育商數的引用並不假定一人逐年的教商固定不變。顯然，一人的教商依其所處教育環境和自身努力而變異。教商的應用最宜於小學階段，到了中學就不大適用了。計算教育商數的公式如下：

$$\text{教育商數} = \frac{\text{教育年齡}}{\text{實在年齡}} \times 100$$

## 七 B 分數

麥柯爾的B分數，其功用與年齡量表中的智力商數和教育商數同。B分數表示一人的能力或程度與同年齡兒童的比較。一人的T分數，假定T量表是根據十二歲兒童編造的，表示他在十二歲兒童中所居的地位，一人的B分數表示他在同年齡兒童中所居的地位。假定一人的智力或教育程度恰等於同年齡兒童的平均則其智力或教育B分數應為50。所以一人的B分數實在就是根據他同年齡兒童所編造的T量表上的T分數。顯然的，表十亦可以拿來解釋B分數的。普通求B分數的方法，是在T分數上加一個校正數。校正數各測驗不同，在同一測驗中各年齡不同。假定T分數是根據十二歲兒童的，則年齡小於十二歲半的兒童，其校正數為正數，年齡大於十二歲半的，其校正數為負數，十二歲半的兒童其校正數為0。換言之，十二歲半的兒童，其T分數與B分數是相等的。

各年齡校正數，各種測驗說明書中均有規定，一查就行了。現在舉一個例子以說明校正數的意義及由來。在廖世承團體智力測驗量表甲上，得41分者其T分數為50，得35分者其T分數為46，得49分者，其T分數為54。根據測驗結果，十一歲兒童（平均年齡應為十一歲半）在該測驗上的平均分數恰為35，十三歲兒童（平均年齡應為十三歲半）的平均分數恰為49。設有一兒童年齡恰為十一歲半，在測驗上得35分其T分數為46，表示他的程度比十二歲兒童的平均程度低四個T。但是這個兒童的分數恰與同年齡兒童的平均分數相等，照定義，他的B分數應該是50， $50 - 46 = 4$ ，這就是他的校正數，亦是其他十一歲半兒童的校正數。再假設有一個兒童，年齡恰為十三歲半，在測驗上得49分其T分數為54，表示他的



程度比十二歲兒童的平均程度高四個T，但是他的分數恰與同年齡兒童的平均分數相等，照定義他的B分數亦應該是50。  $50 + 24 = 74$ ，這就是他校正分數，亦是其他十三歲半兒童的校正分數。

從上面的說明我們可以明瞭B校正分數就是T 50與各年齡兒童平均T分數的差別。所以要決定各年齡的B校正分數，須先求出各年齡的平均T分數，即各年齡的T分數常模。在討論年齡分數時我們曾提及求年齡常模之困難，為年齡分數的一種限制。這種限制B分數亦不能避免。

## 八 成績商數

成績商數代表一人學業與智力的比較。這種分數一方面可以代表學生的努力程度，一方面可以代表教學的效率。因此有人提議引用成績商數為報告學生成績和考查教師效率的根據。

成績商數的公式如下：

$$A.Q. = \frac{E.A.}{M.A.} \times 100$$

另一種寫法是：

$$A.Q. = \frac{E.Q.}{I.Q.} \times 100$$

假使教育測驗和智力測驗並非同時舉行的，計算成績商數祇能用第二公式。

成績商數確實代表教育上一種重要的概念。但這種分數常被人誤解而且本身亦有缺點，茲分別論之。

一、有人根據能力不能超過能量的臆說，以為一人的成績商數不能超過一（或一百）。這種臆說在理論上和事實上都不能成立。智力測驗所測量的並不是一種完全天賦而不受環境影響的能量，所謂「智力」實在不過是各種能力的平均而已。教育測驗祇能表示某一方面的能力。很自然的，某一方面的能力有超過或低於平均能力的可能，所以成績商數可以超過或低於一的數量。事實上兒童的成績商數有超過一的，有低於一的，並不限於一以內的。

二、根據能力不能超過能量的臆說，有人以為假定努力和其他因素的影響能夠除去，則智力與學業之間應呈完全的正相關。這種說法亦是一種誤解。智力與學業之間，大致而論具有正相關的關係但相關並不完全。而且智力與各學科間的相關大有高低深淺之別。假設智力與某科作業是毫無相關的，則智商與成績商數之間，呈負相關之關係，乃是無可避免的結果了。

從應用成績商數所得結果，智商高的兒童，成績商數低，智商低的兒童，成績商數高。一部分原因當然是現在學校編級不能充分顧到兒童的智力，一部分原因恐怕是各科作業與普通智力不能有完全的相關。為克服或減少這種困難，有兩種提議值得考慮的。第一種提議就是求成績商數時，公式中的除數，不要一律用普通智力測驗的分數。譬如求讀法的成績商數，除數可用文字智力測驗的分數，求數理學科的成

績商數，可用數理智力測驗分數。第二種提議是成績商數的公式，除數不用智齡而用同智齡者的平均教齡。這種成績商數有人提議稱爲「迴歸成績商數」(Regression Accomplishment Quotient)，其公式如下：

$$A.Q. = \frac{E.A.}{E.A.}$$

這種迴歸成績商數可以表示一人的學業與同智力程度兒童的平均學業的比較。

三、成績商數本身的缺點就是信度太低。成績商數既然根據教育年齡與智力年齡，假如這兩種分數信度不高，則成績商數的信度當然更低。更重要的事實就是現在所謂智力測驗與教育測驗所測量的對象，一部分或大部分是相同的，此兩種測驗結果的差別，一部分或大部分是隨機的，不可靠的。

據克萊的觀察，現行智力測驗與匯合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績商數信度極低。智力測驗與單獨學科測驗的內容相同的部分當然要少些，因此各學科的成績商數亦比較可靠些。還有團體的成績商數當然要比個人的成績商數可靠。根據這種觀察，克萊主張將成績商數廢棄不用。(註二)

四、在前面討論年齡分數時，我們說年齡分數不適用於中學階段，成績商數既然是根據年齡分數的，所以成績商數亦不適用於中學階段。

五、要使教育年齡與智力年齡可以比較，則兩種測驗的常模須根據相似的團體。這個條件事實上未必能適合。

總結以上關於成績商數的討論，我們可以說成績商數代表教育上一種很重要而有用的概念。但計算成績商數的方法，尚有許多缺點，尤其是信度太低。假使成績商數的應用限於團體的測量，限於小學階段，限於成績商數特別高或特別低的學生，則或可避免或減少解釋上的錯誤。

成績商數的低落，除掉上面所說因為智力與學習各科所須要的能力並不完全相關，所以對於智力高的兒童為不可避免的現象外，還有許多別的原因，如學生學習不努力，學級編製失當，教師效率太低等。無論如何學生成績商數太低，乃是教學與學生不相適應的表徵，我們應該進一步研究其原因之所在，以圖補救而謀改進。

教育成績與智力的比較當然不限於商數的方法，而且不限於教齡與智齡的比較。我們亦可以用教育分數與智力分數的差數來表示。麥柯爾的F分數就是一種差數，其公式為

$$F = T_e - T_i + 50$$

公式內F代表 Effort 或 Efficiency，可稱為努力分數或效率分數， $T_e$ 是教育測驗T分數， $T_i$ 是智力測驗T分數。假如 $T_e$ 和 $T_i$ 相等則F分數為50。

另外一種方法是用標準分數，求教育測驗上和智力測驗上標準分數的差數。為避免負數計，可加上

一百或五十。公式如下：

$$F = z_0 - z_1 + 50$$

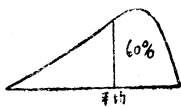
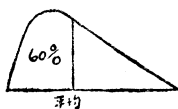
賽孟司的「勤學指數」(Index of Studiousness)就是用上列公式計算的。(註三)

F分數當然與成業商數受同樣的限制。

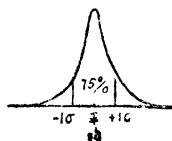
## 九 分數的比較性

上面我們解釋測驗上幾種通用的分數。現在所要討論的是從不同的測驗所得的年齡分數，年級分數，標準分數，百分數及各種商數是否可以直接比較。換言之，從兩種測驗得來價值相等的分數是否具有相同的意義。我們的回答是兩個分數的比較性不能任意假定，而應該進一步去考查各該測驗標準化時所根據的團體是否有比較的可能。第一、兩種測驗的標準化是否根據相似的團體？假如甲種智力測驗的標準化根據比較聰明的兒童，乙種智力測驗根據比較愚笨的兒童，則從兩種測驗所得的智力年齡分數或智力商數或T分數B分數就不能比較。第二、兩種測驗標準化所根據的團體其能力分配是否相似？

甲



乙



們的偏斜程度和峻峭程度是否相似兩個分配的偏斜程度不同，其平均數就不能比較，兩個分配的峻峭程度不同，其標準差（或其他差異度量數）就不能比較。上圖甲表示兩個分配的偏斜度不同，乙表示兩個分配的峻峭程度不同。

年齡及年級常模既是根據平均數的，所以兩種測驗標準化時所根據的團體祇要（一）取樣相似，（二）偏斜程度相似，就可認為有比較之可能。標準分數、T分數、百分數、商數既是根據分配的差異度的，所以兩種測驗標準化時所根據的團體，除須適合上述兩個條件外，還須峻峭程度相等，才可認為有比較之可能。

註一 Freeman, F. N. *Mental Tests*, pp. 275-281.

註二 Kelley, T. L. *Interpretation of Educational Measurements* pp. 21-25.

註三 Symonds, P. M. *Measurement in Secondary Education* pp. 521-525.

## 第十講 測驗結果的應用

測驗是學校行政上教學上一種重要的工具。但測驗本身並不能解決教育上的問題，並不能提高教學的效能，問題是在學校行政人員及教師怎樣利用測驗的結果。測驗在教育上之應用，大別之，可有兩端，即團體調查與個人診斷。茲分別論之。

### 一 團體調查

一、團體間的比較或團體成績與測驗常模的比較 根據測驗成績以作團體間之比較，必須注意這些團體是否有比較之可能。在解釋測驗結果時，對於下列事項必須顧及。

(一)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。教育測驗普通採用年級常模，所以各校之間的比較或一校成績與測驗常模的比較，普通亦以年級為單位。但各校同一年級學生的年齡往往不甚一致，有時差異很大。僅憑各年級兒童的成績來比較，而不顧到他們的年齡，極容易發生錯誤的解釋。下列假設的事實，可以明白表示這種錯誤的性質：

甲校初中二的平均分數

四〇·一

乙校初中二的平均分數

四二·八

初中二的測驗常模

四〇・〇

初中二測驗常模所根據兒童的平均年齡

……………十五歲六個月

初中三的測驗常模

四五・〇

初中三測驗常模所根據兒童的平均年齡

……………十六歲六個月

甲校初中二學生的平均年齡

……………十五歲七個月

乙校初中二學生的平均年齡

……………十六歲五個月

以甲乙兩校初中二的平均成績互相比較，乙校成績顯然較甲校為優。再以甲乙兩校初中二的成績與常模比較，則乙校成績超過常模，甲校成績僅與常模相等。但我們如考查甲乙兩校初中二的平均年齡，以及常模所根據學生的平均年齡，結論便不同了。甲校初中二的平均年齡與初中二常模所根據學生的平均年齡約相同，而甲校初中二的成績與初中二常模又恰相等。這樣看來，甲校成績可算正常。乙校初中二的平均年齡差不多等於初中三常模所根據學生的平均年齡，但前者平均成績為四二・八，後者為四五。這樣看來，乙校成績實在正常以下了。

(二)學生的智力 各校學生智力不等。據許多調查，兒童的智力與家庭職業，學校所在區域很有關係。比較同年齡或同年級兒童的學業而不顧及其智力，顯然不是公允的辦法。

(三)學生以往的學歷 學生現在的程度，當然受以往教育的影響。學生現在所在學校，當然不應替



學生以往所在學校領功或受過。譬如我們比較兩個城市中學學生的成績，照測驗結果，城市甲成績較優。但城市甲的小學比城市乙的小學辦得好，我們似不應斷定城市甲的中學效能比城市乙的高。

(四) 所用測驗是否公允 各校課程或教學的測重點不盡相同，用少數教育測驗去比較兩個學校的成績，往往不能得到公允的結論，除非所用的測驗，範圍廣博而足以概括教育結果的全部。

二、一個團體各種學科成績的比較 中小學教育應該着重各科平均的發展，不應過分偏倚。一班或一校各科成績的比較，可以顯示對於各學科的注重，是否得其平衡，是一種很有價值的比較。

三、考查學生的進步 譬如要考查一班學生在一學期中的進步，可在學期之初與學期之終舉行測驗，兩次測驗成績的差別，就代表學生的進步。兩次所用測驗應該性質相同而難度相等，普通可以用一種測驗的兩個複份。

四、考查教學的效率 如其他條件相等，則根據學生的教育成績，以比較兩個學校或兩位教師的效率，是很正當的辦法。但是其他條件往往是不相等的。要公允地比較兩個學校或兩位教師的教學效率，除掉兩校或兩班學生在測驗時的成績外，還須顧到下列各項事實：

(一) 這兩班或兩校學生在學年或學期開始時，學業程度是否相等？

(二) 這兩班或兩校學生智力程度是否相等？

(三) 這兩班或兩校學生年齡是否相等？

(四)所用測驗對於這兩班或兩校是否同樣地適宜？

麥柯爾曾建議以學期(或學年)初與學期(或學年)終學生成績商數或F分數的差別，來表示教學的效率。這種方法對於以上(一)(二)(三)三點均能顧到。祇要所用測驗選擇適當，對於第四點亦可適合。

五、比較教學設施的優劣 決定一種教學設施的利弊，最後的標準，是這種設施在學生方面引起的變化。譬如主張能力分組的說能力分組可以促進學習的效率。這種說法是否成立，還須用實驗來決定。他如兩種教材或教法的優劣，亦須借重實驗方法，才能作正確之判斷。在實驗過程中標準測驗乃不可缺少的工具，其功用有二：(一)為等組之用。(二)為度量試驗因子所產生的效果之用。茲舉例以說明之：

用實驗法中之等組法以比較兩種算術教學法之優劣，其手續大致如下：

- (一)用一種算術測驗，測量一班兒童，以考查其算術程度。如能加用一種智力測驗則尤好。
- (二)根據測驗結果，用配偶法將這班兒童分為甲乙兩組，這兩組的算術成績及智力程度約相等。
- (三)用第一種算術教學法於甲組，用第二種算術教學法於乙組。兩組教師最好是同一個人。
- (四)經過相當時期，再用一種算術測驗度量兩組的程度。計算其差別。
- (五)應用統計方法以決定所得差別是否可靠，以斷定兩種方法之優劣。

## 二 個人診斷

一、以爲編級之根據 在班級教學制度之下，學生之適當編級，一方面可減少教學上的困難，一方面可促進學生學習之效果。所謂適當之編級，第一要使一班學生程度整齊；第二要使一班學生進步的速率大致相等。茲就安插新生、決定升降、能力分組、特別班的組織，四點分別討論之。

(一)安插新生 學校對於新生的安插，大約不外兩種根據。第一是入學考試，第二是以往學業成績。學校對於學生成績或以分數計算或以等第表示，因各校標準不一，各教師寬嚴不等，這種成績分數並無一致的意義，因此不能作爲安插新生可靠的根據。入學考試普通由教師臨時出題，選材和方式兩方面都不無問題。最大的缺點是考試結果並沒有常模可助解釋。標準測驗其效度和信度均經確定，且有各種常模可資比較，作爲安插新生之根據，應較學業成績和入學考試更爲妥善。

(二)決定升降 標準測驗的結果可以幫助學生升降的決定，較諸僅憑教師意見自然可靠得多。但其他事實如年齡、健康等亦應兼顧。智力測驗或教育測驗結果都可作爲決定升降的一種參考。就小學而論，低年級似應着重智力測驗，高年級則教育測驗較爲重要，如兩種測驗並用而各予以相當的重量則更爲妥善。

決定學生升降如要顧到教師的意見則須使教師意見與測驗結果有比較及綜合之可能。麥柯爾曾有一極簡便的方法，介紹如下（參看表十三）。

1, 用標準測驗（智力測驗或教育測驗）測量全班學生，求各生之年級分數（智力年級分數  $G_i$ ，

或教育年級分數  $G_e$ 。)

2. 將全班學生照年級分數 ( $G_i$  或  $G_e$ ) 依次排列。

3. 請教師將全班學生照其應予升級之程度，依次排列。

4. 教師列為第一學生給以一班中最高之年級分數，教師列為第二學生給以一班中次高的年級分數，餘類推。這些年級分數，稱為  $G_t$ 。

5. 將  $G_i$  或  $G_e$  ( $G_i$  與  $G_e$  並用亦可) 與  $G_t$  平均，得調級  $G$  分數 (Promotion  $G$  score)，簡稱  $G_p$ 。如以為教師意見不應與測驗結果相等並重，則在求調級  $G$  分數時可用適宜之均衡方法。

關於決定學生的升降，麥柯爾有兩條建議，足資參考，節錄如下：

1. 任何學生不應退級或留級，除非他的教育程度在退入或留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之百分之二十五點以上。如更慎重些，可規定百分之五十點（即中數）以上。

(三) 能力分組 以上討論兩點都是關於編級的問題。所謂編級是照兒童的教育程度或智力程度，編入相當的年級，使一級之內兒童的程度大致相等。現在討論的能力分組，乃是學生已經依其教育程度

第十三 表示計算  $G_t$ ,  $G_p$  的方法

學生	$G_i$	教師意見	$G_t$	$G_p = \frac{G_i + G_t}{2}$
甲	6.0	2	5.6	5.8
乙	5.6	3	5.4	5.5
丙	5.4	1	6.0	5.7
丁	5.0	4	5.0	5.0

編入相當年級後，再依其聰明程度或進步速率編入相當組別。能力分組的根據應該是一種商數，如智力商數或教育商數或B分數。能力分組祇能在人數較多的學校舉行。

能力分組的利弊，是一個議論紛歧的問題。不過有一點是贊成和反對兩方面所公認的，就是能力分組不過是改進教學的一種條件，而非足夠的條件。換句話說，分組以後，還須改變教法，改組教材，以適合各組兒童之能力，才能收實際的效果。

編級和分組都是減少班級教學制度困難的幾種辦法。許多教育家主張根本推翻班級教學制度而代以個別教學方法。在個別教學之下，每個兒童得依照其能力進展而不受其他兒童之牽掣。道爾頓制 (Dalton Plan) 和 文奈狄卡制 (Winnetka Plan) 都是最著名的實例。

(四) 特別班的組織 依據兒童能力所組織的特別班，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辦法。所不同者，特別班中的兒童在能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，尤其是低能班兒童的選定。特別班的組織當然可以根據其他需要，如視力、聽力上或語言有缺陷的兒童。

二、以為指導之根據 學校中之指導工作，以促進學生身心各方面健全發展為目的，應該是一種整個的設施。但為討論便利計，我們不妨將學校中指導工作，分為三個部分，或三個方面，即性格指導、教育指導和職業指導。

(一) 性格指導 性格指導以培養健全性格，保持心理健康，和促進社會適應爲目的。有效的性格指導，必須根據兒童環境的調查和個性的了解。要了解兒童個性標準測驗、心理問卷、和評判量表都是有效的工具。

(二) 教育指導 教育指導，簡單的說，是對於學生學業問題的指導。學校中教育指導的工作，可分爲入學指導、修學指導、升學指導三方面。一個人的求學過程，大概有幾個重要的關鍵。小學是國民基本教育，人人應該享受的，可不成問題。小學畢業初中畢業或高中畢業後，各有升學與就業兩條途徑，而升學又有學校的選擇，科系的選擇等問題。這些問題的適當解決，與學生前途有莫大關係，不應草率從事，而應根據各方面的觀察，經過縝密的考慮，再作最後的決定。同時，這些問題往往不是學生個人所能妥善解決的，學校與家庭，都應負指導的責任。這種指導需要最迫急的時候，固然在學業告一段落的時候，但平日對於這些問題，即應注意，不可等到最後關頭，才着手籌劃。各種心理測驗的方法和工具，可以幫助我們了解學生的能力、特長、興趣和性格。對於學生的個性有了正確的認識，則施行指導便比較有根據，有把握。

(三) 職業指導 據美國全國職業指導會，「職業指導，乃是供給事實、經驗、及意見，去幫助個人選擇職業、預備職業、獲得職業、及改進職業。」何清儒說「職業指導即是輔助個人，解決各種職業問題，求人與事的適合，以發展個性，促進社會的工作。對於選擇、訓練、謀業、以及升進改業等之調劑，均應有相當的協助。所居的立場，是以個人爲中心，所取的方法，是個別的處置。」(註一) 簡單地說，職業指導，乃是應用科學的

方法，根據可靠的事實，幫助個人在職業上達到一個圓滿的適應。

職業指導主要的依據，不外對人的研究或個性的考查，和對事的分析或職業的調查兩項。一人之個性包括體格、智力、特殊能力、興趣、性格各方面，而這些方面與職業均有相當之關係。標準測驗、心理問卷、評量表，都是考查個性有效的工具。

三、以爲教學之根據。教學是達到教育目標的過程，測驗是考查教育目標已否達到的一種工具。所以教學與測驗實有不可分離的關係。測驗在教學上之應用，可分幾點來說明。

(一)了解個性。「因材施教」「注重個別」都是現代教育上極重要的概念，因此發現學生的才能，考查學生的個性，乃現代教師極重要的任務了。

(二)考查成績。在開始教學時，學生程度如何？爲學習一種新材料所必需的基本智識技能已否具備？對於某段教材已否了解，已否嫻熟？對於規定目標已否達到，相距多少？學生的成績與其能力是否相稱？這些問題的解答，有賴於測驗者正多。

(三)診斷困難。學習上的困難，適應上的困難，必須詳加診斷，教學的設施，才不致如暗中摸索，漫無依據。測驗是診斷困難最經濟最有效的一種工具。

(四)引起動機。使學習者知道自己的成績，爲引起學習動機有效的方法，已有實驗證明。所以測驗的結果應該使兒童知道，最好作成圖表，以示進步的狀況，而資客觀的比較。

(五)改進考試 舊式考試的弊端我們在第二講已約略提及。因此有人注意到如何利用測驗的原則，以改進考試的方式和記分的方法。所謂新式考試或客觀考試，學校已多有採用者。新式考試的利益，簡言之，不外取樣適當、信度較高、定分客觀、校閱便利。關於新式考試的方式與編製已有專書討論，我們祇得從略了。

學校中最通行的記分制度大概還是百分制。百分制的最大缺點，就是分數的分配，缺乏共同的標準，因此各教師的分數無從比較。分析言之，百分制記分往往對於一班學生中心分數所在和分數的參差程度，缺乏共同的標準。因為中心所在缺乏標準，甲教師的分數全班平均六十五分，乙教師的分數全班平均八十五分。因為參差程度缺乏標準，丙教師的分數全班限於一二十分之內，丁教師的分數全班優劣相差六七十分之多。這兩種缺點都有校正之必要。茲將沈有乾（註二）所提出之方法介紹於下：

假設一班學生二十人，考算術兩次，每次五十題，每人做對題數如下：

第一次	50	48	47	46	46	45	45	44	44	44	43	43	43	43	42	42	42	41	41
第二次	40	40	39	37	35	35	33	31	30	30	29	28	27	25	23	20	17	12	2

若照百分法記分，每題得二分，則結果如下：

第一次	100	96	94	92	92	90	90	88	88	88	86	86	86	86	84	84	84	82	82
第二次	80	80	78	74	70	70	66	62	60	60	58	56	54	50	46	40	34	24	4





照記分習慣九十分以上爲甲等，八十分至九十分乙等，七十分至八十分丙等，六十分至七十分丁等，六十分以下戊等，則兩次考試等級的分配如下：

第一次考試

等級	人數	百分比
甲	2	10%
乙	3	15%
丙	10	50%
丁	5	25%
戊	0	0%

第二次考試

等級	人數	百分比
甲	0	0%
乙	6	30%
丙	9	45%
丁	3	15%
戊	2	10%

上面所規定中數以七十五爲標準，中部分分之八限減以二十五爲標準，不是絕對的。有時全班程度特別好或特別壞，有時全班程度特別齊整或特別參差，在校正時都應顧到。所以校正方法應當加以變通。沈有乾規定辦法如下：

(1) 百分制分數的全班中數普通可定爲七十五分。若全班程度特別好，中數可酌量提高，但以加到八十分爲限。若全班程度特別壞，中數可酌量減低，但以減到七十分爲限。

(2) 百分制分數的全班中部十分之八的限域普通可定爲二十五分。若全班程度特別齊整，這限域可酌量縮減，但以減到二十分爲限。若全班程度特別參差，這限域可酌量擴充，但以加到三十分爲限。

另一種通用的記分制度是採用常態分配的面積比例，不給分數，只分等級（普通五等），各等級限定人數，不得由教師任意支配。採用這種方法，我們應該注意一班學生程度的分配，不一定盡適合常態。常態分配的面積比例祇給我們一種大約的指南。賽門司所提出的方法頗值得採用。（註三）茲根據前例兩次算術考試分數，說明其方法如下：

第一步，計算兩次分數的平均數和標準差。

第一次  $M = 44.05$        $\sigma = 2.29$

第二次  $M = 28.15$        $\sigma = 9.45$

第二步，照下列標準差限域規定分數等級。

甲  $+1.5\sigma$  以上

乙 從  $+0.5\sigma$  到  $+1.5\sigma$

丙 從  $-0.5\sigma$  到  $+0.5\sigma$

丁 從  $-1.5\sigma$  到  $-0.5\sigma$

戊  $-1.5\sigma$  以下

第三步，將上列標準差限域化成分數。例如第一次考試平均分數為 44.05，標準差為 2.29，則  $+1.5\sigma = 44.05 + (1.5 \times 2.29) = 47.49$  爲簡便計，以整數計算，得結果如下：

第一次考試

甲 48 或以上

乙 46—47

丙 43—45

丁 41—42

戊 40 或以下

第二次考試

甲 43 或以上

乙 33—43

丙 24—32

丁 14—23

戊 13 或以下

第四步，照第三步所列標準定各人之等級。兩次分數的分配如下：

第一次考試

等級	人數	百分比
甲	2	10%
乙	3	15%
丙	10	50%
丁	5	25%
戊	0	0%

第二次考試

等級	人數	百分比
甲	0	0%
乙	7	35%
丙	8	40%
丁	3	15%
戊	2	10%

上列等級分配與照沈有乾規定辦法所得結果，幾乎完全一致了。

註一 何清儒職業指導學（商務）一一頁。

註二 沈有乾教育心理（正中）二八五——二九〇頁。

註三 Symonds, P. M. *Measurement in Secondary Education* pp. 513-516.

## 第十一講 學校設備與設施的測量

在第一講裏我們說過評量一個學校的優劣，有兩種着手的方法。第一是學生成績的測量，第二是學校設備及設施的測量。前面幾講所討論的都是關於學生成績的測量，現在介紹測量學校設備及設施的方法。

學校設備及設施的測量，有評判量表 (rating scale) 評量單 (Check list) 及記分片 (Score card) 幾種方法。這幾種方法可說是大同小異。方法是將評量的對象分析爲若干特點或方面，然後對於每一特點或方面規定評量的標準，評量的結果或以等級表示或以分數表示，最後綜合各特點或各方面的評量，求得一個總評量結果。

### 一 學校設備的測量

就學校設備說，對於校舍、桌椅、記錄報告、圖書館、實驗室、教便物、教科書各方面，都可以採用記分片以爲評量之工具。校舍的記分片經美國哥倫比亞大學師範學院教授施菊野 (G. D. Strayer) 和安革霍 (N. L. Engelhart) 二氏之倡導，發展最早，亦最完善。教科書記分片，因實際上之需要，美國各學校採用者亦頗多。茲就校舍及教科書記分片各舉一例以示一般。

一校舍記分片 施菊野安革霍二氏所編小學校舍記分片經邵爽秋略加修改，其大綱如下：（各項目旁分數係標準分數）

I. 地址	125
1. 坐落	55
2. 排水	30
3. 大小形狀	40
II. 房屋	165
1. 地位	25
2. 外部構造	60
3. 內部構造	80
III. 衛生設備	280
1. 取溫換氣	80
2. 防火	65
3. 去污	20
4. 燈光	20
5. 電用設備	15
6. 水料	30
7. 廁所	50
IV. 教室	290
1. 位置	35
2. 構造粉飾	95
3. 光線	85
4. 藏衣室	25
5. 設備	50
V. 特別室	140
1. 普通用的	65
2. 辦公室	35
3. 其他特別室	40

上列係一個大綱，1. 2. 3. 4. 各項目下尙有小項目，此外各項目記分標準均有詳細規定，可參閱邵爽秋教育調查應用表格。

這種記分片的特點，就是將校舍分析爲若干方面，評量時祇須對這些方面個別注意，而且有標準可資遵循，各方面之最高分數亦經規定，如此，主觀的成分自可減少。爲增加評量結果之準確起見，至少須有三人評量，然後求三人評量結果之平均。依評量總分，我們可以將校舍分爲幾個等級，如分數在七百五十以上者爲優等，五百至七百五十分者爲中等，二百五十至五百分者爲下等，二百五十分以下者爲劣等。安革霍曾有下列之建議：九百分以上者可視爲建築和設備優良之校舍；七百至九百分者可認爲相當滿意，

但應注意分數較低的方面以謀改進；六百至七百分者需要相當之改進，五百至六百分者必需大量改造和修理才可合用；五百分以下者不能認為合用而且不值得改造和修理。

二、教科書記分片 教科書記分片可以用作選擇教科書之依據，比較普通的評判客觀而周詳。但反對者以為採用記分片以評量教科書不免機械化，而不易顧到各學校之特殊需要。大概評量教科書之標準依適用年級，學科性質，及教學側重點而異。所以各年級及各科教科書應有特殊的記分片，而記分片上各項目應占之分量（或分數）不宜呆板規定，宜由各學校依其需要，斟酌情形，自行決定。如此則伸縮性較大或可避免機械化之譏。

批評教科書的標準主要的不外形式和內容兩方面。吳研因（註一）曾介紹英國彭勃格（Bamberger）的「低級兒童用書形式評點表」抄錄其大綱如下：

I. 大小	10分
II. 封面	20分
1. 封面顏色	10分
2. 飽和	2分
3. 鮮明	2分
4. 封面畫	6分
III. 書名	5分
IV. 分量	20分
1. 頁數	5分
2. 課數	5分
3. 兒童見聞材料佔全書份量百分比	10分
V. 插圖	38分
1. 插圖占全書面積百分比	7分
2. 插圖大小	3分
3. 插圖顏色	20分
4. 插圖的精細	5分
5. 特性	3分
VI. 頁面	7分
1. 頁面佔面積百分比	2分
2. 頁邊面積	1分
3. 每行長短	2分
4. 每頁平均行數	2分
總計	100分



表中對於各項目的評量，定有詳細標準，讀者可閱吳氏原文。包括形式和內容雙方的記分片可以美國渥  
鐵斯(B. M. Otis)的內容學科教科書記分片爲例，介紹於下：

I. 對於本地情形的適合性		計一百十分
1. 觀察	二十分	
2. 教學	二十五分	
3. 兒童	三十分	
4. 班次	十分	
5. 設備	二十分	
6. 學期	五分	
II. 教材		計四百分
1. 兒童經驗	六十五分	
2. 教學目的	五十分	
3. 個性差異	四十五分	
4. 選擇均衡	一百分	
5. 道德與公民價值	六十五分	
6. 真實性	三十五分	
7. 文體	四十分	

III. 佈置和組織		計一百二十分
1. 分幾部份	四十分	
2. 設計教學法	八十分	
IV. 對於教學的幫助		計一百七十分
1. 應用	九十分	
2. 選擇的餘地	五十分	
3. 索引	十分	
4. 註解	十分	
5. 目錄	十分	
V. 形式		計一百五十分
1. 美觀	二十五分	
2. 插圖	四十五分	
3. 印刷	三十分	
4. 裝訂	二十五分	
5. 紙張	二十五分	

VI. 其他

1. 著者	計五十分
2. 出版者	三十分
	五十分

3. 序言	五十分
4. 出版期	十分
總計	一千分

吳研因根據勃柯二氏的意思參照個人的見解擬定一個「批評我國小學教科書的分數表」抄錄如下，以供參考：

甲、形式方面

一、書本大小厚薄	二〇分
二、封面、封面蓋	二〇分
三、插圖大小顏色	九〇分
四、頁面天地頭行間字體	二〇分
五、印刷裝訂	二〇分
六、紙張	一〇分
七、其他特點	二〇分
乙、內容方面	計八〇〇分
一、適合本國情形時代需要	一二〇分

二、適合教育目標與有道德的公民價值	一〇〇分
三、適合兒童經驗與個性	一〇〇分
四、選擇均衡適當	七〇分
五、正確或有充分的真實性	三〇分
六、文字通暢條達優美	一四〇分
七、組織和佈置適當	八〇分
八、分量充實足用	四〇分
九、便於教學	一二〇分
總計	一〇〇〇分

吳氏這個記分表只是一個項目大綱，而各項目的記分標準尚待詳細規定。

## 二 學校設施的測量

就學校設施說行政組織、訓導機構、財務處置、教學效能、課程綱要各方面都可利用記分片或類似方法從事評量。

一、教學效率評量表 教學效率評量表已經編就的種類很多。就功用論有自衡表及評量表兩種，前者是教師用來自己評量自己的，後者是視導員用來評量教師的。就方式論，有問題式、等第式、及記分式三種。茲舉數例如下：

1. 羅格(H. O. Rugg)現任教師考核表(註二) 羅氏教師考核表將教師的行爲分爲(1)教學的技能，(2)教室管理的技能，(3)團體的活動，(4)進修的工作，(5)個人及社會的特性五組，每組下列有許多問題，對於每問題的評量以上中下三等級表示。這張表教師用以自衡或行政人員用以評量教師均適用。

2. 鮑愛斯(R. C. Boyace)教師效率記分表(註三) 本表內容分(1)人格，(2)社交能力與專門預備，(3)學校管理，(4)教學技能，(5)教學結果五大項目，各大項目下再分析爲若干小項目，共四十四項。評量結果分優上中下劣五等表示之。

3. 邵爽秋教學效率測量表 邵編教學效率測量表內容分(1)教室管理，(2)儀容言動，(3)教材

表十四 部爽秋教學效率測量表(一)

一、教室管理 (155)			
上課 退課	(9) 遵守時間	(6) 遲到或早退三分鐘以上	(3) 不依時間
出入 教室	(12) 安靜活潑秩序井然	(8) 略有喧嘩秩序略有紊亂	(4) 喧嘩毫無秩序
空氣	(30) 能充分利用窗戶調節空氣	(20) 1. 不能充分利用窗戶調節空氣 2. 無窗戶設備	(10) 有窗戶不知利用
坐次	(22) 坐位先後依學生身長視力聽力之差異而定	(11) 坐位排列毫無標準後坐兒童聽講或抄寫時顯有視聽上之困難	
教便 物	(16) 不受經濟限制之教便物均已搜集或攜帶	(8) 忘帶某項教便物	
物品 收發	(15) 用極經濟之傳遞方法收發用品	(10) 1. 有人收發但無次序 2. 未有物品收發	(5) 個別取送毫無秩序
課內 秩序	(33) 無喧嘩擾害他人等事發生	(22) 此等事發現一次	(11) 此等事發現一次以上
學生 姿勢	(18) 學生坐定時身體上無惹人注意之態度	(12) 五人以下發現不良姿勢	(6) 五人以上發現不良姿勢如叉手伏案斜倚後仰等
二、儀容言動 (192)			
態度	(15) 和藹可親	(10) 和平	(5) 暴戾討厭
	(21) 老成練達	(14) 態度大方	(7) 羞怯畏縮
舉動	(15) 適當活潑	(10) 有無謂之舉動	(5) 有極不適當之舉動
體格	(15) 精壯堅實	(10) 適中	(5) 纖弱羸瘦
	(15) 姿勢正直	(10) 略有不正	(5) 躬腰聳肩或伸頸
衣履	(9) 衣履修潔樸素	(6) 中平	(3) 衣履眩耀或污垢
語音	(36) 清晰悅耳高低適度	(24) 略有含糊	(12) 尖粗刺耳
語法	(21) 極有倫次	(14) 略有倫次	(7) 顛倒錯亂
語調	(21) 要點加重語勢並有語調	(14) 略有變化	(7) 單調全無變化
語義	(21) 易於領悉	(16) 略用深語	(8) 用語太深



及教法，(4)學生反應四大組，每組之下各有項目若干。評量方法以分數表示，各項最高分數均有規定，最高總分數為一百分。茲將該表錄前以資參考。

關於本表之編製，邵氏曾有說明，其主要步驟如下：

(1)搜羅條目 編製測量表時，第一步為搜羅條目，如「上課退課」，「出入教室」之類。此類條目當該表初製時，共有一百餘條，概從通常所用教學測量表、教學法著作，及作者經驗中搜羅而來者。

(2)精選條目 此百餘條目，經廣州中山大學教育研究所同人逐條討論，最後用公決法，選出得票最多者三十四條為構造該表之基本條目。

(3)分化條目 次將各條目就其性質分為三等或兩等。如「語言」則分為「清晰悅耳高低適度」，「略有含糊」，「尖粗刺耳」三項。「發問」則分為「發問能引起學生思想或興趣」及「呆板發問」二項是。

(4)決定總分 因此表條目繁多，每條又有差等，一百分斷不敷分配，故決定以一百分為該表上之總分數。

(5)分配分數 分配分數之目的，在該表上，一方面欲將各條目之最高分數湊成百分，他方面又欲將每條目各等間之分數排成倍數的順序（如 15, 10, 5, 或 27, 18, 9）各條目應得之最高分數係根據專家意見視各條目之輕重決定。

二、課程評量表 美國哥倫比亞大學師範學院教授勃倫納(H. B. Brunner)和史厥突麥(E. B. Stratemeyer)二人根據豐富的經驗，釐訂評量課程綱要的標準。他們的標準逐年修改，下面所介紹的標準是一九三七年所發表的。

標準分哲學、內容、活動、學生工作之考核四部分。每部分若干項目，每項目再分若干條目，總共有一〇七條之多。這些條目就是評量的標準。評量分最優、甚好、好、勉強、壞、五等及「未列入」一項。評量者先就各條目一一評量，根據各條目之評量對於各項目定一總評量，再根據各部分下各項目之評量，對於各部分定一總評量，最後對於整個課程綱要定一總評量。茲將該評量表大綱詳述如下：（除「社會哲學」項下條日照譯外，其他項目下條目從略）

### 1. 哲學

1. 社會哲學 社會哲學應促進民主社會的最終目標，應認識社會的變動性，並要求學校為促進社會進步的機關。

- (1) 是否以民主社會為理想？
- (2) 是否認識社會上各種組織，因新環境的要求與吾人新的了解，應不斷的改變？
- (3) 是否以生活為適應動的世界的過程？
- (4) 是否以社會生活為個人發展所必需？

這些勢力和問題有無應付的準備？

(6) 是否承認學校為改進社會的機關？

2. 教育哲學 教育哲學應根據社會哲學，並應為決定課程內容的主要依據。教育之主要目的在幫助個人獲得圓滿的發展，以增加其改進社會之能力。

3. 學習原則 課程應根據最正確的心理學原則。

### II. 內容

1. 正確 教材須真實正確。
2. 實用 教材須能幫助生活上問題的解決。
3. 適宜 教材須能促進個人的發展和社會的利益，並適合兒童的成熟階段。

4. 組織 教材的組織應以兒童經驗為中心，以幫助兒童興趣的發現和發展，適應其需要，並獲得豐富的經驗。

### III. 活動

1. 自動 活動須能激發兒童自動的願望。
2. 興趣和需要 活動須能滿足兒童的需要，依據兒童的興趣，以期獲得最好的發展，所以活動須與兒童當時的經驗密切連繫。

3. 社會價值 活動須能供給經驗以助兒童成為社會上有用的份子。

4. 真實性 活動須從實際生活情境中選擇，而為兒童所認為有興趣而重要者。
5. 變化性 活動須多變化。

6. 方式 須使團體中每一份子有完成工作單元的願望。
7. 最後活動 最後活動須使團體和個人都能滿足他們預期的目的。

### IV. 學生工作的考核

1. 目的 考核的目的在(1)了解個別兒童，(2)以為改進教學之依據，(3)以決定教育目標之已否實現。

2. 多變化 考核應採用各種方法，各種工具。一方面由教師考核兒童的工作，一方面由兒童考核自己的工作。

3. 效度 考核的效度被決於下列因素：(1)考核的方式接近自然的情境；(2)個人認識考核的需要和目的，並參加考核的工作；(3)行為各方面的考核應互相參照，以期獲得對於整個兒童的認識。

4. 生長的方面 考核須包括體格、情緒、社會、以及智能各方面的發展。

5. 解釋 課程綱要對於各種考核結果的解釋應有具體的說明。

麥柯爾在其近著測量學 (Measurement, Macmillan Co, 1939) 一書內介紹一種學校設施問



卷，爲考核學校課程（廣義的，包括一切設施在內）一種客觀的工具。這種問卷是要兒童回答的，從小學四年級至初中三均適用。問卷內共有一〇五個問題，回答方式是就「是」「否」「兩答中選擇一個」。據麥氏說，這種問卷可以鑑別合乎民主主義和背乎民主主義的教育設施。牠不僅是一種測量工具，如運用得當，且有促進民主主義教育的功能。本問卷的記分方法麥氏有詳細的規定。關於本問卷內容，可閱麥氏原書。

陳鶴琴在小學教師月刊第三卷第一期發表一文，題曰活教育與死教育，列舉活教育（或新教育）與死教育（或舊教育）的十大區別，並就課程、教學、教師、兒童、行政、設備六方面，備有兩種教育的詳細對照。這種對照表祇須稍加整理，並對各部分的輕重加以斟酌，不難成爲一種極有用的教育設施評量表。茲節錄陳氏原文如下：

## 十大區別

### 活教育

- 一、一切設施一切活動以兒童做中心，當主體。學校裏一切活動差不多都是兒童的活動。
- 二、教育的目的在培養做人的態度，養成優良的習慣，發現內在的興趣，獲得求知的方法，訓練人生的基本技能。
- 三、一切教學，集中在「做」，做中學，做中教，做中求進步。

第十一講 學校設備與設施的測量

### 死教育

- 一、一切設施一切活動，教師（包括校長）是中心是主體，學校裏一切活動差不多都是教師的活動。
- 二、教育的目的，在灌輸許多無意義的零星知識，養成許多無關重要的零星技能。
- 三、一切教學，集中在「聽」，教師口裏講，兒童用耳聽。

四、分組學習，共同研討。

五、以愛以德來感化兒童。

六、兒童自訂法則來管理自己。

七、課程是根據兒童的心理和社會的需要來編訂的，教材也是根據兒童的心理和社會的變遷來選定的，所以課程是有伸縮性，教材是有活動性，而可隨時更改的。

八、兒童天真爛漫，活潑可愛，工作時很靜很忙，遊戲時很起勁很高興。

九、師生共同生活，教學相長。

十、學校是社會的中心，師生集中力量，改造環境，服務社會。

### 活教育與死教育的詳細對照

#### 活教育

##### 一、課程

1. 以大自然大社會作主要的教材，以課本作參考資料，這是直接的活知識，是直接的經驗。

2. 各科混合或互相關聯。

個人學習，班級教授。

以威以畏來約束兒童。

教師以個人主見來管理兒童。

固定的課程，呆板的教材，不問兒童能否瞭解，不管與時令是否適合，祇是一節一節的，上一課一課的教。

兒童呆板，暮氣沉沉，不好動，不好問，儼然像個小老人。

師生界限分明，隔膜橫生。

校牆高築，學校與社會毫無關係。

#### 死教育

以課本作主要教材，是間接的死知識，是間接的經驗。

各科獨立而不相聯絡。

3. 不受時間的限制，沒有分節的時間表，時間倒爲功課所支配。

4. 內容豐富。

5. 生氣勃勃。

6. 兒童自己做的。

7. 整個的，有目標的。

8. 有意義的。

9. 兒童瞭解的。

### 二、教學

1. 多在戶外。

2. 領導學生自勵研討。

3. 啓發式，誘導式。

4. 自動的。

5. 數兒童。

### 三、教師

1. 笑嘻嘻地和藹可親。

2. 聲音悅耳。

功課受時間的限制，一節授一課，不管科目的長短，時間一到，即須停課。

內容簡單。

乾燥無味。

現成的，由教師代做的。

片段的，沒有系統的。

無意義的。

兒童不瞭解的。

整天在室內。

只會照着課本某講，學生懂不懂不問。

注入式，填鴨式。

被動的。

教書。

板著臉，威嚴可怕。

聲音粗糙。

3. 說話有禮，多鼓勵。
4. 低音清晰。
5. 行動輕快。
6. 立得筆正，坐得挺直。
7. 衣服整潔，面目清楚。
8. 態度從容。
9. 精神飽滿。
10. 創造能力。
11. 健身。
12. 快樂樂觀。
13. 研究精神。
14. 樂業。
15. 互助合作。
16. 慈愛。
17. 負責。
18. 教學有技能。
19. 了解兒童心理。

隨便漫罵。

大聲喊叫。

走路拖地。

立得傾斜，坐得駝背。

衣冠不整，頭髮蓬鬆。

脾氣暴發。

沒精打彩。

模仿。

多病。

憂愁悲觀。

苟且因循。

思遷。

孤獨利己。

冷酷。

敷衍。

教學呆板。

不明瞭兒童心理。

#### 四、兒童

1. 活潑天真獨出心裁。
2. 自己找材料的。
3. 興趣濃厚的。
4. 身體健康的。
5. 好問好奇的。
6. 知道求知的方法而活用知識的。

#### 五、行政

1. 學生自己管理的。
2. 考核成績在活動。
3. 教學目的，在培養做人優良習慣，服務合作的精神。
4. 尊重兒童的人格，訓導的方式是友誼的。
5. 師生共同生活的。
6. 教師直接參加各種活動的。
7. 有組織有力量的。
8. 學校與學生，站在同一戰線上，向學業上進攻。

#### 六、設備

#### 第十一講 學校設備與設施的測量

呆板，不活動的，死讀書的。  
模仿的。

做事讀書毫無興趣的。

身體軟弱的。

唯唯諾諾的。

學了許多死的書本知識，而不會應用的。

學生由教師管理的。

考核成績在紙片。

教學目的在灌輸知識，養成技能。

訓導的方式，是防賊式的。

師生界線分得很嚴的。

對於各種活動，教師站在指揮的地位。

學生是一盤散沙毫無組織的。

學校怕學生有組織而與學校對抗。

1. 課桌椅分開，可以移動。
2. 座位討論式的一組一組排列著的。
3. 圖書教具很豐富的。
4. 學校的環境，整齊美麗。
5. 校內的佈置是學生做的。
6. 佈置的材料，利用自然物和兒童的成績。

課桌椅是兩張相連，死釘在地板上的。

座位聽講式的一排一排向着教師排列的。

圖書教具很簡陋的。

學校的環境雜亂骯髒。

校內的佈置，是教師做的。

佈置的材料是化錢買來的。

註一 吳研因批評小學教科書的標準教育與學第一卷第十期。

註二 關於該表詳細內容可閱雷震清教育視導之理論與實際二〇四頁。

註三 關於該表詳細內容可閱杜佐周教育與學校行政原理（商務）一九四頁。

## 第十二講 測驗運動的回顧與前瞻

教育上的測驗運動到現在大約有四十年的歷史。我們回顧已往，遠矚將來，深信測驗運動雖然就發展的方向說，應該有一種重要的轉變，或者已在這轉變之中；就所用的工具說，還有擴展和改進的必要，但測驗並不是教育上一時的風尚，而將成爲教育上一種永久重要的方法和工具。茲就測驗運動的成就和限制作一簡括之討論以爲本書之結束。

### 一 成就

測驗的發展對於促進教育之科學的研究有很大的功績，對於教育設施，尤其是教學方法和教材的改進，有重要的貢獻。這是無人可以否認的。分析言之，測驗運動對於教育之貢獻有五方面。

(一)使教育者對於受教的兒童有進一步的了解。教育的設施欲求其有效，對於施教的對象，須有清楚的了解。一位教育家說，「我們不能教育兒童除非我們能了解兒童。」測驗供給了施教者一種比較正確而可靠的工具，以度量受教者的智慧、才能、學業、興趣、態度和性格。測驗結果顯示了個人之間的差異，引起了教育者對於個別兒童的注意。

(二)使教育者對於教育目標有較確實的認識。編造教育測驗之第一步驟就是確定教學的目標。

含糊的規定，如促進健康培養美德、發展高尚興趣、養成良好習慣，對於編造測驗並無多大幫助。我們必須進一步將這些目標演繹為兒童行為上的變易，使牠更具體、更確定，才能作為編造測驗的依據。所以測驗的編造直接引起了教育目標的分析和釐定。

(三)使教育者對於學習的過程有更清楚的了解。學習不是一蹴就成的，必需經過一定的步驟，在學習的過程中常遇到許多困難。教師對於學習的過程有清楚的了解，教學才不致如以盲引盲，才不致如在暗中摸索。測驗在這方面有很大的貢獻。例如讀法測驗的應用，使我們更清楚地認識朗讀和默讀的區別。閱讀時眼動的測量，使我們對於閱讀的過程有正確的了解。各種診斷測驗，使教師對於兒童在學習過程中所遭遇的困難，能詳細地加以分析。

(四)使教育者有一種客觀的依據以度量教育的設施。整個教育的設施實以兒童為中心。一種教育設施須在兒童行為上或心理上引起預期的結果，才是有效的。除非這些結果可以客觀地加以度量，則教育上一切設施的價值，便無從決定。

(五)養成教育者客觀的態度。測驗運動最普遍的影響就是客觀態度的養成。教育者初以為一班中兒童的學力，大致齊整，但測驗結果顯示了意想不到的差異。教師初以為普通考試足以可靠地甄別學生的學業程度，但標準測驗的結果，暴露了普通考試和教師定分的不可靠。這些事實使教育者深深覺悟到主觀見解之不足持，和尋求客觀證據之必要。



## 二 批評

在前三十年當中（大致自一九〇〇年至一九三〇年）測驗運動發展的迅速和所獲成就的豐偉，不免引起了教育者對於測驗過度的信仰和熱忱。有些人對於測驗的效度和信度有過高的評價；對於測驗的效用，有過奢的期望；對於測驗結果的解釋不免過於機械。有些人不明瞭測驗的真諦，盲目地模倣地加以採用，浪費了教師學生的時間，而對於教學的效果並不能發生任何的影響，這些都是不容諱言的流弊。近十年來經過遠識教育者（包括測驗專家在內）的探討，批評的態度已代替了盲目的信仰；而測驗運動的發展，遂發生重大的轉變。這種轉變乃基於下列幾種覺悟和認識：

（一）測驗的範圍太狹窄，許多重要的教育結果，尚不能用客觀的方法加以測量。

（二）測驗的方法流於機械，工具過於單純。所謂「測驗」差不多限於問答的方式，而測驗的情境往往控制失當，不能引起兒童自然的反應。

（三）測驗結果的解釋，往往忽略了行爲的完整性，有機性和綜合性。因此對於兒童行爲的動機和行爲不能有正確的了解。

（四）測驗與教學不能取得密切的連繫。測驗成爲教育上之裝飾品而不能產生真實效能。

### 三 基本的認識

從上面的討論，我們對於測驗在教育上的地位，應有幾種基本的認識。

(一) 測驗乃教學的一種工具，用以促進教育目標之實現者。一切測量工具的價值都取決於此標準。凡有促進教育目標實現之功效者都是有價值的，不然便是沒有價值的。

(二) 測驗的對象須是兒童在實際生活情境中行爲的自然單位。所謂教育目標就是我們在兒童行爲上所希望引起的種種變易。測驗既是幫助實現教育目標之工具，便應該以兒童行爲上的變易爲對象。例如學校中衛生教學的目標在養成兒童衛生的習慣。要考查衛生教學目標之已否實現，須從兒童日常生活中，觀察其生活習慣是否合乎衛生原則，紙筆一類測驗並無多大功效。

(三) 測驗的內容須與教育目標相符。不然便喪失其效度。

(四) 測驗須能測量一切重要的教育目標，不然一部分教育目標，便有被忽視的危險。因爲智識技能的測量較易於着手，初期的測驗不免偏於這兩方面。近十年來測驗的範圍已經大大擴展，許多從前認爲無法測量的抽象的教育結果，如態度、興趣、思考、欣賞、品格等等，已有方法可加以考查了。

(五) 測驗不宜拘於紙筆測驗一類，而應採用多種的方式。許多重要的教學目標現在尙不能用客觀的方法或工具去考查。在這些客觀的方法或工具尙未完成以前，這些教學目標仍不得不用主觀的方法

去考查。我們一方面要創造新的方法，一方面要改進舊的方法，以適應新的需要。

的。

(六)測驗應以其效度爲主要條件。以往的測驗偏重客觀性和信度而犧牲了效度，實在是得不償失的。

(七)測驗爲度量一切教育設施的效果的工具。一切教育設施的價值均取決於測驗的結果。麥柯爾說「除一個問題外，每個教育問題的最後答案均須取決於教育測驗者。」

(八)測驗與教學應相輔而行。教學離了測驗，教學有如在暗中摸索，測驗離了教學，更成爲毫無意義的工作。

(九)教師應參加測驗工作。一切教育結果的考查應由教師與測驗專家雙方合作。

(十)行爲的性質是完整的、有機的、綜合的，而測驗祇能局部進行。但分析的測量並不一定妨礙整個兒童的了解，問題是在測驗結果的如何解釋。整個的兒童或一樁教育經驗的全部結果是無法整個地加以測量的，不過對於一個兒童的各方面或一樁教育經驗的結果的各方面分別地加以測量後，將這些局部的結果，加以妥適的綜合的解釋，則對於整個兒童或一樁教育經驗的全部結果，仍可有正確的了解。

## 附錄一 教育測驗參考書目

下列書目都是關於智力及教育測驗的通著。專門著作及研究報告統未列入。

1. Freeman, F. N. *Mental Tests*, Revised Edition, Houghton Mifflin, 1939

敘述智力測驗的歷史、方法和應用。

1) Pincher, R. *Intelligence Testing*, Revised Edition, Henry Holt, 1931

第一編敘述智力測驗的理論和歷史，第二編敘述智力測驗的方法和工具，第三編敘述應用智力測驗所得的結果。

3) Peterson, J. C. *Early Conceptions and Tests of Intelligence*, World Book Co., 1925

敘述比納智力測驗和比納對於智力的概念甚詳。

四、McCall, W. A. *How To Measure in Education*, Macmillan, 1922

第一編討論測驗在教育上之應用，第二編討論編造測驗的方法，第三編敘述統計法和圖表法。書中對於力量表的編造有詳細的說明。

五、McCall, W. A. *Measurement*, Macmillan, 1939

本書為前書之增訂本。觀點比較前進，新添材料不少。陳選善在教育雜誌第三十卷第三期曾有本書之

介紹。

六、McCall, W. A. *How to Experiment in Education*, Macmillan, 1923

本書是討論實驗方法的。書中第五章介紹著者所發明的T B C F量表。

七、Wilson, G. M. and Hoke, K. J. *How to Measure*, Revised Edition, Macmillan

本書介紹智力測驗及各科教育測驗甚詳。

八、Green, H. A. and Jorgensen, A. N. *The Use and Interpretation of Educational Tests*,

Longmans Green & Co. 1929

本書專為教師寫的，內容很實際。對於測驗在教學上的應用，尤其對於診斷測驗及補救教學，討論甚詳。

九、Lee, J. M. A. *Guide to Measurement in Secondary Education*, Appleton-Century, 1936

本書專為中學校長及教師寫的，內容很切於實用。

十、Ruch, G. M. & Stoddard, G. D. *Tests and Measurements in High School Instruction*,

World Book Co., 1927

介紹中學適用的各種測驗及編造測驗的原理。

十一、Symonds, P. M. *Measurement in Secondary Education*, Macmillan, 1928

介紹中學適用的各種測驗，討論測驗結果的計算、解釋和應用。

十二 Wood B. D. Measurement in Higher Education, World Book Co. 1923

敘述哥倫比亞大學應用智力測驗之經過及結果。

十三 Terman, L. M. The Measurement of Intelligence, Houghton Mifflin, 1916

這是第一次司丹福訂正比納西蒙智力量表的指導書。關於修訂經過有簡單的敘述。

十四 Terman, L. M. & Merrill, M. A. Measuring Intelligence, Houghton Mifflin, 1937

這是第二次司丹福訂正比納西蒙智力量表的指導書以及修訂經過的說明。

十五 Ruck, G. M. The Objective or New Tyte Examination, Scott, Foresman, 1929

十六 Odell, C. W. Traditional Examinations and New Type Tests, Century, 1928

以上兩書是討論新式考試的。

十七 Symonds, P. M. Diagnosing Personality and Conduct, Century, 1931

本書討論測量性格的各種方法和工具，最爲詳盡。

十八 Review of Educational Research

十九 Psychological Bulletin

以上兩種雜誌每年有一二期專門敘述在最近一年或數年之中測驗的發展，並附有詳盡的書和論文的目錄。

二十、王書林心理和教育測量，商務

二十一、陳選善教育測驗，商務

以上兩書是專門討論測驗的中文書籍

二十二、測驗雜誌

這是專門討論測驗的月刊，由中國測驗學會編印

## 附錄二 測驗專門名詞英漢對照表

Age	年齡
chronological age	曆齡
Educational age	教育年齡
Mental age	智力年齡
Factor	因子
Factor analysis	因子分析
General Factor	普遍因子
Group Factor	團體因子
Specific Factor	特殊因子
Norm	常模
Age norm	年齡常模
Grade norm	年級常模
Percentile norm	百分數常模
Standard Deviation norm	標準差常模
Quotient	商數
Accomplishment Quotient	成績商數
Educational Quotient	教育商數
Intelligence Quotient	智力商數
Reliability	信度
Reliability Coefficient	信度係數



Sampling	取樣
Random sampling	隨機取樣
Extensive sampling	廣博的取樣
Intensive sampling	深入的取樣
Scale	量表
Age scale	年齡量表
Grade scale	年級量表
Percentile scale	百分位量表
Performance scale	作業量表
Point scale	績分量表
Quality scale	品質量表
T scale	T 量表
Score	分數
Age score	年齡分數
Derived score	引伸分數
Grade score	年級分數
Percentile score	百分數
Raw score	原來分數
Standard score	標準分數
Standard	標準
Unit	單位
Validity	效度
Validity Coefficient	效度係數

### 附錄三 中西人名對照表

Aldrich, C. G.	安特律其
Ayres, L. P.	愛里斯
Bamberger, F. E.	彭勃格
Bernreuter, R. G.	朋路透
Binet, A.	比納
Bobertag, O.	鮑勃太格
Bolton, T. L.	波爾敦
Boyace, R. C.	鮑愛斯
Brown, W.	白朗
Brunner, H. B.	勃魯納
Buckingham, B. R.	柏金漢
Burt, C.	柏推
Cattell, J. M.	卡泰爾
Courtis, S. A.	科底斯
Cunningham, K. S.	克能海
Doll, E. A.	道爾
Douglass, H. R.	道格拉斯
Elliott, E. C.	伊立歐忒
Engelhardt, N. L.	安革霍
Fisher, G.	菲奢

Freeman, F. N.	費里門
Galton, F.	高爾登
Gesell, A.	格賽兒
Goddard, H. H.	高達德
Goodenough, F. L.	古納夫
Gray, W. S.	葛雷
Hartshorne, H.	哈之兄
Herring, J. P.	黑林
Hillegas, M. B.	希里格司
Kelley, R. L.	克來
Kelley, T. L.	克萊
Kilpatrick, W. H.	克伯崔
Kirkpatrick, E. A.	克伯崔克
Kuhlmann, K.	庫爾門
Lee, J. M.	李
Mann, H.	邁恩
May, M. A.	梅
McCall, W. A.	麥柯爾
Merrill, M. A.	滿利
Monroe, W. S.	門羅
Norsworthy, N.	諾斯維塞
Otis, A. S.	渥鐵斯
Otis, E. M.	渥鐵斯

Paterson, D. C.	派忒孫
Pintner, R.	賓特納
Porteus, S. D.	鮑德氏
Pressey, S. L.	濮來西
Rice, J. M.	來斯
Ruch, G. M.	露西
Simon, T.	西蒙
Spearman, C.	斯比門
Starch, D.	斯達奇
Stern, W.	斯旦恩
Stone, C. W.	斯東
Stratemeyer, F. B.	史厥突麥
Strayer, G. D.	施菊野
Strong, E. K.	施厥浪
Studebaker, J. W.	施都特培克
Stutsman, R.	斯達次門
Symonds, P. M.	賽門司
Terman, E. L.	德爾滿
Terman, L. M.	推孟
Thorndike, E. L.	桑戴克
Thurstone, L. L.	賽斯東
Van Wagenen, M. J.	范魏格寧
Wood, B. D.	吳德
Woody, C.	吳特
Wundt, W.	馮德

中華民國三十三年六月出版

教育講  
話叢書

# 教育測驗講話

實價國幣一百二十元

外加運費郵費

版權所有  
不准翻印

編者	陳選善
主編者	陳選善
發行人	陸高誼
出版者	世界書局
發行所	世界書局

