

新法考試

校 林初書王 譯 人宗漪浦
初 洪 夏 宗 明 黃

行 印 局 書 中 正

5213
518
2

黃浦
明漪
宗人
編譯

新
法
考
試

正中書局印行

原序

考試在教育實施中要算是最古和最重要的了。牠是學校工作之一，具有各種重要的功用，如決定教學方法的成敗，供給成業的標準，和激起優良的學校作業等等。牠不僅是有用的必需的教室中的臂助，而且對於教師的工作有密切的關係，所以任何有關考試的討論，都能引起他們極注意的反應。凡進步的教師，對於考試的問題，都很關心。

考試對於教師及其工作雖早已發生直接的關係，可是我們對其地位和改進的注意，還是較近的事。考試的方法，尚不能同其他教育方法和實施並駕齊驅，這是很明顯的。現在雖已有許多知識可應用於測驗，但大多數的教師自編測驗，恐怕依然是舊式的。近幾年來，討論考試方法的專書和手冊，已出版了五六種。討論考試的期刊，也日漸增加了。師資訓練機關中，現在也開着測驗方法的學程了。教師如無新法考試的知識，再也不能予以原諒的了。

本書的目的，在用作師範生考試方法等科目的教本，同時並欲供給在職教師及其他欲知新式測驗的編製和應用的人的閱覽。教師的需要，著者是始終關懷着的。討論務求簡單而切於實際。理論的和複雜的統計方法，都設法避免。本書是著者為許多在職教師及師範生所開的一個學程的講稿，內容尚覺通俗而有益。各章的

取材和排列都與教師當前的重大問題相符。著者相信，本書對於教師及其他有應用機會的人們，是很有價值的，因為現在每一教師都應熟知各種新的測驗方法。

在每章之末，附有若干習題，以資溫習和討論。這種計劃與他書不同。習題都採取正誤語句的形式，前面冠以「+」「-」「？」的記號，以代表正確，錯誤，與疑惑不決。希望讀者能勝任愉快地把每句加以記號，且證明其記號為無誤。著者曾應用此法於好幾種學科，且知其能供給極良好的複習和引起極生動的討論。（下略）

亞爾保德·蘭

夏序

教育科學化的運動在中國已有十幾年的歷史，在這十幾年中比較成熟的是測驗運動。我們用技術眼光診斷中國教育的實施，不能不說是幼稚。建築設備，在教費困難的情況下，固然不能談科學化；如課程，行政方法，事務手續，皆無需和經費問題發生絕對關係的，依然是墨守成規。甚至學校圖書館，還在應用舊式目錄；學校賬目，還多半用舊法的簿冊。教育科學化這樣的不進步，自然影響測驗運動不少。所可慶幸者，測驗在師範學校課程中的位置的穩定了；測驗法比較一切主觀考試法為可靠已無人懷疑了。在教育科學化運動最可樂觀最有成功可能的或是測驗運動。

國內出版界不乏關於心理或教育測驗的專書，但求其淺顯易解而又能明其所以然足以適應教師需要的還不多。浦黃兩君最近譯著的新法考試一書，可以彌補這個缺點。本書的特色是意義不求深奧，材料不取冗煩，而測驗的理論却面面俱到。兩君執筆時，於一字一句發生疑問，不惜各方就商詳細推敲，可見確能忠於其職。雖然本書曾經雜誌的介紹，或著作者的引用，但兩君的譯本是包括本書的全文。在測驗運動正在趨向復興的時候，不能不算是一種貢獻。

二十四年四月夏承楓序於南京

新法考試

杜序

考試的功用甚爲廣博。單就學校方面而言，我們分組、分班、及分年級，均非應用考試不可。他若測量學習的成績和評斷教學的效率等，尤須以考試爲主要的方法。此外，如診察學習上的困難，以爲實施指導的參考，研究教學上的問題，以爲改進教育的張本，亦均賴考試之適當的利用。新近在政界、軍界、工界、或商界各方面，應用考試來選擇人材的，一天普遍一天，而且甚富成效。可見考試不特爲解決各種教育問題的必要工具，且是一種推進各種社會事業的良好法則。

可是考試的功用之究能充分表現與否，又視其方法之是否適當以爲斷。若是一種考試，僅憑主觀的意見，出題四五，以爲測量全部學識的根據，其所得的成績，必不十分可靠。再者，記分既無客觀的標準，評斷又乏具體的量表，則其不公平處，亦必難於幸免。故考試固當重視，其方法尤宜研究。我個人有感於這種工作的重要，屢欲編述考試與測驗一書，以爲研究本問題者的參考。但因課務繁劇，遲延數年，尙未着手；言之實非常慚愧！今同學黃明宗、浦漪人二君課暇譯成蘭(J. Lang)氏新著新法考試一書，我閱讀之餘，如卸自己的重負，不勝喜慰之至。蘭氏爲教育測驗專家，其書內容豐富，說理詳盡。首述考試的歷史和功用；次述優良考試的特徵，和傳統考試與新式考試的同異；又次，分述各種新式測驗的方法；最後，討論新式測驗的應用，及施行測驗、記分和分列等第等

問題。舉例完備妥適，極合實際應用。倘依之以改良考試的方法，其成績必有可觀者。至於黃浦二君的譯文，不特忠實，而且暢達，誠可謂名著佳譯。將來書成，其有貢獻於社會各界，必是不可限量的；故余甚樂爲之序。

杜佐周序於廈門大學。

民國廿四年，一月，十六日。

王序

自有教育以來，就有考試的制度。不過舊法考試，即所謂論文式的考試，有許多缺點，其結果之不足信，已有許多實驗證明。因此近來有人主張取消考試。其實考試的流弊是因為考試的方法之不良，而考試制度在教育上仍占有重要的地位，所以大多數教育家之主張，乃是改良考試之方法。

標準的教育測驗是應此種主張與需要而起的。但是標準測驗在應用方面有許多限制。其最重要的一點為測驗之內容常不能符合教師當時教學的情形。新法考試一方可以避免舊法考試之流弊，一方可以補測驗方法之不足，是教師所必須具有的知識。黃明宗君與浦潑人君自入中大以來，從事於教育的科學之研究，已四年於茲，深知新法考試之重要與國中教師對於此種知識之缺乏，乃合譯美國教授 (Ablett R. Janes) 所著之新法考試 (Modern Methods in Written Examination) 以資提倡。此書對於各種考試方法，均有詳細說明，正如客柏來教授 (Ellwood P. Cubberley) 所云，是「一本很可用的教科書」。譯稿全部，書林幸得展閱，覺兩君對於文字內容兩方面，都有充分的了解，譯筆不但忠實，且頗流利，深為可賀。

王書林序於南京中央大學

廿四年四月十一日

新
法
考
試

譯者言

本書原名 *Modern Methods in Written Examination* 係亞爾保德·蘭 (Albert R. Lang) 所著。氏爲教育測驗專家，在加利福尼亞 (California) 州立師範大學 (State Teachers College) 教育學院 (School of Education) 担任院長。

每章譯完後，均由譯者互相校閱。校閱時不但注意於錯誤之改正，尤力求行文之一致。

原書舉例，間有與國情不合處，均由譯者設法改易。凡改易之例，均有「キ」爲記，以資識別。

凡原書不易明瞭之處，均由譯者詳加註釋，（標明譯註字樣）以求節省讀者時間與精力。

本書譯成後，承王書林夏滿初兩先生詳爲校閱，杜紀堂先生亦予以不少指正，不勝感謝。付梓時，又蒙同學姜光漢張崇業施懿德諸君閱讀一過，並給予許多幫助，亦於此誌謝。

譯者識

新法考試目次

第一章 考試的歷史……………一

考試的起源——中古時代大學的影響——一八四五年的波士頓考試——荷斯曼論文字考試——智力測驗的發展——標準測驗的發展——新法考試的興起——考試的重要。

溫習與討論問題

第二章 考試的功用……………一五

考試的功用——測驗知識的保持——成績的決定——日常工作的刺激——溫習的動機——供給客觀的標準——教學效率的測量——教學效率的改進——特殊困難的診斷——智能的培養

溫習與討論問題

第三章 優良考試的特徵……………三三

優良考試——正確性——可靠性——客觀性——廣博性——便利性——實用性——融洽

性。

溫習與討論問題

第四章 傳統的論文式考試……………四五

論文式考試的永久地位——作文的正確測量——編造和施行的便利——對於優等生的價值——論文式考試的缺點——論文式考試的主觀記分——論文式考試的問題範圍太隘——記分上不和干的因子——學生的反應——結果的用處太隘——論文式考試的改良方法。

溫習與討論問題

第五章 新式考試……………六一

新式測驗的主要特徵——客觀測驗的普通形式——測驗形式的選用——客觀測驗的特殊價值——學生爲什麼願受新式測驗——推理的測驗——新式測驗中的作弊。

溫習與討論問題

第六章 正誤測驗……………七七

正誤測驗——正誤測驗的優點——刺激適當的心理作用與態度——猜度的因子——證明

的方法——錯誤語句的影響——編造正誤測驗的技術——編製正確的語句——正確的語句改爲錯誤的語句——規定學生做記號的方法——正誤測驗的記分——施行正誤測驗的方法——應用正誤測驗的結果

溫習與討論問題

第七章 選答測驗與對偶測驗 九九

選答測驗——選答測驗的優點——選答測驗的批評——編製選答測驗的技術——正確語句的編製和排列——易混答案的擬定——學生做記號的規定——選答測驗的記分——施行選答測驗的方法——選答測驗的變化——對偶測驗——編造對偶測驗的技術——正確語句的編製——分語句爲兩部——學生做記號的規定——對偶測驗的施行和記分——對偶測驗的變化——選答測驗和對偶測驗的結果之應用

溫習與討論問題

第八章 回憶測驗 一一三

回憶測驗——簡答測驗——編製簡答測驗的技術——簡答測驗的施行與記分——簡答測驗的變化——填字測驗——填字測驗的價值——編造填字測驗的技術——測驗材料的準

備——選擇空缺的答案字——規定學生填寫的方法——填字測驗的施行與記分——應用
回憶測驗的結果

溫習與討論問題

第九章 綜合測驗與雜式測驗……………一四五

綜合測驗——綜合測驗的優點——編造綜合測驗的技術——編造正確的語句——將各種
語句改成適宜的測驗形式——規定學生做記號的方法——規定測驗報告的形式——雜式
測驗——重組測驗——指認測驗——證明測驗——其他的雜式測驗——應用綜合測驗與
雜式測驗的結果

溫習與討論問題

第十章 新式測驗在各種學科上的應用……………一六九

新式測驗在各學科上的應用——讀法——國語——算學——社會科學——自然科學——
外國語——其他各科——高等學府中的新式測驗

溫習與討論問題

第十一章 測驗分數化為等第的方法……………一二九

分數與等第的區別——分數制度——常態分配曲線的應用——次數分配——化測驗分數為等第的方法——標準差法——標準差法舉例——化分數為等第——分數的分組計算——比例法——成績分等的相對次序。

溫習與討論問題

第十一章 測驗的施行問題 一四五

測驗的施行問題——教育行政機關舉行的考試——教師參預普通考試工作——給與教師局部應用測驗的便利——考試的次數時間與排列——通知與不通知的測驗——開書測驗——口述測驗——能力分組在測驗結果上的影響——作弊與無監考制——考試的免除——考試分數對於最後等第的影響——測驗結果應求最大應用。

溫習與討論問題

新法考試

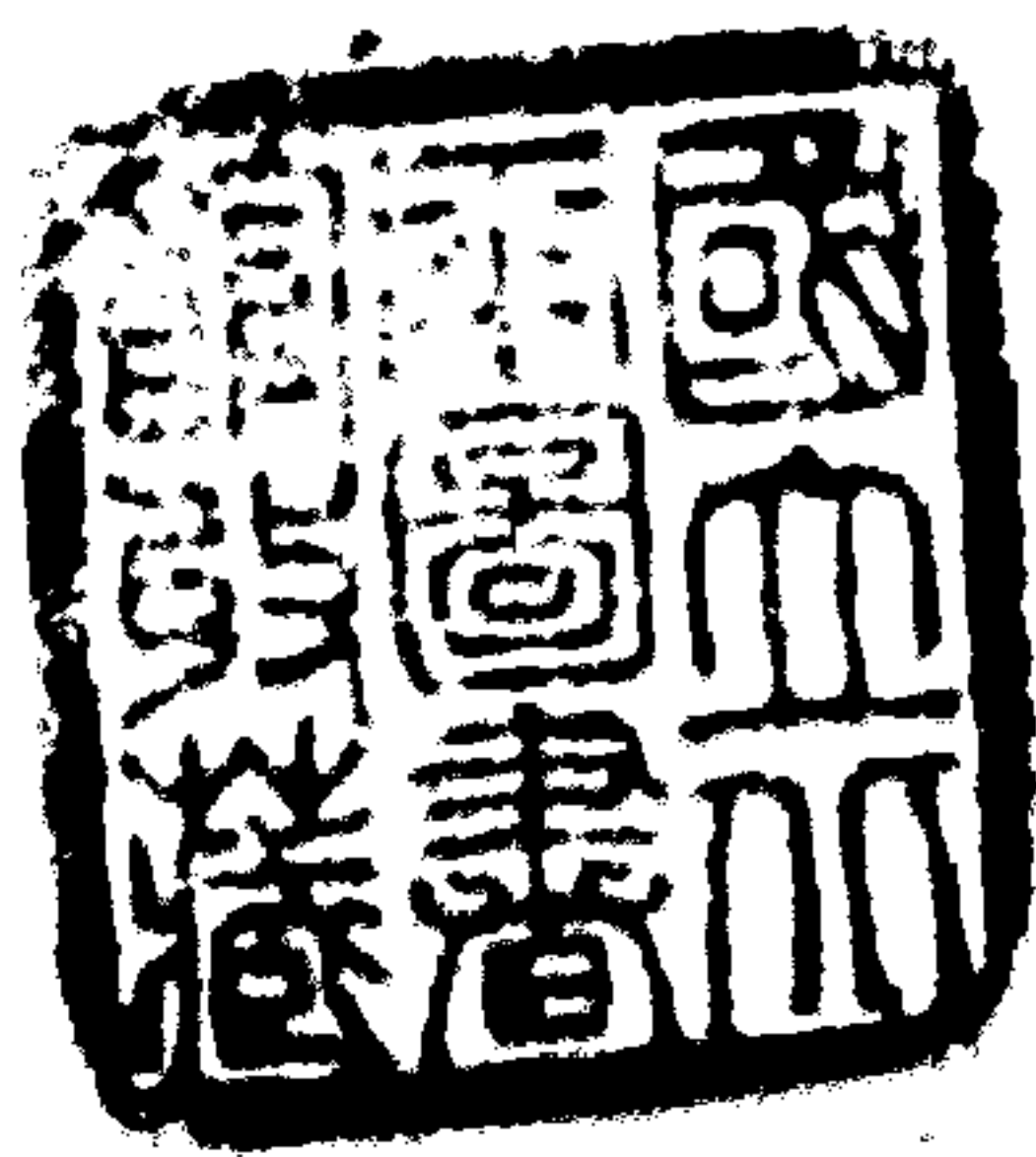
★

新法考試

第一章 考試的歷史

考試的起源——考試的山來已久，牠到底創始於什麼時期，什麼地方，已無從稽考的了。牠的起源在古代時候，早已失傳了。我們在舊約全書中，曾讀着一種考試，當以法蓮人（Ephraimites）要渡過約但河（Jordan）的時候，基列人（Gileadites）便要考驗他們讀“Shibboleth”那個字的能力，倘使以法蓮人不能發“sh”那個音，而只會說“Sibboleth”，便將其就地處死。在最凶惡的那天，據說有四萬二千個人考試都是失敗的。

這個悲慘的試驗當然不好算考試的先例。遠在紀元前二千二百年的時候，中國已有精密的國家考試制度了，其目的在考選官吏。上古時代初民的冠禮，已經含有心理特質和身體特質的測驗在內了。他們以為必須通過了這種測驗，才可擔當成人的責任。在紀元前約五百年的時候，雅典人和斯巴達人已舉行各種定期考試，以測驗青年的技能和能量。斯巴達人舉行種種正式的考試以測驗忍苦的能力，其方式是在阿提密斯·奧西亞（Antemis Orthia）的祭壇前執行鞭笞，蘇格獵底最著名的教學法，便是使他的學生回答彈智竭慮的問



題。古代基督教徒的問答教學法，一問一答，是一種連續的考試方法。所以考試自上古發明以後，其間經歷不少的年代，演進不已，一直保留到現在。

中古時代大學的影響——中古時代的大學，有許多方面為近代大學所取法的，其中考試的觀念要算是很顯著的一種了。最早的大學不過是有志於教學的人所組織的公會或團體。這種教學公會與當時各種職業公會一樣，分為學徒 (Apprentice) 職工 (Journeyman) 和工頭 (Master) 三個階級。從下一個階段晉陞到上一個階段，必須經過一種考試。學徒期終了的時候，給以一種考試，以測驗其解釋及說明各種名詞的能力，及格的授以學士的學位。職工期終了的時候，也給以一種考試，以測驗其對於一種論題的討論與辯駁的能力，及格的授以碩士或博士的學位。波倫亞大學 (Bologna) 早在 1119 年已有這種考試的記錄。十三世紀之末，巴黎大學也有了。廓美紐斯 (Comenius) 的大教授法 (Great Didactic) 中所計劃的大學是僅僅訓練那些考試及格的人的。

古代大學中的考試，僅是一種口試。大學中的文字考試 (Written examination) 似乎到了十八世紀初葉才發達的。但在中國則早在第七或第八世紀已有文字考試了。大學中第一次舉行文字考試大約是在一七〇二年；開這風氣之先的實為英國的劍橋大學。後來隔了不久，牛津大學以及其他許多英國的學校便都相繼仿效了。當清教徒遠赴美洲的時候，英國的考試制度已斐然可觀了。所以中古大學以考試制度貢獻於西方

文明，似乎是毫無疑義的了。

一八四五年的波士頓考試——在美國史上第一個考試，是一八四五年的波士頓（Boston）考試。在那時波士頓學校委員會有一種任務，便是每年須舉行一次正式的學校考查，這種考查包括一種學生的口試在內。嗣後學校漸漸發達，學生漸漸增加，委員會要想對於學生逐一考驗，勢所難能。所以便採取了一種僅考各校第一班（最高班）的計劃。但因學校繼續的發達，連這個辦法也變得不可能了。結果這種每年的學校考查，終於變為敷衍的和形式的例行公事，而無甚意義了。到了一八四五年，學校委員會決定組織一種副委員會（Sub-Committee），負責辦理波士頓各學校高級部的年考。副委員會進行這種工作時，十分澈底，十分嚴格。牠決定舉行一種文字考試，編印了許多的問題，凡校中所授科目，無不包括在內。問題所以要這樣決定，為的是要考試益臻公允，給各個兒童以均等的機會；免除有暗示性的問題；留一個永久的考試記錄；斷定兒童何者未知；何者已知；考驗兒童對於文字發表的敏捷程度；並測驗造句、標點和拼法的能力。這種考試包含天文、定義、地理、文法、歷史及自然哲學等科。在那時候波士頓共有十九個公立高級小學（Grammar School）這種考試只施諸各校的最高級，共有五百三十個兒童；在那時已佔波士頓學校登記學生數的百分之七以上了。學生的平均年齡約在十四歲左右，這個考試共有一百五十四個印就的問題，其中關於文法者十四題；關於自然哲學者二十題；關於其他各科者各三十題。這樣看來，波士頓一八四五年的考試實為一種包羅極廣的文字考試，那

是很顯然的了。

波士頓考試的施行手續是規定得極謹慎的。他們謹防任何學校探聽問題的性質，所以問題的編製和印刷都是很秘密的。這種秘密的計劃只有三個委員去執行。早晨八點鐘，他們毫無聲息的突然到各校去舉行考試。在考試以前，他們把學生的書籍和參考材料統統收了起來，令學生的坐位彼此相隔很遠，以防互通聲息；通知學生只有一小時的考試時間；再告訴他們不必爲了書法的美觀而多費時間，因爲書法和記分是無關的；然後分發當天所要測驗的問題。一切手續都完備了，才叫學生開始工作。一小時後，主試的委員便收集了試卷，再到第二個學校去，在那裏也舉行一小時的考試；然後再至第三個學校；在上午一共要考三個學校。到了下午，每個委員再去考另外三個學校；這樣，他們於一天以內得在各校考完一種科目。第二天，他們帶着另一學科的問題，以同樣的方法，繼續舉行考試，直至各校的考試程序完全完成爲止。

關於問題的編製、記分，以及結果的解釋，委員會都加以慎重的考慮。在籌備考試的時候，各種問題都是照難度而排列的，其中有極簡單的，就是最笨的兒童也能明瞭，也能解答；也有極難的，以便試驗優異兒童的能力。關於記分方面，預先定了記分的規則。這些規則雖然不一定嚴格的應用，却是一種很好的南針。他們對於疑難是很注意的，凡問題因懷疑而不答的，比瞎猜而答錯的，在記分上便宜得多。根據這種考試的結果，委員會便依照答對題數的多寡評判成績，以定學校等級的高下。委員會對於學校分等的說明，對於考試結果所抱之信心，

都是很保守的 (Conservative)。他們以爲他們分等的根據，是學生在考試上所表示的智力的大概情形，而非學校的絕對成績。他們相信考試所測量的，僅爲知識上的活動和成就，如果其他的特質像良心、守秩序、宗教意識等等都加入計算，或許可以改變等級的秩序。

委員會根據考試所得編成各種說明和觀察，爲現代教師極有趣極有益的讀物。考特韋爾 (Carpwell) 柯蒂斯 (Curtis) (註1) 二氏曾將高級小學委員會的報告，全部重印一過。我們對於一八四五年波士頓人的各種敘述，不僅要驚異莫名，並且要感動不已。這個波士頓的考試，其完美和精密，在美國教育史上實爲一件最堪注意的事情。

荷勒斯曼論文字考試——一八四五年的波士頓考試給荷勒斯曼 (Horace Mann) 以極深刻的印象。在那個時候，氏任麻省教育局總幹事。我們讀了高級小學委員會的報告，立刻便會想到荷氏對於考試的計劃和報告的編造是極有關係的。無論如何，他對於此次考試是極有興趣的，而且相信在教育測量上開了一個新紀元。當氏任普通學校雜誌 (The Common School Journal) 編輯的時候，曾把高級小學委員會的報告擇要發表，並加以討論，後來又把考試的題目刊入。他所選擇的要點和所下的批評，都載於考特韋爾和柯蒂斯合著的教育的過去與現在 (Then and Now in Education) (1845—1923) 一書中。這些偉論是新鮮的，是增益知識的，亦即荷勒斯曼氏成爲教育行政名家的原因所在。

荷勒斯曼在他的批評中，以為文字考試較口試有明顯的優點，簡括之，其理由如下：

(1) 文字考試是公允的——牠不但給全班學生以同一的問題，並給各校同年級的學生以同一的問題。口試則不然，對不同的學生提出不同的問題，欲求各個問題之難易相等是不可能的。再者，給學生以許多問題足以增加測驗其能力的機會，這也是很明顯的。『每個問題，就是一個部分的測驗，所以問題愈多，測驗愈為完善。』

(2) 文字考試對於學生是比較公平的——在三十個兒童的班上，如舉行一小時的口試，每個兒童只不過分到二分鐘罷了。倘使是一小時的文字考試，則三十個學生每人都有六十分鐘了，不會引起驚慌而致遭失敗了。

(3) 文字考試比較完善——以二三個問題去考試兒童，正如寓言中斯各拉斯各(Scholarthikos)的計劃一樣，他想賣房屋，卻帶着一塊磚頭走到街上作為樣子。審慎的考試者和審慎的主顧一樣，當然要希望獲得充分的了解。

(4) 文字考試可以防止教師無謂的干涉——教師插入暗示或暗示性的語句，足以破壞問題的整

(註一)這裏的材料與下面的資料係採自 O. W. Caldwell 與 S. A. Curtis, *Then and Now in Education*, 1845-1923, 於 1923 年

由 World Book Company 出版。

個目的。至於文字試驗，在主試者發問題的時候，教師們不得不袖手旁觀。

(5) 文字考試可以決定教學效率。——學生是否受到真誠的完善教學，可以從文字考試中看出來。文字考試能夠表示教學的性質，和其能夠表示學生造詣的程度一樣的明確。

(6) 文字考試可以防止偏袒。——每個學生回答同樣的問題，不但可以免除偏袒，而且可以使人不致懷疑有無偏袒。

(7) 文字考試的結果對於大眾較為有用。——口試僅給在場的人以一種真確的觀念；而印就的問題和寫就的答案，其結果可以公諸大眾，並可留有永久的記錄。用口試方法，不同的委員，其優良的標準亦各不同，『如果各人的尺都作十二吋長，那麼究竟用那一個人的尺來量呢？這就成爲一個重要的問題了。不同的人心中有不同的判斷標準，其情形同尺一樣，究竟用那個的標準好呢？』

(8) 文字考試可以表示問題的難易。——印就的問題，其難易程度可以一望而知。如果問題太難了，在失敗的時候，可以予以公平的處置。倘使學生對於書中的事實，回答得敏捷而準確，而對於那些關於『原理的應用和關聯』的問題，不能回答，那麼其答便在教師了。這種重要的知識只有在考查問題時才能得到；換句話說，只有在採用文字考試時才能得到。

荷勒斯曼最後的結論，以爲文字考試的優於口試，已至爲明顯，所以任何學校委員會均不應再『貿然採

用以前不適當的和不確切的方法了。『不過對當時辦學者而言，他太超時代了，所以他的預測無從證實。』波士頓的試驗，僅為歷史上的陳跡，對於當時一般的實際情形，並沒有多大的影響。

智力測驗的發展——我們現在所知道的智力測驗，比較上還是一種新的發展。這種測驗是由於個別差異的實驗研究而發明的，在英國高爾登（Galton）於一八六九年就研究個別差異了。此後約五十年光景，在美國有加推爾（Cattell）和桑代克（Thorndike）二氏首先注意到這個問題。至於個別差異的測驗，乃為最近四十年來的事。在開首的十年左右，這種測量工作大部還只限於研究大學生的差異。從事這種工作的為各大學的心理實驗室。學校採用智力測驗，大約是在二十五年之前。創始這種工作的是法國心理學家比納（Alfred Binet）。他在與另一法人西蒙（Simon）合作以前，從事研究這種實際的心理測驗，已有十五年之久了。到一九〇八年，他們合編了一個比納——西蒙量表，用以測量學生的智力。這個量表曾在一九一一年加以修正，其後傳入德國、比國和美國。這個量表譯成英文後，高達德博士（Dr. H. H. Goddard）把牠在美國紐澤塞省（New Jersey）外蘭因（Vineland）地方的低能訓練院中，首先應用。在初行的時候，這種測驗大抵只用於低能訓練機關、監獄、感化學校和青年法庭等等，對於美國的一般兒童並不通用的。迨一九一三年斯旦福（Stanford）大學的推孟博士（Dr. Lewis M. Terman）始將比納量表加以修正和擴充。這種修正量表自一九一六年由推孟出版後，應用極廣。以上所討論的是一種個別測驗。

當世界大戰的時候，團體測驗就應運而生，用以測驗二百餘萬新兵的心理能力。在大戰前數年中，關於團體智力測驗方面已做過好幾種工作了，其最足注意的便是歐蒂斯博士 (Dr. A. S. Otis) 的工作。軍隊測驗對於智力測驗的推行，助力極大。現在社會上已經有許多不同的測驗了，自幼稚園起至大學止，無不具備。庫爾門氏 (Kuhlmann) 延長了量表的低步組，所以三個月的兒童也可應用測驗以測量其心理能力了。在今日的學校中，心理測驗已變為司空見慣的事了。

標準測驗的發展——就我們的記錄看來，我們現在所知道的標準客觀測驗，其創始者是英國的一個小學校長費許 (Fisher) 牧師。他於一八六四年編了一本『量書』 (Scale book)，中有各科的成績，都是依照內容的優劣而定等第的高下。凡作文、圖畫、法語、文法、歷史、智識、算術、航行、實際科學、聖經、拼法和習字等量表，無不兼收並蓄。等第的排列是依成績的次序的，最高者為『一』，最低者為『五』，而每個等第之中尚有四個中間等第。拼法上的等第是根據默句中錯誤百分比而定的。習字呢，則用了許多性質不同的樣本，同今日所用的差不多。至於其他各科，都備有代表的問題，對於每個正確的答案都規定了應得的等第。據費許氏自己的意見，這本『量書』有下列二個優點：(a) 評判考試成績有一定的標準；(b) 把各等第總加起來，容易算出最後的分數。可惜他這種標準測驗太超時代了，在當時並沒有什麼影響。這個測驗和一八四五年的波士頓考試一樣，祇可為以後測驗運動的先導罷了。

就美國言，標準測驗的開始採用，已在三十年以後。一八九四年萊斯博士 (Dr. J. M. Rice) 編了一個拼法的標準測驗。九年以後，他又編了算術和作文的標準測驗。他爲了要編製拼法測驗的緣故，特地選了五十個字，到各鄉鎮的學校中去測驗學生，叫他們拼這些字。他根據測驗的結果，便得到下面的結論：兒童在校於八年內每天費十五分鐘去學拼字的，其成績同八年中每天費三十分的一樣。但當時的教育者並不同意於他那種進步的測驗方法，他們否認他的結果和他的測驗方法。萊斯和荷勒斯曼費許一樣，他的教育測量的科學方法，也太超時代了，所以他的努力對於當時的教育者並無多大的影響。

所幸者，當萊斯進行他的測驗計劃時，桑代克博士適在哥倫比亞大學讀書。他自從知道了萊斯的工作，便得到極深刻的印像。從此他對於教育測量的可能和需要便感覺極大的興趣。此外，教育家中也漸漸同意於新的測驗方法了；這也是一件很幸運的事。自一九〇八年斯董氏 (Stone) 的算術測驗，一九〇九年桑代克的書法量表相繼出版後，教育界中便有所謂「科學的測量」了。我們如把討論測驗的論文數目統計一下，便可知道自一九一〇年至一九一七年間，人們對於教育測驗的興趣，已有顯著的進步了。（但因大戰的爆發，曾有一度的衰退。）當這個時候，與教育測量有關的著作如依其輕重排列起來，當爲算術、讀法、測量的理論、書法、作文和拼字。至於代數、繪畫、文法、地理、歷史、拉丁文和幾何學等等，也有稍稍加以注意的。現在關於標準測量的書籍，已是汗牛充棟了；舉凡中小學裏所教的各種學科，差不多都有標準測驗了。

新法考試的興起——自從文字考試的可靠性引起了嚴重的懷疑以後，迄今已有二十年了。在這二十年中，我們已經搜集了不少實驗的證據，使人們對於舊法考試的信仰日益減少。同時還有許多研究，也足以證明教師對於論文式的考試所給的分數，並不真確。因此，有的教育者便主張完全廢除舊法的文字考試而以標準測驗代之。但一般辦學的人總以為標準測驗之為用有限，不若用非正式的非標準的測驗易於適應當地的情形。自標準測驗發達後，考試的技術也隨之進步，所以測驗結果的可靠性便大大的增加了。新法考試的興起，大概有二個原因：第一，由於一般人對於舊法考試的不滿；第二，由於專家因編製標準測驗而獲得一種副產物，就是測驗的技術。許多教師已經很靈敏地感覺到，偶然應用測驗法測驗任何一種學科，皆是證標準測驗的真價。所以在文字考試方面，便很快的採用了新的方法。

在過去幾年中，發表了很多的著述，專門討論這種新技術在非標準的，非正式的考試上的應用。除了雜誌上的文章外，尚有六七種專書與許多專論，探討這個問題。大學學生也有以此為專門研究，藉獲高等學位的。師資訓練機關在文字考試方面，也添設了一種專門學程。關於這個問題，不僅已寫了許多的著述；而且這種新的技術，舉凡中小學進步的教師都已採用了。至於其他機關之採用此種新技術者有：美國公務人員管理委員會（The United States Civil Service Commission）大學入學考試委員會（The College Entrance Examination Board）紐約教育局教師考試委員會（The Board of Examiners of The New York

City Board of Education) 以及各省的教育機關應用於他們的種種考試。新法考試或客觀測驗這個名詞，在教育界中已成爲日常的談資了。時至今日，欲做一個近代的教師，非精通這種新法考試不可。

考試的重要——考試的重要可以由兩方面來證明：一方面從考試所費的時間上來證明，另一方面從考試結果的意義上來證明。每學期中，學生方面要準備及參加學期大考和平常的各種小考；教師方面要忙於出題、改卷、記分和分等；兩方面耗費的時間與精力不知要多少！有些教師似乎常常帶着許多試卷，用橡皮線紮着，好像是等待批改的樣子。祇就師生費於考試的時間和精力而論，考試已是一種極重要的學校工作了。不但如此，倘就考試結果的意義而言，將更顯得重要呢。對於考試的結果，不論學生、父母或教師，都很重視的。在學生觀之，這種結果有一種切身利害的關係，足以影響他們的升級、失敗、狀況、學校地位、高中或大學的升晉、學術獎金、學校榮譽和他人的尊敬等等。大多數學生在討論考試題目的時候，往往興高彩烈，而不覺手舞足蹈。父母對於考試，同學生一樣，也有很熱烈的興趣；有時因爲希望和志向的不能達到，往往深感失望。教師亦復如此，對於考試結果，極爲關切。就一方面說，他們對於學生的進步，有一種切身的和職業的興趣；就他方面說，考試的結果，往往足以引起許多對人的不快的問題，足以爲學校和家庭間合作的障礙，亦即足用爲教師效率的反證。這些因子雖然都足以使考試變得格外重要，但其最重要的所在當然還在考試本身所能表現的作用。

溫習與討論問題

下面許多問題，其目的並不在測驗；牠們僅是正誤法在溫習和討論上的一種練習而已。認為對的問題，可在「十」號的外面加個圈；認為錯的問題，可在「一」號的外面加個圈；疑惑不決的問題，可在「？」號的外面加個圈。在做這種記號的時候，須能提出充分的證據以資證明。這些練習，如果能做的得當，實為最好的溫習工作，並能引起有趣的討論。

(1) 十一 ? 考試的觀念由來已久。

(2) 十一 ? 西洋的考試制度是從中古大學中傳遞來的。

(3) 十一 ? 大學中最早的考試是文字的。

(4) 十一 ? 美國在殖民地時代的許多學校，其考試是文字的。

(5) 十一 ? 一八四五年的波士頓考試，在當時是超時代的。

(6) 十一 ? 波士頓的考試是隨便編造的。

(7) 十一 ? 波士頓委員會很謹慎的預防欺騙。

(8) 十一 ? 在波士頓的考試中不注意書法的優劣。

(9) 十一 ? 波士頓的考試是簡單的。

(10) 十一 ? 波士頓考試的試卷之記分是很謹慎的。

- (11) 十一 ? 波士頓的考試對於結果的說明是很保守的。
- (12) 十一 ? 荷勒斯曼提出八個理由，以說明文字考試的優點。
- (13) 十一 ? 波士頓的試驗對於教育的實際，並沒有多大影響。
- (14) 十一 ? 如果荷勒斯曼還是活着，他或許要反對近代測驗的。
- (15) 十一 ? 智力測驗的起源遠在古代。
- (16) 十一 ? 最早的實際的心理測驗是在法國編製的。
- (17) 十一 ? 最早的客觀標準測驗是在美國編製的。
- (18) 十一 ? 討論考試的優良書籍現在還不多。
- (19) 十一 ? 對於論文式的考試在二十年前早已有人懷疑了。
- (20) 十一 ? 新式測驗是標準測驗的副產物。
- (21) 十一 ? 新式測驗運動的傳播，極為迅速。

第二章 考試的功用

考試的功用——任何一種活動，在選定材料與方法以前，就應該決定其功用。所有材料與方法，應與預定目的相符合。現在學校教學的所以低劣，其最大原因，或許是由於每種科目或單元沒有預先安立目標。對於功用如不加以考慮，就是最有價值的活動，也不會產生最大的效率，牠們的各種可能性也只能獲得部分的利用而已。在考試上有一種顯著的困難，就是常常有人把牠們視爲一種一成不變的例行公事——教師或學校當局定期考試一些什麼，學生亦以能通過這一類材料爲盡責。牠們的目的大部僅以決定對於某事或他事的適宜與否爲限；甚至即作爲如此目的，也並不能有周詳的計劃。結果，遂有一部分教師，因爲看不出所用的考試的用途，便極端主張把牠們摒除於教育系統之外。麥克魯魯局長 (Superintendent McAndrew) 的態度便是這樣的，所以在討論考試的結論中說：『從各方面觀察起來，我最後的結論，以爲如果用外加的考試去決定學生的適宜與否，是不誠實的，不公平的，不道德的，而且是含有惡意的。』對於各種考試，許多教師以爲可褒者少而可貶者多。但是我們如能明瞭現在各種不健全的考試所常常發生的顯著的誤用，這種情形便不足爲奇了。

但是有許多，對於這個問題曾詳細的研究過，便以爲適宜的考試能完成許多重要的功用。研究測驗的

人，對於考試功用的詳細分類，雖則意見紛歧，但是普通都承認考試能很有益地達到幾種重要目的。任何一種考試功用的分類，多少總有些重複的地方。每種考試幾乎都包含一種以上的功用。我們要想把考試所能完成的功用，為之一一分離，事實上恐是不可能的。

本章擬將下面各類的考試功用，詳加討論，著者相信，這樣是最有益了：

1. 測驗知識的保持 (Retention)
2. 成績的判斷
3. 日常工作的刺激
4. 溫習的動機
5. 供給客觀的標準
6. 教學效率的測量
7. 教學效率的改進
8. 個別困難的診斷
9. 智能的培養

一個僅在測驗對於某種工作單位的知識的記憶的考試，同時也有普通的診斷價值，因為同時可以顯示班級的或個別的弱點。而一個準備診斷個別困難的測驗，也可考驗知識的記憶程度，不過能將其注意點集中於主要的弱點上，而不失其為診斷的而已。同理為測驗各種功用的考試，雖非為智力研究而設，均能予智能以練習的機會。但考試的各種功用雖是互有關係，却不能不就功用分為各類，庶幾可看出各類考試各有其價值，在各時期的應用途可各循其趨重的目標。所重視的功用與一個考試的形式和內容以及施行的手續，至有關

係。

測驗知識的保持——學生做完了某種工作，對於工作單位所含的知識，到底記住了多少，教師既不能正確的明瞭，學生也缺乏清楚的觀念。教師與學生都孜孜不倦的努力於所習教材的精通。但有許多教師如發現了學生對其所指定的參考書甚少心得，他們將極驚異。許多重要的學校工作，並未熟習，便認為及格聽其通過；到了後來逐漸發現以前所學實未完善，足為學校或職業成功的永久障礙。有效的教學與學習應覆驗自教室工作及指定作業中所保持的知識。

所保持的知識包含一種工作單位中的事實，這種事實學生或能從記憶中回憶之，或能於遇到時認識之。例如工作的單位為南美洲的地理，那末便可測驗學生回想的能力，使其從記憶中說出南美的國名；或者僅僅測驗他由聽了某某幾個國名而知道牠們在南美洲，或見了牠們的名字或簡圖的邊界即能認識是南美洲的國家。要想從記憶中想出南美各國的名字，比看見了或聽見了國名然後想起其是屬於南美洲時，須更加純熟。因為從記憶中想起較從目觀中認識要難得多。就實際目的而言，僅能認識某種事物，對於許多教材，已好算十分純熟了。

有的時候，我們須決定學生對於某種事物間關係的了解。這可以包括某種事物相互間在時間上的關係，事物在普通性質上相對的重要性，事物在某種質量或數量上所佔價值的次序，以及原因與效果的關係等等。

例如有一學生於此，我們可以測驗其對於幾種重大發明或幾次重要戰爭的知識，幾個大都市與其家鄉的比較距離，各國產油數額的相對重要，以及我們在經濟生活中大資本吸收小資本的原因和結果等等。關於這種關係的熟記，可從「記憶中想起關係的能力」或從「目視中認識關係的能力」測驗之。校正測驗(Check-up tests)所測量的知識應儘量與日後生活中應用的相同。倘使教材的熟習能與將來所需要的一樣，那末學生在校正測驗中須得高分數，有時並須得足分，這是很顯然的。不過分數的主要點不在校正測驗本身，而在表示各人相當地位，俾給學生以一種動機去改進學校工作。測驗的功用既是互相重疊的，沒有一種可能性是可以忽視的。

校正測驗的普通診斷價值，是應加重視的。牠能指出全班中的普通缺點，對於這種缺點，將來須加以重視，並須施以更有效的教學方法，牠還能指出學生個別的弱點，對於這種弱點，須用更詳的診斷考試來測驗，以求決定原因之所在，並施以可能的改進方法。倘在全級方面，還沒有獲得某種重要結果，則於全級的工作，尚須加以進一步的注意。倘在個人方面發現特殊缺點，則非從事各種個案研究(Case studies)不可。覆驗學生所學習的工作單位時，考試能完成一種有價值的功用。

成績的決定——除了測量學生某學科特殊工作單位的熟習程度外，有時更須進一步決定其全部成績。在這兩個測驗目的中，因為彼此互相重疊，所以往往不免時起紛擾。但是前者在所習工作單位的限度以內，注

重所記住的知識分量；而後者則注重某一時間內對於某科所成就的一切。明乎此，這種紛擾便不難解除了。前者係指校正測驗，後者係指總額測驗（Inventory test）兩者異曲而同工。

各種考試所以遭受不少苛刻的非難與嚴厲的攻擊，原因由於所謂成績，持論往往不同。決定學生對於某事如何始為適合，乃考試最廣博的目的，有時亦為惟一的目的。在許多教師觀之，此種目的乃深思熟慮的，或是有意識的。無疑的過去的考試過於重視這一點，尤其是那種常用到的不良考試。亦自以為是的認作成績的標準。無論如何，這個功用是合理的，有價值的。一個教師的工作常常須知道學生成績的總額。但現在大家都知道一個教師的主觀的判斷，對於決定各種事項如升級、分班、分組、學校地位的指定，新工作出發點的規定等等，都是不可靠的。設有一新生於此，自一小學轉來，該校的組織不同，標準不同，課程不同，那末總額測驗，便有助於適當地位的確定了。又如一高中學生，自一校轉入他校，欲選習數學，外國語，或其他須有先修課的工作，則總額測驗對於決定其所需各科的準備，大有裨益。一有志教師入教師訓練機關時，總額測驗可用以決定學校課業成績。一人在投身選定的職業以前，可用總額測驗以視其準備的性質。現在有許多教師，在指定新工作單位以前，往往先施行一種預先測驗（Pre-test）拼法的總額測驗乃一普通的例子。此種測驗能告訴我們那幾個字學生已經會拼了，那幾個字還須繼續學習。結果，不論教師或學生，均能以極少的時間與精力，去解決特殊的困難。所以決定學生的成績乃考試最重要的功用之一。

日常工作的刺激——動機 (Motivation) 在現代教師職業的字彙中，是一個常用的名詞。在自然而有趣的形態之下，如何去刺激學生，使從事更有效的學校工作，實為一極重要的教育問題。近幾年來，動機已成了新教學法的基本原理。任何方法，只要能使學校活動含有充分的動機與興趣以從事有效工作者，均應注意。在理論上，學生每天的準備，應受下列各事適當的刺激：活動本身所生的快樂，對於工作的愛好，認識工作的用處，或直接感覺熟習工作的需要。但是，為教師者誰都知道，理想與自然的動機，對於學生，並不能常有適當的刺激，使每天獲得滿意的準備，大多數學生所需要的，為一種固定的當前的動機，這不是活動的本身所能供給的。有許多學校工作，教師們不得不借着人為的誘導，使學生從事學習。此種誘導，應與活動的本身，愈近愈妙。就這目的而言，考試有一被人忽視而極重要的貢獻。

學校工作的刺激，在測驗上成為一合理的有用的部分。此種刺激與活動的本身有親切的自然的聯絡，所以是合理的；能供給當前的合理的動機，所以是有用的。心理學上有一重要定律，就是學生於其進步，如能知道，則所學最佳。根據這種學習法則，學生便會知道測驗所表示其自身的努力結果，並將此結果繪成學習曲線，以與自己最高成績相競賽，實在是很有刺激的。這種進步的記錄，在學生觀之，比暗中摸索，要滿意得多。倘所注意者是進步，而非達到熟練的某種標準，則結果一定會獲得更廣博的要求與更有效的動機。熟練的固定標準，須超出能力較低的人所能達到的，但却不可過高，專為優秀的天才而定。根據測驗結果而在個人間或團體間為

種種競賽，也有極大的裨益。但在舉行時，須注意比賽者的能力，應該不相上下。

以上所述每天工作的種種誘導，包括日常測驗的應用在內。大考對於刺激完善的工作，也極有效。斯古德（F. H. Schuster）的研究，對此有實驗的證據。在兩學年內，斯古德曾將一級學生分成兩組，每組各一百人。兩組的工作，各方面完全相同，所不同者：一組須告訴他們要大考，另一組須告訴他們不要大考，他們的分數完全根據他們的日常工作。對於考試組每星期提醒大考一二次；同日囑另一組努力工作，因為他們並無大考。為欲避免「急來抱佛脚」起見，在考試以前的一二星期，課外的指定工作極為繁重；對於考試組並告其「急來抱佛脚」無補於事。在舉行考試時，先將實驗的意義講給非考試組聽，然後兩組各給以一百八十條理解的新法試題。非考試組因欲證明考試無效，極努力將事。後來考試組與非考試組的分數為：中數 72 與 70 ，均數 72.5 與 95.5 ，兩極距離（Range） 68 至 154 ，與 60 至 124 。觀此，足以證明考試組好得多了，因此斯古德便下一結論，以為大考的結果，極有價值。刺激的日常工作，乃考試的一個重要功用。

溫習的動機——溫習工作之於優良教學，關係極大。有許多優良教學所以無多大希望，就是因為沒有承認心理學上的「遺忘律」（Law of forgetting）而缺少充分的溫習工作。只要遺忘的歷程一開始，學習便不復進行。遺忘的趨向是一種自然的歷程，也是一種幸運的歷程。我們所學的東西，有許多是極瑣細的，有許多是不確切的。我們如把這些東西統統記住了，勢必心勞日拙，而為學習的莫大障礙。遺忘不僅能使我人從事溫

習工作，俾適當地記住某種教材；並能使應該永銘於心的特殊教材，有選擇與組織的可能。對於學校工作的溫習，如無相當注意，決難獲得有效的教學與學習。

要引起溫習工作的動機，考試為最有效的方法。對於合理的考試，如欲為適當的準備，則心靈上僅有匆促而膚淺的活動，如普通所謂「急來抱佛脚」者，尚有未足。所有材料，須熟思一番，批評一番，然後放其過去；一切事物，要沈思熟慮，權衡輕重；各種重要觀念，必比較其重要，而組成一有用計劃；一切材料當依其適宜觀點，而為整個的籌謀；此外還須好好去融化，細細去分類，有條不紊的去記在心頭。用一種近乎考試的目光去觀察教材，便可注意及此。欲使學習有效，必須獲得這種反省的見解。就溫習的動機言，考試具有絕大的功用。

供給客觀的標準——學校工作的需要客觀標準，凡對此曾深思過的人，都知其重要的了。教育工作因為結果的空泛與難解，不知受了多少桎梏。為教師者，每以為他們的班級不及標準，或超過標準，或適合標準。有時他們以為這年度的一班比上年度的一班壞，或則以為較好或相等。他們也許會相信，普通的工作標準是逐年退步的，或是進步的，或是保持常態的。此種見解實為一種主觀的意見，係根據不可靠的普通印象所成，故無論如何是無意義的。一般守舊的人，據其主觀的見解，咀罵現在學校，以為有許多基本工作，不及以前有效。此種心理上的態度，雖全無依據，缺乏可資辨證的事實，但影響所及，是很危險的。適當的客觀標準，對於精確的學級競賽，以及學校工作的批評，關係都很重要。倘欲建立客觀目標，以指導各種教學工作，則此種標準也屬必要。教師

們因爲缺乏客觀標準，對於一級到底存什麼希望，作什麼要求，往往茫然不知所從。

供給客觀的標準包含某種測量。就教育的領域言，測量就是考試的結果。新法測驗的分數所表示的成績，較爲確定。測驗分數如有永久記錄，則級與級之相比，要容易得多。倘使去年一級的歷史測驗，其平均分數爲80。而今年一級爲85。則教師便能獲得一種客觀的證據，斷定今年一級的歷史程度不如去年一級的好。如將此種分數累積起來，便可造成一種標準，教師即可據此獲得一正確觀念，以測一級的程度。用了這種方法，便能使測驗標準化。

坊間所出版的測驗，其定標準的方法是將各種測驗測量了許多人，然後據其結果以造成常模(Norm)或標準；這樣，便可推知任何類似團體在這種測驗上所能達到的程度。就教室的目的而言，這種商業化的標準測驗所造成的常模，用時每生極壞的影響。牠們並非就一般中才以假定平均數，故不能用作一學生或一學級的教育目標。在美國全國，牠們的應用能降低工作的水準。在某一級的學生，其所爲工作，應該勝過全國學生所爲的普通工作。學生如僅能接近或達到此種常模，教師還不應滿足。不但如此，就是當此種常模用作目標時，教師爲欲表示測驗的優異起見，不得不注重那些要測量的東西。這些材料或非最應注意的。測驗所注重的教材，也許有錯誤的時候。有些材料的選入，或許因其本身易於測驗；有些材料在測驗編了後，或許已不合時宜了。應用標準測驗的常模以爲目的，每致永久保存陳舊的材料，並固定傳統的學校課程。

教師工作須隨時變更，以適應現代社會的要求。最適宜的客觀標準之造成，根據教師自編的測驗，這種測驗應參照當時當地有用的貢獻。一切等第足以代表以前學生之測驗分數者，可為一個便利的標準，以定個別學生（受此測驗，作為特殊的個別考試）的等第，或定一班人數極少不能自立可用標準的班級之等第。新法考試對於供給學校工作的客觀標準，實負重要的使命。

教學效率的測量——教學效率的測量是一個陳舊而困難的學校問題。教師與學校當局早已一致承認教師服務有精確測量的必要。成績優良的，應予認可與獎勵。服務不力的，應詳加糾查，或設法改進，或予以淘汰。此種舉動，僅恃個人的秉公判斷，尚覺不足。教學職業的地位，繫於教學的性質，而表現與維持之者，端賴從事教學的人。優良的教師，有時須防公眾的猜疑與反對，教學效率的正式記錄，便有這種防護的功用。

有許多理由能用以證明教師的分等。教師之應分等，已有一致的認識，這裏無置辯的必要。問題所在乃為如何去獲得完滿的測量方法。時至今日，凡充分認識學校事業者，對於用舊的印象法（Impressionistic Methods）以分等教師，已無多大的信仰了。以前的學校行政人員，依據此種方法，默然走入教室，數分鐘後，便恃其臨時的觀察與主觀的判斷，以衡量優劣；復參照他自己對於學校工作的舊經驗，予教師以有效的分等。此種分等的最後結論是何等的可驚訝！這個問題太切身了，太鄭重了，太緊要了，不能僅恃個人的意見與主觀的判斷去解決。牠與教師職業的成功和個人的快樂，關係至切，決不可屈服於不可靠的方法下。不適宜的分等，遠不如

不分等的好。這樣分了等級，在被分等者中間，反足引起種種不滿。教學效率的測量，端賴充分客觀的方法以博信仰。此種方法今日尚在演化的歷程中。

各種考試實供給最豐富的園地，以訂定測量教學效率的客觀方法。愛列亞脫 (Eliot) 判斷哈佛大學教授的教學能力，係根據他們給與學生的各種考試。他這種方法是很對的。考察許多試題後，就可知道教師的測驗技術是否純熟，對於教材是否了解。教師的「教學方法與實際」的知識可用專為那種目的而編的職業測驗以測量之。此種測驗現在已經有了。但一個教師的真正效率，應根據他對於學生的服務而定。最優良的教師能使學生有最大的進步。由此以言，用「學生測驗法」(Student-test Method) 測量教學效率，當然最可靠了。如用此種方法，則於某時期的開始與終了，各予測驗一次，其間所有進步，歸功於教師。此種方法的應用，須極注意，因為有許多非教師能力所能控制的因子，也許要加以考慮。這種因子如學生先天的能力，學生的年齡，學生以前準備的性質，教材的方式與各教師的注意點，學期的長短，教學的分量，輔導的種類，設備以及家庭的扶助等等。每個因子均能使任何教師的工作異於其他教師，所以都應加以相當的注意。關於教學效率的測量，考試能完成一種重要的功用。

教學效率的改進——現在有一種趨勢，以為教學效率的改進，僅限於劣等教師。實則在一種進步的職業如教學中，每個教師都應注意其自身的改進。任何人都應這樣，因為沒有一個教師是一個完人。教學的經驗，日

積月累，遂為職業素養的一部。若干年的教學，在其本身，不一定會增加能力或效率。教師有了教學經驗，有時反而有減低效率之可能。經驗這個因子，其價值須視其貢獻於將來生長之道以為斷。最重要的一事乃為教師於服務期內的助益範圍。在教學生活中改進效率，乃從事此職業的各個分子所應時常注意的。

在改進教學效率的諸種方法中，很少比考試的可能性更大的了。考試之貢獻於教學效率的改進，不一其道。教師如將測驗結果詳細研究一下，便能知道學生的成績，而明瞭其缺點與困難之所在。既具此種知識，教學便益能適應學生的能力與需要了。測驗結果同時並能用以決定各種教學法的效率，因此教師能獲得一種指導，藉以選擇最好的方法。再者，有效的教學，須隨時明瞭學生的進步，一個優良考試的編制使教師不得不自決課程與各個單元的目標；完善的選擇並組織教材；精嫻一切必需的材料。搜集考試材料以表示班級成績的進步，可為日常討論與教學極好的準備。學生對於考試的反應性質有助於教學效率。此種性質能表示何處並未充分注意，或僅涉及皮毛；以及教學上的種種弱點。考試技術的熟習，就其本身而論，並非自己改進的成功。考試的最大功用，便是教學效率的改進。

特殊困難的診斷——有效的教學有待於學習上特殊困難的診斷。在這方面，教學與醫術的實習，極為相似。優良的教師與優良的醫生一樣，須知道診斷的方法。醫生用了普通的考驗方法，便能指出那些人身體上的缺陷，須待更完善的測量。他詳察各種特殊的情形，直至明白特殊的困難為止。經過了一番周詳的診斷，他便能

開出最好的藥方，或採取正確的方法。教師也應以同樣的科學態度與科學方法，將學習上的各種缺點，詳加診斷。屈拉巴博士 (Dr. Trabue) 將所有情形詳加衡量之後，曾這樣說：『……教育上特殊病態的診斷並長期實驗以發現形式訓練對於學生有益抑有害，——就是個案工作，為應用教學測量的第二步——都應由固定的教師就教室中為之。』倘無完善的測驗計劃，此事的實現是不可能的。

教育的診斷有屬於全級的，有屬於個別的。全級的診斷須研究普通的總額或調查考試 (Survey or amination) 中各個問題的答案。測驗的最後目的在改進教學，這也是一種最直接的方法。測驗能顯示全級學生的共同困難。團體的共同弱點，既經發現，其原因普通歸於教學的缺點，而不歸於學習的缺點。在那種情形之下，補救的方法，莫如改進教學。這樣，可使教師分析其工作，以求結果所以寡效的原因。例如已經舉行了一種語言測驗。教師研究各份試卷的個別錯誤後，指出必需的修正，計劃補救的方法。然後將試卷發回學生，囑其溫習。各生將他自己的錯誤列成一表，而注意各個錯誤的修正。在多數情形之下，此種修正法 (Method of correction) 既可應用，且合實際。在少數情形之下，教師尚須將所有困難，為更完善的分析，並採用特殊的補救方法。再舉一例，假如有一學生，數學能力極為薄弱。既經測驗，便可顯示若干基本運算的小錯誤，如小數的誤點，零的錯用，某項分數的誤解，加減乘除習慣上的謬誤，以及其他種種。一個默讀測驗可以顯示學生閱讀問題，不能充分了解。對此的補救方法須靠個別教學，非傳統的團體教學所能為功。此種診斷工作，與上面所討論的

教學效率之改進，關係極密。調查測驗可以顯示某種學生較其他學生程度低劣。就這些個別情形而言，缺點在乎學生本身，並不在教學的方法，凡發生特殊問題的學生，都應受進一步的診斷。調查測驗中學生所誤答的問題，俱應詳加分析。這樣一來，所有的困難，便有線索了。

除診斷工作外，特殊的測驗也不可少。此種測驗，應注意弱點所在的全部工作單位。所有教材須分爲若干要素或成分，而分別加以觀察。普通的調查測驗，在顯示什麼尚未學習。特殊的診斷測驗則在顯示爲什麼尚未學習。教師熟知學校工作所以低劣的最普通原因後，便會感覺特殊考試的需要。這樣，便可減少許多盲目分組之弊，研究與經驗的結果，此種知識便日積月累，增長不已。每個教師都應熟習推孟（Terman）與亞馬克（Almack）在學校兒童衛生（The Hygiene of the School Child）一書（一九二九年版）中所討論的教材。此種知識，實爲推知學校工作低劣的身體原因之基礎。學習的不良，或由於營養不足，或由於聽覺或視覺有缺陷，或由於患神經病，或由於身體上的其他妨礙。教育的診斷，須明瞭學生的身體基礎。學習的不良有時亦由於學生普通智能的不足。要決定學生的智能，須賴完善的智力測驗。

此外教師尙有其他事件須待查出；如測驗方法的誤解，閱讀能力的低劣，主要的先修的知識或技能之缺乏，對於學校工作的錯誤態度，興趣或注意的缺乏，讀書習慣的不正確等等。困難的原因既然診斷了出來，補救方法常極簡單，或至少能應用科學方法來處置。關於教育診斷與補救方法的舉行，考試能完成一種不可少的

功用。

智能的培養——考試能使應考者養成某種「智能的習慣」，此種習慣在日常生活應用極廣。所謂智能係指應用，注意堅忍，自恃，與推證等等。考試能供給一種情境，在此種情境之下，我人能由應用所有知識而獲得練習。考試也能訓練注意的能力。我們決不能找得一羣在考試時不十分專心的學生，他們攻求問題，大都是費盡心力的。他們工作時，我們又覺其堅忍持久，非至終了不止。有些學生因為意志的驅使，人家做完後，尚孜孜不倦的繼續工作。末後他們常有幾個答案，須加證明。這樣，便指引他們去和專家討論。在考試時，學生不得不依靠他自己的判斷與心智歷程（Mental process）。這樣，便可發展其自恃心。凡此種種智能，均為無價之寶，他們都能由考試中獲得經濟的練習。關於考試功用的價值，哈佛大學校長羅惠爾（Townell）曾有極精透的意見，茲摘錄如下：

……獲得為應付考試用的技術，最為有益。……一個律師在出庭辯護時，當前每遇到一種出乎預料的證據問題，必須立刻想到他以前經過的各種案例。……這就是應付考試了。一個醫生突然請去治療病人，情形也是相同。……在病榻之旁，他也須應付一種考試。我們一生中間常常因被迫而竭盡所能去搜集以前所獲的知識。此種應付的能力能從練習中培養之。迨事過境遷，我們往往自己向著自己，我們為什麼不把緊要的所在，牢牢記住，學生在考完以後，老是這樣的。追憶得迅速，完全，正確的技術，是心能訓練中極

重要的一部。此種技術，乃一種特殊的技術，與僅求胸羅萬卷不可同日而語。有些人，所有知識好像準備好的。應用時，一呼便出。有些人，須經若干時的沈思熟慮，方能將知識搬出。有些人只能在閉戶獨學時，應用其知識……能於需要時知識回憶得敏捷的價值，在事業家比在學問家與著作家為大……在教育上時間與思想的耗費，沒有再比這樣值得的了……結論……考試適當的應用，乃為教育歷程中最有生氣的一部，不過欲用以獲得良好結果，此種技術極為複雜，極為困難。

在實際上，每種職業都含考試在內，此種考試的應付，端賴人類智力的應用。除其他功用外，僅就培養智能而言，考試在教育系統上，佔一重要地位。

溫習與討論問題

- (1) 十一 ? 一種活動的功用應決定於方法與材料計劃好了以後。
- (2) 十一 ? 專家對於測驗的主要目的有一致的意見。
- (3) 十一 ? 每種考試只能完成一種功用。
- (4) 十一 ? 一般趨勢過分重視教材的精通。
- (5) 十一 ? 校正測驗須與知識後來的應用相符。
- (6) 十一 ? 校正測驗與總額測驗應該分別清楚。

- (7) 十一 ? 成績的決定為測驗最普通的功用。
- (8) 十一 ? 總額測驗僅可用於升級之用。
- (9) 十一 ? 利用測驗以引起日常工作的動機在教育上是不健全的。
- (10) 十一 ? 知道進步能刺激更大的效率。
- (11) 十一 ? 適宜的溫習工作，普通極感缺乏。
- (12) 十一 ? 考試能鼓勵『急來抱佛脚』而不能鼓勵有價值的溫習。
- (13) 十一 ? 教育工作因缺乏客觀的目標常受桎梏。
- (14) 十一 ? 教師所用非正式的測驗作為標準測驗的初步。
- (15) 十一 ? 教學效率的測量在教師極為有益。
- (16) 十一 ? 在理論上，學生的進步為教師分等可靠的根據。
- (17) 十一 ? 學生測驗法對於教師分等有相當的限度。
- (18) 十一 ? 服務改進的問題僅限於劣等教師。
- (19) 十一 ? 考試在改進教學的可能性中佔高的地位。
- (20) 十一 ? 教育診斷是徒費時間的。

新法考試

- (21) 十一 ? 測量對於教育診斷極關重要。
- (22) 十一 ? 有價值的心能特質可由考試中培養之。

第三章 優良考試的特徵

優良考試——一個考試的優劣，由於個人的判斷而定。判斷的方法，同評判一個故事，一張圖畫，一曲音樂，一位演說家，或一天的工作一樣。在演說競賽中，所謂最好的演說家就是各位評判員所認為最好的一人。然而富有經驗的評判員並不信賴一般主觀的印象。他們先定了幾種重要的性質以為標準，競賽者中誰具有這些性質而至最高的程度，就判定誰最好。批評一個故事也是如此，其或優或劣，完全看牠有無某種適當的性質以為斷。這樣看來，選作標準的諸種性質，務求其為合理的評判標準，這是很重要的。譬如硬度可作為評判岩石的一個適當性質，却不能以此來斷定考試的優良。現在測驗專家所公認的性質，可為衡量考試最有用和最正當的標準者，有下列數種：

- (1) 正確性 (Validity)
- (2) 可靠性 (Reliability)
- (3) 客觀性 (Objectivity)
- (4) 廣博性 (Comprehensiveness)
- (5) 便利性 (Facility)

(6) 實用性 (Utility)

(7) 融洽性 (Rapport)

這七種性質為一個優良考試的適當特徵。一個考試的價值，須以其具有這些特徵的程度，如何來決定。本章將專討論這些重要的性質。

正確性——任何事物的質量或數量，其決定方法必須是合理的，能解釋的。由頭部的指數 (Cephaloindex dex) 以測量心理的能量，大多數人已知其不可信了。因為這種方法，除表示頭的形狀外，便不能表示什麼了。以重量來表示健康，或者計算頭核 (cephalic index) 以度量誠實，這也公認其不可信了。因這種方法所測量的是重量和頭蓋骨的構造，而非健康和誠實。所以這些測量縱名之曰健康測驗或誠實測驗，而這種事實總是不變更的。大概一個測驗的名稱，其本身並不能保證這個測驗就能測量被命名的東西。我們往往認為一個稱做默讀的測驗就是測量默讀能力的，一個命名歷史的考試就是測量歷史的。在這種情形下，我們就有點類乎老農了；老農需要許多天文的知識，但却不願行星名詞的如何發現。這就是說，我們每易把名實相混，而不加區別。實在說來，一個名義上稱為算術的測驗，實際上很容易變為一個默讀測驗。一個測驗應該測量其所要測量的東西，所謂測驗的正確性，就是看牠能夠達到這個目的的程度如何。

為求測驗的正確，編造測驗者應牢記他所希望測驗的事實以及所期達到的功用。測驗當與一科目的教

學目標相符合，又必與所預期的功用相表裏。雖然所謂功用往往重複含混，並且即顧及各種功用亦未必牽動一個測驗的本質，但因目的的不同，而取材的型式以及實施測驗的手續却有區別。校正測驗未能用作總額測驗，而兩者又均有異於診斷測驗。為鼓勵逐日的學習，其測驗當簡短，不絕的舉行，並且不先通知。倘若用不常舉行並且預告的方法，便不妥當了。上章所述測驗的各項功用，在編造測驗求其正確有效，皆不可忽視。雖然如記憶的功用與思考的功用之不易辨別，當於第五章再詳論及，但測驗應用事實於相當境遇時的能力和僅測驗其是否獲得事實，却顯有不同。惟有於擬定問題時在問題的形式上多所變化，後當詳論之。

測驗所包含的內容，應富有價值而且極關重要的。其中最重題目的可從學程大綱，基本教科書，和最低標準中決定之。除此以外，我們應該注意免除書法，拼法，語法，淨潔，整齊，書寫速率等不相干的因子。除非這個考試是專門測量這些因子中的某一種的。這些因子既與所要測量的東西無關，就不應變為測驗的一部分。如果要計算這些因子的分數，應該把牠們當作一件另外的單獨的事。我們知道一個正確的歷史測驗，僅僅測量歷史上的成就，而不及其他。如將歷史以外的因子也放入測量的範圍，那就變為不正確了。

我們很相信，如果一個測驗能與該科學程，該科基本教科書中的材料，以及專家所定的最低標準相符合，而其形式與施行手續復能與所完成的功用相一致，那麼這個測驗就很正確了。請富有經驗的人擔任批評，是一個考驗正確性的良好方法。至於數量可用相關方法決定正確性的程度。其實，教師如能審慎的選擇學科上

重要的材料，除去不相干的因子的影響，充分注意測驗的功用，那麼這個測驗，對於實際的目的，大概是很正確的。

可靠性——同類的量表應用的結果彼此相距懸殊，便均不可靠。因為他們的結果有了差異。我們對於這種結果不能信任。例如一隻鐘，測量時刻，如不正確，那就不可靠了。考試的道理也是如此。一個不可靠的測量方法是騙人的，牠能使人發生錯誤的判斷。一個測驗，不論其所測量者為何，總應測量得準確而一致。所謂測驗的可靠性，就是看牠達到這個目的的程度如何。

上面已經說過，把測驗所要測量的因子明白記住，把其他不相干的因子除去或減少，足以增加測驗的正確性。其實，注意了這二點，也足以增加測驗的可靠性。除此而外，每個測驗還應由許多獨立的題目所組成。這就是說，每個問題不應該倚靠其他任何問題，或受其影響。問題的難易，相差應該很廣，以便充分適應全班學生各種不同的能力或成就。就一方面言，一個測驗應該是很難的，須使全班學生無一能得足分。這一點是很需要的，因為這樣才可以測量中材以上的成就差異。如果一個測驗只有普通的難度，中材以上的人都得了足分，那麼優良學生的差異依然沒有測量出來。就另一方面言，一個測驗須是很容易的，使全班學生無一得到零分。這一點也很需要，因為這樣才可以測量中才以下的差異。如果測驗太難了，有些人得不到分數，那麼能力低下的學生的差異，依然沒有測量出來。凡是參加測驗的人，都應該得到正的分數。所以測驗中的題目，都用同一難度是

不適當的，這是很顯然的了。這種同一難度的測驗，不能指示各學生的差異。如果分數集中於分配的上端，這就表示測驗太容易；如果集中於分配的下端，這就表示太難。至於一個適當的測驗，則從最低分數至最高分數，散佈於全距之間，形成一正常的分配。一個測驗只要有許多獨立的題目，其結果又能表示常態分配的分數，教師便可深信其非常可靠了。測驗的可靠性也可用相關方法在數量上決定之。不過，我們並不叫從事實地教學的教師去計算非標準測驗的可靠性。

客觀性——各人的判斷不同。我們都不免有癖性，成見，和偏愛。我們主觀的判斷，每受氣質的影響。譬如各人估量飛機的高度，室中的溫度，汽車的速度，以及房屋的大小與物件的輕重，其結果每大不相同。有些人常常估量得太高，有些人則太低，恐怕沒有人估量得正好的。天文學家欲免個人的錯誤起見，必須修正他觀察的結果。他記錄一顆星的運行，比實際上變動的一刹那，不是快了一點，就是慢了一點。所以觀察的人，都有改正公式以減少其反應時間的差異。至於教師評判試卷時標準的不同，以及分別等級時的寬嚴互異，這是盡人皆知的。一個考試應該除去或減少記分上的主觀判斷，所謂考試的客觀性，就是看牠達到這個目的的程度如何。

在一個客觀的考試裏，記分者的思想和情感並不影響記分。不論記分者為誰，同一答案應受同樣批評。不過，絕對的客觀的量表或測量標準自然是沒有的，就是寒暑表和碼尺也僅能使判斷減少主觀而已。牠們的精確，有賴於量尺上單位的精細，以及應用時的審慎。牠們不過是幫助人類的判斷，而不能替代人類的判斷。人類

的判斷，如無外物的幫助，是不可靠的。例如高度測量的精確，靠着測量機上細小的單位。測驗的情形也是如此。一個具有許多精細單位的測驗，其精確程度，遠勝於僅具少量粗糙單位的測驗。所以一個考試的編造，應該使任何人所記的分數都能一致。

廣博性——有許多事物，為其本性所限，我們不能測量牠的全體。在這種情形下，就不能不以樣子 (Specimens or Samples) 來估量全體了。譬如礦的性質，由化驗礦苗的樣子而決定。穀的優劣，也由考驗其樣子而推度。對於人也是如此，要想把一個人的能力，成就，或對於某學科所有的知識，一一加以測量，完全無遺，也是不可能的。我們所能為力的，僅不過從他的能力，成就，或知識中取出樣子而已。因此，他對於某學科的全般知識，也只能從所取的樣子中推知。所以一個考試，不過是從牠所要測量的東西中取出樣子。我們對於牠的結果能相信到何種程度，大都要看牠取樣的完善與否而定。取樣不當，不但要發生錯誤，且使被試者橫遭不公。一個考試應能為適當的測量。牠的範圍應該很廣博而含有代表性，所謂考試的廣博性，就是看牠達到這個目的的程度如何。

把考試中每個問題作為個別的和孤立的樣子看待，很為有益。計算結果時，所依據的優良樣子愈多，結果愈為準確。一個考試應有適當的代表問題，這是教師所應知道的。所謂代表的意義，就是說這些問題能包括所要測驗的材料，而其難度又適合被測驗的團體的能力。問題的意義如已見於其他題目，不可再行重複。此外，要

使考試含有廣博性，問題的形式和數量也應加以考慮。問題的數量，在任何考試，自然都有限制。對於測驗的編製，施行和記分所需的時間，我們也應注意。每個問題所代表者如尚未見於其他問題，都應該列入。同類的和論點相同的取樣，為數雖多，但不能增加結果的可靠性。問題的數量大了，固足增多變化問題形式的機會；不過一個短短的測驗要是包含着精選的問題，其廣博性要比一個長而不顧問題性質的測驗來得大。實際上，一個考試已達到了某種程度，就是再加多題目，對其結果也極少影響；結果上的差異也不值得特別注意。美國公務人員管理委員會（註一）已經知道一個四十分鐘的良好測驗，實際上與六十分鐘的良好測驗，結果相同。所以我們可以說，一個測驗如增加了題目，也不變動學生的相對地位，則其廣博性或適當性便很大了。

便利性——一種方法固可用以獲得良好的結果，但也許因過形複雜反致不切實用。我們知道一部機器在能用以後，尚須經過長時間的試用，然後方能運用便捷，毫無困難，至此方能傳之於市。汽車和無線電經過長期應用後，才變為簡單便宜，幾乎盡人能使用，盡人能享有。運用便捷這個性質，應用於考試上同應用於機械的發明上一樣。一個考試的施行和記分，都應力求便捷，所謂考試的便利性，就是看牠所具這個要件的程度如何。一個考試的排列，應謀師生雙方時間上和精力上的經濟。我們應使長篇答案的寫述，變為不重要而且不可能。學生書寫的速率與持久能力，考試中應不予注意，否則反易混淆。預先準備簡短而確定的答案，可以免除這種混淆的情形，同時且可使記分簡單化。教師批改試卷，往往精力費得太多，這是由於批改的時候，先要讀許

多頁編製拙劣的材料，然後再決定這種冗長的討論之價值。批改試卷，即在人數適中的級上，也常成爲繁重的工作。所以一個測驗的排列，應使記分能於機械化的態度下，利用答案紙 (Answers) 我們應使記分成爲例行的抄寫工作。關於考試的進行手續，學生如能得明白而確定的說明，施行就很便利。但一般的說明，縱不錯誤，也晦澀而不全。對於所要做的事，如有明確的觀念，則事無大小，均易於成功。明確的說明，能減少學生工作上的困難，而其結果，教師也易於記分。測驗的便利性也需代價的。於可能時，每個學生最好都有印發的問題，這樣要比從黑板上看或聽教師口述便利得多。總之，無論何事，凡足使測驗進行較易，記分較便者，都能使測驗更臻完善。

實用性——當此工業時代，我們的思想常盤旋於善用事物的概念中。大家都要問，『這爲的是什麼？』——『這有什麼用？』不過，當這崇尚實用的時代，人們所做的無用之事却也很多。無用的東西所以存在於教材及學校方法中者，都是由於無知，冷漠，自私，或習慣等原因。我們雖常常舉行標準測驗，並用牠來記分，但因結果無用，便常中途廢棄。一個考試的結果，如不加以說明和應用，並不能達到牠的整個目的。因此，考試急宜努力於標準的建立，同這種標準相比較，才可以確切測量學生的成就，或看出團體和個人的缺點，以及應採的補救方法。所謂考試的實用性，就是看牠達到這個目的的程度如何。

(註一) 一個四十分鐘的測驗與一個六十分鐘的測驗，其相關係數爲 .912。見 1926 年的 Civil Service yearbook, p. 52.

一個考試應使能與某種參照點(Point of Reference)相較，藉以估計某生的情形。代表『一點沒有』的零點，在高度和重量的絕對測驗中用之，而不用於教育測量上。例如要用一個絕對的名詞來表示一個學生有多少歷史的知識，這是不可能的。教育測量是相對的，其數量是因他人成就的多少而變換的。所以在相對的教育測量上，算術均數(Arithmetic mean)和標準差(Standard deviation)是最穩固的參照點。一個良好的考試，倘以算術均數為原始點，標準差為測量單位，那末成績的量尺上單位就相等，牠便可就相對成績量表的相等單位來分別學生的優劣了。等級與分數應行分清，這是很重要的。分數是測驗上的成就，而等級是分數與某種標準或參照點的關係之說明。測驗的實用性，實與教育診斷有關，也與化分數為有意義的等級的適當性有關。化分數為等級的方法，當在十一章中討論之。

融洽性——一個人對於一件工作的態度，很能影響其作業的性質。做一椿享樂的事往往比做一椿沒趣的事要易有成績。因此，現代的各種教學法，無不努力使工作發生興趣。優良的教師總是斟酌理，盡量使學生感到舒適，且與教學情景相諧和。考試一事，學生常常視為畏途，而且極不信任。為避免這種情形計，一個考試應該創造一種富有興趣與安適的感情。所謂融洽性，就是看牠達到這個目的的程度如何。

編造考試應顧到學生或然的態度。一個測驗應盡量謀興趣的豐富。牠的構造應使學生對於結果的公平，表示滿意。有的時候，很難使學生相信等級的公平。若等級剛在及格以下，或較競爭者稍遜一等，則這種情形尤

爲顯然。其實，考試的結果未嘗不可使其非常公平，直至毫無爭論的餘地。所應注意的，測驗應該實際而合理。考試開始時，最好用幾個容易的問題，使全體都能回答。這樣足以引起學生自信力，而且對於以後較難的題目，也可有適當的準備。

溫習與討論問題

- (1) 十一 ? 考試是以絕對的測量來估定價值。
- (2) 十一 ? 測驗的名稱足以保證其所要測量的東西。
- (3) 十一 ? 正確性就是不同的測量，互相一致的意思。
- (4) 十一 ? 一個正確的測驗必定是可靠的。
- (5) 十一 ? 可靠性的意義就是指準確而一致的測量。
- (6) 十一 ? 測驗的可靠性，可以數量決定之。
- (7) 十一 ? 各人估量遠近，極相一致。
- (8) 十一 ? 客觀性的意義就是使測驗的目的明顯。
- (9) 十一 ? 一個客觀的測驗能除去記分上主觀的判斷。
- (10) 十一 ? 測驗不過是取樣。

- | | | | |
|------|----|---|-----------------------|
| (11) | 十一 | ? | 最廣博的測驗，必須有極多題目。 |
| (12) | 十一 | ? | 便利性的意義是指測驗便於舉行，易於記分。 |
| (13) | 十一 | ? | 教師應費多量的時間於批改試卷。 |
| (14) | 十一 | ? | 教師對於測驗應該不加任何說明。 |
| (15) | 十一 | ? | 考試的結果應加以說明。 |
| (16) | 十一 | ? | 一個人的歷史知識，不能以絕對的名詞決定之。 |
| (17) | 十一 | ? | 實用性的意義就是指把分數化為有意義的等級。 |
| (18) | 十一 | ? | 測驗的態度足以影響測驗的成就。 |
| (19) | 十一 | ? | 測驗開始的幾個問題應該難。 |
| (20) | 十一 | ? | 融洽性的意義是指測驗所生的興趣和安適。 |

新
法
考
試

第四章 傳統的論文式考試

論文式考試的永久地位——有的教育家主張以標準測驗代替論文式的測驗，不過標準測驗雖云發達，尚不足以代替非標準的考試之需要。教師常欲測驗含有地方性或特殊興趣的材料。他又時常覺得測驗小單元的工作是很適宜的。這就需要教師自編測驗了。標準測驗並不適於這個目的。這種自編測驗的應用，雖然所費過大，記分困難，說明麻煩，但在教育測量上確已有很重要的貢獻，而現在也仍有相當的地位。此外，因新式的或客觀的非標準測驗日益發達，有的熱心家頗相信傳統的論文式考試已成明日黃花，毫無實際的價值了。這也是一個錯誤的觀念。其實三種測驗（標準測驗，論文式測驗，和客觀的非標準測驗）各有特殊的價值，也各有某種的限制，我們不應採用一種而摒棄其他，却須三者並用，以互相調劑其缺陷，這才是合理的態度。

傳統的論文式考試與教育實施的關係，極為密切。任何教師都熟悉牠，而且現在依然採用牠。牠曾遭受過極嚴厲的批評，不過，我們雖能發現牠的各種缺點，但仍有多優點，足以保證牠有改良和應用的價值。所以牠仍將為教師所繼續採用。不過對於牠的缺點應詳加討論，以便校正或減少，牠的優點應予承認，以便利利用。總之，凡足以使牠改良的方法，都應採用。關於論文式考試的各種優點、缺點，以及改良的方法，本章都要討論。

作文的正確測量——論文式的考試是作文能力的正確測量。要測量文字發表的造詣，唯一的方法就是

叫學生做一篇文章。欲達這個目的，除用論文式的考試外，別無他道。就作文本身的性質而論，不易用客觀的測量。文章要寫得準確、合理和有趣，其中所包含的因子，極爲複雜。所謂文字的發表，其實就是輯辭、方法、段落、書法、標點、文句構造、文體和諧、字彙等因子的複合。新式測驗僅能測量這些因子的技術上的知識，而不能代替舊式考試以測量作文的能力。

所以論文式考試宜於測量作文的能力，這就是牠的優點之一。不過，當這種考試用以測量作文以外的因子時，這種優點便有變爲缺點的可能。舊式測驗既極宜於測量作文能力，則常用以測驗其他成績時，每多把作文能力和其他學業成績互相混雜。論文式考試的歷史分數，大約是歷史成績和寫作能力的一種混合分數。在這種情形之下，因爲時間的不足，不能審慎寫作，或不能及時完卷，文章也許做得很壞。論文式的試卷，因爲書寫的匆忙與疏忽，常多寫錯的字和不合文法的結構。

編造和施行的便利——論文式的考試，編製和施行都很容易，而且對於任何地方性的或普遍性的教材都能適用，這是牠另一種優點。在測驗中，恐怕沒有比編製五個或十個論文式的試題再簡單了。不過唯其編造太易，教師所費的時間和心力便往往過少。其實，編造論文式的試題，雖是一樁簡單的事，但欲求其優良無疵，却需精密的思考。至於施行論文式考試時所發生的困難和障礙，也是無足輕重的。問題既可板書（按即寫在黑板上）也可口述，印發學生使各得一份，也無不可。無論如何，這是無關緊要的。教師所欲測驗的任何教材，牠們

都易適用；對於教師所在工作的任何狀況，都有很充分的適應性，以應付其需要。論文式考試既易於編造和施行，故教師對之便生滿足之感，有時，且因此引起他們一種信仰，以為牠已能充分適應各種測驗的需要。

對於優等生的價值——論文式考試對於優等生有幾種特殊的價值。牠能供給機會以顯示學生的推理方法，創造才能，以及組織材料的能力等等。這種價值在他種測驗中，從沒有這樣明顯。此外，牠允許學生自由作答，故於成熟的學生很有利益。蓋自由作答足使學生獲得一種機會，以練習識別力和評判力。而且在論文式考試中，學生可按自己能力的強弱以說明問題和陳述答案，故從這裏也發生幾種益處。學生在應付論文式的考試時，可依照他自己的智力能量。

試題中如用下列項目，可以激發和啓示學生的思想：

分類……就是依照某種共同特徵，把事物齊集、排列、分配和組合。

比較……就是辨別許多事物的異同。

批評……就是對於題目下良好的判斷和審慎的分析。

定義……就是決定題目的範圍，並確立其意義。

敘述……就是把形成題目的性質和情形，加以選擇和描寫。

討論……就是把題目詳細研究，提出正反兩面的考慮，並引證理論或理由以維護其所持之立場或態度。

說明……就是刪除足使題目模稜的各點，使其明白通暢。

例證……就是舉例以說明題目。

解釋……就是以個人的信仰或判斷，闡發論題的意義。

辯證……就是要表明事物的合理與否或根據所在。

大綱……就是在討論、辯論或程序中，揭示大綱，或指明其要點。

溫習……就是把題目細讀一遍，且加以精究。

摘要……就是以簡括的嚴密的形式陳述題目。

重述……就是把某個題目詳加發揮。

證明……就是證明一事之真偽。

上表，尚可以其他相似問題之包含原因、結果、理由等等者擴充之。這些問題果能審慎編製，而又適合學生能力，確有很多的價值，足以證明論文式考試對於優等生的功用。

論文式考試的缺點，——論文式考試雖有上述許多優點，但同時也有不少缺點。除了便利一點外，牠缺乏前章所討論的優良考試的各種特徵。客觀性和廣博性為可靠性所必需的條件，而論文式考試獨付闕如。除了測量作文能力以外，論文式測驗缺乏正確性和實用性，融洽性也不完備。為討論便利計，牠的最大缺點，可分

類如下：

1. 主觀的記分。
2. 狹隘的範圍。
3. 不相干的因子。
4. 學生的反應。
5. 狹隘的用處。

論文式考試的主觀記分——論文式考試有一個最大的缺點，就是缺乏客觀性。試卷的優劣，不足以決定分數。考試的分數不過是分等者主觀判斷的結果。關於這點，實驗的證明，已不一而足。司且區 (Sturton) 和 伊里歐 (Elliot) 二氏的研究實為極重要的參考資料，足以表示教師對於論文式試卷所批的分數是變化無常的。在一個實驗中，(註一) 二氏以二個中學學生的英文試卷，請 15 位英文教員批分數。這些教師對於一份卷子所批的分數自 54 至 98 不等；另一份卷子則自 50 至 98 不等。這二份試卷，又叫 威斯康辛 (Wisconsin) 大學英文教學班上 68 位學生去批分數。他們所批的分數，一份卷子自 72 至 98 不等；其他一份則自 65 至 98 不等。這二份試卷，又請該大學教育測量班上 68 位高才生去評定分數。他們所定的分數，一本自 68 至 98，另一本自 64 至 98。在另一個實驗中 (註二) 司且區 和 伊里歐 二氏選定了一份平面幾何的試卷來記分。所以要選擇幾何因為普通以

爲算學試卷的記分，總是具有算學的精確。他們請了二百位算學教師去定那份平面幾何試卷的分數。結果自100分至80分不等。他們又請200位歷史教師定一份歷史試卷的分數，結果自60分至30分不等。這幾種實驗可以表示：教師對於試卷所定的分數，差異很大，而且這種差異，各科都是有的。

馬頓氏 (Morton) (註三) 也曾將教師檢定考試的記分，作一種類似的實驗。他在俄海俄 (Ohio) 州一個府教育局的教師檢定考試中，任意的抽出三份試卷來。一份是數學，一份是地理，還有一份是教學的理論和實際。每份有十個問題，選作八個。他把這些問題和答案，同原稿一樣的印刷了，送給該州的各府教育局長，再由教育局長轉請府內負有這種考試責任的考試員，批定分數。數學試卷有20位考試員定分，他們所給的分數自100至80不等。地理試卷有20位考試員定分，分數自100至80不等。教育理論和實際的試卷也有20位考試員定分，分

(註一) D. Starch and E. C. Elliot. "The Reliability of Grading High-School Work in English" 載在 *School Review*, Vol. 20, p.p. 442-570 (1912年9月)

(註二) D. Starch and E. C. Elliot. "Reliability of Grading High-School Work in Mathematics," 載在 *School Review*, Vol. 21, p.p. 554-49, (1913年4月)

(註三) R. L. Morton. "The Examination Method of Licensing Teachers," 載在 *Educational Administration and Supervision* Vol. 6, p.p. 421-32 (1920年9月)

數自55至94不等。如把每個問題的最低分數總加起來，以百分比論，那麼所得的分數是：數學28分，地理16分，理論與實際32分。如把每題的最高分數總加起來，每種卷子都是100。

論文式考試的記分，真是徒勞無益的，看烏特氏 (Wood) (註1) 的報告，愈覺明白。烏氏的報告是說五六個哥倫比亞大學的歷史教授，在1920年之夏，批定許多歷史試卷的情形。他的報告中說，在這五六個專家中，有一個於批定了幾本試卷以後，就找出一本他認為模範的試卷，作為指南。這本試卷復混入其他許多卷子中。後來這本「混入的模範試卷」由其他各記分者給予分數，自60分至80分不等。

此外，我們還能引證許多其他的實驗，以表明論文式考試的分數，依據於記分人者多，而依據於受試人者少。陶格蘭斯 (Douglas) (註2) 發現同一教師，隔了相當時期，如重記試卷的分數，就前後懸殊了。其他的實驗也歸於同一的結論，公認各人對於同一試卷，在不同時間內，重給分數時，差異很大。所以論文式試卷的分數很不可靠，這一點已無庸置辯了。將牠作為學生作業的正確測量，不能獲得人們多大的信仰。關於這個缺點，記分人並無何種不名譽。人類無標準的判斷本來是不準確的，既易陷錯誤，且易被欺騙。各人的標準互殊，他們的偏見亦不能相同。即使同一個人，也是時時變更的。教師對於論文式考試的記分有很大的差異，固屬可驚，但知道了他們記了相反的分數時，當尤覺驚奇了。實驗的證據，僅足以證明從人類無標準的判斷中，能合理的希望些什麼。

各教師所給分數不同的一個大因子，是因為他們對於考試中每個問題所下的衡量或價值不同。摩明 (Robert Comin) (註三) 發現教師對於試題難度的估計，其不可靠同其他主觀的判斷一樣。他請了 8 位教師按難度排列 50 個算術問題。這些教師是六年級的教師，這些問題是適合於六年級學生的問題，然而他發現這些問題可以被各教師歸入各種不同的難度之內。例如第二十三個問題，自第一級至第二十一級之間，各級都有。這些教師又估計他們自己的學生解答這些問題的能力。平均的估計是這樣：如果每一問題能給以十分鐘的時間，那麼有 85.8% 的學生能把牠們全數解答。於是便將測驗施行。實地測驗的結果，僅有 54.3% 的學生解答了這些問題。這個研究足以表示教師既不能判斷問題的難度，也不能估計學生解答問題的能力。對於問題重要性的判斷，也是同樣的各不相同，各教師非但對於各試題的難度或重要性估計多有不同，就是對於難度或重要性已經決定的問題，各人所下的衡量或價值仍然不一致。答準一個困難的問題，比答準一個容易的問題，表示更大的能力。答錯一個容易的問題，所表示的無能，比答錯一個困難的問題為甚。計算最後的分數

(註一) Ben Wood, "The Measurement of College Work," 載在 *Educational Administration and Supervision*

(註二) H. R. Douglass, "Modern methods in High School Teaching," p. 368.

(註三) Robert Comin, "Teachers, Estimates of the Abilities of Pupils," 載在 *School and Society*, Vol. 3.

時，應該着重最難的問題呢，還是應該着重最容易的問題？關於這點，意見紛歧。所以主觀的記分爲論文式考試的最大缺點。

論文式考試的問題範圍太隘——論文式考試的另一個顯著缺點，就是問題的範圍太狹隘。模範的論文式考試僅有五個乃至十個的問題。這幾個問題太少了，不足以爲學生真正能力的適當的取樣。證據既然不足，取樣亦復粗劣簡單。僅須編製少數的問題，反足使教師草率從事。對於問題的選擇如僅略加些微的思索，或不加思索，這些問題便無意義之可言，同課程的目標更是不相符合。論文式測驗的範圍既太狹隘，則對於問題的選擇，自然難以公平。一個學生或許偶然記住了測驗所包括的一部分教材，一個問題的措辭或許很不明白，以致學生不懂或誤解，這都是不公平的原因。在一個論文式考試中，學生常常不知道所要回答的是什麼，也不很明白回答至什麼地步才算完畢。至於教師呢，他要想批定學生對於一個簡單問題所答的冗長而複雜的答案，其困難比較學生要想知道問題還大。因此記分上易於錯誤。問題很少時，一個題目的不公平，影響總分數很大。如果加多題目，使論文式考試臻於適當，則又失之太長，而不合實際目的了。總之，廣博性這個原理，在論文式考試中，顯然是談不到的。

記分上不相干的因子——在最近出版的一本新聞雜誌上，述及美國某州立大學的女生。她們在聽講以後，逗留著與教師談話。上課的時候，她們坐在靠近講台的地方，聽了教師談話的話，就作啞啞的媚笑；看着教師

的奇形怪狀，故作笑容；當教師講到要點時，她們杏眼斜睨，含情脈脈。這樣一來，這些詭詐的女生，所得的分數，便較毫不動人的男生高得多了。另一州立大學的校長曾勸告其教師，謹防大多數男女同學的學校中所流行的機詐的影響。上述的論斷，含有很多的真理，非普通所能想像。許多學生一般的態度，也大有影響於班級地位和考試分數。有的學生甚至在初小低年級的時候，已經學會了博得教師好評的各種姿態。他們雖則愚笨與疏忽，也要故意裝着聰明和注意的樣子。教師的判斷，在有意無意間，多少要受外來因子的影響。這種因子與評判學生的分數，毫無關係。論文式測驗的記分，易受外來因子的影響，這是牠最大缺點之一。

爲便利計，影響考試分數的不相干的因子，可以分爲下列三方面：

技術方面——技術方面的不相干的因子，包括不流利的國語、錯字、別字、書法和整潔等等。在試卷上學生所用的國語，常足以影響成績的分數。標點用得當，文法的結構不良或不準確，都可以使分數降低。這種情形對於有語言困難的學生特別不公平。不能用文字把自己的意見表白清楚的人，成績分數常被減少。此外，寫錯了字，考試分數也常常因而減少。現在有一種趨勢，就是文筆優美的試卷，所得的分數較文筆拙劣的要高；書法拙劣的試卷，所得的分數較書法挺秀的要低。一個層次分明，簡淨扼要，或外表整潔，形式美觀的考卷，大概易於獲得高的分數。有的時候，答案的長短，亦爲一重要的因子。有些寫得慢的學生，當時間有限而不能完卷時，分數就要減低。

原諒方面——原諒方面的不相干的因子，包括進步，努力，能力，以及校外工作等等。有的時候，教師因為注意了學生已有的進步，所記的分數就超過了學生成績所應得的。有的時候，又往往注意到學生特殊的努力和專心。學生因為努力的緣故，每可獲得實在能力所不能求得的分數。教師也許會覺得某生的能力實在比分數所表示的要好，但因某種理由，不能如願以償，這樣一想，分數就提高了。對於非常聰明和非常能幹的學生，或在別方面曾表示過出類拔萃的技能者，特別會發生這種情形。現在有一種趨勢，就是對於做校外工作或從事課外活動的學生，教師記分時往往曲予原諒。

人的方面——個人方面的不相干的因子包括下列各種影響：如對於學生之一般的意見，他的禮貌，莊重態度，善意，誠實，舉止，興趣，誠心，與一般態度；以及回答教師的癖見，給學生並未述及的意見以分數，分數對於學生將有的影響，父母的勢力和地位，與教師的諂媚等等。一個學生從前的學校成績，他人的批評，日常工作，的印象，以前與教師共同工作的品質，以及能力測驗的結果，有時大可以決定考試的成績。有些學生曾將自己所做的工作互相對調，然後繳卷，但也照樣的獲得或高或低的分數，好像繳自己的工作一樣。開首幾個問題上所獲得的一般印象，有時往往為決定總分數的重要因子。學生的文雅，溫柔，和沉靜，都有相當影響。外貌佯裝着考慮或沉思，常常發生作用。

學生的反應——論文式考試的另一缺點，就是學生對牠常生不快的反應。有許多人想到了將臨的考試，

就起極大的杞憂。他們在試期以前就憂慮到考試的結果。試期日近，他們的不快之感也隨之而增。結果，試期當前的時候，他們或許已到身心俱憊的地步了。在這種情形之下，再加上一二小時的寫作，更足引起神經上的緊張。他們在機械的抄寫上和答案形式的計劃上，所費的時間與考試所考的知識相較，往往失其平衡。在教師方面也須費幾小時的工夫，去批改試卷。因此，論文考試在編製上和施行上雖有便利，也被一筆抹煞了。學生對於考試常生一種有害的情緒，這是爲了要求無益與不確定的結果而遭到的極大犧牲。

論文式考試的性質，常使學生覺得結果不公平。學生自然不能知道牠的不可靠性，像專門研究過論文式考試的人一樣；但他把他的分數和旁人的相較，就會看出其中的不公平來。如果他把分數差異的原因去問教師，也得不到滿意的解釋，因爲這常常是無理可講的。於是他對於考試就形成一種不信任的態度。所以，學校裏的任何方法，如果牠的性質有使學生感覺待遇不公的機會，我們對牠就應該懷疑了。除此以外，論文式考試還鼓勵學生「虛張聲勢」。在論文式的考試中，憑着作文的能力，可以強不知以爲知。真正的義蘊儘可不知，而試卷上却可滿寫着似是而非的幌子。爲教師者常爲這種假裝的言辭所惑，而受其欺騙。「虛張聲勢」是一種欺騙的形式，有的學生行之極熟。論文式考試的本身，就適合於這種形式。與「虛張聲勢」相魯衛者，便是「急來抱佛脚」。論文式考試也鼓勵「急來抱佛脚」；因爲牠的問題既不多，涉及的題旨便很少。學生們往往利用機遇，預猜題目，憑着「急來抱佛脚」的方法去預備。這種匆促的表面的閱讀，並無永久的價值。牠的唯一合理的

的功用是在材料缺乏的時候。一個優良的考試應能鼓勵有系統的溫習，而論文式的考試則不能。論文式測驗的所以能一急來抱佛脚，一因為如此預備的材料，縱然猜不到正確的要點，也常能用以一虛張聲勢，一論文式考試的性質易於鼓勵或養成這些不良的反應，這是由於缺乏融洽性的緣故。

結果的用處太隘——縱把傳統的論文式考試的結果，看作是知識或能力的絕對衡量，其用處也很有限。百分一數的意義，就是學生僅有完全知識的四分之三。每個學生的絕對知識認為與其他的學生無關。測驗上的零分，是表示在所測驗的範圍內毫無知識或能力，而100分則表示完全的知識或能力。不過，什麼是完全的知識？學生應該知道多少？有多少知識才算及格？諸如此類的問題，純粹一任教師或學校當局主觀的獨斷，毫無客觀的標準。學生所得的分數，依據於學生的知識者少，而依據於記分人的寬嚴和考試的難易者多。如果考試容易，或記分人寬鬆，則學生易於及格，而且能得高的等級。如果考試困難，或記分人嚴緊，則學生就不易及格，或僅能得到很低的等級。分數的用處貴能說明同一考試上各人所得成績的關係。牠至多只能表示學生在某一考試上的成績，而不能表示無知與全知間的地位。論文式考試頗難說明所得的分數，因為牠的本身就不適於建立常模或標準。所以論文式測驗的結果，並無多大的意義或川處。

論文式考試的改良方法——要想改良論文式的考試，有幾件必須做的事情。考試所要達到的真正目的，應先行決定。考試中每一題目，要考察其是否合於所欲達到的目的。問題應與課程上教材的目標相適合。關於

這點，可從下列各項中得到有益的建議：學校當局所訂定的最低標準大綱；考試範圍內的各種教科書和課程；課程組織專家所做的課程研究；標準測驗和量表上所有的題目；從教學過程上教師每天所累積的問題。考試的內容應為牠的目的所支配。考試應該適合於所要測驗的學生。牠的難度應在被測驗學級的能力範圍以內。更要者，則為制定規則以防記分上不相干的因子。如果要注意到拼法，書法，作文等等，應該把牠們另行記分，不使與考試的目的相混。凡與記分有關的各種因子，應預先確定其價值。這種因子就是答對的原則，部分準確的答案，材料的組織力，整潔等等之計分。如把各種試卷分次記分，每次注意一個因子，能得最好的結果。在批改試卷以前，如寫出準確的答案，用作標準，以與學生的答案相對照，這是一個很好的辦法；可使記分程序中始終用同一標準。在這裏，也應如第十一章中討論的一樣，將分數和等級加以區別。

溫習與討論問題

- (1) 十一？ 現在僅有少數教師用論文式考試。
- (2) 十一？ 論文式考試仍有實際的價值。
- (3) 十一？ 標準測驗為非標準測驗的良好代替。
- (4) 十一？ 論文式考試不適於作文測驗。
- (5) 十一？ 論文式考試編製頗易。

- (6) 十一？ 論文式考試對於優等生並無用處。
- (7) 十一？ 論文式考試缺乏優良考試的多數特性。
- (8) 十一？ 論文式考試的記分易於客觀。
- (9) 十一？ 教師對於測驗問題的相對難度，判斷極為一致。
- (10) 十一？ 在論文式考試中有不公平的因子。
- (11) 十一？ 有的學生每假裝各種姿態以影響教師的記分。
- (12) 十一？ 有的教師在分等時能避免各種不相干的因子。
- (13) 十一？ 記分上各種不相干的因子，都是屬於人的方面。
- (14) 十一？ 論文式考試易引起學生的不良反應。
- (15) 十一？ 「急來抱佛脚」有時是可用的。
- (16) 十一？ 論文式考試的分數，易於說明。
- (17) 十一？ 各校應廢除論文式考試。
- (18) 十一？ 論文式考試可以改良。
- (19) 十一？ 論文式測驗的主要缺點是取樣太受限制和記分太主觀。

新法考試

第五章 新式考試

新式測驗的主要特徵——新式測驗與論文式的考試相較，有幾個顯著的特徵。牠的編製問題非常巧妙，所以可用最少的時間，測量出學生最多的知識。編製的方法在使學生少用書寫，或簡直不用。學生對於答案的形式可以不很注意，因為這種答案只包括一個單字或名詞，一個數目，一個校對符號，一個圈，或一條線。總之凡足以減少學生寫答的諸種方法，無不採用。這樣能使一個測驗，包括很多的題目。通常的題數總在75至100之間；150或150個以上，也常有的；至於50個以下的，就很少了。問題既多，足使測驗廣博而公平。各題答案所需的時間既速，可使測驗在一個合理的時間內——一小時左右——完成。問題通常是印刷的，答案就由學生寫在問題紙上。新式測驗的批閱是很快的，可以應用標準答案或由學生自己批改。記分的手續純粹是機械的，是一種抄錄的例行工作，無論什麼人都能機械地照做。記分者主觀的判斷並無影響。任何人去記分，分數總是一樣的。新式測驗的記分，既極客觀，所以我們有時叫牠客觀測驗。

客觀測驗的普通形式——客觀測驗有二個普通形式，一為回憶式(Recall)，一為認識式(Recognition)。回憶測驗由許多句子或段落組成，其中重要的部分，留着空白，囑令學生填寫，使成正確完全的語句。回憶這種問題的速率，各生不同，每視其對於教材熟習程度和個人差別而異。最常用的回憶測驗，為簡答測驗(

Single-Answer test) 和填字測驗 (Completion test)。

簡答測驗有許多問題，每個問題只用一個單字就可以回答，至多也不過用一句話在測驗卷的邊上留着空白，以便學生填寫答案，這樣排列，可使教師用正確的答案紙板對照，以核對他的錯誤，藉定測驗的分數。分數就是正確答案的總數。下面是一個簡答測驗的例：

× 中華民國第一任的正式大總統是誰？

× 新疆省的省會是什麼？

× 武昌在那一省？

爲什麼有春夏秋冬四季的分別？

× 萍鄉以什麼著名？

巴拿馬運河 (Panama Canal) 如何使商業繁盛？

填字測驗與簡答測驗不同，牠在一句中缺一字或數字，或者在一段中缺一字或數字。這種測驗的答案，批閱時難於應用標準答案紙，因爲牠們依着所缺各部的地位，四散在卷子的各處。茲舉一個填字測驗的例子於下：

× 法國的國都是……

在美國以南的國家是……

發現太平洋的是……

哥倫布大約在……年生於……城。他要向着……航行，以達……

……所以他去游說……，……等國的元首，請求幫助。大家都……

幫助。最後……國的女皇願意供給……，才能整裝遠征。

在認識測驗中，問題下寫着許多預擬的答案，叫學生從中認出準確的或最好的一個。認識式中最常用的是選答，對偶和正誤等測驗。

選答測驗包括許多句子，每句各有幾個預擬的答案，以便填入。預擬的答案通常自二個以至七個。最好的選答測驗，每句有五個可能的答案，叫學生在最對的和最好的一個下面劃一條線，或者把牠的符號數寫在該句的前面。爲欲應用標準答案紙，以便批閱起見，學生填寫答案的地方，是預先規定好的。分數就是準確答案的總數。以下的例子，便是選答測驗。

……殖民於惹米斯敦 (Jamestown) 的是 (1) 英國人 (2) 荷蘭人 (3) 西班牙人 (4) 法國人 (5) 意大利人。

……世界金鑽最多的地方是 (1) 巴庫 (2) 古巴 (3) 墨西哥 (4) 開普敦 (5) 馬尼刺。

……世界上最大的城市是(1)巴黎(2)紐約(3)倫敦(4)芝加哥(5)倍諾斯愛勒(Buenos Aires)
如把上列各題的形式略加改變，使包括着知識的應用，就比較難答而更有價值，如：

……在1901年，惹米斯頓的居民中最多的是(1)荷蘭人(2)法國人(3)英國人(4)意大利人(5)西班牙人。

……我們要開金礦，最好在(1)開普敦(2)墨西哥(3)巴庫(4)古巴(5)馬尼刺。
……世界上最大的城市，在(1)紐約省(2)英格蘭(3)法蘭西(4)麻塞邱斯 Massachusetts (5)中國。

對偶測驗由兩組相關的名詞或事實組成，其次序是隨機排列的。學生須指出其相屬的各條目。在第一組的名詞前留着空白，在第二組前則寫有數字。學生把第二組中某一條的數字，寫在第一組中與其相屬的一條前空白內。做對的總數，就是所得的分數。下面是一個對偶測驗的例子：

第一組

第二組

- | | |
|--------------------------|-----------|
| …… <u>哥倫布</u> (Columbus) | (1)英國物理學家 |
| …… <u>俾斯麥</u> (Bismark) | (2)美洲發現人 |
| …… <u>柏格森</u> (Bergson) | (3)法國大將 |

- ……麥哲倫 (Magellan) (4) 葡國探險家
- ……莎士比亞 (Shakespeare) (5) 英國博物學家
- ……達爾文 (Darwin) (6) 德國政治家
- ……牛頓 (Newton) (7) 英國文學家
- ……福煦 (Foch) (8) 法國哲學家

正誤測驗包括許多語句，其中大約一半是對的，一半是錯的，學生須分別指出。測驗分數通常是正確的答
案數減去錯誤的答案數。以下是正誤測驗的例：

- 十一 鐵的利用較銅爲早。
- 十一 腓尼基人曾以語音的著作貢獻於世。
- 十一 據稗史的記載，來喀古士 (Lycurgus) 是斯巴達的立法者。
- 十一 德累科 (Droco) 編纂斯巴達法典。
- 十一 伯里克里斯 (Pericles) 是一個亞典的皇帝。
- 十一 亞歷山大是一個埃及人。

測驗形式的選用——在不同的情形之下，測驗究用何種形式，教師須詳加考慮，然後決定。有的時候，形式

的選用，僅取決於教師和學生的喜歡與否。正誤法是師生所共同喜歡的，所以應用常極普遍。然而測驗形式的選用，不應祇顧到師生的喜歡與否，應加考慮的事項甚多，如所要測驗的材料性質，測驗中所重視的心理能力，變化的原則，測驗所要完成的重要功用，以及舉行測驗的便利等等。

所要測驗的材料性質，對於測驗的選擇，關係至重。有的材料，其本身最適用於某種形式。教師們在應用各式測驗的經驗中，定會感覺到這一點。變化的原則也須注意。測驗有了變化，方能增加測驗的興趣，否則便要感覺單調。回憶和認識二式應該兼施並用，使這二種能力都得運用。回憶一事的能力和認識一事不同，前者比較困難。除了回憶和認識簡單事實知識外，教師尚須注意各種心理的能力，如判斷力，關係的知識，和學識在實際情況上的應用等等。至於何種測驗適合於測量何種能力，當於以下各章討論之。選擇測驗形式時，在第二章中所討論的測驗的重要功用，也要記着。教師不難發現，有些形式比其他形式易於達到某些功用。

教師的意見，有時常為學校的設備所限，如果沒有印刷的便利，便只好用板書法或口述法。對偶式和填字式不易用口述法，但可以寫在黑板上。正誤式不適用於板書法，但很可以用口述法。如有印刷測驗的工具，則揭示測驗題目的方法，就無關於方式的選用了。每個學生各有一份題目紙，則任何方式都能令人滿意。至於某種特殊形式的選定，應視整個測驗情景以為斷。許多教師發展其編製各種測驗手續能力，以適應於各殊的材料，能量及功能。

客觀測驗的特殊價值——新式測驗在教育實施上所以成爲重要部分，因爲牠具有特殊的價值。亞爾馬

克(Almack)和蘭(Lang)(註1)列舉新式考試的優點如下：

1. 新式考試免除記分的主觀——記分都用同一的方法。不論記分者爲誰，其最後結果，總是符合的。記分者判斷良好答案構成的條件，不能因試卷或時間而改變，學生也不能因教師的喜歡便得高的分數；大家都受公平的待遇。

2. 新式考試免除模稜兩可的答案——學生的答案要簡明潔淨和確定。他們如用搪塞的言詞或近乎正確的答案，來裝飾試卷，那是不能通過的。他只有「知道」與「不知道」，決無模稜兩可的餘地。

3. 新式考試免除不相干的反應——和測驗不相關的答案一概刪去。

4. 新式考試所包括的材料，範圍甚廣——所含的材料比論文式考試要多五倍至二十倍。

5. 新式考試有精細的記分單位——問題甚多，所以每個問題對於總分數的影響極小。學生或教師，或因誤解問題，或因問題含糊，或因誤記分數而發生的錯誤，在實質上並不影響總分。

6. 新式考試的記分單位是標準化的——記什麼分數，以及如何記分，都有明確的規定。答案只有正誤兩方面，給分與否，並無討論的餘地。

7. 新式考試中不相干的因子，不致影響總分數——判斷一個答案的準確與否，不受書法，綴法，整潔，

人格，或學生勢力等等的影響。

8. 新式考試易於施行和批改——牠們施行所需的時間，較論文式考試為短；牠們並不需要多寫，所以不致使學生多耗精力。牠們可由學生用標準答案紙片，自己批閱。而且同一測驗可以用好幾次。

9. 新式考試可以產生標準——教師如把各班的記錄都保存起來，便能逐年比較學生的成績，及自己教學的效率。

10. 新法考試是富有興趣的——教師喜歡編造客觀的測驗，學生喜歡受客觀測驗。他們同迷津一樣的有趣。

11. 新式考試是富有適應性的——其中有一種或一種以上的形式可適合任何學科。

詹姆氏 (James) (註二) 舉出新式測驗所以優於舊式測驗的理由如下：

1. 材料的範圍較廣。
2. 學生易於信服分等的公平。
3. 書寫較少，思想時間加多。

(註一) J. C. Almack and A. R. Lany, "The Beginning Teacher," p.p. 225—28。

(註二) B. D. James, "The Modern Test," School and Society, Vol. 23, p.p. 309—13, (1924年2月)

4. 不能以流利的文筆來掩飾「不知。」
5. 學生興趣較好。
6. 可以明瞭學生以後生活上所需的準備。
7. 節省教師的精力。
8. 需要規律的和時常的溫習。
9. 在最終分等時，可以免除憂慮。

新式測驗的明確有定，師生都感滿意。學生明白所問的是什麼；當他們照答的時候，他們已經知道了。新式測驗對於有語言困難的外國學生，也能適用；對於文字不很清楚的低級兒童，也可適合。新式測驗能創造與實際生活相符合的情境。牠能使學生對於某種問題作公正的觀察，為直接的反應。我們生活上最多的情狀是「是」、「非」、「正」、「誤」的反應，和其中最優的選擇。在實際生活上需要寫一篇文章的反應是很少的；即或有之，也是「開書」式的，其中材料可於需要時臨時參考。新式測驗的特殊價值，不勝枚舉，都是很重要的。

學生為什麼願受新式測驗——據著者的經驗，學生極願意受新式測驗，最足表示的一個實例，是教育原理班上一百二十一個學生。這一班學生中，有一一五人，願受新式測驗，五人願受論文式考試，還有一人不置可否。克列斯 (Klise) (註1) 在初級心理班的三三七個學生中，發現有二六五人或百分之七十九願受新式測

驗，六一人或百分之十八願受論文式考試，其他十一人或百分之三則疑惑不決。這六十五個學生曾述願意受新式考試的理由，茲以贊成人數多寡為序列之於下：

1. 學生可以考核自己的知識。
2. 材料有較多而較大的變化。
3. 比較迫近生活情景。
4. 措辭與記分無關。
5. 需要思考和判斷。
6. 可得比較公平的等第。
7. 不能「急來抱佛脚」。
8. 不能「強不知以為知」。
9. 學生不感受時間的匆促。
10. 減少緊張和興奮的情緒。

(註1) N. M. Kline, "Students' Opinions of Types of Examinations," 載在 *School and Society*, Vol. 24, p. 23—24.

(1926年7月3日)

11. 誤解問題的機會較少。

12. 學生能明瞭錯誤所在。

因這種種理由，大多數的學生願受新式測驗，而不願受論文式考試。他們舉出很好的理由，來說明願受新式測驗的緣故。新式測驗的融洽性很高，這是優良考試的一個重要特徵。

推理的測驗——有人根據二種理由以批評新式考試，一種是認為新式考試只要記憶，一種是認為新式考試不能測驗推理能力。這種批評至少根據二個可疑的假設，一個是認思考和記憶為絕不相關的心理作用，另一個是認問題本身的形式就足以引起推理。就現在心理學的知識而言，要在測驗上明分思考和記憶是不可能的。烏特 (Wood) (註1) 說得好：「事物並不孤立的存在心中，我們由思考來記憶，也由記憶事實來思考……：至今日為止，已做過的和已報告的實驗研究，都表示某範圍內的知識測量，與該範圍內思考材料的能力或智力有很大的關係。」更進一步言之，推理並不受測驗形式的嚴格束縛，牠大半視學生固有經驗及反應而定。一個學生的思考問題，或許是另一學生的事實問題，須視其反應的準備情形而異。所以測驗的問題，不能離開了被測驗的某一學生，而分為思考問題或記憶問題。

對於成熟的學生，論文式考試能夠供給機會，以訓練創造的能力，若在新式測驗便不能明白具有。有些價值已於上章論及。至於推理的問題，並不如一般人所言。一個新式測驗的問題，可包括很多的推理，而一個論文

式考試也可以僅需極少的推理。論文式的考試問題，其編製雖可供給注意學生推理順序的機會，但如將教師所編造的思考和事實問題一加研究，就知這樣區分，成功殊少。新式測驗能夠表示推理的結果，所以，以不能正確測量推理為理由而反對新式測驗，是不健全的。

貝脫孫 (Paterson) (註一) 主持一個實驗，這個實驗表明論文式和新式測驗，都趨向於測量同一的心理作用。他給一班學生以二個同內容的論文式考試，而尋得兩考試中結果相符合的範圍，相關係數是 $.52$ 。後來他又給他們以一個內容相同的新式考試，而發現結果與論文式考試相符合的範圍，同論文式考試自相符合的範圍一樣，相關係數也是 $.52$ 。論文式考試的結果與新式測驗的結果，其互相符合的程度，既與論文式考試自相符合的程度一樣，則這兩種測驗，不論其測量思想或知識，所測的必為相同的東西，那是很顯然的了。此外，他又發現二個新式測驗的結果，較兩個論文式考試的結果，更易互相符合。其相關係數是 $.67$ 。所以除了兩種考試都測量同一東西外，新式測驗所測量的且較舊式為佳，這是很明顯的了。

新式測驗中的作弊——我們能使新式測驗較論文式考試難於作弊。有的學生往往極其能事以謀作弊，

(註一) 採自 Ben Wood 的 Measurement in Higher Education P. P. 162—63.

(註二) D. G. Paterson, "Do New and Old Examinations Measure Different Functions?" 載在 School and Society, Vol.

24, p. p. 246—48, (1926年8月21日)

所以考試務須盡量防護。在擁擠的教室中，一個學生只要向旁人的試卷一瞥，便很容易偷看到文章的內容。至於新式測驗，雖在擁擠的情況之下，也能預防作弊的機會。譬如測驗的頁數，可不標明數目，並坐的學生可給以不同的卷子而令其同時工作。如果問題上也不註明數目，學生便難於就旁人的試卷核對問題。聰明的教師，能在新式測驗上想出許多預防作弊的方法，以應付特殊的環境。

而且在新式測驗中檢查作弊也比論文式考試容易。在論文式考試中，學生能抄襲文章的內容而變換語句，以毀滅其犯罪的證據。在新式測驗中要防免破綻就比較爲難了。材料未曾準備過的學生，每將錯誤的和正確的答案一併抄襲。他不能辨別孰是孰非。所以有作弊嫌疑的試卷，可與認爲所從抄襲的試卷相對，能夠算出他們同錯的數目。如果兩份試卷上同錯的答案數大於假定的機遇數，那麼便可很合理的假定他有抄襲的行爲。在二份試卷上，有過多的同錯答數，這是作弊的積極證據。

盤特 (Charles Bird) (註1) 曾舉有應用這種方法以檢查客觀測驗上作弊的例證。他在一個二六六個學生的心理學班上，舉行二小時客觀測驗，共有一四九個問題，其中有四個學生，似乎在偷看旁人的卷子，一小時後，警告其中一個學生而令其調換位置，測驗終了之後，把四個學生的試卷，與認爲所從抄襲的旁人的試卷兩相核對。結果，同錯的題數各爲17, 25, 28及31。受警告的學生，在第一小時內有十五個同錯的題目，但在第二個小時內，僅有二個。這就可很顯然知道他在第一小時內作過弊了。他又隨機的取出二十個學生的試卷，

與其隣坐學生相較，平均有四個相同的錯誤。這個證明足以毫無懷疑的顯示四個學生的作弊。以後再取一百份卷子，兩相核對，結果平均有 $\frac{1}{3}$ 個共同的錯誤。用統計的方法，可以算出每個學生作弊的機遇。新式測驗能以算術的精確來檢查作弊。檢查與判決的正確，其本身即應為預防作弊的重要因子。

關於作弊問題，新式測驗比論文式考試具有明顯的優點。在論文式考試上，教師的猜疑大部份是自己的意念，和主觀的判斷，通常並沒有充分的證據，以資明證。證明的責任，完全在告發者的身上。有力的犯罪的證據尚須加以補充。新式測驗能供給客觀的證據，使學生無從否認。隨便的教師對於這一類的測驗，如應用不慎，則也會發生極大的作弊，但這是由於教師的關係，並非測驗本性所致。

溫習與討論問題

- (1) 十一？ 批閱新式測驗必須逐句閱讀。
- (2) 十一？ 新式測驗的記分，須有極大的判斷力。

(註1) Charles Bird, "The Detection of Cheating in Objective Examinations," 載在 *School and Society*, Vol. 5, p. p. 261-62. (1927年2月26日)

Charles Bird "An Improved Method of Detecting Cheating in Objective Examinations," 載在 *Journal of Educational Research*, Vol. 19, p. p. 341-18. (1929年5月)

- (3) 十一？ 有二十五個問題，就可構成一個優良的新式測驗。
- (4) 十一？ 客觀測驗有二種普通形式。
- (5) 十一？ 簡答測驗與正誤測驗相同。
- (6) 十一？ 填字測驗的記分較對偶測驗為難。
- (7) 十一？ 五中選一的簡答測驗較二中擇一的為優。
- (8) 十一？ 師生通常都喜歡填字測驗。
- (9) 十一？ 選用測驗形式時僅須顧到教師的願意。
- (10) 十一？ 新式測驗惟一的優點就在記分。
- (11) 十一？ 新式測驗對外國學生亦能公平。
- (12) 十一？ 新式測驗是與生活情景相適應的。
- (13) 十一？ 新式測驗的融洽性很低。
- (14) 十一？ 新式測驗可以測驗推理能力。
- (15) 十一？ 新式測驗中檢查作弊較舊式確實可靠。
- (16) 十一？ 新式測驗易防作弊。

(17) 十一？新式測驗利用印刷的便利，日見普通。

第六章 正誤測驗

正誤測驗——正誤測驗(true-false test)包含許多語句，其中有正確的，也有錯誤的。正確的語句與學生已經學習的事實或普通原理相合。錯誤的語句則與公認的真理不符。此種測驗的目的在使學生認識語句的正誤，並應用其知識以決定何者為正，何者為誤。從辨別正誤的能力上，可以正確地測量出學生的學科成績。

正誤測驗在新法考試中，應用最為普遍。有些教師，應用得最多而知道得最詳的，恐怕只有這種測驗。他們聽見人家談到客觀測驗，便立即想到正誤測驗或正負測驗——因為學生作答時有時用正負號，故名。優良測驗的種類很多，他們僅知此一種，實在不幸得很。有些專家以為正誤測驗在通行的新式測驗中，並不是最好的一種。無論如何，僅用了正誤測驗，而擯棄其他各種測驗，是不應當的。如以第三章中優良考試的標準來評品，則正誤測驗在客觀測驗中，地位是很高的。但我們贊同的為廣博的應用，而非獨占的應用。

正誤測驗的優點——正誤測驗的應用所以普遍，以及在客觀測驗中所以最為著名，並非偶然的事情。正誤測驗在新式測驗中，最為簡單，也最易適應。編造既不麻煩，記分亦很容易。舉凡教師批閱論文式試卷的種種辛勞，正誤測驗可減除大半。牠的著名，一半也為了這一點。牠於各項教材以及第二章中所討論的各種測驗功用，都有實際的利益。牠能激起適當的心理作用與態度。倘或沒有印刷的便利，可藉口述法以施行之。正誤測驗

的應用便利，乃一般教師普通的要求。在學生方面，正誤測驗的形式也最孚衆望。這不但因為他們對於牠最爲熟悉；也是因為我們覺得易於明瞭，做起來簡單，確定，公平，節省時間精力，有趣而且完善。就每個題目而言，正誤測驗的可靠性，雖較其他幾種的測驗爲低，但就全體言，可靠性的差異是很小的。牠在很短的時間內，可以包括很多的教材。一個極完善而極可靠的正誤測驗，可以在十五分鐘至半點鐘內施行。

分數的分配依照常態分配曲線。學生的真正差異，正誤測驗比論文式考試表示得格外清楚。克勞福特（Crawford）與蘭奈爾多（Raynaldo）（註1）的研究，對於這點已說得十分明白。他們根據學生的等級地位，在一個團體中選出知識豐富與知識缺乏的二種學生。假定A、B兩個學生是知識豐富的，C、D、E、F四個學生是知識缺乏的。他們用正誤測驗與論文式測驗來測驗這些學生，以決定何者易於分別其真正的差異。實驗的結果，學生真正差異的表示，正誤測驗較論文式測驗要確實可靠。

刺激適當的心理作用與態度——正誤測驗最大的優點，便在刺激適當的心理作用與態度。要準備這種測驗，須多知道種種事實。廣博的知識是精確的見解之基礎。正誤測驗如編製完善，能測量廣博的知識，也能測量特殊的事實。要確切的認識事實，廣博的知識也是很重要的。這方面的需要，便使讀書時不得不精求了解與明辨，於是便漸漸養成優良的學習方法。著者因為供給大學學生正誤語句表，——與本書每章末後的討論溫習題相似——以便應用於指定作業的準備，獲得很多的優良結果。因為這種語句表能使學生應用已得的知

識，並供給引入討論的問題。正誤測驗不但要求各種事實的完善知識，而且使此種知識應用於新的情境。例如，學生已教過了廣義的真理，如人類中各個種族都有合羣的本能，都能從事社會的接觸。一個應用此種真理的優良正誤語句，可以如：

十 ① 山皮拉斯印第安人 (San Blas Indians) 避免社會的接觸。

一個學生雖然以前並不知道山皮拉斯印第安人，其情境對於他雖然完全是新的，但須能應用普通的原理，對於這個語句有正確的反應。此種語句須有心理的作用，與僅需普通真理的語句不同；例如：

② 一 各個種族都有社會的本能。

十 ① 有些種族缺乏社會的本能。

顯然的，第一個正誤的語句，比其他二個好。下面的語句，除知道事實外，還需要推理：

十 一 清教徒與遊方人 (Pilgrims) 是同一種人。

十 一 幸我們抵制日貨，因為牠不耐用。

正誤測驗能就事實的實際關係作廣博的討論。一種語句對於學生雖然完全是新的，但可應用以前學過的各種事實，真理，原理等，以決定其正誤。於是不得不注重完美的知識，以便適應隨時發生的情境。有用知識的獲得與此種知識的應用，都有賴於深刻的思考。此種思考，能使學生從事判斷，反省，以及形成意見。記憶與思想

關係的密切，前已明白指出了。

正誤測驗不但能刺激心理作用，使其活動；且能引起懷疑的態度，與訓練生活上正誤的敘述。懷疑的態度是科學的思想與工作的一部，文化的成功，端賴於此。平常每有一種傾向，即盲目的接受各種問題，無所懷疑，而於印刷品爲尤甚。發問，善疑，與研究的習慣，應該設法培養。正誤測驗便有助於此種懷疑態度的養成。牠並不是矯揉造作的方法。我人一生中，無時不須要去分辨各種問題的是非正誤。對於各種觀念與建議詳述是非的時候，其中便會有思想的作用；而此種觀念與建議的本身，能引起解決緊急的問題。對於是非真偽的辨別，實爲必須學習的一件事。

猜度的因子——對於正誤測驗有一種很重要的批評，就是猜度正確反應的機會太大。猜度的因子可以產生二種不良的結果。第一，能養成全恃猜度而不求可靠證據的習慣。最美滿的思考，雖無異正確的猜度，但是以已知的事實爲根據的。猜度而不了解或觀察相關的事實，在教育上，在科學上，都有未合。正誤測驗應注意記熟學科中適宜的事實，以便猜度時有適當的基礎。其他事情亦莫不如此，任何學問的權威者，便是在那種學問內最會猜度的人，因爲他能夠利用最多的事實，使他的結論完滿無疵。

(註1) C. C. Crawford and D. A. Raynald. "Some Experimental Comparisons of True-False Tests and Traditional

Examinations," *Journal of School Review*, Vol. 33, p. 698-706. (1925年11月)

在正誤測驗中，學生每每僅恃猜度，並相信命運，因為其中有一半的機遇可以猜正。有些專家以為學生在不知道的時候，應教其猜度，並將猜的題目加以標識。在記分的方法上，對於純粹的猜度，可加以修正，所以在實質上並不影響測驗結果的正確性或可靠性。但是猜度方法能養成一種不好的習慣，即驟下結論，而不能用可靠的事實來證明。此種方法在應付非常事變，誠屬必要，但不應當在學校裏教學。其他的專家也有反對猜度方法的。他們警告學生，倘使不能確知正確，不要猜度。但不用猜度，也有一個困難，就是過於謹慎的學生，便因此陷於不利。他們對於不能絕對有把握的問題，往往完全放棄。根據一知半解的知識而猜度時，正確的多，而錯誤的少。所以學生得了這種機遇，便可獲得不應得的益處。羅偉（Lowry）與克勞福特（註一）發現，當學生本來不很有把握的答案，再經思考而改掉的時候，其改正的機遇大約是二與一之比。著者以為如將猜度的因子向學生解釋清楚，並告以猜錯了要扣分的辦法，則猜度這件事，便常任學生斟酌情形，自行決定。

第二，猜度中機遇的因素，對於學生或許會發生不公平。普通都假定所有的錯誤都是猜度。除了有些事實是學習錯的，或語句的用字易於引起錯誤外，這個假定大概是正確的。不過在任何情形之下，這種事實又是很少遇到的。又有一種假定，以為學生所猜對的與猜錯的，個數相同，所以做正的分數是所做的題數減去錯誤數的二倍。在理論上，倘使猜度的次數很多，則猜對與猜錯的機會相等。猜度的次數如果不多，那就不同了。這正與投擲銅元的情形相似。倘使一枚銅元，投擲的次數很多，則正面與反面發現的機遇大約相等。但如僅投擲十次，

則正反面發現的機遇，便大有變動了。

證明的方法——如以銅元的正面代表準確的猜度，反面代表錯誤的猜度，五個學生每人各投十次，這十次假定其完全是猜度的，則結果如下：

學生A	學生B	學生C	學生D	學生E
誤	正	正	正	誤
正	正	誤	誤	正
正	正	誤	正	正
誤	誤	正	誤	正
正	誤	正	正	正
誤	誤	誤	誤	正
誤	誤	誤	誤	誤
正	正	誤	正	正

(註1) Lowc, M. L. and Crawford, C. C. "First Impression Versus Second Thought in True-False Tests,"

Journal of Educational Psychology, Vol. 20, p. p. 192—95 (1929 #32)

正	正	誤	誤	誤
誤	正	誤	誤	正
5 誤	4 誤	7 誤	6 誤	3 誤
$2 \times \text{誤} = 10$	$2 \times \text{誤} = 8$	$2 \times \text{誤} = 14$	$2 \times \text{誤} = 12$	$2 \times \text{誤} = 6$

在這裏所須注意的，就是將錯誤二倍了以後，五個學生中只有學生A的分數把猜度正確地校正了。學生C與學生D減得太多，學生B與學生E減得太少。就五個學生的五十個猜度全數而言，適有二十五個猜對，二十五個猜錯。這與「猜度的數目愈大，正誤的數目愈易相等」那句話，適相符合。一個未受教的人，純用猜度的方法去做一百個問題，也會偶然地大約做正一半，做錯一半。照此情形，將錯誤二倍了，然後從做的題數中減去其積所得的分數大約是零，正等於應得的數目。純用猜度最不公平的是優秀的學生。因為猜度的問題為數不多，他們的猜對與猜錯之數，遂不易相等。一個長的測驗，使猜度的題數增加，遂使猜度的錯誤減少。

再者，一個長的測驗在全部分數上，除減少猜度的影響。測驗的長度，對於減力最後分數上猜度的影響，可從上述例證中的二種理論假定上看出來。假定全部測驗包括二十五個問題，每個學生應有十五分的真正分數，因為那個分數就是每個學生所確知的題數。他們再猜度其他十個。用了猜度的方法，學生A得真分數15。B得17，C得11，D得13，E得19。只要知道了每人的真分數15，這些分數的不可靠，便很易看出。現在再假

定全體測驗包含一百個問題。學生A得 ∞ 分，B ∞ 分，C ∞ 分，D ∞ 分，E ∞ 分。不公平的情形雖仍未完全免除，但較含有二十五題的測驗，要少得多了。

兩倍錯誤的方法，並不是在任何情形之下，都能得到真分數。但是牠比僅僅計算做正的數目，較為接近真分數。在任何情形之下，一班學生的相對地位，總是相同的。克利斯頓孫 (Christensen) (註1) 建議，雙重測驗可以作為考試中校正猜度的方法。在正誤測驗以後，稍隔幾天，便接以選答測驗 (Multiple-Choice test)。最後的分數是兩種測驗中做對題數的總和。此種方法雖似合理，但用同樣的材料編造與應用兩種測驗，時間上便為人所反對。正誤測驗如有充分的長度，適當的編製，其分數又能校正猜度，則猜度的因子，並不多麼嚴重。

錯誤語句的影響——關於正誤測驗，還有一種批評，就是錯誤的語句，能引起錯誤的印象。從這個理由攻擊正誤測驗的人，一則以為牠能形成不正確的觀念，適足以滋混亂；一則以為從教育的觀點上說，牠是不適宜的。他們以為只有正確的與精密的東西，方可放在學生的面前。普通言之，這是一個健全的教學原理。起初的印象，應當力求精確。只有正確的東西，才應去練習，才應去學習。但真理在教完以後，便應當用應用方法之一，便是辨別何者為正，何者為誤。能夠這樣辨別，也是一種教學目的。這點在本章開始時，早已討論及之。因為有錯誤印

(註1) Christensen, A. M. "A Suggestion as to Correcting Guessing in Examinations," 載於 *Journal of Educational*

象的可能，而批評正誤測驗，實係由於不能分別教學與測驗所致。教學與測驗是二件不同的事情，各有不同的技術。就是在教學中，在說『不是這件東西』的時候，每有助於表明『這是什麼東西』。測驗的施行，在假定學生形成了正確印象以後，故可測量學習已達何種程度。

蘭滿斯 (Remmers) 與蘭滿斯 (Remmers) (註一) 爲欲決定正誤問題中錯誤語句的影響，曾舉行一個實驗。他們對於正誤測驗中『錯誤的轉移』 (Negative Carry-over)，或錯誤的聯絡，並沒有找到證據。白拉特 (Ballard) (註二) 於學習與復試的研究中發現，學生在錯誤的語句上所獲的進步，比正確的語句還大。所以，以引起錯誤印象爲理由而反對正誤測驗，並沒有健全的根據。

編造正誤測驗的技術——正誤測驗有時產生不圓滿的結果，並引起反對的評論，係由於沒有適當的編製。語句的數目也許太少。有幾句語句或許不僅一種解釋。有幾句或許包含二種或以上不同的觀念。有幾句或能引入或暗示正確的答案。正確的語句數或遠多於錯誤的語句，或反而行之。正誤語句的排列，多少有一定的系統。語句的次序，或許會排得使前一語句暗示後一語句的答案。語句的內容，或許太瑣屑，太無關輕重。凡此種種因子，均能使正誤測驗失其效用。僅僅選擇幾句正確的與錯誤的語句，把牠們組成測驗的形式，決不能產生優良的正誤測驗。有了技術的知識與編製的經驗，才能編造這類真正第一流的測驗。爲敘述便利起見，編製正誤測驗的技術，可分三大步驟討論：

1. 編製正確的語句。
2. 正確的語句改爲錯誤的語句。
3. 規定學生做記號的方法。

編製正確的語句——編造正誤測驗的第一步，就是編製許多正確的語句，其中包含所要測驗的教材。一個優良的測驗，須有七十五至一百五十個題目。常用的短時間測驗，其結果是累計的，題目可以較少。因爲這些題目含有學程中所重視的主要要素，也含有許多不甚重要的細點，所以許多的測驗材料，應從教學過程中去搜集。就是將以後應該考試的各點，日常就加以注意，這是一個很好的計劃。這樣可以使教學與測驗互相聯絡，以後並可節省不少時間。

由所選教材中編製正確的語句，有數點應加注意。各種語句，應注意其是否真正重要。只努力於搜集多量題目，每使測驗爲不重要的事實與不合理的細節所混亂。外加的題目，倘無內在的價值，只值得化費時間或精力。凡僅僅重述已經列入的材料語句，都應刪除。各種語句應愈簡明愈好。每一語句最好只佔一行。牠們應當

(註1) Remmers, H.H. and Remmers, E. M. "The Negative-Suggestion Effect of True-False Questions." 譯文

Journal of Educational Psychology, Vol.17, p.p. 52-56. (1926年1月)

(註2) Ballard, P. B. "The New Examiner," p.p. 96-98. (1924)

確定，明晰，含糊的，模稜的，複雜的語句，應該避掉。每一語句應限於一句句子，表達一種意思。不要說華盛頓林肯是美國的大總統，而應分別敘述，一句說華盛頓，另一句說林肯。否定的語句應竭力避免。兩重的否定永不可用。否定的語句易於導入錯誤，兩重的否定易於使人糾纏。廣義的總括也應避除。要使各個語句真實，常須加以形容。所以各種語句的正誤，常由有無形容詞以暗示之。各種形容詞如「一切」「每個」「時常」「永久」等等，不應常用，因為牠們容易決定學生反應的性質。凡足以激起思想的，應求其多。有些語句應該說明已經教過的基本原理。第 50 頁中舉有這樣的一個例子。要編造一個絕對正確的語句表，使與以上所述注意點相合，須加思考和注意。

圓滿的正確語句表編成以後，就須排列適宜於測驗的次序。牠們的排列可以依照牠們的難度，重要性，論理的次序，或隨機的順序。由教師判斷各個語句的相對難度，是不可靠的，這早已用統計的方法精密的決定過了。難度的次序，在非標準的教室測驗裏是不合實際的，也是不必要的。以教師的意見判斷測驗题目的相對重要性，也是不可靠的。難易的次序與重要性的次序，可以置之不顧。最普通的排列法是依照論理的次序；易言之，就是依照各個題目自學程中選出時的次序。這是一種自然的簡易的排列法。不過，也含有引進 (Leads) 的危險性。編造各種語句，稍不注意，便弄成彼此倚賴。如果把牠們連續的排列起來，每能暗示種種線索，以助疑難题目的解答。若免此弊，則論理次序的排列，其優點誠有足多者。依照隨機的順序，也是可能的。這樣排列，一個語句

不致暗示其他連續語句的解答。依照隨機的次序排列各個語句，很是容易，先把牠們標以號數，次將相當的號數記在紙條上面，統統都攪亂，然後依照抽出號數的次序排列起來。在任何情形之下，把少數極容易的語句——無論那個被試者都能正確回答的——排列在測驗的開端，籍以引起學生的自信，並使發生適宜的融洽性。這個方法總是很好的。

正確的語句改爲錯誤的語句——編造正確測驗的第二步，是將大約半數的題目改爲錯誤的語句。對於要使其變成錯誤的語句的選擇，應該愈隨機愈好。學生遇到一個語句，其形式是正確的，抑是錯誤的，應爲機遇的問題，而非教師主觀判斷的結果。所以如果僅僅選擇某種語句，以爲在某些理由上是應該使牠變成錯誤，並依照此種方法，繼續爲之，直至所選的題目已夠，且都變成錯誤的語句爲止。這樣的選擇方法，勢難滿意。但是一個語句，如不使其錯誤過於明顯，便不能改成錯誤的形式，或則因爲班級的討論或其他特殊的理由，曾用錯誤的形式作過特殊的敘述，則此種機遇的方法只好認作例外了。機遇的選擇應當作爲定例，因爲牠能供給能力的隨機的樣子 (Random sampling)，以決定語句的正誤，能使主觀的系統的選擇法失其可能，俾不損害測驗；並能便於選擇語句而改成錯誤，便於依照隨機的次序而排列，便於使正誤語句的數目大略相等，以及便於保持語句使適合於測驗的次序。我們看了下列的討論，對此便很明白了。

正確的語句表依照需要排好之後，那些預備改成錯誤的語句，便很易隨機決定，只要每個語句各投一次

骰子好了。骰子上的奇數，可代表錯誤的語句，偶數可代表正確的語句。投了骰子，如果轉出偶數，這個語句仍保持正確的形式；如果轉出奇數，這個語句便改成錯誤的形式。此外也可投擲銅元，正面代表正確的，反面代表錯誤的，也無不可。又如在豆類或紙條上，註明了字，也可用以代表正誤。這些豆類或紙條，一半註明正確的符號，一半註明錯誤的符號，可從一個匣子或袋子裏面一次一次的抽出來，以決定語句的形式。其他相似的機遇方法，也可應用。此種方法不但使語句由機遇的選擇改為錯誤，以及依照機遇的次序排列，並能使測驗中正誤語句的數目，自動的相等。

要將正確的語句改為錯誤，有幾點應加注意。最後決定的語句，應該使牠確實的錯誤，不應僅是一部分的錯誤。如在編製正確的語句時，能使牠們確實的正確，對此很有助力。否定的語句，應該竭力避免。有的時候，否定的應用，似乎是使一個語句變成錯誤的唯一方法。只有在少數的特殊情形下，一個正確的語句，可籍改為否定的以轉成錯誤。倘使錯誤的語句中用了幾句否定的形式，則在正確的語句中，也應有幾句否定的形式，二者的數目大約相等。將正確的語句改為肯定的錯誤語句，因練習而得便利。各種語句應該自然。牠們的錯誤不可在編製上有所暗示。倘不注意避免，此種情形每易發生。凡公認的正確語句中的一切特點，在最後決定的錯誤的語句中，都應具備。

規定學生做記號的方法——編製正誤測驗的第三步，就是規定學生做記號的方法。關於這層須注意二

點。第一須便於學生，第二須便於記分。一種方法是由學生在預定的空白裏填寫指明的文字或符號。平常所用指明的文字和符號，如其或假，是或非，正或誤，O. R. 或 N, T 或 F, R 或 W, 十或一，以及十或〇。寫真與假，似乎太費時間。倘寫「與」，有些學生的「與」有時又難於分別。其中有些學生故意寫得模糊，尤其是在不知道正確反應的時候。十與〇的符號似較優於十與一。倘使學生交互批改卷子，他們易於把一改成十。學生如要寫入答案，可採用下列的形式：

說明：下面的語句中，有正確的，也有錯誤的。你認為正確的語句，在前面做一加號（+）；錯誤的，在前面做一零號（〇）。分數是所做題數減去做錯題數的二倍。

.....減法中我們要用被加數。

.....乘法的結果叫做積。

..... $\frac{3}{4}$ 是金元一元的四分之一。

還有一種方法，是將指明的符號寫在各個語句的前面，所以學生要表明答案時，只要在適當的指明符號下劃一條線，或沿着牠的四周畫一個圈。這樣，時間既極經濟，批閱亦復簡單。下面的形式，便是這種方法的例子：

說明：在下面的語句中，有正確的，也有錯誤的。你認為正確的語句，把「」圈起來；錯誤的，把「」圈起來。分數是所做題數減去做錯題數的兩倍。

III 六打雞蛋就是七十二個。

IV 一英寸見方就是一方英寸。

V 除法的結果叫做餘數。

VI 分數的分子與分母乘以同數，其值增大。

規定學生做記號的方法，須注意：書寫減至最低限度，答案清楚明晰，答案的排列應使易與答案紙對照及校正。

正誤測驗的記分——批閱正誤測驗時，應用答案紙。答案紙的編製，只要將一份已做的測驗卷，加上正確的符號，最好把牠的邊摺轉或裁下，使上面的答案與學生的答案相靠近。這樣，在檢查與標準答案不相符的答案時，較為便利。凡不相符的答案，應在旁做一符號，以表示錯誤。沒有做的應加上符號。做錯的倘做V符號，未做的可做Q符號。這樣，做錯與未做的分別相加，便比較容易了。有些人以為倘將所有卷子的第一頁，統統改完了，再改第二頁，比較要便利得多。有些人寧願改完了一個學生的全份卷子，然後再改第二份。這個問題的決定，須視那個方法執着答案紙與測驗卷的次數最少，與那個方法時間最為經濟。所有的卷子都批改完了，便將每份卷子的做錯題數與未做題數，分別相加。測驗中的全體題數減去未做題數，便是所做的題數。將做錯題數兩倍了以後，以所做題數減之，這便是測驗最後的總分了。例如一個測驗共有一百題，一個學生做錯八個，未做四個，

則分數的計算如下：

$$100\text{題} - 4\text{未做題} = 96\text{做題}$$

$$2 \times 8\text{做錯題} = 16$$

$$80 \dots \dots \text{最後分數}$$

精度的校正公式 (Correction formula) 普通是 $R - W$ ，意即做正題數減去做錯題數。這與所做題數減去做錯題數的兩倍相同，因為：

$$\text{分數} = \text{所做題數} - 2W$$

$$\text{所做題數} = R + W$$

$$\text{分數} = R + W - 2W = R - W$$

化分數為等第的方法，當於第十一章中討論之。

最近有一種測驗的方法，可使記分的整個問題簡單得多，並可無須編造及應用答案紙，這便是克拉潑與
· 衡 (Clapp-Young) 的自己記分法 (Self-Marking Device)。用了這種方法，一個學生可在許多小方紙上，擇寫一 \times ，以表示他的答案。把這些小方紙與複寫紙精密排列着，則摺疊紙的裏面二頁，便將符號印出。倘使答案是對的，每個印出的符號落在小方紙以內；倘使錯的，落在外邊。摺疊紙啓開，只要數有 \times 的小方紙，便獲得

分數了。學生做完了測驗以後，在一二分鐘內，便能統計他們自己的分數。此種方法在答案小冊（Answer Booklet）（註一）一類的小書中可以找得。學生在任何一種正誤測驗或選答測驗裏面，都可應用之。其中有一百個題目，每個題目的答案有五個選擇的機會。自己記分答案小冊（Self-Marking Answer Booklet）的應用便利與節省時間，使此種方法成爲教師測驗設備的固定部分。此種小冊的定價，是很便宜的。

施行正誤測驗的方法——施行正誤測驗最好的方法，乃爲印刷充分的卷子，使每個學生都有一份。倘使沒有印刷的便利，用口述法，也可滿意的施行。板書法也可應用，但結果不甚圓滿，所以不能博得人家的贊成。此法教師寫得太多，黑板的地位，通常每嫌太小，不能做這類的工作。就是黑板的地位夠了，教師也願意寫了，但要使黑板上所寫的許多語句，不使有幾個學生比其他學生預先看見，也極困難。所以正誤測驗最好不用板書法。

正誤測驗極宜用口述法。（註二）就是印刷很便利的時候，有些教師有時也願以口述法施行測驗。他們以爲這種方法不必特別去編造測驗的材料，與平常的教學程序又無多大差異，所以施行起來很爲容易。學生平時雖喜歡印刷法，但使他們練習判斷口述的語句，也未嘗無益。在生活情境中，筆寫的語句與口述的語句，須同樣去辨別其正誤。此種方法在學生方面，雖然比較困難些，但仍不失爲優良的方法。應用口述法，教師須注意各個學生反應時所需時間的差異。有些學生的反應較其他學生要遲緩得多。各個語句教師不可讀得太快，每句都須讀二遍。每個學生對於每個語句，都應有充分明瞭的機會。學生須在格式一律的卷子上寫各種反應標準。

答案可放在同樣的紙上，以便與學生的答案相合。考卷倘使由學生交互校正，則卷子不一律也不關重要。在教師口讀各種語句以前，學生在他們的卷子上，應寫出號數，俾與口述的語句相符。教師說出號數，並讀着語句。學生在每題以後，用適當的符號，表明語句的正誤。教師應該知道，用口述法施行正誤測驗，有時每發生一種作弊的情形，即優良的學生，常用預定的記號，把某句應做的正號或誤號暗示他人。口述法雖可偶用，但通常所用的方法，每個學生仍應發一印刷的測驗卷子。

應用正誤測驗的結果——測驗結果如不應用，便失掉測驗的目的。第三章中曾述實用性為優良考試的特徵之一。測驗的施行，僅應為了牠能夠完成某種功能。測驗的功用，在第二章中曾詳細論及。測驗的結果，如不用以改進學習與教學，就是最好的測驗，也徒費時間。正誤測驗的易於適合各種測驗目的，及其如何刺激適宜的心理作用與態度，本章前已論及。

正誤測驗的總分數，能用以決定學生的相對成績，及其學業的地位與等第。正誤測驗，因為不論用口述的形式，或印刷的形式，均甚便利，所以對於指定工作的是否熟習，實為最良的檢查。牠又能用以引起日常工作的

(註一) The Clapp-young Self-Marking Tests-Answer Booklet, Houghton Mifflin Company, Boston, 1929

(註二) H. C. Lehman發表一文名“The Oral Versus the Mimeographed True-False,”載於School and Society, V.

L. 30. P.P. 470—72 (1929年10月5日)。此以為口頭的表现與印刷的表现同樣有效。

動機。要達這個目的，不須施行日常的測驗；所期望的結果可由施行五分鐘的臨時測驗中得之。例如一個教師決意要每隔四天在其級中施行一簡短的正誤測驗，少數重要的語句，可在每天的日常準備中設法編製。手頭可準備四張卡片，三張表示本日沒有測驗，一張表示有一測驗。在上課時間開始的時候，就這些卡片中抽出一張，以決定當日是否要施行測驗。所以學生並不知道何時要測驗，因而須繼續的準備着。將學生臨時測驗的結果繪而成圖，也能引起動機。

從正誤測驗中所得的結果，也有很大的診斷價值。爲着診斷與補救教學的目的，而施行測驗，並不含有神秘與奇異的意義。任何測驗，如教師對於牠的錯誤答案，願加審慎的研究，都有診斷的價值。各種問題如集中於許多可能的弱點上面，並將各個錯誤繪製成表，俾易於觀察與分析，則一個測驗對於診斷的目的，便格外有用了。倘使教師檢查測驗的各種語句，發現其中某某幾個，大部分的學生都是錯的，則便可知關於這些做錯的語句有關係的特殊工作所應特別努力教學的，在教學上沒有重視，或有缺點。少數學生所做錯的語句，也可以表示個別弱點的性質，並能指示何處需要個別的教學以校正之。要正誤測驗有完滿的應用，對於各個語句的反應，應悉心研究，俾達改進與校正教學的目的。這樣分析問題，可以指示所應注重的的工作，以及各個學生所需的特殊注意。在養成測驗技術的進程中，測驗結果的應用，不能過於注重。

溫習與討論問題

- (1) 十一？ 正誤測驗是應用最不普遍的新式測驗。
- (2) 十一？ 正誤測驗如以優良考試的特徵來評品，地位很低。
- (3) 十一？ 正誤測驗的適應性很大。
- (4) 十一？ 正誤測驗能刺激適宜的心理歷程。
- (5) 十一？ 正誤測驗含自然引進的妙用。
- (6) 十一？ 正誤測驗供給生活情境的優良訓練。
- (7) 十一？ 根據可靠的證據去猜度，乃適宜的特質。
- (8) 十一？ 將錯誤的反應兩倍之是校正猜度的精確方法。
- (9) 十一？ 測驗愈長，猜度的錯誤愈大。
- (10) 十一？ 正誤測驗在教育上是不健全的。
- (11) 十一？ 正誤測驗應有幾個含糊的語句。
- (12) 十一？ 正誤測驗中應避免否定的語句。
- (13) 十一？ 正誤的語句，應依其重要性的大小，排列次序。
- (14) 十一？ 改成錯誤的語句，應隨機選擇。

(15) 十一？學生記分用的方法，是很重要的。

(16) 十一？正誤測驗應依照機械的方法記分。

(17) 十一？以板書法施行正誤測驗是圓滿的。

(18) 十一？學生願意用口述法施行正誤測驗。

(19) 十一？知道如何應用正誤測驗的結果，比知道如何編造良好的正誤測驗為重要。

新法考試

第七章 選答測驗與對偶測驗

選答測驗——選答測驗包含許多語句，每句預擬幾個答案，以備選答。預擬的答案自五個至七個或七個以上不等。普通為五個。其中只有一個是準確的或最好的，其餘都是不準確的。這些不準確的答案，都是易於混淆的，學生須在許多易混的答案中認出一個準確的。這個準確的答案，因雜在易混的答案中，便變成模糊難辨。這樣的編製就需要判斷和辨優的能力了。能於許多錯誤的答案中選擇一個正確的，很可測驗出學生關於所考的材料的成绩。

選答測驗具有很大的用處。牠是最優良的客觀測驗之一。牠雖很著名而且常用，但其為用，終不及正誤測驗之大。這大部份是因為編製較難的緣故。要編許多易混的答案，不及改正確語句為錯誤語句那樣容易。編製上既較正誤測驗為難，所以應用難廣。雖然如此，選答測驗總應正式列為測驗程序之一。其中實含有優良考試的特徵。

選答測驗的優點——選答測驗有許多顯著的優點。牠的編製雖較正誤測驗為難，而施行和記分卻是同樣容易。牠能用標準答案紙記分，容易而捷便。牠能適應各種材料，差不多大多數學科都能適用。牠受各學生的歡迎，幾可與正誤測驗並駕齊驅，有的學生且有願受選答測驗的傾向。他們覺得牠和正誤測驗一樣，也很容易。

明確，公平，有趣而且徹底。至於猜度的機會較少，尤為特色。猜度的因素的減少，與每個語句下預擬答案的增加成比例。如果預給了五個可以選擇的答案，那麼純用猜度而猜中的機會是五分之一；但在正誤測驗中則猜中與猜錯的機會相等。做選答測驗不及做正誤測驗那樣快，牠的速率視測驗的難度和學生的反應時間而異，不過牠總算是一個便捷的測驗方法。

選答測驗能引起具有實用性的心理態度和活動。這種測驗需熟知事實，需應用知識，需有健全的判斷，和事實的精確估量。牠包含着比較和辨別的能力，同點和關係的發現，以及良好教學和觀察的必要。牠所創設的情景並不是空中樓閣，而是自然而切實際的。我們在生活的情景中，常須於許多可能之事中，選擇其一。選答測驗的形式和易混的答案的性質，其編造方法，可使練習適應現代生活中複雜和變化的環境所需要的心理活動。選答測驗非僅測量孤立的事實，也測量獲得學識的應用能力。

選答測驗的批評——對於選答測驗有兩個重要的批評，就是猜中答案的可能性，和錯誤印像的影響。此二者也是正誤測驗的重要缺點，已在前章討論正誤測驗時述及。前章的評論，大體也可用於選答測驗。不過，在選答測驗中猜中的機遇較少，按照每語句下預擬的答案數而遞減。如預擬的答案僅有二個，則猜度的因素與正誤測驗完全相同。如有五個，那麼猜中的機遇便減少至五分之一了。其分數可用下列公式，以校正猜度的因素。

$$\text{校正後的分數} = R - \frac{W}{N-1}$$

在這裏 N 代表預擬的答案數。如果只有二個預擬的答案，那麼校正的公式就變為正誤測驗中校正猜度因素的公式了；那就是以 2 代之 N ，公式就變為 $R - \frac{W}{2}$ ，與正誤測驗一樣了。如果每語句下有三個預擬的答案，其校正式為

$$R - \frac{W}{2}; \text{有五個則為}$$

$$R - \frac{W}{4};$$

餘類推。不過在這裏我們已經看出來，如果答案數有四個或四個以上，則猜度的因子大大減少，足使猜度的校正數變為無足重輕。至於「錯誤印象」的因子是不顯著的，並不足以使這種測驗，遭人反對。

有些人以施行測驗的立場來反對選答測驗。他們覺得要編製一個完善的選答測驗，較他種測驗為難。牠不適於口述的方法。如果沒有印刷的便利，就不能應用牠。牠有許多易混的答案，佔據很多的地位，故欲包含相當的語句數量，就要很多的印刷。不過，這些批評並不足以證實這種反對選答測驗的判斷。選答測驗的優點實勝於牠的缺點。

編製選答測驗的技術——一個優良的選答測驗，須注意適宜的編製。以隨便的態度，東拉西湊，結果決難

圓滿。編製不良的選答測驗，實是數見不鮮。推其不良的原因，或由於缺乏這種測驗的知識和經驗，或由於缺少注意。不論原因如何，牠的缺點常是很明顯的。有的時候，排列不相一致。譬如在同一測驗中，易混的答案，有的寫在語句之前；有的在語句之後；有的嵌在中間。牠的排列，有的成縱列；而有的又成橫列。所出的題目，有的為陳述式；有的為問題式。凡此種種，都是不良的方法。排列上這樣的雜亂無章，使學生神昏目眩。與此相附而生的，有一種缺點，就是有時缺乏適宜的條件，以致學生不能很便捷的做記號，教師不能很便捷的記分。

就是易混的答案，其本身的編製，也常屬不良。準確答案和不準確答案的彼此相異，竟可暗示線索。關於這點，在下述例子中可見一斑。下例採自高年級歷史選答測驗，在每個題目中最好的答案和其他最壞的答案不同之點，就是牠總是最長的一個：

清教徒何以入美洲？

1. 爲了要做意大利人。
2. 爲了在本國沒有政治上、宗教上和軍事上的自由。
3. 爲了要參加政治機關。

★歐洲大戰的導火線是什麼？

1. 奧太子斐迪南在塞爾維亞被刺。

2. 德法是世仇。

3. 法國經濟紊亂。

黑人何以視林肯爲己友？

1. 因爲他維持美利堅合衆國。

2. 因爲他是著黑奴解放論的一個人。

3. 因爲他是擁有奴隸的人。

隋煬帝爲什麼開運河？

1. 便交通。

2. 防匈奴。

3. 到揚州去看瓊花。

文法的結構上，或易混答案中含有不同的概念，都可給學生以辨認準確反應的線索。有些易混的答案，是不自然的，或是不合理的，或是錯誤過顯的，以致學生們可以看出準確答案而無何種特殊困難。有些易混的答案，也許部份準確，含糊不明，是非無定。或是措詞太長含義過多，不合測驗的目的。所以要完善的編製一個選答測驗，首須明瞭編製的技術，且須有充分的練習，以臻精巧。爲討論便利計，選答測驗的編製，可分下列三大步驟：

1. 正確語句的編造和排列。
2. 易混答案的擬定。
3. 學生做記號的規定。

正確語句的編造和排列——編造選答測驗的第一步，就是許多正確語句的編造和排列。此種正確語句，應具有前章中對於正誤測驗所建議的特質，牠們應該是重要的，獨立的，簡單的，清晰的，明確的和正面的。不重要的細節應行除去，學程的精義和要點應予重視。選答測驗有許多良好的材料，可於班級討論中和教師預備教學工作時採集之。測驗材料以及教材的編製應該互相溝通。含糊模稜以及駁雜的材料應該除去。每一句子應該簡單清晰而限於一義，不可有反面的語句和廣漠的概說。有許多題目，足以刺激思想和說明已學之根本真理者，應行列入。語句到了最後編製的時候，可以隨機排列，或依照理論的程序。

易混答案的擬定——編製選答測驗的第二步，就是在每一語句下擬定幾個似是而非的或似優而劣的答案。五個答案便可構成一個良好的測驗。擬定這些易混的測驗，最好要記着，無學識的學生通常依照一定的方法以決定所選的答案，他們用反面的方法，根據某種理由，除去自己認為最不準確的答案；他們在編製上找尋足以決疑的線索；或是選定他們所最熟悉的和最注意的答案。教師於擬定易混的答案時，必須殫智竭慮，使這些無學識的學生所用的方法歸於無用。準確的答案和易混的答案間，任何必需的差異，均應使無學識的學

生趨向於選擇錯誤的答案；這就是說，如果我們不能使正確與錯誤的答案具有同樣的吸引力或熟悉性，寧可使錯誤的答案較正確的答案更覺引人，更爲熟悉。

在擬定選答測驗中易混的答案時，教師應注意幾點。對於學生，易混的答案應與準確的答案同樣熟悉。舉例言之，如果答案所包括的是許多菜名，那麼每個菜名對於學生應該同樣的著名；同樣的普通。像 *Peperya*, *Cherimoya* 等，除非與準確的答案同樣著名，不應列入易混的答案之中。易混的答案與標準的答案應屬同一概念。如果準確的答案是菜名，那麼各種預擬的答案均應爲菜名。總之，錯誤的答案和準確的答案應該討論同一題目和同一材料；應該一樣的引人注意，一樣的簡單；同樣的合理，同樣的自然。編擬易混的不準確的答案時，難免有使答案含義模糊的危險。所以爲教師者，慎勿使不準確的答案，有部份的準確或有準確的可能。測驗編就以後，最好請另一教師來檢查其中有無含義模糊和不合理的答案。如果有些答案有部份的準確，那末是最好的答案測驗而非準確的答案測驗了。除了複選測驗 (*Plural Choice Test*) 外，每一語句，只應有一個準確的答案。錯誤顯然全無困難的答案應行除去。易混的和準確的答案均應簡單，需字很多的語句，應該採用別一式的測驗。是非不定的成分，應行除去。不定前置詞 *and* 的應用應該注意，如果用不得當，可直接表示出準確答案來。舉例言之，如果某一語句的措辭，其準確答案僅需一「*or*」字，而有些或全部易混的答案，均應用「*and*」，那麼，準確的答案已由構造上明確的表示出來了；或者真正易混的答案數至少因此減少了。易混的答案

如僅係措辭上的錯誤，不應列入。優良的易混的答案，常可自得自平日教室討論中學生所屢答的許多錯誤答案。機警的教師，無論何時，都注意着測驗的材料。

每語句下所需的易混答案選定以後，應該隨機的把準確的答案排列其間，排列之法可以擲骰子決定之。如果骰子上擲出「一」來，那麼把準確的答案列在各易混答案之首；如果擲出「二」來，列在第二；每句都以此決定。除擲骰子方法外，也可在豆類或紙條上依次寫着與每語句所預擬的答案的相當數字，看每次抽取的數目，以定每句準確的答案在易混答案中的地位。這步做完了，易混答案的編製可告完成。

學生做記號的規定——編製選答測驗的第三步，就是規定學生做的記號。通常我們叫學生在準確答案下劃一條線。這個方法便於學生而不便於記分。如果用這個方法來做記號，那麼批閱時不便應用標準答案紙。答案的分佈太無規則，只有透明的或雕孔的答案紙 (transparent or Cast-out key) 才能適用。預備與應用均屬不便。比較妥善的方法還是在預擬的答案前，冠以數字，然後叫學生在每語句之前寫下準確答案前的數字，這個方法既便於學生，記分時又易於應用普通標準答案紙。不過用這個方法時，學生每發生錯誤，所寫的數字，或許與所選定的答案不符。爲了要免除這一點，我們可以叫學生一方面在答案下劃線，一方面再以準確的數字表示。這樣一來，我們對於學生的意思有懷疑時，便可將數字和劃着線的答案兩相核對。下列是一個虛擬的選答測驗：

說明：下面是幾句不完全的語句，每句下均附有幾個預擬的答案，以備選擇。選定一個你認為最好的答案，把牠前面的數字寫在牠所屬的語句前面空白上。分數就是做對的題數。

……減法所得的結果叫做：①被減數 ②積 ③和 ④減數 ⑤餘數。

……K等於①10. ②500. ③100. ④1000. ⑤1500.

選答測驗的記分——如果自閱答案紙尚未應用（參看30頁），選答測驗應該用標準答案紙來批閱。標準答案紙之製造就是把一份測驗的卷子，加上正確的記號，答案紙的邊應當裁去，或反摺在後面；那麼上面的答案就可與學生所做的記號相湊近。學生試卷和標準答案紙間的視距，也可因此減少。這樣，記分既然容易，而學生所做的記號與標準答案互相比較時，又可減少錯誤的機會。與標準答案紙不同的答案，在上面作一個錯誤的記號。試卷批改完畢，把錯誤的記號數總加起來。如果測驗所包含的語句，僅有二個選擇機會，那麼最後的分數，是做的題數減去錯誤數之二倍。如果測驗所包括的語句，僅有三個選擇機會，那麼最後的分數，就是做的題數減去做錯題數的二分之一。如果每個語句下所預擬的答案超過三個，那麼猜度的校正可不必顧到，最後的分數，就是做正的題數。既然大多數選答測驗的答案數總在三個以上，所以其分數通常總是學生作對的總數。總分數算出之後，再用第十一章中所討論的方法之一來定等級。

施行選答測驗的方法——施行選答測驗最適宜的方法，莫如印就充分的試卷，使每生各得一份。口述法

不能應用，因為用此法學生在決定準確答案時，要記着許多預擬的答案，不勝困難。板書法也不切實際，一則因為教師寫得太多，一則因為有些學生有比其他學生先看到的可能。有些教師，當缺乏印刷的便利時，曾合用口述和板書兩法，極著成效。用此法施行測驗時，每語句下所預擬的答案，可寫在黑板上，並冠以字母，如下式：

1. (a) 秘魯。 (b) 墨西哥。 (c) 智利。 (d) 阿根廷。 (e) 巴西。

於是叫學生在格式一致的試卷上先寫出語句的順序，教師然後說出順序，並讀出語句如下：

1. 布宜諾斯亞利斯 (Buenos Aires) 是 國的京城。

學生在黑板上語句順序「1」字的後面，看着預擬的答案，並在試卷上「1」的後面，寫下他們所認為準確答案前面的字母如：1 (d)；每個語句，都如此類推。每語句應由教師朗讀二次，勿使學生着忙。試卷可應用標準答案紙來批改，或者叫學生互換試卷，由教師說明每句下與標準答案相符的字母，令學生照着批改。在這裏我們應該注意語句的順序是否正確，和學生在每句中是否把預擬的答案聯在其所屬的語句下。應用這種口述板書的混合法時，克拉潑的自閱答案冊之應用很為便利，並可節省不少時間。

選答測驗的變化——選答測驗有好幾種可能的變化。易混的答案自然能多能少。每個題目的構造，也有陳述式或問題式的不同。除了這種形式上的部分變化外，測驗的性質也可稍變。複選測驗就是普通變化之一。複選測驗與一般選答測驗不同之點，在複選測驗有二個或二個以上的準確答案。易混的答案也比較多。每個

預擬的答案，都屬同一概念。所有準確的答案，學生都須選出。下面是一個選答測驗的例子：

- 楊子雲著 ①法言 ②史記 ③漢書 ④三國志 ⑤論衡 ⑥太玄經 ⑦方言 ⑧宋儒學案
⑨山海經

雜選測驗 (Compound choice test) 是另一種變化。這種測驗除了預擬的答案有二個或二個以上不同概念外，其他都和複選測驗相同。學生須於每個概念下選出準確的答案。下面是一個雜選測驗的例。

富士雅馬 (Fujiyama) 是 (1) 河 (2) 湖 (3) 城 (4) 山 (5) 運河；

在 (a) 俄國 (b) 中國 (c) 日本 (d) 非洲 (e) 歐洲。

雜選測驗的應用，不及複選測驗那樣普通。

徐爾史氏 (Giles) (註一) 所建議的一種選答測驗的變體，就是劣答測驗 (Worst-answer test) 這種測驗包含許多語句，每句都有幾個預擬的答案，其中錯誤的只有一個，或簡直沒有。學生如認為毫無錯誤，便在語句前畫一個圓圈；否則就將錯誤答案前的數字寫在各該語句之前，以表示何者為錯。氏曾舉一個劣答測驗的例子如下：

郎法陸 (Longfellow) (1) 生於1807年，(2) 進的是包多音 (Bowdoin) 大學，(3) 著Thanatopsis，
(4) 著Evangeline。

劣答測驗是脫胎於普通選答測驗的一種有趣的變體，偶一用之，未嘗無益。

對偶測驗——就某一點言，對偶測驗是選答測驗的一種變化。這種測驗包含許多分組的語句，學生須根據自己的識別而將其湊合。這就是要學生就其所知，把兩件有關係的事實聯絡起來；這些事實構成的材料，有人物和事業，事業和日期，事業和地方，名詞和定義，國家和都會，城市和特性，著作家和著作，引句和原著，引句和著者，人物和書籍，單字和詞類，原因和結果，字彙和外國文中之同義字，算學問題和答數，同義字以及反義字等等。對偶測驗在最好的客觀形式中，佔很高地位。故應常常應用。

對偶測驗具有幾種性質，足使牠的價值和用處增大。編製和記分便捷而簡單。學生喜歡牠的簡單，又喜歡牠作記號的容易。牠用不着書寫，程序上又富有興趣。做這種測驗的程序，同解決迷津時所需要的程序一樣。牠供給一種便捷的方法，以測驗許多準確的知識。此外，對偶測驗能適用於多種材料，而尤宜於社會科學。將分離語句，作適當湊合的能力，包含有關係材料的廣識，辨別判斷的應用，以及關係的了解等。對偶測驗所具的各種特徵，已使其躋於優良考試之列。

編造對偶測驗的技術——因對偶測驗取材容易，常有編製不當的危險。關於這點，同論文式的考試一樣。我們已經說過，匆促的出問題，實為許多論文式考試不良的原因。編造一個良好的對偶測驗，務須小心翼翼。此中伏有極度的困難不容率爾操觚的。例如猜度的因素須詳加考慮。又如學生在一長列的題目中，往往上下找

尋適當的對句，以致空費時間。凡此種種因素，編造測驗時都應注意。對偶測驗的編造與正誤測驗和選答測驗一樣，可分做三大步驟如下：

1. 正確語句的編製。
2. 分語句為二部。
3. 學生做記號的規定。

正確語句的編製——編造對偶測驗的第一步，就是編製許多同式的正確語句。十句至二十句，已可構成優良的對偶測驗。如果語句太少，譬如在十句以下，那麼因猜度而湊對的可能性太大，測驗難於可靠。如果語句太多，譬如在二十句以上，那麼題目太長，不適於眼距，逐題找尋適當語句，難免空耗時間。如果包括的語句很多，應該分為不同的對偶組。舉例述之，如有三十句語句要用於對偶式，便應把牠們分為兩個十五題的測驗。為說明便利起見，可用下面的準確語句為例：

1. 哥倫布 (Columbus) 於 1492 年發現美洲。
2. 瓦斯科·達伽馬 (Vasco da Gama) 於 1497 年發現東至印度的路徑。
3. 來夫·厄立克孫 (Leif Ericson) 於 1000 年抵北美大陸。
4. 巴托羅繆·地亞士 (Bartholomew Diaz) 於 1487 年至好望角。

5. 加伯拉爾 (Cabral) 於1500年至巴西海岸。
 6. 拔爾波亞 (Balboa) 於1513年橫過巴拿馬地峽。
 7. 麥哲倫 (Magellan) 於1522年作第一次環球的航行。
 8. 逢退·對雷溫 (Ponce de Leon) 於1512年遠征佛羅利達 (Florida)
 9. 科德司 (Cortez) 於1519年在墨西哥作第一次的永久居留。
 10. 比撒羅 (Pizarro) 於1532年征服秘魯。
 11. 對索托 (De Soto) 於1514年入密士失必河。
 12. 科洛那多 (Coronado) 於1572年建立聖大非鎮。(town of Santa Fe)
 13. 卡退 (Cartier) 於1534年發現聖魯倫斯河 (St. Lawrence)
 14. 德類克 (Drake) 是1580年環航世界的第一個英國人。
 15. 刺里 (Raleigh) 於1584年移民於勿爾吉尼河 (Virginia)
- 分語句爲兩部——編造對偶測驗的第二個步驟，就是分語句爲二部。這僅是主詞和謂詞的劃分，如下面的例子。

哥倫布

於1492年發現美洲。

瓦斯科·達伽馬 於1497年發現東至印度的路徑。

來夫·厄立克孫 於1000年抵北美大陸。

巴托羅繆·地亞士 於1487年至好望角。

加伯拉爾 於1500年至巴西海岸。

拔爾波亞 於1513年橫過巴拿馬地峽。

麥哲倫 於1519年作第一次環球的航行。

蓬退·對雷溫 於1519年遠征佛羅利達。

科德司 於1519年在墨西哥作第一次永久居留。

比撒羅 於1539年征服秘魯。

對索托 於1541年入密士失必河。

科洛那多 於1542年建立聖大非鎮。

卡退 於1581年發現聖魯倫斯河。

德類克 是於1580年環航世界的第一個英國人。

刺里 於1584年移民於勿爾吉尼河。

一個對偶測驗的排列，每可使一個以上的主詞或謂詞與其他一個主詞或謂詞作準確的對偶。這點應該避免。所以對於主詞或謂詞應詳加校核，務使其中的準確對偶不致在一個以上。含了好幾個準確的對偶，那麼測驗愈趨困難，辨識愈需精細。可有幾個對偶的題目，如不將其對偶正確，則某些問題的對偶將極感困難。這一點足使測驗難施行於小學生，而對於中學生，亦只應稍稍採用。

對偶測驗並不常含整句，為求測驗的簡明起見，往往把題目縮短。在這種情形下，題目中除了對偶的要素外，其餘什麼都可略去，簡直可以縮成兩組單字。如果要對偶的要素十分明瞭，對偶測驗本無需整句。上述語句中，有的可以縮短如下。

哥倫布

美洲

瓦斯科·達伽馬

印度

來夫·厄立克孫

北美洲

地亞士

好望角

加伯拉爾

巴西

拔爾波亞

巴拿馬海峽

達退·對雷溫

佛羅利達

科德司

墨西哥

比撒羅

秘魯

刺里

勿爾吉尼河

爲欲儘量免去猜度的因子起見，有時可以增加謂詞，使多於主詞，那麼對偶時將有剩餘的謂詞。這個方法在僅有少量題目需要對偶時，尤覺需要。例如在上列謂詞的表中，可加下面的謂詞，而不加主詞：

第一個生在美土的英國小孩。

從1499年至1508年沿着南美大部分海岸航行。

有的時候也可以增加主詞而不增加謂字。

把語句分爲主詞和謂詞二部之後，謂詞應以機遇的次數排列。排列的方法，可將謂詞或其相當的數字寫在不同紙條之上；次將紙條攪亂，然後依其拈出之次序排列之。不然，按年代的遠近，筆劃的多少，或字母的次序來排列，也不失爲機遇的排列。用筆劃的多少或字母的次序常與隨機的次序同樣的便於學生。我們如用隨機的次序來排列主詞而不排列謂詞也無不可。

學生做記號的規定——編製一個對偶測驗的第三步，就是規定學生做記號的方法。最簡單的方法便是每個主詞之前冠以數字，然後叫學生把適當謂詞的數字寫在其所屬的主詞之前，下述的形式，便是對偶測驗

的例子：

說明：把第二組題目前的數字寫在第一組相當題目之前，以表示第二組各題屬於第一組的那一題。分數就是做對的題數。

第一組

第二組

- | | |
|----------|--|
| 哥命布 | (1) 於1584年移民於 <u>勿爾吉尼河</u> 。 |
| 瓦斯科·達伽馬 | (2) 於1531年發現 <u>聖魯倫斯河</u> 。 |
| 來夫·厄立克孫 | (3) 於1587年至 <u>好望角</u> 。 |
| 巴拉羅繆·地亞士 | (4) 於1512年遠征 <u>佛羅利達</u> 。 |
| 加伯拉爾 | (5) 於1510年建立 <u>聖大非鎮</u> 。 |
| 拔爾波亞 | (6) 於1000年抵 <u>北美大陸</u> 。 |
| 麥哲倫 | (7) 於1514年入 <u>密士失必河</u> 。 |
| 蓬退·對雷溫 | (8) 是於1580年環航世界的 <u>第一個英國人</u> 。 |
| 科德司 | (9) 於1497年發現 <u>東至印度的路徑</u> 。 |
| 比撒羅 | (10) 於1519年在 <u>墨西哥</u> 作 <u>第一次永久居留</u> 。 |

對索托

(11) 於1522年作第一次環球的航行。

科洛那多

(12) 於1492年發現美洲。

卡退

(13) 於1500年至巴西海岸。

德類克

(14) 於1532年征服祕魯。

刺里

(15) 於1513年橫過巴拿馬地峽。

對偶測驗的施行和記分——施行對偶測驗最適宜的方法，就是印就充分的試卷，俾每生各得一份。欲以口述法施行對偶測驗，是不可能的。這類測驗很短，很可用板書法來施行。如用板書法，第一組的題目應冠以數字，第二組則冠以字母。學生可在紙上寫下數字以代表主詞，而於數字之後加以代表謂詞的字母，此種謂詞應與各該主詞相聯絡。對偶測驗批閱時，可用標準答案紙。分數就是正確對偶的總和。

對偶測驗的變化——複對測驗(Compound Matching test)為對偶測驗的一種變化，與一般對偶測驗不同之處，就是對偶的組數在二個以上。上述語句，可改成複答測驗如下：

……哥侖布

(1) 勿爾吉尼河。

(a) 1000。

……瓦斯科·達伽馬

(2) 好望角。

(p) 1487。

……厄立克孫

(3) 佛羅利達。

(o) 1492。

……地亞士	(4) 聖大非鎮。	(d) 1497.
……科洛那多	(5) 北美大陸	(e) 1500.
……加伯拉爾	(6) 密士失必河	(f) 1512.
……拔爾波亞	(7) 印度	(g) 1513.
……對索托	(8) 墨西哥	(h) 1519.
……達退·對雷溫	(9) 美洲	(i) 1532.
……科德司	(10) 巴西	(j) 1541.
……比撒羅	(11) 秘魯	(k) 1572.
……刺里	(12) 巴拿馬地峽	(l) 1584.

選答測驗和對偶測驗的結果之應用——除非能從結果中獲得有價值的事物，就算施行第一等的測驗也屬枉然，這是我們所應時常服膺的。為教師者應常顧到測驗之最大目的，及其結果之應用。選答測驗和對偶測驗並不及正誤測驗那樣完全適用於第二章中所討論的各種測驗的功用。這是因為牠們並不適用於一切教材，而同時編製上和施行上也比較困難的緣故。所以，牠們對於檢查日常工作的純熟與否，以及激勵日常工作等等，為用頗有限止。從牠們所得的總分數，可以善用之以決定學生的相對成績及其學業地位和等級等。

但這種測驗，往往含題太少，故專用此種測驗，不能達到目的。但是，牠們能夠刺激和引起適宜的心理態度，以及含有實際性的心理歷程。兩者均與學生成績有關，應加考慮。他們的優點，本章前已論及。

從選答測驗和對偶測驗上所得的結果，具有診斷的價值，教師應該認識和注意。以前已經說過，一個為診斷和補救教學而計劃的測驗，其性質並不格外複雜，也無特別的異點。有效的診斷大部分是將各個問題的反應，加以分析的研究，而把測驗的材料集中於某幾種困難上面。一個需要精細的實在知識和訓練判斷對偶的對偶測驗，如將其每個錯誤，加以研究——例如將許多政府的功能與美國政府各部的名稱對偶——也能夠表示那幾點在教學上應更加注意。一個選答測驗如在易混的答案中包括普通的錯誤，也能顯示應加校正之弱點。舉例言之，如果教師所編製的英文選答測驗，包括下面的題目，那麼最普通的語言錯誤，以及某些學生所獨犯的語言錯誤，能由檢查學生對於每題所答的答案中決定之。

……I don't like……of books① Those sort② That kind③ Them kind④ that kind ⑤ those kinds.

根據測驗的結果，個別工作和班級工作和語言錯誤的改進法，都能計劃一種方法，以校正已有的缺點。

這種測驗如用某種方法來編造，可以顯示不能應用已知事實於適切情景的狀況。把下列各題的答案，加以分析，便可知道何處需要更有效的教學。

……南加利福尼亞是一個大產橘區，因為：①地是平坦的②水是純潔的③人民很多的④氣候很適宜的⑤地是肥沃的。

……四季的不同是因為：①洋流②雨量③風向④太陽的地位⑤大氣的密度。

僅僅注意學生做錯的總數，像平常的所為，仍屬無用。一個真正錯誤的研究能顯示補救教學的性質。選答測驗在易混的答案中，包括了普通的錯誤，實為診斷學生各種錯誤的最好方法。對於各個學生應重視什麼，應特別注意什麼，均可由此看出。至於測驗的應用與整個的測驗方法有密切關係之說，不能過於重視，也不能常常重視。

溫習與討論問題

- (1) 十一？ 選答測驗的編製較正誤測驗為易。
- (2) 十一？ 選答測驗在優良考試中，地位很高。
- (3) 十一？ 猜度的因子，在選答測驗中較在正誤測驗中為大。
- (4) 十一？ 選答測驗所假設的是不自然的情景。
- (5) 十一？ 選答測驗極易用口述法施行。
- (6) 十一？ 易混的答案應該錯誤得很顯著。

- (7) 十一？ 在選答測驗中，正確語句的編製，同正誤測驗相同。
- (8) 十一？ 二個易混的答案，儘可構成一個最好的選答測驗。
- (9) 十一？ 易混的答案應與正確的答案具有相同的特質。
- (10) 十一？ 學生在班上所答的錯誤答案，不可編入測驗的易混的答案中。
- (11) 十一？ 選答測驗的記分，能設法使其與正誤測驗同樣容易。
- (12) 十一？ 複選測驗與選答測驗不同。
- (13) 十一？ 對偶測驗乃選答測驗的一種變化。
- (14) 十一？ 有三十句語句儘可構成一個良好的對偶測驗。
- (15) 十一？ 對偶測驗中包含主詞和謂詞兩部。
- (16) 十一？ 猜度的因子，可用加多謂詞的方法減少之。
- (17) 十一？ 謂詞的排列次序，應與主詞的次序相顛倒。
- (18) 十一？ 對偶測驗能以標準答案紙批閱。
- (19) 十一？ 複對測驗需要複句。
- (20) 十一？ 教學的改進乃需要測驗之唯一理由。

翻
法
考
試

三
三
三

章八章 回憶測驗

回憶測驗——前兩章所討論的測驗叫做認識測驗 (recognition tests) 牠們所以這樣命名，是由於牠們所需的反應形式。在正誤測驗中，學生須認識或檢查語句的正誤。在選答測驗與對偶測驗中，學生須於許多預擬的答案中認識或發現其正確的。在認識測驗中，學生不要填充什麼。他們只須認明測驗中所說的是什麼，或從許多預擬的答案中選出一個。回憶測驗所以這樣命名，也是由於牠所需的反應形式。在回憶測驗中，答案是不預擬的，學生必定要回憶學科的智識，而以自己的答案填入。從他心智上的努力，他必定要想到許多可能的答案，而從中加以選擇。傳統的論文式考試，便可列為回憶測驗。回憶測驗的兩種普通客觀形式為：(1) 簡答測驗 (Single-Answer test) (2) 填字測驗 (Completion test)。

簡答測驗——簡答測驗就是問答測驗 (question answer test) 其對於每個問題的答案包含一個單字，至多也不過一短句。這種測驗，答案的表明極為簡單，並不用完全的句子或段落，此外便與論文式的考試相同。這是簡單而常用的回憶測驗，實際上，牠並不是新的或近代的測驗。牠成立已久，教師無不熟知之。牠雖應用已久，但其正式列入新法測驗的原因，則由其近來的改進，而使其益趨客觀，則尤為著稱。

簡答測驗具有許多明顯的有價值的特點。其重要的優點是在答案的需要清楚簡短。答案的簡潔明確，須

賴學習的恰當精審。這樣足以獎勵真知灼見，而貶抑一知半解。學習的澈底與否，可由簡答測驗測量之。牠能培養回憶與精確應用所習教材的能力。簡短的答案，能使在短的測驗時間內，測驗多量的教材。簡答測驗學生所能反應的速率，較認識測驗為遲，但卻比較困難。若與傳統的論文式考試相較，就算快了。學生因需自己填寫答案，故不能不有創造與自信。事實上，在簡答測驗中，並沒有猜正的機會。牠對於各種教材，都能適應。牠適用於口述的考試，也適用於筆寫的考試。牠有論文式考試的各種優點，而減少其許多弱點。

簡答測驗並非沒有缺點。牠不及認識測驗的客觀。如欲編製得使達到優良考試所必要的客觀性，比較不易。在批閱方面，也有相當的困難。有許多答案，可以有各種不同的可能性。學生喜歡簡答測驗，雖較傳統的論文式考試為甚，但仍不如認識測驗。學生所以不甚喜愛，大半由於回憶事實較認識事實為難，並且需要較多的書寫。然而一個編制完善的簡答測驗，仍不失為一種最優良的考試，教師應當設法應用。

編製簡答測驗的技術——編製簡答測驗的技術，雖甚簡單，但也須十分注意，務使客觀性愈高愈好。每個問題的編製，只應容一個正確的答案，而此種答案又須一二字可以包括。編製簡答測驗的第一步，就是寫成許多簡要的語句，其中包括所要測驗的材料。第二步是將此種語句改成只需數字的答案的問題，而寫答案的空白，須使便於批改。這種測驗的實際編製，便是編造一論文式考試的問題表，並寫成簡單語句的答案，於是將此種答案改成僅需數字的答案的問題。下面的形式便是簡答測驗。

說明：下面幾個問題，你們都要回答。在每個問題前面的空白內，填寫一個或幾個字的正確答案。答對的題數，就是所得的分數。

..... 中國第一商場是那裏？

..... 世界最大的漁場是那三區？

..... 我國進口貨最多的是什麼？

..... 煤鐵開採最多的是那一國？

..... 颶風的原因是什麼？

..... 大平原 (Great Plains) 上所有的草倘使一旦都枯死了，什麼工業受最大的打擊？

..... 全世界產銀最多的是那一國？

..... 構成生物的單位是什麼？

..... 最適於植物生育的是什麼土壤？

有的時候，空白上畫着點或線，或寫着正確答案的第一字，或二者兼而有之，以指示正確答案的字數。在這

種情形之下，其形式如下：

哥..... 發現新大陸的是誰？

這種點或線的方法，可使學生對於正確的答案，得一線索。牠着重答案的確切性，而減少似是而非的答案的紛擾。答案的一致，對於批改與記分，很覺得便利。但是此種暗示的方法，也有一種缺點，就是使答案過於明顯。所以這個方法，在實際上，並不十分受人們的贊同。

簡答測驗的施行與記分——施行這種測驗的最好方法，是使每個學生均得一印刷的卷子。對於簡答測驗，教師需多量的書寫，所以施行時難於應用板書法。如把問題寫在黑板上面，有些學生常有比其他預先看到的可能。以口述法施行簡答測驗，結果極易圓滿。學生只要在他的卷子上面，寫明與問題號數相同的數目，然後跟着教師口述的問題，將答案寫在數目之後。每個問題須讀兩遍，在讀第二個問題以前，每個學生都須給以充分的時間去寫答案。印刷即甚便利，但偶而以口述法施行簡答測驗，仍很有益。

在簡答測驗的記分中，優良的而非正確的答案，雖非教師所希冀的確切反應，也應給以分數。記分的是意義而非確切的文字。通常的記分方法，是把答案看作全正或全誤。在回憶測驗中，各種不同而可接受的答案，都有得分的價值。倘使答案是完全正確的，給予二分，倘使一部分正確，給予一分；倘使完全錯誤，則給予零分。或是，最好的答案三分，部分正確而非最好的二分，含糊的一分，全誤的零分。最圓滿的方法或許是做對一題給予一分，將各分相加而成總分。記分時亦可應用答案紙。

簡答測驗的變化——普通的答案測驗包括許多完全的問題。其形式可以變動，使與對偶測驗相符合，不

過對偶測驗減少一組，由學生填入。變形的簡答測驗的組成，與對偶測驗一樣，可以為人物與事實，事實與時間，事實與地方，名稱與定義，國家與首都，城市與特性，作者與著作，引文與著作，人物與書籍，單字與詞類，原因與結果，字彙與外國語中的相當字，算術問題與答數，同義字，異義字，以及其他等等。例如，給學生一個書名表，叫他們把作者的姓名填入，舉例如下：

書名	作者
《史記》
《漢書》
《後漢書》
《資治通鑑》
《紅樓夢》
《聊齋誌異》
《儒林外史》
《水滸》
《吶喊》

同 境 測 驗

平民權初步

此種變相的簡答測驗，可以編成許多種類。牠們用口述法，板書法，和印刷法，都很易施行。牠們可以適應許多種的教材。所以教師對於這種變相的簡答測驗，可以廣為應用。

填字測驗——在許多客觀測驗中，填字測驗最爲人所習知，在教師方面應用也極普遍。填字測驗與簡答測驗不同的主要點，在測驗的形式與複雜的變化。牠係由不完全的語句所組成，而非由許多問題所組成。牠雖能與簡答測驗同樣的簡單，但卻能變得複雜與多含。牠包括一套的句子或段落，而將其中重要的字空出，叫學生在空白的地方填入相當的文字，以完成各個語句，而使其具有正確精當的意義。學生回憶基本的觀念或文字，以完成一種思想的能力，實爲測量學生是否熟知教材的良好方法。於需要時應用知識的能力，也能用填字測驗測量出來。所以教師應當採用這種填字測驗。在客觀的考試中，這種測驗要算是最好的一種了。

填字測驗的價值——填字測驗有幾個優點，與簡答測驗相同。學生僅需少量的書寫，可使在一個測驗時間內，測量很多的教材。填寫的速度，各個學生差異很大，測驗的難度也是如此。填寫測驗易於適應各種教材，極適合於普遍的應用。編製優良的填字測驗，可以切實減少猜中正確反應的機遇。欲於多量的教材中回憶所需的正確反應，則學習非精確與澈底不可。對於這種測驗，可使『急來抱佛脚』的準備，歸於無效。牠供給求得精確知識的直接動機，也鼓勵每天專心準備教室工作。

填字測驗較簡答測驗更易激發思想。位於空缺前後或相與緊接的字句，能刺激強烈的思想活動，且足以決定學生反應的適當與否。句子的上下文所選擇與應用的字，須十分注意其確切的意義。這是填字測驗的一個重要的優點。清楚的思想的表現，須具有豐富的有意義的字量。在某種情形下，填字測驗注重表現的精確，而不需多量的書寫。就這點而言，牠具有論文式考試的優點，而沒有其缺點。學生用了最少的書寫，可以包含長而複雜的歷程和程序，或一種科目的長的討論。連續的工作或事情，可用點子橫寫，預備學生填充。這個特點，乃填字測驗的顯著價值之一。

從填字測驗的思想過程的性質看來，牠有時被評為智力測驗或字彙測驗 (Vocabulary test) 而非成業測驗 (Achievement test)。這也有一部分的理由。正確的應用已學的教材，乃普通智力的測量。應用已經獲得的東西的能力，又為成業所不可缺。就這個觀點言，填字測驗在測驗順序上實具有一種重要的功能。一般學生對於牠，普通不如其他客觀測驗那樣喜愛。這是由於牠較難的緣故。至於喜愛牠的學生則以為牠不易擾擾；猜度較少，故較公平；不易引起錯誤的印象；較為澈底；應用所學的東西需較多的思想與能力。要編製一個優良的填字測驗，需要很多的思想與時間。就批閱而言，牠在客觀測驗中是最慢的一種。但如逐題而論，牠在所有的新法考試中或許要算最可靠的了。

編造填字測驗的技術——要編造一個優良的填字測驗，需多量的時間與真實的思想。倘使不是精心編

造，往往有極多可能的答案。有了許多可能的答案，測驗的客觀性便減少，記分也覺困難。此外還有一種危險，即過重機械的聯結，而輕視思想的過程。在補充空缺時，須練習健全的判斷。空缺每使測驗太難太易，或失其價值。為便利計，編造填字測驗的技術可分下列三大步驟：

1. 測驗材料的準備。
2. 選擇空缺的答案字。
3. 規定學生填寫的方法。

測驗材料的準備——編造填字測驗的第一步，就是準備測驗材料。這個步驟包括將測驗材料編成各種重要語句在內。此種語句可以是獨立的句子，也可以是成一段落的許多關聯的句子。有的時候，各種語句係從教科書中逐字摘錄出來。但這樣一來，便有過重機械的記憶的危險了。不過這不是常有的事情。若以不同的形式編造各種語句，較用教科書中的形式要好得多。許多問題的編造，可依照論文式考試的格式，並為每個問題寫一適宜的答案。此種答案便可用作填字測驗的材料。編製測驗材料最應注意的要點，即既經決定的語句，要重要，要適當，不要過於依附教科書。下面所舉的即可用作實例：

亞馬遜河流經巴西，為全世界最大而次長的河流。

有三種基阿那(Guanas)：英國人，荷蘭人，法國人。

盧內熱約里 (Rio de Janeiro) 爲全世界通行葡萄牙語最大的城。

全世界產白金的最大河流是在哥倫比亞。

巴拿馬帽的草辦產於厄瓜多爾。

亞馬遜河發源於秘魯。

在南美諸國中，只有哥倫比亞兩面濱着大洋。

厄瓜多爾橫跨着赤道。

哥倫比亞的首都是波哥大 (Bogotá)。

厄瓜多爾的統治階級是純粹的西班牙血統。

厄瓜多爾的主要出口貨是可可 (Cacao)，其種子可做可可與諸果律漿 (Chocolate)。

在南北美洲中，歐人所首建的都會是秘魯的利馬 (Lima)。

秘魯擁有全世界最高標準軌幅的鐵路。

*阿根廷於1816年離西班牙獨立，自建共和國。

*智利是全世界最狹長的國家，首都曰森地亞哥 (Santiago)。

哥倫布於1492年作第三次航行時，發現委內瑞拉海岸。1499年亞美利瓜味斯浦奇 (Amerigo Vespucci)

探險至委內瑞拉與哥倫比亞海岸，在馬拉開巴湖 (Lake Maracaybo) 的沿岸，發現一用木排建築的印第安人的村落，便名之曰委內瑞拉，意即小威尼斯 (Little Venice)。

選擇空缺的答案字——測驗材料編製好了，第二步便是選擇空缺的答案字。這一個步驟，可以決定測驗的價值，故需審慎的判斷。要選擇空缺的字，有幾點應加注意。第一，凡足以使語句所欲表達的意義含糊的字不應缺掉。例如「亞馬遜河發源於祕魯」中缺掉了亞馬遜河與祕魯，則這個語句便顯然的變為無價值了，因為缺掉牠們語句的原意便盡失去了。「……河發源於……」這個語句能指任何地方的任何河流，所以沒有意義。一個語句如空缺過多，而致空泛無定，對於測驗目的，便失其價值。

第二，凡對於受測驗的學生太難或太易填寫的字不應缺掉。填寫的難度隨空缺的字數與性質而異。所以空缺的字如由句子的前後文決定得太清楚，或暗示得太明顯，便失去測驗的價值，因為這樣便太易填充了。這可以用前面許多實例中的最後一個來證明。倘使「亞美利瓜」，「村落」，「沿岸」，「委內瑞拉」等字缺掉，便變成不完全的語句，如「1492年……味斯浦奇探險至委內瑞拉與哥倫比亞海岸，在馬拉開巴湖的……發現一用木排建築的印第安人的……便名之曰……意即小威尼斯。這樣的空缺便太易填充了。所要填寫的字，可由其前後文或其他關係極切的字來決定或暗示。無論如何，這樣的空缺是沒有意義的。所以這種語句便失去了測驗的可能性。反之，同樣的語句，也可改得太難。例如缺掉了「一」

「亞美利瓜，味斯浦奇，」「委內瑞拉，」「哥倫比亞，」「印第安，」「馬拉開巴湖，」「委內瑞拉，」「小威尼斯，」便變成如下的不完全語句：「……年，……探險至……與……海岸在……的沿岸發現一用木排建築的……村落，便名之曰……，意即……。」

這種語句僅留這樣一個大綱，當然太難達到測驗的目的了。

教師用了適宜的空缺，便能使填字測驗的難易，恰到好處。所以上面所述的注意點，應加以嚴密的考慮。各個空缺應當是對於思想的關鍵字。不重要的字，或對於思想關係極小的字，不應空掉。一個填字測驗不可在施行前幾分鐘內編製。牠須預先幾天編製好了，擱置在旁，然後再加考察，以求可能的改進。倘隔了若干天，教師用以測驗自己，便可發現有許多地方須待修改。經過了充分的練習，教師便會熟能生巧，留出適宜的空缺，使成一個良好的填字測驗。

規定學生填寫的方法——通常關於填字測驗的填寫方法，並沒有特殊的注意。不完全的語句，或是隨機排列，或是依照論理的次序排列，學生則就空白的地方，逐一填寫。這對於學生已是無可再便的了。但在雜亂情形之下，批改四散於考卷的各個答案，就比較困難了。有時答案量尺的編造，可用透明的材料，如透明角質，臘紙，或薄紙；或在紙板上開許多小孔，使與測驗的空缺相合。在小孔的下面寫適當的答案。但這種方法也是非常拙劣而不圓滿的。比較優良的方法，莫如將空缺標明數目，在測驗的邊上留着與空缺數字相當的答案地位。此種

方法對於學生填寫，雖不甚便，但在記分上卻有外加的便利來補償。於是答案紙的編造，祇須在某一份測驗卷上寫着正確的答案，用以與學生的答案相對照。下面的形式，便是填字測驗：

說明：下面是幾個句子，其中有某幾字是缺着的。在邊上標着數字的地方，填入最適當的字，以補足語句中與數字相當的空缺，使牠成一完全而正確的語句。用正確的字填充的空缺數，就是所得的分數。

- (1) (1) 流經 (2)
- (2) 為全世界最大而次長的河流。
- (3) 有三種基阿那 (3)
- (4) (4) (5)
- (5) (6) 為全世界通行葡萄牙語
- (6) 最大的城。
- (7) 全世界產白金的最大河流是
在 (7)
- (8) 巴拿馬帽的草辦產於 (8)
- (9) 亞馬遜河發源於 (9)

- (10)在南美諸國中，只有(10) 兩面濱着大洋。
- (11)厄瓜多爾橫跨着(11)
- (12)哥倫比亞的首都是(12)
- (13)厄瓜多爾的統治階級是純粹的(13) 血統。
- (14)厄瓜多爾的主要出口貨是(14) 其種
- (15)子可做(15) 與(16)
- (16)在南北美洲中，歐人所首建的都會是
- (17)(17) 的
- (18)(18)
- (19)(19) 擁有全世界最高標度的鐵路。
- (20)阿根廷於1842年離(20) 獨立，
- (21)自建(21) 國。
- (22)智利(22) 是，全世界最狹長的國家，
- (23)首都曰.....。

- (24).....(24).....於1498年作第三次航行時，發現委內瑞拉海岸。
- (25).....1499年(25).....探險至委內瑞拉與
- (26).....(26).....海岸；在馬拉開巴湖的沿岸發現一用木排建築的印第安人村落。
- (27).....便名之曰委內瑞拉，意即(27).....。

填字測驗的答案地位與簡答測驗一樣，其排列有時用點或線，或正確的答案的第一字，或兩者兼而有之，以指明正確反應的數目。例如，以前諸題中的第一個可寫成：

(1)亞.....(1)亞——————流經

(2)巴.....(2)巴.....，為全世界最大而次長的河流。

這個方法雖注重確切的答案，協助學生決定其反應的是否正確，並減少答案的紛歧，而使便於批閱與記分，但因牠使答案過於明瞭，便損害測驗的價值。牠是將一部份認識的要素與回憶相聯絡。有的時候，關於特殊困難的教材的測驗上，此種方法可以通用，在一般的應用上，我們便不應採用牠。此外缺的空白，最好長度相同，這樣便可避免以空白的長短來暗示填充的字的長短了。

填字測驗的施行與記分——施行填字測驗最圓滿的方法，莫如印刷充分的本子，分發各個學生，使在上面做他的工作。用口述法施行填字測驗，不能獲得成功。要把一個不完全的語句，記在心頭，未免太難了。要把口

述的完全而有意義的語句或問題在心上考慮已屬難能；若空缺了重要的字致失去語句的意義，便須記住許多斷續的字與四散的空白，學生將益覺困難了。所以如由口述，則變成屬於記憶方面的心能測驗，而非教材成績的可靠測驗了。如由板書法施行填字測驗，則可以獲得優良的結果。其缺點與其他測驗應用板書法時同。除教師須要書寫外，在施行測驗前，尚須使學生不預見工作。倘用此法施行測驗，黑板上的空白，應寫明號數。學生考卷上所寫的號數，應與黑板上空缺的號數相同。這些號數的排列應依照三四一六頁所示卷子的邊上的答案形式。學生在考卷上註明號數的地方，寫一個或幾個最適當的字，以補充每句中與號數相當的空缺，俾成一完全而正確的語句。倘使學生用格式一律的紙，則能用答案紙批改答案。如果教師讀着每個空缺的正確答案，學生隨着批改卷子，這樣也是可以的。

(學級)六年級 (科目)地理
 診斷分數表 (譯註一)

(日期)1929年11月1日
 (工作單元)南美洲

學生		分	
張祖良	王蔚	郭一齋	楊振亞
張其昌	張恩祖	方天錫	周祖燾
黃芸	顧加	陸撲誠	高一峯
楊浩如	陶容芳	陳心農	朱維明
李元溥	姜琮	安思汝	蔣麗貞
倪家璞	李博文	呂嘉華	華明
趙其美	施自安	楊景明	許慶祺
謝天漢	薛芹	王仁壽	費昌福
馮仰賢			
總分			

診斷分數表 (續)

(學級)六年級
(科目)地理

(日期)1929年11月1日
(工作單元)南美洲

題目	張祖良	王蔚	郭一齋	楊振亞	畢其昌	張魁祖	方天錫	周望祖	黃芸	顧珊	陸樸誠	高一峯	楊浩如	陶容芳	陳心豐	朱維明	喬人俊	李元厚	姜涼	安思汝	蔣麗貞	倪家璞	李博文	呂嘉華	崔明	趙其廷	施自安	楊景明	許慶祺	謝天漢	薛芥	王仁壽	費昌福	馮仰賢	總分
15	✓	✓	✓	✓										✓	✓	✓	✓	✓	✓		✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	15	
16	✓						✓				✓				✓	✓	✓	✓										✓	✓	✓	✓	✓	✓	✓	10
17	✓									✓		✓			✓	✓	✓	✓	✓		✓				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	13
18		✓	✓						✓		✓	✓	✓		✓	✓	✓	✓	✓		✓			✓			✓							✓	14
19	✓							✓			✓			✓	✓	✓	✓	✓	✓					✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	13
20	✓								✓								✓	✓										✓	✓	✓	✓	✓	✓	✓	5
21	✓																✓	✓																✓	3
22		✓		✓												✓	✓											✓					✓	✓	9
23			✓							✓		✓	✓							✓														✓	7

的方法來記答案的分數，總比用不同的分數來評量答案要圓滿得多。在可能的時候，回憶測驗的記分，應該用答案紙，因為牠的正確性和便利性都比較大些。

應用回憶測驗的結果——教師在着手編造回憶測驗以前，第一件事便應決定爲什麼要施行這種測驗，並如何應用其結果。測驗倘使不能證明其本身能完成第二章中所討論的各種功用，即使施行也無目的。簡答測驗適用於檢查日常工作，以及引起日常專心準備的動機。不論用口述的形式，或印刷的形式，他對於各種教材都有廣博的適應性，編造與施行也比較容易。填字測驗對於刺激日常工作與檢查指定作業不甚適用，雖適用於各種教材，但卻難於編造與施行。牠不能應用口述法。簡答測驗與填字測驗都能引起精確學習的動機。填字測驗尤須用字正確，表述適當。牠還能供應一種方法，以測量描寫種種連續的與複雜的動作或事實的能力。回憶測驗所需的心能作用和態度，與認識測驗略有不同。這些測驗的特殊優點，本章均已詳細討論過。至於回憶測驗的總分，很可用以計算學生相對的成績，與決定其學業地位和等第。

回憶測驗的結果，有重要的診斷價值。例如教師施行一個填字測驗，包含題目如下：

- (1) ……キ中國的山脈起於帕米爾高原東部的(1) ……。
- (2) ……キ山西礦產以(2) ……。三項爲大宗。
- (3) ……キ(3) ……。鐵路爲歐亞交通最捷的途徑。

- (4).....キ(4)倫敦跨(4).....河的兩岸。
- (5).....キ(5)日本人屬(5).....族。
- (6).....古巴現在的獨立，係獲(6).....幫助。

為改進教學與學習起見，對於這種測驗，應注重做錯的題目與錯誤的種類，而不應僅知每個學生全部測驗的總分。對於測驗中每個題目的錯誤應加注意，俾於必要處予以校正的教學。補救教學的效率如何，應由以缺點為中心的另一種測驗決定之。回憶測驗的診斷價值是不應忽視的。

倘能計劃一種格式，以表示錯誤的題目與做錯的學生，則回憶測驗結果的分析便容易了。一三七—四〇頁的診斷分數表為達到這種目的很便利的格式。假如一三四—六頁的填字測驗已經施行過了，檢查出的錯誤也已列入診斷的分數表內，這樣便表示每個學生所做的錯誤，以及全級學生所共同錯誤的範圍；這樣並能指示重教全級時應特別注意的在那裏，每個學生須特殊教學的又在那裏。教師如果不應用這種方法去分析測驗結果，並從而做校正的工作，則測驗的價值便要失去不少。這種診斷的分數表供給很有價值的知識。其他測驗也能適用。當教師編造一個優良的測驗，要備嘗艱辛；學生受此測驗，又要費去寶貴的時間，所以結果的全部價值都應利用。

溫習與討論問題

- (1) + - ? 回憶測驗是暗示答案的。
- (2) + - ? 傳統的論文式考試是一種回憶測驗。
- (3) + - ? 簡答測驗與論文式考試相同。
- (4) + - ? 簡答測驗的主要弱點是缺乏客觀性。
- (5) + - ? 簡答測驗極適用板書法。
- (6) + - ? 填字測驗在客觀測驗中最高劣。
- (7) + - ? 填字測驗所能包括的材料有限。
- (8) + - ? 一個編製完善的填字測驗，能將猜度的因子減至最少。
- (9) + - ? 回憶比認識難。
- (10) + - ? 填字測驗適用於長而複雜的歷程。
- (11) + - ? 學生通常喜歡填字測驗，而不喜歡其他測驗。
- (12) + - ? 一個優良的填字測驗很難編造。
- (13) + - ? 通常，測驗中的語句，應異於教科書中語句的形式。
- (14) + - ? 各個空缺應含有重要的答案字。

- (15) + - ? 各個空缺應使語句的原意含糊。
- (16) + - ? 空缺的字應由其前後文很清楚地暗示出來。
- (17) + - ? 填字測驗批閱時應用標準答案。
- (18) + - ? 點和線的方法應變成普通的實用方法。
- (19) + - ? 填字測驗易於採用口述法。
- (20) + - ? 對於正確答案的同義字應給以足分。
- (21) + - ? 各種答案平常只應視作非正即誤。
- (22) + - ? 回憶測驗的結果僅有極少的診斷價值。

第九章 綜合測驗與雜式測驗

綜合測驗——綜合測驗(Combination test)包括二種或二種以上不同形式的測驗。牠由各個特殊的部份組成，每個部份代表一不同的形式，故有時又名混合測驗(Composite test)牠又常稱彙選測驗(Diversity tests)，因為牠將許多相同的測驗方法排列為若干組，以求產生一綜合的測量結果。聯合數種測驗形式而成一單獨測驗，乃比較冗長與廣博的學期與學年考試的妥善辦法。各式的測驗，既可綜合而成單一測驗，則綜合測驗應有特殊的考慮。

綜合測驗的優點——綜合測驗便於編造。在計劃一種廣博的考試時，我們每發現某種重要教材，其本身適於某一種形式的測驗，而不適於其他形式。用了幾種不同的形式，便可選擇特殊教材的性質所需的特殊形式。在這方面，一個考試頗與高爾夫球(GOLF)的遊戲相似。打高爾夫球時，有一種便利，就是有許多手杖，使我們得照球之地位與方向的性質而選擇之。考試的施行也是這樣，有了各種的測驗形式，從這種形式根據教材的性質與需要而加以選擇，便要便利得多。一個單獨的考試用了幾種形式，不僅使測驗易於編造，並且使考試增加興趣。變化足以引起興趣，單調易於使人乏味。一個很長的考試，如只限於一種測驗形式，不待做完，早已感覺單調。因為這個緣故，綜合測驗的變化性，其本身具有顯明的優點。牠能增加考試的融洽性；而且不同的形式對

於教材須有不同的心理反應。認識測驗所需認識事物的能力，異於回憶測驗所需回憶事物的能力。綜合測驗能由各種不同形式的測驗而開拓更廣的心理能力。對於教材的各種心能反應，如能使其活動，有所取樣（sampled），則結果更爲可靠。綜合測驗對於成績可爲廣博的測量。所以適於完成第二章中所討論的各種測驗功用。總之，綜合測驗的主要優點，在編造簡易，增加考試的融洽性，含有範圍較廣的心理反應，以及給予更可靠的結果以實行各種測驗的功用。凡較長的考試如含有大單元的教材，則此種優點能保證其可用。

編造綜合測驗的技術——編造綜合測驗的技術，包含編造組成綜合測驗的各個測驗的技術在內。但此外還要加上幾種專門的細則，此種細則大半是關於考試的各種適當選擇與排列。各種教材要選擇適宜的測驗形式，此種形式要排成良好的程序，便須要知道各種正確的方法。編造綜合測驗的技術，可以分爲下列三大步驟：

1. 編造正確的語句。
2. 將各種語句改成適宜的測驗形式。
3. 規定學生做記號的方法。

編造正確的語句——編造綜合測驗的第一步，是編造一個包含許多正確語句的表，課程的各個重要部份都應包括在內。凡是要測驗的教材，應專心檢閱一番，以便選擇所應採入的各種要節。有些語句可採自學生

於學程中讀過的補充讀物。包括這種材料可以引起一種優良的動機，使學生能夠專心研究教師所指定的參考書。這些語句的敘述形式，應與教科書中的不同。多用教科書的語句，趨向於鼓勵記憶的方法，而不能鼓勵從學得的教材中去推理。教師如能在教學歷程中，逐日將適宜的測驗材料摘入考試記錄簿或卡片，則正確語句的編製，要簡單得多。

所擬的各種語句，應較測驗中實際應用的為多，最好比所需的能多一半以至二倍。其中須要刪去的，必定不少，如含糊的，衝突的，依賴其他語句而成意義的，與其他語句重複的，多含的，冗長的，用字欠妥的，不關重要的，以及含有生僻或不常見的字的。這些語句及其類似的語句，倘使不能改進，都應刪除。只有那些能夠適應以前各章中可用語句的標準的，才可保留，以作考試之用。這些語句應能代表一種適當的難易範圍。有些要非常容易，使全體學生都能回答。有些要非常困難，使僅有少數學生能回答得正確。難度適中的語句，比較要多些。通常容易的，困難的，以及難易適中的語句，為數最好大約相等。

將各種語句改成適宜的測驗形式——編造綜合測驗的第二步，是將各種語句改成適宜的測驗形式。關於這個步驟，有幾點須加考慮。第一，慎選應用的測驗形式。這事有關於教師個人的愛好，教材的性質，施行測驗的方法，以及測驗所欲完成的主要功用。如果用口述法施行考試，最好限於正誤測驗與簡答測驗。如果口述法與板書法並用，便可加入對偶式與選答式。綜合測驗太長了，所以不能單用板書法來施行。倘印刷便利，則任何

測驗形式，均可採用。施行綜合測驗最好的方法，是將其印刷，使每個被測驗的學生，都能獲得一份。這種綜合考試，應當包括認識式與回憶式。正誤式，選答式，填字式與其他形式如對偶式，簡答式，或雜式之一，可以構成一個優良的綜合測驗。

第二，每種測驗式所包括的問題數應有約略的估定。這雖然不免有些由教師獨斷，但須視規定的考試時間與學生對於各式測驗的反應速度而定。教師應根據其經驗，以估定學生反應各式測驗的速度。正誤測驗是最快。認識測驗較回憶測驗迅速。假使一個六年級的考試定四十五分鐘，其中正誤式，選答式，填字式同時並用，而每式所費時間，欲求相等；那末可以編造一個七十五個正誤題，六十個選答題，以及四十五個填字題的考試。教師如果以為六十個選答題不易編製，則可減至四十題，而將正誤題增至一百個。教師可應用便捷的方法，以規定一個考試各部的適當長度。

第三，將正確的語句改成預定的形式。各種語句都應詳加檢閱，以求其適於何式測驗。你不難發現，有些語句易於適用某種形式，而不適用其他形式。本身適於選答式的語句，較正誤式或填字式為少。正誤式比較最易應用。一個語句所用的測驗形式決定以後，便應在前面做一記號，例如「正」代表正誤測驗，「選」代表選答測驗，「填」代表填字測驗，餘依此類推。所有語句都這樣標明了記號，凡欲採取同一形式的，應歸在一處，並依照以前各章所述的方法，將其逐一改成各種測驗式。有的時候，各式測驗，不問異同，將其混合在一處。這種方法

便叫做混雜測驗 (Omnibus Test) 所以這樣命名，因為其中的問題含有混雜的性質。混雜測驗不適於非標準的教室考試。要達教室考試的目的，應將各種測驗形式，各自排列成部。

規定學生做記號的方法——編造綜合測驗的第三步，是規定學生做記號的方法。在測驗卷的前面，應有關於全體測驗的普通說明或指導。此種說明應包括測驗所由組成的各個部數，以及做測驗時所用的方法等。下面的例子便是規定這種形式的一種指導：

社會科學考試

……年級

姓名……

日期……

(每面都要寫你的姓名)

一般的說明： 這個考試包括四部。各部都要做得快。在你沒有做完全部測驗以前，不要爲了困難問題多費時間。各部的說明都要仔細去看。測驗中每個題目都要用心閱讀。

第一部

說明： 下面許多問題，有幾個是正確的，有幾個是錯誤的。你認爲是正確的，沿「+」的四週畫一圓圈；認爲是錯誤的，沿「-」的四週畫一圓圈。分數是所做題數減去做錯題數的兩倍。

- 十一 暹羅居民除土著外，以中國人爲最多。
- 十一 絲的輸出額，以法國爲第一。
- 十一 歐洲最大之水爲窩瓦河。
- 十一 中國的貴州產鉛最多。
- 十一 蒙古有戈壁沙漠。

測驗的其他各部，也應依照所包括的測驗形式，冠以相同的說明。各種形式的適當說明，及其做記號的方法，在前幾章中已經論及。如把一份測驗卷子，加以正確記號，可以作爲批改的標準。如將用作標準的卷子的邊緣摺轉，使正確的答案可以靠近學生的答案，則閱卷較易。標準答案上所做的記號，以及測驗卷子的批改，最好都用紅的或藍的鉛筆。所用的方法與以前所述的各種個別測驗的方法相同。（註一）

規定測驗報告的形式——一個優良考試的編造，很費思想與時間，所以應有多餘的本子，留作將來之用。這就是說，考試卷子不能發還學生，測驗材料應妥爲保存。要替代發還考卷，便應另想一種方法，將結果報告學生。這種報告的形式，應包括學生所欲知道的結果的一切事項。如將這種形式編排於測驗第一面的下部，便很

（註一）對於綜合測驗的各部記分，如感有興趣，可參考 W. A. McCullough How to Measure in Education, Chaps. 9-10; 或 Charles

易撕下，以發還學生，則極爲便利。下面所示的排列，尙覺滿意。

注意：這方小紙是記載你的考試結果的。考卷概不發還。在下面指明白的空白上，寫你的姓名與日期。

姓名：.....

日期：.....

社會科學考試

六年級

第一部分數：.....參加測驗的人數.....

第二部分數：.....學級平均數.....

第三部分數：.....標準差.....

第四部分數：.....你在全級中的等第.....

總分：.....你的五分位數.....

備註：

測驗批完以後，便應將結果錄入發還的報告單中。教師應注意每個學生已否將自己的姓名寫在發還的單上。於是把這些單子撕下，用以計算其他事項。將這些單子依照分數的多寡排列，便能排成一種材料，由以計算學級平均數，標準差，以及個人的等第等等。這是處置材料的一種很便的方法。把所要知道的事項求出，並記

入報告單後，便可將報告單依姓氏的筆畫排列，並將材料記入教師的學級記錄簿。於必要時，還可把牠們依照學生的座位排列，使易於發還。考卷的下部，可以摺疊起來，至姓名爲止，使除所欲發還的學生外，不能知道其中一切事項。

發還報告單的時候，最好將考試的結果，乘機與學生作詳細的討論。化分數爲等第的方法，應加以說明。所批最高與最低的分數，每個等第記號的分數距離，以及每個等第內的學生數，應寫在黑板上，俾學生知道。他們將樂於知道這些事項。關於測驗的一般批評以及對於學生所發問題的答覆，應當自由坦白。學生如能澈底明瞭全部的方法，將益信考試的正直與分數的公平。養成對於考試的優良態度，是第三章中所討論的融洽性之一，也應視爲測驗技術的一種。

雜式測驗——應用的雜式測驗 (Miscellaneous types of tests) 有好幾種。牠們的應用較有限制，大都僅適用於特種的材料。有的時候，這些雜式中的一種，也能用作極優良的測驗，以測量特殊的事物。因爲可以應用而教師應熟知的測驗有：

重組測驗 (rearrangement test)：

指認測驗 (Identification test)。

證明測驗 (Verification test)。

重組測驗——重組測驗包括許多隨機排列的同類事物，叫學生依照某種標準，將其重行排列。牠有時又稱連續測驗 (Continuity test)，因為各種事物須依照其連續的性質而排列。例如一個依照筆劃多少或字母先後為序的域名表，可令學生按其大小，重行排列。一句句子中的各個字，可顛倒了命學生排成適宜的次序。一串隨機排列着的事實，可囑學生依照年代的先後加以重組。完成某種工作的各種動作，可隨機排列成表，叫學生依照進行方法的程序，重行排列。一個隨機排成的食物表，可依煮熟所需時間的長短，重新排列。有的時候，這種測驗所欲測量的，為關於問題中某種性質的相對差異，或關係之普通欣賞，而非測量每個問題所含的知識確量。如遇這種情形，重組測驗實為一種完滿的方法。一個優良的重組測驗，有自六個至十個的問題，也儘夠了。下面便是一個重組測驗的例子：

說明：下面是美國的許多域名。選出你認為最大的城，在其前面的線上寫一(1)；再選出次大的城，在其前面的線上寫一(2)；餘類推，直至你依照牠們的大小排完為止。(域名係照原文為首字母的次序而排列)。

巴爾的摩爾 (Baltimore)。

波士頓 (Boston)。

芝加哥 (Chicago)。

克勒維蘭 (Cleveland)。

底得律 (Detroit)。

洛司阿格勒司 (Los angles)。

紐約 (New York)。

非勒特爾非 (Philadelphia)。

舊金山 (San Francisco)。

聖路易 (St. Louis)。

重組測驗的記分比較複雜，因其有各種不同的正確程度。一個把非勒特爾非記作(5)的學生比記作(9)的學生，雖則同是錯誤，但對於城的大小觀念，較為清楚。以數字註明洛司阿格勒司大於克勒維蘭與克勒維蘭大於聖路易的學生，其各個城排列的次序，雖然完全錯誤，但對於關係的觀念，還算清楚。所以僅僅計算做正數的總和，並不是適宜的記分方法。下面的記分方法，對於不同的正確程度，計算極為精確，並能查出測驗所要考試的，易言之，即關係的知識。將標準答案放在學生所定的等級旁邊，並在他的卷子上，就測驗中第一題的對面，把他寫在標準答案中標明(1)的問題上的數字寫上。其下，在測驗中第二題的對面，把學生寫於標明(2)的標準題上的數字寫上，其餘依此類推，直至全部的測驗造成一個校正項(Checking Column)為止。

下面的例子，便是以上面的城市表，說明這種方法。(譯註二)

標準答案	學生所定的等第	校正項
8	5 巴爾的摩爾	1
6	8 波士頓	3
2	6 支加哥	2
6	5 克勒維蘭	4
4	4 底得律	6
5	6 洛司阿格勒司	5
1	1 紐約	10
3	2 非勒特爾非	9
10	7 舊金山	8
7	10 聖路易	7

從校正項計算分數，可在校正項中把數字(1)勾銷，並在其對面寫未勾銷的各項數。——在這個情形下為9項(即3, 2, 4, 6, 5, 10, 9, 8, 7)。次及數字(2)，把牠勾銷了，在其對面寫未勾銷的各項數。——在這個情形

下爲7項(即4, 6, 5, 10, 9, 8, 7)。如是依照數的次序前進, 至校正項中各個數字都勾銷完爲止。其法表示如下:(譯註二)

校正項

X	9
X	7
X	7
X	6

(譯註一)紐約的標準答案爲1, 而學生所定等第亦爲1, 故在校正項中第一項寫「1」字; 芝加哥的標準答案爲2, 而學生所定等第爲3, 故在校正項中第二項寫「3」字; 舊金山的標準答案爲10, 而學生所定等第爲7, 故在校正項第十項寫「7」字; 以此類推。

(譯註二)在校正項中分爲二行, 第一行各數爲學生所定等第。先把此項中的「1」數勾銷, 並在其對面寫下「1」數; 下未勾銷的項數, 在本例中適爲9。次及2數, 亦在其對面寫下2數; 下未勾銷的項數, 在本例中適爲7。次及3數, 4數等亦如之。惟3數在本例之第一行中, 適在2數之上, 故從2數, 亦在對面寫下7字。又如第一項中的7數, 下面並無一項, 即學生以此城列爲第十, 故未勾銷的項數爲0。因此8, 9, 10數之未勾銷項數亦爲0。

5	4	4
5	4	4
10	0	0
8	0	0
8	0	0
7	0	0
<hr style="width: 50%; margin: 0 auto;"/>		
37 = 分數		

問題之數，不論多少，其最高分數之求得，可將校正項中各數依照正確的次序排列之，而後再依照上述的方法記分。最高分數也可利用公式計算：

$$\frac{n^2 - n}{2} \text{ 或 } \frac{n(n-1)}{2}$$

公式中 n 代表測驗中的問題數。在十個問題的測驗中，其最高分數為：

$$\frac{10(10-1)}{2} \text{ 或 } 45$$

倘使校正項與數的次序完全相倒，分數當然就等於零了。

倘使直接把題目依適宜的次序寫下，而不用代替的號數，則可直接應用批改重組測驗的同樣方法。這樣，

可叫學生依照大小的次序，把下面的十個美國城市寫下。學生的答案表與標準答案表，並舉例如下：

標準答案表

學生答案表

紐約	紐約	0
芝加哥	非勒特爾非	7
非勒特爾非	芝加哥	7
底得律	底得律	0
洛司阿格勒司	克勒維蘭	0
克勒維蘭	洛司阿格勒司	0
聖路易	舊金山	0
巴爾的摩爾	波士頓	0
波士頓	巴爾的摩爾	0
舊金山	聖路易	0

37 = 分數

要批學生答案表的分數，須在其上勾銷發現於標準答案表上的第一題，並在其旁將發現於其下的未勾

鉛的問題數寫上。復次，在學生答案表上，勾銷其發現於標準答案表上的第二題，並在其旁將發現於其下的未勾銷的問題數寫上。如此循序以進，至各題均已記分爲止。然後將各題的分數總加起來。（註一）

指認測驗——指認測驗通常包括一幅地圖，表解，圖形，圖畫，或其他種圖的表示，由學生指認其中某種性質。牠與對偶測驗相似，有時又名聯對測驗（Paired-associates test）。用一暗射地圖，可命學生指認其中的某政治區域，地質，城市，富源，物產，以及其他等等。一個幾何圖形可用以指認各種性質。一幅表解或圖畫可用以指認各個特殊部分的地位。指認測驗特別適用於測驗地理，自然科學，以及幾何中的某種教材。回憶式或認識式均可用於指認測驗。下面的例子，可用以說明牠的形式：

說明：下面是一幅南美洲的地圖，各個國家均表示於略圖中，並各各註明數字。圖旁係一南美各國的國名表，其前均留有空白。把適當的數字（圖中註着的）寫在圖旁各個國名前的空白上。分數就是做正的題數

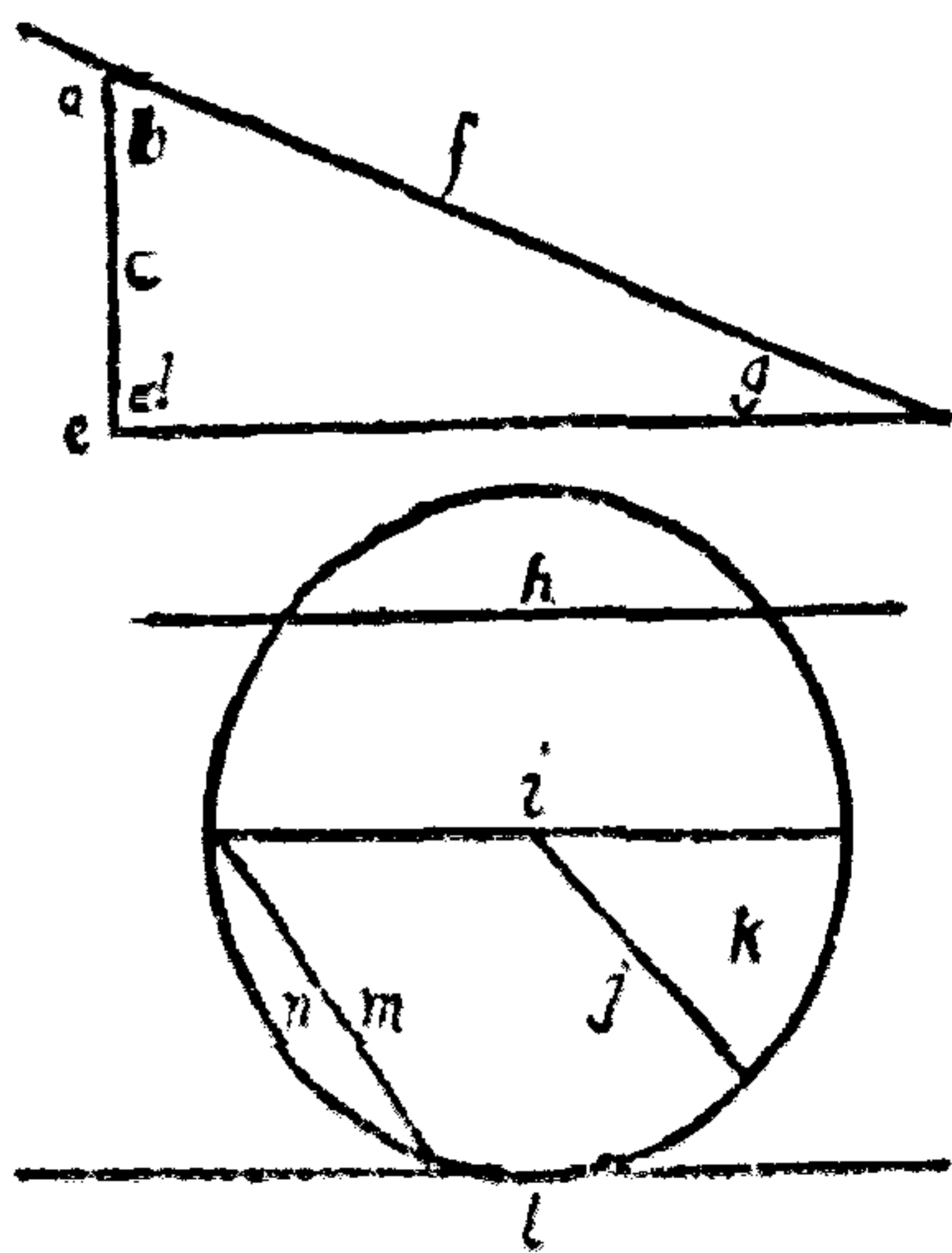
（註一）採自 Howard E. Wilson 的論文，載於 School Review, Vol. 38, No. 2, (1930年2月)



-阿根廷 (Argentina)
-玻利維亞 (Bolivia)
-巴西 (Brazil)
-智利 (Chile)
-哥倫比亞 (Colombia)
-厄瓜多爾 (Ecuador)
-圭亞那 (Guiana)
-巴拉圭 (Paraguay)
-祕魯 (Peru)
-烏路圭 (Uruguay)
-委內瑞拉 (Venezuela)

說明： 下面是兩個幾何的圖形，用英文字母代表各個不同的部份。旁邊為許多名詞，其前各留空白。在圖旁各個名詞前的空白上，以適當的字母（見圖）指認各部的名字。分數就是做正的題數。

- 銳角
- 弦
- 直徑
- 斜邊
- 股
- 鈍角
- 半徑
- 直角
- 割線
- 扇形
- 切線
- 頂點



上面兩個例子代表指認測驗中的認識式，如將數字與字母寫在旁邊，其後留以空白，然後囑令學生將適宜的名詞填入，使與地圖和幾何圖形上的數字與字母相當。這樣，便將認識式改成較難的回憶式了。有少數問

題，不在圖的表示以內，則認識式便更難了。例如，南美洲地圖的問題中，可以包括中美洲諸國，如宏都拉斯（Honduras）、尼加拉瓜（Nicaragua）、與薩爾瓦多（Salvador）。各種說明應重述一遍，並應告知學生，旁邊所列諸國，地圖上不一定都能找到。指認測驗如照上述的形式編造，批改時可以應用標準答案，很為迅速。這樣測驗因要繪圖，所以很難編造。將圖掛於黑板上面，已很圓滿，不必印刷分發。對於某種教材，指認測驗是一種優良的形式。有的時候，牠是適應各種測驗需要的唯一形式。

證明測驗——證明測驗包括許多正確的語句，囑令學生用主要的理由證明每個語句的真理所在。牠有時又叫判斷測驗（Judgment test），因為要從事種種必要的比較與分別，以及搜集主要的證據，以證實一個語句或證明其正確，便須加以判斷。答案的多變使證明測驗的記分異常主觀。牠的記分，可以用「非對即錯」的方法，或者也可以依照答案假定的優良程度而給以分數。下面便是證明測驗的例子：

說明：下面是關於地理上的幾句語句。在每個語句的空白上，將各該語句所以正確的主要理由，加以簡要的說明。說得最好的理由得二分，部分正確的得一分，沒有說的得零分，說錯的得負一分。

1, 南美洲的厄瓜多爾晝夜的時間常常是約略相等的。

理由：.....

2, 黃河易於潰決。

理由……

3. 錢塘江的湖水最大。

理由……

其他

其他的雜式測驗——在雜式測驗中，有幾種是教師不常用的，如相似測驗（Similarities）類比測驗（analogies），同義字測驗（Synonyms），反義字測驗（Antonyms），與刪字測驗（Cross-out）等。相似測驗學生須由許多雜類的字中選出與其他已知的字的相同字。例如在下面的例題中，希望學生能認識前三個名詞都是省會，而武昌這個名詞，應從其他各個名詞中選出，因為牠也是省會，故與前三個名詞相類。

✱杭州，鎮江，安慶……上海，漢口，武昌，九江。

類比測驗，學生須求出某種環境下事物間關係的類似（在他種環境下並不相合）。例如，在下面的例中，學生在第一題內須填寫華盛頓，因牠與美國的關係，猶南京與中華民國的關係；第二題內須填寫太和族，因牠與日本的關係，猶盎格魯撒克遜族與英國的關係。

✱中華民國：南京∴美國……

✱英國：盎格魯撒克遜族∴日本……

同義字測驗，須使學生認識意義相同或近似的字。例如在下面的例子內，學生須選「明白」，因牠同意義與「清楚」相似。

清楚………糊塗，明白，遮蓋，不定。

反義字測驗或相反測驗 (Opposites test)，須使學生認識意義相反的字。在下例中，學生應選「稀少」，因其意義適與「稠密」相反。

稠密………堅實，緊密，密集，稀少。

刪字測驗包割錯誤在內，學生須查出而刪去之，舉例如下：

巴拿馬運河屬於美國與巴拿馬。

江鎮，漢口，杭州與廣州都是省會。

教師以少用這些雜式測驗為是，但對於為着特殊目的而編的各種測驗式，則應熟知之。凡富於創造精神的教師，如欲編造一種測驗，以應用於普通測驗所不適用的特殊教材，則此種知識很為有用。

應用綜合測驗與雜式測驗的結果——有些雜式測驗的特殊價值，本章前已論及。重組測驗能用以考驗各種關係的欣賞或知識。指認測驗能顯示學生對於某種材料的熟習程度。證明測驗可以供給一種方法，以顯示學生見了一種知識或一個結論時，能否證明其是非真偽的能力。雜式測驗如用以刺激心理反應，則與各個

個別的測驗一樣，有時也能有很好的應用。牠們的特殊價值，有時使牠們用作綜合測驗的一部。構成綜合測驗的各種普通方式，其價值與應用，前已討論及之。

綜合測驗最廣博的用途，是在決定學生的學業成績。所以學期中和學期的客觀考試，通常採用綜合的形式。此種考試的結果，常常用作一種基礎，以定奪學業的等第，並決定失敗或通過的學生與分數的多寡。這種用途極為重要，所以第十一章完全是討論測驗分數如何可用以決定一個學校制度中所採用的各種分數。與此種價值連同的應用便是升級，分級，分組，規定新工作的起點，以及特殊的研究如決定各種學習方法的效率等等。倘欲獲得學習的可靠測量，便當應用綜合測驗。學業成績的決定，乃綜合測驗的合理的與重要的功用。

但是倘將綜合測驗的結果，如通常所為一樣，僅僅應用於分數的分配與學生的分等，而不推廣其用途，實在是錯的。各種總分應有一永久的記錄，以便年復一年的保留下去。這樣，便能建立種種客觀標準，以便日後用於級際的比較。一個綜合測驗的分數能確定的顯示成績，並能供應極有價值的各該科教學目的。綜合測驗中各項分數的學級等量之建立，也是極有價值的。牠能作為一種實際的標準，後來施行此種同樣的考試時，便可用以公平的把各個學生或小團體，加以分等。關於客觀標準的供給，教師能發現綜合測驗有極有價值的應用。

測驗的應用如果至獲得總分而止，便要失去許多測驗價值。倘使不注意各個問題的個別反應，綜合測驗的結果不能稱為完滿的應用。注意了這些反應，便能明白全級或個人的優點與弱點，而引入更有效的教學。考

卷如果發還了學生，則測驗在將來的應用價值便不能獲得。但各種考卷應在個別的談話中與學生一同檢討一番。測驗的最後目的在改進教學與學習，而這種方法對於達到這種目的，最爲有價值。教師應牢牢記着，測驗的各種功用，並不是互相排斥的。在主要的功用以外，還有別種價值，也應當利用。測驗的各種功用既是互掩的，所以當考慮各種測驗結果時，第二章中所討論的種種應用，都不應忽視。

溫習與討論問題

- (1) 十 — ？ 每個測驗均應採用綜合式。
- (2) 十 — ？ 教材的性質與測驗的選擇無關。
- (3) 十 — ？ 綜合測驗的多變性足以增加融洽性。
- (4) 十 — ？ 所有的客觀測驗能引起同樣的心理反應。
- (5) 十 — ？ 教師在着手編造測驗以前，不應考慮測驗材料。
- (6) 十 — ？ 編擬的語句應較實際需用的爲多。
- (7) 十 — ？ 各個語句應代表各種不同的難度。
- (8) 十 — ？ 一個綜合測驗應包括回憶式與認識式兩種。
- (9) 十 — ？ 一個考試與其太長，無寧太短。

- (10) 十 - ? 非標準的測驗不採用混雜測驗。
- (11) 十 - ? 綜合測驗的各部，都應有分別的說明。
- (12) 十 - ? 應准許學生保留他們的考卷。
- (13) 十 - ? 教師不應抽出時間與學生討論達到考試等第的方法。
- (14) 十 - ? 重組測驗包括各種不同的問題。
- (15) 十 - ? 重組測驗的記分較對偶測驗複雜。
- (16) 十 - ? 指認測驗需要圖的表示。
- (17) 十 - ? 指認測驗可為記憶式或認識式。
- (18) 十 - ? 指認測驗包括各種正確的語句。
- (19) 十 - ? 指認測驗需主觀的記分。
- (20) 十 - ? 在教師最常用的諸測驗中，有相似式，類似式，同義字式，反義字式，與刪字式。

新法考試

第十章 新式測驗在各種學科上的應用

新式測驗在各學科上的應用——在以前幾章裏，對於新式測驗的編造，施行和記分等各種方法，都已講過了。爲使方法明瞭起見，例子也舉過了。在這一章裏，因欲幫助教師能夠把新法測驗應用於各種學科起見，還要另加說明。我們要明白，新式測驗應用於教材的範圍，比沒有經驗的教師所見到的要廣大，以下所舉的例子，係用以說明牠們對於讀法、國語、算學、社會科、自然科、外國語以及其他各科的適應性。這種例子，不過聊供參考，並不完全，也不能把牠們的用處，盡量表白出來。爲教師者如果應用得宜，幾乎任何學科都能適用的。教師能藉着經驗，培養各種能力，以靈巧的適應各種特殊教材與特殊目的。

讀法——讀法過程的根本要素，是閱讀的速率，和對於所讀材料的了解。這二種要素的非標準的測驗，應爲平日教學的一部分。亞爾馬克和蘭二氏，關於讀法科速率和了解的測驗，有幾個建議如下：（註一）

（一）選擇一篇四百字左右的文章，材料只要合於比所測驗的年級低一級的程度就適用了。出十個或十個以上的問題，其中包括全文的主要思想，然後把開始讀起的地方指示學生，並且告訴他們這次練習的目的在測驗他們閱讀的速率，和對於所讀材料的了解。當聽說『起』的時候，大家就開始閱讀，直到一分鐘完了才停止。當停止的記號發了出來，學生就在剛剛讀到的字上做一記號。照着這個方法，把全文

讀完。讀完之後，把問題寫在黑板上，叫學生回答。這個工作做畢以後，叫學生各自計算每分鐘內所讀的字數。每個學生閱讀的速率和答準的題數，應該與其姓名兩兩相對的寫在黑板上。這種測驗每月可以舉行兩三次；有時也不妨變更一點，讀了二次課文以後，再回答問題或測驗速率，也無不可。

(2) 叫學生把地理、歷史或衛生等科內指定的部分默讀四五分鐘。在測驗終了的時候，叫他們計算所讀的字數，並回答許多根據所讀材料而出的問題。其速率和了解可照建議一的方法計算而得。

(3) 一個默讀和朗讀的混合測驗，可編造如下：叫每個學生朗讀一二分鐘，當「停止」的記號發出的時候，在最後讀的一個字上做一個記號，叫他們計算字數，然後從終了的地方起接着默讀，時間與朗讀相等。再叫他們計算字數，然後把分數同姓名相對着寫在黑板上。至於各種課文的了解，可用「思想的問題」來測驗。這二種分數的比較是討論朗讀與默讀的良好根據。

課文中思想的獲得與否，很容易用新式考試來測驗。有些測驗的方法，可以下面的例子來表示。以下所建議的方法任何年級都可應用。牠們須適應於正在讀的特殊材料及正在受測驗的團體的能力。教師所處的地位最可以做這種必需的適應。讀法測驗的編製，有的是為課外的閱讀，有的是為課內的工作。

新式測驗應用在讀法上的例子

(註1) J. C. Airmack and A. R. Lang. "The Beginning Teacher", p.p. 266—67 Boston, 1928

正誤法

*(王冕——高級復興國語教科書第一冊第九十兩課)

十一•王冕是一個醉心利祿的人。

十一•王冕小的時候爲秦家牧牛。

十一•王冕後來成爲一個畫家。

*(知道正義的兒子——同上第廿七八兩課)

十一•波斯老人有四個兒子。

十一•長子能夠不在暗地裏傷害他的仇人。

十一•次子能夠見義勇爲。

十一•寶劍名馬結果送給幼子。

選答法

*(王冕——見上)

……王冕是很①勢利的②嶮崎磊落的③鄙吝的④小心翼翼的⑤粗魯的。

……王冕小的時候①家境很好②父母很有錢③讀了十幾年的書④爲人家牧牛⑤不事生產。

……王冕後來成爲①政治家②帶兵官③畫家④富翁⑤教師。

キ（知道正義的兒子——見上）

……次子報告的事情①不苟取②見仇不報③旅館客人④寶劍名馬⑤救落水的孩子。

……波斯老人後來把寶劍送給①長子②次子③幼子④公益機關⑤自己作養老金。

「長子把全數歸還明珠的故事說了之後，波斯老人說這是應該做的事，沒有甚麼可貴，若是你暗地裏拿了他的明珠，你想你是一個怎麼樣的人呢？」

……於下列各字中，在最足以表示長子當時的感覺的那個字下，劃一條線，①恐懼②興高彩烈③驕傲④憤怒⑤心平氣和。

填字法

キ（王冕——見上）

王冕生長在……縣。

王冕小的時候爲人家……

王冕後來成爲一個……。

（知道正義的兒童——見上）

波斯老人把財產分做……份。

三個月期滿了，三個兒子都回來向波斯老人報告……

波斯老人聽了……的話，就把……送給他。

簡答法

キ（王冕——見上）

王冕爲什麼不能繼續讀書？……

王冕的性情怎樣？……

王冕後來成爲什麼？……

（知道正義的兒子——見上）

波斯老人有幾個兒子？……

波斯老人把家產分作幾份？……

波斯老人把名馬寶劍送給那一個？……

同義字法

……キ胆怯①心驚②心虛③懦弱④乏力⑤害臊。

新式測驗在各種學科上的應用

……
× 膈膈① 忸怩② 怕羞③ 汗顏④ 不安⑤ 不耐煩。

刪字法

× 我看見鳥一株樹

× 樹上有的一鳥巢

× 她歡愛她的父親

國語(註一)——國語包括文學與語法。一個文學的學程，須使學生對於詞藻美麗或情節生動的名著，能夠欣賞和熟習。現在雖然還沒有令人滿意的欣賞測驗，不過對於文學作品的欣賞，必須與其熟習有密切的關係。至於學生對於文學的熟習和普通知識，則可用新式測驗來測量。

新式測驗應用在文學上的例子

正誤法

十一 × 「吳剛伐桂」是古時候的一個神話。

十一 × 西廂記是施耐庵做的。

十一 × 賈寶玉是紅樓夢裏的主角。

(譯註一)此處所謂國語，在原著中為英語，在本書中則指漢文，所舉例子，係仿原著體裁而加以改動。

選答法

……紅樓夢的作者是①施耐庵②曹雪芹③董解元④羅貫中⑤吳敬梓

……宋江是下列一書中的主角①紅樓夢②水滸③西廂記④三國志⑤儒林外史

……杜甫石壕吏的詩是描寫①逃難的人②苛稅③水災④拉夫⑤年老人

……胡適著有①嘗試集②女神③草兒④隔膜⑤寄小讀者。

對偶法

……吳敬梓 ①史通

……曹雪芹 ②史記

……董解元 ③女神

……羅貫中 ④彷徨

……施耐庵 ⑤隔膜

……劉 鶚 ⑥上下古今談

……吳稚暉 ⑦老殘遊記

……葉紹鈞 ⑧水滸

新式測驗在各種學科上的應用

……魯迅

⑨西廂記

……郭沫若

⑩紅樓夢

……司馬遷

⑪三國志

……劉知幾

⑫儒林外史

填字法

①……阿Q正傳是①……著的。

②……曹操是②……裏的主角。

③……劉勰著的一部不朽之作叫③……

⑬天方夜譚

⑭今古奇觀

⑯文史通義

簡答法

司馬遷著的是關於那一方面的書？……

長恨歌是那一個作的？……

辛西廂記的男主角是誰？……

一個語法的學程，不論對於思想或感情的發表，或者是口頭的，或者是文字的，都應使學生熟練。口頭發表的能力難於測驗，所以考試祇應用在文字的發表上。文字發表所包括的有作文，字形和書法等三項。書法能力的最好測量，就是把學生所寫的字同標準量表相比較，字形就是把適當的筆劃按照準確的順序書寫而成字，所以一個字的形狀非正即誤；主觀的判斷同字形的記分無關，故記分是極客觀的。至於測量字形的能力的方，莫善於普通教室中所舉行的聽寫，不過新式測驗法在某種理由下而似乎合宜時，也可以應用。作文能力的測量須藉論文式的測驗，因為論文式的測驗中，學生才有以寫述方式來發表理想的機會。記作文分數最好的方法，是把牠同標準作文量表相比較，或者與同班中其他的作文相比較，這種作文須據評判結果，可以代表各種成績的。語法的能力雖然是多方面的，卻很容易用新式考試去測驗，因為這種能力普通屬於技術的性質，大部分不出知識的領域。

新式測驗應用在語法上的例子

正誤法

（文法）

十一 形容詞是用以形容名詞或代名詞的。

新式測驗在各種學科上的應用

十一 一個他動詞必須有一個賓詞去補足牠的意義。

十一 一句陳述的句子，一定是問一個問題的。

十一 幸副詞是用以形容名詞的。

(字形) (譯註一)

十一 幸正

十一 幸真

十一 幸狗

十一 幸馬

(句子的意義)

十一 幸彼方讀書

十一 幸彼寫

十一 幸現在的人

選答法

(譯註一) 此處在原著中爲拼法，本書改爲字形，例子也都換過——按英文的字形也決定於拼法。

(文法)

……把字與字，短句與短句連絡起來的是①名詞②動詞③介詞④形容詞⑤接續詞。

……キ表示女性的代名詞是①她②他③它④牠⑤此。

(準確的應用)

……キ他…殺•①要②想③看④去⑤被

……キ我…煩惱•(1)見(2)要(3)被(4)很(5)想

(字形)

……キ(1)武(2)或(3)武(4)或(5)武

……キ(1)華(2)華(3)華(4)華(5)華

(字彙)

……キ胆怯的意思是①害臊②小心③胆子小④怕冷⑤躲避

……キ滔滔的意思是(1)水流貌(2)雨下貌(3)漫天貌(4)水名(5)水下滴聲。

對偶法

(文法)

新式測驗在各種學科上的應用

- ……キ普通名詞 (1) 指事物之數量程度。
- ……キ固有名詞 (2) 多用以指示事物之性及態。
- ……キ疑問代名詞 (3) 指示一切事物或前文之名詞。
- ……キ指示代名詞 (4) 連接兩詞或兩句。
- ……キ形容詞 (5) 狀動詞之態。
- ……キ副詞 (6) 代不知之人或憑虛起問。
- ……キ接續詞 (7) 其名為其人其物之所專有。
- ……キ示紀形容詞 (8) 其名通用於同類之物。

填字法

(文法)

- ①……キ集……而成句；集……而成文。
- ②……キ短句有兩種；一為……一為……。

簡答法

(文法)

キ指示事物之數量程度的是什麼？……
キ以於、與、爲、等字屬於那一種詞類？……

刪字法

(準確的應用)

キ他正在讀看報。

キ那座塔很長高。

(改正錯誤)

……キ他在什麼？

……キ我喜歡這一塊樣子。

……キ我沒有不看見太陽，

算學——一般人常以爲新式考試對於測驗算術 (arithmetic) 和代數的能力，並無實際的用處，對於測驗幾何的能力，爲用也極有限。他們的理由是算學 (mathematic) 在各學科中最爲客觀，牠的答案或是或非，毫無懷疑的餘地。所以就是在其他各科上贊成新式測驗的人，也持這種意見。實際上，算學雖具較大的客觀性，不過，有的研究已經告訴我們，教師記算學試卷的分數，同記他科的論文式試卷有同樣的差異。如果每個

問題的記分，取非正即誤的方法，雖然能夠免除這種差異，但在某幾種算學的材料裏，應用新式測驗的方法，在規定時間內，可以包括更多的內容。在算學的學程中，應該培養運算基本四則的技能和解答問題的能力。新式考試很可用以測驗這種技能和能力。

新式測驗應用在算學上的例子

正誤法

(算術)

- 十一 分數中在橫線以上的數目叫做分母。
- 十一 $\frac{3}{4}$ 是一塊金元的四分之一。
- 十一 體積是用立方來表示的。
- 十一 六打雞蛋，等於 72 個。

(代數)

- 十一 $2a - b$ 是個二項式。
- 十一 $2(b - c)$ 等於 $2b - 2c$ 。
- 十一 $\frac{a+b}{c} = \frac{a}{c} + \frac{b}{c}$ 。

十一 如 $3x - 7 = 14 - 4x$, 且 $x = 3$,

(幾何)

十一 鈍角大於 90° 。

十一 角的大小視兩邊的長度而定。

十一 一直角等於一平角的四分之一。

十一 角是可以加減來計算的。

選答法

……以小數點來表示就是 ① 0.05 ② 0.5 ③ 50 ④ 0.15 ⑤ 15。

……如果一數被別數所除，牠的答數叫做 ① 被除數 ② 積 ③ 商 ④ 餘數 ⑤ 和。

……一個公司按照股票所發給的金額，叫做 ① 紅利 ② 獎金 ③ 票面價額 ④ 經手費 ⑤ 佣錢。

(代數)

……如果 $5x - 4$ 等於 16, 則 x 等於 ① 20 ② 12 ③ 4 ④ $\frac{12}{5}$ ⑤ $\frac{16}{5}$

……一個三項式的例子是 ① $2a - b$ ② $2a + b - c$ ③ $3x$ ④ $\frac{x}{3}$ ⑤ $3(a - b)$

……當~~y~~等於 $\frac{y}{x}$ 的時候，則y等於① $\frac{x}{z}$ ② $\frac{z}{x}$ ③ $\frac{x}{z}$ ④ $\frac{z}{x}$ ⑤ $\frac{z}{x}$

(幾何)

……一個四邊形內各角之和為① 320° ② 160° ③ 180° ④ 90° ⑤ 360°

……沙堆的形狀是①角柱體②矩形③圓錐④圓⑤角錐體。

……圓的四分之一叫做①扇形②弧③象限④半徑⑤三角形。

對偶法

(算術)

…… 245 (1) CCLXII

…… 950 (2) CDLXXXVII

…… 693 (3) LXXVII

…… 483 (4) CML

…… 139 (5) DXXXVI

……77 (6) DCXXVII

……1000 (7) DCCCLXXIV

..... 874	(8) XCIV
..... 325	(9) CXI
..... 570	(10) M
..... 111	(11) CCXLV
..... 262	(12) CXXXIX
..... 919	(13) DLXX
..... 536	(14) CMXXX
..... 98	(15) CCCXXV

(幾何)

.....圓的面積	(1) 直徑的三方的四分之一乘3, 1416.
.....橢圓的面積	(2) 長短二軸的乘積 $\times 0.7854$.
.....正多角形的面積	(3) $\frac{1}{2}$ (底邊 \times 高)
.....平行四邊形的面積	(4) $\frac{1}{2}$ (各邊之和 \times 邊心距)
.....梯形的面積	(5) 直徑 $\times 3, 1416$.

新式測驗在各種學科上的應用

- ……三角形的面積
 - ……圓周
 - ……圓的直徑
 - ……圓柱的側面積
 - ……圓錐或角錐的側面積
 - ……球面積
 - ……圓柱的體積
 - ……球體積
 - ……圓錐的體積
- 填字法
- (6) 半徑的平方 $\times 3.1416$.
 - (7) 直徑 \times 圓周
 - (8) $\frac{1}{2}$ (底的面積 \times 高)
 - (9) 高 $\times \frac{1}{2}$ (平行二邊之和)
 - (10) 圓周 $\div 3.1416$
 - (11) 圓底的周 \times 高
 - (12) $\frac{1}{2}$ (圓底的周 \times 斜高)
 - (13) 底面積 \times 高
 - (14) 底 \times 高

(算術)

- (1) ……如果一盞燈每晚用油一品脫(Pint), 問一加侖(Gallon)能敷……晚之用。
- (2) ……百分之幾的百分(%)不過是……的另一個名詞。
- (3) ……利息加……稱做本利和。

(代數)

①……方程式的任何一項，可以自左邊移至右邊，或自右邊移至左邊，只要牠的符號……。

②……如把 $7XY$ ， $-3XY$ ，和 $-4XY$ 相加，其總和是……。

(幾何)

- (1)……以三直線圍成的圖形是一個……。
- (2)……一個小於直角的角叫做……角。
- (3)……直角三角形中，直角以外二銳角之和等於……度。

簡答法

(算術)

沿關閉圖形的全距叫做什麼？……

800元的百分之七是多少？……

(代數)

包括二項的多項式叫做什麼？……

如果 a 等於 1, b 等於 2, c 等於 3, 則 $a-bc$ 等於多少？……

新式測驗在各種學科上的應用

(幾何)

在直角三角形中，與直角相對的叫做什麼邊？……
等角三角形中每一角有幾度？……

指認法

(幾何) 參看157頁

社會科學——前幾章所舉的例子，有許多是屬於社會科學的。這是因為社會科學範圍內的教材，如歷史、公民、地理、經濟、社會、政治等，其本身就適用於新式測驗的方式。茲為篇幅所限，勢難每科都舉例子。下面所舉的，不過說明測驗在社會科學的各方面都有應用的可能罷了。

新式測驗應用在社會科學上的例子

正誤法

(地理)

十一 洛陽在河北省

十一 黃石公園在哥羅拉多 (Colorado)

十一 緬甸在澳洲。

十一 蘇必利爾湖 (Superior) 沿岸有許多鐵礦。

(公民)

十一 中國國民黨中央執行委員會由全國人民選出。

十一 國民政府主席由國民代表大會產生。

十一 美國總統的選舉在三月裏舉行。

十一 中國現在的政治是政黨政治。

(歷史)

十一 紐約初爲英人所佔。

十一 美國的南北戰爭使移民運動進行遲緩。

十一 1882年美國國會禁止中國學生入境。

十一 秦始皇因爲要遊玩江南而築長城；隋煬帝因爲要防禦外寇而開運河。

選答法

(地理)

……美國太平洋沿岸最大的城是①西雅圖 (Seattle) ②舊金山 (San Francisco) ③聖第亞哥 (San

新式測驗在各種學科上的應用

Diego) ④洛司阿格勒司(Los Angeles) ⑤波特蘭(Portland)。

……南美洲最大的國家是①秘魯②巴西③阿根廷④哥倫比亞⑤厄瓜多爾。

……世界最高的高原在①印度②中國③英國④美國⑤德國

(公民)

……一個人第一個接觸的社會是①城鎮②鄰里③家庭④教堂⑤學校。

……司法部包括①市長②上議院③警察④省長⑤大總統。

……中國國民黨所規定的政府權限是①地方分權②中央集權③中央與地方均權④永遠一黨專政⑤勞農階級專政。

(歷史)

……同美洲的發現，很有關係的一樁發明是①火藥②印刷③地圖的構製④指南針⑤望遠鏡。

……中國第一個通西域的是①張騫②班超③司馬相如④馬援⑤諸葛亮。

對偶法

(地理)

……里斯本(Lisbon)

①葡萄牙

……阿根廷

①棉花

.....	不魯舍拉 (Brussels)	②	羅馬尼亞	中國	②	茶
.....	哥本哈根 (Copenhagen)	③	比利時	烏拉圭	③	錫
.....	馬德里 (Madrid)	④	俄羅斯	巴拉圭	④	家畜
.....	百倫 (Berne)	⑤	波蘭	哥倫比亞	⑤	金
.....	布達佩斯 (Budapest)	⑥	瑞士	密士失必	⑥	羊
.....	柏爾格刺德 (Belgrade)	⑦	西班牙	秘魯	⑦	油
.....	那不勒斯 (Naples)	⑧	意大利	厄瓜多爾	⑧	糖
.....	華沙 (Warsaw)	⑧	西伯利亞	玻利維亞	⑨	波羅蜜
.....	莫斯科 (Moscow)	⑩	匈牙利	明尼蘇達 (Minnesota)	⑩	橘
.....	不加勒斯多 (Bucharest)	⑪	奧大利	加那大	⑪	咖啡
.....	維也納 (Vienna)	⑫	丹麥	加里福尼亞	⑫	軟木
				智利	⑬	橡皮
				巴西	⑭	可可
						⑮	硝石

新式測驗在各種學科上的應用

(公民) (附註一)

…… 軍政部長

① 陳紹寬

…… 國府主席

② 孔祥熙

…… 內政部長

③ 蔣介石

…… 行政院長

④ 汪精衛

…… 軍事委員長

⑤ 李元鼎

…… 實業部長

⑥ 陳公博

…… 海軍部長

⑦ 林森

…… 財政部長

⑧ 何應欽

…… 審計部長

⑨ 黃紹雄

(歷史)

…… 柏爾(Bell)

① 織布機

(附註一) 原著中的例子亦為當時的副總統與部長等。

- ……愛迪生 (Edison) ② 旋轉印刷機
- ……飛爾德 (Field) ③ 刈禾機
- ……福爾敦 (Fulton) ④ 縫紉機
- ……谷第耳 (Goodyear) ⑤ 電話
- ……和 (Hoe) ⑥ 冰藏車
- ……蒙 (Howe) ⑦ 橡皮和硫法
- ……麥柯密克 (McCormick) ⑧ 飛機
- ……馬可尼 (Marconi) ⑨ 拍照機
- ……麥根塔勒 (Mergenthaler) ⑩ 無線電
- ……模斯 (Morse) ⑪ 白熱燈
- ……瓦特 (Watt) ⑫ 輪船
- ……衛斯亭豪斯 (Westinghouse) ⑬ 排字機
- ……輝特尼 (Whitney) ⑭ 電報
- ……來特 (Wright) ⑮ 用氣壓的汽車制動機

新式測驗在各種學科上的應用

填字法

(地理)

①……景德鎮以產①……著名。

②……印度是②……的屬地。

③……美索不達米亞在③……與④……兩河的流域。

(公民)

①……英國是一個①……帝國

②……國民黨把……權歸之於民，而

③……辛把……權歸之於政府。

(歷史)

①……辛①……發現美洲。

⑩ 大西洋的海底電線

⑪ 軋棉機

⑫ 蒸汽機

②……キ②……在中國做了官，回去著東方見聞記。

③……キ訂立片馬條約的是③……與

④……④……兩國

簡答法

(地理)

……キ揚子江發源於何省?

……キ中國那一省人口最密?

……南美洲那一國最大?

(公民)

……キ國民政府的組織，與歐美各國政府組織最大不同之點何在?

……キ爲什麼要有訓政時期?

……キ人民有那幾種政權?

(歷史)

……キ開運河的是誰?

……キ拿破崙最後被打敗的地方是那裏？

……キ歐洲大戰最先開火的是那兩國？

指認法

(地理)キ

……黑龍江

……雲南

……浙江

……遼寧

……外蒙古

……廣西

……江蘇

……吉林

……新疆

……廣東

……山東

……熱河

……西藏

……福建

……河北

……察哈爾

……綏遠

……青海

……江西

……北平

……甘肅

……四川

……安徽

……山西

……寧夏

……貴州

……河南

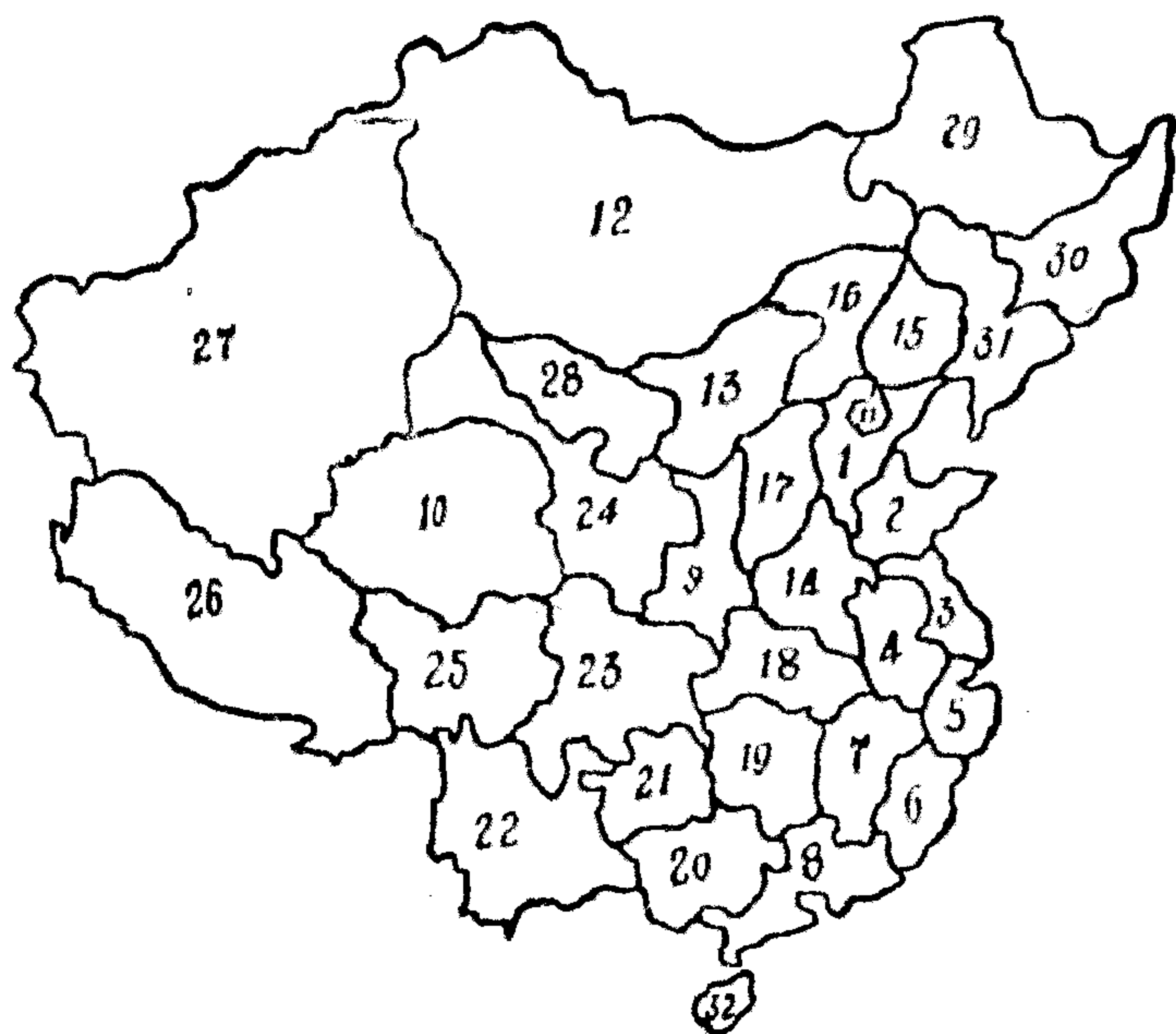
……陝西

……西康

……湖南

……湖北

……海南島



(衛生)

新式測驗在各種學科上的應用

自然科學——自然科學包括農業、植物、衛生、生理、動物、化學、地質、物理及自然研究等科目。牠們的教材是精確的，所以問題的答案，務須簡潔。嚴格說起來，並不適用於傳統的論文式考試。所以就一般而論，自然科學的考試早已採用客觀的方式了。不過那時所謂客觀同現在的意義不同。在自然科學方面，新式測驗就使不能很顯然的比舊式測驗更為客觀，但所包含的範圍更為完全，批改試卷的時間也可減少。所以自然科學的教材，很適用於新法考試的方法。就算篇幅很多，也不必把每科的新式測驗盡行列舉。下面所舉的例子，儘可表示新式測驗在各種自然科學上的應用了。

新式測驗應用在自然科學上的例子

正誤法

十一 一塊普通的布，只要浸在熱水裏，就可以消毒。

十一 急救法可用以代替醫生。

十一 當傷處出血頗多時，所出的血，能使傷處清潔。

(生理)

十一 脾的部位是在胃後的左側。

十一 鼻爲味覺所必需。

十一 膽囊分泌膽汁。

(化學)

十一 冰的融化是一種化學變化。

十一 氯氣是沒有顏色的。

十一 綠礬就是硫酸銅。

(農業)

十一 肥的母雞，一定很會生卵。

十一 用各種方法耕種是最安穩的制度。

十一 各種昆蟲都應一律撲滅。

選答法

(化學)

……大理石含炭的百分率是①6%②24%③3%④12%⑤18%

……硫化鋅的顏色是①紫②黃③白④橙⑤黑。

……氮的分子式爲① N_2 ② N_3 ③ N_4 ④ N_5 ⑤ N 。

(衛生)

……瘧疾的發生是因爲①蒼蠅②不清潔的水③熱的天氣④蚊子⑤潮濕的空氣。

……肌肉的發達是靠著①休息②運動③藥物④含有澱粉質的食物⑤治療。

……微生物的發現者是①巴斯德 (Pasteur) ②拉齊爾 (Lozear) ③戈加斯 (Gorgas) ④拍爾曼 (Pelman) ⑤愛迪生 (Edison)

(農業)

……能由插木生長的樹是①松樹②橡樹③柳樹④橘樹⑤胡桃樹。

……一種可以肥田的植物是①苜蓿②玉蜀黍③蕃薯④煙草⑤麥。

……累蒿痕 (Leghorn) 是一種①山羊②牛③鴨④綿羊⑤雞。

對偶法

(化學)

- | | |
|---------|-----------------------------|
| ……臭氧 | (1) NaOH |
| ……二氧化二氮 | (2) H_2O_2 |
| ……氯化氫 | (3) NH_3 |
| ……硫酸 | (4) H_2S |
| ……硃精 | (5) HCl |
| ……二氧化氮 | (6) NH_4Cl |
| ……二氧化鐵 | (7) CO |
| ……二硫化碳 | (8) H_2SO_4 |
| ……碳酸鉀 | (9) NaCl |
| ……氯化鈉 | (10) NaOH |
| ……氯化銻 | (11) HgO |

……氧化錄

(12) O_2

……氧化銀

(13) K_2CO_3

……一氧化鉛

(14) CS_2

……水

(15) H_2O

(16) NO_2

(17) CO_2

(18) PbO

(19) Ag_2O

填字法

(生理)

①……食物在口內①……同

②……②……混合起來然後

③……嚥下,由③……作用,把

④……④……帶下

新式測驗在各種學科上的應用

(自然研究)

- ①……雜草的種子大多數為①……所散播。
- ②……一種②……可以吸住鐵針。
- ③……花的③……上附著花粉。

(化學)

- ①……鐵的生鏽,因為牠同空氣中的①……相化合。
- ②……由於生石灰與礬砂的交互作用,便產生②……

簡答法

(化學)

- ……一公升水有多少立方厘米?
- ……氯酸鉀的顏色怎樣?
- ……醇的分子式怎樣?

(自然研究)

- ……豚鼠吃何種食物?

……熊怎樣過冬？

……大多數的花裏面，顏色最鮮妍的是那一部分？

(物理)

……檢驗蓄電池通常用什麼儀器？

……長距離的電線是什麼東西做的？

指認法

(植物)

……雄蕊

……花瓣

……心皮

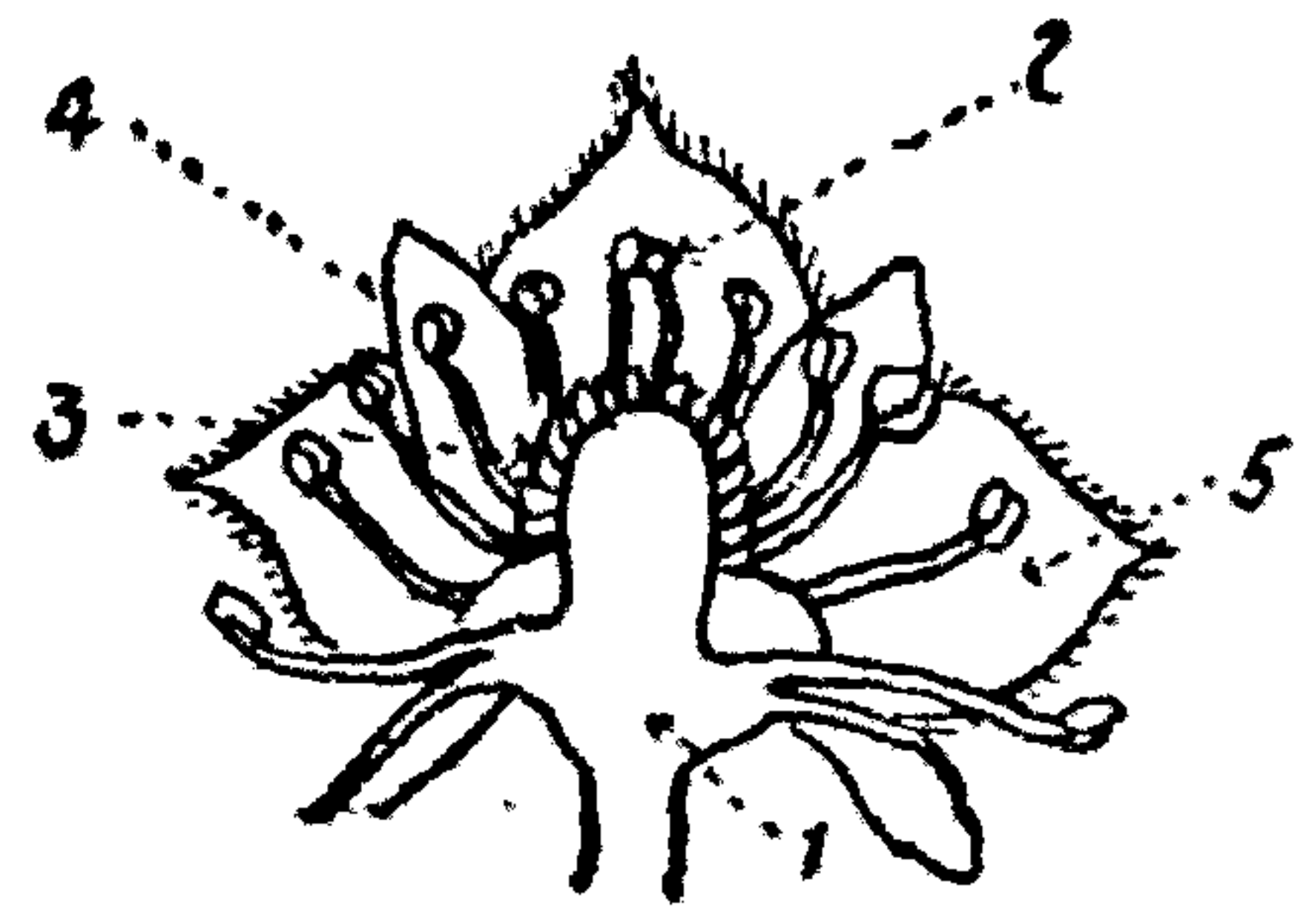
……萼片

……花托

重組法

(物理)

新式測驗在各種學科上的應用



密度大小的次序：

……	瑪瑙	……	水銀	……	鉑
……	蜂蠟	……	金鋼石	……	銀
……	黃銅	……	金	……	銅
……	銅	……	鐵		
……	木栓	……	錫		

外國語（譯註一）——外國語的口頭發表和文字發表，其適用於客觀的測驗，並不比國語容易。關於這種能力的測量，實在需要論文式的測驗；在論文式的測驗中，學生才能把自己真正的表現出來。不過，如文法，字彙，準確的應用和閱讀的了解等，很可用新式測驗來測量。下面的說明，以英文為例，但對於法文，德文，拉丁文等也同樣的適用。

新式測驗應用在外國語上的例子

正誤法

（譯註一）在原著中，以西班牙文為外國語的例子譯者以英語在我國為最普通的外國語，故即以之為例。所舉各個例子，雖與原著異，但體裁則仍相似。

(文法)キ

十一 在英語中以 *sion* 爲末尾的名詞，多半是抽象名詞。

十一 “They” 是祇代表人類的。

十一 名詞之末加 *s* 的，表示複數。

(讀法)

爲欲測驗學生對於所選讀的英文的了解，可以英文或國語編造許多句子，其中有的是對的，有的是錯的。方法同一二頁上的讀法一樣。

選答法

(文法)キ

……一個表示所有格的代名詞是 ① You ② it ③ Their ④ they ⑤ him.

……下面是一個副詞 ① good ② beautiful ③ Nice ④ my ⑤ Very.

(字彙)キ

…… Friend ① 兄弟 ② 親戚 ③ 朋友 ④ 姊妹 ⑤ 叔伯。

…… Strong ① 勢力 ② 強權 ③ 大 ④ 強健 ⑤ 高。

新式測驗在各種學科上的應用

(讀法)

爲欲測驗學生的了解，可從英文的選文上用國語或英文編造許多選答的句子，方法同二二頁上的閱讀一樣。

對偶法

(文法) *

(字彙)

.....To	①前置詞	also	①悲傷的
.....Cup	②連續代名詞	ball	②手指
.....What	③名詞	city	③字
.....That	④發問代名詞	easy	④強健
.....and	⑤接續詞	friend	⑤新
.....going	⑥序數	finger	⑥信
.....good	⑦分詞	ink	⑦牛乳
.....his	⑧不定代名詞	lily	⑧少壯的
.....his	⑨指示代名詞	milk	⑨風

.....Very	⑩ 形容詞	New	⑩ 鉛筆
.....To see	⑪ 物主位字	pencil	⑪ 亦
.....The	⑫ 無定冠字	strong	⑫ 窗
.....a	⑬ 副詞	Sad	⑬ 空虛的
	⑭ 基數	Vain	⑭ 墨水
	⑮ 無定式動字	Window	⑮ 百合
	⑯ 有定冠詞	Word	⑯ 球
		Wind	⑰ 朋友
		Young	⑱ 容易的
			⑲ 城
			⑳ 廣闊的
			㉑ 和平
			㉒ 水
			㉓ 廉價的

填字法

* He.....My brother.

* We shall.....home.

* The story.....never have been told.

重組法

* See a I man.

* He me to Visit Comes.

其他各科——新式測驗可以應用於各種教材，各科的教師，都可以找到許多有用的參考資料和各科測驗的樣子。在這裏，祇能在藝術、音樂、機械畫、家政和木工等科的範圍內，略舉數例而已。

新式測驗應用在其他各科上的例子

正誤法

(藝術)

十一 綠的顏色是由黃和藍混合而得的。

十一 紅的補色是橙色。

十一 一種顏色當遠看的時候，似乎淺淡一點。

(音樂)

十一 抄寫樂曲的五線譜，包含五條橫線，四條間隙。

十一 一個音符可以表示音的高低及長短。

十一 登堡 (Tempo) 是拍節加速的意思。

(機械畫)

十一 一個圓柱側面的影是射在牠自己的等高線之外。

十一 楷書的斜度應該為 45° 。

十一 正射形是畫在三個平面之上的。

(家政)

十一 富有蛋白質的食物，應該在低溫度中煮之。

十一 乾酪可以代肉。

十一 要使麵包發酵，最好放在清涼的地方。

十一 白洋布是絲線織的。

(木工)

- 十一 銳利的器具是鐵做的。
- 十一 木屑應該用以填塞釘眼。
- 十一 買普通的釘時，常以斤計算。

選答法

(藝術)

- ……同臀至膝的距離相等的是①肩至手②肩至臀③臀至足④頭至手⑤胸至臀。
- ……藍的補色是①青紫色②橙色③紅色④黃色⑤綠色。

(音樂)

- ……一個八分音符是①J②♪③♪④♪⑤♪
- ……△這個記號的意義是①漸強②快板③漸弱④正板⑤慢板。

(家政)

- ……一種富有鐵質的食物是①菠薐②牛奶油③糖④雀麥⑤蜜。
- ……在蕃茄湯裏加一點蘇打為的是①增加香味②保持色澤③加厚濃度④防止過沸⑤調和酸性。

……製樹膠的東西是①動物的體素②穀類③蘋果的果醬④水母⑤橡樹汁。
……夏衣最合宜的質料是①麻②絲③毛④棉⑤革。

(機械畫)

……畫一個圓的時候，應該用①曲線板②圓規③兩腳規④分度器⑤三角板。
……在圖畫板上用丁字規的時候，通常的擺法總是從①上面②下面③左面④右面⑤靠窗的一面。

(木工)

……選擇一條鉋柄，應該看牠的①美觀②堅硬③顏色④光澤⑤勻稱。
……要把一塊木頭磨平，應該用①砂皮紙②金鋼砂的輪子③磨石④鉋刀⑤油砥石。

對偶法

(家政)

……鹹肉 ①可可質

……牛奶油 ②葡萄

……穀類 ③綿羊

……諸果律 ④棉花的種子

……鮮牛酪

⑤ 酸澀的牛奶

……酒石精

⑥ 草的種子

……樹膠

⑦ 小牛

……玉蜀黍粉

⑧ 植物的脂肪

……通心麵

⑨ 甘蔗

……乳部

⑩ 玉蜀黍

……羊肉

⑪ 動物的體素

……人造牛酪

⑫ 麥粉糊

……糖

⑬ 果汁

……茶

⑭ 醬液

……贖肉

⑮ 豬

填字法

(藝術)

①……青紫的顏色是由①……及多量的②……混和而成。

②……紅的補色是③……

(音樂)

①……一個音符可以牠的①……來表示一音歷時之長短。

②……一個音符可以牠的②……來表示音的高低。

③……譜表是一種記號，可用以決定在③……上的音調的

④……④……與

⑤……⑤……。

(家政)

①……良好的肉的顏色是①……。

②……肉腐臭的原因是②……。

③……絲是③……吐的。

(建築畫)

①……窗棂的寬度視①……的厚度而定。

②……兩②……相交處用簷溜。

新式測驗在各種學科上的應用

(木工)

①……橡樹因①……而著名。

②……極薄的材料要用②……鋸。

指認法

(家政)

……尻骨

……脛後部

……臀之後部

……臀之中部

……胸部

……頸部

……前部肋骨

……後部肋骨

……臀之表部

……Ratleran(譯註1)

……腰窩

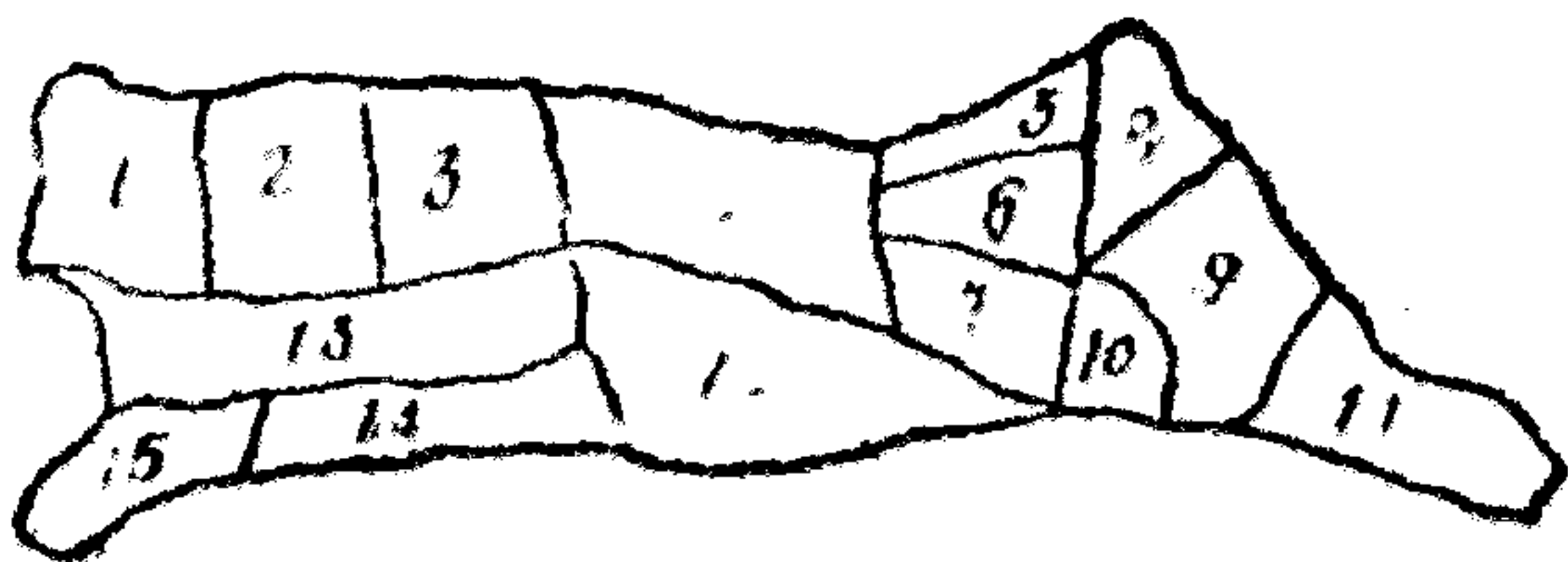
……股肉

……脛前部

……腰肉

重組法

(譯註1) 本字無相當譯名，故暫用原名。



(家政)

烹調時間的長短：

……龍鬚菜(老的)

……豆

……甜菜(老的)

……黃芽菜(嫩的)

……花甘藍

……玉蜀黍

……蕃芋

步驟的次序：

使菠蘿蜜成爲冰凍果汁的步驟：

……過濾

……放在一旁，讓其冷卻

……加菠蘿蜜

新式測驗在各種學科上的應用

……加沸水和糖

……把膠汁浸入冷水內

……使其冷卻

……繼續攪拌，至膠汁融解而後止。

……加檸檬汁

高等學府中的新式測驗——新式測驗應用於高等學府中，同應用於中小學一樣。在大學中，應用新式測驗的試驗，已經不少了。討論新式考試應用在高等文藝或專門課程方面的文字，有價值的頗多。哥倫比亞大學裏，烏特氏的工作，尤為重要。（註一）

溫習與討論問題

(1) 十一？ 新式測驗的適應性很大。

(2) 十一？ 在讀法中，速率比了解難以測驗。

(3) 十一？ 課文中思想的獲得與否，很容易用新式測驗去測量。

(4) 十一？ 文學的欣賞，很可以用正誤法去測量。

(註一) Ben Wood 的 Measurement in Higher Education, p.p. 162 - 63, (1923)

- (5) 十一？文學的欣賞與知識，一定是正的相關。
- (6) 十一？口頭發表的能力，難以很客觀的把他測量出來。
- (7) 十一？作文的能力，最好用舊式測驗去測量。
- (8) 十一？算學的客觀性很高，所以不需要客觀的測驗。
- (9) 十一？新式測驗對於社會科學特別適用。
- (10) 十一？自然科學的考試，早已有客觀的傾向。
- (11) 十一？新式測驗對於測驗外國語的能力，完全適用。
- (12) 十一？舊式測驗對於審美藝術和應用藝術，都能適用。
- (13) 十一？用新式測驗的試驗，在小學中常常舉行。
- (14) 十一？中學和大學的測驗方法，極不相同。
- (15) 十一？有許多文官考試是用新式測驗的。

新
社
考
試

第十一章 測驗分數化爲等第的方法

分數與等第的區別——文字考試的普通缺點，在不能分別分數與等第。一個學生在一個測驗上所得的分數常作爲他的考試等第。在許多教師觀之，一個測驗的記分與分等是同一件事。實際上，分數與等第並不同，故在二者中間，須有清楚的分別。分數是學生在測驗上所得的點分數（Points）。牠代表他的作業或成績。牠隨着測驗的難度與題數而異。倘爲論文式測驗，牠又隨改卷者的寬嚴而異。分數的本身是很少意義的。僅知學生獲得八十分，除了知道八十分以外，不能表示其他的意義。牠不能顯出學生的優劣。欲使分數發生意義，尚須加以解釋。等第是根據某種標準對於分數所下的解釋。對於普通非標準的考試，解釋分數的標準決定於受測驗者的全體的成績。等第是有確定意義的。等第的目的，在藉其他學生的成績以表示某個學生的相對成績，所以牠能表示相對的地位。

分數與等第的混雜，在論文式的考試爲尤甚。有些教師以爲一個客觀測驗應有一百個問題，使其分數可用百分等第來表示。這種謬誤的觀念，係由於假定考試爲成績的絕對測量——百分之零的意思是零分，或對於被測驗的事物表示無能；百分之百的意思是完全的成績。根據這種假定，學生的等第便視爲測驗上所得的分數，或在改卷者的判斷認爲圓滿完成的全部測驗的比例分。爲欲決定成績，有些分數如四十分，五十分，七十

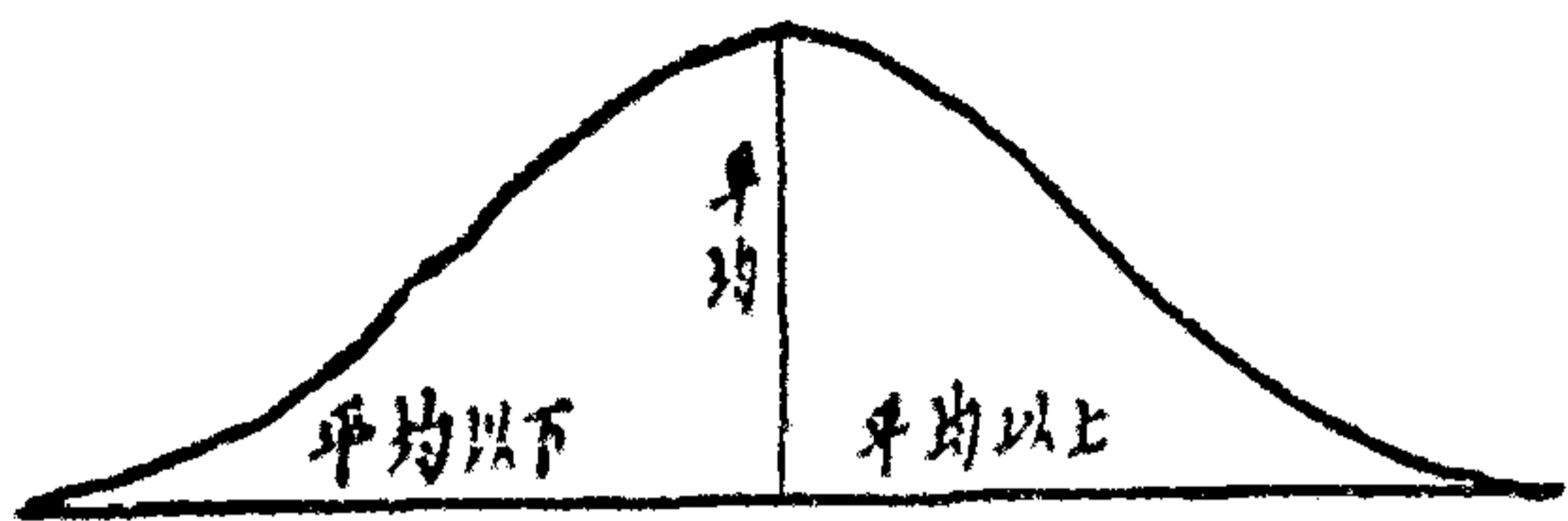
分等，便勉強用作及格等第。這種方法不能真正分別測驗分數與等第。

分數制度——欲化測驗分數為等第，須將通行的分數制度加以考慮。分數是一種判斷的單位，用以記載學生的工作。牠是報告或寫出等第的正式形式，各校所用的符號以及各個等第的距離，差異很大。傳統的百分制度，還是最佔勢力。在百分制度中，各個百分的分數可為 $\frac{1}{10}$ 的倍數，如50, 60, 70, 80, 90, 100；或為 $\frac{1}{5}$ 的倍數如45, 50, 55；餘類推。較小的單位如1, 2, 3等，不過一種假的精確。因為傳統的百分數用絕對的名詞來解釋，故漸有脫離等第判斷的趨向。百分法已漸為文字、字母、或數字的制度所替代。用相對的文字表示成績的各種程度的，如超，優，中，可，劣，或甲，乙，丙，丁，戊。用A, B, C, D, E英文字

母制度的，也日見其多。有些學校用1, 2, 3, 4, 5等數字。無論如何，為一致起見，為使一般人明瞭起見，須將分數化到合於本地的分數制度。本地的分數制度雖須加以考慮，但比較上並不重要。重要的一件事是化分數為等第的方法。等第倘使決定後，將牠化成記分的方法是很簡單的。

常態分配曲線

常態分配曲線的應用——常態分配曲線，(Normal distribution curve)是化分數為等第的一個有用南針。牠能指示在常態或自然的情形下等第的分配常態。根據研究的結果，我們知道人類的特質——包括心理能力與學校成績——大約形成常態



分配曲線。這種常態分配曲線略呈鐘形。近於中等或平均能力的個人，都集中在一處。中間的百分之五十，則密集於平均數。從平均數起，曲線便漸向左右兩端傾斜，表示離開平均數的距離愈增，人數愈減。在平均數上下的相等距離內，所佔的人數均相等。曲線由中間分為兩半，其形態亦相同。在一大羣的常態學生中，等第的分配，應希望其近於離中律 (The law of deviation from an average)

根據常態分配，便能計算應得某等第的學生數的比例或百分數。在理論上，這些比例能由二項展開式 (

$(\frac{1}{2} + \frac{1}{2})^n$ 以決定。其乘幕總是比等級數小一。例如在五分點或五分等制下，各種等第的學生比例，能計算如下：

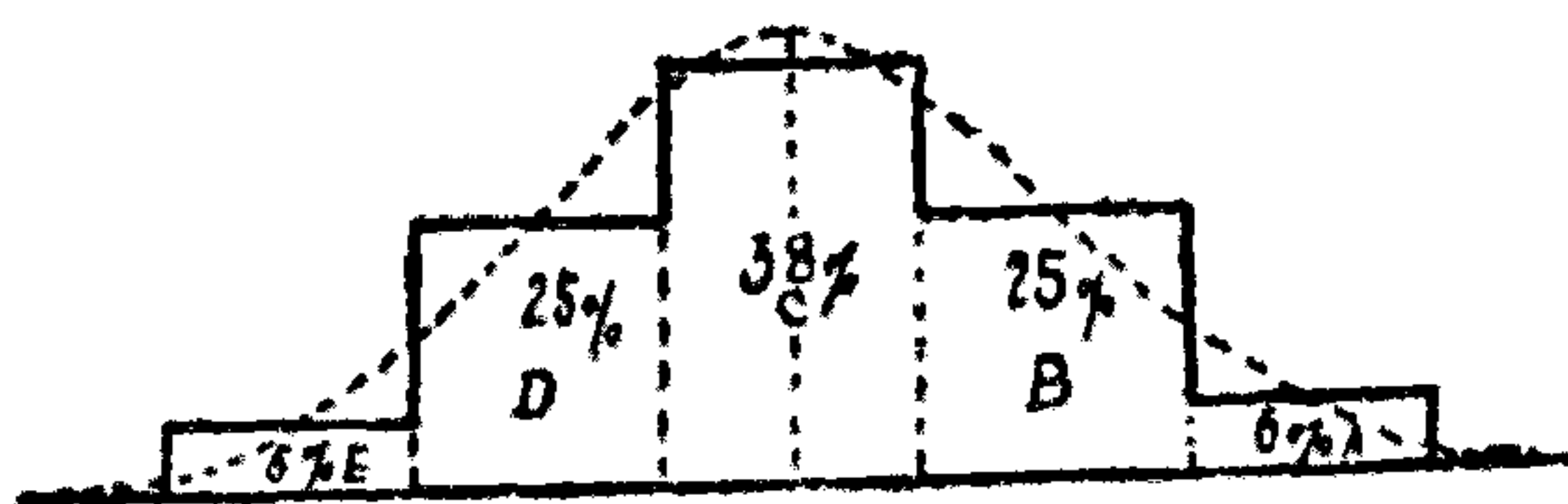
$$\left(\frac{1}{2} + \frac{1}{2}\right)^4 = \frac{1}{16} + \frac{4}{16} + \frac{6}{16} + \frac{4}{16} + \frac{1}{16};$$

$$\text{或 } .0625 + .250 + .3750 + .250 + .0625。$$

將小數以下第三位酌量進位，並以百分表示之，則關於五分點或五分等制，我們能發現有百分之六的學生，應希望其獲得 A, 甲, 1, 超或 100, 等；百分之二十五應得 B, 乙, 2, 優或 80 等；百分之三十八應得 C, 丙, 3, 中或 70 等；百分之二十五應得 D, 丁, 4, 可或 60 等；百分之六應得 E, 戊, 5, 劣或 50 等。上圖是根據常態分配曲線以解釋五分等制的。

四分點或二十五分制的比例，能由展開 $(\frac{1}{2} + \frac{1}{2})^n$ 決定之。如

五分點等第分配



$$\left(\frac{1}{2} + \frac{1}{2}\right)^3 = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8};$$

或 $.125 + .375 + .375 + .125。$

將小數以下第三位酌量進位，並以百分數表示之，則我們知道百分之十二應得1，百分之三十八得2，百分之三十八得3，百分之十二得4。六分點或六分等制的比例，能由展開 $\left(\frac{1}{2} + \frac{1}{2}\right)^5$ 決定之如下：

$$\left(\frac{1}{2} + \frac{1}{2}\right)^5 = \frac{1}{32} + \frac{5}{32} + \frac{10}{32} + \frac{10}{32} + \frac{5}{32} + \frac{1}{32};$$

或 $.03125 + .15625 + .3125 + .3125 + .15625 + .03125。$

將小數以下第三位酌量進位，並以百分數表示之，我們知道百分之三應得1，百分之十六得2，百分之三十一得3，百分之三十一得4，百分之十六得5，百分之三得6。

在實際上，與剛才所計算的理論上的分配，很多差異。最普通的方法是百分之十得A，百分之二十得B。百分之四十得C，百分之二十得D，以及百分之十得E。其他的分配如百分之五得A，百分之十五得B，百分之六十得C，百分之十五得D，或百分之七得A，百分之十八得B，百分之五十得C，百分之十八得D，百分之七得E等等，也有人應用的。有的時候，可用限度，如百分之五至十得A，百分之二十至三十得B，百分之四十至六十得C，百分之二十至三十得D，以及百分之五至十得E。常態分配曲線對於化分數為等第的客觀方法，實為根本的基礎。

次數分配——不論用那種方法去化測驗分數為等第，第一步便是依照分數的大小排列次序，使成一次數分配。(frequency distribution) 這樣便使分數便於運算。從次數分配的性質上，教師能看出測驗本身的優劣，與被測驗者的情形。次數的分配如集中於上端，我們便可知測驗太易，優秀的學生並未予以適宜的測量。次數的分配如集中於下端，我們便可知測驗太難，低劣的學生並未予以適宜的測量。這也足以表示在該團體內，優秀學生或低劣學生有過多的情形。一個團體中能力分配的性質，只能用智力測驗的分數來決定。下面的次數分配，可以指示次數所自產生的測驗的情形：

$X =$ 分數

$f =$ 次數

測驗一

測驗二

測驗三

X	f	X	f	X	f
40.....	I	32.....	IIII	25.....	I
36.....	I	30.....	IIII	23.....	II
32.....	II	27.....	IIII	22.....	I
30.....	I	25.....	IIII	21.....	II
28.....	I	24.....	IIII	20.....	I

測驗分數化為等第的方法

26..... I	23..... III	19..... III
24..... I	22..... I	18..... I
22..... III	20..... I	17..... III
20..... I	17..... I	16..... III
18..... III	14..... I	14..... III
16..... II		9..... III
14..... II		
12..... I		
6..... I		
4..... I		

看了上面的例子，我們知道測驗一是優良的，測驗二是太易，測驗三是太難。教師應依據分數分配的性质，估計測驗的優劣；倘使測驗不能使分數適當的分配於全距離，不應便引為滿意。分數集中的地方大約應在全距離的中間，不應如測驗二與測驗三那樣偏於任何一端。

化測驗分數為等第的方法——由分數計算等第，應採取一種確定的客觀的方法。讀者想還記得，新法考

試的最大優點，在記分合於客觀的標準，以及各個教師對於各個學生所得的分數，意見必相一致。許多教師的記分雖然相同，但只要對各種分數所指定的等第不同，測驗結果的客觀性，便有許多要變得空泛而無足輕重。上面已經講過，等第較測驗分數更為重要，更有意義。學生對於他們的等第的興趣較分數為濃厚。如欲發展與維持測驗之適當的融洽性，須使學生知道決定等第的方法是公平無私的。如欲使等第具有比較的价值，則須有一致的與可靠的方法。由分數計算等第的客觀方法，與決定測驗分數的客觀方法，是同樣的重要。現在要討論兩種化分數為等第的客觀方法；即標準差法（Sigma or Standard deviation method）與比例法（Proportion method）。

標準差法——化測驗分數為等第的最好方法之一為標準差法。凡欲完成考試技術的人，務須熟習此法。標準差法對於沒有統計訓練的人，起初雖似複雜而混雜；但實際極易學習。就是小學的學生，也容易學會，且能知其合理與公平。教師熟習以後，自能熱心應用。

在教育統計學上，標準差是常用到的名詞。正如均數代表一點或一值，為各分數向其集中，而標準差則代表離開均數之一距離，為各分數向其分配，標準差實是一種最好的離中趨勢測量法。牠又名為Sigma，因為牠最常用的符號是一個希臘字母 σ 。這個字母與英文中S相當。簡寫的SD也是常用的。算術均數（Arithmetic mean）或簡單平均數（simple average）的求法，係以次數除分數的總和。其公式如下：

$$M = \frac{\sum X}{N}$$

公式中M代表均數，X代表分數，N代表次數。

標準差的求法，係把每個分數與均數之差平方起來，求出這些平方的總和，再依次數除之；然後將其結果開方。其公式如下：

$$SD = \sqrt{\frac{\sum D^2}{N}}$$

公式中SD代表標準差，d代表每個分數與均數之差，N代表次數。在實際上，簡捷的方法可以先假定了一個均數，再從這個假定均數上着手計算，而加以必要的校正。這個方法較用真正均數(Real mean)與許多小數來計算，要簡單些。從假定均數(Guessed Mean)以計算真正均數校的正公式如下：

$$M = GM + \frac{\sum D}{N}$$

公式中GM代表假定均數。至於自假定均數計算標準差的公式則為：

$$SD = \sqrt{\frac{\sum D^2}{N} - \left(\frac{\sum D}{N}\right)^2}$$

各種符號的意義，與以前各個公式相同，惟d係代表一個分數與假定均數之差。假定愈適當——即假定均數愈接近真正均數——則計算時的數目愈小。

標準差法舉例——標準差法可以以 1935 年 1 頁上測驗一的分數的分配為例。這個分配是六年級算術正誤測驗的結果。均數假定為 20。在計算方面，X 行爲各種測驗分數，f 行爲造成各種分數的學生數，d 行爲各種分數與假定均數之差。若那個分數比假定均數大 15，故置 +15 於 d 行以代之。若那個分數比假定均數小 15，故置 -15 於 d 行以代之。其他的分數也是如此。f d 行爲 f 行的數與 d 行的數之積。得 50 分的只有一個學生，所以 f 行的 1 乘 d 行的 +15 得 +15，即 f d 行；得 50 分的有兩個學生，所以 f 行的 2 乘 d 行的 0 得 0，即 f d 行；其他的分數依此類推。

12 行爲 d 行的數之平方乘 f 行的數。d 行中 +15 的平方爲 225，乘以 f 行的 1 得 225，即 12 行；d 行中 -15 的平方爲 225，乘以 f 行的 2 得 450，即 12 行；餘依此類推。求 Σd^2 的數，尚有一比較簡便的方法，即以 d 行的數，乘 f d 行的數。例如 f d 行中的 -15 乘 d 行中的 -15 得 +225，結果與上面計算方法相同。

$$\Sigma f b \quad \Sigma fd \quad \Sigma fd^2 \quad M = GM + \frac{\Sigma d}{N} = 22 + \frac{-14}{23}$$

$$40 \quad 1 \quad + 18 \quad + 18 \quad 324$$

$$36 \quad 1 \quad + 14 \quad + 14 \quad 196 \quad = 22 - .609 = 21.4 \text{ 或 } 21$$

$$32 \quad 2 \quad + 10 \quad + 20 \quad 200$$

$$M = 21$$

$$30 \quad 1 \quad + 8 \quad + 8 \quad 64$$

測驗分數化爲等第的方法

28.	1	+	6	+	6	36
26	1	+	4	+	4	16
24	1	+	2	+	2	4

$$SD = \sqrt{\frac{\sum d^2}{N} - \left(\frac{\sum d}{N}\right)^2}$$

GM22 4 0 + 72

22	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
						<u>23(N)</u> - 86 1572 ($\sum d^2$)

$$= \sqrt{\frac{1572}{23} - \left(\frac{-14}{23}\right)^2}$$

$$= \sqrt{68.2 - (-.609)^2} = \sqrt{68.2 - .36}$$

$$= \sqrt{67.84} = 8.29 \text{ 或 } 8$$

$$SD = 8$$

$$\frac{+72}{-14(\sum d)}$$

這裏須注意， d 行與 f 行中所有的數大於 GM 者為正，其前冠以 $(+)$ 號；小於 GM 者為負，其前冠以 $(-)$ 號；在 f 行內，正數的和為 $+72$ ，負數的和為 -86 ，正負兩數相加為 -14 ，即 d 的總和 $(M2)$ ，或者說與假定均數 $(Q1)$ 的差數的總和 Q_2 的總和或者說與假定均數的平方差的總和為 156 。以各項的數字代入上列各公式的種種符號，得真正均數為 21.4 ，標準差為 8.29 ，若取其整數，則為 21 與 8 。

化分數為等第——均數與標準差決定後，分數便很易化成等第。取均數 21 為等第 C 或平均等第的中點，取標準差 8 為各等第中所包含的連續分數的個數。

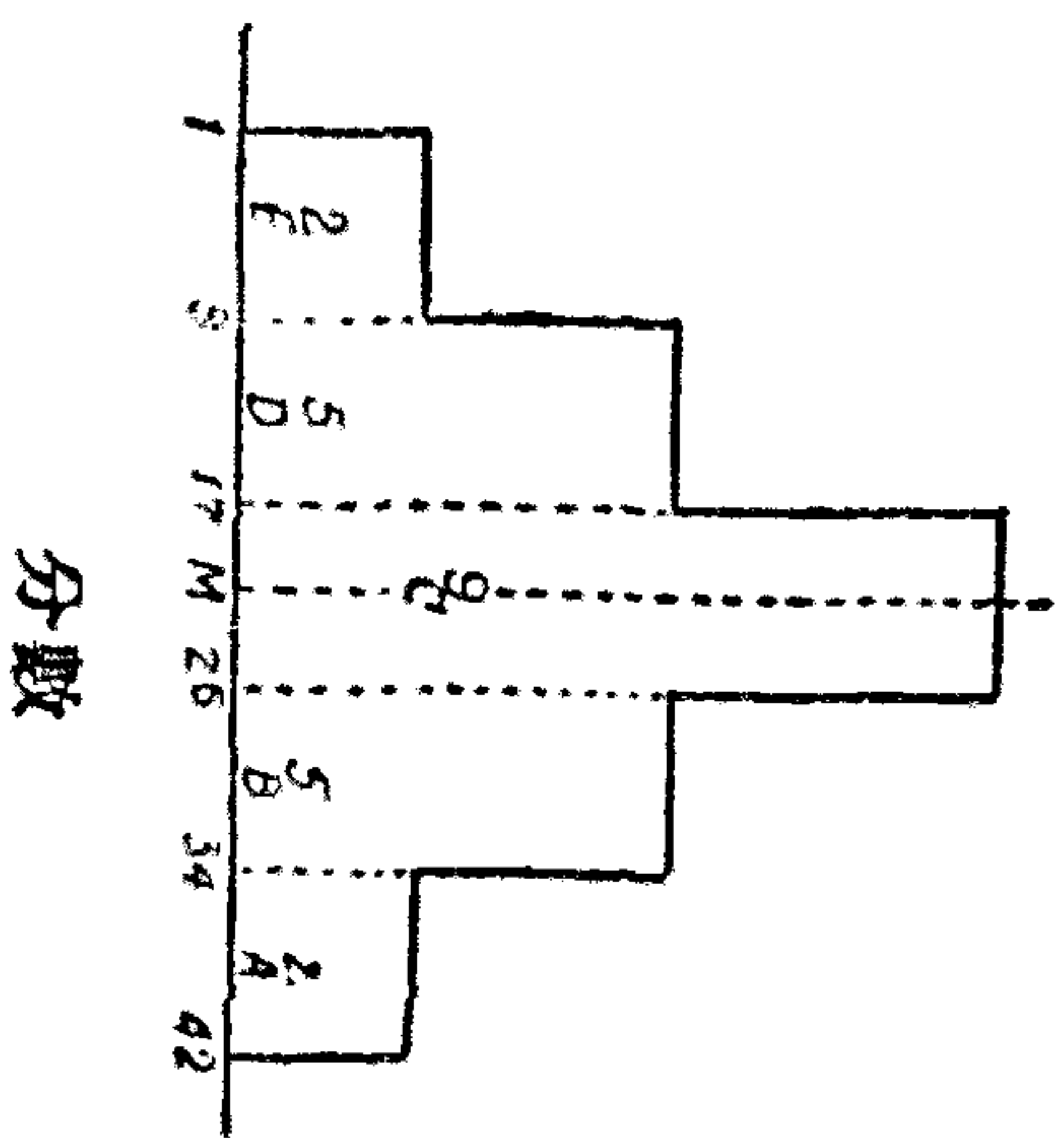
等第 C 從均數 21 下分延長至其上 4 分，如下： $17, 18, 19, 20, 21, 22, 23, 24, 25$ 。當標準差為偶數時，在等第 C 上總要加一分數，使均數為中點。所以在這個 C 的等第內有九個分數。

等第 B 包括其次的八個分數，第一個數比等第 C 的末一數大 1 ，為： $26, 27, 28, 29, 30, 31, 32, 33$ ；等第 A 包括再其次的八個分數，為： $34, 35, 36, 37, 38, 39, 40, 41$ ；等第 D 則包括適在等第 C 下的八個分數，為： $9, 10, 11, 12, 13, 14, 15, 16$ ；等第 E 包括適在等第 D 下的八個分數，為： $1, 2, 3, 4, 5, 6, 7, 8$ 。所以根據這個算術測驗，可得各個等第如下：

分數 · 等第 每個等第的學生數

34—41	A	2
26—33	B	5
17—25	C	9
9—16	D	5
1—8	E	2
		23

M=21
SD=8



此種方法對於學生與教師均極公正。分等的理由是很充分的。每個等第中的分數距離是相同的。B在C上，距離與D在C下相同；A在C上，距離與E在C下相同。有的時候，學生所得的分數無一能遠超平均數而得到A，所以B便為最高的等第。同樣也可沒有分數剛好低至獲得E的地步。全體學生的測驗成績可以決定平均等第或等第C的數值，也可決定每個等第所含分數的限度。結果，各個等第的分界並不致成爲主觀的事了。自一個測驗中所得的等第的分配，不一定常與這裏的結果同樣成爲適宜的比例。但測驗如果優良，對於一般團體的學生，各個等第能得一完善的比配。

分數的分組計算——關於均數與標準差的計算，教師應已知其有藉於分數的分組。(Grouping of

(Scores) 各分數的距離如果很大，分組之法更爲重要。分數倘不分組，計算的數字便很大，並很難統制。分了組以後，每個數便變得很小，計算起來簡單得多。結果不但還算迅速，並且精確可靠。一個初中的社會科學班正誤測驗所得分數的分配，可爲分組方法的例子。各項分數，以及每個分數內所佔的學生，列表如下：

X	f	X	f	X	f	X	f	X	f	X	f
105.....1		95.....5		77.....3		67.....5		56.....1		41.....1	
103.....1		87.....2		75.....4		65.....4		55.....3		37.....1	
101.....1		85.....5		74.....4		64.....1		53.....6		33.....3	
97.....1		83.....4		73.....3		62.....1		51.....2		31.....2	
95.....2		81.....1		71.....4		61.....3		49.....1		27.....1	
94.....1		80.....1		69.....2		59.....8		47.....2		24.....1	
93.....1		79.....3		68.....1		57.....4		45.....2		22.....1	

每個分數倘如 227 頁那樣分別處理，則計算上列分配的均數與標準差，既困難，又費時。但如將各項分數分了組，計算起來便容易而簡單了。每個分配最好不少於 15 組，也不多於 30 組。先求得各分數的距離，以 15 除之，以所得商數代表各組內的組距。這樣，組數便能在 15 至 30 這個距離以內了。分組最好用奇數，使其中一個分數

爲全距的中點。倘使商數爲偶數，則適在其下的奇數應作爲各組內的組距數。分數距離的求法，係從所得的最高分數減去最低分數。例如在上面的分配中，自100（最高分數）減去92（最低分數），則98便爲距離。以98除93，使得商數9.9，或進爲10。倘以此數用作每組內分數的個數，則這個分配便分成10組。以下第二個最低奇數爲96，倘以此數用作各組的組距，便分成10組。在此種情形之下，所有分數可以96個爲一組，或10個爲一組，分配的組數仍在以上所述的96至98的範圍以內。

照普通的習慣，在分組時，以組數的倍數作爲每組的中點，例如在現在這個題目裏面，如以10爲組距時，第一組的分數爲102, 103, 104, 105, 106, 107, 108，因爲105爲10的倍數；第二組的分數爲95, 96, 97, 98, 99, 100, 101；餘類推，直至所有分數完全包括在內爲止。在這裏須注意，d行中各數是連續的。這些數代表包含假定均數的一組與其他各組之差。這些組都成連續的次序。假定均數是所在組內的中點。在這個實例中，假定均數爲90。分組的校正數是以分數的組數乘均數公式中d之總和與求得的標準差。其他方面解法與101頁中的問題相同。

以7爲組距

f

d

fd

fd²

i = 組距

$$102-108 \quad 2 + 5 + 10 \quad 50 \quad M = GM + \left(\frac{\sum fd}{N} \right) i$$

95-101	4	+	4	+	16	64	$= 70 + \left(\frac{-44}{103} \right) 7$
88-94	7	+	8	+	21	63	$= 70 - 2.99 = 67$
81-87	12	+	2	+	24	48	$M = 67$
74-80	15	+	1	+	15	15	

GM67-73 15 0 + 86

60-66	9	-	1	-	9	9	$SD = i \sqrt{\frac{\sum d^2}{N} - \left(\frac{\sum d}{N} \right)^2}$
53-59	24	-	2	-	48	96	$= 7 \sqrt{\frac{726}{103} - \left(\frac{-44}{103} \right)^2}$
46-52	3	-	3	-	9	27	$= 7 \sqrt{7.05 - .18}$
39-45	3	-	4	-	12	48	$= 7 \sqrt{6.97}$
32-38	4	-	5	-	20	100	$= 7 \times 2.64$
25-31	3	-	6	-	18	108	$= 18.48 \text{ 或 } 18$
18-24	2	-	7	-	14	98	$SD = 18$
	103(N)				130	726($\sum d^2$)	
					+	86	
					-	44($\sum d$)	

測驗分數化為等第的方法

以67為均數，18為標準差，這個社會科學的測驗可得等第如下：

分數 等第 每個等第的學生數 $M=67$

95-112 A 6 $\sigma=18$

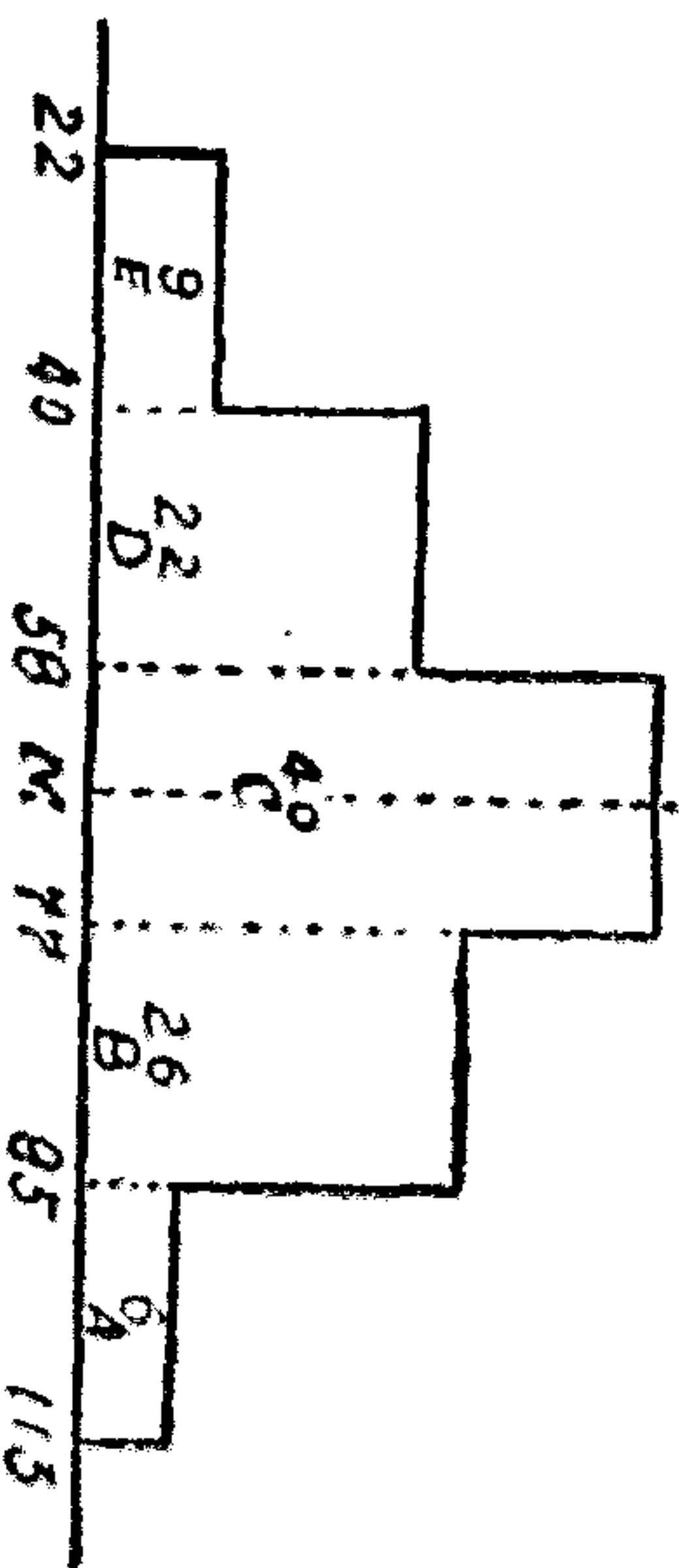
77-94 B 26

58-76 C 40

40-57 D 22

22-39 E 9

103



分數

倘欲應用正負數而為更精細的分等，這些等第的每組可各以六個分數，分為三個小組，如下頁所示。

如無絕對的零點以為計算學生成績的起點，則均數與標準差便為分各分數為等第組數最有用的原始點。均數與標準差實是一種最穩妥的參照點，以化分數為等第而變成有意義，並因而詮釋與說明其意義。等第計算之可能的貢獻應使有普遍的利用。

分數	等第	每個等第的學生數
107-112	A+	0
95-112 { 101-106	A	2
95-100	A-	3
89-94	B+	7
77-94 { 83-88	B	11
77-82	B-	8
71-76	C+	15
58-76 { 64-70	C	13
58-63	C-	12
52-57	D+	14
40-57 { 46-51	D	5
40-45	D-	3

測驗分數化為等第的方法

34—39	E+	7
28—33	E	6
22—27	E-	$\frac{3}{103}$

比例法——化分數為等第的比例法，應用較為普遍。用了這個方法，各個等第可依照百分標準求出，如二頁所述的便是。倘使採五分等第的分配 6—25—38—25—6 為標準，則學生中最高的百分之六在測驗上得 A；其次的百分之二十五得 B；再次的百分之三十八得 C；再次的百分之二十五得 D；最低的百分之六得 E。這個方法可以下面五年階歷史簡答測驗的分數分配為例。

X	F	
36	1	
34	2	3A
33	3	
32	2	
31	2	7B

$32 \times \frac{6}{100} = 1.9$	或 2A
$32 \times \frac{25}{100} = 8.0$	或 8B

29	4	$32 \times \frac{38}{100} = 12.2$ 或 12C	
27	1	$32 \times \frac{25}{100} = 8.0$ 或 8D	$32 \times \frac{6}{100} = 1.9$ 或 2A
26	2		$32 \times \frac{25}{100} = 8.0$ 或 8B
25	3	$32 \times \frac{6}{100} = 1.9$ 或 2E	

24 3 13C

23 2

22 1 分數 等第

20 1 34, 35, 36.....A

17 1 31, 32, 33 B

14 1 24, 25, 26, 27, 28, 29, 30.....C

8 1 7D 16, 17, 18, 19, 20, 21, 22, 23)

6 2 2E 8, 9, 10, 11, 12, 13, 14, 15 }.....D

32 1, 6, 7.....E

測驗分數化為等第的方法

24 3 13C 倘使30這個分數因無人得到，便將其勉強加

23 2 於C組，等第C的分數距離便為24—30，或

22 1 7分。

20 1 分數 等第 學生

17 1 5D 38—44 A 0

14 1 1E 31—37 B 10

8 1 17—23 D 5

6 2 3E— 10—16 E 1

32 3-11 E— 3

這個後來指定的等第，雖不如前述的比例法那樣對稱分配，但與實際情形或較適合。關於這些相同的測驗結果，標準差法得 0A, 10B, 16C, 3D與3E。這個測驗似乎太易，不能分別 A B 兩等的學生。

成績分等的相對次序——欲依照百分的分配標準分一級為五個等第組，教師不得不下許多主觀的判斷。學級也許太小，也許不十分常態，不能嚴格應用標準。測驗也許不能顯示學生方面所存在的種種差異。我們

不能在事前很可靠的預測有多少學生測驗一定失敗，一個也不會多，一個也不會少；並也不能預測獲得A的學生數相同。百分標準是對於一個團體的推測，而非實在的情形。所以牠也不過是教師評判的指導，却不可機械的依附牠。欲將記分不客觀的論文式考試或口述考試分成等第，牠便有特殊的價值。教師就成績的相對次序，而將論文式考卷或學生的努力分別等第，較給牠們以一個具體的數值來得簡單可靠。學生等第排好以後，分等的方法便可舉例如下：

學生姓名	測驗等次	等第
黃琪	1	A
趙俊庭	2	A
6%		
王其昌	3	A*
陳學廣	4	B
張濠傳	5	B
余頌年	6	B
駱霞	7	B

A的總數=3

王其昌的測驗成績較近於趙俊庭，而較遠於陳學廣，故給以A，而不給以B。

B的總數=7

(附註一) 本表除學生姓名與原著不同外其他均無稍異。

周家倫	8	B
周允文	9	B
吳述伯	10	B
25% 胡景文	11	C*
鮑純一	12 /	C
張爲之	13	C
程瑞芳	14	C
陶信芳	15	C
許法祖	16	C
姜良規	17	C
陳 玠	18	C
陸元亮	19	C
王大倫	20	C
林綬昌	21	C

胡景文的成績極近鮑純一，稍遜於吳述伯，故列入C等，而不列入B等。

C的總數=17

測驗分數化爲等第的方法

趙燕	22	C
曹宜春	23	C
濮漪	24	C

28% 吳聿良 25 C

張勉之 26 C*

劉本漢 27 C* 吳聿良,張勉之與劉本漢三生的成績大約相同,但較郭景祺稍優。張勉之,劉本漢與吳聿良均列入C組。

郭景祺 28 D 入C組。

于啓庸 29 D

徐平江 30 D

凌頌聲 31 D D的總數=6

艾振鏞 32 D

芮秀容 33 D

25% 朱文 34 E

丁紹先 35 E

朱文的成績並不遠勝於丁紹先,而遠遜於芮秀容,所以朱文與丁紹先均列入E。

溫習和討論問題

- (1) 十一？ 測驗分數與測驗等第是兩件不同的事。
- (2) 十一？ 百分等第是錯誤的。
- (3) 十一？ 測驗記號就是測驗分數。
- (4) 十一？ 各校所用的等第符號極爲一致。
- (5) 十一？ 常態分配曲線是替代判斷的。
- (6) 十一？ 分數的次數分配是不甚需要的。
- (7) 十一？ 教師的記分如果一致，測驗的分等也必一致。
- (8) 十一？ 化分數爲等第的標準差法是主觀的。
- (9) 十一？ 均數與中數是相同的。
- (10) 十一？ 學生通常以爲標準差法是不公平的。
- (11) 十一？ 分組法增加教師計算的麻煩。
- (12) 十一？ 每組所包括的各分數的個數須爲偶數。

測驗分數化爲等第的方法

(13) 十一？ 均數與標準差是將分數分成等第組最有用的分點。

(14) 十一？ 比例法在每個等第組中保持同一的分數距離。

(15) 十一？ 以成績的相對次序，去分論文式測驗的等級，是一種優良的方法。

第十一章 測驗的施行問題

測驗的施行問題——測驗的施行問題關係到二方面，一方面是教育行政機關所舉行的考試，一方面是教師在教室內所應用的考試。吾人既了解考試的重要與功用，以及測驗技術的新方法之發展，學校管理在這種工作上更當負責進行。關於考試施行所含的問題，如教育行政機關的測驗計畫，教師參與普通考試工作；教師的局部測驗的容納；考試的次數，時間，與排列；通知與不通知；開書與口述測驗的應用；能力分組在測驗結果上的影響；作弊與無監考制；免試的問題；考試分數在最後等第上的影響；測驗結果應求最大應用等等。本章的目的便在討論這些問題。

教育行政機關舉行的考試——教育行政機關舉行的考試，包括由州、府、市、視學區以及一校區所施行的考試。這種考試的性質與範圍根據施行機關實施時之需要與各種可能的便利。州教育機關早已應用各種考試，使其管帶下的鄉村學校標準化，防止對於州定課程之疏忽，決定八年級制之適宜性，維持法定學校課程的與各校的公認工作以及考驗公立學校的教師。有幾州已將此種責任的大部分委於府教育機關。州府教育行政機關所施行的測驗計畫，因為選任官員的時常更迭，一貫政策的缺乏，以及考試事務沒有專門訓練的人才，便受到很大的障礙。有的時候，常常以舊式測驗——遠遜於進步的教師所熟習的測驗——施於學校，所以

教師不得不使其學生將正常的學校工作暫時停止，而匆匆地去準備州府的陳舊的考試。州府考試所視為必需的，便應為優良考試的模範，不論其普通形式，內容，與特殊目的，都應為教師特殊指導的先驅。

州府視學區與一校區所舉行的測量，乃指導改進學校狀況的重要因子。除了新法考試對於改進教學的有價值的貢獻外，市與府的學校當局，還不得不更注意於教育測量的增加。公共學校的財產捐，雖尚未使人厭惡不堪，但至少已經有些不悅意了。教育上繼續不絕的公共輸將，不能僅恃對於公共學校制度的誠樸的信仰與感情的忠實。我們必須使捐款的人看到足以引起信仰的事實，俾其知道用於學校的金錢是一種優良的投資。僅是學校行政對於學校所抱的信仰或意見，並不能使捐助經費的人深信不疑。學校行政人員一定要正式報告事實，使教育董事部與社會充分明瞭學校中的教育狀況。這些報告必須要根據可靠的測驗結果，用生動的描寫，清楚的解釋，圖式的表示。教育主管機關應採納優良測驗的可能貢獻。

教師參預普通考試工作——教師參預教育行政機關舉行的普通考試的編造與施行，有兩重的價值。在一方面是產生更有用更適宜的考試。在他方面教師自身能因此獲得訓練，增加經驗。教師能有助於決定所欲測驗的成績單元；在最適於考試的諸單元中選擇材料；對於測驗的文字，形式與說明，注意是否在被測驗學生的了解範圍內；供給重要的榜證以解釋測驗結果。教師對於合作的測驗，將益信其公平與優良。凡此種種，均能造成一優良的考試，非普通集中的與武斷的方法所能望及。教師參預了測驗程序，便能熟知測驗各級的較良

方法；廣知所教的各種教材；明瞭學科的各部與他部以及與全體的關係；更能感知各部的比較意義；在合作者的工作中發展深切的了解與濃厚的興趣；以及由團體的接觸與討論中獲得偶然的益處。合作的測驗程序對於職業改進的貢獻至大。

奧利恆 (Orleans) 與西萊 (Sealy) (註1) 曾於1927年之春，在紐約路易斯府第一鄉村視學區 (First Rural Supervisory District of Lewis county) 指導一饒有興趣的合作測驗順序。他們的書 (註一) 中所報告的為一二個教師所參加的測驗設計。他們事業的成功，可以表示合作測驗下所能做的工作；就是在相反的狀況下，亦復如此。教師參加普通考試為合乎實際的提議，係一極明顯的事。

給與教師局部應用測驗的便利——合作的測驗不是一方面的事，乃教師參加普通考試工作的結果。而教師局部應用教室測驗當然亦需行政的協助。學校所用大部分的測驗是教師在本級應用的結果。在這方面，教育行政機關應多貢獻必要的指導與建議，並能供給新法測驗所必需的用品與設備。

一般教師因為問卷的難於複印，常放棄而不舉行測驗。所以在教室測驗方面，對於教師最大的鼓勵或助力，是使測驗的印刷獲得種種便利，以供教室之用。市上所有的印刷機，形式甚多，功用與價格亦不一致。每個學校，應量其經濟能力，設法購置一部，以便應用。

每個高深的研究機關，每個大規模的中等學校與每個完善的小學，如果可能，都應設一測驗部。這個部的

設置應使教師便於前往詢問各種新式測驗的編造。測驗部所做的工作如製版分發適量的卷子；聚集卷子準備學生應用；考完後由答案小冊總計分數，或應用教師供給的標準答案批改卷子；並化分數為等第。這種事情，耗費並不算鉅，因為我們如果觀察一下，便知材料的所費比學生用的論文式考試小冊與卷子並不費多少，測驗的印刷與記分為機械式的事情，可藉書記與學生的勞力為之，故比較不甚費錢，而教師却可節省許多時間與精力，以從事較高的服務；並且這種事情對於改進考試與教學的貢獻，價值無窮。教師應該知道，製版和印材料的費用所耗甚少，而測驗功效却大。

考試的次數時間與排列——考試施行的次數，每隨教師而大異。有些教師對於考試的價值有深切的欣賞。他們大約每一二星期施行一次小考，每月或每六星期施行一次較長的測驗。他們以為對於目前的指定作業，施以簡短的，日常的，或每週的小考，在學生方面，能引起較優良的學習；使學生感覺其進步，供給日常記錄的良好基礎；在小學校裏，每天有了學習的動機，並能改少許多訓育問題。還有些教師不能認識考試的價值，或是他們想避免一個優良考試所必需的勞苦。此種教師除學期考試外，便不施行什麼測驗，有時甚而連學期考試也不舉行。要確定的說測驗應施行多少次數，那是不可能的。牠們的次數須根據牠們的功用，性質，以及當地的種種情形。牠們不可過於重視，也不可忽略。較長考試的施行，最好在教材完成一自然單元的時候，而不必在規

定的時間。

一個考試的長短須根據牠的目的，所用的測驗種類，學生的成熟情形，以及所期望的可靠程度。每星期舉行一二次測驗的教師，通常只測驗五分鐘或十分鐘。這些測驗在單獨的時候，雖都不甚可靠，但累積的結果便可靠了；況且短的測驗，或許也會有一二種能和長的測驗同樣的引起學習。普通長的測驗應當限於一課的時間。在無論那種情形之下，一個考試不應長至二課以上的時間。因為比這樣再長的考試，就摻入耐久的因子，易於失去若干的正確性。在平常的測驗中，應儘量使學生有充分的時間，以完成其工作。如果不是爲了特殊的原因，須以速率爲主要點時，時間上不應有嚴格的限制。全體學生在一課時間內都能切實做完的客觀測驗，大都能夠產生優良的結果。

學期考試的排列，教師間的意見，至爲紛歧。有些教師以爲在每學期的終了，須劃出兩三天來，以兩倍的時間從事考試。此種方法有二種教師贊成，一種是舉行論文式考試而需二課時間的，一種是要騰出一點鐘來批閱測驗及作學期報告的。有些教師贊成在最後一星期仍照原來上課時間，舉行考試，至於其他的日子，仍舊進行平常的工作。倘從後面的方法，須避免考試的集中，這樣，在一日之內，學生不致負擔過重。要做到這層，學期考試的排列大約可以如下：

第一時的課

星期二

或星期三

第二時的課	星期四	或星期五
第三時的課	星期一	或星期二
第四時的課	星期三	或星期四
第五時的課	星期二	或星期五
第六時的課	星期一	或星期四
第七時的課	星期二	或星期三

這樣的排列，可為教師的指導。如有充足的理由，要將其更動，當然不必拘拘於此。

通知與不通知的測驗——比較重要而長些的測驗，通常應該預先通知學生。一個考試的主要價值之一是在激起工作的溫習，以及教材的較大單元之組織。倘不預先通知，此種價值，便要失掉。學生應該知道何時舉行考試，俾獲得一種機會與動機，以從事有價值的溫習。學生喜歡通知的測驗，因為對於牠們可以有特殊的溫習。倘以不通知的簡短的臨時考試，考驗新近的指定作業，也有幾種優點。牠可以引起日常的準備，養成學生的機警，以及顯示每天所做的工作究達何種程度。主要的測驗雖應通知，但有時也須不予通知，而突然給以簡短的測驗。

開書測驗——開書測驗 (the open-book test) 准試學生參考教科書，筆記簿，以及其他必要的材料。

這種測驗雖然有些教師已經表示贊同，但為用依然不廣。牠具有幾點實際的價值。牠能測驗的是迅速找到適宜參考資料的技能，正確選擇並採用相關要點的判斷力，以及藉着參考材料的協助以作清晰思想的能力。不論在校內或校外，欲應付所遇的問題而獲成功，這些才能實是很重要的，故對於牠們應有相當的注意。施行開書測驗有一種方法，與其他測驗同，所有工作均須在教室內做，但得享有應用參考材料的特權。另一種方法是每一二星期將全體的或部分的測驗問題，在考試日期以前給與學生，以便他們得利用閒暇，從事準備，並得在教室以外，博覽羣籍。前一種方法須能敏捷應用參考書，並須會做獨立工作。後一種方法與生活的經驗極相符合。證明測驗與正誤測驗均適用這兩種方法。常用開書測驗雖然太費時間，太覺麻煩，但為培養特殊的才能計，仍須偶然應用。

口述測驗——口述的考試，從前雖曾普遍的應用過，但現在除了估定個人是否適宜較高學位與職業業務外，幾已摒棄不用。這是由於以文字考試為教育測量的方法具有很大的優點所致。將口述考試用於全級，以測驗學業的成績，實有多方面的限制。全級中任何學生所回答的許多問題，機會是很不平均的。他們所須回答的問題，難度極不一致。他們的反應大都靠了所準備的材料偶然倖中。以口述考試為基礎，而判斷學生對於某學科熟習程度，是不可信的。荷勒斯曼在其文字考試優於口述考試的分析中，曾指出此種以及其他各種缺點，第一章中曾經提及。

然而口述測驗也有很充分的價值，值得我們注意。牠能測驗口頭表現的技能，回憶必要知識迅速而正確的能力，以及於頃刻間獲得清晰思路的能力。口述考試與開書測驗一樣，也應偶而應用，因為牠注重特殊的才能。口述測驗的問題應採取論文式，並須預先妥為編製。取材應以使全體學生均能熟悉的重要內容為限，並應採取可以激發思想的形式。每個問題的指與學生，應取隨機的方法。口述測驗的分等，最好由教師根據被測驗學生特殊能力的成績次序。要使全級學生嚴密注意各個同學所答的每個問題，要檢查教師對於各個成績所下的判斷，最好使每個學生根據其對於全體學生相對純熟的印象，一一為之分等。

能力分組在測驗結果上的影響——前章曾述，學生在測驗上的等第，係代表他與同級中其他學生比較所得的相對成績，想讀者還能記得。同樣的分數，等第可以較高，也可以較低，因為牠是根據全級的成績，有了這種成績，才可解釋等第。一個學生，倘在中等程度以下的學級裏，衡以那級的標準，使會獲得高的等第；別一個學生雖得同樣的分數，但因在程度較高的學級裏，故衡以該級的標準，只能獲得低的等第了。在範圍較大的學校裏面，常將單一的學級，根據學生能力的高下，分成幾組。如把測驗分數化成記號以代全部的等第，則所有失敗的人，恐將集中於能力較低的組，而較高的記號亦將為能力較高的組所包辦了。此種分組的目的是在使能力大約相等的學生互相競賽，故相對成績的解釋，須根據同組的學生，而非根據他組的學生。每種能力的組，自己有一個平均數或e等的分數，組內其他的分數，則分配於這個均數的上下。所以各組所得的等第，意義並不相

同。

如何調和測驗結果上能力分組的影響，以使各種等第互相一致，實為教師方面的一個難題。欲達這個目的，可用複記號以適應之；易言之，即以一種記號表示學生在組內的相對成績，另以一種副記號表示他所屬一組的相對能力。例如，假定將一級學生依其能力分成上中下三組，而以 α 、 β 、 γ 代表那三組。於是 $\alpha\alpha$ 就是代表上組的 A 號， $\beta\beta$ 就是代表中組的 A 號， $\gamma\gamma$ 就是代表下組的 A 號。這個原則可以適用於任何分數制度以及任何的能力組數。這種複記號可以表示能力分組，也可以表示等第，故能調和測驗結果上能力分組的影響，而合於各種實用的目的。

作弊與無監考制——作弊在考試中是一個常遇到的問題，可為教師與學生生動的討論資料。每個學校都難保沒有作弊，不過程度上有些差異而已。有的學校，作弊風氣極盛；有的學校，則已大減。有些教師對於這個問題，處置得很輕，或竟置若罔聞；還有些教師則予以嚴重的注意，並用斷然的態度來對付。這個問題最圓滿的解決方法是一件應該鄭重考慮的事。有些教師已採用一種無監考制來應付牠，在此制之下，學生須培養誠實的理想，然後將此種理想與他們的榮譽心陶溶於一爐，而使正確與誠實的意識互相密切一致。用了這種制度，在舉行考試的時候，教師可以不必在教室內。有的時候，每個學生須寫誓約書，以莊嚴地保證考試時不彼此互助。有許多人十分懷疑這種誓約的方法，以為牠對於某種學生，在作弊之外，反而加上了裝假。

在事實上，全部的無監考制能否成功，是很可懷疑的。常有少數學生，其注意獲得滿意的等第，甚於獲得達到等第的方法。他們對於沒有充分預備的考試，很強烈地感到及格的需要與重要，以至視應用誠實的方法以達到那個目的，好像是很少結果的。在這種學生的心目中，只求目的，不擇手段，所以他們根據種種理由，為他們的方法辯護，他們的理由是：考試的本身是不公平的，或不正常的；教師的分等測驗更是不正確的；能以取巧狡黠的方法在考試中混過教師，還算較為聰明；大家既然都以作弊為自衛的方法，便不妨也和其光而同其塵。無監考制沒有注意到人類的劣根性。牠不完全適於全體的學生。有許多學生以為牠每使不誠實的學生，占得便宜；對於誠實的學生，反而極不公平。倘將引起作弊的種種引誘設法移去，使學生難於舞弊，乃為一種較佳的方法。教師於考試時留在教室內察看，排列座位使便於獨立工作，以及應用第五章中所討論的方法以減少作弊，並不是誣蔑學生的誠實。嚴防作弊的教師，較將學生的作弊置若罔聞的教師，往往更易為學生所尊崇，並更為學生所歡迎。

考試的免除——有的時候，一個學校的制度，或一個教師，採用一種方法，只要學生合於某種要求，便可免除考試。免除考試的特權，通常給與學業上達到特殊標準的學生，作為一種獎勵。有時也可以獎勵不缺席與優良的品行。在這種免除考試的觀念中，含有一種假定，就是免除考試的可能性，能夠刺激學生為更大的努力，並獲得更優良的日常結果。

有幾個研究家（註一）曾研究免考制（exemption system）在學業上的影響。他們覺得免考制在實質上並不影響所做工作的性質或分量，而且學業上假定的進步每較真正的進步為明顯。他們發現許多教師在免考制下有一種很大的趨勢，就是要想加多免考的等第，勉強將低的等第化成高的等第，而於學業上顯著的進步，並不注意。這種高分數的批給，也許為的是嘉獎某種學生，或避免多改考卷之勞。結果，在等第的分配中，免考各等第所集合的地位，足以破壞分配的常態。

免考制除了這種不好的影響外，還把考試自身的功用，看得狹隘而不完全，牠將考試視為應該省去的東西。牠忽視考試對於學生的各種積極的教育價值。以優良的學業為根據，而免除各種考試，常常限制少數優良學生日常工作的動機，允許他們在考試的時候空耗其寶貴的光陰，並剝奪其行將臨頭的考試所引起的溫習與重組教材的刺激。至於以不缺席與行為做根據而免除考試，格外不合科學，所以更可批評了。就第二章中所討論的考試功用觀之，免除考試很難證明其價值。

考試分數對於最後等第的影響——最後的等第應當代表學生在科學中相對的地位。每個問題對於這種等第所有的影響，應以對於這個問題的自信，作為某學科中學生成績的測量。第四章中所討論的許多主觀的不相關的因子，往往影響最後的等第。有幾種因子如行為、態度、勤懇、進步，以及其他類似的情形，在最後等第的計算，並無地位。有些教師因學生缺席而減低其等第，這種方法不但不足以刺激學生的出席，並且足以抹煞

最後等第的原意。這種因子具有間接的影響，足以牽涉學生的學科成績。關於這方面可以不必多加考慮了。有幾種更有關係的因子，如學期報告，論文，筆記，以及實驗等，是否在不必需的時候，就應該加入最後等第的計算，這都是很可疑的。因為這種問題，學生單獨做的工作到底有多少，從他人所得的幫助又有多少，是很難決定的。因為這種工作難於給以可靠的等第，所以我們以為教師應以和悅的態度，把牠們視為絕對必需的事，而不予以等第。這些工作如果沒有完全做好，只應在最後的等第上做一未完的記號，而不應加以其他的影響。

有些教師對於最後等第的決定，極注重日常的工作。他們於平常復習的記分，極為注意。但這種分數能否正確可靠，殊屬可疑。他們多少總有些根據悅人的聲音笑貌，以及無關係的個人的因子，而不必都根據真正的學問與成績。在小的學級中，討論的機會較多，教師對於學生的認識，也較為深切，則於日常工作的記分，自能比較可信。但學級一大，便不能以日常工作為重要的因子了。況且級中的工作，其目的是為教學與學習，並非為測量。過於重視日常復習的記分，便足以摧殘學級討論中所極需要的自然與自發。學期考試，平常測驗，以及各種

(註1) C. J. Anderson, "Is the Exemption System Worth While?" 載於 *School and Society* Vol. 3p.p.357-60(19

16年3月4日)

H. Kimmel, "Exemption from Examinations and grades" 載於 *School and Society* Vol. 8, p.p. 112-14

(1918年7月27日)

小考，乃為決定最後等第最有價值的因子。在大的學級中，只有牠們才是可以深信不疑的惟一分子。客觀的平常測驗與學期考試的結果，能供給學業成績最可靠的證據。牠們對於每種學程的最後等第，均應有極重大的影響。許多教師已經覺得，把每個測驗的分數總加起來，並任其遞加成最後的等第，是極圓滿的方法。

測驗結果應求最大效用——本書所時刻注意到的是明白為什麼要施行測驗，以及注意充分利用測驗的結果。學校當局負有一種特殊的責任，須注意將各種測驗結果，不論是教育行政機關所主持的，或學校教師所施行的，為最大的應用。測驗分數在行政上的幾種應用，本章前面已經討論過。牠們的惟一的價值，在將學校制度的地位，顯示於教育董事部與社會之前，藉以表示教育上所獲的進步，以及證明學校經費與範圍有擴充的必要等等。所有行政上的應用，僅在直接或間接的增進學生較優良的教育機會。所以牠們的最大價值乃得自教學視導中的直接應用。改進教學與學習的最佳方法，可自測驗結果中得之。視導與測驗有同等的功用，均能增進教學的效率。在實際上，牠們應當互相攜手。以測驗結果為根據的視導，才是積極的、建設的。牠能開闢科學的門徑，以從事有益的討論、建議與實驗。倘使不去應用測驗的結果，視導僅為一種捉迷藏的方法，縱然無可反對，可是有時毫無價值。測驗結果的充分應用，在可能時，須注重在學校視導方面的應用。第二章中所討論的優良考試的發展及其功用的清楚觀念，足以增加學校管理的責任，這種責任就是注意測驗結果的充分應用。

溫習與練習問題

測驗的施行問題

- (1) 十一？ 有許多州府的考試都是舊式的。
- (2) 十一？ 教育行政機關在測驗上的責任是在逐漸增加。
- (3) 十一？ 教師參預測驗，會產生不好的考試。
- (4) 十一？ 教師參預測驗，能獲得有用的訓練。
- (5) 十一？ 奧利恆與西萊，覺得教師參預測驗，是很切實際的。
- (6) 十一？ 合作的測驗是一件相互的事情。
- (7) 十一？ 印刷機是學校必需的設備。
- (8) 十一？ 新式的學校應設一測驗部。
- (9) 十一？ 測驗的次數每隨教師而大異。
- (10) 十一？ 考試應在規定的時間舉行。
- (11) 十一？ 長的測驗通例應限於一課的時間。
- (12) 十一？ 重要的測驗應預先幾日通知。
- (13) 十一？ 不通知的測驗易於刺激日當的準備。
- (14) 十一？ 開書測驗是耗費時間的。

- (15) 十一？口述測驗如偶而應用，價值很大。
- (16) 十一？能力分組在測驗結果上的影響是無足輕重的。
- (17) 十一？考試上的作弊是一個沒有解決的問題。
- (18) 十一？無監考制已經成效卓著。
- (19) 十一？通常有少數學生應該免除考試。
- (20) 十一？日常的復習在最後分數中所佔的成分較測驗為重要。



物版出新局書中正

變態心理學	現代人口問題	最近歐洲政治史	中國土地新方案	最近政治思想史	唯生論	國學入門	教育研究法	英國與其殖民地	掀天動地的蘇俄革命	水滸傳與中國社會	興國英雄加富爾	中國小說的起源及其演變	兒童音樂故事	丙康	風箏
雷學燾	阿桑峯	袁道豐	殷震夏	薛品源	陳立夫	蔣梅笙	朱智賢	姚定慶	陳樂橋	陶孟武	王開基	胡鑄琛	宋壽昌	梅心如	陳澤風
精裝大洋二元五角	精裝大洋二元五角	精裝大洋二元五角	精裝大洋二元五角	精裝大洋二元五角	甲種實價六角 乙種實價四角	九角	一元四角	一元	八角	六角	九角	四角	二角	一元六角	四角

川遊漫記	現代外交家傳記	近代各國外交政策	中國今日之邊疆問題	軍縮戰債賠款	世界軍備	山地行軍	戰鬥綱要表解	積極防空	期待	毋寧死	恨世者	波蘭的故事	偏見集
陳友琴	周子亞	周鯨生	袁道豐	史無弓	鄧陸夫	彭萊父	周鏡民	王平陵	方于六	趙少侯	鍾憲民	梁實秋	
四角五分	三角	五角	三角五分	四角	五角五分	一角五分	七角	一元二角	四角五分	六角	六角五分	六角	

新生活叢書已出二十種
每種實價一角五分

版權所有

(89)

中華民國二十四年六月初版

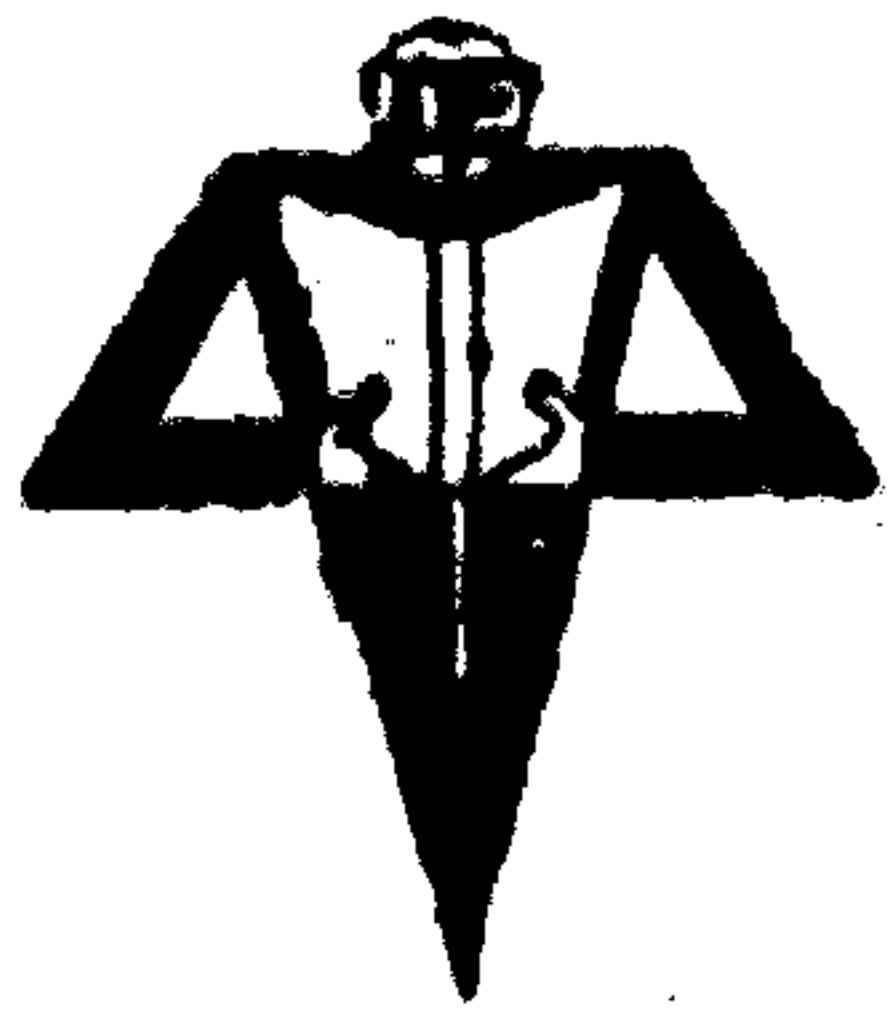
新法考試

全一册定價銀壹元
(外埠酌加運費郵費)

原 著 者
編 譯 者
校 訂 者
發 行 人
印 刷 所
發 行 所

南	正	正	吳	夏	王	黃	浦	Albert R. Lans
京	中	中	秉	湛	明	宗	人	
太	書	書						
平								
路	局	局	常	初	林	宗	人	

0.85
7/10



新法考試：全一册定價銀一元

