

527.11
508

新考
論試
著煊集史

行發局書智民海上

豐

史美煊著

考
試
新
論

上海民智書局發行

中華民國二十二年六月初版

新 試 考 論

（每册定價大洋五角五分）

（外埠酌加郵費匯費）

著 者 史 美 煊

印 刷 者 民 智 印 刷 所

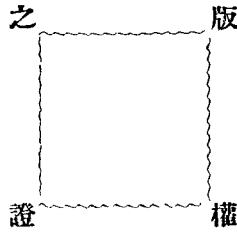
發 行 者 民 智 書 局

分 發 行 處 民 智 書 局

分 售 處 海內外各大書坊

總發行所 民 智 書 局

上海河南路中市
二〇〇至二〇二號



憲	日	之	深	服	總	貴	余
法	所	新	慚	務	理	之	常
則	希	著	愧	於	所	係	謂
功	望	丈	耳	政	理	統	學
業	之	足	史	府	想	然	校
更	建	補	美	主	之	後	與
無	設	益	煊	持	考		任
量	方	於	君	考	試		用
矣	針	教	著	政	方		宜
	有	育	考	者	克		吏
民	所	倘	試	四	實		兩
傳	貢	更	新	年	現		種
聞	獻	進	論	尚	惜		考
題	以	一	為	無	願		試
章	弼	步	中	所	奢		必
年	成	於	國	發	而		須
春	玉	賢	空	明	力		有
月	梳	平	前	珠	小		一

自序

學校考試，向來是認爲例行的事務，無人注意。直至近二十餘年，教育的研究日趨於科學化，考試方法遂大受批評。若干調查均證明考試分數之不可靠。於是標準測驗代之而興。此種新發展，一部分是兒童研究的興趣之結果，一部分則由於教育統計之最近的發達，乃屬考試方法演進之第一步。這種測驗運動，在近代教育史上佔一重要的地位，是大家公認的事實。歐美各國姑且不論，即在我國，已經出版的標準測驗不下數十種，專論測驗的著作亦頗多，由此可想見其盛況。

這種新的標準測驗之優點甚多，凡曾稍加研究或應用它的人都曉得，但亦有其弱點，如不能常用、結果不易解釋、等等。所以最近的趨勢注意非標準測驗，即所謂新法考試。此種測驗之產生，可說是考試方法演進之第二步。它具有標準測驗之特長，而無其缺欠，故頗風行一時。

本書之目的，就在詳述此種新法考試之意義、價值、編造、施行、校閱、與應用，旁及優良的考試之標準、舊式考試之改良法、分數制度、等等。至於教育測驗之高深的理論與複雜統計，則一概從略。所以本書不但便於師範學校的學生與在職的教員閱讀之用，即工商界選擇職員與行政當局考試公務員時，如能加以參考，不無裨益。

著者在二年以前，即已開始收集材料，着手編輯，但因種種關係，屢作屢輟，到現在方才草畢，然不滿意之處甚多，終因友人之催促，乃交書局付印，尚望讀者不吝賜教，幸甚。

本書之完成，得力於附錄中各參考書處甚多，未能一一註釋，謹此聲明，並示感謝。書中有數章曾於中國測驗學會出版之測驗上發表。一切譯名，多沿用舊有者，間有少數新譯者，然均列入附錄譯名對照表中。李君家湛代抄全稿，殊為感激，特此誌謝。

考試新論目錄

戴序

自序

第一章 緒論……………一

考試之批評及其答辯——考試之目的——考試之種類

第二章 優良的考試之要素……………一七

特效——可靠性——客觀性——廣博性——便易性——用途——融洽性

第三章 考試分數之研究……………三一

學校的分數之研究——教員的分數之研究——校閱同一試卷的結果
之研究——同一教員前後所定的分數之比較——問題的價值之研

究

第四章 舊式考試之改進……………五—

舊式考試之優點—舊式考試之限制—舊式考試之改良法

第五章 新法考試……………六五

新法考試之優點—新法考試之限制—新法考試之種類及其舉
例

第六章 新法考試之編造……………八一

新法考試之編造—新法考試之施行—新法考試之校閱—結果之
處理

第七章 追憶測驗……………一〇五

一字答測驗—填字測驗

第八章 真偽測驗……………一二七

真偽測驗之意義與性質—真偽測驗之優點與限制—真偽測驗之

編造——真偽測驗之施行與校閱——真偽測驗之變像——猜度與反面暗示的影響

第九章 選答測驗…………… 一三二

選答測驗之意義與性質——選答測驗之優點與限制——選答測驗之編造——選答測驗之施行與校閱——選答測驗之變像

第十章 雜式測驗…………… 一四三

對偶測驗——類比測驗——重組測驗——指認測驗——訂誤測驗——

第十一章 綜合測驗…………… 一五七

綜合測驗之優點與限制——綜合測驗之編造——綜合測驗之施行與校閱

第十二章 分數制度…………… 一六五

測驗分數與等級分數之區別——分數制度——化測驗分數爲等級分數法

附錄 譯名對照表……………一七九

參考書目……………一八九

考試新論

史美煊著

第一章 緒論

考試之批評及其答辯

考試並不是一種近代風行或新發明的方法，它的淵源甚古。在多數的學校中，成績之測量，已有其相當的地位。大概世界最早的考試制度，當推我國古代的官吏考試制。三代以上（約在西歷紀元前二二〇〇年）我國就用此法取士。他如希臘與羅馬，亦曾實行此法。但是現代歐美各國的學校考試，則發源於中世紀的大學考試。

考試雖有這樣悠久的歷史，而其價值却未確定。對於考試之批評，由來甚久。一七六六年牛津大學有一學生以爲其學位考試是一絕對的趣劇，包括一很容易的西伯來文問題與一歷史題。美國哈拉斯梅恩 (Torace Mann) 在一八四五年亦表示對於當時的考試方法不滿。然而討論之烈要算近世。近數十年

來，研究考試的文字，佔歐美教育著作之一重要部分。標準測驗之產生，即由於此。

我國自五四運動以後，有一部分教育家與在校的學生頗反對考試制度，甚至少數學校已經實行廢除，但是教育當局則力主保存之，國民政府更確立考試制度，以拔取一般官吏。故考試之存廢問題，實有研究之必要。

造成此問題之大原因，自然是由於若干調查考試與分數的價值的結果之公布。這些調查皆證明教員所給的分數之不可靠與差異，將於本書第二章詳述之。

我們歸納反對考試之理由，不外下列數點：

(一) 考試有礙於被試的健康，因為考試促成過度努力、氣餒、焦急、及其他不應有的身心影響。考試疲勞的結果，往往發生神經衰弱、神經興奮、恐怖不安等症。據茂慈(Mezi)的調查，在二百名女子小學生中，其卒業前一月，與卒業後體重之比較，平均減少七百五十克。

(二)考試的問題之內容，不適合公認的教育目標，却鼓勵強記。完全記憶事實，而無推理或其他高深的思想歷程。此種強記不但不能持久，反傷身心。

(三)考試常使本身成爲目標，學生以爲求學之最大目的是考試及格，而非精通各種學科。

(四)考試鼓勵作弊。考試時，學生每有夾帶、抄襲、竊視、或他種欺騙的舉動。

(五)考試造成應用不良的國語與惡劣的書法之壞習慣，因爲要想數量衆多，書寫迅速，遂致忽略良好的形式。

(六)考試之時間可作他種用途，如上課、溫習、研究、等等。

(七)教學的結果是無形的，無法測量。

(八)考試非必須的，優良的教員可以不用考試評判其學生的工作。

我們細閱上述八點，就知道它們都是反對考試之誤用，而非反對考試本

身。換言之，我們可改善考試之形式與方法，而不必廢除之。所以我們對於諸反對理由之總答辯，是編造、施行、校閱、及應用得法的考試，對於這些弱點均可避免。茲再分別答辯如後：

考試的結果妨礙被試的身心福利，誠屬不可諱言的事實。但是修學以及他種學校工作與校外的一切工作也是一樣，却沒有人主張因此事而完全廢除各種學校工作或校外工作。若在衛生的情況下——如溫度得宜、空氣流通、光線充足、座位適當之教室中——舉行考試，而考試的時間又不太長，實可避免此種弊端。

考試之所以不常符合公認的教育目標，乃由於編造者之無知與粗率。草擬測量事實的知識之問題，自比編撰測量推理能力、適應能力、應用能力，及心理發展者為易。然而測量後數種能力之考試，亦屬可能的事。故救濟的方法不是廢除考試，却是明定目標與給予教員以編造符合目標的考試之訓練。

若說考試鼓勵強記，至少可有二種解釋。第一，強記視考試的情形而定

。如果單純事實知識之測量達其正當地位，即僅爲考試之一很小的機能，也可減少強記，使之不佔重要地位。第二，所謂強記之某部分並不是完全不需要。學習之一原則，是常須溫習已習得者。倘使考試鼓勵溫習，則已達到一合理的目的。其實，強記總比不記爲佳。再自另一見地觀之，強記是在較短的時間內閱讀巨量的材料而保存之，乃屬一有價值的能力。律師預備案情，醫生診治病症，以及他種職業中，均很需要此種能力。

的確，有時考試本身成爲一目標。然而考試是否成爲惟一的目標，全恃教員重視考試的程度而定。如使學生認考試之成敗爲莫大的榮辱，這種態度與結果自不佳妙。在他方面，如果學生視考試爲課業之一部分，僅佔學期或學年分數之一小部分，考試非但不礙目標，且爲一合理的。尤有進者，若考試能精秘的包括已教的材料，則考試的目標，很類似精通課程的內容之目的。這些目的自然不能認爲不必需。

學生在考試時，每有作弊或欺騙情事，但是他們在其他日常工作，遊戲

、交際、及各種活動中，也如此作，却無人以為應該阻止學生工作、遊戲、交際、等等。如果考試編造得法，使被試難於作弊，並且對於作弊者予以嚴厲的處罰，大可避免此種弱點。苟欲達到此步，必須教員方面之批評的與審慎的判斷。學生是否欺騙，全恃三種原素而定：即其所受的德育、欺騙之需要、及欺騙而不被查出之機會。倘若校風不利於作弊，使被試感覺分數之增加、不敵自尊之損失；教員不太着重考試，使無欺騙之必要，而考試之編造，又使不易欺騙，自然無庸顧慮此點。

考試常注重於書寫之分量與速率，而忽略良好的形式。故考試不僅一任惡劣的書法、標點、文法、等等不加糾正，甚且發展之。其實，施行得法的考試可免此弊。後面將提出應用文字的良好習慣與發展表述的能力乃屬考試之一機能。簡要的救濟辦法，是予以充分的時間，訂出高的標準，及側重於合理的目標。

關於利用考試的時間作他種活動一點，也與對於他種活動同一見解。換

言之，有些教員用太多的時間測驗學生的工作，最好將其一部分用在他種活動上。在他方面，有些教員對於此事所支配的時間太少。考試的時間之比例，依科目或所教的科目之某方面、學生以前的預備及其成熟程度、教學的方法、與其他原素而異，故無定例。在較小的班級中，教員易用他法與各生接觸，故考試的時間可少。依照各種理論與實驗的證據，都證明時常測驗的學生比不常測驗的學生更達較高的成績標準。

『教育的產品是無形的，且是無法測量的』一語，我們承認其有相當的真理。若把它們與工業及科學的出品比較，實是無形的。然而它們的區別是屬於程度的而非性質的。在某些工作方面——如教育之類——的確不能測量其結果達一極精密且完全的程度。以前對於正確的科學也是同樣的情形，就是現在，它們的測量之準度也非絕對的，却僅比較高。此事自不能使我們主張廢除教育測量，却應引起研究與改良測量工具之決心。雖然心理的發展與成績不能達到高度，重量、面積、溫度、等等測量之客觀、可靠、與特效，然

可發展至一較現時爲完全的較高程度。加之，有了比較不精密的測量，總比沒有測量更佳。

對於考試不必需之答覆，我們祇須說明一點，即一切關於此事之實驗的證據，都證明此點錯誤。偶或有少數優秀教員可以不用正式考試估計其學生的成績。又或有少數教員能不借助於考試而鼓勵學生或發現其教學的弱點。然而這種教員佔極少數，多數的教員皆須應用考試爲其評判學生工作與增進其教學效率之根據。

現在我們可引證哈佛大學校長羅維爾 (Lowell)的話作結，他在考試之技術一文中說：「結論……：是使用得法的考試爲教育歷程中之一重要部分，但是應用它們使生最佳的結果之技術是極其複雜且繁難。」

考試之目的

任何活動必先決定其目的，然後才好選擇方法與材料。若不考慮其目標，雖最有價值的活動也不能很有效率，而其可能性僅部分利用。考試之一顯

著的障礙，就是把它單作一例行事務，教職員按期舉行之，學生則盲目受試。所以我們必須詳細研究其功用，然後方可避免其缺陷而儘量利用之。關於考試之目的，各家的見解不一，但有一共同之點，即一優良的考試，常具數種重要之目的。然而各種目的每有互相重複的或可再加劃分的，我們必須加以注意。例如，一測量知識之記憶的考試，也有某種診斷的價值。茲將考試之主要目的分述如下：

(一) 測量學生的成績與能力。教員對於學生的成績與能力，每乏精確的認識。學生對於其本人的成績與能力，也不很明瞭。教員與學生雙方均以爲所知者甚多或能力甚大，究竟其實際如何，却不得而知。但是此事與日後_的學業很有密切的關係。故有效的學習與教學，必須測量學生之知識與能力。至於測量的方法，則隨需要而定；或用追憶法、或用認識法、有時或須知道學生明瞭事物間的關係之程度、或應用等等能力。此種測驗的結果，一方面可覘學生的學業進步，以定分級、升級、畢業、等等之是否適宜；他方面

指示我們以學生的能力地位，作將來教學的根據。考試之測量的功用自很重要，但是以前未免過於重視，幾認為考試之惟一的目的，致引起許多反對，我們不可不留意之。

(二)供給客觀的標準。凡屬稍加思索的人，皆知供給學業以客觀的標準之重要。如欲精確的比較諸年級或年齡，或對學業作一可靠的判斷，必須有一精確的客觀成績標準。科學的課度標準之決定，也須借重它。然而供給客觀的成績標準，須作某種測量。在教育範圍內，測量是根據於考試之結果。測驗的分數表示某年級、某年齡、或某組別的學生所作的或已成就的。累積這些記錄，可製成一正確的標準。他日應用，祇須將各年齡或年級的成績和它對比。

(三)診斷特殊的困難。有效的教學必須診斷學習之特殊困難，尤其是對於工作不滿意者。教學好像行醫。醫生必先診察患者之病源而後醫治之。教員也應採取同樣的科學態度與方法，診斷學習的缺點或弱點，診斷的作用是

要發現全班共同的困難或個人的弱點，最需要何種教學。它注意於學生所未習得的及其未曾習得的原因。它要顯出學生已習得的材料和其能力、態度、性向、興趣、與其他影響學業成功之特質。它的目的是救濟工作。自然此種救濟工作不限於狹義的特殊學科，却為廣大範圍的教育與職業指導。但是若無測驗，則此事必不可能。

教育診斷涉及某班或某個人。班級診斷僅須分析一普通調查攷試各問題之答案。至於學科的弱點之診斷，則應集中考試問題於弱點所在的工作單位。測驗之最後的目的，是改良教學，而診斷則為達到此事之最直接的方法。有了測驗結果，就可引起教員分析其工作，以發現教學不良之原因。

(四)引起學習與溫習的動機。近年來引起動機已成許多新教學法之基本原則。凡能達到此目的之方法，均應重視之。自理論而言，學生對於工作之刺激，應於愛好工作、認識工作的價值、或精通工作的需要之直覺中求之，然而理論的與自然的動機，對於學生不常有效，多數學生需要一更明瞭且直

接的動機。教員必須借助於人爲的興趣。考試適符合此條件。成績記錄爲學習之一大鼓勵，已經實驗心理學反覆證明。依據心理學的定律，學生知其工作將被批評，並明瞭其進步的情形，必更努力。所以知道結果的學生之成績比不知結果的更佳，考試則告知學生以其本人與全班的成績，使他們明瞭自己與全班的學業程度，促其競爭心。根據考試的結果，可引起組別間、個人間、並與本人已往的記錄競爭。

溫習工作爲良好教學之一要素。很多優良的教學之失敗，就因爲缺少溫習。學習尙未告終，遺忘即已開始。考試却爲引起溫習工作之一良法。對於考試之充分的預備，實不僅強記已也，必須熟思且批評的閱讀全部材料，考試予學生以統觀科目的全體、集中注意、練習、組織、應用所學者之機會。

(五)訓練文字發表力。主張舊式論文考試之一最大的理由，即其教授學生組織其觀念且用良好的國語表述之之高大的價值。然而實際的結果却不如，教員每發現考試對於國語訓練生反面的影響，其成績較平常的尤劣。此

事已在前面提及。所以我們如欲考試達到此種目的，最好把它劃出一部分作爲此種訓練，並附說明如下：

『此種考試的分數是根據下列原素：（一）思想的表現，（二）良好語句與節段結構，（三）文法與默寫的錯誤，及（四）其他。』

（六）測量與改善教學效率。教學效率之測量是一舊而難的問題。教員與教育家久已承認精確的測量教員工作之必需。異常的成績應加獎勵。不良的服務須加淘汰或改進。但欲求一滿意的測量法，殊不易得。依舊式的方法，一校長或視學走進一教室，參觀數分鐘，於是根據所得，對這教員加以批評。此法過於主觀。

考試予發明客觀的測量教學法以豐富的園地，一研究考試的問題，即顯出教員是否熟習測驗技術、了解教材、及教材之選擇是否得當。至於教員對於教學法之知識，也可用專爲此事而編的職業測驗測量之。然而教員之真效率應以他對於學生之服務而定。最好的教員乃是使學生進步最大的。自此種

觀點而言，以測驗學生法度量教學效率，很為可靠。如用此法，則於某定期之開始和終了，各施行測驗一次，而以其間的進步歸於教員。不過應用此法評判教員時，應很慎重，因為尚有許多他種原素，如學生的能力、教材的難易，等等，影響教學效率。

普通以為教學效率之改良僅限於劣等教員，其實優良教員也應隨時改進。因為無一教員為一完人，教學經驗都是逐漸累積而成的。劣等教員自應明瞭其教學之缺點，使知改良之必需；而優良教員也須知道其成就，使更加努力，否則必難改良之。

在許多改良教學效率的方法中，考試之可能性很大。一詳細研究測驗的結果，教員即明白學生之成績地位及其缺點與困難。有了此種知識，自可使教學適應學生的需要與能力。測驗的結果不但可用以評判教員之優劣，且可用以定奪各種教學法之效率。於是指導教員選擇最佳方法。

考試之種類

考試的種類繁多，其最普通者爲下列四種：

(一)口試 考試方法以口試爲最古，而反對口試者，則較反對筆試爲尤烈。然而口試實爲日常課業之一重要部分，要在教員運用得法。但若用作測驗之工具，則不甚可靠，其理甚明，無待贅述。

(二)舊式論文考試 此乃指通用的筆試而言，久爲教員估計學生的成績之工具。大概現在最常用的考試方法要算它。至於它的優劣點及改良的方法將在第四章詳論之。

(三)標準測驗 標準測驗卽由專家編造之釐訂標準的考試。它的歷史甚暫，僅有二十餘年。奈斯 (Nis) 諸人實爲其先鋒。現在不但出版的各種標準測驗甚多，且極風行。舉凡舊式考試之弱點，標準測驗幾均避免之，無怪其受教育人士之歡迎。

(四)新法考試 這是教員模倣標準測驗而編的考試。它一方面避免了舊式考試之弱點，而他方面又彌補了標準測驗之限制，最宜於教員之用。本

書之主旨，就在專論此種考試方法。

第二章 優良的考試之要素

考試或測驗之優劣，依個人的評判而定。批評考試的方法，和批評藝術的方法頗相似。富有經驗的批評家，均不借重一般的印像。他們預先定出某些重要的特質，作為評判的標準，以資比較。然標準隨事物之性質而異，標準測驗與非標準測驗之性質既不相同，其標準亦因之稍異。本章所討論的諸要素是以標準測驗為根據，但其中之大部分要點，皆可應用於新式的客觀測驗，而於上述二種測驗之選擇或編造都有裨益。茲列舉測驗專家所公認為估計一考試的價值之標準如下：

- (一) 特效
- (二) 可靠性
- (三) 客觀性
- (四) 廣博性

(五) 便易性

(六) 用途

(七) 融洽性

大概此七種特質爲一優良的考試所必具的標準。一考試具有這些特質的程度之高低，定奪其價值。茲分別討論之。

特效

優良的考試之首一最重要的標準是它所具的特效。關於特效的意義，各人有各人的解釋。簡言之，特效就是一考試測量它所想測量的事物或能力之程度。如果一測驗的確測量它所欲測量的，則此測驗有特效，反之則無。比方，我們皆知重量是一優良的度量法，但是用以測量一人的誠實程度，即無特效。測驗的名稱，不一定保證它測量該事物或能力。時常一優良的測驗，一經審慎的考察，就發現其測量他種能力——不是它所要測量的——達此種程度，致破壞後者之測量。例如，一測驗雖名爲算術測驗，其實測量默讀能

力，而非加減能力，此種測驗自無特效之可言。

測驗的特效之高低，每隨其可靠性、客觀性、廣博性、及其他原素而異。增進測驗的特效之方法，不外下列三種：（一）遵照專家的意見，（二）根據實驗的結果，及（三）二法兼用。第一法包括專家的評判、課程、或課本之內容與形式的分析、及教育團體對於學程課程最低標準、等等之建議。依照此法，最重要的測驗問題可依課程大綱、基本教科書、及最低的標準決定之。所選擇或包含的材料應有價值且關緊要。

第二法則包括社會用途的實驗的研究、最常見的錯誤之研究、計算各年齡與年級的學生答對各題之百分比、及核算本測驗和他種標準之相關度。此法中之社會用途，可舉默字測驗中之最常用的字與算術測驗中之商業中算術的需要之研究爲例。

標準測驗之編造，均採用約制的實驗法。通常的學校考試，大都根據專家的意見，因實驗法既繁而又難，對於教員不甚經濟及可行。所以教員的考

試，常兼用下列三種標準：（一）本地的學程，（二）分析所用的課本，及（三）本人對於應側重、包含、與除外諸點之判斷。測驗的特效法之指導的原則是：測驗須符合實際的教學。編造測驗時，教員應把他所想測量的能力及科目的教育目標明白存在心中，目標雖然不影響測驗的性質，却常使它所包括的教材之種類與形式大異。故測驗必須符合科目之目標。

測驗的問題之排列，對於其特效有很大的影響。問題的難度之判定，須經勞力與價昂的實驗。然為簡便起見，亦可採用教員的判斷。二三教員之共同的評判，自較一人的為優。論到難度，凡屬一切皆答對或答錯的測驗問題均無特效，故都不應包括在測驗中。但為鼓勵起見，測驗之首數題應易，而為防止被試得完全的分數起見，測驗之末應有數難題，換言之，特效的排列法是問題由易而漸難。

測驗中若有完備的說明且多用舉例及示範，都是增進測驗問題之特效的方法。

特效的測驗應避免或減少各種無關的瑣碎原素，如不合文法、書法端正、意義含糊、或其他不明瞭處。除非此種測驗是為測量此事而編底，這種原素不應成爲測驗之一部分；苟欲測量之，可使成一單獨的測驗。

此外尚有一事須記在心中，即長的測驗比較短的測驗更有特效。短的考試常有種種缺點：如不能十分包括全範圍；過於側重所測量的知識，而忽略教材中未被測量的其他同等重要的部分；引入機遇原素，一學生幸而知道此特殊問題，但對其他知識則毫無所知，亦得良好的分數。並且測驗若長，縱使其中許多個別問題錯誤或無價值，其消極的影響亦爲良好的問題所抵銷，而得一很特效的量數。

可靠性

優良的考試之次一重要特質爲其可靠性。可靠性是一考試測量它所測量的事物或能力之準度。簡言之，可靠性就是一學生對於一測驗先後或二次所得的分數之一致的程度。倘若結果絕對一致，此測驗即完全可靠，否則不可

靠。一不可靠的測驗工具，造成判斷的錯誤。

可靠性是一較特效為狹義的名詞，它是特效之一方面。特效包括可靠性，但可靠性不定包括特效。此點常混亂。茲舉例說明之。例如施行一特效且可靠的國文測驗，而將其結果標作手工的測驗，其實在的分數仍屬可靠，但因此種錯誤的命名，却完全無特效。特效與可靠性既如此密切相關，故測驗之實際編造時，前者之準備，大都顧及二者。

就理論而言，一測驗如欲完全可靠，必須無限的長。除非所測量的能力極有限，一考試可包括且測驗其中的一切材料，此測驗不算完全可靠。但嚴格言之，就是這種測驗，亦不完全可靠，因它尚測量其他事物或能力。既然一考試不能包括太大的內容，故可靠程度常視取樣之優劣或廣狹為轉移。他事相等，取樣愈廣，可靠性愈高。因此增加一測驗之長度——增多問題數目或加長測驗時間——可靠性常增。測驗之長度則以考試之目的、已讀的教材之分量、及學生之成熟程度為根據。就實際的工作時間而言，中學以及小學

低年級以九十分鐘至五十或二十分鐘爲準。

前已言之，若把所想測量的原素明白存在心中，且淘汰一切無關的分子，則測驗的特效大增。此法亦同樣的增加測驗之可靠性。測驗問題的難度範圍應很廣，足予全班的能力或成績的範圍以充分表現之機會。一方面測驗應很難，以免任何人得完全的分數。測驗若僅普通的難度，則高能力的學生均得完全的分數——因未達其能力限度——於是優秀學生之差別仍未測出。在他方面，測驗須甚易，以免任何人得零分。問題若太難，有些人的能力不及，必都得零分，其差別亦未測出。如果分數分配於全距離中，自最低的以至最高的，此即一優良的測驗。

客觀性

客觀性也是優良的考試之一要素。所謂客觀性，乃指一測驗不受施行者與校閱者之個人差律或主觀的影響而得同樣的結果而言。換言之，一切經驗的閱卷者，對於一測驗的一切答案之正誤的判斷，意見皆一致——即均予以

同樣的分數——此測驗即客觀。在客觀的測驗中，一切主觀原素均被淘汰。個人的評判不很相同。我們各有我們的成見、癖性及偏心。這些原素均足影響我們的主觀判斷。故對同一事物，各人的估計不同；有人太過，有人不及，而無人恰對。教員評閱試卷之顯著的差異，已屬人皆知之的事實。新法測驗之問題，如真偽、填字、選答、等等，大都避免施行與校閱之主觀的影響。

在客觀的考試中，校閱者的感想與感情，對於分數都無影響。無論何人校閱，同樣的答案均得同樣的分數。自然，教育測量中，並無一絕對客觀的測量表或標準，各種測驗僅使判斷較不主觀。其準度視量表的單位之適宜及應用之慎重而定。它們祇輔助人類的判斷而不替代之。人類的判斷，若無輔助，是不可靠底。

廣博性

廣博性是指一考試應包括一很廣博且足代表全體的範圍而言。有些事物

因其特有的性質，不能完全測驗之。遇此種情形時，必藉取樣以度全體。一人之能力其全部的成績、或其對任何科目之知識，皆不能完全測量之。所能底是其能力或成績或知識之取樣。由他對於該科的取樣之表示，以推測其總知識。故一考試總不完全，僅爲它所欲測量的知識或能力之取樣而已。即或一測驗能包括所有的教材，其測量仍不完全，因爲答復問題時，心理原素——如粗心、急促、疏忽、體弱、努力、與動機之變更、等等——常有升降。至於此種結果之可信的程度，大都視取樣之完全與否而定。不精密的取樣，對於被試很不公平。

現行的考試採用二種不同的取樣的見解。第一種爲內包的取樣說，以舊式論文考試代表之。第二種爲外延的取樣說，乃新法客觀考試或測驗之特徵。前者常含五至十個廣義的問題，必須詳細答復。後者則包括五十以至二百多個狹義的問題。他事相等，外延的取樣更爲可靠。

最好的辦法是把一考試中之各問題當作一個別的且獨立的取樣。他事相

等，測驗愈表，取樣愈精；取樣愈精，結果愈確。教員應確知一考試有充分數目的代表的問題。所謂『代表的』，意即足以代表教材之某範圍。但是測驗的問題不可重複。編造一廣博的考試時，應注意問題之種類與數目二者。自然，任何考試之題數皆有一限度。採用一問題，乃因其代表他問所未代表的某事物。巨數的同類的取樣，並不大增結果之價值。

便易性

優良的考試，一方面應使教員易於施行及校閱之，而他方面應使學生易於受試。此即測驗之便易性。換言之，測驗應節省教員與學生雙方面的時間及精力。學生對於新式的測驗，沒有對於舊式考試那樣的熟習。測驗的問題比普通論文式的问题尤為複雜。故一測驗必須備有說明，使學生知道希望他作何事。然而教員應記得一教育測驗是為測量成績底，不是測驗說明的了解力。

測驗之編造比較舊式考試更費時間，但是此種時間之一部分，却可在校

閱時節省得之。故測驗問題之答法務須客觀且經濟，例如祇須作一符號，不必書寫字句。如此則考試的結果不受書寫的速率之影響，而同時校閱的事務亦簡化，因為祇要一答案標準，雖然不懂該科知識的人，亦可校閱。並且此法很能節省校閱之時間與精力。校閱最好用一標準紙或機械法行之，大概閱卷的時間不應超出考試時間的五分之一。

便易性也包括測驗的費用在內。如屬可能，最好各人都有一本試卷，比用口述法或黑版法為優，總之，凡使測驗易於施行與校閱之原素，皆使之較佳，不過我們同時也應把節省經濟一點存在心中，不可為便易之故，耗費很多金錢。

用途

關於標準測驗之解釋，常模很有價值。故一優良的考試，必須具有常模，藉以比較個人的或團體的成績。然而教員編造其本人的測驗時，却無現成的常模，常模對於解釋此種測驗之結果雖很有用，但是產生此種標準之耗費

，使之得不償失。教員可由累積的記錄，製造本校或本地的常模，並且其解釋很合乎實際。

優良的考試應有替換式。所謂『替換式』，就是難度略同而問題互異之數種同樣的測驗。替換式之數目愈多，則測驗之用途愈大。替換式對於教員極其有用；例如教員常須編一新考試測驗缺席的或不及格的學生，但前後二測驗之難度，難免不一致。如有替換式，可免此弊。每一考試常有疑難的情形，例如一學生的分數比其普通的成績特別壞。遇此情形時，必須重測，而替換式正適應此種要求。在同一次的考試中，可輪流分配替換式於各人，以防止學生抄襲、竊視、或其他種作弊法，而鼓勵澈底的研究。

優良的考試須能顯出個人的或團體的弱點及缺陷與應採行的補救方法。此即測驗之診斷作用。診斷不僅要顯出各人對於已習的教材之差異，且須顯出其能力、態度、興趣、等等。診斷之最大作用是求出學生需要何種教學，它注意於學生所未習得者及其未習得之原因，以便應付之。

融洽性

個人對於一種活動的態度之性質，與其成績之特性有很大的關係。一喜歡的工作比較一厭惡的工作常作得好些。現代的教學法每努力使工作具有興趣，即因此故。學生視考試爲畏途，且懷疑其價值。考試應有充分的趣味與刺激，並造成一種安適的感情。於是他們必深覺受試有趣且有價值。

編造考試時，應顧及學生的態度。考試須使他們滿意於結果之公平。考試之首數題若甚易，則可造成學生之自信心與對於難題之正當的『心向』。

考
試
新
論

第三章 考試分數之研究

考試分數，在普通一般人看來，是完全正確且可靠，並無錯誤；即或偶有差異，亦不過毫釐之別，沒有很大的關係。孰知依近年教育家之研究，其結果適與常人之忖度相反。此乃標準測驗產生之一大成因。現在把幾種重要的調查分述如下：

學校的分數之研究

柯特爾之研究 柯特爾(Catell)是研究這個問題的先鋒。他在一九一一年調查了某三小學校畢業生之算術分數，其分配列後：乙校學生有三分之二在七十八分以下；三分之一在七十八分以上。甲校學生有三分之一在七十九分以下；三分之二在七十九分以上；三分之一在八十四分以上。丙校學生有三分之一在八十二分以下；三分之二在八十四分以上；三分之一在八十八分以上。就上述的情形而言，丙校學生的算術程度最高，甲校次之，而乙校最劣。

此種分數如果正確，則集此三校學生於一堂教學時，丙校學生的算術能力應較其他二校的學生爲優。此三校畢業生後果入同一中學肄業。柯特爾一調查他們的算術成績，竟與以前的結果成反比例。乙校學生原爲最壞的，平均成績較爲優良；而丙校學生原爲最優的，平均反最劣。由此可知三校的分數很不準確。

克雷之研究 上述的調查，單祇研究算術一科。一九一三年克雷(Keel)曾作一各科的比較之研究。他先搜集了某四小學校的六年級生之各科成績。後來此四班學生同入另一學校之七年級。他又把他們在第七級的成績收集起來，將六七兩年級的分數作一比較。他發現丙校所給的國語、習字、歷史、等科之分數，都高於丁校的。丁校之乙等的價值，幾與甲校之甲等相等。

教員的分數之研究

約翰生之研究

芝加哥大學附屬中學的校長約翰生(Johnson)也曾研究這個

問題。他統計該校一九〇七年至一九〇八年與一九〇八年至一九〇九年兩年

間各科教員所給的分數之分配。結果如下（見第一表）：英文科不及格的學生佔百分之15.5，歷史科不及格者僅佔百分之8.1。二者相較，英文科不及格的學生幾二倍於歷史科的。而英文教員所給的不及格，將近二倍於家政教

第一表

芝加哥大學附屬中學各科的分數之分配

科 別	人 數	E		D		C		B		A	
		人 數	百 分 比	人 數	百 分 比	人 數	百 分 比	人 數	百 分 比	人 數	百 分 比
拉丁文與希臘文	886	94	10.6	143	16.1	282	31.8	208	23.5	159	17.9
德文	416	35	8.4	81	19.5	110	26.4	119	28.6	71	17.1
法文	475	52	10.9	89	18.7	157	33.0	133	28.0	44	9.3
英文	1514	235	15.5	329	21.7	497	32.8	354	23.4	99	6.5
歷史	825	67	8.1	131	15.9	258	31.2	248	30.0	121	14.7
數學	1466	212	14.5	370	25.2	405	27.6	310	21.1	169	11.5
理科	672	56	8.3	113	16.8	186	27.7	219	32.6	98	14.6
手工與圖畫	867	76	8.8	122	14.1	289	33.3	287	33.1	93	10.7
家政	176	10	5.7	4	2.3	48	27.3	91	51.7	23	13.1
總數	7297	837		1382		2282		1969		877	
平均數			11.5		18.9		30.6		27.0		12.0

員的。在優等的成績方面，法文科得甲等的學生佔百分之 9.3，德文科得甲等的學生却佔百分之 27.1，亦幾二倍之多。拉丁文與希臘文的教員所給的甲等約三倍於英文教員的。但照該校的課程標準，英文與歷史佔同等的地位，法文與德文也佔同等的地位，似不應有此種大差別，其原因仍不外乎分數不正確之關係。

邁爾之研究 一九〇八年邁爾 (Meyer) 研究密梭理大學各科的教授所給的分數，結果如下(見第二表)。我們試看此表，哲學科的甲等佔百分之 55，而

第二表
密梭理大學各科的分數之分配

科 目	A %	B %	C %	F %	總 數
哲 學.....	55	33	10	2	623
經 濟 學.....	39	37	19	5	461
德 文.....	26	38	25	11	941
機 械 畫.....	18	29	41	12	538
英 文.....	9	28	35	28	1028
化 學.....	1	11	60	28	1903

化學科的甲等僅佔百分之一，相差至 54。至於不及格者，哲學科只有百分之 2，而英文與化學二科中，各佔百分之 28，不為不多。由此可見同一校中，各科的分數之分配，也不一致。

恆理克遜之研究

恆理克遜 (Hendrickson) 曾舉一大中學校中各部定學生的

分數之標準互異的另一例

第三表
一九二七年六月凡那士中學各部的分數之分配

部 別	A %	B %	C %	D %	E %	分數總數
藝術	29	32	29	8	2	302
商業	21	39	33	3	4	348
英文	10	27	23	22	13	984
歷史	13	27	36	16	8	710
家庭經濟	21	36	23	10	5	288
語言	19	32	24	7	18	323
數學	9	27	32	22	10	596
機械藝術	26	45	20	6	3	462
音樂	40	36	16	8	0	684
體育	18	58	20	3	1	886
理科	22	33	30	9	6	469
諸部	20	36	26	11	7	6016
初級中學	15	33	29	16	7	2966
高級中學	26	31	26	9	8	2153
專門部	13	28	31	17	11	3046
普通	26	44	22	6	2	2970

他的結果如下（見第三表）。他解釋此表說：總分數的百分之 50 為 A 或 B 或可入大學，而 D 與 E 僅佔百分之 20。換言之，分配很偏斜於最高的分數。但是該校所施行的智力測驗與教育測驗，結果均近

於常態曲線。並且此地的智力程度也近於常態。

盧赤之研究

盧赤 (R. C. H.) 也作過一類似的調查，第四表即其結果。此表表

示阿理根大學附中六星期內六五九種分數之研究。P. T. 是指一切實習教員的

第四表
六星期內阿理根大學附中的教員所給的等級分數

教員	百分比				
	A *	B	C	D	E
甲.....	48.7	44.7	6.6	0.0	0.0
乙.....	28.6	58.6	12.8	0.0	0.0
丙.....	15.2	66.3	15.7	1.7	1.1
丁.....	41.8	48.6	6.7	2.9	0.0
戊.....	8.9	78.6	12.5	0.0	0.0
己.....	30.0	55.0	15.0	0.0	0.0
P. T.	46.8	37.0	8.4	3.2	4.5
總共.....	32.0	54.0	11.0	1.6	1.4
學校標準...	6.25	25.0	37.5	25.0	6.25

分數之總和。總分數則包括正式教員（以甲、乙、丙、等代表之）與實習教員二者。學校標準是教員會議所採為該校之正式標準。

不顧學校標準分數之存在，就全體而論，分數異常偏斜向上。教員戊更證明一有趣的情形。在前次報告期內，曾將各教員的分數分配及其圖解貼在教員室中。這位教員所給的分數，在百分之五十以上。為要避免此

種情形，他極力減少A分數，致在各教員之下，而所用的方法，則很簡單，即改A爲B。

狄爾朋之研究

狄爾朋 (Dearborn) 曾調查同一校內，同一班次，教授同一

課程的二教員所給的分數之分配。二人中之一人，給全班學生百分之 α 以最優等，而無一不及格者。其他一人不但不給最優等，而不及格者竟多至百分之14。

芬克斯頓之研究

芬克斯頓 (Finkelsien) 調查某校某年級上下二學期的分數

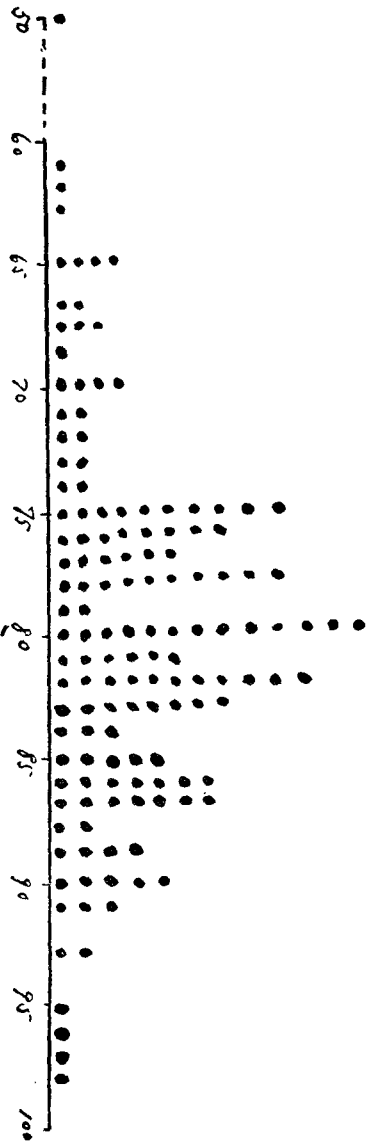
之分配，列爲一表（見第五表）。該級的學生和科目都未變更，不過上學期歸教員甲教授，下學期由教員乙教授。但是分數相差很遠。五十九分以下的與免考的人數之百分數，幾爲一與三之比。

第五表
二教員對於一班學生之同一
科目的分數的分配

分 數	教員甲	教員乙
0—59.....	9.5	3.5
60—59.....	25.5	24.9
70—79.....	32.3	34.1
80—89.....	26.6	43.7
90—100.....	6.1	3.8
免考者.....	12.5	37.7
學生總數.....	263	25.7

校閱同一試卷之研究

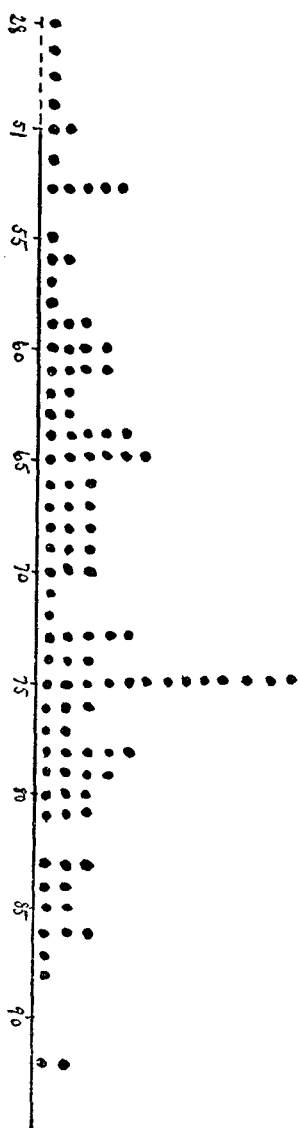
斯達赤與依理特之研究 斯達赤(Starch)與依理特(Elliot)曾作數種苦心的探索。他們把兩本中學一年級的大考英文試卷，按照原樣摹印出來，雖其中之一字一劃，以及改削處等等，都與原本無異。然後寄與各校中學一年級的英文教員，請其依本人常用的記分法評定分數。統計一百四十二位英文教員的分數之分配如下(見第一圖)：



第一圖 一四二位英文教員對於同一英文試卷所給的分數 (甲卷)

甲卷之最少的分數是50，最高的分數是98。乙卷的分數則由64至98。此二卷的分數之距離，皆在三四十分之間。一種試卷不容有二種不同的分數，今竟有這樣大的差異；有人以為這答案最佳，給以九十八分，有人以為它最壞，給以五十分。這還可靠麼？

我們或者以為語言一科，無一定的標準，所以有許大的差異。若為準確的學科，如數學之類，必無此種現象。因此斯達赤與依理特又做了一數學的



第二圖 五位中學數學教員對於同一幾何試卷所給的分數之分配

實驗，證明此說。他們拿一本中學的幾何試卷，仿前法印就，請一百十六位中學的數學主任教員，根據百分法，評定分數（見第二圖）。一看此圖。即知數學的分數，也是毫無標準。圖中最低的分數是 2% ，而最高的分數爲 92% ，差至 64% 。此一百十六人中，給 90% 分以上者二人，給 30% 分以下者一人；給 20% 分以上者二十人，給 0% 分以下者亦二十人；認該卷爲及格者四十七人，認它爲不及格者六十九人。意見之參差，較英文分數之分配爲尤甚！

他們還仿前法，研究一本美國史的年考試卷。七十位教員評閱該卷，一人定爲 55% 分，另一人却給 0% 分；十人把它列在 80% 分以上，其他十人則把它列在 55% 分以下。結果與上述者略同。

盧赤之研究 盧赤曾作二種類似的研究。第一種實驗是把一地理問題之三種不同的答案，請九十多位教員定分。這三種答案是選自一班的試卷中的，代表最優的、最壞的、及中等的試卷。學生的答案是逐字抄下油印的，保留其一切錯誤。第六表歸納九十一位教員重閱三答案之分數。

第六表

一地理問題之三種答案的原來的分數與91位教員重定的分數

分 數	答 案 一	答 案 二	答 案 三
20.....	1		9
19.....			3
18.....	1		21
17.....	1		12
16.....	2		17
15.....	10		15
14.....	2		3
13.....	8		1
12.....	14		3
11.....	5		1
10.....	24		3
9.....	5		1
8.....	7		1
7.....	3		
6.....	4		
5.....	1		
4.....			
3.....	2		1
2.....	1	3	
1.....		1	
0.....		87	
人數	91	91	91
平均數	10.9	0.1	16.1
均方差	3.2	0.4	2.9
原來的分數	16	2	12

第二種實驗是自一第七年級的美國史班中取出三本試卷，油印之，除保留一切拼字與文法的錯誤外，還儘量保留其原來的式樣。這些試卷也是班中之最優的、中等的、與最劣的。一一五位教員評定的結果見第七表，與斯達

第七表
一美國史之三試卷的原來的分數與115位教員
重定的分數

分數	第一卷	第二卷	第三卷
100.....	6		
95.....	33		
90.....	32	1	
85.....	22	12	1
80.....	15	13	4
75.....	6	29	9
70.....	1	18	5
65.....		16	20
60.....		12	16
55.....		9	19
50.....		3	13
45.....		2	8
40.....			14
35.....			5
30.....			0
25.....			1
人數	115	115	115
平均數	88.7	70.3	56.6
均方差	0.6	9.9	12.3
原來的分數	100	88	67

赤和依理特的報告之性質相似。

吳德之研究

吳德(Wood)研究一九二二年六月大學入學考試委員會之代數，

和幾何考試，更顯出舊式考試的分數之不可靠。他用隨機取樣法選出代數和

幾何試卷各四百份，請二人各自記分二次。代數和幾何之第一次與第二次記分的相關度皆很高，約 $\cdot 98$ 與 $\cdot 96$ 。記分雖如此客觀，然考試本身之可靠性則很低：代數考試之一半的相關度僅 $\cdot 61$ ，而幾何考試的僅 $\cdot 52$ 。用白朗的公式 *Brown's formula*，全部代數考試之可靠性不過 $\cdot 76$ ，而全部幾何考試的祇 $\cdot 58$ 。

吳德更告訴我們一可笑的事。一九二〇年夏間若干大學歷史教授校閱一組歷史試卷。其中有一教授，校閱數卷後，為本人之便利起見，填就一份他視為模範的試卷。後於無意中，此卷混入學生的試卷內。他一閱卷者不知此事，却把它當作學生的試卷評閱，竟予以不及格的分數。於是所有的閱卷者均評閱之，所給的分數差到40至90。

柏拉德之研究 柏拉德 (*Ballard*) 曾舉出一事實，證明教員的分數受已往的成績之影響。他說某英語專修學校之甲乙二生異常親密。他們同班，每二星期作文一篇。作文時，他們彼此交換意見，互相指導，故論文實很相似。甲

生之首次論文的等第爲『最優』，乙生的是『尙佳』。此後二人之地位常與以前相同。某次他們把彼此的文章交換——甲抄乙的作爲甲的，乙抄甲的作爲乙的——而分數則仍舊。

法倍爾與盧赤之研究

法倍爾 (Pauber) 和盧赤曾作一實驗，取一內容價值很

第八表

法倍爾研究形式對於一美國史試卷的影響之結果

教員號碼	1	2	3	4	5	6	7	8	9	10	11	12
試卷一甲	96	92	93	91	91	100	82	92	91	100	76	94
試卷一乙(內容)	95	92	93	91	92	100	83	92	91	100	76	94
試卷一乙之扣分：												
清潔	10	2	5	5						4	4	1
形式		2		5						4	5	2
拼字		2		10						4	5	2
文法		2		5						14	3	1
扣除總數	10	8	5	25	10	0	10	0	0	26	16	1
試卷一乙之最後的分數	85	84	88	66	82	100	73	92	91	74	60	88

高的八年級歷史試卷，稱它爲試卷一甲。然後改變此卷，用一粗心狀態書寫之，作種種文法、標點、拚字、等等錯誤。有些部分匆匆劃去，且有少數墨水漬。這個重寫的試卷稱爲一乙。內容或思想則未變更。

十二位教員給試卷一甲的分數之距離爲76至100，即24分。試卷一乙的距離爲60至100，即40分。此額外的20分之距離，則由無關的原素而生。

鮑爾頓之研究 以上的調查都證明教員的分數不一致，但最近鮑爾頓 (Bo-

ton) 根據他的研究之結果否認此事。他請若干六年級的算術教員編一考試並施行之。然後用取樣法選出其中之二十四本試卷，交由二十二位教員校閱之。分數的平均差約百分之五。第九表總括其結果。

我們對於這個研究應注意下列三點：(一) 他用很客觀的科目——算術。(二) 他用很少數的教員二十二人——大可減少差異。(三) 他用平均差代替極端的差異(全脫離)，距離最高的爲 $\frac{1}{2}$ ，而全距離之中數約25。

第九表

摘錄鮑爾頓研究教員的分數之資料

(甲)	(乙)	(丙)	(丁)	(戊)	(己)
學生	最低分數	最高分數	全距離	二十二位 教員之 平均數	平均差
1	77.5	100	22.5	88.7	3.6
2	75.0	90	15.0	85.0	4.3
3	70.0	95	25.0	88.7	3.5
4	43.0	71	28.0	57.8	5.9
5	65.0	95	30.0	84.6	6.4
6	25.0	70	45.0	51.0	10.5
7	74.0	91	17.0	84.8	3.1
8	73.0	93	20.0	84.7	5.4
9	85.0	100	15.0	93.8	2.7
10	71.0	95	24.0	89.6	3.4
11	66.0	80	14.0	77.5	2.8
12	84.0	90	6.0	88.0	1.4
13	46.0	85	39.0	71.5	7.8
14	67.0	84	17.0	74.4	2.3
15	47.0	74	27.0	62.7	4.6
16	82.0	98	16.0	91.5	3.9
17	53.0	95	42.0	78.9	8.4
18	70.0	91	21.0	77.9	4.4
19	37.5	78	40.5	53.5	7.2
20	43.0	71	28.0	55.3	4.0
21	48.0	88	40.0	74.5	8.5
22	35.0	58	23.0	45.1	4.7
23	43.0	75	32.0	65.3	7.3
24	17.5	47	29.5	27.2	5.8
中數	65.5	89	24.5	77.7	4.5

斯達赤之研究

斯達赤對於同一教員前後所定的分數，曾作一度考察。他

同一教員前後所定的分數之比較

拿一些試卷給一位教員評閱，將分數記在另一紙上。隔若干日，俟此教員盡忘前次所定的分數後，再請他評閱原卷。於是對照二次的結果，發現許多不同的地方：分數大有出入，或高或低，而平均差爲1.75。

道格拉斯之研究 道格拉斯(Douglas)曾作一類似的實驗。他請四位中學歷史教員校閱二十八本美國史試卷。卷上祇標號碼，不書學生姓名；評定分數後，即抄在另一紙上，立刻擦去卷上的分數痕跡。四個月以後，仍請他們再加評閱。此四位教員的二次分數之相關度爲.77，.74，.82，及.76。有一教員第一次定某卷的分數爲64，而第二次却定爲89，竟差到25分。此四位教員之二次分數的平均差爲5.7，6.4，5.2及5.8。

問題的價值之研究

殷格利士之研究 上述的調查都着眼於分數上，殷格利士(Inglis)則自問題之價值方面入手，研究這個問題。他把十問幾何的試題，請許多中學數學教員評定各題應得的分數，而使總數爲一百分。在一二二位教員中，有二十人

第十表

教員估計考試問題的價值之差異

所定的價值	給各問題的價值之教員數										總數	百分比
	1	2	3	4	5	6	7	8	9	10		
0						1					1	0.1
1					1						1	0.1
2		1									4	0.4
3	1		1					1		1	2	0.2
4		3						1		1	5	0.5
5	14	15	7	1	16	2	8	6	8	19	96	8.8
6	4	7	1	1	6		2	1		13	35	3.4
7	6	3	1	3	6	4	6	2	4	8	43	4.2
8	23	26	15	14	29	6	4	8	7	8	160	15.7
8.5		1					1	1	1		4	0.4
9	3	5	5	4	4		3	3	2	2	31	3.0
10	34	30	44	24	27	34	31	24	16	18	284	27.6
11	2	3	4	6	1	3	2	2	4	1	28	2.8
11.5					1	1	1	1	1		4	0.4
12	8	6	16	23	6	21	16	24	21	10	151	14.8
12.5		1				1	2	2	2		8	0.8
13	3		2	2	2	3	5	2	2	1	22	2.2
14	1			1		3	1	1	1		8	0.8
15	3	1	5	16	4	18	9	21	26	7	110	10.9
16				2		2		1	3	1	9	0.9
17						1					1	0.1
18			1	1			1	1	2	1	7	0.7
19												0.0
20				3		2			1	1	7	0.7
25				1							1	0.1
全距離	2	13	15	20	14	20	13	6	18	18	25	
平均數	8.9	8.3	10.	11.1	8.5	11.5	10.1	11.1	11.6	8.6	10.	
平均差	2.0	1.9	1.6	2.5	2.0	2.3	2.3	2.5	2.8	2.8	2.4	

考試新論

四八

贊成每問各給十分。其餘一〇二人的分數之分配詳第十表。

我們可以看出最常用的是5，8，10，12與15。價值之距離很大，最大的
是20，最小的為12。

後來他用此十題試驗一班中學生，得各題之比較的價值如下：

總分	試題
187	1
191	2
135	3
102	4
260	5
8	6
20	7
15	8
0	9
298	10

此十題中，以第九題為最難，第十題最易。第四五兩題之難度，相差不過一倍，但有二教員的分數，却成一與二十五之比。各教員的意見，都與實際的價值不符。

考
試
新
論

第四章 舊式考試之改進

舊式考試係指普通的論文式考試而言，例如『試述某事』、『討論某問題』、『批評某方法』一類的問題。近一二十年來，許多教育專家曾努力證明此種考試之弱點及標準測驗之優點，因此有些人以爲它已經不合時宜，並無任何實際的價值，而主張廢除之，代以標準測驗。於是它的特長遂被忽視了。此種態度乃一誤解，實屬不當！我們祇須稍加思索，卽知標準測驗之發展，並未淘汰非標準的考試與舊式考試之需要。教員常需測量地方性質的或特別興趣的材料，常需測量工作之小單位。標準測驗對於這些事之測量皆不相宜。所以一完全的測驗程序中，應三者俱備，不能單用一種，因爲它們不但併行不悖，且互相補充。換言之，我們不是要決定何者較優而單獨採用之，乃是決定對於某種情形，何者最相宜且最有價值，於是採用之的問題。

舊式考試在教育的方法中，已有悠久的歷史。就現在的教員而論，他們

更熟習此種測驗，並且現時大部分採用的，仍爲此種考試，這是毫無疑義的。雖然目前關於教育測驗的著作很多，教員却常不留意，還有許多人不知其目標、限制、及施行法。故強勉教員採用它，實屬不智。舊式論文考試雖有種種缺點，但若加以改善，仍可利用之。本章之目的，就是討論它的優劣點及其改進法。

舊式考試之優點

論文式考試是作文能力之一有效的測量。測量筆述的發表力之惟一方法，就是叫學生作一篇論文。關於此事，別無他法能替代舊式考試之地位。就文章之性質而言，作文實不易客觀的測量，因爲優良的文章之原素異常複雜，包括措辭、用字、文法、章節、文格、標點、書法、等等。所以測量作文能力之適當性，實爲此種考試之一大優點。

此種考試之另一優點，是它比較他種考試更易編造與施行，及其適於任何地方的或普通的教材之特性。事之簡單莫過於編造舊式考試，因爲它所包

含的問題甚少，普通約五題至十題，教員祇須在短促的時間內，檢閱已授的教材，即可擬成之。但因編造之容易，常誘惑教員少用時間與精力，遂致結果不佳。此種考試之施行，亦極便易。它的問題可書於黑板上，或用口述法，而無須印刷或其他手續（但印刷或油印亦可）。印刷或油印爲一實際的障礙，許多學校，尤其小的學校，無油印機或印刷的經費。

舊式考試對於高級學生有特殊的價值。它供給學生以表現推理、創造、組織、解釋、分析、判斷、機智、等等思想歷程之機會。他種測驗之這種價值則不如此顯著，因其僅測量獲得的知識或事實，且常用很不連屬的形式，祇略爲涉及各種思想歷程。此種測驗則常予學生以相當程度的自由選擇其答案形式之機會，而不予以限制。對於成熟的學生，此法頗有優點，不過對於小學生的價值，尙屬疑問。許多問題不能算是絕對錯或完全對，必須加以辨別與判斷，舊式考試適給以此種機會。它聽憑學生解釋其問題，並依其能力的程度作答，故很公平。

舊式考試避免機遇、猜度、與反面暗示之影響。一毫無學科知識的學生，對於新式考試，全憑猜度或反面暗示，亦可答對若干問。論文式考試則無此弊。

舊式考試之限制

以上所述的，都是舊式考試之優點，但是此種考試亦有許多弱點。大概我們對於它的限制皆已熟知，茲特簡述如下：

論文式考試之首一劣點，就是它缺乏客觀性。試卷之優劣不能決定分數。考試的分數，乃是校閱者主觀的判斷之結果，故極不公允。此事已經許多實驗的調查證明了。斯達赤與依理特之研究爲此種實驗中之最著名者。

論文式考試之另一顯著的弱點，是它的問題之有限的範圍。它的範圍雖與所測量的教材之範圍同樣大，而測驗本身則比較有限。普通的論文式考試，多半是五題至十題。此數實在太少，不足爲學生的真能力之精密的且可靠的取樣。取樣既不優良，又很簡單，因此教員常在考試前數分鐘草草擬就。

對於擬題既如此不加思考，則問題易爲不重要者，甚或在學科之目的以外，自不待言。因爲此種考試之有限的範圍，遂造成一極不公平的結果。某生平日的成績雖不佳，但是幸而記得教材之某部分，恰巧教員選作此次考試的問題適爲該部分，於是獲得優良的分數。反之，一成績頗佳的學生，適不幸而忽略該部，遂致不及格。憑藉此種分數，以定他們的學業之高下，是否可靠？問題之選擇一不得當或組織不佳，每致意義含糊或生誤解，學生常不知所詢者爲何事，或不知答至何種程度，方稱完備。教員解釋學生之冗長的答案，比較學生領悟題旨爲尤難，以至校閱試卷，常生錯誤。

舊式考試缺乏特效，其成績常受無關的瑣碎原素之影響。例如在男女同學的學校中，教員對於男生的分數常嚴，而對女生的分數則寬。又如學生平日的態度，也很影響他的考試分數。此種原素可分下列三種：

技術的原素 此條包括不良的國文、書法、清潔、等等。文章之優劣，或致增減分數。因此這種考試不能測量發能力不佳的學生，因其難於組

織其思想，而以簡明的文字發表之，其所知者或多於其已答者。文章之長短，有時也是一重要原素；於是書寫的速率，發生很大的影響。能力相同而書寫速率互異的學生，成績遂不同。且因大部分時間用在書寫上，致少思考的工夫。

掩飾的原素 此條包括進步、努力、能力、在外工作、等等。例如二人的答案相同，教員以爲甲生異常用功而加其分數，乙生極其怠惰而減其分數。又如對於在外作工的學生則寬，反之則嚴。

私人的原素 此條包括學生之態度，如謙恭、阿諛等，及父兄之地位的影響。例如富貴子弟則多給分數，常提質問的學生則少給分數，等等。

舊式考試阻止系統的與有價值的溫習。因爲此種測驗僅涉及全範圍中之小部分，學生每於考試之前，猜度教員將出的問題，而溫習該部分，忽略同等重要的他方面。

此種考試尚有一缺點，就是學生對它所取之不快的反感。在考試之前，

他們焦急；考試將到時，他們不安；而考試時，他們處於一身心疲乏的狀態。學生用很多的時間在書寫及計畫答案形式上。此種時間及教員閱卷之額外數小時，將此種考試之編造與施行之便易完全抵銷。

論文式考試之性質常使學生覺其結果不公平。他們不十分了解其不可靠性，却把他們的分數互相比較，見同樣的答案得不相同的分數，故以爲不可靠且不公平。如果他們質問教員，亦得不着滿意的解釋，因爲實無可解釋。它造成學生對於考試之一不信仰的態度。此外論文式考試還鼓勵作弊。藉作文能力，學生能掩飾其無知。它也鼓勵強記。因爲問題的數目有限，且僅涉及少數普通的題目，學生願意碰碰機遇，猜度題目，而從事強記。此種讀書法毫無永久的價值。

若視舊式考試之結果爲代表知識或能力之一絕對的量數，其用途則很有限。七十五分視作學生知道四分之三的完全知識。這是假定各人的絕對知識是各自獨立的。零分爲毫無知識或能力，百分爲完全的知識或能力。究竟何

爲完全的知識，何爲及格的知識，學生應知若干，等等，純爲教員武斷的事務。學生所得的分數，乃視給分者之寬嚴與考試之難易而定，並非根據學生的知識。如果考試容易或給分者寬鬆，學生卽及格，且得高分數；反之，卽不及格或得低分數。分數至多表示學生在此考試上所作的，並不表示他們在零分知識與完全知識間之地位。因此種考試的分數難於解釋，故其結果無大意義或用途。

舊式考試之改良法

依以上所述的，我們編造舊式考試時，必須竭力避免其弱點，而善用其特長，方收效果。茲將編造此種考試應注意之點，分述如下：

編造此種考試之首一應考慮的事，就是決定其目的，如編造標準測驗然。各問題均應自目的之見地觀之。我們應決定欲測量何種能力、何種推理歷程、或關於某科之了解程度。範圍既經決定，然後可依照它草擬問題。故考試之內容，必須符合教育之目標。教員編造論文式考試時，應參照學校所編

的最低限度大綱、各種課本、專家的課程研究、標準測驗與量表的問題、教學時聚集的問題、等等。

考試的問題，必須適合被試的程度，其難度應以全班的能力範圍為標準，而在學生的能力之內。問題不可過於廣泛，使學生不知其何時方達到教員所希望的程度；但亦不宜過狹，致用『是』或『非』即可答覆。一切問題須使各人皆作同樣的解釋。如屬思想問題，更應提醒學生，其答案須表現審慎的思考。

至於問題的性質，不可限於一律，致太單調。問題之開始，如用下列名詞，大可引起並表現學生之思想歷程。

分類……此法要求根據某共同的特性，將事實集合、排列、分配、與組織。
比較……此法要求偵察事實之同異。

批評……此法要求優良的判斷與一題目之慎重的分析。

界說……此法要求決定一題目之一定的界限或限度，而予以明確的定義。

敘述……此法要求選擇與描寫一題目之特有的性質。

討論……此法要求對於一問題之贊否兩方面，予以精密的考察，而援引贊助一方面的理由。

說明……此法要求辨明一問題之含糊點，使之清晰。

詮釋……此法要求說明所討論的問題之例證。

解釋……此法要求就個人的信仰或判斷，引出一問題之真意。

申辯……此法要求表明一事是合理的或確實的。

略述……此法要求提出一討論、一理由、或一歷程之要點。

評述……此法要求慎重的閱讀一問題，而作批評的考察。

總括……此法要求用簡明的形式表出一問題。

追溯……此法要求詳述一問題之發展與進步經過。

證實……此法要求證明一事為真確的之證據。

應用……此法要求應用原則於新情境。

上表可再擴充而包括原因、結果、何故、諸類的問題。如果此種問題曾經慎重編造且不超出學生的能力，將有許多高貴的價值。它們說明舊式考試對於高級學生之用途。

考試之長度應依教材的範圍、用作考試的時間、學生的能力與經驗、等等而定，但須使全班均忙碌。

此種考試應指導並鼓勵學生在作答之前，至少閱讀問題二次，而多用時間思索及組織材料，以養成思想的習慣。有時此種初步組織可於心中行之；有時可作筆記大綱或用他法，以爲最後的答案之根據。不過對於這方面用太多的時間，亦屬可慮。究竟需用若干時間於此方面，向無定例，大概可用全部考試時間的百分之二十至二十五。

做法說明爲標準測驗之一要素。此種說明減少測量之變誤，故於舊式考試有同樣的影響與助益。因此舊式考試之前，亦須有做法說明，指明下列各事：

(甲)時限 說明時限，則學生可支配全部時間於各問題，不致用太多的時間於首數問，而無餘時答覆最後數問。定奪考試之時限，應以考試之長度與可以用作考試之時間爲準。

(乙)答法 依次序回答或先答易者。

(丙)所希望的答案長度或完備程度 對於普通或廣泛性質的問題，如『討論』或『詳述』之類，尤須如此說明。

(丁)其他 例如文法、書法、等等。若採用這些標準，則應說明分數之某比例是以它們爲根據。

此種考試之施行，可用黑板法或口述法，如能印刷或油印則更佳。但施行時，須於時限達到前，下一警告，說：『時間快到了，尙餘幾分鐘！』此種警告的理由是予學生以在餘時結束或完成其討論及書出其計劃中之最重要部分的機會，而不使其答案成爲不完備的形式。至於此種警告應在時限前若干分鐘，則隨測驗與時限之長度而定。最好仍餘十分之一的時間，卽下警告

，例如五十分鐘的考試，則在五分鐘前，餘類推。

此種考試應儘量減少校閱與記分的困難。最好每問祇許一個對的答案，則校閱時不致意見紛歧。除內容外（即思想），其他無關的原素，如衣履清潔、態度、年齡、性別、等等，均不應包括在分數之內。校閱此種考試時，宜先書出答案標準；那種答案給予分數，那種不給分數。然後參照此種模範，比較學生的試卷。校閱時，最好逐問校閱，先閱讀各試卷之第一問答案，再閱次問之各答，因為每一試卷常引起新的校閱問題，若不如此，閱卷者之意見易變更，致分數發生差異。

已經校閱的試卷應早發回學生，因為過若干日後，學生的印象模糊，已忘記關於此考試之某些事實——如為何他們如此答法——遂致喪失其興趣。若在一二日後就發回試卷，教員和學生詳細討論之，可矯正許多錯誤的觀念及教授種種不了解的問題。發回試卷時，教員應解釋各人在全班中之地位，例如宣布班中之最高的分數、平均的分數、及最低的分數。再或將全體分數

的分配寫在黑版上，說明若干人得甲、乙、丙、丁、各等。但是無論如何宣布全班的分數時，除提出得最高的分數或在某種分數以上之姓名外，總以不表明個人的分數為宜。

第五章 新法考試

標準測驗之流行，造成所謂新法考試之迅速的發展。『新法』一詞是用以區別舊式論文考試的。此種考試是一教員爲適應其特殊需要而編造的非標準測驗。換言之，即教員採用編造標準測驗之技術所編的普通考試。此種測驗一方面利用標準測驗之某些價值，而他方面同時又保留舊式考試之許多優點。其實，它是標準測驗之一演進。學生受新法考試時，書寫甚少，僅需劃橫、劃圈、填字、或作正負號，不必作長篇巨論。因此它又稱爲『短答』測驗。如果它的問題均經審慎編造，使各題均祇一個對的答案，亦名『客觀』考試。現在上至大學，下至小學，都常採用此種新法考試。幾乎一切課程均適用之。任何教員皆可依本書之指導，編造新法測驗。

對於一種考試所希望的价值，爲決定所欲採用的考試之種類的一重要考慮。教員若採用新法考試，他們應確知其未犧牲舊式考試所具的任何重要價

值。盧赤歸納考試最重要的價值如下：（一）引起動機，（二）測量，及（三）訓練文字之發表力。他比較新舊二種考試，表示（一）它們對於刺激兒童的工作有相等的價值，（二）作爲一嚴格的測量法，新式優於舊式，較爲精確且更有效率。最後，關於文字表述的訓練，的確爲教學之一大目的。最好偶或舉行此種能力之考試，而不應專用新法考試。在他方面，教員須把何爲某種考試之目標，明白存在心中。除非是專爲測量上述的能力，採用新法考試較爲有利。換言之，一時祇作一事；或測驗事實之知識，或度量組織與發表事實之能力，但不可混爲一談。

關於新法考試之價值與限制，已於前章討論舊式考試之優劣點時，間接涉及。茲再分論如後：

新法考試之價值

新法考試之價值甚多，其首一優點就是內容廣博。如果一種考試能用數百問題測量被試之某科知識，則有澈底包括一課程中之材料的較大可能性。

新法考試之範圍很廣大，能包涵比舊式考試更多的材料，約五倍至十倍於舊式者。且因學生少需書寫，能多用時間思考，故可答覆較多的問題。問題一多，則校閱時縱有一二錯誤，亦不致影響最後的分數。更因此種考試富於彈性，可適應隨時的需要及任何科目或教材，如地方知識、特種教學、等等。

新法考試之次一特點就是客觀。此點前已討論，各種實驗均反覆證明舊式考試之主觀，因此不可靠。但是新法考試的問題之答案的正誤，皆有一定的標準：是則是，非則非；何者給分，何者不給分，都已明白規定。閱卷者祇須用製就的答案標準與學生的試卷比較。故對同一試卷，無論何人校閱之，皆得相同的結果，絕不容閱卷者參加個人私見。在被試方面，則強迫其答案簡明且精確，使他們無法作弊。在舊式考試中，他們如果不能回答一問題，可假作誤解題義，給以似是而非的含糊答案，以逃過難關。在新法考試中，則知之為知之，不知為不知，絕不能掩飾其無知。

至於其他一切無關係的瑣碎原素，如作文能力、書寫速度、清潔、品性

、等等，亦均淘汰殆盡，對於分數不生影響。若不如此，則對發表能力不佳的學生極不公平，因為他們難於組織其思想，而以明白的文字表述之。他們所知的每多於其答案中所表述的。於是成績中雜入該科知識以外的能力在內。在另一方面，書寫一多，則能力雖同而書寫速率互異的學生，成績竟不相同。此乃測量書寫的能力，而非教材知識。

時間與精力之經濟為新法考試之又一優點。在舊式考試中，書寫動作與修飾文字等等耗時太多，約佔全部考試時間之一半或四分之三。而新法考試則因約制反應，在很少的時間內，能答很多的問題。時間既短，書寫又少，被試自然不易疲勞。校閱時，閱卷者只須注意一字、一數、或一符號，約比校閱舊式者快二倍至五倍，數分鐘能閱完數百題。倘使學生誠實可靠，更可命之在教員監督下交換閱卷，則節省時間與精力尤多。

新法考試之第四種優點，是其教育的價值。此種測驗鼓勵系統的溫習。因為舊式問題僅涉及一科目的巨數問題中之極小部分，學生每猜度教員欲出

的問題之範圍而溫習之，遂忽略其他同等重要部分。有些人甚至不溫習。至於新法考試則非全部溫習不可，因其範圍廣，不限於某一部分。

新法考試很符合真實生活情境。生活中常有要求『是』與『非』、或此『真』彼『僞』、或此乃少數可能中之『最佳者』的情境，絕少不准參攷書籍，全憑記憶作長篇論文之事實。

若任學生交換校閱其試卷，大增他們的興趣與教益。閱卷時或閱卷後，可以討論共同的錯誤，加以解釋，則學生立刻知道何處他們必須再加研究。

第五，新法考試之難度有一定的標準，可以互相對照。如果保留諸連續班級之記錄，加以整理與統計，可產生常模。於是教員可用以比較其逐年逐班的學生及其本人之教學成績。

最後，新法考試常受學生歡迎。它鼓勵學生之興趣，不像論文式考試之乏味。因為它的明確性，使學生甚為滿意。學生明白考試所要求的，並且知道何時已經達到此種要求，不似他們在舊式考試中之毫無把握。此外他們更

信仰此種考試的分數之正確與公平，教員無表示偏心或私見之機會。因為這種考試極為客觀，學生比較同樣的答案，皆得相同的分數，遂毫不懷疑；反之，學生比較舊式考試的試卷，發現同樣答案得不同的分數，故以為不可靠。

新法考試之限制

新法考試雖有上述種種優點，但亦有它的弱點。有人批評新法考試祇要求記憶，而不測量推理能力。此種批評至少是根據於二種假定的。其一是思想與記憶是顯然的獨立的心理機能；另一種是問題本身之形式足以引起推理。要想在測驗中明確的分別思想與記憶，乃是現代心理學的知識所辦不到的。吳德說：「事實在心中並不孤獨的存在。我們藉思想而記憶，我們亦藉記憶事實以思想……現在的各種實驗研究，皆證明一範圍內的知識之測量，與思想該範圍內的材料之智慧或能力有一很高的關係。」進而言之，推理並非生而繫於考試的問題之形式，它更倚恃被試的反應與已往的經驗。問題之

爲思想的或事實的，視學生的反應之準備而定。新法考試或需很多的推理，論文式考試或要求很少的推理。同一問題，對於此人爲思想問題，而對於另一人或爲事實問題；對於某人初次爲思想問題，二次則爲記憶問題。

潘德生(Paterson)曾作一實驗，證明舊式考試與新法測驗皆測量同樣的心理機能。他給一組學生以一包括同樣教材之二舊式考試，發現二考試的結果之範圍一致，相關係數爲.52。然後他又給該組學生以同樣教材之新法測驗，發現結果之一致的程度與論文式考試相一致的程度同，相關係數也是.52。因爲舊式考試的結果和新法考試的結果相同，顯然二者測量同一事物，無論其爲思想或知識。他又發現新法考試之結果，彼此的一致性優於二論文式考試相互的一致性，相關係數爲.67。故除二者測量同樣的事物外，新法的測量且優於舊式者。

又有人反對新法考試的問題測量孤獨的事實，而非課程之連貫的知識。然而此事不確。它可將化學中之一複雜的方法或歷史中某一事變分成若干重

要部分，而用新法問題測驗之。根據某人對某事的要點之知識的取樣，我們就知道他究竟明白全部歷程與否，而不必要求作長篇論文。

不過此種考試使學生無發揮其思想之機會，到是一種大缺憾，因為除了作符號或填字外，別無所事事。受此種測驗時，學生喪失解釋或說明及總括材料之經驗，然而這些能力都是很重要的。再則對於糾紛的問題，如哲學問題，此法不甚適用。

新法考試之編造及施行，自較舊式的更為繁難。故被試的人數如在五十以下，採用此法，實不經濟。但自經濟的觀點而言，校閱時所節省的時間與精力，大可與之相抵銷。並且新式問題雖難編擬，但如擬有充分的數目，却可逐年逐月反覆應用。練習和記憶以前的測驗內容，對於結果之影響，不如教員想像的那樣重大。倘若試卷不落在學生手中，大可避免想像的危險。至於需要印刷或油印，自是一實際的限制，然而多數新法測驗，皆可用口述法或黑版法施行之。

新法測驗含有猜度原素或機遇成功及反面暗示之影響。關於此點，將於另章中詳論之。

在新法考試中，學生易於作弊，因為答案祇需一字或一符號，更易竊視及抄襲。其實不然，我們可設法避免作弊的機會。凡屬鄰座的學生，皆給以不同的測驗式，再或雖予以相同的測驗，却不註明頁數，則他們很難抄襲。在舊式考試中，偵查作弊，殊非易事，因為他們偷看以後，祇須改頭換面，稍易文字，即偵查不出。而在新法考試中，則易查出。因為學生對於錯的答案與對於對的答案，將同樣的抄襲，因其不辨正誤。於是遇有可疑的試卷，可與被抄的試卷對比，計算其相同的錯誤數。若相同之數，多於機遇數以上，即可信其抄襲。

柏爾德(Bird)曾提出一偵查客觀測驗中作弊的方法之證例。他給一心理學班二六六學生以二小時的新法測驗，共計有一四九個問題。全班中似有四人偷看鄰座的試卷。經一小時後，曾警告一學生，且命其掉換座位。測驗後

，將此四人的試卷及其鄰座的試卷皆標明。於是計算其結果，相同的錯誤數爲十七、二五、二八、與三一。曾被警告的學生，第一小時有十五個相同的錯誤，而第二小時却僅二個。顯然在第一小時，他抄襲鄰座的。柏爾德又用隨機取樣法拿二十學生的試卷，與其鄰座的比較，平均相同的錯誤數爲四。此卽證明這四人作弊。

新法考試之種類及其舉例

新法考試，大別之，可分爲二類：（一）追憶類，及（二）認識類。所謂追憶類，是指被試供給其本人的答案一類的問題而言。答復此種問題時，學生必須回想已習的材料，然後才能作答。此類之最常用者爲一字答與填字測驗。

認識類則皆有二個以上的答案，被試只須識出何爲對的答案，何爲錯的答案。此類之較普通者爲真偽、選答、對偶類比、重組、指認、諸式。

茲將各式分別舉例如下：

(一) 追憶類：

(1) 一字答測驗

一九一九年世界大戰和約簽字於

硝酸的符號是

供求律大都應用於

凡爾賽

HNO_3

物價

(2) 填字測驗

$$(a+b)^2 = a^2 + 2ab + b^2$$

英語中字的次序是 主詞 在前， 賓詞 在後。

國民政府之下，設立法、司法、行政、考試與監察五院。

(二) 認識類：

(1) 真偽測驗

……心跳的常態速率是七十。

……凡屬人類均有社會本能。

.....命分之上部分叫做分母。

(2) 選答測驗

(二答) 中國最大的口岸是 (1) 天津 (2) 上海..... ()

(三答) 馬爾薩斯研究 (1) 失業問題 (2) 人口問題 (3) 婚姻

問題..... ()

(四答) 助燃燒的氣體是 (1) 淡氣 (2) 炭氣 (3) 輕氣 (4) 養

氣..... ()

(五答) 人類與其他胎生動物最相似的時期為 (1) 成人期 (2) 嬰

孩期 (3) 幼年期 (4) 老年期 (5) 胚胎期..... ()

(七答) $5X-4=16$, X等於 (1) 20 (2) 9 (3) 12 (4) 4 (5) 1

(6) $\frac{12}{5}$ (7) $\frac{16}{5}$ ()

(3) 辨別測驗

_____金 1 A

- | | |
|-------|-------|
| ___ 鈣 | 2 Au |
| ___ 鈉 | 3 C |
| ___ 銅 | 4 Cu |
| ___ 錒 | 5 Fe |
| ___ 銻 | 6 Hg |
| ___ 鋅 | 7 I |
| ___ 錫 | 8 Na |
| ___ 鐵 | 9 P |
| ___ 硫 | 10 S |
| ___ 碘 | 11 Sb |
| ___ 磷 | 12 Ti |

(4) 類比測驗

畫之於夜像之黑於 白

道教：老子：耶教：耶穌

汽車：馬車：機器腳踏車：（四輪車，火車，腳踏車）

X：3X：：6：（99, 20, 27, 12, 18, 30, 6）

(5) 側線與鑿

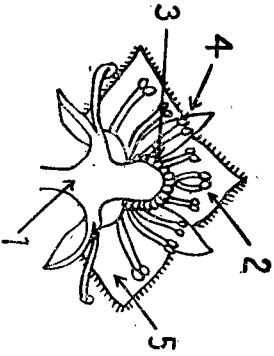
依城之大小定下列各城之次第，最大的是1，次大的為2，餘類推。

揚州 無錫 蘇州

鎮江 上海 通州

江陰 南京 常州

(6) 距臨與鑿



-蕊
-瓣
-心皮
-萼
-花房

(7) 訂譯測驗

中國大宗出產是牛油

$17 \times 7 = 109$

紅血輪之最大機能是殺菌

考
試
新
論

第六章 新法考試之編造

如欲新法考試有價值，編造時必須異常慎重。沒有受過相當訓練的教員，不應冒昧的從事於此種考試之編造。許多人的失敗就因為這個原故。這種教員應該一時研究一式，自最簡單且易編造的着手。經過相當的練習後，再試編第二種，於是第三種第四種。直至一二年，一切測驗式幾均用過，則經驗豐富，自可運用自如。其實，雖屬曾經研究測驗編造法的教員，也應如此。

在未分別討論各種新法考試之前，我們似應先考慮應用於一切測驗或多數測驗之普通原則。這些原則，有些已於前面各章間接提及，茲再詳述之。

新法考試之編造，約分編造、施行、校閱、及處理結果四大步驟，其詳情如後：

新法考試之編造

新法考試之編造步驟，又可分爲下列數種程序：

(一) 決定目的 測驗之目的是編造時的指南針。目的既定，則以後的工作有所根據，故把它列在第一。所以編造測驗時，首須定奪它是測量難度、準度、或速度。

(二) 擬題 考試的問題宜於舉行考試之前數日擬就。擬定之後，可放在一旁，經若干日後，再作第二次的考慮。於是以前的印象已淡，可顯出許多弱點，以便加以修正。擬題時，如能得同事之協作或批評尤佳，則可選二人同意者爲考試的問題。

爲免除臨時慌忙起見，教員應隨時留意收集優良的問題，例如本人用過的、他人或他地用過的、研究某題時心中想到的、學生提出的、等等。遇有這類問題時，即應錄下保留，放在妥當的地方，使不致遺失，而於應用時，能在短時間內集合一考試所需的問題。經濟的方法是採用卡片制。每一問題書一卡片。此法遠勝於連續的寫在紙上，因爲它們不但可以隨意分類、排列

、移動、增減、或棄去，而無須重寫問題，且便於查閱或備日後之用。關於已經用過的問題，並須註明其適用的程度。卡片的式樣，最好用三吋長五吋寬的圖書館用者，形狀如下：

史——15

辛亥革命起義於武昌

所擬的問題須能代表或包括各要點，淘汰一切瑣碎的問題，例如偶然的問題。而問題的性質與形式應屬多方面的，舉凡測量事實、原則、關係、應用、解釋、等等問題，與真偽、選答、填字、諸式，都須應有盡有。這樣不

但包括週到，且不單調，並增興趣。

除國文測驗外，一切問題均應沒有文字的困難，因為若不如此，則所得的結果為語言能力之表示，而非測量該科的知識。造句務求簡明，不宜冗長或太複雜，致意義含糊，必使各人皆同樣了解。遇不易明瞭處，能附圖解更佳。兩可的問題不可用；誤引的問題不可用；雙關的問題不可用。用字須妥當而淺易，適合被試的程度。選擇日常使用的字或詞代替專門或典雅的字或詞。不常見的字或詞不可用；太專門的字或詞不可用；非本科的專門術語尤不可用。

所有的問題均應遵守優良的文字表述、文法、標點、等等規則。例如別字、俗字、或俗語都不可用。

語義須屬正面，不宜用反面語氣，尤不可用雙反面，致為被試所忽視，或增其閱讀的困難。一問之中不可含有二義。

各問題須自成一個獨立的單位，不可倚恃他問以知其意義。故同一測驗中

，不可有一問答復他問，或暗示他問之答案。

對的答案與錯的答案皆不可過於顯明，致無教材知識的人都能辨出。

如欲測量學生之全部能力，可於下列二法中，任擇其一。第一種方法是在考試內加入一二很易的問題與一二很難的問題，使無零分與完全分數。如此則最劣與最優的學生均被測量。第二法是定一時限，使劣等生能作數題，而優等生不能答完。

所擬的題數應多於實際需要者，約較最後題數多百分之二十五至五十，以便選擇其中最佳者，及平衡各着重點。一旦遇有不適用者，也可拿其餘的補充。問題一多，則一二無心的錯誤，對於最後的結果，影響亦不甚大，可避免機遇成功。此外尚有一優點，即編造替換式。

(三) 排列問題 擬就問題後，次步就是排列諸題。問題之排列，務求經濟，以能節省施行、校閱、及學生受試的時間與精力為最佳。測驗的問題普通都是橫排，但標誌答案的地方與方法，則因人而異。有人把答案的地位

對的答案發現的次序，應當或前或後，切不可按一定的規律排列。最好依機遇法排列之，例如編真偽測驗時，則以擲銅元法決定之。其法是以銅元正面代表真，反面代表偽，依擲出的正反面，定該題之應真偽。又如編一四答的選答測驗，即以拈鬮法決定之。書若干一、二、三、四紙條置於帽中，抽出何數，則把對的答案排在某處。抽出之數爲四，就列在第四。

至於全部問題之排列法，則有三種主張：(甲)依教材之次序，(乙)依價值之輕重，及(丙)依難易的程度。第一法的缺點就是前面的問題常暗示後面問題之答案。第二法是把最有價值的問題排列在前，予各人以答復的機會。但是最有價值的問題或爲最難的，若排列在前，則費時太多，致礙他題的答復。第三法是最易的問題最前，最難的最後。依據經驗，測驗開始的數題應易，則被試不致心慌意亂，可以順應測驗動境。普通多用此法。

(四)辨別難度 如按難度排列問題，次步自然是辨別難度。評定客觀的難度很不容易，編造新法考試時，不妨採用簡便的辦法。其法如下：

依五級制或十級制，根據個人的意見或數位教員之共同意見，

(甲)定一切很容易致各人幾皆答對的問題爲「1」。

(乙)定一切很繁難致各人幾皆答錯的問題爲「5」或「10」。

(丙)定二極端間的問題之等級爲「2」至「4」或「2」至「9」。

(丁)照所定的次序排列各問題。

(五)化分替換式 如果問題甚多，可把它們分成數替換式。所謂替換式，就是難度約略相等的數式測驗。這種測驗的優點甚多。第一，它們可避練習的影響，因爲每次皆施行同樣測驗；被試記得其中的某些問題，雖無進步，第二次的分數亦必高於第一次的。第二，因爲難度相等，故各次的結果均可比較。第三，遇考試時有學生缺席，可作補考之用。第四，可逐年輪流應用，而無須每次皆費時編造測驗。最後，同時使用二式，可以防止學生作弊。

至於化分的方法，通常採用隨機取樣法。例如，欲分二式，則將奇數題

分爲一式，偶數題又爲一式。倘需難度十分相等，可用下法：

甲式：

第一問

第四問

第五問

第八問

餘類推

乙式：

第二問

第三問

第六問

第七問

餘類推

此法較奇偶數題各爲一式法稍優，因爲免除難度之系統的差異；否則第一式必較易，第二式較難。

(六)編製說明 新法考試必須附有說明，指導學生如何做。說明之詳略，應依被試的經驗、年齡、和智力而異。對於一組未曾受過此種考試的學生，必須用一很詳細的說明，此後則可逐漸簡略。換言之，說明之詳略，可與施行考試的次數成正比例。就年齡與智力而論，如果被試年幼或智力低

，說明必須詳細。

說明之編製，以簡明為原則，文字務求淺易，使被試充分了解。各種說明均應指明下列各項；

(甲) 考試之種類，如填字測驗、等等。

(乙) 答的次序，或先易而後難，或依順序作答。最好的辦法是告訴學生先答一切確知的問題，並警告他們不可用太多時間於某一問題。

(丙) 答的方法與答的地方，如填字、填數、劃圈、等等，以及答在題前、題後、等等。最好附以示範或舉例。對於年幼的兒童，此種示範須有一二印就的，一二在教員指導下作答的，使他們確知做法。

(丁) 關於學生無把握的或猶豫的問題之答法，應有特殊的說明，例如答錯須扣分。根據近年的實驗結果，最好禁止猜度。

(戊)時限，並告訴學生應否趕快工作。

(己)宣布結果，如此則更增興趣，鼓勵競爭。

(七)製造答案標準

答案標準之製法，依問題排列之形式而異。如果

答案在一定的地方，且已排列成行，製法就很簡單。祇須取一空白試卷，將對的答案填入，裁成紙條，即成答案標準。第三圖表示應用此種標準紙校閱一頁選答測驗試卷之情形，旁邊的標記表明答錯者與未作者

試卷	標準
1.	2
2.	3
3.	4
4.	
5.	1
6.	3
7.	5
8.	1
9.	4
10.	1
11.	4

第三圖 答案標準之一頁

如欲它們耐久，則製法稍異。其手續如下：

(甲)取一厚紙，裁成長與試卷齊而寬約一吋之紙條若干。

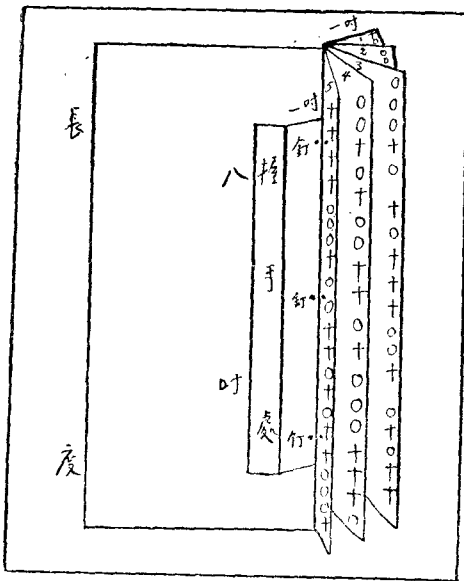
(乙)取一空白試卷，填就對的答案。

(丙)把裁好的紙條放在填就答案的試卷旁，比照答案的地位，將對的答案抄入。填完一頁填次頁。

(丁)每條上端書試卷之頁數。第一頁放在最後，餘依次排列。

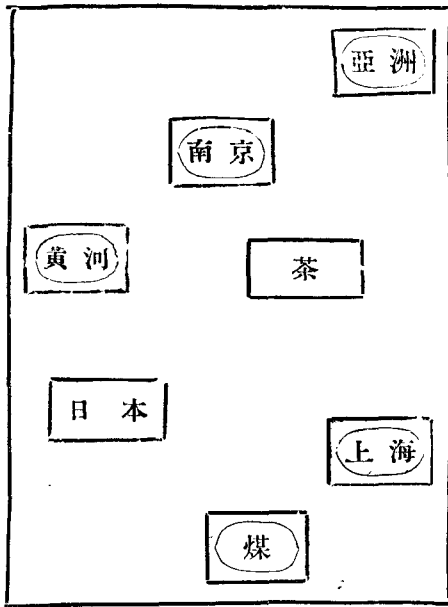
(戊)將諸頁釘在一處。

諸頁既已釘就，即成一完全的答案標準。第四圖表示一校閱真偽測驗之標準。

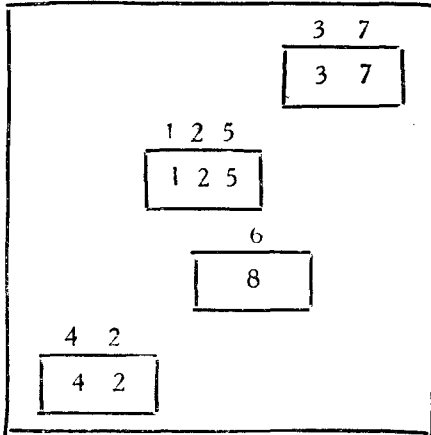


第四圖 真偽測驗之答案標準

倘使答案散在各處，則上法不適用，必須另製裁洞標準紙校閱之。此種標準之製法是把一張白紙或厚紙放在試卷上，在對的答案地方裁一洞，使下面的答案顯出。校閱時祇須計算對的題數。若是算術測驗，則將答案寫在洞的上邊。第五圖為二種裁洞標準紙舉例。甲圖是在對的答案上劃圈，故未劃圈的就是答錯的。乙圖為一校閱算術測驗之標準紙，答數不符者即錯。



甲



乙

紙準標洞裁 圖五第

(八)定閱卷規則 關於閱卷，我們祇須注意數點：

(甲)不給部分分數，對即對，錯即錯，無所謂半對或半錯。

(乙)每問給一分，填字測驗則每一空白給一分。

(丙)除真偽或二答測驗採用『對減錯』公式校正機遇成功外，其餘各式均以答對題數為分數。

新法考試之施行

此種測驗之施行應注意二點。一方面須使測驗的動境不變，則各次的結果皆可比較，而分數亦不致因動境的影響有所升降。他方面，此種考試必須易於施行。換言之，須使學生易於受試，教員易於舉行。施行之法，不外下列三種：

(一)口述法 此法是叫被試預備一張白紙(最好一律的大小)，寫好問題的號數，如1、2、3、等等。然後教員高聲朗誦一題，被試即在該問號數後作答。此法甚為簡便，但有數種缺點。因為有些人不能十分了解口述的

問題，必須自行閱讀才能完全明白。並且語句若長，被試尤難記憶。至於填字與對偶測驗，此法幾完全不適用。

(二) 黑板法 此法是把問題完全寫在黑板上，被試則照前法備一白紙，書就題數，然後依法作答。此法可免上述諸弊，但在實際上也有種種困難。例如問題佔地太多，黑板不敷；須於上課前寫好，課堂又恐不空；學生竊視；書寫費時，等等。

(三) 印刷法 此法是將問題印刷、油印、或複寫，使每人皆有一份。不過此法雖佳，却不甚經濟，有些學校沒有這種設備。

此種考試之施行，除速率測驗外，最好任各生作完，不必規定時限。萬一要定時限，應以學生的成熟程度與材料之多寡為標準。大概普通的考試，至多以一班上課的時間為限，不宜更長於此，致引起疲勞種種原素之影響。

至於計時，可用有秒針的錶。然最要者是予學生以按時起停之訓練。既經預備後，各生均舉手，面向教員。直至教員發出開始的命令，即立刻工作

。同樣的，教員發出停止的口號，即刻停做，舉起手來。

施行時，尚有一事應加注意；即防止學生作弊。就理論而言，新法考試不易作弊，因為此種問題作答非常迅速，少有竊視的機會；且因題數甚多，偷看很難準確。普通防弊法是予鄰座學生以替換式，問題既不相同，自無法抄襲。但是若無替換式，此法失效。救濟的方法是嚴厲監督可疑的學生及互換頁數。所謂互換頁數，就是變更其次序，例如一、二、三，二、三、一，或三、一、二，等等。

新法考試之校閱

此種考試之校閱，較舊式考試便利多了。校閱之法有二：（一）學生校閱，及（二）教員校閱。前法是由教員監督學生行之，可增學生的興趣，並能視作溫習。而其最大的優點却在指示正確的答案，並可取學生的意見備參攷，以便改良。如用此法，可於考試完畢時或考試之次日行之，則印像尤存，易收效果。校閱時，先命學生交換試卷，即甲閱乙卷，乙閱丙卷，等等。然後

教員逐問口述對的答案，詢問全班的意見，如有異議，則加討論。否則叫學生在錯的答案旁，作一記號。凡屬未答的，另作一號。無號者即爲答對的問題。最後求出做對與做錯的總題數，代入公式，即得校正的分數。採用此法時，教員似宜檢閱一過，以免錯誤。而計算分數亦宜由教員行之，恐學生不懂應用機遇公式之方法。

但是有些教員怕學生校閱發生弊端，如改答、不準確、等等，甯肯稍費時間，自行校閱。採用此法，則取製就的答案標準放在試卷旁，互相對比。凡屬答案不同的即錯，則標明之，未做者另用一號標出之。凡無號者即算答對。依前法求出校正的分數。

如爲填字測驗，可有數種不同的答案，則宜逐問校閱，以免意見參差，致結果不可靠。

校閱時，教員宜用顏色筆，如此則答案之正誤，一目了然，不致混亂。最好用紅鉛筆。

閱卷後，須將答案標準撕去或秘密保存，以免流入學生手中。如由學生校閱，即須收回試卷。

各卷既經閱畢，則求每卷之答對的、答錯的、與未答的題數，把它們記在卷面適當的地方。如果不用機遇公式，則取做對的題數為總分數；否則將各數代入公式中，求出應得的分數。此種分數通稱之為測驗分數(Test Score)。

結果之處理

處理測驗的結果時，每涉及教育統計，但是一般的教員都缺乏此種統計的知識。故為避免這種困難起見，此地僅討論極不可少而又極簡單的方法，其餘概置不論。

分數計算完竣後，第一步工作就是考慮如何將這些結果歸成一最易解釋的形式，俾能一瞥即知全班的一般狀況。何為最高的分數？何為最低的分數？何種分數足以代表全班？這就是求全部數量的工作。此種數量之求法，我們可在此地敘述二種：即順序分配與次數分配。未經分配的分數，都是亂雜

無章的，難於看出其中的意義，故首先把它們排成順序的分配。所謂順序分配，就是把所有的分數，按着大小的次序排列，小者在前，大者在後。例如

第十一表 順序分配表

135	116	97
133	114	96
130	112	96
126	109	93
121	108	86
120	107	79
120	106	76
120	104	
120	104	
119	104	
119	103	
116	102	

第十一表。一看此表，就知最大的分數是135分，最小的為76分。但是人數若多，統計亦感麻煩。最好列成次數分配，較為簡單。例如第十二表。此表第一行是分數之組距。因為全距離很長，故以 ∞ 為組距。定了組距後，就把各人的分數

，歸入各組距內。例如，183則歸入181—189組距內。餘類推。

上述二法都是表示全部數量的，此外尚須求一數目足以表示全體成績之大概情形，即統計學中所謂點量數。此種量數之求法甚多，現在我們祇討論

第十二表
編造之分配次數表示表

分數	次數	次數
181—189	I I I	3
172—180	I	1
163—171	I I	2
154—162	I I I I	4
145— 53		0
136—144	I I	2
127—135	I I I I	3
118—126	I I I I	2
109—117	I I I I	1
100—108	I I I I	3
91— 99	I I I I	1
82— 90	I I I I	2
73— 81	I I I I	1
64— 72	I	1
		$26 = N = \Sigma f$

平均數。所謂平均數，就是用次數的總數除分數的總數之商數。此種平均數為考試中之一新目標。在舊式考試中，教員的目標是得百分的人愈多愈佳，不及格的愈少愈佳。平均數則用作全班之代表。此班的平均數可與用同樣測驗的他班平均數比較。而教員的目的應是提高一班的平均數。

其求法如下：上表的分數總數

為3469，人數為26。用26除3469，得133.4，即平均數。但是人數太多，數目太大，此法頗感麻煩，另有簡法。其法如下：

第十三表 平均數求法

分數	次數	差數	差數×次數
181—189	3	48	144
172—180	1	40	40
163—171	2	32	64
154—162	4	24	96
145—153	0	16	0
136—144	2	8	16
127—135	3	0	360
118—126	2	-8	-16
109—117	1	-16	-16
100—108	3	-24	-72
91—99	1	-32	-32
82—90	2	-40	-80
73—81	1	-48	-48
64—72	1	-56	-56
			<u>-320</u>

$$\begin{array}{r}
 \text{人數} = 26 \\
 \text{假定平均數} = 131 \\
 \text{平均數} = 131 + 1.5 = 132.5
 \end{array}
 \qquad
 \begin{array}{r}
 360 \\
 -320 \\
 \hline
 40 \\
 \hline
 26 \\
 \hline
 = 1.5
 \end{array}$$

點，在假定數平均數下八點。餘類推。

(3) 用差數乘次數。例如 189 有 3，總數為 144。156 有 1，總數為 156。

(1) 取分配中任一級

距為假定平均數，例如 131，為

127—135 組距之

中點。

(2) 求假定平均數的

離中差。例如 13

6—144 組距的中

點，在假定平均

數上八點；118

—126 組距的中

餘類推。

(4) 正差數的總數為360，負差數的總數為-320。正負相消，餘數為40。用人數26除之，得1.5。

(5) $131 \times 1.5 = 196.5$ ，即真正平均數。

既已求得上述諸量數後，教員就可宣布測驗的結果，例如最高的分數、平均的分數、與最低的分數。但是宣布分數時，不可說出個人的分數。不過有時可宣布得最高分數的人名或在某分數以上的學生姓名。如將分數之全部分配書於黑板上，或繪成圖解，則尤明顯。

倘使教員欲知問題的價值與難度，以備日後的參考，他們可再作二種統計。茲先述決定問題優劣法如下：

- (1) 選出全班最優的二十學生與最劣的二十學生之試卷。
- (2) 分別計算這些學生答對各題的人數。
- (3) 求優組與劣組之差數。

此表中最好的問題，就是最易的問題，因為「是」最大的差數。第三問是適中的問題。第四問最劣，因為「優」組答對的人數反比劣組少。

至於求問題難度之方法，祇須計算全班答錯各題

第十五表 難度決定法

學生	問題號數													總	
	1	2	3	4	5	6	7	8	9	10	11	12	13		
A					X					X				0	3
B					X			X		X				0	5
C											X			1	
D									X	X				5	
E						0			0		X			5	
F			X		X		X		X		X			9	
G			X	X	X		X	X		X	X			9	
H				X	X					X	X			2	
總			1	1	5	1	1	2	3	3	7	4	6	34	

的人數。例如第十五表中，第一二兩題無人答錯，自然最易；第十一題竟有七人答錯，其難可知。

第十四表

表示決定測驗題的優劣之方法

問題號數	全班二十學生之對的答案數		差數
	最好學生之對的答案數	最劣學生之對的答案數	
1	18	11	+7
2	15	9	+6
3	12	12	0
4	8	10	-2
5	7	5	+2

此表還有一種用途，即其診斷的價值。它告訴我們何人作得壞及何人作得好與其好壞所在處。例如，學生C僅答錯一問題，學生F却有九錯。

考
試
新
論

第七章 追憶測驗

追憶測驗之命名，是由它所要求的反應而產生的。這種測驗，並無建議的答案，必須學生追憶該科之知識，而供給其本人的答案。且在追憶時，他應回想一切可能的答案，加以選擇。答復此種問題之成敗，全恃追憶及應用已經澈底習得的原理、方法、事實、等等之能量。因此，這種測驗頗宜於測量學生精通與組織材料之能力。舊式論文考試可列爲此法之一，不過它的答案太無限制。比較客觀的二種是：(1)一字答測驗，及(2)填字測驗。茲分述如後：

一字答測驗

一字答測驗是由若干問題或不完整的語句組織而成。它祇要求一字或多一詞作答。列如，我們若想測量學生對於晴雨表的知識，論文式的問題是：『何爲晴雨表？』一字答式則爲：

測量空氣壓力的器具叫做什麼？

或 測量空氣壓力的器具叫做……

此地僅有一個對的答案，即『晴雨表』。除了答案簡單外，此法頗類似論文式考試。學校教員常採用之。

優劣點 一字答測驗具有若干有價值的特性。它的首一重大優點，就是答案的明潔簡單之條件。答案之明確，必需學習正確且精密。因此它輕視含糊的概括，而着重於學習的透澈之測量。它鼓勵精確的追憶與應用已習得的材料的能力之發展。答案的限制既嚴，故有價值的答案甚少，校閱因之頗易。且因答案簡短，使能在比較短促的時間內，包括或測量巨數的教材。學生回答此類問題之速度，雖較認識式的爲慢，却比舊式考試的更快。它適於多種教材，並少猜度的危險。總之，它有舊式考試之優點，而避免其弱點。

然而一字答測驗不是沒有缺點的。它不及認識式那樣客觀。若自客觀性的立場言之，這種測驗不易編造，亦難校閱，常有多種公認的答案。因爲追

憶較認識爲難，且需更多的書寫，所以沒有認識式那樣受學生歡迎。且因需要追憶，每易促成學生強記或呆記所學者。

編造法 一字答測驗之編造法很簡單，然須異常慎重，使之極近客觀，方有最大的價值。每個問題均須如此組織，使之只有一個對的答案，而此答案又僅一字或一詞。編造此種測驗之第一步工作，是就所要測量的範圍內之各問題，各擬一對的答案。次步則把這些答案組成一字答式的問題。例如你想測量學生對於經濟初步中『價值』一詞的知識，原來的問題是『何爲價值？』而對的答案爲『價值是用以表示二商品自由交換之比率。』現在可用此答編成一字答式問題如下：『何物表示二商品自由交換之比率？』答案自爲『價值』一詞。

• 第三步是將問題排列成一定的次序，而將答案的地位設在一定的地方，例如右邊或題前，而以虛綫表明之。茲舉例如下：

說明：下列爲若干一字答問題，在各題前面虛綫上書出對的答案，祇可用一字或至多一詞。分數爲做對題數。

……(一)何事使南京異於我國其他城市？

施行與校閱 口述法尙適於此種測驗之施行，不過採用此種施行法時，每問應朗誦二遍，且予以充分的書寫時間。黑板法却不很相宜，因為書寫費時，且需很多黑板。最好的施行法還是印刷(或油印)法，予各人一份卷子。

校閱一字答測驗，最好是用製就的答案標準，與試卷對照。如無此種標準，則先閱若干試卷，錄出各答，定其價值，然後再正式校閱。校閱時，對於凡屬同等價值的答案，縱非教員所希望的，亦應給完全的分數。因為所要校閱的是答案的意義，而非正確的字。為免除意見參差起見，宜逐問校閱。至於計分的方法，通常是最好的答案給二分；部分對的給一分；錯的則不給分。

一字答測驗之變像 此種測驗可擬成問題式或填字式，但亦可採用對偶的形式。如取此種形式，則刪去其中之一行，命學生填出。此種變像的一字答測驗，如同對偶測驗一般，可由人物與事變、事變與日期、事變與地方、等

等組成之。例如舉出若干日期命學生填寫其相關的事變如下：

日期	事	變
1911
1914
1918
其他	事	變

此種變像的一字答測驗之可能性甚大，它們適於多種教材及任何施行法。

填字測驗

填字的方法，我國私塾早已採用。但是用作正式測驗法，則始於愛賓哈士 (Ebbinghaus)。他首先建議用作心理測驗法。現在此法已由智力測驗應用到教育測驗，成爲客觀測驗中之一最熟習者。學校教員常使用之。填字測驗與一字答測驗之區別，大都在其形式與其複雜程度。雖然二者可同樣的簡單，

而填字測驗却可更爲複雜。此種測驗包括若干真確的語句或節段，而刪去某些重要觀念或字或詞，命學生依據其知識之範圍及不完全的語句的其餘部分之暗示，填入相當的字或詞（最好是刪去的），以完成語句之正確意義。此法是假定學生如知道全體的意思，即可藉填入之字，完成語句。追憶重要的觀念以完成一思想之能力，是學生熟習教材與否之一優良的量數。有時此種測驗也測量被試應用其知識之能力。

優劣點 填字測驗有些優點與一字答測驗的相同。此法供給異於他法之表述。它無須學生思想並書出全句，祇要他們供給數字。然而字之選擇，却表示學生能自行表述的程度。此法書寫甚少，能在較短的時間內包括或測量大量的教材。學生作答的速度，則視學生的程度與問題的難度而異。它易於適應各式教材，對於需要適當的名詞或事實知識者最相宜。如果編造得法，幾完全淘汰猜對答案之機遇。欲從巨量教材中，追憶明確的反應，學習必須澈底且完備，強記無效。

此法比一字答式較有引起思想之可能性，由刪去的字或詞之前後密切相銜接的語句部分，刺激活潑的思想活動，以決定學生的答案之是否適當。語句之上下文則要求應用或選擇詞字時，須相當注意其正確的意義。此乃填字測驗之一重要價值。明確的思想之表述，需要一豐富且有意義的字彙。被試因此發現各字皆有一明確的意義，而在某種情形下，常有數個可能的字。於是鼓勵學生權衡各字之價值，而依其所代表的觀念選擇之。此法着重於表述之正確性，而不要求巨量書寫。在此點上，它有舊式考試之優點，而無其弱點。它也可敘述連續的動作或事變，以測量學生之知識。

填字測驗的結果供給重要診斷之可能性。分析其結果，即知何問答錯及何人答錯。如製成診斷表，可顯各人答錯的及全班共同答錯的程度。它指示全班需要側重處及各人需要特別指導處。

因為此種測驗中發生的思想程序之性質，有時曾被批評為一智力測驗，而非成績測驗。此語也有一部分理由。它是正確應用已習得的教材之普通智

力的測量。並且刪字太易，則無價值；但刪字太多，却又太難。故此種測驗沒有他種客觀考試之受學生歡迎。所以編一優良的填字測驗很費思索與時間。而校閱時，亦比他種繁而且慢。然就個別問題而論，它是新法考試之最可靠者。

編造法 優良的填字測驗之編造，需要巨量的時間與真正的思想，已如上述。此種測驗，除非慎重擬題，可以承認的答案必多。巨數答案之可能，減少測驗之客觀性，而使校閱複雜。刪字時需要良好的判斷。刪節不得其當，每使測驗太易、太難、或無價值。茲分述編造的程序如下：

(一) 豫備測驗材料 編造填字測驗之第一步驟就是預備材料。這個步驟包括收集與草擬所測量的範圍之重要語句。這些語句或為各自獨立的，或為一組連續的節段。有時語句是完全抄自課本的，不過此種辦法有過於着重機械記憶之危險，切宜注意。所以最好的方法，還是自編語句。先擬一組問題，次作各題的答案。然後用這些答案作測驗的材料。此步之主要考慮是檢

查最後的語句是否是重要的，並且簡單而自然。

至於每一測驗的題數。至少應在二十五題以上，結果方可靠。以時間而論，每二十分鐘約可作七十題。編造時，很可以此為參攷。

(二)選擇應刪的觀念 預備材料後，第二步驟就是選擇應刪的重要詞字。這是關係測驗的價值之步驟，異常重要，必須運用審慎的判斷。一不當，則全部失敗。刪節詞字時，應該注意數點。第一，凡屬妨礙語句所欲表述的意義之詞字不宜刪去。例如『孫中山是國民黨的總理』一語，若刪去『孫中山』與『總理』，此語顯然無價值。因為如此刪節，此語之原意已失，填入『共產黨是國民黨的敵人』亦可。

第二，不應如此刪節，使被試易於或難於填充。填字測驗之難度，依刪去的詞字之數目與性質而異。語句之上下文若明白決定或暗示答案，則太易填充。在他方面，同樣的語句，若刪節太多，則難於答復。

(三)排列 通常對於填字測驗之排列法，皆不甚注意，大都遇刪節之

處，即留一空白地位。學生受試時，則遇空白即填。此法對於學生很為便利。但是空白處散佈在試卷之各方，校閱頗感困難，惟有用裁洞標準紙行之。然而製造此種標準紙手續繁難。另有妥當的方法，就是給予各空白以一號數，更將同樣的號數排列在該問之前，使成直行。此法對於年幼的學生不很相宜，但對高年級的學生却不困難，而校閱則便利多了。茲舉例如下：

說明：此乃一填字測驗。下列語句中，皆刪去一字或詞。將應填的字或詞，寫在該問前面與空白的號數相同之號數內。分數為填對的空白數。

- (1) (甲) 上海為 (1) 之最大城市，是 (2) 之中心。
 (2)

施行與校閱 口述法不宜於施行此種測驗，因為語句若很長，被試心不易記憶。黑版法可行，但其弱點如在一字答式所述有同。最好的方法，仍是印刷法。

填字測驗可由教員校閱之，而不宜於學生校閱。校閱法與一字答式校閱

法同。首先略閱各卷，定各答的價值，然後細閱。

第七章 追憶測驗

考
試
新
論

第八章 真偽測驗

真偽測驗之意義與性質

真偽測驗亦名正誤測驗或是非測驗——同異測驗爲其變像——是客觀考試中最普通且常用的一種。以學校論，現在下自小學，上至中學與大學，幾都採用之。就學科言，對於各種科目，莫不適用。範圍既廣，影響亦大，實有特殊研究之必要。本章的目標就是討論它的各種原理，以作編造或改良這類測驗之指導。

所謂真偽測驗，乃是列舉某一學科中之若干依承認的標準或真或偽的語句，命被試根據他所具的該科知識，辨別其真偽並標出之；然後用承認的標準與之比較，定其答案之錯或對。此種承認的標準是指公認的科學方法、理論、專家的意見、或其他根據而言。通常則以某科數教員共同承認某句爲真或偽作根據。

真偽測驗之回答，初看起來，似甚簡單。但是真偽之判斷所經過的心理歷程決不如此簡單。它們要被試把習得的事實或原理應用到新情境，甚且發生很高深的思想與推理。至於推理的程度，則視編造測驗者草擬思想的問題之機智而定。

真偽測驗可以測量（一）文字的記憶，與（二）推理能力之某方面。文字的記憶即指簡單的事實知識。此類材料約有下列多種：長度（或深度）、重量、面積、容量、日期（或年代）、時間、符號、數目、定義、假設、定理（或定律）、公式、引證、位置、錢價、各種名稱（人名、動物、植物、礦物、化合物、元素、城市、江、湖、山、街道、器具、等）、事變、等等。

此法也可用以測量被試對於某科之一般推理能力。吳德報告某些法律考試「不是專為測量法律知識與法規，却也測量『推理能力』及『有效的應用法律知識於複雜的法律問題之能力』。這些新式法律考試測量推理能力，的確優於舊式法律考試，已在結果之分析中充分證實了。」不過編造這種問題時，

語句中須有某種證據，足爲被試之推理的根據。

真偽測驗之優點與限制

此種測驗之優點甚多，凡屬客觀考試之特長，它都具有。茲舉其顯著者略述之。

(一)客觀 無論是編造或答復或校閱，皆有一定的標準，私人的意見無法滲入，故結果可靠。

(二)取樣廣博 因爲答復與校閱均迅速，故在較短的時間內能包括或測量巨數的問題。問題既多，自能代表學科之多方面，甚至全體。於是被試得一較公平的測量。

(三)經濟 被試的反應既加限制，則節省很多的時間與精力，而用於答案之思考上。擬題雖稍費時，但校閱則極經濟且容易。

(四)刺激適當的心理歷程與態度 要預備受這種測驗，必須廣知範圍內的諸事實。因此引起精讀與優良的修學法之尋求。它不但鼓勵此種心理歷

程的活動，却也刺激疑問的態度與生活中真偽語之訓練。疑問的態度為科學思想與工作之一方面。普通常有接受真語而不懷疑之趨勢。其實應該發展此種懷疑與調查的習慣，真偽測驗適養成此種良好的心理態度。

(五)富有彈性 此法可以順應各地、各學制、各校、或各科之要求，不必像標準測驗之劃一。

(六)有教育價值 此種測驗的問題既多，範圍自廣，實無異於溫習一次。如在考試後，對於懷疑的問題加以討論，格外有價值。

(七)受學生的歡迎 據專家的研究，學生與教員均認考試為不愉快的工作，不過就新舊二種考試比較而言，新法考試稍受歡迎，而真偽測驗又為最受歡迎之一種。

其弱點如下：

(一)難於編造 如欲此種測驗有真正的價值，避免一切含糊、兩可、及無意義的問題，殊難編造。

(二) 忽略文字的訓練 被試僅須書一字或作一號作答，不能表示其組織材料的能力。此點已於第五章論及。

(三) 結果非診斷的 被試如果答錯，不知其思想歷程錯在何處。

(四) 某些科目不適用 凡屬真偽的性質未確定的材料——如哲學中之糾紛的問題然——及一切非判別真偽所能解釋的問題均不適用。

(五) 含有精度原素及反面暗示的影響 此點將於後面討論之。

真偽測驗之編造

真偽測驗有時產生不良的結果，且引起反對的批評。推其原因，皆由於編造不得其法。語句或太少。某些語句或有數種解釋。某些語句或引出或暗示對的答案。真句或偽句過多。真偽句之排列或有一定的系統。這些原素均足破壞真偽測驗之用途。僅僅選擇若干真偽語句，編成測驗形式，不能產生優良的真偽測驗。如欲編造一真正第一流的此種測驗，必先熟知其編造之技術與實際。編造真偽測驗之程序略如下述：

- (一) 擬題
- (二) 定真偽
- (三) 辨難度
- (四) 定格式與題數
- (五) 編說明
- (六) 製答案標準
- (七) 定校閱規則

以上諸項，已於第六章述及，茲僅論其特殊之點。編造真偽測驗之第一步，是草擬巨數真確語句。擬題時應該注意數點。語句以簡短爲妙，魏迪曼 (Weidemann) 建議每句祇須十至二十字。不可用半真半偽或意義兩可的語句。不可用『詭題』或『迷題』。避免暗示答案的詞字。有些詞字預決學生的答案，即魏迪曼所謂『特殊決定原素』 (Specific determiners) 是也。其意即指測驗中之一字或一詞，使句義真多於偽或偽多於真，致機遇公式不能矯正其分數。魏

迪曼發現(甲)三句『常常』與『未嘗』語句中有二句是偽的，(乙)三句程度或比較的語句有二句是真的，及(丙)三句原因或理由的語句中有二句是偽的。對於此種原素，有些人主張完全避免。但亦有他法約制之：(1)凡含有特殊決定原素之問題，使之真偽約各半。(2)不可用『不』字將一語句由真改爲偽或改爲爲真。(3)同一句中，不可用二特殊決定原素。

擬題之後，次步就是定問題之真偽：孰應爲真句，孰應爲偽句。最好是用機遇法決定之。例如拋擲銅元，正面代表真句，反面代表偽句。若擲出正面，該句即定爲真者；否則改成偽句。但是改爲偽句時，必須注意一點，卽改爲偽句若太明顯，則不必更改，甯可使之仍爲真句。至於真句題數宜各半。如此可避免機遇的成功，否則學生一發現真多於偽(或偽多於真)，遇不知真偽時卽猜其爲真的(或偽的)，分數必較其應得者爲高。

第三步工作是排列問題。排列的方法，不外下列四種：難度、價值、論理的次序、及機遇法。此四法中，以難度排列法爲最常用，前章已詳論之，

茲贅述。

第四步是定測驗的格式與題數。據魏迪曼的研究，真偽測驗之最好的形式如下：答案的地位應在語句之前，而以虛線表示之。問題的號數則在虛線後。如果語句甚長，第二行須退後二格。例如：

.....69.	x	x	x	x	x	x	x	x	x
	x	x	x	x	x	x	x	x	x

標答的方法計有二種：(甲)填充，如書真偽、是非、正誤、十一、或十○、等等。(乙)在印就的答上作號，如○、△、一、x、等等。上述二法相較，前者優於後者，而尤以十一與十○法最佳。凡屬習過代數者，皆覺十一(即正負號)有意義，惟學生閱卷時，太易改一爲十，故以○代一爲宜。

至於測驗之長度，可參照下列研究。盧赤估計中學與大學學生每分鐘能答十至十五問真偽題，小學生約答五至八問。就題數而言，每答一百問真偽題約需時102分鐘。關於測驗的總題數，吳德主張以二百問爲最低限度。此

數似覺稍多。一般測驗編造者都以爲至少需七十五問，而以一百問爲適當。若問題甚多，不妨分作二或數替換式。

第五步是編說明，已於第六章述及，茲僅舉例如下：

這是一真僞測驗，共計五十題。其中約有半數是真的，半數是僞的。詳細閱讀各句。如果你認爲是真的，在該句前虛線上作「+」號，否則作「○」號

依次回答。對於懷疑的問題，切勿猜度。

不許談話或詢問。

時限是二十分鐘。

工作前，先看下列：

十 凡人皆有死。

第六步是製答問標準，第七步是定校閱規則，均已於第六章述及，無須重述。不過真僞測驗的分數，要用機遇公式校正之。公式如下：

$$S = R - W \quad \text{或} \quad S = A - 2W$$

此公式中，S 爲校正的分數，R 爲做對題數，W 爲做錯題數。例如一測驗計有 137 題，某生做對 111 題，做錯 26 題，代入公式如下：

$$(1) R - W = S = 111 - 26 = 85 \text{ (校正的分數)}$$

$$(2) A - 2W = S = 137 - 2(26) = 85$$

真偽測驗之施行與校閱

真偽測驗之最好的施行是印刷法，使每人都有一份試卷。如無此種設備，口述法亦很滿意。黑板法雖亦可用，但不甚佳，因教員需巨量書寫，而黑板也有限。

真偽測驗之校閱，可由教員或學生行之。若非學期考試，似宜由學生校閱，既可增加興趣，又可矯正錯誤。

真偽測驗之變像

是非測驗爲真爲測驗式之一種。它與真偽測驗之惟一異點，就在它採取

詢問形式，而真偽測驗則用說明的形式。在小學校中，此種測驗似較真偽式爲優，因它不致混亂思想，且無錯誤的印象。茲舉例如下：

是 犬有四足否？

同異測驗亦屬此類之一變像。它的目的是要學生認識詞字的意義之同異，例如：

大。巨

高 × 低

精度與反面暗示的影響

對於真偽測驗之第一批評，就是它含有猜對正確的答案之機遇。精度原素造成二種不良的結果。第一，它鼓勵純粹精度之習慣。鼓勵精度實屬不合教育原理且非科學的。第二，精度的結果，對於學生或不公正。通常是假定一切錯答都是精度的，並也假定猜對與猜錯的題數恰相等，故真正分數是對的題數減錯的題數。就理論而言，題數若多，猜對數與猜數將相等，但題數少，則有大差別。

現在的測驗學者對於此事，意見不一。吳德反對猜度，普通的教員也反對猜度。他們以為猜度不能得真正的分數，且養成不良的習慣。麥柯爾 (McCall) 却贊成猜度，以為猜度愈多，則校正機遇的結果愈準確可靠。但據狄格 (De Graff) 與盧赤研究的結果，最好不許猜度，並採用校正機遇的公式校正其結果。能而根本的辦法還要在編擬問題中求之，即問題須如此編造，使被試若依暗示或純粹猜度，必致答錯，則可免去此種影響。

關於真偽測驗之另一反對的理由，就是它的反面暗示的影響。所謂反面暗示，乃指偽句給與的錯誤的聯想或印像而言。近代教育心理學頗重視錯誤的練習之危險。許多人以為此種考試中之偽句是危險的教育學，因其在學生心中造成錯的神經結。既能一測驗中有半數語句是錯的，顯然有造成負的或反面的影響之大機會。但是這些批評家都祇知其一，不知其二，須知此種考試中之真句在學生心中，確定真理的事實，使其印像加深。並且他們未曾思及心理學中之其他原則，例如『心向』。反面學習的危險大都視學生受試時的

心理態度而異。假若把一頁字——其中有半數錯字——教學生學習，自然有很多的錯誤練習。但是真偽測驗的情形適相反，其名稱與做法說明及被試已往的經驗，均告訴他們其中有錯的，使之採取批評的、反對的、與思考的態度。他們須知約有半數是偽的，其工作就是尋出這些偽句。進而言之，實際生活亦復如此。社會情形中有許多事物須加判斷，故此法對於發展兒童之辨別力很有益。考試之後，如加討論，更無錯的聯想之慮。

再就實驗方面而言，柏拉德用史地真偽測驗考試數班十三歲的學生，後又將同樣問題編成短答式考驗之，並在班中討論。最後再用追憶式施行同樣問題。根據此種實驗的結果，他歸納說，兒童自真偽測驗所習得者較自他種考試所得者為多，而尤奇者，自錯句所得者多於自真句者。

此外羅博斯女士 (Miss Roberts) 與盧赤及雷馬士等 (Remmers) 之研究，亦均證明真偽測驗之影響是一正面暗示的現象。

考
試
新
論

第九章 選答測驗

選答測驗之意義與性質

選答測驗是亞迪斯博士 (Dr. O.H.S.) 發明的。它包括若干真確語句，每句之後，各有數答(自二答以至七答，最普遍的是四或五答)，但是其中僅有一個是正確的或最佳的，其餘都是錯誤的或次佳的。它的做法是要被試運用其知識與判斷，選出對的或最佳的答案並標明之。此種測驗要求被試的判斷力與辨別力。被試能自數種錯答中選出對的答案，是他對於教材之一優良的測量。例如：

何物表示二商品自由交換之比率： (1) 勞工代價 (2) 需要

(3) 價值 (4) 供給…………… ()

如被試答『價值』，即算答對，否則答錯。此乃假定被試若了解題義，立刻知所選答；而他方面則假定：凡不知何為『價值』的人，將覺諸答皆適用，以致

答錯。

選答測驗頗值得廣大的應用。它是客觀考試法中之一最佳者。但是它雖優良且常用，却不敵真偽測驗應用之普遍，這種情形大都是因為較難編造的原故。預備若干偽答，沒有改真句為偽句那樣容易。

選答測驗之優點與限制

選答測驗有若干顯著的優點。它雖比真偽測驗較難預備，却易施行與校閱。祇須用一答案標準，即可迅速且容易的校閱。此種測驗適於各種學科與教材。它為真偽測驗一樣的受學生歡迎。他們覺其易答、明白、公平、有趣、且澈底。

選答測驗刺激一實際性質的心理態度與活動。它要求事實之熟識與慎重的估值、健全的判斷、與知識之應用。它含有比較、辨別、偵察類似點與關係及觀察之要求。它的情境不是人為的，都是自然的與實際的。生活情形中，常需自一組可能的事物中加以選擇。測驗之形式與答案之性質可如此編造

，使予適應現代生活之複雜的且變異的環境所需的各種心理活動以練習的機會。選答測驗不僅測量孤立事實之具有與否，且測驗事實之推理。

對於選答測驗有二大批評，即猜度對的答案之可能性與錯誤的印象之影響。此二點亦為真偽測驗之缺陷，前已討論了，其同樣的解釋可應用於此地。然而選答測驗之猜度的機會，却較真偽測驗的為少，猜度的原素依答案數為比例而減少。如有充分數的答案，猜度的可能性實不足為反對的理由。例如五答式，猜對的機會，僅為五與一之比。再則我們可用機遇公式矯正其猜度的影響。但是普通對於四答以上的選答測驗之結果多不校正，因為根據實驗的結果，校正的與未校正的並無多大的差別。

至於錯誤的印象之影響——即使學生混亂且於其心中造成錯誤的觀念——我們雖應承認其有某種危險，但一仔細研究理論的觀點與現有的證據，則表示危險甚少。如果測驗是關於學生所未學習的材料，自易發生混亂與錯誤觀念，但是學生若熟習所包括的教材，則不致有此結果。

選答測驗之編造

優良的選答測驗，不能隨意草擬，必須慎重編造。它的編造步驟約分下列數種：

(一)擬題 編造選答測驗之第一步是草擬若干真句。此種真句必須具備前章對於真偽測驗所建議的諸特性。它們應該都是重要的、不相關聯的、正面的、簡單且明瞭，以能刺激思想及證明基本真理的爲尙。凡屬含糊的、不確定的、及複雜性質的材料，均不可用。擬就以後，可把它們隨機排列或依論理的次序排列。

(二)編造 編造選答測驗之次步是草擬對的答案及似是而非的錯答。編造時，我們心中應記得凡無學科知識的人常依一定的方法選答，必須避免此點，使之不能成功——即被試若如此選答，必致答錯——結果才有價值。編造時，我們應當注意下列數點。錯答須與對答一樣的熟識。它們均應爲學生所常見者，不可相差太遠，致供暗示的線索。例如答案都屬通都大邑

，不可雜入一鄉僻的地名。錯答應與對答皆屬同類的概念。例如對答爲菓品，則所有的答案均應爲菓品。一切答案的長度與難度要互相彷彿。例如答案都很短，對的答案不可特別長。錯答與對答應同樣的自然、合理、吸引人、且互有關係。顯明的錯答不可用。部分對的答案宜少用。

至於測驗的題數，則隨測驗的時限、學生之成熟程度、與答案數而互異。約言之，每一測驗之最低限度應有五十題。就時限與答案數而言，每分鐘約答五個七答題、六個五答題、或七個三答題。自學生的工作速率立論，最慢的學生每分鐘至少能答三題。普通學生每分鐘能答五至六題（三答至五答者）。每分鐘作四題爲一很合理的標準。

(三) 排列次序 答案已既編就後，對的答案發現的次序應依機遇法排列之。例如用擲骰法。擲出爲1，則列在第一；擲出爲5，則排在第五；等等。若依答案筆劃之多寡，定其次序之先後，亦一良法。

(四) 定格式與編說明 編造選答測驗之最後一步，是定測驗之格式與

編製做法說明。普通的答法是命被試在對的答案下劃一橫。此法對於學生很為便利，但是校閱則感困難，因為答案的地位不整齊，製造答案標準頗費手續。較優的方法是給各答一號數，而於每問題之前或後印一虛線或括弧，使成一單行，作為答案地位。被試回答時，祇須把對的答案之號數填在虛線上或括弧內。此法對於學生並不很繁雜，而校閱則異常便利。茲舉例如下：

說明：下列為若干不完全的語句，後面各有數答，以便選擇。你認某答為最好的答案，就把它前面的號數填在題前的虛線上。（或題後的括弧內）。分數為做對題數。

..... (3) M等於 (1) 10 (2) 100 (3) 500 (4) 1000 (5) 1500
 或： (3) M等於 (1) 10 (2) 100 (3) 500 (4) 1000
 (5) 1500

選答測驗之施行為校閱

選答測驗之施行，亦以印刷法為最佳，給予各人一份試卷。黑版法雖亦

可行，但不甚佳，因為所佔的黑版太多、書寫費時，且有學生預見問題諸弊。口述法則完全不適用，因為要被試同時記憶許多答案，並決定孰為對答，殊屬不易。結果或竟為記憶力之考試。

選答測驗可由教員或學生校閱之。至於製造答案標準與校閱試卷的方法，則與真偽測驗的相似，故從略。惟因此種測驗或含機遇成功原素，計分方法稍異於他種者。選答測驗計分時，普通多採用下列公式，以校正其機遇：

$$S = R - \frac{W}{(N-1)}$$

此公式中， S 為校正的分數， R 為答對的題數， W 為答錯的題數， N 為每問的答案數，如三答、五答、等等。若是三答式測驗，公式為： $S = R - \frac{1}{2}W$ 。茲舉例說明如下：

130 (最高分數)

5 (未做題數)

125 (實做題數)

$\frac{24}{101}$ (答錯題數)

101 (答對題數)

$\frac{12}{89}$ (24 ÷ 2 或答錯題數之半數)

89 (校正的分數)

對於四答、五答、以至七答的測驗，可推演相似的公式，如四答者為對的題數減去三分之一答錯題數等等。但照最近的趨勢，三答以上者多不校正機遇——即以為對題數為最後的分數——其原因前已言之。

選答測驗之變像

選答測驗可有多種的變像。答案之數目自可變動，由三答以至七答。測驗問題可編成語句式或詢問式。除了這些形式的變像外，尚有性質之變像。一二答測驗亦可算為選答測驗之一種，然而普通都歸入真偽測驗類，其實錯誤。根據最近實驗的結果，如用同樣的問題施行二答測驗與真偽測驗，後者似

覺較難。若被試完全缺乏所測量的知識，二式的機遇相等，倘使被試稍具該種知識，二答式因為建議較多的知識，較易答對。茲舉例如下：

中國首次革命是在 庚子年 辛亥年

選答測驗之另一變像為最好答案測驗，亦名判斷測驗或證明測驗。此法為測驗中所常用者，宜於原則之推理與辨別，不適用於事實問題。它的做法是示被試以若干事實或原則與某些原因，要他們選出使之真確的一原因。例如：

說明：下面為若干原則或事實，後面都附有使之真確的原因，但是其中僅有一個是最佳的。選出最佳的原因，把它前面的號數，填在後面的括弧內。

(25) 蕨科植物在自然中佔一重要的地位，因為：

- (1) 附於它根上的微菌使淡氣回到土中
- (2) 它們能在貧瘠不能種植其他植物的土地生長

(3) 乾草的經濟價值很大..... ()

最壞答案測驗又爲此法之一變像。它包括若干語句，各句皆有數答，但是其中僅有一個錯的或無一錯的，要被試選出錯答或最壞的答，並標明之。例如：

說明：下面爲若干語句，後面各附有數答，其中僅有一個錯的或最壞的。把此答的號數填入後面括弧內。

- (10) 孔子 (1) 生於周 (2) 受業於老聃 (3) 周遊列國
(4) 歿於漢..... ()

諸選答測驗亦爲選答測驗之一變像。此種測驗與普通選答測驗之異點，就在它有二個以上對的答案，而不祇一個對答。它的做法是命被試選出一切對的答案，例如：

生活之必需品是：水，衣，屋，金錢，空氣，食物

選答測驗還有一種變像，即複選答測驗。此種測驗中，含有二類以上不

同性質的答案，命被試各於其中選一對答。例如：

富士是——：山，湖，城； 在——：中國，日本，俄國

考
試
新
論

第十章 雜式測驗

除了上述各式測驗外，尚有數式測驗，因其不常見，遂爲教員所忽略。這些測驗之用途都很有限，多數僅適於特種科目或教材，然而有時却是最佳的測驗法。它們雖與上述者頗多類似之處，但均具有特殊的性質，值得分別討論。本章所論者，計分下列五種：

- (一) 對偶測驗
 - (二) 類比測驗
 - (三) 重組測驗
 - (四) 指認測驗
 - (五) 訂誤測驗
- 對偶測驗

對偶測驗也可算是選答測驗之一變像，故亦稱選擇測驗。它是測量被試

明瞭事物間關係之程度或能力。它的做法是列舉二行事物，命被試辨出第二行中某事物與第一行中某事物之關係，且標明之。

對偶測驗具有數種有價值的特性。編造與校閱都很簡單且公正。如果慎重編造，幾可使之完全客觀，其客觀性實高於其他測驗。學生亦喜其簡單及答覆便易，而無書寫的繁難。它是一迅速測量廣大範圍的正確知識之方法。它頗適於社會科學之教材，而尤宜於史地一類的科目，用以測量人物與事變、事變與日期、事變與地方、國名與都城、城市與其特徵、事變之因果、名詞與定義、外國字與本國譯名、發明家與其發明、名著與原作者、引證與原著者、器具與用途、等等之聯想。

對偶測驗之編造，約分下列三步：

(一)擬題。編造對偶測驗之第一步驟，是草擬若干真確語句。每一問題所包含的項目，以十個至二十個為最得當。如在十個以下，則含有猜度原素，機遇成功太易。萬一不易收集充分的數目，甯可採用不完全的對偶法，其

解釋詳後。項目若在二十個以上，則太費時間。項目甚多時，可分成若干組，例如三十項目，則分成十五個項目一組之二組。

每一問題中之諸項目，應屬同類性質，如此則無精度之綫索，否則易生機遇成功之危險。例如，某題第一行包括若干人名與少數日期，則頗易於第二行中求得與此少數日期相關的事變。

(二)分行。編造對偶測驗之次步，是把諸語句分成二行。此法很爲簡單，可將語句之主詞與賓詞各分爲一部。例如，『哥倫布於一四九二年發現美洲』一語，可分成：

· 哥倫布 一 於一四九二年發現美洲

但在對偶測驗中，不常採用完全的語句。普通都是使它極其簡單，刪去一切形容詞、助詞、等等，祇留主詞與賓詞中之重要字或詞，於是成爲二行單字。比方，上例若注重事變，則成下列形式：

哥倫布 一 美洲

如重視時代，則形式如下：

亞倫布 一 一四九二年

(三)隨機排列。語句分行之後，第三步就是把賓詞(即第二項之項目)隨機排列，例如依筆劃多寡、年代先後、或純粹機遇法。

(四)定測驗的格式與編做法說明。編造對偶測驗之第四步，即最後一步，是決定測驗的格式及編造做法說明。最普通的方法是在第一行與第二行之間，留一空白，而以虛線連接之，作為答案的地位。例如：

哥倫布.....美洲

但是第一行之詞字長短互異，故空白地位亦參差不齊，校閱時，稍感困難。所以較好的方法是在第一行前面留一空白，而以虛線表示之；第二行各項目，則各予一號數。學生回答時，祇須把第二行對的項目之號數寫在第一行該相關的項目前虛線上。例如：

下面為若干名著與著者。在著者人名表中，求出各書之原著者，把他前

面的號數寫在書名前虛線上。

..... 思維術 (1) 羅素

..... 創化論 (2) 杜威

..... 哲學中之科學方法 (3) 柏格森

對偶測驗之施行，自以印刷法爲最佳，黑版法亦尙可行，但是口述法則完全不適用。至於校閱，可用答案標準行之。計分的方法則以每對一對偶給一分爲原則。

對偶測驗有二種變像，卽不完全的對偶法與複對偶法。所謂不完全的對偶法，是在第二行諸對的答案中，增加若干似是而非之無關係的項目。它的目的是要避免精度原素，並增加測驗之難度。倘使可以配合的對偶甚少，尤宜如此，結果才有價值。

複對偶法是要配合三行以上的事物。此乃包括較廣範圍的知識之方法，其難度大增，不宜於小學初年級學生。茲舉例如下：

下面有三表：第一表中都是歷史上重要人物，第二表中都是重要的事變，第三表中都是日期。求出第二行與第三行各項目與第一行各人物之關係，把它們前面的符號寫在第一行有關係的人物前面的虛線上。

(1) 哥倫布 (1) 發現美洲 (子) 一四九二年

類比測驗

類比測驗是測量被試明瞭各種原則、理論、事實等等之關係的知識。它的目的是要學生察出某環境下事物間的關係之類似性。它的做法是列舉諸有關係與無關係的理論或事實，命被試依據某一類比而推測其同樣關係者——即他一類比。此法宜於測量歷史、地理、數學諸科之知識。

類比測驗雖屬優良的測驗法之一種，但是除非其中之次一類比與首一類比之關係完全相同，此法即喪失其價值或意義。校閱試卷時，亦感繁難，採用追憶式時尤其如此，答案每有出人意外者，頗難處理。有人批評此法為測量智力，却也有一部分理由。

類比測驗之編造，可採用追憶、選答、或刪字之形式。編造此種測驗之第一步是集合若干相關聯的類比。第二步是刪去類比中之一項，普通都把該項列在最後。例如：

下列各問中都有三個名詞與一空白。依照前面二名詞之類比，推測與後面一名詞成同樣的關係者，把它填入空白中。

(1) 孔子：孔教：耶穌：.....

若用選答法，則求出若干與最後一項類似的錯答，一同列入，命被試辨別之。例如：

下列各題均有二類比。根據前一類比，在後面括弧內選出與後者成同樣類比者，在它下面劃一橫。

(15) 孔子：孔教：老聃：(佛教, 耶教, 道教, 回教)

口述法、黑板法、及印刷法，對於類比測驗均適用，而以最後一種最佳。它的分教是做對的題數為標準。

重組測驗

重組測驗度量被試重組或重排事物之能力，例如年代之先後、動作之次序、雜句、等等。它包括一組隨機排列的同類事物表，命被試依某一標準或根據重排其次第。比方，命被試依人口之多寡或面積之大小重排一城名表之次序。故有時亦稱連續測驗。此種測驗最適於歷史、地理、物理、化學諸科，而於他種科目，則不相宜。

重組測驗之編造法，約分下列三大步驟：

(一)擬題。編造重組測驗之第一步也是草擬問題。每問所包含的項目，不宜超過十個以上，編以六個最得當。

(二)隨機排列。擬題之後，第二步是把各項目依機遇次序排列之。他如照筆劃多寡為先後等法亦可。

(三)定測驗的格式與編做法說明。此步與以前各測驗中所述者相同。答案的地位，以題前為最佳。茲舉例如下：

下列爲一城名表。選出你認爲最大的城，在它前面虛線上寫一(1)字，選出次大的城，在它前面虛線上寫一(2)字。餘類推。

.....上海
倫敦
巴黎
紐約

重組測驗之施行，以印刷法爲最佳，黑板法亦可行，但口述法却不適用。至於校閱，則較他種測驗爲麻煩，因其計分法大異。如果單以做對的題數爲分數，手續自然簡單，但是此種測驗必須接受不同的正確程度之答案，故有一特殊的計分法，即以學生答案的次序與真答之次序的差數爲根據。計算的方法如下：用每問最大的差數爲該問之總分數。所謂最大的差數，就是真次序與一完全相反的次序之差數。換言之，應列在第一位的，列在最後了，應排在最後的，排在最前了。茲附差數表如下：

項目數	差數
2	2
3	4
4	8
5	12
6	18
7	24
8	32
9	40
10	50

計算分數時，先求出學生答案的次序與真次序之差數。再將最大的差數減去此數，餘數即實得的分數。例如：

標準答案	甲 生		乙 生		丙 生		丁 生	
	次序	差數	次序	差數	次序	差數	次序	差數
2	2	0	1	1	3	1	3	1
4	4	0	3	1	4	0	1	3
3	3	0	2	1	2	1	2	1
1	1	0	4	3	1	0	4	3
差數總數		0		6		2		8

依照此表，甲生的分數應是8，因此題的項目為4，最大的差數是8，而甲生答案的次序與真次序之差數為0。依前法遂成 $8-0=8$ 。乙生的分數為2 ($8-6=2$)，丙生為6 ($8-2=6$)，而丁生為0 ($8-8=0$)。

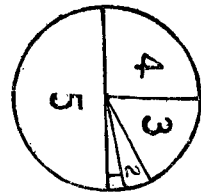
初看起來，此種計分法似極繁雜，却很易明白，而一經熟習之後，且很簡易。

指認測驗

指認測驗表顯被試對於教材之熟習程度，通常包括一圖、表、繪畫、等等，使被試指認其中的某特性。它很類似對偶測驗。例如，用一暗射地圖，命學生指認其中的某些政治區域、地質、城市、物產、等等。此種測驗特別適於地理、歷史、幾何、手工、及自然科學之用，對於文學、外國語、公民諸科，則不相宜。它的最大限制，就是難於編造，且不易印刷，因需繪圖的原故。

指認測驗之形式，可採用追憶式或認識式。茲舉例如下：

下面是一圓形，分成五部，代表五大洋，其名稱列左。五部的大小，是以諸洋的面積為比例。求出各部所代表的洋名，把該部的號數填在洋名前虛線上。



.....	南冰洋	大西洋	印度洋
.....	北冰洋	太平洋		

上例爲認識式，這些問題亦可改成追憶式，即刪去附表，命學生自行填入洋名。如欲使問題較難，機遇成功甚少，附表中可多列數種關係的事物。不過如此辦理時，說明中必須明白告訴學生。

口述的施行法，對於指認測驗不甚適用。其他二法却均可行，但以黑版法較優，因爲此種測驗的文字甚少，不大佔黑版面積，並且教員本人繪圖不致錯誤。在他方面，印刷法雖佳，但印圖不易，恐或印錯。至於計分方法，亦以每對一項給一分爲原則。

訂誤測驗

訂誤測驗包括若干錯誤與未錯的語句，却可由更改、加入、或刪去一二

字而訂正之。但是此種應訂正之處，並未標明，必須學生自行求得之，然後加以訂正。有時它祇要學生指出錯誤處，但是普通都要學生加入正確的字。這種測驗雖無他種之流行，却也有若干優點。它比別種測驗更善於度量完全的或整個的思想及純粹記憶事實以外的其他能力。另一優點，就是它毫無猜度的機會或機遇原素。在他方面，此種測驗和填字測驗一樣的難於編造。學生訂正處，常非教員心中應改正者。

在擬題方面，對於真偽測驗之各種建議，頗適於此法。每問祇可有一錯誤處。全部測驗中，錯句與對句的數目之比例，應隨時變化：有時巨數對句中夾少數錯句，有時二者之數目略等，更有時多數為錯句，或甚至全部為錯句。訂誤測驗之例如下：

下面的語句，多數是對的，小數是錯的。尋出錯句並改正之。1. 紅血輪的功用是殺菌。

三種施行法，對於此種測驗皆適用，而以印刷法為最佳。至於校閱則較

難，可參看關於填字測驗之建議。計分的方法是改對一句給一分；若原為對句，而未更改或改錯，亦給一分。

第十一章 綜合測驗

綜合測驗是一包括二種或二種以上的不同的測驗式之測驗。通常多稱它爲彙選測驗，因爲它排列若干類似的測驗式成組，以產生一統一的測量的結果。聯合數測驗式成一獨立的測驗，爲學期與學年考試之正當辦法。所以此種測驗實有特殊考慮之價值。

綜合測驗之優點與限制

綜合測驗便於編造。我們計劃一考試時，常發現某種重要教材特別適於某測驗式。因此，若用數測驗式構成一考試，則可依教材之特性與需要，選擇適當的測驗式以表述之。此種測驗不但使測驗之編造便易，却也增加一考試之興趣。一式不免過於單調，多式則易引起趣味。一很長的一式考試，在未作完之前，卽已使人感覺乏味或單調。因此綜合測驗之殊異的性質本身就是一顯著的優點。綜合測驗之各式要求對於教材之不同的心理反應。認識測

驗所要求的認識事實之能力，異於追憶測驗所需要的追憶事實之能力。綜合測驗藉不同的測驗式，度量較廣範圍的心理能力。如果對於教材之各種心理反應皆生作用，並且都有取樣，結果自更可靠。

綜合測驗爲成績之一綜合的量數。它告訴我們以學生之比較的地位。通常教員如欲求得學生對於所有的科目之總等第，必須舉行若干考試，然後集合這些測驗之結果爲一綜合的分數。但是此種辦法，不很公平，因爲各測驗之難度互異。一測驗易而他測驗難。容易的測驗之分數甚至二倍於困難的測驗之分數。綜合測驗不但避免此弊，並且費時少，省力多。

然而綜合測驗有一弱點，即被試容易混亂。因爲作一種測驗時，心向完全集中於該式。若數分鐘一變，每易弄錯。但是此種限制可以避免，即將各式測驗劃分清楚，更附以明晰的做法說明。

綜合測驗之編造

綜合測驗之編造法，包括它所含的個別測驗之編造法。此外尚有某些專

門的細則。這些細則大都是關於考試的各部分之適當的選擇與排列。爲各種教材選擇適當的測驗式，並將它們排列成良好的形式，需要某些正確方法之知識。綜合測驗之編造法可分下列三大步驟：

(一)擬題。編造綜合測驗之第一步，是擬一綜合的真句表。課程之各重要方面均應包括。事先應慎重評閱所測量的教材，以便選擇應包括的諸要節。有些語句可取材於補充讀物中。包括此種材料可促進學生審慎研究教員所指定的參考書。語句之結構却須異於原書者。多用原書之語句，則鼓勵記憶而非推理。如果教員採用本書第六章所述的卡片制，記載逐日教學中所見到的適當測驗材料，則擬題手續甚爲簡單。

所擬的語句，應多於測驗中實際要用的，最好約比所需者多一半至二倍。其中必有許多刪去不用的，例如語義含糊、意見衝突、恃他句知其意義、重複、複雜、太長、造句不佳、無關重要、以及含有不常見的字或詞之語句。這些語句，除非改正，均應淘汰。所擬的語句，應代表一良好的難度距離。

。有些語句應如此容易，使所有的學生皆能答復；有些應如此困難，僅極少數人能答對。難度適中的問題則應多。關於其他各點，可參看本書第六章。

(二)改成適當的測驗式。編造綜合測驗之次步是將諸語句改成適當的測驗式。此步必須考慮數點。第一，選擇應用的測驗式。此事關乎教員之好惡、教材之性質、施行之方法、及測驗之目的。如果用口述法施行此種考試，似應限於真偽式與一字答式。若口述法與黑版法併用，可加上對偶式與選答式。單用黑版法，則常感覺此種考試太長。最好的方法是油印之，使各人皆有一份試卷。如有此種油印的設備，則可包括任何測驗式。此種考試必須包括追憶與認識二大類。真偽式、選答式、填字式、與一他式——例如對偶式、一字答式、或一雜式——構成一優良的考試。

第二，定奪各式之長度。換言之，決定每一測驗式應包括的問題之概數。此乃教員獨斷的事，大都視考試的時限與學生答覆各式問題之速度而定。教員應根據本人的經驗與專家實驗的結果，定奪學生反應各式問題之速度。

真偽式最快。認識式較追憶式爲速。約言之，二十真偽題、十選答題、與七填字題所需的時間略等。我們若以此爲標準，則編造一三十分鐘的真偽、選答、及填字的綜合考試時，可用一百真偽題、五十選答題、與三十五填字題。

第三，將諸語句改成各測驗式。既已決定須採用的測驗式後，就應審慎的披覽各語句，求出其適於某式，即在該句前作一標誌，例如適於真偽式者，則在句前書一『真』字，適於選答式者，則在句前寫一『選』字，餘類推。各句俱已標誌後，就把同式的歸併在一處。然後依本書以前各章所述的方法，將語句改成各該測驗式。

(三)定格式及編說明。編造綜合測驗之第三步是決定問題排列之形式與編製做法說明。關於排列之形式，已於以前各章詳論之，茲不重述。至於說明，則須略加解釋。因爲綜合測驗包括數種不同的測驗式，故在正式的測驗之前，應有關於全部測驗之普通說明。此種說明宜簡短，告訴被試以測驗之

目的、所包含的部分、各部的性質、等等。

各個測驗之前，須附特殊的說明。此種說明之編製法，亦已討論了，茲從略。現在且舉二種說明之例如下：

社會科學考試
第六年級

姓名.....

日期.....

普通說明：

此種考試包括四部。迅速回答各部。不必多用時間躊躇難題。在未作各部測驗前，仔細閱各部之特殊說明及各個問題。

第一 部

說明：下面為若干真偽題，有些語句是真的，有些是偽的。如你認該句意義真確，則在題前虛綫上作一“十”，否則作一“○”。分數是做對題數減做錯題數。

.....1 _____

.....2 _____

.....3 _____

.....4 _____

.....5 _____

.....5 _____

綜合測驗之施行與校閱

綜合測驗之施行，亦以印刷法爲最佳，但其他二法也可行。然而我們須認清一點，即各部不必都限於同一方法。我們可依各部之特性，選擇其施行法，例如，宜於口述者，採用口述法。故一考試之第一部用口述法，而第二部或採黑板法，第三部或取印刷法。

至於校閱，祇須製一答案標準，可由教員或學生校閱之。答案標準之製法，已詳第六章中。最簡便的辦法，是正確的標出一空白試卷，而用之與學生的試卷對照。

如欲分數畧等，不致任某一部的佔很多的地位或數量，最好須加以均衡。但是統計的均衡法甚爲繁難。我們可用一簡便的方法，即各用一指數乘各部的總分數（或除之），使之畧等。前面曾經提及以一百真偽題、五十選答題、及三十五填字題組成一綜合的考試，我們就用它爲例，說明此法。此三式的題數，以真偽式爲最多，計一百題，選答式祇及其半，填字式約爲其三分

之一。於是二倍選答式分數，三倍填字式分數，則三式的分數略等。

此種精密的考試，自然很費精力與時間，因此用過之後，仍須留作將來之用。換言之，試卷概不退回學生，而測驗的材料應保存在安全的地方。故應製一報告單，告訴學生以其結果，以代退回試卷。此種報告單必須包含學生欲知道的一切事。最好將此單附在試卷第一頁底下，則校閱後，即可將結果抄入，撕下給學生。茲舉例如下：

至於其他手續，已詳見本書第六章結果之處理一節中，毋庸重述。

注意：此單記載你的成績。試卷概不退回。	
。寫上你的姓名及測驗的日期。	
姓名	日期
社會科學考試	
第六年級	
第一部分數	被測人數
第二部分數	班級平均數
第三部分數	均方差
第四部分數	你在全班之等第
總分數	
附註：	

第十二章 分數制度

測驗分數與等級分數之區別

筆試之一常見的錯誤，就是不分別測驗分數 (Score) 與等級分數 (Grade)。普通多視一學生的測驗分數爲其考試的分數。在一般教員看來，二者亦爲同一事物。其實，測驗分數與等級分數有大區別。測驗分數是一學生對於某一測驗或考試所得的點分數。例如，一學生默寫一百字。他的測驗分數就是默對的字數。此種分數雖然代表學生的作業或成績，却隨測驗之難度及題數而異。如爲論文式考試，更因閱卷者之寬嚴，而生差別。測驗分數本身毫無意義。倘欲使之具有價值，必須加以解釋。等級分數則爲測驗分數所化成的學校分數。它是根據某一標準之測驗分數的詮釋。對於普通非正式的考試，解釋測驗分數之標準是受試的全體之成績。等級分數之目的是敘述學生在全體中之相對的成績。它表示學生的比較的地位。測驗分數必須化成等級分數

，然後纔有價值及明確的意義。

這兩種分數之混亂，尤其與論文式考試相連。因此有些教員以爲一客觀測驗應有一百問題，使其分數爲百分分數。此種錯誤觀念是由於假定一考試之成績是一絕對的量數：零分爲缺乏被測的能力，一百分爲完全的成績。根據此種假定，學生的測驗分數遂被認爲其等級分數。這都是由於未能分別二種分數之故。

分數制度

現在各學校所採用的分數制度，約有數十百種，乃是毫無疑義的事。據阿底爾(O'Neil)調查美國意大利諸州二八一個中學所用的分數制度，不下百餘種。但是種類雖多，大別之，約分下列二大類：

(一) 絕對的量表制，如百分分數制。

(二) 相對的價值、等級、或常態曲線制。

百分制 第一類之最著名者爲百分分數制，現在各學校仍多沿用之。所謂

百分制，就是把學生的等級列在零至一百之間。此種量表還假定一分數爲及格分數，例如六十分或七十分。此種武斷的與固定的及格分數是這種量表之絕對性的另一證明。它未計及學生達到、超過、或不及及格分數之數目，爲考試、科目、或閱卷標準之難度的函數。此種分數之單位各別：有以十爲單位者，如五十、六十、七十、等等；有以五爲單位者，如四十五、五十、五十五、等等；更有以一爲單位者，如八十一、八十二、八十三、等等。單位愈小，分數愈不可靠。此制之最大的弱點，就是予人以錯誤的準度之印像。換言之，人類的能力實不能區分學生的成績至此種精微的程度。

字母或數字制 因爲鑒於百分制之弊端，有些學校遂採用字母或數字分數制。此制對於學生的成績，不作詳細的區分，祇作優劣之辨別。它所用的代表字或符號，爲最優、優、中、及格、劣；甲、乙、丙、丁、戊；A、B、C、D、E、或1、2、3、4、5。此制之中，又因等級之多寡，分爲若干種，如三級至七級。此制較百分制略優，因爲等級比較含混，容易辨別，爲人力

所能及。然而採用此制時，亦常以百分數為根據，普通的標準如下：

A = 90 - 100

B = 80 - 89

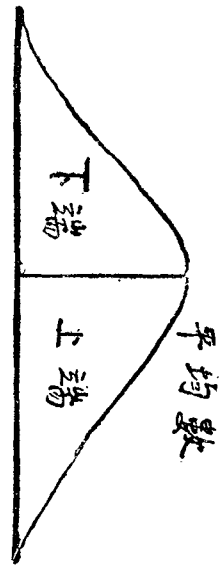
C = 70 - 79

D = 60 - 69

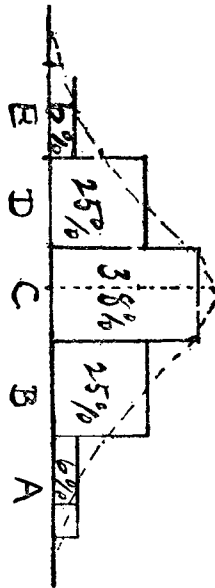
E = 60以下

不過此制若不以百分數為根據，或遇不能根據百分數時，各等級應佔的百分比，則無法分配。

常態曲線制 常態分配曲綫制是一比較新而合理的分數制度。各種調查均表示人類之多數特性的分配，約近於常態分配曲綫。故分數之分配，亦宜以此為根據。此種常態分配曲綫是一鐘形，個體集中於中間，自中間成斜形分配於兩端，距離愈遠，分配愈小，如下圖：



根據常態分配，我們可以計算得各等級分數的學生之百分比，例如五級制的百分比如下：



此種百分比是依據機率原理計算的。如果採用他種分級制，百分比之分配則如下表(第十六表)：

對於某些班次，尤其是小的班次、非平均的組別、或選擇的組別，此法

第十五表 各等級分數應佔百分比

分級數	字 母 等 級						
	A	B	C	D	E	F	G
3	25%	50%	25%				
4	12%	38%	38%	12%			
5	6%	25%	38%	25%	6%		
6	3%	16%	31%	31%	16%	3%	
7	2%	9%	23%	31%	23%	9%	2%

則太機械且武斷，因為它們的分配不一定近於常態。對於此種班次，有幾種救濟的方法：

(一) 規定數種分配制，如一常態分配與數偏斜分配，任教員酌量每次考試的情形，選用其中之任一種。

(二) 定一分數分配之伸縮的限度，使教員得以斟酌當時的情形，自由伸縮，如第十六表。

第十六表 各等級分數的百分比之限度

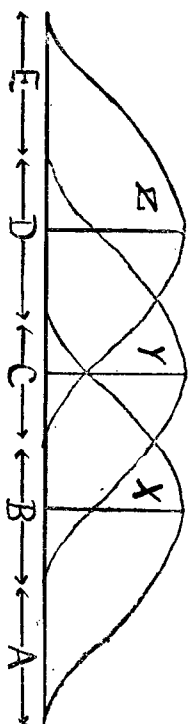
分級數	字 母 分 數						
	A	B	C	D	E	F	G
3	16%—25%	50%—68%	16%—25%				
4	7%—15%	35%—43%	35%—43%	7%—15%			
5	2%—15%	15%—24%	26%—50%	15%—24%	2%—15%		
6	1%—5%	12%—16%	30%—37%	30%—37%	12%—16%	1%—5%	
7	1%—5%	6%—10%	20%—24%	23%—38%	20%—24%	6%—10%	1%—5%

(二) 根據智力測驗的分數之分配。但是智力測驗對於某些科目，並非一良好的指導。

(四) 根據標準測驗的分數之分配。

如果學生是按能力分組，且用五級分數制，分數又如何分配呢？對於此種情形，亦有數種辦法。有人主張仍照樣各給△至□的五級分數。此法似不公平，因為優組得□的甚或優於劣組得△的，但不幸而分入優組，竟至不及

格。有人主張優組祇給A與B，中組僅給B與C，劣組則給D與E。此法未曾計及學生之努力與否，似不妥當。更有人主張照五級制給分，但標明其組別，如 $A_x, B_x, C_x, \dots, A_y, B_y, C_y, \dots, A_z, B_z, C_z, \dots$ 。此法有一困難，即各組需用一不同的測驗或考試，並且轉學時亦感不便。最好的辦法是以分配之重疊現象為根據，例如下圖：



依照此圖，優組可給A、B與C，中組給B、C與D，劣組給C、D與E。

化測驗分數為等級分數法

由測驗分數計算等級分數，必須採用一明確且客觀的方法。茲述二法：
 即等級法與均方差法。

等級法 此法亦稱比例法，是根據常態曲線（機率曲線）的觀念，它先規定

各字母等級應佔的百分比，即依此將測驗分數化成等級分數。例如採用 5-38-25-6 的五級分配制，則最高的百分之六為 A，次百分之二十五為 B，次百分之三十八為 C，次百分之二十五為 D，而最低的百分之六為 E。茲舉例說明如下：

第十七表
化測驗分數為等級分數之等級法

百分比	學生	測驗分數	測驗等級	等級分數
6%	A	40	1	A
25%	B	36	2	B
	C	32	3	
	D	31	4	
	E	30	5	
	F	28	6	
	G	26	7	
	38%	H	25	
I		24	9	
J		23	10	
K		22	11	
L		21	12	
M		20	13	
N		19	14	
O		18	15	
P		17	16	
25%		Q	16	17
	R	15	18	
	S	14	19	
	T	13	20	
	U	12	21	
	V	11	22	
	W	10	23	
6%	W	6	23	E

人數 = 23

$6\% \times 23 = 1.38$ 或 1, $6\% \times 3 = 1.38$ 或 1.

$25\% \times 23 = 5.75$ 或 6.

$38\% \times 23 = 8.7$ 或 9.

$25\% \times 23 = 5.75$ 或 6.

均方差法

此法為化測驗分數成等級分數之一最佳的方法。凡欲精通考試

技術的人，均應明白此法。初看起來，此法對於無統計訓練的人，似很複雜且混亂，其實很易學習。

均方差為教育統計中之一常用的名詞。它代表分數分配於平均數前後之距離。均方差之求法，是平方各分數與平均數之差數，然後將這些平方的差數加起來，以分數總數除之，再開方即得。其公式如下：

$$S.D. = \sqrt{\frac{\text{Sum of } d^2}{N}}$$

S.D. 為均方差，Sum of d^2 為差數平方之總數，N 為分數（或人數）總數。如自假定的平均數計算之，公式如下：

$$S.D. = \sqrt{\frac{\text{Sum of } d^2}{N} - \left(\frac{\text{Sum of } d}{N}\right)^2}$$

茲舉例說明如下：下列分數為一六年級算術真偽測驗之結果。假定平均數為 22。x 行為測驗分數，1 行為分數之次數，d 行為各分數與假定平均數之差數。例如 40 比假定平均數 22 大 18，故作 +18。4 比 22 小 18，故作 -18。1d 行為次數乘差數之積。祇有一人得 40 分，1 × 18 = 18。一人得 32 分，2 × 10 =

第十八表 均方差求法

X	f	d	fd	fd ²
40	1	+18	+18	324
36	1	+14	+14	196
32	2	+10	+20	200
30	1	+8	+8	64
28	1	+6	+6	36
26	1	+4	+4	16
24	1	+2	+2	4
22	4	0	+72	
20	1	-2	-2	4
18	3	-4	-12	48
16	2	-6	-12	72
14	2	-8	-16	128
12	1	-10	-10	100
6	1	-16	-16	256
4	1	-18	-18	324
	23		-86	1572

20; 餘類推。f² 行爲次數乘差數平方之積。例如，+18之平方爲324，以1乘之，仍爲324；+14自乘爲+196，以1乘之，得196；+10乘之，得200；+8乘之，得64；+6乘之，得36；+4乘之，得16；+2乘之，得4；餘類推。此行之簡便求法是以d行之數乘f行之數，結果相同。

$$M = GM + \frac{\text{Sum of } d}{N} = 22 + \frac{-86 - (+72)}{23} = 22$$

$$+ \frac{-14}{23} = 22 + (-.609) = 21.4 \text{ 或 } 21. \quad M = 21.$$

$$S.D. = \sqrt{\frac{\text{Sum of } d^2}{N} - \left(\frac{\text{Sum of } d}{N}\right)^2} = \sqrt{\frac{1572}{23}}$$

$$- \left(\frac{-14}{23}\right)^2 = \sqrt{68.2 - (-.609)^2} = \sqrt{68.2 -$$

$$.36} = \sqrt{67.84} = 8.2 \text{ 或 } 8. \quad S.D. = 8.$$

我們可以看出行與行中，在假定平均數以上者為正的，而在其下者為負的。正的總數為 +72，負的總數為 186。正負相消得 110。行的總數為 1572。代入公式，平均數為 21.4，均方差為 8.29，取整數為 21 與 8。

既已求得平均數與均方差後，即可將測驗分數化成等級分數。取平均數 21 為 C 分數之中點，再根據均方差計算各等級應包括的分數。

以一均方差為單位，C 級的分數應延長至平均數上下各 4 分如下：17, 18, 19, 20, 21, 22, 23, 24, 25。B 級將包括 26, 27, 28, 29, 30, 31, 32, 33。A 級為 34, 35, 36, 37, 38, 39, 40, 41。D 級為 9, 10, 11, 12, 13, 14, 15, 16。而 E 級為 1, 2, 3, 4, 5, 6, 7, 8。簡括之，成表如下：

A = 34 以上

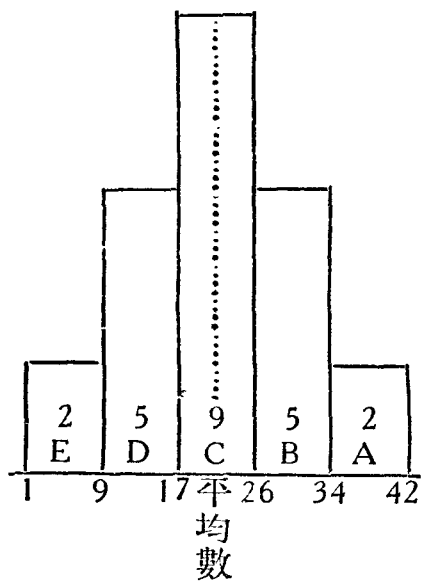
B = 26 — 33

C = 17 — 25

D = 9 — 16

E=9以下

依此標準，化上述測驗分數爲等級分數如下圖：



考
試
新
論

附錄一

譯名對照表

一劃

一字答測驗 Single-answer or simple-recall test

二劃

二答測驗 Alternative-response test

三劃

口述法 Dictation method

四劃

分配 Distribution

心向 Mental set

比例法 Proportion method

附 錄

分數制度 Grading or marking system

內包取樣 Intensive sampling

反面暗示的影響 Negative suggestion effect

五劃

用途 Utility

平均數 Average or mean or M.

平均差 Average or mean deviation or A. D.

主觀性 Subjectivity

可靠性 Reliability

可靠係數 Reliability coefficient

正誤測驗 Alternative or right-wrong test

外延取樣 Extensive sampling

六劃

次數與次數分配 Frequency and frequency distribution

考試與考試制度 Examination and examination system

印刷法 Printing or mimeographing method

全距離 Range

同異測驗 Same-opposite test

百分分數制 Percentage grading system

七劃

成績 Accomplishment

作弊 Bluffing

均方差 Standard deviation or S. D.

均衡法 Weighing

八劃

取樣 Sampling

附 錄

非正式考試或測驗 Informal examination or test

非標準測驗 Unstandardized test

九劃

計時 Timing (of test)

客觀 Objective

客觀性 Objectivity

便易性 Facility

相關 Correlation

相關係數 Correlation coefficient

訂誤測驗 Correction-of-error or incorrect statements test

指認測驗 Identification test

重組測驗 Rearrangement test

追憶測驗 Recall test

是非測驗 Yes-no test

客觀考試或測驗 Objective examination or test

十劃

差異 Difference or variation

時限 Time limit

特效 Validity

連續測驗 Continuity test

個人差律 Personal equation

真偽測驗 True-False test

特殊決定原素 Specific determiners

十一劃

猜度 Guessing

常模 Norm

偏斜分配 Skew distribution

假定平均數 Assured average or mean

常態分配曲線 Normal distribution curve

十二劃

強記 Cramming

診斷 Diagnosis

測量 Measure or measurement

測驗 Test

測驗式 Test Form

黑板法 Blackboard method

替換式 Comparable or duplicate or equivalent form

等級法 Ranking method

等級分數 Grades

等級次序 Rank order

順序分配 Order distribution

答案標準 Answer keys or stencils

集中趨勢 Central tendency

短答測驗 Short-answer test

測驗問題 Test item or exercise

測驗分數 Test score

測驗之施行 Administration of test

測驗之編造 Construction of test

測驗之校閱 Scoring of test

最好答案測驗 Best-answer test

十三劃

準度 Accuracy

彙選測驗 Batteries or teams of test or Battery test

填字測驗 Completion test

新法考試或測驗 New Type examination or test

十四劃

說明 Direction or instruction

廣博性 Comprehensiveness

綜合測驗 Combination or composite test

選答測驗 Multiple-answer or multiple-choice or multiple-response test

認識測驗 Recognition test

對偶測驗 Matching test

對減錯法 Right-minus-wrong method or formula

十五劃

標準 Criterion or standard

標準測驗 Standard or standardized test

論文考試 Essay examination

隨機取樣 Random sampling

複對偶測驗 Compound matching test

複選答測驗 Compound multiple-answer test

諸選答測驗 Plural multiple-answer test

十六劃

機遇 Chance

融洽性 Rapport

十七劃

點分數 Point score

舊式考試 Traditional or discussion examination

十九劃

附錄

考試新論

難度 Difficulty

類比測驗 Analogies test

參 考 書 目

關於新法考試之著作，我國尙無出版者，即短篇文字亦頗少見，故本書目中均爲西文書籍。（中國測驗學會之測驗雜誌將發行新法考試專號，但是現在尙未出版，將來讀者可以參考。）

本書目僅列有關於新法考試之著作，至於關於一般的教育測驗與統計之專籍，則均未列入。所舉書目雖不完備，但除散見於報章雜誌之短文外，主要著作都已錄入。

Ballard, P. B., *The New Examiner*, Hodder & Stoughton, London, 1924.

Bardy, J., *An Investigation of the written Examination As a Measure of Achievement with Particular Reference to General Science*, University of Penn., Philadelphia, 1923.

Prinkley, S. G., *Values of New-Type Examinations in the High School with Special Reference to History*, Teachers College Contributions to Education, No. 161, Columbia

ia University, N. Y., 1924.

Bursch, J. F., & Melzer, H, The New Examination; Its Construction and Use, School of Vocational Education, Oregon State Agricultural College Bulletin, No. 422, Oregon State Agricultural College, Corvallis, Oregon, 1926.

Butler, W. F., The Value of Informal Tests in Supervision, First Yearbook of the Dept. of Elementary School Principals, National Education Assn. of the United States, Washington, 1922.

Chapman, J. C., Trade Tests, Henry Holt & Co., N. Y., 1921.

Douglass, H. R., Modern Methods in High School Teaching, Houghton Mifflin Co., Boston, 1926.

Ellis, R. E., Standardizing Teachers' Examinations and the Distribution of Class Marks, The Public School Publishing Co., Bloomington, Ill., 1927.

Fenton, N, & Worcester, D. A., An Introduction to Educational Measurements, Ginn & Co., Boston. 1928.

Filer, H. A, & O'Rourke, L. J. Annual Reports of the Chief Examiner and Director of

- Research of the U. S. Civil Service Commission for the Fiscal Year Ended June 30, 1923, Government Printing Office, Washington, 1923.
- Finkelstein, I. E., The Marking System in Theory and Practice, Educational Psychology Monographs, No. 10, Warwick & York, Baltimore, 1913.
- Fiske, T. S., Annual Reports of the Secretary of the College Entrance Examination Board, 1921-24. 431 W. 117th St., N. Y.
- Gray, W. S., et al., Informal Reading Tests, 24th Yearbook of the National Society for the Study of Education, Part I, Public School Publishing Co, Bloomington, Ill., 1925
- Greene, C. E. New-Type Tests, Research Monograph, No. 3, Public Schools Denver, 1926.
- Hatch, R. W., How Can Examinations in History Be Improved?, Horrace Mann Studies in Education, Columbia University, N. Y., 1922.
- Hayhurst, E. R. How to Write an Examiner, Ohio State University, Cooperative Supply Co., Columbus, Ohio, 1922.
- Hopkins, L. T., The Construction and Use of Objective Examinations, University of Colorado, Boulder, 1926.

- Kelly, F. J., *Teachers' Marks: Their Variability and Standardization*, Teachers College Contributions to Education, No. 66, Columbia University, N. Y., 1914.
- Lang, A. R., *Modern Methods in Written Examinations*, Houghton Mifflin Co., Boston, 1930
- Mazah, W. R., et al, *Report of the Commission on New Types of Examinations, etc.*, The College Entrance Examination Board, N. Y., 1923.
- McCall, W. A., *How to Measure in Education*, The Mcmillan Co., N. Y., 1927,
- Miller, C. D., *Objective Tests in High School Subjects*, Published by the author, Norman, Oklahoma, 1926,
- Monroe, W. S., *Directing Learning in the High School*, Double-Day, Page & Co., Garden City, 1927.
-, *Written Examinations and Their Improvement*, Univ. of Ill. Bulletin, Vol 20, No. 7, Bureau of Educational Research Bulletin, No 9, Univ. of Ill., Urbana, 1922.
-, *DeVoss, J. C., & Kelly, F. J., Educational Tests and Measurements*, Houghton Mifflin Co., Boston, Revised Edition 1924.

-, and **Sonders, L. B.**, *The Present Status of Written Examinations and Suggestions for their Improvement*, Univ. of Ill. Bulletin, vol. 21, No. 13, Bureau of Educational Research Bulletin, No. 17, Univ. of Ill., Urbana, 1923.
-, & **Carter, R. E.**, *The Use of Different kinds of Thought Questions in Secondary Schools and Their Relative Difficulty for Students*, Univ. of Ill. Bulletin, vol. 20, No. 34, Bureau of Educational Research Bulletin, No. 14, Univ. of Ill., Urbana 1923.
- Odell, C. W.**, *Traditional Examinations and New-Type Test*, Century Co., N. Y., 1928.
-, *Objective Measurement of Information*, Bureau of Educational Research, Univ. of Ill., Circular No. 44, 1926.
- Orleans, J. S.**, *Manual on the Local Construction and Uses of Objective Tests*, Univ. of the State of New York Bulletin, No. 893, 1928.
-, & **Sealy, G. A.**, *Objective Tests*, World Book Co., Yorkers, 1928.
- Paterson, D. G.**, *The Preparation and use of New-Type Examinations*, World Book Co., Yorkers, 1925.

- Rush, G. M., *The Improvement of the Written Examination*, Scott, Foresman & Co., Chicago, Chicago, 1924.
-, *The Objective or New-Type Examination*, Scott, Foresman, & C., Chicago, 1929.
-, et al, *Objective Examination Methods in the Social Studies*, Scott, Foresman & Co., Chicago, 1926
-, & Stoddard, G. D., *Tests and Measurements in High School Instruction*, World Book Co., Yonkers, 1927.
- Rugg, H O., *A Primer of Graphics and Statistics*, Houghton Mifflin & Co., Boston, 1925
- Russell, C, *Classroom Tests*, Ginn & Co., Boston, 1926.
- Spence, R., B., *The Improvement of College Marking Systems. Teachers College 'Contributions to Education*, No. 252, Columbia Univ., N. Y., 1927.
- Stormzand, M. J., *American History Teaching and Testing*, The Macmillan Co., N. Y., 1926
- Strickland, V. L., *Objective Tests*, Kansas State Agricultural College, Manhattan, Kansas, *Bulletin XVIII*, No. 2.
- Symonds, P. M., *Measurement in Secondary Education*, The Macmillan Co., N. Y., 1927.

- Topp, H. A, *Trade Tests In Education, Teachers College Contributions to Education*, No. 11
5, Columbia Univ., N. Y., 1921.
- Weidemann, C. C., *How to Construct the True-False Examination, Teacher College Contributions to Education*, No. 225, Columbia Univ, N. Y., 1926.
- Whitten, C. W., *Report on Standardizing Teachers Marks, 6th Yearbook of the National Association of Secondary School Principals*, The George Banta Publishing Co., Menasha, Wisconsin, 1922.
- Wilson, G. M., & Hoke, K J, *How to Measure, The Macmillan Co., N. Y.*, 1928. (Revised Edition)
- Wood, B D, *Measurement in Higher Education, World Book Co., Yonkers, 1923.*
....., *New York Experiments With New-Type Modern Language Tests, The Macmillan Co., N. Y., 1927.*