

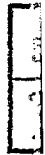
教育叢著

第八種

教育雜誌社編輯

測驗之學理的研究

上海商務印書館發行



520.8
210
311

測驗之學理的研究



3 0537 7130 3

~~000102~~
000082



六	教
週	育
年	雜
彙	誌
刊	十

目次

智力測驗之論爭與教育學說	(一)
測驗之種類	(一九)
心理測驗	(三五)
智力測驗法的標準	(五五)
智力測驗法的研究	(六九)
教育測驗	(八一)

821.3
357

測驗之心理學之研究

測驗之學理的研究

智力測驗之論爭與教育學說

孟憲承

裴葛萊 (W. C. Bagley) 在哥倫比亞師範院教授中，算一位比較的「舊派」的學者。他在談論裏，對於美國學校過重智力測驗的趨勢，早有微辭。本年二月，他公然在芝加哥大學教育教師協會宣讀一篇論文，題為教育的定命論或德謨克拉斯與西奧·T·O. 這文一發表，激怒了多少心理學專家，惹起了幾篇嚴重的抗辯，與教育學說，很有關係。原文冗長繁衍，這裏單就文中要義，提出幾條。

(一)他說智力測驗運動過盛的趨勢，把學校看作一種甄別的，檢定的機關，而把教育反似輕視；又根據兒童智力的限制，要將他們教育的機會，也限制之。他援引許多文學來作證，如：

美國教育研究雜誌裏說：

「高的智力，和藍眼睛，高鼻子一般，是與生俱生的。我們的學校，無意中，一年一年，淘汰掉智力比較薄弱的兒童，而保留那些穎才，能得到智慧上的成就的。所以學校是教育機關，也是一個甄別機關；那入過學校的兒童，可說他們受過教育了，也可以說他們受過檢定了。」

推孟 (Terman) 說：

① Bagley: Educational Determinism, or Democracy and the I. Q. Educational Administration and Supervision, Vol. VIII, No. 5, School and Society, April 8, 1922.

「在事實上，兒童能受教育的限度，可用智力測驗在入學的第一年測定。測驗幾次，到入學的第五年或第六年末，可以精確地找出這個限度。這時候職業訓練和職業指導，就應當開始。」

英國一個教育雜誌說：

「據盧思克 (Rusk) 教授的調查，英國兒童，百分之七十，都不能發達到十四歲以上的智力，所以他們的繼續的教育，是虛耗的。其餘百分之三十中，能在大學受優越學位的，只有四個……將來如將智力測驗，施行於三歲至十四歲的兒童，可以找出其中那幾個是能受高級教育的；把愚笨的汰去，就是給天才兒的機會增多，這是教育上一大經濟。」

(一) 裴葛萊 先要問，那「一般智力」(general intelligence) 究竟是什麼？我們現在把他比之藍眼睛、高鼻子，而不知眼和鼻是生理的結構，至於智力，不是結構，而是機能。二十餘年前的舊說，是以爲智力高下，視大腦灰白質裏細胞的多少

的。後來我們知道智力差異的生理基本，在神經接合狀況的不同。晚近又知道神經質和接合點的發達，也視神經系以外幾種筋絡與腺的刺激變化。所以智力不但是生理結構，并且還不是某項結構的直接機能。說者乃比之藍眼與高鼻，太簡單了。

(二)智力的性質，科學還沒有定義，如何去測驗呢？不要緊，哥爾文(Colvin)在本年全國教育研究會年刊裏說：「我們從不直接測驗天賦的智慧；我們不過測驗獲得的智慧。却於一團有相同經驗的兒童中，測出他們能力的差異，因而推斷他們稟賦的差異。」所以智力測驗的假定，是那被測驗的人，對於一種測驗材料，所有環境、經驗、教育、刺激，都相同的。實則智力愈高，經驗的差異愈多，測驗的正確就愈少。現在最簡單地要測驗有若干智力與絕無智力的差別，上項假定最可適用。到智力年齡漸高，這假定的可適用性漸減。這時測驗，自然仍能測定某項能力，或某類能力，但是其中經驗的成分多了，他的影響大了。我們若以爲所測驗的，

仍是一個與生俱生，毫沒變換的東西。叫做智力，就大錯了。如何可以執着這樣一個得數，來斷定各個兒童能受教育的限度呢？

(四)然而推孟顯然要依據這種得數，來限制大多數兒童的普通教育於五六年内。今假定兒童到了一個限度，智力停了，不再生長，豈便能說，一切教育，也停了，不再發生效力嗎？定命論者自然說，否，否，不過智慧的教育，確停止了。裴葛萊於此有兩行辯論：(甲)智力生長，是無限的，所以教育效力也無限。他說智力生長有二種。從嬰兒誕生起，智力自然的向上的增加力量直到假定的智力年齡限度。同時這智力也忙着在那裏獲得經驗。那自然的生長即停止了，難道他繼續獲得經驗的一種生長也停了不成？如不停，我們不叫他智力的生長，叫他什麼？這種生長，我們可稱他「橫的生長」；那自然的生長，就可稱爲「縱的生長」。即如定命論者所說，縱的生長是有限的，他們也應說橫的生長是無限的。(乙)再說智慧的教育，確有限度一層，且問所謂智慧的教育，是否指抽象觀念底教育？若是指抽象觀念，

那末在橫的生長內，智力不能繼續的獲得他嗎？任指一個平民——智力是平常或平常以下一點的——在他日常生活中，處理他所熟悉的問題，不繼續增加他的能力嗎？

(五)我們對於智力生長的限度，既未確定，而就拿這假定的限度，來作為「檢定」兒童的標準，來判決十二歲兒童的教育機會。平凡的兒童，就打上一個永久低劣的圖記，說他不能受更廣的教育，只須狹窄地學習一種職業，將來做些機械的事務。把較高的教育權利，留給天才的兒童，以為有功於教育的經濟和社會的幸福了。這樣辦，裴葛萊以為不惟無功於社會幸福，而大有罪於德謨克拉西。人常說，為種族的強健計，那疲癯殘疾的人，是應當淘汰的；醫學者不聽其死亡，而用技術去延續他們的生命，似乎是違反種族的利益的。然而醫學者從不這樣想，他們忠於自己的理想，努力自己的學術，今日結果怎樣呢？平均生命年度延長了，人生的疾苦慘痛減少了。教育者不應忠於他的理想改進他的藝術，來提高一般兒童

的智慧嗎？倒憑着 I. O. 來做淘汰的標準，減縮平民教育的機會，還說有功於社會幸福呢！講到平民主義，定命論者豈是反平民主義者？他們說平民主義，只在教育機會的平衡 (equality of opportunity)。我們只要給每人以適宜的機會，使能盡量發展他固有的稟賦，這個發展了做個工匠，那個做個實業領袖；這個做個書記生，那個做個政治家；恰如其分的發展各人稟賦，就是平民教育了。不想一德謨克拉西的要求，不但在各人職業的效率，而在提高一般程度，到最高集合的思想感情，去參與平民政治的生活。現在憑着 I. O. 來擇將來的領袖，而用最高的教育機會來訓練他們，且莫說智慧闊闊的流毒，比貴胄閱閱資產閱閱，還要利害；也慢說智慧領袖，也須有平民的教育，來誘導他們的同情和理想；單說這些領袖，要保其忠實負責任，要不要平民有知識去選舉他們，批評他們，督促他們？（智力測驗，施行於軍隊，最著成效。我們要記得兵士對於官長，是沒有這選舉批評督促的職能的。）況且平民的領袖，又不限於智慧一種品質，那道德的社會的品質

如同情、勇敢、堅忍、勤勞等，亦領袖所不可缺。富於這些品質的人，智力比較差一點，也不是絕沒有做領袖的可能，就可斷定他智力限度，而不予以較高的教育機會嗎？

(二)總之定命論者給我們一個消極的哲學，只告訴我們教育不能做這樣那樣，却想不到教育的確能做這樣那樣。那有平常智力或比平常智力較低一點的男女生徒，的確能學習幾種極複雜的技能，如讀法、書法、計算等；的確能學習幾種抽象的原理，如地球繞日而行的，納稅而無代議是暴虐的等。而且許多原理，在往代要頂高的智力纔能了解的，現代已成爲平民的公物了；在往代只有牛頓、達爾文能領會的，現代「匹夫匹婦」都懂了。學校不能增加原始稟賦的容量，却能增加那獲得的智慧的總積。有史以來，人類的原始智力，並沒有顯著的增高——有許多理由，使我們這樣相信的——然這六七千年中，人類的文化確進步了。然則教育已往所成就，還不過他將來所能成就的極微的朕兆罷了。教育者不應奮興，不

應努力，而反失望嗎？

二

這樣公開的攻擊，智力測驗派的心理學者，那裏會沒有猛烈的答辯？韋迫爾教授（G. M. Whipple）跟着就在學校與社會週報裏，發表一篇批評裴葛萊主張的論文。^①韋迫爾說，第一，我們先要問事實是怎麼樣？第二，要問我們所有的事實，能否徵實裴葛萊那種哲學？據他的意見，裴葛萊把所有的事實，盡行誤解；他的哲學，更是含糊。

先說事實，心理學者對於智力的性質，確乎沒有定論，但是因此便不能測驗他的表象嗎？物理學者，何嘗知道光與電的性質，然而可以測驗他們的現象。現在智

① Whipple: Educational Determining; A Discussion of Prof. Bagley's Address at

Chicago, School and Society, June 3, 1922.

力測驗裏，也不過測定行爲的幾方面，而從此推攷那一種東西——謂之「智力」——的程度。除了這智力以外，自然還有許多原素，能影響兒童學習的進步的。但是我們深信我們的智力測驗，實能精確地測出學習進步上最重要的一個原素。至於智力測驗，在教育上已奏的成績，更不待辯。最明瞭的，學校生徒的分級分組，較有效率了，也有伸縮了。說智力測驗派心理學者，要減縮學兒的教育機會，不過是一種辯術上的危詞聳聽罷了。

說到哲學，韋迫爾說他自己不是一個哲學者，但也有幾項明白直截的信條：1、他相信智力有根本的，且比較永久的差異，已經證明。2、他相信這種差異，大半是遺傳的。3、他相信智力測驗——正當地編製、施行、解釋的——是測定這種差異的一個有價值的方法。4、他相信對於智力差異的生徒，施以同樣教材，同樣教法，同樣速率，的確虛糜許多時間、金錢和勞力。5、他相信在美國現在教育制度下，天才兒所得的教育機會，確比平凡兒爲少；必須施行甄別，方有平衡。6、他相信教育

機會的平衡，確為德謨克拉西真精神的保障。

裴葛萊原文裏，引及盧思克教授 (R. R. Rusk) 的話，又惱了這位蘇格蘭的測驗家，也連忙在十月裏的學校與社會上發表他的辨正。

盧思克說，^①裴葛萊怕智力測驗，危及平民教育，證明他把智力測驗或平民教育誤解了，或把兩者都誤解了。平民教育所要求的，是兒童不論貴賤，不論貧富，能得到他所最能獲益的教育；而智力測驗，正所以幫助教育達到這個目標。平民教育的意思，決不是說，每個低能兒，都應受大學教育的。現在我們選擇中等學校的生徒，除普通學科試驗外，又施行智力測驗，實是要推廣選擇的範圍，使那以前沒有最好的教育機會的一種兒童，這時有以表見他們的能力——這總不是反德

^①Rusk: Dr. Bagley and Educational Determinism, School and Society, October

14, 1922.

謨克拉西的手續吧！英國智力測驗家，從沒有說兒童十二歲以上，可以沒有更高的教育，他們只主張那教育要有益於兒童自身和社會；所以要甄別他們的智力，智力較低的給他們一種實用的教育。至於修學期限一問題，本來豈智力測驗家所能解決，這全要看一國社會的——即政治的和經濟的——情形如何而定的。

對於裴葛萊文的反響，上述以外，已發表的文字還有聽說哥倫比亞師範院蓋茲教授 (Gates) 近作一演講，題為心理的柔性 (Psychological Tender-mindedness) 就是批評裴葛萊說的。我沒有見到，無從轉述。

上邊所述，限於敷陳，沒有批判，却也無須批判。裴葛萊對於智力測驗，原不是作籠統的批評，他不過就其中假定的——照他看來——危及平民教育的一項，提出討論。這個分際，他的讀者應看清楚的。他說：「我決沒有意否認智力測驗已有的成績，和他將來能有的更大的成績。他積極的建設的方面，我竭誠歡迎；他如有消極的破壞的地方，他自己應負證明的責任。」韋迫爾駁裴葛萊，也說：「他所說

的，有許多地方，智力測驗家是同意的。」後來裴葛萊答韋迫爾的文裏，^①自謂有最堅確的兩點：（一）智力度愈高，測驗的正確愈少；（二）智力有縱的生長，有橫的生長。這兩點韋迫爾均沒有提到，而他所謂「含糊」的哲學，就建設在這兩點上。至於這兩點究竟能否成立，既尙待專家的論定，我們這裏也不能臆斷了。

三

我對於這個論爭，加以注意，注意了又爲之傳述，自然也不是因爲他於智力測驗自身，有何重大的影響；却是因爲他代表教育學說上兩個見地是我們所要思考的。

盧思克說：「裴葛萊對於智力測驗的攻擊，不過是關係教育效力有限無限

① Bagley: Educational Determinism Again; A Rejoinder to Prof. Whipple's Reply,

School and Society, August 6, 1922.

的老爭論中的一章。」就是這句話了；這不是一個重要的爭論嗎？亞丹斯教授 (Adams) 在他的教育學說進化論裏有一段話：「對於海巴脫 (Herbart) 學說最普通的一個批評，就是他給人一種教育的狂熱……然而這種狂熱，不是沒有可以尊重和贊許的所在。現代教育學說中最缺少的，就是一點希望的原素。我們給遺傳和環境的勢力說圍住，對於教育的事業，失望了，灰心了。那教育家對於自己事業，已沒有了信仰，而只覺其無能，更從何盼望他們有偉大的成就？」^①所以教育效力問題，值得透澈的討論。

從來解答這個問題有 Nature vs-nurture (稟賦與教養) 的二派。在現代學說中，倫敦大學 訥恩教授 (T. P. Nunn) 爲題了兩個名詞：前者他叫他高爾登派，後者他叫他新海巴脫派 (Galtonians vs. Neo-Herbertians)。^②分述如下：

^① Adams: *The Evolution of Educational Theory*, p. 326.

^② Nunn: *Education, Its Date and First Principles*, Chap. IX.

高爾登派的學者，以爲各個人教育的限度，早在各人遺傳的稟質上決定不移了。借裴葛萊的名詞，這就是一種「定命論」。這派採取的辯證，有兩大根據。第一，根據心靈與身體相關。他們檢察心靈的品質，和身體的品質，有很高的相關率，而推考他們是同一起原的。身體品質是遺傳的；心靈品質，也不能從教養產出。第二，根據嚴密的遺傳研究。高爾登遺傳的天才的結論中，第一條就是說原始能力高的人，就是有社會階級的障礙，也容易向上的。他又研究兩類的雙生兒；一類稟質相同而環境異的，一類稟質相異而環境同的。如前一類雙生兒中相對的差異增加，後一類中，他們的類似增加，就可證明環境的影響了。而所得結果，均反是。其他重要研究，桑戴克 (Thorndike) 教育心理學三卷十一十三兩章，敘得頗詳。桑戴克自己調查過紐約五十對的雙生兒，他的結論，也說這些雙生兒原始稟質的類似或差異，全由於種子細胞的性質。他們的身心，都受這種子細胞的影響，至於環境上的差異，如紐約學校之所有的，影響於他們的身心者很少。①依據這種結論，

我們對於教育的效力，不由不失望；而人羣改進的責任，只可付之優生學了。

以那樣謹嚴的科學方法，得那樣精確的事實，還是不能折服新海巴脫派。他們的中堅，爲英國海瓦特博士(F. H. Hayward)著教育與遺傳的鬼魅(Education and the Heredity Spectre)一書。他大膽地說，心靈的遺傳，如一個鬼魅，只要我們不拘着抽象的統計，而追求到人生具體的事實上去，他就消滅不見了。海巴脫的哲學，本來以爲心靈裏只有獲得的觀念(Vorstellungen)，沒有先天的智力。這些觀念，全從經驗和社會生活得來，全可受教育的支配；教育功能，所以無限。這種舊說，海瓦特自然也不盡取，他至少承認心靈也自有他傳遺的稟質。但是這些稟質，在幼時很柔黏的(plastic)，可以拿教育來陶鑄的。所以稟賦極劣的兒童，經過教育機關如 Barnado Homes，可以得到很健全的發育。甚至劣下的民族，如 Murray 島人，不會計算六個以上的數目，而他們的兒童，受了蘇格蘭教師的訓

●Thorndike: Educational Psychology, Vol. III, Chap. XI.

練，竟表顯很滿意的數學能力。這種例多不勝舉，新海巴脫派所以說，教育效力是無限的。

如今裴葛萊的批評智力測驗，我看就帶些新海巴脫派的色彩。平心而論，教育者對於己身的事業，誠能得到一個信仰的福音，一種樂觀的哲學，豈不甚善？無如「梓匠輪輿，能與人規矩，不能使人巧」是格言也是常識。「教育只能引導與輔助，不能創造。天賦所沒有的，教育不能供給他；天賦所有的，教育也不能完全摧折他。」這理羅孫克蘭茲（Rosenkranz）三十年前已闡發了。①個性遺傳的差異，教育可能的限制，也已經科學的事實確切證明了。藥石雖苦，不能不嘗；真理雖苦，不能否認。現代科學的詔示，往往使人生呈現一種黯淡的灰色，原不獨教育上為然。雖如詹姆斯說，各人學說，因氣質的剛柔而異其趨向。然一方有真理，而一方違着真理去求慰安，究竟是思想的不徹底。

① Rosenkranz: *Philosophy of Education*, p. 47.

然而韋伯爾也承認的，美國學校的濫用智力測驗，有時也是愛之適以害之，所以裴葛萊之論，多少給他們一些針砭。同時或者也給那悲觀過度的教育者一些奮興。這種言論，在今日美國教育界，諒也不是沒有一部分同情的反應。

附言——這文脫稿後，又見 *The New Republic* 週報（一九二二年十月至十一月各號）載 *Walter Lippmann*, "The Reliability of Intelligence Tests" 一文。雖是一個“*Layman*”的通俗的論述和批評。却和 *Lippmann* 他文一樣，寫得警動有趣，值得一閱。此外專家的論文，近來見到的，還有 *Educational Review*（十一月分）所載哥爾文教授（*Dr. S. Colvin*）的“*The Present Status of Intelligence Testing*”亦明瞭，亦翔實，我們應給他鄭重介紹於讀者的。

測驗之種類

解中蓀

測驗的產生並不久，但是推行却很廣，全世界都有了他的蹤跡，聲價更一天一天高起來！現在中國教育界也起首注意測驗了。不過測驗的種類很多，範圍也很廣，很容易令人淆混；若是起初對於他全體的概念沒有明瞭，以後恐怕要常常發生誤解。這篇文字把測驗的種類略說一說，給讀者一個大體的概念；至關於測驗詳細的內容，這裏暫且不說。

測驗是什麼？是求出人類能力（普通及特殊）的數量，應用到實際生活上去。譬如說某人手腕的運動，敏捷到一種什麼數量；測出以後，以為選擇事業，支配人才的根據。這就是測驗重要的用意。測驗形式雖有不同，用意總相差不遠。

測驗與純粹心理學試驗不同。純粹心理學試驗，是要求出種種心理現象的因果，以為學理的根據；如試驗動物能否辨色，習慣的養成和保存究以何種作用為

最要，本能怎樣發達，怎樣退化等等。這都是研究心理的現象，不以實用為目的。測驗屬於應用心理學範圍，多以實用為依歸。用個比喻來說：純粹心理學試驗好像純粹化學的試驗，只求原子化合等關係；測驗好像工業化學，要試出幾多綠氣最宜造成幾多漂白粉。在英文中，關於前一類的，名曰 *psychological experiment* or *laboratory work* 等；關於後一類的，名曰 *tests*。（測驗）

測驗有兩種方向：一種是測量人的普通智慧；如經過某種測驗，知道甲的普通能力比乙高些，張家孩子的智慧在中人以上，李家孩子在中人以下等等。一種是測量人的特殊能力；如王兒有圖畫的能力，可以學圖畫；趙兒計數敏捷，可以學會計等等。但是，普通智慧是什麼？測驗普通智慧是測驗什麼東西？起初智慧測驗行於歐洲的時候，心理學家西旦（*Stein*）解釋普通智慧，就是普通適應力。後來測驗傳到美洲。美洲有些心理學家覺得「適應力」太含混，沒有內容，不能作為解說。他們承認普通智慧乃心理各種作用的總和，是一種數的概念（*mathematic*

conception) 完全屬於實用的問題。行爲成功的便是智慧高，行爲失敗的便是智慧低。這樣說來，所謂「普通」也不成爲普通，不過各種特殊能力集合的數的概念罷了。

測驗的種類雖很繁複，分起來可得四大類：(一)智慧測驗(intelligence tests)，(二)心理測驗(psychological tests)，(三)教育測驗(educational tests) (四)職業測驗(vocational tests)。現在把四種的內容略略說明於下：

一 智慧測驗

智慧測驗所測量的，便是所謂普通能力；完全是測驗先天的遺傳。人類的個性不同，有的很聰明，有的很笨，這是最明顯的事實。但是某甲的天生智慧究竟在人類中占一個什麼地位？用什麼標準來判斷他的高下？在從前心理學未昌明的時代，多半是迷信的判斷；如中國的算命，什麼干支配什麼干支便是聰明，什麼干支

配什麼干支便是愚笨；又如看相，把眼睛的大小，手指的長短，當作判別智愚的根據。西洋的骨相學（Phrenology）也差不多和看相一樣。現在心理學發達，智慧測驗也連帶發生。根據心理學的所得，定出測驗的方法，立下數的標準；確如尺寸一樣，可以用來丈量人類天生的智慧了。

關於智慧測驗的詳細內容，另篇再和讀者討論；這裏舉出一種最有價值的測驗，作為代表的實例；就是皮奈西門（Binet Simon）的測驗。皮奈氏的方法是以年齡做標準。從三歲起，到十七及十八歲，每歲有幾個測驗，通過一個測驗，便算幾個月智齡；然後再以智齡總數比實足年齡，求出智慧比例數（intelligence quotient）。智慧比例數愈高，就是智慧愈高；譬如有個兒童，他的年齡是六歲另一個月。他的智齡是八歲另五個月，那嗎，他智慧比例數就是以六歲一個月（即七十二個月）除八歲五個月（一百零一個月）得一三八。這是很聰明了。據推孟（Terman）的標準，一百四十以上為才智出衆的人，一百二十至一百四十為特別

聰明的人一百一十至一百二十爲聰明的人九十至一百一十爲普通的中才八十至九十爲愚笨的人七十至八十爲特別愚笨的人七十以下爲低能兒。所以一三八是特別聰明的了。(以上所述，是經推孟 Terman 修正以後的算法。)

皮奈氏測驗的用法，只限於個人，就是一個主試同時只能試驗一個被試。使用的方法，要照他所說明的做，(中文的說明有南高印的皮奈西門智力試驗法說明書和景海女學的皮奈氏智力試驗法，英文的說明有 The Measurement of Intelligence) 不能隨意增減字句；因爲這是有一定的標準，字句變換，結果便要受影響，標準便不能適用了。而且試的時候，別的條件愈同一愈好。試驗的地位和周圍的環境，都要單純清靜；不然，便要分被試時的注意，結果就不可靠了。

皮奈氏測驗的內容很複雜，各方面都試到；如理解，記憶，注意等等都包含其中。這是他的特長。前邊說過，所謂普通智慧，不過各種心理作用的總和；這個測驗，能各方面都顧到，結果自然是正確的。

二 心理測驗

心理測驗，測驗什麼東西呢？就是前邊所說各項心理作用。如果我們承認智慧是各項心理作用的總和，那麼這類測驗和智慧測驗沒有截然的分別；不過這類的測驗每種只試驗某種特殊作用，和智慧測驗試驗各方面不同。據心理學研究，每人各項能力的高下，不是一樣的；讀書記憶能力好的人，學習音樂或竟不成功；所以各項能力的特殊測驗，確有他的價值。不過有一點須請讀者注意，就是人的各項心理作用，不是各個孤立的，總是幾項互相關聯；不過有些關係大，有些關係小罷了。譬如令被試者背幾句詩，看來好像專試記憶，其實所試不止於記憶；誦讀時候的注意，對於詩句的了解，聯合上下文意義的方法，都有關係。用心理學上名辭說起來，實在關於記憶、注意、知覺、聯合四種作用。不過記憶方面最重要最明顯，所以通常說是測驗記憶；若是把他當作專測驗單獨的記憶，那就謬誤了。

這類測驗的計算不以年齡為標準，只計算某人在某測驗中得幾分。這個測驗的中數是幾分，某人比中數高幾何，或低幾何。如測驗兒童，每歲立一個標準，就是說七歲兒童中數為幾分，八歲為幾分，九歲為幾分……某兒年八歲，所得分數若干，比其中數高或低；如比中數高，就是說在這個測驗，他的能力高；比中數低，就是能力低。

現在舉幾個例看看。(一)測驗記憶的限度。主試者讀出幾個數目字，令被試者記著，然後照樣背出來。從很少的試起，直到不能記憶為止。測出被試者記憶的限度在什麼地方。據 Smedley 試驗，結果如下：

年 齡	聽的 記憶 限度	視的 記憶 限度
7	5	5
8	5	5
9	5	6
10	6	6
11	6	6
12	6	7
13	6	7
14	6	7
15	6	7
16	6	7
17	7	8
18	6	7
19	7	8

(二)自由聯念測驗。這個測驗要測量聯念的敏捷。通常所謂聰明人，思想很銳利；聯念敏捷，思想便銳利了。測驗的方法，是限定時間，令被試者寫出或說出腦中所想起的字。據 Play 試驗，普通兒童三分鐘可寫出的字數如下表：

寫出字數		年 齡
女	男	
23.7	23.0	8
31.0	26.9	9
32.2	29.7	10
36.8	33.3	11
36.6	34.2	12
38.3	33.9	13
39.1	33.3	14
40.2	40.0	15
40.9	33.3	16
41.6	42.8	17
47.1	48.9	18

假設有個兒童，年齡八歲，他的記憶限度為六，自由聯念能寫出二十八字，便見得這個兒童的這兩種能力比普通兒童高些了。

以上所說是心理測驗的大概。若是承認他是和智慧測驗不能分，便合起來，通稱爲智力測驗也可以。

三 教育測驗

教育測驗是測量教育的結果。未有教育測驗以前，批評一種教育制度，或一個學校的好壞，都是以個人的意見爲斷。個人的意見不是一致的，所以甲判斷以爲好的，乙以爲不好，真實情形，總是不能發見。教育到現在還沒有成爲科學，實在因爲太主觀的緣故。教育測驗出，然後判斷有了客觀標準，與研究教育主持教育者以莫大的利益。中國教育要想改良進步，教育測驗可以做個好幫助，切不要把他輕視了。

我們知道，受了某種教育多少時，這種能力總要發生些影響；但是影響究竟有幾何，非測量不行。從前惟一的測量方法，便是考試；然而考試記分有幾種弊病：（甲）教師記分無一定標準。據 Carter 和 Kelly 的調查，同一學生在同一課程中，兩個教師所記之分數，相差甚遠；一班學生在同一課程，經兩個教師評判，各人

所列次序之前後也大不相同，足見這種考試記分是沒有價值。(乙)題目不能試出真能力，同一課程同一教材，若非同一教師，所出試題必然不同，其中必有許多不能試出真能力；所以結果不可靠。(丙)試題的難易不齊。所出試題未經標準測量過，難易的分量，教師自己也沒把握，不過靠著猜度。以不齊的試題，記同樣分數，自然不能得到真價值。若教師僅僅憑著猜度，定出難易，又是不可靠。因為這三種緣故，所以考試記分法，不能把人所受教育的結果測量出來。教育測驗把各科都立出一個標準，做一支量教育結果的尺。如國民三年級應通過某種讀法測驗；假設不能通過，便足見這國民學校讀法教的不好。這種判斷是很可靠的。因為用同一尺寸來量，這個國民學校的讀法特別比別校短些，自然是不行了。

現在舉幾種教育測驗做實例(一)叩提斯算術測驗(The Courtis Standard Tests for Measuring Arithmetic)法分加減乘除四組。每組有二十四個同難易的試題，每組限定八分鐘，令小學校兒童計算。他的標準如下表：

究研的理學之驗測

八 年 級	七 年 級	六 年 級	五 年 級	四 年 級	三 年 級	二 年 級	一 年 級	年 級	
11.6	10.9	9.8	8.6	7.4				速度	加 法
76	75	73	70	64				正確	
12.9	11.6	10.3	9	7.4				速度	減 法
87	86	85	83	80				正確	
11.5	10.2	9.1	7.5	6.2				速度	乘 法
81	80	78	75	67				正確	
10.7	9.6	8.2	6.1	4.6				速度	除 法
91	90	87	77	57				正確	

這表上就是說國民四年級兒童加法應當做七個多些，其中應當百分之六十一四是正確的；減法也應做七個多些，其中應當百分之八十是正確的；其餘類推。假設有一班四年級生中數加法和減法速度是八，正確是八十，足見得這班算術特別好了。

(二)中國書法 scale (等格)江蘇第一師範附屬小學校所作中國書法的等格，也是教育測驗標準的一種。他們把每種字（楷書大字、楷書小字、行書大字、行書小字）分出若干等格，國民三年級應在第幾格，四年級應在第幾格等等。詳細內容請看他們所出的印刷品小學校第十號。

關於教育測驗的標準，美國差不多各科都有了；中國除去書法等格以外，却沒有看到別種。許多事業，只候人去做，願研究教育者努力！

四 職業測驗

職業測驗，在各種測驗中，產生較遲，現在還在嬰兒時代。

職業測驗要測出某人具有某種特殊能力，合於某種職業；以爲選擇職業，僱用人工的標準。最初人類職業的分配，是一種世襲制，「士之子恆爲士，工之子恆爲工」……無所選擇。如遇著能力和興趣不合的人，也要強所難能。這是現在公認爲不合理的辦法。但是現今通用試行法，實在也不經濟；因爲僱用者和被僱者都沒有把握，完全是試試看。等到試了幾星期或幾個月，果然不合，再換試別種。因此，有許多終身不得相當的職業，真是可惜！這種試行的僱用法，對於僱用者及被僱者，雙方都受損害；實業效率上也受很大的影響。有了職業測驗，人類天才各得適當職業以表現；實業的效率也可以大增。某種職業測驗須包含某種職業必不可少之習慣和能力。這種習慣和能力是這種職業中最有名的人所有；心理學家把這種習慣和能力分析出來，立爲標準。能通過這種測驗，便能適於這種職業；否則，必不適宜，不必候試了許多時纔知道。各種職業須要各種特殊才能，假使天性不

相近，學習是沒有十分功效的。譬如電話局職員，聯接電話，要記憶和注意強，手腕靈敏纔行。所以許多人記憶力弱，注意不易引起，或是手腕遲鈍，要做接電話的事，非失敗不可！據說紐約電話局每人至少每點鐘要接二百個電話，多至六百個，非有這種特殊能力的人，如何能做這種職業？

Minsterberg 的電話生的測驗做一個實例。

Minsterberg 先從三十個接電話最有效率女子中，試得一個測驗標準，後來用以測驗新入局的女子。果然是他們在測驗裏能力高的人，接電話的能力也是高；在測驗裏能力低的，接電話的能力也是低。他測驗四種：(甲)聽的記憶限度。讀數目字給被試者聽，讀完之後，令她寫出來。從四個數到十二個數，試她究竟能記得幾個數。(乙)注意敏捷和正確。限定六分鐘令她把一頁印刷物上的 A 字母劃去。試她究竟能劃去多少，其中正確無誤的有多少。(丙)空間的位置。試她極速的運動，是不是能在正確的位置上。(丁)令她把一堆雜亂的卡片，照著次序理齊。試

她動作敏捷的程度。

最近歐戰中產生了一種新測驗，便是軍事測驗 *Army Tests*。測驗某人配不配做軍官，某人合於炮兵還是合於馬兵。他的試法和職業測驗差不多，也是分出炮兵或馬兵所不可少的能力，作為測驗的標準。因為軍事的性質，在這一方面看起來，和職業一樣，所以可以放在職業測驗裏，不必另立一門。

讀者如要詳細研究，可讀以下各書：

小學校第十號

皮奈西門智力試驗法 皮奈西門智力試驗法說明書

智力測驗法

Whipple: *Manual of Mental and Physical Tests.*

Terman: *The Measurement of Intelligence.*

Monroe, De Vosand Kelly: *Educational Tests and Measurements.*

Minsterberg: *Psychology and Industrial Efficiency.*

Hunter: *General Psychology.*

心理測驗

陳鶴琴講

楊賢江記
華超

測驗學之研究

近數百年來，科學發達的結果，農業上、工業上、商業上、醫業上，無一不用科學方法來處理。譬如治病，從前是用試行錯誤法，病人喫了這種藥方不對，就換別種藥方來喫，這是完全不是科學的。現在則不然了，醫生治病，先求病源。如瘧疾的起因，知爲蚊子所傳染，就把從前種種的猜想，以爲瘧病是由於開窗而睡，或由於晚上乘涼而起的，一概撇去了。再如種田，從前也不知怎樣種法，應種什麼，只是盲目的做去。現在則知考究土質，配合肥料等種種方法了。商業上，有許多商家費了極多的金錢，利用廣告以引起一般人的注意，也是應用科學方法的結果。總之，十九世紀二十世紀是科學發達的時期，在各種事業上都要用科學的方法。現在試問教育上是否也用科學的方法？像我國從前的教三字經讀四書，固然不是科學的。就

是現在換了教科書，也何嘗能適合兒童的心理？不合兒童的心理，便不能說是科學的。又如兒童的智慧，也無鑑別的方法。懸算命看相的，規定一個人是適於習商，一個人是適於讀書，這個豈能可靠？西洋在十八世紀時，有一種叫骨相學的，只看人的頭骨如何，以為可以規定他的運命。其實也是不足相信的。惟今日所欲講的心理測驗，才是科學的方法，在教育上算是最新的利器。不過心理測驗的發達，還只是十幾年以內的事，而在最近的六七年間算最發達。

普通所行考試方法來評定學生成績的，很不正確，可舉一例來證明。美國有一位心理學家為要研究這種考試方法，特拏一本中學生的幾何試卷，請一百十六位教員評定分數。諸君須知幾何試卷，是有得數可憑，批評結果，應得是差不多的。豈知大大不然。有個教員填了九十二分，有個教員只填了二十八分，還有二十個教員所填分數在八十分以上，再有二十個教員所填分數在六十分以下。可知同是一本幾何試卷，請一百十六個教員來看，結果已如此懸殊，則別種沒有得數可

憑的試卷如文學一類，更是不能準確了。所以這樣不準確的，便因為各個人的見解不畫一，沒有一定的標準。就在同一個人，也要因為改卷時的態度及時候，而有差別。總之這種方法，全憑主觀，簡直沒有標準可言。現今所提倡的心理測驗，便是用科學方法，求客觀的標準，以救濟主觀方法的弊病的。以下且講心理測驗的用途。

(一) 心理測驗與教育

(1) 甄別智愚 德莫克拉西教育的真義，是要使人人得到同等的教育機會。不管他是聰明的，是愚笨的，或是富貴的，是貧賤的，總之當有受教育的機會，以發展各人之所長。但是怎樣曉得某兒聰明，某兒愚笨呢？這就要用心理測驗來分辨出來。所以心理測驗可以幫助德莫克拉西教育的實現。

(2) 改良考試 我國近兩年來，常有因考試而起風潮的現象，這個實因為向來所行的考試方法不好的緣故。美國哥倫比亞大學在一九一八年起利用

心理測驗作入學考試，南京高師及附中附小現在也採用這個方法。其他學校也有試行的。我今且說明心理測驗所以能準確及公平的地方，請諸位看我所分發的測驗。

第一爲數目校對 這個試驗在三分鐘以內行之。所以試驗學生反應的快慢及正確不正確的。美國銀行中辦事者及商家司賬者以及校對人都可應用這個試驗。

數目校對

假使第(二)行的數目與第(一)行的數目一樣的，那末你就在第三行中的方格子內作一個記號如「 \surd 」，若不是一樣的，作一個記號如「 \times 」

(1)	(2)	(3)	(1)	(2)	(3)
650	650	<input type="checkbox"/>	10243686	10243586	<input type="checkbox"/>
041	044	<input type="checkbox"/>	659012534	659021354	<input type="checkbox"/>

湖之學理之研究

2579	2579	<input type="checkbox"/>	383172902	381872902	<input type="checkbox"/>
3281	3281	<input type="checkbox"/>	631027594	631027594	<input type="checkbox"/>
55190	55102	<input type="checkbox"/>	2499901354	2499901534	<input type="checkbox"/>
39190	39190	<input type="checkbox"/>	2261059310	2261659310	<input type="checkbox"/>
658049	650349	<input type="checkbox"/>	2911038227	2911038227	<input type="checkbox"/>
3290517	3290517	<input type="checkbox"/>	313377752	313377752	<input type="checkbox"/>
63015991	63019991	<input type="checkbox"/>	1012938567	1012938567	<input type="checkbox"/>
39007106	39007106	<input type="checkbox"/>	7166220988	7162220988	<input type="checkbox"/>
69931087	69931087	<input type="checkbox"/>	3177628449	3177632449	<input type="checkbox"/>
251024818	251004418	<input type="checkbox"/>	468672663	468672663	<input type="checkbox"/>
299056013	299056013	<input type="checkbox"/>	9104529003	9194529003	<input type="checkbox"/>
36015992	36015992	<input type="checkbox"/>	3484657120	3484657210	<input type="checkbox"/>
3910066482	391006482	<input type="checkbox"/>	8588172556	8581722556	<input type="checkbox"/>
3510273301	3510273301	<input type="checkbox"/>	-3120166671	3120166671	<input type="checkbox"/>

263136996	263136996	<input type="checkbox"/>	7611345879	7611345879	<input type="checkbox"/>
451152903	451152903	<input type="checkbox"/>	26557239164	26557239164	<input type="checkbox"/>
2859016275	3295016725	<input type="checkbox"/>	8819002341	8819002341	<input type="checkbox"/>
532039144	532039144	<input type="checkbox"/>	6571018034	6571018034	<input type="checkbox"/>
61558529	61588529	<input type="checkbox"/>	38779762514	38779765214	<input type="checkbox"/>
211915883	219915883	<input type="checkbox"/>	39008126557	39008126657	<input type="checkbox"/>
670413822	670143822	<input type="checkbox"/>	75658100398	75658100398	<input type="checkbox"/>
17298591	17198691	<input type="checkbox"/>	41181900726	41181900726	<input type="checkbox"/>
36482991	36482991	<input type="checkbox"/>	6553920817	6543920871	<input type="checkbox"/>

第二為謬誤測驗 這個試驗在二分鐘內行之，在每個圖裏有一件是錯的，要學生畫去他。譬如第一排第三圖是手套。看他能不能看出多了一隻手指，并應當畫去那一隻。又如未排第二圖，是一個人用力拉一隻狗，要他看出謬誤的地方并且畫去他。

究研的理學之驗測



諸位看了這一種測驗，可以曉得心理測驗並不是一定十分複雜而需要儀器的。不過也應當曉得雖只是這一張紙，而要造成這些材料，却是費了許多精力。外國有一位心理學家用了兩年工夫才做成一個填字測驗，可見真不容易。而所以可貴的，就在做好了應用起來，非常簡便。

第三是詞句重組。這個試驗內，我們故意把一句句子內的字顛倒錯亂，并且還多了一個字，而且這個字也是很含混，一時不容易驟然分辨出來的。我們要試驗學生國文的程度，就看他能否看出那個多了的字。也還可以試驗他的判斷力。

詞句重組

(陳鶴琴)

- 1, 今我有天病明。
- 2, 姓你麼我什?
- 3, 兩有弟弟我個無。

- 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4,
- 玩你中外到想邊去麼。
- 學書報生讀學進校。
- 火比輪車快得船行跑。
- 鷄明啼天就一月。
- 有個洋地大上五球江。
- 面煙對們我當力反應酒竭。
- 多人都是人的我友朋好。
- 報道知可看進識以增。
- 十茶喫鐘中二點飯。
- 見活來高生日近程增度低。
- 你要塗亂請不糊講我。
- 嬉有游人兒嬉應成遊有無也童。

第四是默字測驗。這是試驗學生認識的字，能不能從記憶中寫出來。譬如「我們要試驗學生能不能寫痛苦兩個字，我們先對他解釋說：『我心裏覺得很痛苦，』他懂了，看他能否寫出來。這個測驗裏上邊的數目字，是表示一百個學生中能默出來的人數。如痛苦兩個字，便有八十八人能默出。

默字測驗（高小二用）

（陳鶴琴）

- 88, 1 痛苦 我心裏覺得很——。
- 87, 2 要緊 這樁事情非常——。
- 83, 3 互相 我們應當——幫助，——親愛。
- 80, 4 報酬 他該你做事，你應當——他——，就是報答的意思。
- 79, 5 更動 ——，就是掉換的意思。
- 78, 6 議論 我們大家來——這件事情，——，就是討論的意思。
- 77, 7 牆壁 這個——。

37,	46,	49,	51,	56,	62,	64,	65,	67,	69,	72,	74,
19	18	17	16	15	14	13	12	11	10	9	8
影響	滋養	獎勵	野蠻	殘廢	喧譁	期限	柔軟	兇猛	翻譯	眼鏡	組織
你們很用心讀書，凡百不用功的學生進來，你們就能——他，	雞蛋的——料很多。	他做事很熱心，我們應當——他。	這個小孩子——，不講理，亂罵亂打。	那個兵的手，被仇敵斬斷了，他所以變成——無用了。	你不要——亂擾。——，就是亂鬧亂擾。	——，就是一定的日期；比方，他在此地做功課有一定——的。	——體操。他的身體很——。	這隻狗很——。——，就是利害的意思。	我說英文，你替我——成中文。	帶的——。	我們大家——一個會來研究中文。

感化他。

39, 20 判斷 我們不要——這個人好,那個人不好——,就是批評。

第五是算學測驗 這是試驗學生算學能力,看他能做對多少。美國關於算學測驗,共有三種,加減乘除的試驗,都極詳細,可惜太多了,不能印出來。

算學測驗

(陳鶴琴)

(1) 猪油三百二十錢一斤。問二百個錢,能買猪油多少。

答

(2) 桃子八十錢一斤。問一百二十個錢,有幾斤桃子可買。

答

(3) 四個人合組一網球場,買網球六個,每個價銀大洋五角;網拍四個,每個價銀三元;網一條,價銀一元;問四個人每人應當拿出幾元。

答

(4) 在十分鐘的時間內我能跑四里路現在有十二里路問要跑多少時間

答

(5) 從上海到蘇州約二百華里。火車行二小時即到，輪船每小時的速率，抵當火車的一半，人力車每小時的速率，抵當輪船的五分之一，問輪船和人力車須各行幾小時可到。

答

(6) 甲跑一百二十碼，要十二秒。乙跑一百二十碼，祇要十秒。現在甲乙二人同時賽跑，問乙到一百二十碼終止點的時候，甲跑幾碼了。

答

(7) 問什麼數目自乘之後，所得的數目等於2乘8。

答

(8) 柴要一元能買七百斤，問買四百五十斤柴，要多少錢。但洋價每元作一千

四百文。

答

(9) 有一個學生，把他自己所有的銀錢用掉三分之二之後，現在還剩八元；問當初共有多少。

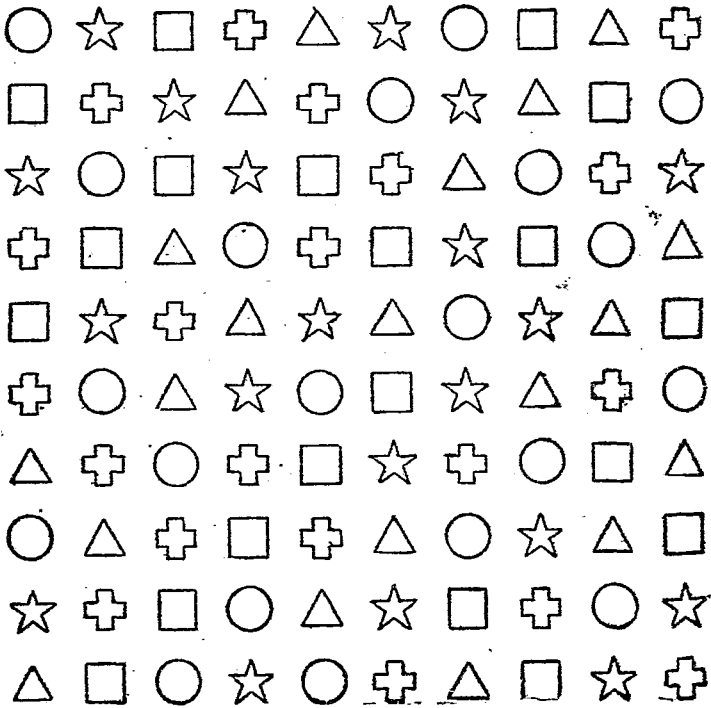
答

(10) 早晨六時有一貨車，從上海開往南京，這部貨車每一點鐘能行六十里。息了二點鐘後，有一部快車，也從上海開往南京，不過每一點鐘，他比貨車快二十里，問幾點鐘始能追上。

答

第六是形式交替 這個試驗，是要學生在下面一百個圖形中，填上相當的數字。譬如見了星形，就填1，見了方形就填3。看他在一分半時間以內能做對多少。

究研的理學之驗測



總之，心理測驗的好處，在畫一不變，今天我試驗得這樣結果，明天依舊這樣。不但我是如此，便是別人用了同樣方法試驗時，也可得到同樣的結果。因為他有了客觀的標準，便可把主觀的影響消滅了。

(3) 分別班次 心理測驗既能分別智愚，所以就依據測驗的結果，作為分別班次的標準。西方學校裏邊，多設立特別班，便為適應這種需要，使聰明的努力發展，愚笨的循序進行，身體殘缺的，也可學得一種本領，自謀生活。

(4) 補助教授 平常教育者不能曉得自己的教法究竟是對不對，有進步沒進步。現在用了心理測驗，也可以曉得了。譬如要考查一種功課，還是由教師讀給學生聽好呢？還是由學生自己看好呢？那就可用這兩法來試驗。假定當初行了讀法，可得七十分，行了看法，可得六十分；但過了半年，行讀法的得了八十五分，而行看法的有九十分。可見讀法只加了十五分，看法則加了三十分，就曉得看法是較好了。還有在學生一方面，也可有法知道自己學習的進步。譬如書

法一項，現在已有了標準。學生只要把自己的書法和那個標準對照，就得知到了什麼程度了。

(5) 分別才能 各人的興趣才能是各各不同的。有的長於音樂，有的長於算術，有的長於辭令，有的長於機巧。我們一定要因材施教，才能把他們的特長發展起來。從前因為沒有分到才能的方法，不知埋沒了多少天才，實在於促進文化上是大有不利的。

(二) 心理測驗與軍事

美國於一九一七年與德奧戰爭時，要在一年之內徵集二百萬兵士，但當初編制軍隊是行試行錯誤法，並沒一定標準，很有許多弊病。到後來政府特召集全國心理學家，用心理測驗來甄別一百七十餘萬的受驗者。試驗的好處，在乎(1)甄別智愚。使智的升遷，愚的淘汰。(2)支配才能。使各得其所，無濫竽的弊病。當年美軍所以能夠戰勝，未始不是心理測驗的功效。

(三) 心理測驗與職業

職業上的心理測驗，美德兩國極爲盛行。美國有個大公司，甄別和升退雇員，均用心理測驗。有幾處軍械局利用心理測驗後，各部做工的效率因之大增。我國工商各業，招收學徒，毫無科學的標準。某夥試了幾次不克勝任，就把他辭去，另換一個。再不對再換，這樣做法，實在太不經濟，我國工商業的所以不發達，這個用人不當也是一種原因。

(四) 心理測驗與社會問題

犯罪爲社會中一個主要問題，應當設法解決。用了心理測驗，就可尋出犯罪的原由，是否因爲智力薄弱不能謀生的緣故。再如娼妓問題，也可用心理測驗來考究婦女流爲娼妓的原因。

(五) 心理測驗與人種學

人類的智力是否相等？黃種與白種究竟那種強？這就可用心理測驗來規定。同

以黃種論，日本人的智力是否比我們強？同以中國人論，南方人是否比北方人強？假使日本果比我們強，南方人也比北方人強，又要問強在什麼地方？這就和心理測驗有關係了。

關於心理測驗的用處，講到這裏為止。美國已有四五種雜誌，專來研究心理測驗，兄弟甚望我國教育界也來加以研究。但有一句話要聲明，就是諸君看了上面幾種測驗，不要以為這很容易得很，便想回去試驗，這是要不得的。因為這是一種科學，非有精密的研究，不能收效。今日的講演，不過引起大家研究的興趣罷了。

記者案：關於心理測驗的專書，這位陳先生和廖茂如先生已經做了一部叫做智力測驗法。列為南京高等師範學校叢書第二種。現在已由商務印書館出版。

智力測驗法的標準

朱光潛

這兩年來，智力測驗法在我國學校裏也逐漸露頭角了。智力測驗原來對於教育是極重要；不過若想得精確的結果，非有縝密周到的方法和堅忍忠實的精神，斷難成功。現在這種方法和精神似乎都還欠缺一點，所以應用智力測驗的人大半祇勉強做到一個「驗」字，至於較困難而且較重要的「測」字都丟在背後去了。「測」是智力測驗法的精華。「驗」的目的也不過是「測」。驗而不測，全是徒勞無補。這篇文章所討論的問題就是：就目前我國學校情形而論，我們應該採用什麼方法去「測」智力？

在未討論方法以前，我們應該對於智力測驗的目的先有一個明瞭觀念。兒童智力發達與年齡增長，通常成正比例。教育應該根據兒童智力發達的順序。我們必定先知道在指定年齡以內，兒童智力到什麼一步階級，實際上能夠學什樣，我

們纔可以教他學什麼。倘若他實際上每日祇能學一課，我們拿成人的眼光來猜度，以爲他可以學兩課，因而就教他兩課，那麼，結果就不免食而不化；反之，他實在每日能學兩課，我們如果以爲他祇可以讀一課，因而就教他讀一課，那麼，結果就不免停滯發展。總之：教育第一步就是要求出兒童的智力程度。智力程度通常固然同年齡相稱；但是有時智力也許比年齡跑得快一點，有時又許慢一點。所以我們如果祇能知道兒童的年齡，決不能斷定他們在這個指定年齡以內，智力已經發達到了什麼程度。智力測驗的目的就是求在指定年齡以內兒童智力已經發達到了什麼程度，并且把這個程度用數量表出。

目的定了，現在來討論達到這個目的所必需的方法。上面說智力發達時，我們用過「程度」「數量」等字。說到程度數量，我們定要有標準單位，例如測長要用尺寸，測重要用斤兩，測溫度要用寒暑表。假使沒有標準單位，我們祇能估得一個大概，不能測出精確結果。譬如拿溫度說，我們如果不用寒暑表，我們祇能說「昨天

很熱，今天很冷。」不能斷定冷到什麼度數，熱到什麼度數。測智力也是一理。我們必定先規定一種標準，纔能配說測。沒有這個標準，我們的眼睛不管什麼亮，我們祇能估，不能測，——祇能說某生聰明，某生笨拙，不能用數目字來表示他們的聰明或笨拙的程度。

舊式考試的目的也有一部分是測驗智力。不過因為沒有普遍的客觀的標準，所以免不了去兩層弊病：（一）出題計分，純以主試者個人意見為標準，所以結果不能精確。例如以同一試卷請甲乙丙三人計分，甲給七十分，乙或給五十分，丙或給九十分。我們究竟當依那一個結果？（二）舊式考試純以比較同學級中各學生程度為目的，所以結果無普遍的價值。例如張生在甲校一年級考得八十分，李生在乙校一年級也考得八十分，他們倆的智力不必實在沒有懸殊。我們所以不信任舊式考試，所以費精力來規劃智力測驗法，就是因為這兩層大缺點。

現在讓我們問一聲：我國現行的智力測驗法比較舊式考試根本好得多遠？

們是估，還是測？就對於這個問題所發表的文字而論，似乎新式測驗和舊式考試沒有大的分別，因為我們還沒有規定一個普遍的客觀標準。先就測驗問題（problems）說，南高陳廖兩先生的智力測驗法裏面僅分「智力測驗容易類」、「智力測驗普通類」、「智力測驗高深類」三種。究竟智力發達的順序是否能如此切作三大段？假使能切作三大段，自幾歲至幾歲纔可用容易類測驗，自幾歲至幾歲纔可用普通類測驗，自幾歲至幾歲纔可用高深類測驗？這些根本問題還沒有解決，我們怎樣配談測驗？從智力測驗法第十四章看，許多測驗都好像有萬能的性質，從國民小學一直用到高等師範。例如二百二十八頁的模型再認，一個測驗，從國民小學三年級用到中學三年級；二百三十六頁的迷津測驗「乙」一個測驗，從國民小學生用到心理專科學生。假使迷津測驗和國民小學合式，拿來測驗心理學專科學生，不太便易了麼？假使迷津測驗和心理專科合式，拿來測驗小學生，不太失公道麼？如果這兩層都使得，那麼，我們可以反對智力是隨年齡發達了。再

進一步說，假定測驗問題都已經依年齡次第分配好了，在指定年齡以內，大多數兒童能夠做成多少？或者拿術語說，某年齡的工作常準（norms of performance）是什麼？沒有這個工作常準，我們用什麼去測兒童智力是否照常規發達？照上面情形看起來，規定智力測驗標準一件事業在教育界要算是當務之急了。現在我們來商議規定的方法。規定標準當分兩層說。第一層是求測驗問題的標準（standardization of tests）第二層是求在指定年齡以內，兒童的工作常準。

一 規定測驗問題標準法

關於規定測驗問題標準一層，法國採年齡制，英國仿之，美國則採用學級制。但無論採年齡制或學級制，目的都是一樣，就是在未以問題測驗兒童以前，先以兒童測驗問題，看看何種問題與何種年齡最相宜。最先主張問題要受測驗者為法人皮奈（Binet）。他的方法很簡單。先以測驗問題考一大羣兒童，這羣兒童中年

齡要參差不齊，自三歲至成人，每歲中都有幾十或幾百人做代表。得了結果以後，察在某指定年齡以內，兒童中能答某問題者之多寡，以定某問題是否與某指定年齡相宜。如果在指定年齡中所有兒童全體能答某問題，則某問題與指定年齡不合，因為太易；如果在指定年齡中所有兒童全體都不能答某問題，則某問題與指定年齡也不相宜，因為太難。反之，如果在指定年齡中一百個兒童中有六七十個人能答某問題，那麼，某問題用來測驗指定年齡以內的兒童，就非常合式了。例如主試者先念一個四位數，姑且說是 5338，叫兒童跟着念。如果四歲兒童中祇有二三十人能念，五歲兒童中有六七十人能念，六歲兒童中有八九十人能念（以上都以每百人計算），那麼，念四位數這一個測驗不能用在四歲，不能用在六歲，祇能用在五歲了。這是年齡制，學級制不過於定標準時拿學級來代替年齡。用在某學級的測驗，定要有某學級中百分之六七十人能做，纔使得。

說到這裏，問題發生起來了。就我國學校現狀而言，我國規定測驗問題的標準，

應該用年齡制，還是學級制？照個人的意見，學級制在我國似乎不大相宜。美國教育已普及了，兒童入學升級通常都依一定年限。同一學級中兒童年齡相差都很微細。所以用學級制既可省事，而根本上又與年齡制沒有大分別。我國兒童入學升級年齡雖也有空文規定過的，實際上同學級中的兒童年齡相差往往很遠。高等小學第一年級的學齡至少也有十歲至十五歲的差別。因為有這層困難，所以純用學級制，似乎結果不能十分精確。這也是我妄想的。我們把南高陳廖兩先生測驗結束拿來比較，就可見得學級制不能無弊。例如謬誤測驗的結果（智力測驗法，二百十二頁）：高小一得 13.5 分，高小二得 15.35 分，高小三得 12.3 分，女中學四得 12.5 分。從這個結果看起來，中學學生的智力不及高小學生的智力，高小三年級學生的智力不及高小一年級的智力。這個結果可不是個奇聞麼？產生這個結果的原因固然很多，恐怕根據學級制是最重要的錯誤的根源罷。

純用學級制既不免無弊，那麼，我們定要用年齡制麼？這又不然。智力測驗與知

識測驗雖然不同，但有密切的關係，因為智力要應用知識。具同等的智力兩兒童中，甲的知識如果強似乙，則受測驗時甲可以利用知識，所以比乙要佔優勢。因此我們測驗智力時也不能不偶爾照顧到知識。知識深淺通常和學級高低成正比。所以遇同等學級中兒童年齡相差甚遠時，純用學級制來規定測驗標準，固然不妥，但把學級完全一筆勾消，也不免發生微細的錯誤。

既不能純用學級制，又不能純用年齡制，有折中的辦法麼？我們第一步要調查各學校中兒童的實在年齡和學級，做一個年齡學級統計表（Age-grade Scale）。在這個表上我們可以看出某學級中以某年齡為最普通。第二步我們就可以拿年齡為根據，用上述皮奈的方法來規定測驗問題標準。這個折中法也不是特創的。去年湯姆生（Thomson）在英國邵譚堡蘭（Southumberland）地方測驗官費生的工作常準，就用這種辦法（見 British Journal of Psychology, Dec., 1921）。我想不但工作常準，就是測驗問題標準也未嘗不可用折中法規定。

現在把邵譚傑蘭的兒童年齡學級統計表譯一節來說明折中法。

年 齡	級	三級以下	三級	四級	五級	六級	七級	七級以上
十一歲半	13	8						
十二歲半	50	15	4					
十二歲半	222	77	8					
	480	238	33					
	255	439	33					
	39	206	45					
	4	23	9					

觀上表我們可以查出某年齡中的兒童實際上大半在某學級。例如十二歲的兒童在第六級者有 439 人，比在他級者都多些，我們就可以說十二歲的兒童實際上大半在第六級。現在我們要求用在十二歲纔合式的測驗問題，怎樣辦呢？皮奈的方法可以借用了。我們和皮奈的方法略差別一點。皮奈所根據的是各級中的十二歲兒童全體。我們祇根據在第六級的 439 個十二歲兒童；因為大部分十二歲兒童都在第六級。我們測驗智力不能不顧及學級上的差別，我們既已用年

齡做標準了，同學級中年齡又有差別，所以不能顧及全體十二歲兒童的學級差別；既不能顧及全體，所以不得已祇顧多數。什麼可以顧多數呢？我們以問題若干測驗在第六級的439個十二歲的兒童。試驗結果，如果某問題在439人中能答有百分之六七十（230人上下），那麼某問題用在十二歲就很相宜了。餘可例推。

二 規定工作常準(Norms of Performance)法

測驗問題既有標準了，我們纔可以用這些問題來正式測驗兒童的智力。這裏要補述一句。規定問題標準時，我們是拿每個問題來測驗。發見問題甲合於十歲，問題乙合於十歲，問題丙也合於十歲了，然後纔把甲乙丙三題合歸十歲組。現在規定工作常準，我們是拿全組問題來測驗。智力相等的人也許不能答同一問題。例如張生、李生智力年齡 (mental age) 都是十歲，但是對於十歲組甲乙丙三問題，張生也許祇能答甲乙，李生祇能答乙丙。所以就每個問題說，我們雖會規定

每百人中應該有六七十人能答至於就全組問題說每個人能答得全組問題百分之幾，還要待測驗纔能知道。現在姑且假定測驗結果，大多數十歲的兒童對於十歲組問題每人能答對百分之六十五。那麼，我們就認定這個百分之六十五為尋常十歲兒童對於十歲組測驗問題之工作常準。這個工作常準既求出了，我們就可以把他當作尺秤一般用來測一切人的智力。不論什麼人對於十歲組問題祇恰好能答百分之六十五，那麼，不管他是五歲或是五十歲，我們都說他的智力年齡是十歲，因為他恰能夠上十歲兒童的工作常準。由此例推，自三歲至成人，每歲中的工作常準都可依法求出。

上面方法祇能拿工作常準定智力年齡。德人司脫恩 (Stern) 由此再進一步用數目字來表示智力程度。其法以實在年齡 (physical age, 簡稱 P. A.) 除智力年齡 (mental age 簡稱 M. A.) 於是得智力商數 (intelligence quotient 簡稱 I. Q.)。通常兒童智力商數都在 1 的上下，天材兒童智力商數都在 1 以上，

低能的兒童智力商數都在1以下。所以祇知某生智力商數，我們可以斷定他比通常人是聰明些還是笨拙些。例如張生實在年齡爲八，智力年齡爲十二，他的智力商數就等於 $\frac{12}{8}$ 或1.5。如果把張生的智力商數告訴和張生素不相識的心理學者，他一定知道張生是很有天材了。智力商數祇能用在兒童期，用在成人期便不精確。因爲人到成年以後，智力大半已達成熟，不像在兒童期那樣發達得快。所以三十歲人的智力年齡和五十歲人的智力年齡往往沒有分別。現在姑且說某甲實在年齡三十歲，某乙實在年齡五十歲，他們倆的智力年齡都是十六歲，如果求智力商數，甲當得 $16\frac{30}{30}$ ，乙當得 $16\frac{50}{50}$ 。這個智力商數不足憑，顯然可見了。

以上是規定智力測驗標準法的大略。我們一定要照這樣辦法一步一步的做去，纔能說測驗。這個方法看起來似乎很簡單，實行起來可就不容易。還有一層，智力測驗標準不是一成不變的。像金融一樣，他要隨時變遷。所以我們隔幾年就要

把舊標準修改一次。因為這個困難，我們不能直譯外國人所規劃的測驗問題和工作常準來用在我國學校。外國搬來的測驗不但不合時，也不合地；因為各民族的智力發達速度不必平行而且相等。現在應用智力測驗的人對於這個道理似乎疏忽一點，所以許多測驗問題都是從西籍直譯來的。原來我們對於外國所已有的測驗，祇仿效其意而更加以改良，固無不可；但是直譯斷難收效的。皮奈西門測驗初到美國時，陶爾曼（Terman）依舊法把皮奈的問題校正一下，結果發見原有四十九個問題中有兩個完全不能用，有廿九個要更動；例如有一個測驗須降下六歲（譬說在法國用在十四歲的測驗，現在美國祇能用在八歲）（Terman: Measurement of Intelligence, p. 61）。從這個例子看起來，可知我們不能把外國的測驗直譯來坐享其成，還要自己努力。

在未攔筆以前，我們還有一層要奉告教育界諸君。智力測驗不過是教育方法初步。爲什麼要測驗智力，因為懂得兒童的智力深淺。我們纔能設法對付他們。假

使智力已測驗出來了，我們如果不利用測驗的結果，還是徒勞無補。據個人的意見，將來新學制如能以智力年齡代實在年齡做入學升級的限制，智力測驗者的苦心就不辜負了。縱使這層一時不易辦到，天材兒童和低能兒童也應該各受特別的待遇。現在他們大半都混在一起，天材兒童因而不能盡量的發展，低能兒童因而不能循序的發展，雙方交受其害。教育當局對於這層可以不設法補救麼？

智力測驗法的研究

厚生

一 智力與智力測驗法

現在的實驗心理學，本來是先從感官知覺，次從表象及感情的研究而出發的。然到了最近，智的作用，也歸在他處理的範圍中了。理論的心理學與應用心理學，雖互相獨立，而著手研究這問題則同。就是在理論的心理學，為開思考心理的領域，在應用心理學，為顯智力個別的研究，余現在所論，是專屬於應用心理學方面的。像那智力活動的性質的研究構成思考的諸因子的考察，是全然別一問題。換句話說，所要研究的，不是現象的智力，而是能力的智力；詳細些說，是研究智力上個人的差別為目的。故所謂智力測驗法，就是決定個人智力程度的方法。如今被人攻擊的，都說我們不確實曉得智力的一般性質，這智力測驗，果能成功嗎？然這種攻擊，實在是不對的。就豫在科學上：吾人雖不確知電氣這樣東西，也能測驗

電力；雖不了解物質究竟是怎樣，也能對於活力（energy）爲量的研究。所以這樣的攻擊，在實際上，卻於智力的理論的研究，很有益處。故在某徵候能相當於各個人所示智力的差異，而他徵候不相當時，又在前徵候的系列，示高度的相關關係，後系列現低度的相關關係時；吾人所有智力上的知識，可由是而多少增加。如是而理論及應用的兩個研究方面，實是可以互相輔助的。

雖然，所謂智力，究是何物呢？考智力的定義，各學者互有異論，極不一致。然據史脫恩（W. Stern）氏所說「所謂智力，就是個人對於新要求，意識的順應自己所思考的一般的能力，也就是對於人生新問題的一般的精神順應力。」原來所謂智力，是一句概括的說話，若分析起來，就是知覺的精細，辨別的銳敏，記憶的正確，思考的迅速，知識的豐富，判斷的妥當，反應的敏活等機能。此種智力，是各人不同的；在精神發育不完全的人，他個人的差別，尤其顯著；所以測驗尋常兒童和低能兒童智力不同的方法，不單在學問上，就是在實際教育上，也是必要的。其中認這

智力測驗爲必要，而實行測驗的，要算精神病學家。原來常態與病態，是應該立個區別的標準。但他們並不豫先確定常態的標準，所以他的測驗，從心理學上看來，就有好多的缺點。其後漸知對於尋常兒童智力測驗爲必要，於是設定研究精神薄弱者比較的標準，而且更就常態的兒童，也確定智力的等級了。

二 個別式測驗法與系列式測驗法

測驗法最先考案出來的，是個別式測驗法；如用什麼文字填補，反應時間的測驗，聯想實驗，等測驗法，去測驗人的智力程度，都是有的。然用此等個別的測驗法，要想計測史脫恩所說一般的精神順應力，萬萬不能充分。何以呢？因爲行個別測驗法，受檢人應那測驗人的質問而發現的精神作用，決不是單純的，實在是種種部分的條件互相融合的結果。在這作用當中，智力的特別條件，究爲何等程度的動作，實是無從知道。所以像一般的精神順應力的複雜，若專用某一種測驗法，固

然不得完全測定，就使合用種種測驗法，恐怕也未必能完全測定他。因為智力是多方面的能力，是精神全體對於新要求的態度，各種測驗的結果，不過單能測驗他各方面的能力；要想由此而曉得一般的智力程度，非用普遍於智力全部的測驗法，終不能算是智力測驗。也有人說：測驗某一機能，他所得的結果，不就可算一般智力的決定嗎？然所謂智力，是指一般順應力說的；若要據一機能的優良而定他智力的程度，實在是極不妥當。故個別式測驗法，在測驗一一的精神能力，雖極有效；而實沒有可以算是智力測驗法的完全價值。

要想彌補這缺點而考案出來的，就是系列式測驗法。這個方法，先就各年齡用系列式設定智力的標準，比照這個尺度，去測定各人智力程度的。但精神病學者所用測驗的系列，全然缺乏統制他結果的原理，並且不脫主觀的直觀的診斷法的範圍。此外還有客觀的設定智力的程度，案出有學術價值的測驗的，就是皮奈

(Alfred Binet) 西門 (Th. Simon) 法了。法國皮奈及西門兩氏，在一八九〇年

時候，要想作「智力的尺度」而應於兒童各年齡，設定幾種測驗法；凡測驗能合格的，即認他是有相當於其年齡的智力。到了一九〇八年，就有所謂「皮奈西門法」的智力測驗標準，發表於世。實行以後，更發見許多要改良的地方，就在一九一一年，公表改訂法。但這測驗法，是「正否法」，受驗人只有能合格與不能合格的兩途；而不能合格的當中，雖也有種種程度，那便不能測驗出來，這是他的一個缺點；而且測驗法凡分幾組，實用上不便之處，也很不少。後來美國耶寇斯及勃利基斯兩人，又想法改良，以一組的測驗系列，測驗各年齡的智力；並且可以據受驗人答語的程度，用分數來表出他的成績；（一九一四年）這就叫做耶寇斯勃利基斯的分數式智力測驗法。

三 適用智力測驗法的各方面

一、異常的成人 異常的成人，是精神病學者的研究材料；他們也在個別式測

驗法以外，案出系列式測驗法。但他們研究法的內容，脫出智力測驗範圍的很多；他的大部分，與其說是量的測驗法，寧當說是質的測驗法。又雖多少用些量的測驗法，但可與正常人比較的材料，終歸是缺乏的。要決定受驗人的智力確是薄弱，總以先規定正常人的智力程度爲必要；精神病學者沒有這樣的規定，可要算是一個缺點。這個缺點，到了最近，也已多少減去些；像索姆美爾、蘭普爾格等精神病學者，拿這測驗法廣用到普通人，要想得著比較的標準，就是。然照國際醫學會議的決議，可以曉得精神病學者自身的要求，還是未能得到的。因爲他們的議決，明明是說：要得智力上精神病學的測驗的標準，無論如何不能不依賴心理學家。所以這件事情，現在應用心理學家，正在那裏著手考究咧！

二、異常兒 異常兒的問題，在近二三十年來，也爲教育上社會上及醫學上研究的中心。像異常兒童教育法、補助學校制、特別學級制、感化院等問題，近來都有人作論文去討論他。然兒童研究，總以個性爲必要。要行精神的診斷及訓練實訓

等，也不可確實知悉他的個性。要適應此等要求，那末智力的測定，雖不是唯一無二的事件，但至少也為重要因子之一。若拿精神病學的方法用到異常兒，他的缺點，恐怕越發地多。用在異常的成人，尙且沒有相當之規範的標準，而況在異常兒童，要想得著規範的標準，不是更加困難嗎？再加規範的標準，不是有了一個就夠的；必定各種年齡，都一一有了標準，方才可以。譬如異常的九歲兒童的智力缺陷程度，不通比較通常的九歲兒童的智力，始能確定罷了。總而言之，精神病學者，要想作各年齡可為規範的系列測驗法，畢竟沒有達到目的；而心理學家，現在已略得成功；像前面說的皮奈西門的測驗法，就是一個例子。這測驗法，雖然不能算得完全，但他的根本概念，是有永久的價值的，他能夠導引我們得完全的解決，這是可以無疑。

三、正常兒 正常兒的智力測驗，我們若看做他對於精神薄弱者的研究，只有所謂提供比較標準的附屬價值，那就大錯。因為在事實上，正常兒的智力測驗，是

與實際教育有密切的關係。一般的學校的試驗，祇能曉得學生知識及外部的教養程度；至其內部的質稟及精神能力的程度，就不能從試驗的結果看出來了。所以心理測驗可補助試驗方法的不逮，就在這裏。心理測驗的必要，尤在入學試驗爲最；但於通常的學校教授，也是不可少的。何以呢？因爲教育的施行，當以適合兒童個性爲第一要義。現今要求心理測驗的聲浪甚高，就是假定個性的充分研究是必要的原故。

我們若以爲施行智力測驗，就算了解兒童的個性全部，則最爲危險。智力比到他種種精神能力，是容易爲量的處理，所以我們動輒有過分重視智力的傾向。至於我們研究了智力，就大概可以明白他個性的成立，這是不消說得。我們確知個人精神上產物的多少，是跟著智力程度的；更了解一般的學校能率，一方因於智力到某種程度，一方因於別種因子到某種程度；這些都要算是心理測驗法的功績了。

正常兒童的智力研究最初是用個別式測驗法的。這法開始於德國，後來又經美國和法國大加改良擴充。自此以後，在法國有皮奈西門法發現，在英國則應用智力測驗的相關關係法，使智力測驗得越發正確；在美國或為斯泰富特的改訂心理測驗法，或為耶寇斯、勃利基斯的分數式測驗法。這都是改良智力測驗法青出於藍的。

四、普通成人 智力測驗應用於成人的方法，現在還沒有十分完全。但在這新範圍，應該要注意的，實屬不少。孟斯太保說道：對於職業的選擇決定，個性的正確知識，非常重要。并主張美國職業局，當協定心理測驗的事。在職業心理學上，智力測驗法，近來大被應用；如勞動者的選擇，職業適任者及不適任者的鑒別，又於求職者的職業指導，益益確認智力測驗的效果。

從前把智力測驗，考想他是在精神病理學的範圍的；如今已屬於純粹心理學的領域，不單在精神病學者及異常兒童教育者，大有效用。並且從法律家、軍人、一般

教育者，乃至產業經營者，也都不能不用他了。

四 智力測驗法的批評

我們於智力測驗實際的應用時候，像下面說的種種情形，也是一定要考慮的。

一、我們的智力測驗法，還沒有脫了準備的階級。像一般施行的皮奈西門法，也未認做是足以定為公共法規上精神診斷的標準。蘇比阿曼應用所謂「相關關係法」說道：「在我英國，年年由政府決定發表各兒童的精神指數，他的時期，大約也將不遠了。」然從我們想來，這的日腳，恐怕還遠得很呢！

二、智力測驗法，他的施行上，實在不易；要想施行，必定要有十分練習，心理學教養、批評精神。現在普通的教師，所用教育學的質問及試驗，是和心理測驗全然異樣的。當施行心理測驗時，動輒取別異的態度；而教育態度與心理學態度的中間，遂有劃然的區別。所以實行智力測驗，要想有顯著的效果，一定要先對於職員，施

行心理學的訓練，故學校校醫及學校心理顧問是不能沒有的。

三、我們不可過分重視智力測驗的價值，信他是完全無缺的測驗法。要曉得這種測驗，不過可補心理學的、教育的、醫學的考察所不足；決不是可以作此等的代用品的。

教育測驗

— Edward Thorndike 原著 —

劉建陽

教育底任務是變化人類。我們教員和學生將於本年內竭盡時光，使在思想、感情和行動上，有新而且較善的方法，以變化我們自己和別人。這些教室、實驗室和圖書館是我們底工具，輔助我們於知識、好尚和能力方面，改變人性，使他愈加良善。

要勝任這種事業，我們必須明白正確地知道何種變化是已有的，何種變化是應該作的。這種知識底定限和正確到了什麼比例，則教育成績和教育目的底知識，必定變為數量的，具有測驗底形式的。教育是一種人類的機器，因測驗人類本性和造就而受利益，正像機械工程和電氣工程之受賜於使用「尺磅」(Foot pound) ①、「加列里」(calorie) ②、「拂打」(volt) ③和「安培」(ampere) ④一般。

① 方底單位

② 熱底單位

③ 電底阻力單位

④ 電底強度實用單位

在教育上人類性質底測驗是從前所少有的。例如一九一〇年以前，中央政府，省政府和地方政府所報告的教育狀況，祇及於時間和經濟——教職員學生人數，授課日數，校舍和地址底價值，書籍用品底消耗。至於被教者底能力，他們所受的理智，品格和技能底進步，就略而不道，祇留作揣索和崇信底資料而已。

自然我們也有公認的教育成績，這即是每個學生對於每學科或活動所得的分數或等第，升級和畢業和獎賞，以及試驗結果所得當律師、醫生、或教書和服務各種文職底許可狀。這些分數等第祇可當作一種意見看待，並不是什麼測驗。這容易犯着兩種重要缺點：沒人對於所得的結果能確有把握，或分數和等第同實在程度能接近到什麼程度。例如專門學校一年生底代數分數，可以說是表示他先天的數學能力，或後天的數學能力，或數學的博學，或一時的記憶，或教員所命令的他能勤敏地依法作出，或善於預測教員底意旨而迎合他所歡喜的！我們若測驗溫度底高低，或電力底強弱，則一切內行人測量底結果，都是一樣的。但我們

若測驗一年生底拉丁文或代數即最精明的二十個教員必測出二十個不同的樣子。

再舉一個例的爾邦 (Dearborn) 在同一學校，同一班級，教授同一功課的教員中，發見甲教員比其他教員所給某生的分數多過十倍，而落第次數不上十分之一。純克斯坦 (Finkenstein) 也發見了，學習同一學科的學生們上期是甲教員教的，下期是乙教員教的，但這期較那期所得分數多過八十五分的當有三次。

現在一般的結果是可羞。佛斯特爾 (Fosser) 在哈佛低年初級學程中發見某生底英文分數多過拉丁分數三十五倍，米爾 (Meyer) 發見某個教員於五年之間，一千學生中沒有一人落第，而同校之他一個教員底學生每千人中必有三百人不及格。

我們縱能確知我們所測驗的，但所給的分數或等第同實在程度相差很遠。例如以回答某幾項特殊問題和證明某幾項特殊習題為測驗幾何能力的範圍。可

是伊里阿提和斯達遲 (Elliott and Starck) 發現一百個有經驗的數學教員於同一試卷上的答案所分等第之差，是自28至90。

我們可以想到，這樣自28至90的變異，多因於評卷者底態度嚴厲或寬恕而定，在這樣的情形中，主任哪，成績委員哪，連同學生自己哪，都可酌量增加每個教員底分數，表明教員個人的差別。所以生出這些變異來的更重要的原因，是由於各人所注意的性質不同。缺乏正確的評定教育成績的標準。除却各個人的寬或嚴不計，還不能杜免沒有差異或大大地將差異減少。

一個學生若於一年之中，他底英文作文，或初級化學，或英國史，正式得着70分，可是他和我們都不知道這70是代表甚麼，或者這真是不是60，或70，或80。明白規定的測驗單位可資計算的工具是沒有的。

第一步規定教育成績底單位，製定測驗這些成績底工具，已起始在十二年前了。起初試行的時候，即限於讀法，寫字，拼音和算術；這些功課雖屬簡單，可是已佔

去美國每年一千五百萬兒童底大部分工作了。

這些教育單位和等級所包含的假設和實驗是複雜而且繁難的，也太專門了，不易講述，但這些等差自身底性質也可略略地說明一下。例如習字一科，假設一排作標準的範字，第一字得零，因為他沒人認識，沒具有字體底任何美麗或優點。在這排底那端叫作17，代表字體通有的大量底美質。在兩端之間，設有模範字，代表1、2、3、4……，每兩字優劣之差異，同其他任何兩字之差異相等。從第五級到第八級的一千兒童底習字之中，在最優和最劣之間分爲十級，每級——即每十分之一——爲一個單位。

若寫字能力變了，他底進步是在標準等級上從8到了10，此時任何人在任何地當任何時候都能知道確切，好像我們說一個嬰兒從八磅變得十磅重了，或電流從八「安培」增到十「安培」了。將學生底習字平列於標樣表旁邊，兩相對照，找出最相似的一點，然後評定等級，此時多數評判人必相一致——自然是不如以

斗量酒那般精密，但反覆測驗，必能應合任何重要的教育目的。

試研究一個識字測驗。學生看見每行第一字之後隨有別的五字。他須於五字之中找出同首一字同一意義，或差近同義的一字，在下面劃一橫線。測驗起始所用的字是限字典內第一千字內較為重要的字，例如：

Afraid	Full of fear	Possible	Necessary	Raid	Ill
Baby	Manner	Trembling	Little child	Notice	Soft
恐怖	很怕	可能	必須	侵掠	害病
嬰孩	態度	發抖	小孩	佈告	柔軟

繼續下去，漸次用不甚緊要的字，例如代表第十千以內的：

Ambiguous	Offensive	Uncertain	Roomy	Very large	Material
Canyon	Menagerie	Palate	Valley	Gun	Rule
Classify	Arrange	Pacify	Make clear	Recede	Promote

Divulge	Recede	Common	Tell	Repress	Project
模糊	得罪(攻勢)	不定	寬舒	很大	物質
峽谷	動物園	上顎	山谷	槍	規矩
分類	佈置	和緩	講明	退讓	升擢
洩露	退讓	普通	告訴	遏抑	計畫

這樣測驗識字的工具有許多優點。我們現在可以注意兩個最顯著的：計分絕對是客觀的，同一試卷，從任何閱卷者可得同樣分數；在不同的各班或不同的各年，五十次試驗能夠作得確實地一樣繁雜，沒有雷同鈔襲之弊。

譯者案 上列識字測驗底譯文和原文迥乎不同。原文第一表是字典內第一千字以內的，第二表是第十千字以內的，譯文就沒限制了。原文雖是單音字或複音字，每個總算是一個字，譯文都弄成兩字了。讀者可依照這個式像，自己可另作識字測驗表參看心理雜誌第一期張耀翔先生底識字試驗。

從事教育科學的職工，因已編定了教育成績底單位和等第，而心理學家也改良着他們底智力測驗。這兩種科學正在協力設製各項學力試驗方法，如算術學習能力，拼音學習能力，或拉丁學習能力。

用規定好了的心理的和教育的單位，以測驗學童底能力和成績，這是現今小學所通行的。關於學校功課的一般智力，去年曾有一百萬男女孩子們受過測驗。那年對於讀法、寫字、拼音、算術、歷史和地理所施的測驗次數已超過二百萬了。

我們測驗學生底學業成績，和他對於那科的學習能力，然後以學習能力除學業成績，所得之商 (quotient)，謂之總成績 (accomplishment 或簡稱 A. Q.)。這種總成績是一樁重要的測驗。由能力 70 而來的 70 分，由能力 140 而來的 70 自然大不相同。用科學方法辦理的小學內，每個學生每年底總成績都登記着。天資較高的學生比尋常總成績祇近於「一」的學生，應多作功夫。這樣，他們才不染有懶怠和驕誇底習慣。天資低的學生也不因成績不佳而受譴責。他們祇須維持

總成績近於「一」而已。

這種學業測驗，能力測驗，和他們底商數測驗之法，可以發達起來，用在中學、專門學校和專業學校。我們若能用充分的巧妙和思素，製出許多方法，有統一而且精確的測驗，那時考試學生底作文、三角、化學、經濟或二年級的法文等科底成績，自然不憑着那班教員不測的繼幻了。

現在的情形最好的也屬可驚。在某科目的列等和學生底實在成績之間底比率（correlation），雖在美國最好的專門學校，差不多不過80，這即表示學校正式計分十分之六是錯的，即用抽彩法評定學生，也不過如是。若某科目有九百學生及格，一百學生落第，我們很有理由相信，落第學生之一半底實在功夫較之及格學生必更加優良些。

若任學生知道他們底學位、名譽和金錢的報酬多不憑着真正成績，反憑着偶然的機會和不相干的東西，那必減少學生對學校的信仰。學校自然可向他們解

釋，說某次分數多係錯誤，但兩年底平均分數可算確是一般的實在成績。但這樣的教訓是一種或然底理論，不足以慰藉學生——他若於甲學科縱優於乙學科的時候，乙學科能及格而甲學科反不及格，或他底那門功課超過某丙，丙已升級，而他必得多習一年，此時他底怨憤是不易息的。若論到教員，就有兩種情形：一種是頑固觀念，以爲他們所列的甲、乙、丙、60、70、80，是成績底鐵案；一種是草潦，隨便試驗，試驗底結果毫不過問，或是隨便計分，以爲認真也是無益。

如今專門學校和別的專業學校都有了試驗標準和別的成绩考試工具，並且還是可能有用；這是由於最近幾年實驗進步底結果。但是要求精密，必須由於長於各科的教員和善於心理測驗的熟手通力合作，非有高等的創造力和許多實驗不易成功。測驗學生底化學成績較難於測驗酸素底重量。測驗經濟知識底進步較難於計量洋銀價格底變遷。讀拉丁文的能力底確切單位和差等，比拉丁文自身還更複雜些。我們要真能測量成績，分出0、2、3、4……，又能分別各數底特

徵，所謂某科底4，確實是他底2底兩倍，因之能告訴學生自入學之始至畢業時各學科底程度真相。要做到這步，恐怕還須許多年代理。除非達到這步，則一切報告統計都是暗昧的，各學校底比較和教授法都是不可靠的，而教育所產生的成效究無從知悉了。這纔是我們底本務應努力求進步的哩。

向着客觀的測驗方面進行的各個方法都是有利益的。去年哥倫比亞大學考試一年生底「現代文明」備有一種新法，較從前的臆度已進了一步。凡有關係的教員和有學識的學生必不再想回到舊式的考試法了。對於學習能力測驗也是一樣。若不願學生底能力能否容受，一任教者勉強注入，有何益處？若有精巧的實驗，事先將學生底各種能力一一測出，然後因材施教，豈不更好麼？

特殊能力底測驗，如教學或語言之類，還沒在中等以上的學校行過。但普通抽象的智力測驗或學業成績測驗，在近三年內已在許多大學通行了。同時，哥倫比亞專門院主任，大學入學考試委員會長，和布勞大學科爾文教授已開始編製許多

事件和符號，測驗普通能力，作為新生入學考試之一種。同樣有短的考試（Short），到處通行，尤以加拉基工科學校為最；一八九四以來，哥倫比亞心理學部製有心理測驗方法，凡全請求入學的新生，都須受試。

這種測驗底價值已證明了。他確能預知學生在一年級中將來所得的等級。凡愚笨學生之混入專門學校，或由他們底教員底弄巧和家庭底金錢，或由任意發出的保證書底推薦，一經測驗，立刻就可指出，無誤。他可幫助主任或教員會火速地。正確地決定成績惡劣是否由於身體的智識的或道德的原因。他可任人核計相近的 A. Q. 或總成績。例如在某個大學，主任可以召集各項能力試驗分數很高的學生，開一談話會，告訴他們說，「你們底天資較高，成績應該比天資較低的學生優些。某某學科底成績，應該甲和乙之上，否則以不及格論，正像聰明較遜的學生底成績在丙以下的不能升班一般。」

在專業學校的測驗，因限於時間，不能多說，可指的有二年前的加拉基考察團

和他底報告，鼓勵工程學校有重大的測驗運動。

總而言之，在高等教育中，以成績測驗代憶度考試底效果傳播必廣，進行必速。所可慮的不是敵視的批評，乃是過分的希望，這是必須防範的。小學裏面常有許多錯誤，甚至妄誕的方法，躲在教育的科學底旗幟下面遊行。所得的測驗報告，切於實際的固多，然而盜虛聲，像以雷聲閃光底電量似的也很不少。這種事情對於用科學方法測驗的教育者，比對於無論何人，還更覺得可惡些。

現在有三種批評，雖大學教育中正當精確的測驗必得遭受。

第一，有人說，學問應該是爲着學問，這就是說，人們應該爲學問而求學問；在美國境內，注意於分數、獎品、學位等的人們太多了，學生爲着分數而用功的多於爲着真正的成績。無論這個議論有什麼力量，無非是傾向廢除正式的成績考查，或者使他們是真正成績底測驗。我們若保有分數和單位制度，學生將爲着分數和學位而讀書。我們能夠全行不要，或者我們能保有值得追求的部分。兩者都屬合

理，但是第二項似更可取。

第二，有人說，教育們底精力與其用來測驗成績有多麼大，毋寧用來使成績增大，這固不錯。對於許多的教員們和許多的學生們，最好是盡量去教，盡量去學，結果如何，可以付之信心和希望，或者甚至付之慈悲善士！況且有些天資很高的人們，不喜受科學式和業務式的拘限，並覺得這些造作是不相干的，甚至討厭難受；他們不應被要求而測驗他們正在那裏所作的，或用尋常話來說，他們不必知道自己在那裏作什麼。但是這些東西集合起來，不過是表示在教授上一個疏忽而已。我們大多數中，必須知道我們打算教什麼或學什麼，我們已教到那步或學到那步；將受賜於測驗教育目的和結果底工具，而不爲他所累。

第三，有人說，教育祇有較粗淺的部分能受估量品評，至於對人底精神上較爲優美的結果將因測驗而反喪失去了，其喪失底分量，視我們測驗他們爲比例——因之大學將變爲學者製造所，製造律師和醫生，擔保使人滿意，但是絕無陶冶。

還有一部分恐怖是科學和測驗若用到人事上去——家庭、國家、教育、宗教——必污壞人生底美麗，將他底高尚淪入物質主義。我無暇徵引證據，但願諸君相信。這個恐怖是沒根據的，完全立在偽的心理學上。凡屬存在的東西，就存在分量內。測驗他不過是知道他底有變換的分量而已。人雖在前日研究數量的植物，今日看花，不見花中美麗之減少。研究植物，是尋出他們對於系統的關係，精腺底流動和在血液中沙糖底生產，這並減少勇敢和忍耐底美德。若任何美德值得探求，我們知道他更深，測驗也更精，則我們探求他的心思將更親切，自作神祕，並不擡高人底威嚴。教育中的測驗科學底直接而且實用的利益，我們儘可安心地承受，並不危害理想主義。

上面那篇教育測驗是哥倫比亞大學教育院教育心理學教授桑戴克於去年九月在大學開學時的講演。內中很有可注意之點：（一）教育測驗之重要，（二）舊

法考試之不可靠和弊害，(三)識字測驗之實例，(四)新法測驗之近况，(五)測驗總成績後，對於高材生和低能生底成績考查，分兩種標準，(六)測驗不切實如以雷聲當電力之可慮，(七)新法測驗批評之答覆。

我個人底意見以爲一二兩項已不成問題了，無庸討論。第三項測驗方法還在草創時期。外國材料不合國情，可模倣而不可鈔襲。國內許多學校已試行的成法還不夠用。希望各明達的各科教育者和心理學者多多貢獻。第四項新法測驗在中國還很幼稚，偏僻區域更有沒聽見過的。第五項就我所知的，國內似沒人說過。最少是沒有學校已實用過的。測驗之後，若不分別學生智力和能力底高低，施以不相同的成績考查法，則測驗還無補於實際。優能生平常不用功，而考試偏偏勝人，低能生用盡苦力，結果反落人後。一個驕怠，一個懊喪，真是世間不平之事。譯者每當考試完了之後，常遇着學生問道：「先生，我這次及格了麼？祇要及格，便算了事。」這類學生多係優能生，至低能生每當考試之後，垂頭喪氣似的，不敢仰望。我

雖以期望鼓勵前者，以溫語慰藉後者，心中存着兩種不同的待遇，但不敢倡論兩種及格標準。如今既有桑戴克諸人提倡，各校各科急宜採行，在沒用學科制或選科制的學校，尤其應先行兩種及格標準辦法。

第六項也是實情。每當一種新制度倡行的時候，就隨着有聞聲吶喊，搖旗號召，慕形式，忘却實質。中國學生久已苦於舊法的考試。若新法測驗行得不好，反兒加上一樁負擔，壓得較前更加難堪。第七項的批評必不可免，但已有了答覆，我們不必多慮。可慮的還在我們自身方面。總之科學昌明底結果，可以支配一切人事。教育久已成爲一種科學了，教授法也成了一種藝術，新法測驗是兼有兩種性質，一面是科學，他面是藝術。

譯者識

教育叢書目錄

- [1] 新學制的討論(三冊) [2] 新學制中學的課程 [3] 小學的新課程(二冊)
- [4] 小學教育的實際問題 [5] 初級中學教育 [6] 大學校之教育
- [7] 師範教育改造問題 [8] 測驗之學理的研究 [9] 麥柯測驗法
- [10] 皮奈西門智力測驗法二冊 [11] 陸軍用的智力測驗法 [12] 團體智力學力測驗法
- [13] 五項測驗 [14] 測驗與入學考試的改進 [15] 教育統計法
- [16] 現代教育思潮批判 [17] 日本最近教育思潮 [18] 社會教育與個性教育
- [19] 教育與德謨克拉西 [20] 晚近美學說和美的原理 [21] 美育之原理
- [22] 美育實施的方法 [23] 教學之美學的基礎 [24] 教育上之理想國
- [25] 設計教學法實施報告二冊 [26] 三教育家之設計教學法 [27] 設計教學法概要
- [28] 設計教學法的實際 [29] 道爾頓制概要(二冊) [30] 道爾頓制的實際
- [31] 文科試行道爾頓制說明 [32] 教育哲學 [33] 教育之生物學的基礎
- [34] 教育雜文 [35] 教育獨立問題之討論 [36] 教育行政效率問題研究
- [37] 性教育概論 [38] 性教育的理論 [39] 性教育與學校課程
- [40] 男女性之分析 [41] 青年之性的衛生及道德 [42] 巴哥羅底兩性教育觀

[43] 特殊教育之實施
[44] 鄉村教育研究及研究法
[45] 社會學與教育

[46] 教學法

[49] 作文法

[52] 小學法

[55] 工藝法

[58] 協動法

[61] 教育法

[64] 歐美法

[67] 科學法

[70] 田徑法

[73] 訓育法

[76] 教材法

[79] 哲學法

[82] 變態心理學概論

[85] 小學教育參考書(三冊)

[83] 庚子賠款與教育(三冊)

[86] 現行教育法令(三冊)

[84] 教育短評

以上共壹百冊

國立北平圖書館

登錄號 000082 分類號

520.8
210
311

學法概要

學法

法法

研究及討論

學法

日美幼稚園

行實況三冊

視察後感想

面之研究

The Chinese Educational Review Series
 Theoretical Study of Intelligence Tests
 The Commercial Press, Limited
 All rights reserved

中華民國十四年七月初版

回(教育叢著)測驗之學理的研究(一冊)

(每冊定價大洋壹角)

(外埠酌加運費匯費)

此書
 有著作權
 翻印必究

編纂者 教育雜誌社
 發行者 商務印書館

印刷所 上海北河南路北首寶山路
 商務印書館

總發行所 上海棋盤街中市
 商務印書館

分售處 北京 天津 保定 奉天 吉林 龍江
 濟南 太原 開封 西安 南京 杭州
 蘭谿 安慶 蕪湖 南昌 漢口 長沙
 商務印書館

常德 衡州 成都 重慶 廈門 福州
 廣州 潮州 香港 梧州 雲南 貴陽
 張家口 新嘉坡

年	誌	教
彙	十	育
刊	六	雜

• 8