

心理與教育測驗的發展

簡 茂 發

心理與教育測驗是二十世紀一門新興的應用科學。其淵源可溯自三千多年前中國以六藝取士及隋唐之後的科舉制度 (Dubois, 1970)。如以科學研究的觀點來說, 心理與教育測驗的發展歷史尚短, 但其成長與進展甚速, 現已成為心理學在研究及應用上最主要的方法技術之一。世界先進國家在各級各類學校、就業服務部門、工商企業機構、軍事單位及醫院診所等, 均已普遍採用心理測驗, 作為測量人類行為個別差異的科學工具。由於科學技術與統計方法的進步, 心理測驗在編製和應用方面已獲致相當的成果, 尤其在教育科學化的潮流激盪下, 心理與教育測驗的發展更顯現其光明的遠景。本文擬先敘述心理測驗之起源, 然後說明美國心理與教育測驗之發展、我國心理與教育測驗之發展, 進而綜論心理與教育測驗發展的新趨勢。

一、心理測驗之起源

心理測驗萌芽於十九世紀末葉, 係受當時德、英、法三國有關人類心理的科學研究之影響。

(一) 德國實驗心理學研究

實驗心理學之父馮德 (W. Wundt) 於一八七九年在萊比錫大學設立心理實驗室, 利用物理學與生理學的儀器與方法, 測量物理性刺激和主觀的感覺之強度, 進行心物學 (psychophysics) 的實驗研究。此種實驗所使用的科學定量技術和精確記錄方法, 實為現代心理測驗之濫觴。

(二) 英國個別差異研究

達爾文 (C.R. Darwin) 的名著「物種原始」(Origins of Species) 指出生物種屬中有變異存在, 激起優生學者高爾登 (F. Galton) 對人類個別差異研究的興趣。高氏為了研究人類身心特質的遺傳性, 乃發明各種測量工具, 並以評定量表、問卷及自由聯想法, 獲取客觀可靠的量化資料, 進而創用分析個別差異資料的統計方法, 後來由其門生皮爾遜 (K. Pearson) 和斯皮爾曼 (C. Spearman) 予以發揚光大, 為心理測驗奠定了穩固的基石。從此以後, 測驗與統計也結了不解之緣。

(三) 法國異常心理學研究

十九世紀法國人士基於人道的立場, 對智能不足和心理失常者寄予關懷, 主動研究其成因, 並探求適當診療的方法, 以期改善這些不幸者的生活。當時有兩位著名的醫生艾斯奎勒 (E.D. Esquirol) 和施桂音 (E. Seguin), 致力於低能者之鑑別及訓練的實驗研究工作。前者以語文運用的能力作為鑑別低能層次的標準; 後者主張給予低能者充分的感官訓練以發展其動作能力。此等方面努力的成就, 對以後智力測驗之發展, 具有啓迪和推動的作用。一九〇四年, 比奈 (A. Binet) 承法國教育部之委託, 主持智能不足兒童的鑑定工作, 與其同事西蒙 (T. Simon) 合作, 於一九〇五年編成世界上第一套標準化的心理測驗——比西量表 (Binet-Simon Scale), 而被尊稱為智力測驗之父。

二、美國心理與教育測驗之發展

美國心理學者卡特爾 (J.M. Cattell) 是馮德的學生，曾受德國實驗心理學的訓練，後來在劍橋大學講學時，又與高爾登交往，對英國個別差異的研究亦有相當的認識和了解。卡氏由歐返美以後，任教於賓州大學，積極倡導設立心理實驗室及推展心理測驗運動。在一八九〇年發表「心理測驗與測量」 (Mental tests and measurements) 一文，創用「心理測驗」一詞，主張「心理學若不植基於實驗與測量上，決不能夠有自然科學的準確性。」他是使美國心理學由哲學的玄想轉入實際的應用之功臣，在心理測驗發展史上佔有相當重要的地位。當時由於教育上各種革新措施的需要，心理測驗乃成爲不可缺少的科學工具。在此方面從事研究的教育與心理學者承先啓後，人才輩出，形成心理測驗研究發展的風氣。

心理測驗發軔於歐洲，來自德、英、法三國的現代心理學研究之潮流在美國匯合，形成一股巨大的力量，推動心理測驗加速發展。其在二十世紀中的進展，可分爲五個時期：

(一) 開創期 (自一九〇〇至一九一五年)

由於上述四種科學研究的影響，心理測驗的可能性和重要性逐漸被人們所認識。當時教育與心理學者在測驗方面正進行各種初步嘗試性的工作，其研究成果爲以後的發展奠定了基石。在本期中，比較重要的測驗發展事項有三：(一)比西智慧量表 (Binet-Simon Intelligence Scale) 在一九〇五年初次問世，又經一九〇八年和一九一一年兩次的修訂，立即引起各國心理學者的注意，競相翻譯、修訂和應用，其中以美國斯丹佛大學心理學教授推孟 (L.M. Terman) 所主持的修訂工作最足稱道。(二)歐迪思 (A.S. Otis) 等學者由於研究興趣和實際需要，開始嘗試編製團體智力測驗，探討其可行性。(三)在學校教育方面，不同學科的成就測驗相繼出現，如桑戴克 (E.L. Thorndike) 的「書法量表」、史東 (C.W. Stone) 的「算術測驗」、白金漢 (B.R. Buckingham) 的「拼字測驗」等皆是。

(二) 盛行期 (自一九一五至一九三〇年)

大量編製及普遍應用測驗工具，而疏於批判和檢討，以致造成誤用、濫用的流弊，爲本期最顯著的現象。其測驗發展主要的事項如下：(一)一九一六年斯比智慧量表 (Stanford-Binet Intelligence Scale) 出版，樹立修訂測驗的楷模。(二)在第一次世界大戰期間，美國陸軍大規模使用團體智力測驗，有甲、乙兩種量表 (Army Alpha, Army Beta)，其成果相當良好，使智力測驗的概念流傳民間，並加強人們對團體測驗的信心。(三)學校教育工作所需的各種標準化測驗紛紛編製完成，並廣泛加以應用；世界上第一套綜合性成就測驗——斯丹佛成就測驗 (Stanford Achievement Test) 於一九二三年出版，可同時評量學生在各種學科上的一般成就，據以比較不同方面學習成就的高低。(四)在人格衡鑑方面，開始採用自陳量表 (self-report inventories)，如吳偉士 (R.S. Woodworth) 的「個人事實表格」、普萊西 (S. L. Pressey) 的「x-o測驗」等，使測驗方法的應用由智能之測量擴及情性領域。

(三) 批判期 (自一九三〇至一九四五年)

在此一時期中，心理測驗的研究轉向理論基礎的探討，測驗的方法和技術注重信度和效度之提高對各種測驗結果的解釋採取比較審慎的態度，而教育評量的重心已由少數學科知能的測量轉至整個教育目標是否實現的評鑑。此外，衡鑑人格特性及適應狀態的投射技術 (Projective techniques) 由臨床心理學所創用，受到相當的重視，使心理測驗能進一步探索個人心理深層的反應。

(四) 統合期 (自一九四五至一九六〇年)

第二次世界大戰中，美軍測驗計畫之實施又獲致相當的成功，加強了人們大量使用測驗的信心，於是全國性大規模的測驗計畫 (large-scale testing programs)，如 CEEB (College Entrance Examination Board) 和 ACTP (American College Testing Program) 等，受到社會大眾的支持，充分發揮測

驗在人員甄選 (personnel selection)、分類 (classification) 和安置 (placement) 方面的統合功能。各種測驗有計畫的安排配合，且以有系統的方式付諸實施，為本期測驗發展最主要的特徵。

(四) 推廣應用期 (自一九六〇年以後到現在)

在電腦時代的工業社會中，測量和現代生活息息相關，人們常以數字表示個別差異現象，每年成千上萬的兒童、青年及成人接受各種心理測驗，而測驗的結果決定了他們升學和就業的機會，於是物極必反，引起社會大眾對心理測驗的懷疑和攻擊，紛紛批評測驗的使用侵犯了私權 (invasion of privacy)，指責他未必是公平的衡鑑工具。因此，心理測驗的編製、使用以及結果的解釋分析，更需心理學專業知識的配合，務使學校教師、輔導人員和社會大眾皆有充分的認識和了解，然後測驗始能有效地推廣應用於人生各方面的活動，而發揮其最大的功能。

三、我國心理與教育測驗之發展

我國以考試取才已有相當悠久的歷史，隋唐以降的科舉制度對傳統教育的影響至為深遠。歐美各國文官考試制度亦取法於我國，故外國學者論述考試的理念與方法之起源，大致歸之於中國。林一真 (民 73) 曾為中國式智力測驗尋根，指出一些民俗活動如擲子兒、翻鼓、中國結、跳房子、仙人擺渡、九連環、七巧板、燈謎、字陣等，皆蘊含智能的巧思作用，與智力測驗的原理相通。至於以現代科學方法引進歐美心理測驗的技術，進行研究發展工作，雖然僅有大約七十年的歷史，但其研究成果不少，值得一述。路君約和黃堅厚 (民 71) 在其所著「我國心理測驗五十年來之發展」一文中，以中國測驗學會於民國二十年六月在南京成立及於民國四十年六月在台北復會為準，分為三個時期，說明我國心理與教育測驗發展的重要史實。

(一) 發軔期 (民國元年至民國二十年)

我國心理測驗的發展，係從智力測驗的引進開其端，帶動其他各種測驗的研究。民國四年，柯雷頓 (Creighton) 曾在廣東以直譯的智力測驗測量五百餘個兒童，測驗內容包括機械記憶、邏輯記憶、交替、類推等項目。綜其結果，中國男生智力僅及美國男生的百分之八十四，女生僅及百分之七十七，蓋因語文之不同所致。民國七年，華爾科特 (Wallcott) 任教於北京清華學校，曾以一九一六年斯比量表 (Stanford-Binet Intelligence Scale, 1916) 測驗該校高等科四年級學生六十四人，其結果有四十四人智商在一百以上。俞子夷等也在民國七年根據桑戴克的書法量表之編製程序，編成小學生適用的正書中字、小字及行書中字、小字四種量表，開我國教育測驗編製之先河。民國八年，廖世承與陳鶴琴在南京高等師範教育系講授心理測驗，是為我國大專院校設有心理測驗課程之始。民國九年，廖陳二氏在新生入學考試時使用心理測驗，正式介紹應用科學的測驗方法，翌年其合著的「智力測驗法」一書出版。

民國十一年，美國哥倫比亞大學教授麥柯爾 (W.A. McCall) 應中華教育改進社之聘，來華指導編製各種測驗工作，南京高師、北京高師、北京大學、燕京大學等校心理學者及教育學者全力協助之，至國十三年陸續完成智力測驗及教育測驗達二十餘種之多，計有(1)廖世承的團體力測驗、(2)劉延芳的中學智慧測驗、(3)德爾滿 (E.L. Terman) 的非文字智力測驗、(4)德爾滿的機械智力測驗、(5)陳鶴琴的圖形智力測驗、(6)陸志韋訂正的比納西蒙智力測驗等。

民國十七年，國民革命軍統一全國以後，致力於國家建設，社會各方面欣欣向榮，心理測驗也有了新的發展。中國國民黨於中央訓練部設置測驗科，一面編製標準測驗，一面舉行新法考試。民國十九年於中央政治學校設立測驗工作人員養成所，由吳南軒教授主持其事，訓練專才辦理測驗工作，對我國心理測驗之發展，頗有貢獻。

(二) 成長期 (民國二十年至民國四十年)

民國二十年六月二十一日，中國測驗學會在南京中央大學正式成立，參加者有來自全國各地的心理學者及教育行政專家一百三十七人，以研究測驗理論、推行測驗方法、並培植測驗專門人才為宗旨，發行測驗專刊，致力於各種測驗之編訂及研究發展。

在智力測驗方面，中國測驗學會與中央大學心理學系合作完成者，有小學智慧測驗兩種、中學智慧測驗五種、訂正古氏兒童智慧測驗及訂正墨跋幼兒量表等；與人事心理研究社合作者有大學心理測驗五種；與內政部合作者有普通警察個別測驗一種、普通警察團體測驗兩種及警官智慧測驗十種。此外，尚有陸志韋和吳天敏第二次訂正比奈西蒙智力測驗、艾偉修訂賓特勒兒童智慧測驗、黃覺民修訂幼兒智力圖形測驗等。

在教育測驗方面，艾偉於民國十五年受中華教育文化基金董事會贈予科學講座，從事於中小學學科心理的研究，尤注重國語心理的研究，遂編製若干國語及英語測驗，以研究國語文及英語學習心理，其研究結果曾發表為四大著作：一為漢字問題，二為國語問題，三為國文教學心理學，四為英語教學心理學，對我國教育學術之研究，貢獻至鉅。抗戰時，艾氏專心主持中大教育心理研究所，一方面繼續研究學科心理學，一方面擴大與指導所中師生編製心理與教育測驗。陸續發表的測驗，屬於國語文者有中學文白理解力量表、漢字測驗等八種；屬於數學者有小學算術應用題測驗、高中平面幾何測驗等九種；屬於英語文者有中學英語測驗、大一英語理解速度測驗等四種；屬於常識科者有初中常識測驗等九種；並於民國二十五年出版中學學力及高中與大學學力測驗兩種，幾乎包羅全部中小學學科，均由中華書局出版（路君約、黃堅厚，民 71）。

在人格測驗方面，蕭孝嶸於民國二十三年修訂「x-○測驗」，民國二十五年修訂「個人事實表格」，「馬氏人格評定量表」及「勒氏品質評定量表」。此外，尚有沈有乾修訂的情緒態度測驗、興趣測驗各一種及何清儒修訂的主要興趣測驗等。

在心理測驗推廣應用方面，內政部警政司於民國二十四年成立測驗小組，與國立中央大學心理系合作研究警員心理，並進行編製各種心理測驗。民國二十七年改組為警察智力測驗室，民國三十五年警察總署成立後，更加強此項科學技術的研究工作，先後編製完成「普通警察心理測驗」、「特種警察測驗」、「警官智力測驗」、「警官品格測驗」等數種。自民國二十九年至三十六年間，接受測驗的員警，即達二萬三千餘人之多。民國三十六年，我國空軍總司令部成立心理研究室，由程法泌教授主持其事，曾參照美國 Army General Classification Test 編訂四種普通分類測驗，以測量空軍官兵的智力。考試院也成立考試技術改進委員會，負責籌辦心理測驗的研究與實施等工作。

(三) 研究發展與推廣應用期（民國四十年至今）

民國四十年六月十日，中國測驗學會在台灣省立師範學院召開會員大會，重新展開心理測驗研究發展工作，於民國四十一年十一月第二次年會時，發行「測驗專輯」，於民國四十二年十二月舉行第三次年會時，出版「測驗年刊」第一輯，此後每年年會均循例印行年刊一輯，至民國七十六年一月已發行三十四輯，共有論文四百餘篇。

民國四十一年，國防部成立心理研究室，進行分類任職制度的規劃與研究，由路君約、孫敬婉及顧吉衛等教授先後主持普通分類測驗之編製與實施工作。民國四十二年，海軍總部人事署成立供求分類室，由路君約教授主持海軍測驗的編製和實施工作。民國四十四年，陸軍總部人事署成立分類任職室，由徐正穩教授主持陸軍人員分類測驗之編製和實施工作。民國四十五年，國軍退除役官兵就業指導委員會聘請蘇蕪雨、倪亮、蕭世朗、張肖松、湯冠英諸教授，應用陸軍乙種智力量表、班達完形測驗等，檢定不適服現役人員。

民國四十八年，教育部國民教育司與中國測驗學會合作編訂智力測驗，由程法泌、顧吉衛兩位教授主持編製國民智慧測驗甲種（文字）、國民智慧測驗乙種（非文字），於民國四十九年完成；由路君約、黃堅厚兩位教授主持比西智力量表第三次修訂工作，於民國五十二年完成。民國五十六年，教育部委

託國立台灣師範大學教育研究所編製國民中學適用的心理測驗，由賈馥茗教授主持其事，完成國民中學普通能力測驗一種及國民中學學科能力測驗五種，已建立台灣地區常模。民國五十七年，政府明令國民教育延長為九年，並公佈國民中學暫行課程標準，規定國民中學應辦理「指導活動」，必須實施各種心理測驗，以了解學生的智力、性向及興趣等心理特質，於是各大學心理系、教育心理系、教育系、教育研究所等學術研究單位紛紛從事各種心理測驗的編製和修訂，以利輔導工作之推行，已完成的智力測驗達二十餘種之多。民國六十年，教育部指定九所高級中學試辦學生評量與輔導工作，各大學亦應需要而編訂各種心理測驗，計有加州心理成熟測驗第五種、區分性向測驗、白氏職業興趣量表等。

最近數年來，先後完成兩種重要的個別智力測驗之修訂工作，乃是我國測驗學界的兩件大事。比西智力量表的第四次修訂，由教育部邀請路君約、程法泌、黃堅厚、張春興、盧欽銘、簡茂發、黃國彥、林邦傑等二十餘位教授成立工作小組，從民國六十四年四月開始，進行編訂測驗題、取樣及建立常模等，於民國六十五年七月完成。魏氏兒童智力量表（WISC-R）的修訂工作，由國立台灣師範大學教育心理學系主任陳榮華教授主持，參加者有郭為藩、黃堅厚、路君約、簡茂發、盧欽銘、吳武典等教授，自民國六十六年八月開始，至民國六十八年七月完成，共歷時二年。這兩套智力測驗對我國近年來積極推展特殊教育與輔導工作，已充分發揮衡鑑和診斷的功能，貢獻頗大。

在教育成就測驗方面，台灣省教育會與中國測驗學會合作，由師大宗亮東教授領導編製國民學校五、六年級適用的國語默讀、算術、算術應用題、公民、歷史、地理、自然等七種學科測驗，於民國四十一年由台灣省教育會出版。民國四十六年起，台灣省國民學校教研習會在政大程法泌教授指導下，三年中編成國民小學國語、算術、常識、自然、歷史、地理、公民等學科各年級上下學期的成就測驗四十種，參加建立常模的學生樣本多達六十萬人。台灣省立師範學院中等教育輔導委員會亦聘請專家學者設計編製初中適用的學科成就測驗，至民國四十三年已完成出版者計有沈亦珍氏初中英語測驗、唐守謙氏初中作文量表、孫邦正氏初中法量表、宗亮東和朱辛二氏合編初中代數測驗及其他學科測驗等。其後數年間，教育部亦聘請李祖壽、程法泌教授編製初中算術、代數、幾何、理化、博物等五種學科測驗。民國六十二年，教育部指定九所高級中學試辦評量與輔導工作，逐年陸續編製各年級學生適用的國文、英文、數學、三民主義、歷史、地理、物理、化學、生物等學科成就測驗，並建立各科題庫，共有選擇題二萬七千餘則。民國六十八年，國民小學新課程由一年級開始逐年實施，台灣省國民學校教師研習會為求配合，乃聘請學科專家與測驗學者共同指導國小資深優秀教師參與數學及自然科學成就測驗之編製工作，迄民國七十五年六月，完成國小一年級至六年級上下學期各一冊的數學、自然科學成就測驗共二十四種，建立都市、北部、中部、南部、東部及離島等五個地區的T分數、百分位數常模。

各大學心理系、教育系、教育心理系、教育研究所、心理研究所、輔導研究所等單位，對心理測驗的研究發展不遺餘力，貢獻良多，其主要成果如下：

1. 國立台灣大學心理學系所曾編訂大學資格測驗（CQT）、加州心理成熟測驗第五種（CTMM-Level V）史氏職業興趣調查表（SVIB）、基督氏性量表（G-Z-T-S）、卡氏十六人格因素量表（Cattell's 16PF）等測驗。
2. 國立政治大學教育系所及心理系所曾編訂孟氏行為困擾調查表（Mooney Problem Check List）、石爾斯通性格量表（Thurstone Temperamental Schedule）、多元性向測驗（Multiple Aptitude Test）、貝爾適應量表（Bell Adjustment Inventory）、白氏職業興趣量表（Brainand Occupational Preference Inventory）、包何二氏學習習慣態度問卷（Brown-Holtzman Survey of Study Habits and Attitudes）、葛氏非文字測驗（Cattell's Culture Fair Intelligence Test）羅桑二氏語文智力測驗、羅桑二氏非語文智力測驗（Lorge-Thorndike Intelligence Tests）、國民中學國文科、英語科、數學科、物理科、化學成就測驗等。
3. 國立台灣師範大學教育心理系及教育研究所曾編訂普通分類測驗（AGCT）、中學智慧測驗、瑞

文氏非文字推理測驗 (Raven's Progressive Metrics)、魏氏成人智力量表 (WAIS)、魏氏兒童智力量表 (WISC-R)、系列學業性向測驗、美術性向測驗 (修訂 Meier's Art Judgment Test 及 Graves' Design Judgment Test 而成)、青年人格測驗 (MCI)、圖畫觀察測驗 (Harrower's Multiple Choice Rorschach Test)、艾德華斯個人興趣量表 (EPPS)、逆境對話圖冊 (Rosenzweig Picture Frustration Study)、加州人格品質量表 (CPI)、社會興趣調查問卷 (Study of Values)、普通能力測驗 (甲)、國文能力測驗、英語能力測驗、數學能力測驗、機械能力測驗、動作能力測驗、國民中學適用國文成就測驗、英語成就測驗、數學成就測驗、社會科成就測驗、自然科成就測驗等。

4. 私立東海大學曾編訂國語文能力測驗、英語診斷測驗、數學測驗、職業興趣量表等。

行政院青年輔導委員會於民國六十二年六月邀集國內測驗學術界人士組成心理測驗策進委員會，同時委託中國測驗學會編印發行「測驗與輔導」雙月刊，迄民國七十六年四月已刊行八十一期。在編製心理測驗方面，曾委託中國測驗學會修訂完成「青年性向測驗」、「青年興趣測驗」、「工作價值觀量表」與「職業探索量表」。

內政部職業訓練局為配合我國職業訓練與就業輔導之需要，計劃編訂完整的心理測驗工具，乃於民國七十一年聘請學者專家組成職業心理測驗指導小組，研擬長期的發展計劃，並委託中國測驗學會進行有系統的測驗編製或修訂工作，迄民國七十六年已完成修訂通用性向測驗 (General Aptitude Test Battery)、修訂職業興趣量表 (USES Interest Inventory) 修訂基本語文及計算能力測驗 (Basic Occupational Literacy Test)、編製興趣檢核表 (The Interest Check List)、編製工作性格測驗、建立特定職業性向組型等任務。

此外，心理測驗的發展與文官考試制度及方法的研究改進之關係至為密切。考試院為改進考試技術，成立考試技術研究委員會，負責籌辦心理測驗的研究與實施等工作。自民國四十年至四十三年，各年度心理測驗均與高普考同時舉行，四十四年僅普通考試施行心理測驗。民國四十四年高普檢定考試開始採用測驗式試題的科目有國父遺教、憲法、歷史和地理；翌年，物理、化學、政治學、法學通論等科目亦採用新法考試試題；到了民國四十六年，又推廣至行政法、民刑法概要。迄民國四十九年止，一共有高檢部分的新法考試試題二十九種，普檢部分的新法考試試題三十一種，合計六十種。據統計，自民國四十四年至五十六年止，考試院共編印智力測驗八種和人員分類測二十五種，以拔取真才，杜阻倖進之徒。

四、心理與教育測驗發展之新趨勢

心理測驗發展至今，大約有四分之三世紀的歷史，理論的闡發日益精微，編製技術不斷進步，其應用範圍也逐大漸擴大。最近心理測驗的發展，以心理學的理論作基礎，運用統計學的方法和技術，在量的增加和質的提高的兩方面，均有相當可觀的成就。現代心理計量學 (psychometrics) 和差異心理學 (differential psychology) 的研究，已與心理測驗的編製和應用密切結合，使心理測驗成爲一門兼顧理論和實際的應用科學。由於新式電子計分機和電腦在測驗資料的處理方面提供最佳的服務，乃使心理測驗獲致長足的進展。綜觀本世紀以來心理測驗的演進趨勢，可知心理測驗之興起乃由於心理學者在理論方面的研究興趣以及社會方面的實際需要；其研究之內容自智能的測量而擴及整個人格的衡鑑，由表面行爲的觀察而透入深層反應之剖析，從一般學習成果之考查進而注重各方面行爲變化之評量；測驗的對象包括兒童、青年和成人，而應用的範圍已從學校推廣至社會各行業。茲分測驗的理論、編製和應用三方面，說明心理測驗今後發展的趨勢。

(一)理論方面

1. 心理測驗旨在增進個人的了解——現代民主化教育尊重學生的人格，教育情境之安排以學生身心的生長與發展為著眼點，舉凡教材的選擇與教法的運用，均力求配合學生的能力、經驗、興趣與需要，以促進學生完整人格的健全發展。為了解學生各方面個別差異現象，必須配合運用各種不同性質的心理測驗，以獲取正確可靠的量化資料，作為施教和輔導的依據。測驗結果應向學生作充分的解釋分析，使其了解自己，並作明智的抉擇，惟有關測驗的資料宜予適當的保密 (confidentiality)，以維護學生的私權 (protection of privacy)，這是學校使用測驗的基本原則。
2. 兼顧智性與情性之測量——個人是機能性的整體 (functional whole)，要了解一個人，必須從其完整人格各方面行為去觀察、測驗和衡鑑，而個體的行為主要涉及智性 (cognitive) 與情性 (affective) 兩種心理特質。前者屬能力範疇，包括智力、性向和成就；後者為非智能的人格特質，諸如性格、興趣、態度等均屬之。人類在這些心理特性上皆有個別差異存在，其測量的可能性已被肯定，雖然目前對情性特質尚無法加以精確的測量，但測驗學者正在方法上謀求改進，將來可能會有一些突破性的發展。由於智性與情性特質均為構成整個人格的要素，彼此密切相關而不可分，故現代心理學者認為這兩方面的測量一樣重要，應同時兼顧之。
3. 智能是多方面能力的綜合表現——心理學者對智能的性質及其組成因子的見解不盡一致，形成各種智能理論，影響智能測驗之內容及其測驗結果之解釋。然而，異中有同，大致認為智能並非單一因子，而是多方面能力的綜合表現，且智能之發展兼受遺傳和環境的交互作用，性向與成就並無截然區分的界限，故智力測驗、性向測驗與成就測驗固然有其目的和功能上的差異，但其內容殊多重疊之處，三者之運用上應力求相互配合，而不宜有所偏廢。
4. 重視創造力的測量與研究——最近有不少教育與心理學者致力於創造力 (creativity) 的研究，分析創造者的人格特質，並探求創造力與智力、學業成就之間的關聯性，發現創造力不同於一般智能，其可變性較大，以擴散性思考 (divergent thinking) 的能力為主，故無法以傳統的方式測量之，而須改用無固定答案的開放式題目，就受試者作答時所表現的獨創性 (originality) 變通性 (flexibility)、流暢性 (fluency) 和精密性 (elaboration) 四方面，以測量其超乎尋常的思考力。此外，創造性的表現屬情性範疇，尚須以人格測驗定性的方法衡鑑之。
5. 注重測驗性能之實證分析——各種測驗所測量的特質應予以具體明確的說明，行為樣本 (behavior sample) 和效度標準 (validity criterion) 之選取，須依據測驗的性質和功能而決定之。測驗的內容宜包含生活情境中反應性和操作性兩種行為，使其能充分反映個人在所測量的特質上差異情形。至於測驗的因素之確定，可用「多變數分析 (multivariate analysis)」的方法為之，而試題之難度和鑑別度以及測驗之信度和效度，更須以適當的統計方法，進行多次的複核，相互參照比較，以確定測驗真正的性能，進而適切地應用測驗，發揮其評估、診斷和預測的功能。
6. 潛在特質理論之建立與發展——測驗理論是測驗編製、解釋及應用的基礎。鑑於傳統的測驗理論之缺失，現代心理計畫學者乃提出新的測驗理念，建立並發展更適宜的心理測驗理論，此即「潛在特質理論」 (latent trait theory)，或稱「試題反應理論」 (item response theory)。Lord (1952, 1953)，首倡此一理論，Rasch (1966) 提出單參數的對數模式，其後學者，(Bock, 1972 ; Hambleton & Traub, 1973 ; Hambleton & Cook, 1977 ; Lord, 1980 ; Samejima, 1972 ; wright 1977) 加以發揚光大，雖然所用的函數程式不盡相同，但有其共通性，要點有四：(1)心理特質潛藏在個體之內，無法直接觀察，祇能從受試者的行為表現加以推衍得知；(2)估計試題難度時，應摒除受試樣本程度的影響；(3)估計個人特質分數時，應摒除測驗特性的影響；(4)把試題難度及特質分數化成相同的尺度，便於測驗結果的解釋與預測。基於上述觀點，研究者可選擇適當的試題或用加權的方法組成測驗，以增加所估計特質的精確性，也可經由電腦程式來估計潛在特質之值。

(二)測驗編製方面

1. 對測驗編製持審慎態度——目前心理學者對測驗之編製，已持相當審慎的態度，認為它是一種有計畫的專業性工作，必須按照一定的程序，經由設計、命題、預試、項目分析、選題、建立常模等步驟，始能編出一套測驗，而一套優良的測驗有其必備的條件，測驗編製者尚須顧及其信度、效度和實用性等問題。美國心理學會（American Psychological Association）於一九七四年出版「教育與心理測驗的標準」（Standards for Educational and Psychological Tests）一書，對標準測驗的信度、效度、常模等均有明確的規定，已成為心理學者編製測驗的圭臬。布羅斯（O.K. Buros）從一九三八年起編輯「心理測量年鑑」（Mental Measurements Yearbooks），至今已出版八卷，對各種標準測驗均有中肯的評論，可作為一般人選擇測驗與編製者修訂測驗之借鏡或參考。
2. 成立題庫以備測驗取材之用——試題是組成測驗的要素，有了適當的試題，始能構成一套可靠而有效的測驗。現代由於電腦的發明和統計分析方法的進步，大規模的測驗編製機構均事先配合發展測驗的需要，約請專家大量命題，經有計畫的預試，然後進行電腦化的資料處理，求出各試題的難度和鑑別度，再依試題材料的性質及其功能，分門別類予以儲存，成立題庫（item banks），將來編配正式測驗各種複本之試題時，即可通盤考慮各項因素從中取材，使測驗具有相當的適切性，而更能發揮其功能。
3. 研究改進編製成就測驗的新方法——傳統編製成就測驗的方法大多以課程教材內容為其主要的著眼點，力求試題的取材出處符合均勻的比例分配，據以編成「內容取向的成就測驗」（content-oriented achievement tests）。鑑於正常的教學與評量須兼顧教材內容和教學目標，成就測驗之編製除了切合教材內容外，應依據教學目標，就學生行為變化的不同層面加以分析，注重認知、情意及技能等心理歷程之評量，而編成「歷程取向的成就測驗」（process-oriented achievement tests）。因此，在編製學科成就測驗的時候，通常先分析課程的教材內容及行為目標，進而形成「雙向細目表」（two-way specification table），作為命題的藍圖，如此可提高其內容效，發揮評量的功能。最近十餘年來，有些學者（Anderson, 1972；Bormuth, 1970；Gronlund, 1985；Hively et al., 1973；Millman, 1973；Popham, 1978, 1981）指出傳統編製成就測驗方法的缺點和限制，引介「領域參照測驗」（domain-referenced testing）的理念，發展一些編製成就測驗的新方法，諸如語文轉換法、試題架構法、詳述目標法、層面設計法等（黃國彥，民73）。
4. 開拓建立常模的新途徑——測驗結果的解釋必須有適用的常模作為參照的基準。常模的適用性取決於標準化樣本的代表性和人數的多寡。從前測驗求取常模的範圍往往過於廣泛，晚近的趨勢則根據測驗目的選擇適切的樣本以建立常模，除全國性常模（national norm）外，另有地區性常模（local norms）、學校常模（school norms）以及其他各種分類的常模。為了便於兩種以上測驗分數之間的比較，最近測驗學者採用「雙重標準化」（dual standardization）的方式，參酌測驗的性質、功能和將來應用的範圍，選取具有共同代表性的樣本，以建立常模。另有一些測驗學者針對測驗結果之解釋與交互應用的需要，提出建立「校準常模」（calibrated norms）的理論與方法，以性質相同而已標準化的舊測驗為定準，基於「相等百分位數」（equipercentile）原理，據以建立新測驗的常模。在成就測驗方面，除傳統的建立常模之方法外，測驗學者發展一種「試題抽樣」（item sampling）的方法，分別求出學生在各類試題通過的百分比，以比較受試者在不同方面學習成就的水準，從教育評量的觀點而言，此種方法更具有實用的價值。

(三)測驗應用方面

1. 建立完整的測驗計畫——學校為了瞭解學生，以便因材施教，必須使用心理測驗。由於心理測驗的種類繁多，其性質和功能各有不同，在實施和應用方面宜作適當的安排，始能前後連貫，相輔相成。因此，建立學校測驗計畫（school testing program）是首要工作。一項完整的測驗計畫應包

- 括八個步驟：確定計畫的目的、選擇適當的測驗、實施測驗、計分、分析並解釋測驗分數、應用測驗結果、評鑑測驗計畫的成果，登錄資料和提出報告。測驗計畫的施行，是全校教師共同的責任，必須以合作的方式完成之。
2. 測驗的實施、計分和解釋分析電腦化——資訊時代的來臨，電腦在學校等機構普遍使用，對心理與教育測驗的研究發展及推廣應用之影響頗大。電腦可以儲存大量的試題，視受試者的背景條件、能力水準及作答反應情形等，取用最適當組合的試題，進行有彈性且具最大效率的施測，如此可在符合測驗標準化的要求下，充分適應個別差異，使每個受試者都能盡力而為，作最大或最切合實情的表現，此即所謂「適性測驗」(tailored testing) 或「個別化測量」(individualized testing)。使用電腦施測的結果，可套用現成的計分及資料處理的程式，發揮自動化功能，打印出測驗分數及其經由統計分析的結果，一貫作業，可達到正確、快速、簡便、充分、有效的科學要求。
 3. 兼顧團體性和個案資料的分析——學校對學生測驗結果的分析，可從廣面的 (extensive) 和深入的 (intensive) 兩個角度同時進行，一方面從大量的數據中尋求一般的法則 (nomothetic approach)，以認識學生團體普遍的心理現象，另一方面就個別事例加以精深的探究 (idiographic approach)，而對某一特定學生各種心理特性之間的交互關係獲取透徹的了解。兩者在方法上相互補充，在結果上互相印證，使心理測驗的應用更能切實有效。
 4. 綜合各項資料解釋測驗結果——各種測驗結果之解釋，必須與其他有關資料相互配合，對照比較，始能充分顯示其意義，而有助於真正了解受試者的能力組型和人格適應的狀態。單一分數祇能表示籠統的結果，而多項測驗分數所組成的側面剖析圖 (profile)，可顯示個人在各種心理特質上優劣參差的情形，無論是臨床上的診斷或實際上的應用，都更具有價值。對測驗結果的解釋，依其方向和著眼點之不同，可分為四個層面：(1)描述身心屬性的現狀；(2)追溯其現狀的形成原因及其演變之過程；(3)以過去和現有資料為基礎，預測其未來可能的發展；(4)根據既定的標準，對行為表現作價值判斷，並就其出路的抉擇，提供評鑑性的建議。
 5. 注重測驗結果之適當運用——一般人由於對測驗的基本理論與技術未能深切認識，而容易產生誤解、錯用和濫用。其實，心理測驗並非萬能的科學工具，各種測驗均有其特殊的用途和限制，唯有在它的限制範圍內，由適當的人員，依照它的特性，作合理的運用，始能發揮其最大的功能。除了測驗本身有其功能上的限制外，測驗的結果因受外在許多因素之影響，必然造成誤差，故我們在分析測驗結果時，必利用各種有關資料加以研判，尤其應考慮測驗的信度和效度，據以估計測量誤差的大小，推斷其可信的程度，進而能適當地運用測驗結果，以解決一些實際的問題。
 6. 心理測驗與教學評量密切配合——無論教學前、教學中或教學後的評量，均可採用心理測量的科學方法，以了解學生身心發展的成熟程度或起點行為；隨時掌握學生在學習過程中行為變化的情形，診斷其學習困難之處，並分析其原因，及時給予補救教學及學習輔導；最後衡鑑學生累積的成就水準或終點行為，據以評定其成績，並考查其努力程度和進步情形。有些學者 (Glaser & Klaus, 1962 ; Popham, 1978, 1981) 檢討「常模參照評量」(norm-referenced evaluation) 的缺失，倡導「標準參照評量」(criterion-referenced evaluation) 的方法和「精熟學習」(mastery learning) 的精神，已形成在此方面繼續研究發展的一股潮流，藉測驗與評量，以維持教學水準，促進教育的正常發展。

參考文獻

- 王鳳喈：由本刊一至十輯論文總目錄看測驗在中國之進展。測驗年刊，民國五十二年，第十輯，第一至五頁。
- 林一真：潛在特質理論簡介——測驗編製的發展新趨勢。見路君約編：我國測驗的發展。中國測驗學會，民國七十一年，第六十一至七十頁。
- 林一真：中國式智力測驗。張老師月刊，民國七十三年，第十三卷第四期，第三十七至四十五頁。
- 候璠：測驗之重要及趨勢。教育文摘月刊，民國五十年，第六卷第十期，第一至四頁。
- 張肖松：心理測驗簡史。測驗年刊，民國四十三年，年第二輯，第三至十頁。
- 黃堅厚：近年我國測驗工作的進展。測驗年刊，民國五十九年，第十七輯，第一至五頁。
- 黃國彥：編製成就測驗的新方法。測驗年刊，民國七十三年，第三十一輯，第六十九至八十二頁。
- 程法泌：五十年來心理測驗在中國的發展。見中國教育學會主編：近五十年來之中國教育。台北市，復興書局，民國六十六年，第一三六至一八一頁。
- 路君約、黃堅厚：我國心理測驗五十年來之發展。見路君約編：我國測驗的發展。中國測驗學會，民國七十一年，第五至十六頁。
- 簡茂發：心理測驗的發展趨勢，見龔寶善主編：昨日今日與明日的教育。台北市，台灣開明書局，民國六十六年，第七五二至七六六頁。
- 簡茂發：心理測驗問題之檢討。測驗年刊，民國七十四年，第三十二輯，第一九九至二〇二頁。
- Aiken, L.R. Psychological testing and assessment.(4th ed.) Boston : Allyn and Bacon, 1982.
- American Psychological Association. Standards for educational and psychological tests. Washington, D. C. : Author, 1974.
- Anastasi, A. Psychological testing. (5th ed.) New York : Macmillan, 1982.
- Anderson, R.C. How to construct achievement tests to assess comprehension. Review of Educational Research, 1972,42, 145-170.
- Bloom, B.S., Hastings, J.T.,& Madaus, G.F. Handbook of formative and summative evaluation of student learning. New York : McGraw-Hill, 1971.
- Bock, R.D. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. Psychometrika, 1972,37,29-51.
- Bormuth, J.R. On the theory of achievement test items. Chicago : University of Chicago Press, 1970.
- Buros, O.K. The eighth mental measurements yearbook. Highland Park, N.J. : Gryphon Press, 1978.
- DuBois, P.H. A history of psychological testing. Boston : Allyn and Bacon, 1970.
- Glaser, R. Instructional technology and the measurement of learning outcomes : Some questions. American Psychologist, 1963, 18, 519-521.
- Glaser, R. and Klaus, D.J. Proficiency measurements : Assessing human performance. In R.M. Gagne (Ed.), Psychological principles in systems development. New York : Holt, Rinehart & Winston, 1962, 419-474.
- Gronlund, N.E. Measurement and evaluation in teaching. (5th ed.)New York : Macmillan, 1985.
- Hambleton, R.K., & Traub, R.E. Analysis of empirical data using two logistic latent trait models. British Journal of Mathematical and Statistical Psychology, 1973, 26, 205-211.
- Hambleton, R.K., & Cook, L.L. Latent trait models and their use in the analysis of educational test data. Journal of Educational Measurement, 1977, 14, 75-96.
- Hively, W., Maxwell, G., Rabehl, G., Senson, D., & Lundin, S. Domain-referenced curriculum evaluation : A technical handbook and a case study from the Minnemast Project. CSC Monograph

- Series in Evaluation, No. 1. Los Angeles : Center for the Study of Evaluation, University of California, 1973.
- Holtzman, W.H. The changing world of mental measurement. American Psychologist, 1971,26,546-553.
- Hopkins, K.D., & Stanley, J.C. Educational and psychological measurement and evaluation. (6th ed.) Englewood Cliffs, N.J. : Prentice-Hall, 1981.
- Lindquist, E.F.(Ed.)Educational measurement. Washington, D.C. : American Council on Education, 1951.
- Lord, F.M. A Theory of test scores. Psychometric Monograph, 1952, No.7.
- Lord, F.M. The relation of test score to the trait underlying the test. Educational and Psychological Measurement, 1953,13,517-548.
- Lord, F.M. Applications of item response theory to practical testing problems. Hillsdale, N.J. : Erlbaum, 1980.
- Mehrens, W.A., & Lehmann, I.J. Measurement and evaluation in education and psychology. (3rd ed.) New York : Holt, Rinehart and Winston, 1984.
- Millman, J. Passing score and test lengths for domainreferenced measure. Review of Educational Research, 1973,43,205-216.
- Nunnally, J.C. Psychometric theory. (2nd ed.) New York : McGraw-Hill, 1978.
- Popham, W.J. Criterion-referenced measurement. Englewood Cliffs, N.J. : Prentice-Hall, 1978.
- Popham, W.J. Modern educational measurement. Englewood Cliffs, N.J. : PRENTICE-HALL, 1978.
- Popham, W.J., & Husek, T.R. Implications of criterionreferenced measurement. Journal of Educational Measurement, 1969,6,1-9.
- Rasch, G. An item analysis which takes individual differences into account. British Journal of Mathematical and Statistical Psychology, 1966,9,49-57.
- Samejima, F. A general model for free-response data. Psychometric Monograph, 1972, No.18.
- Sundberg, N.D. Assessment of persons. Englewood Cliffs, N.J. : Prentice-Hall, 1977.
- Thorndike, R.L.(Ed.) Educational measurement. (2nd ed.) Washington, D.C. : American Council on Education, 1971.
- Thorndike, R.L., & Hagen, E.P. Measurement and evaluation in psychology and education. (4th ed.) New York : Wiley, 1977.
- Thorndike, R.L. Hagen, E.P., & Sattler, J.M. Technical manual for the Stanford-Binet Intelligence Scale : Fourth Edition. Chicago : Riverside, 1986.
- Wechsler, D. The measurement and appraisal of adult intelligence. (4th ed.) Baltimore : Williams & Wilkins, 1958.
- Wechsler, D. Manual for the Wechsler Intelligence Scale for Children-Revised. New York : The Psychological Corporation, 1974.
- Wright, B.D. Solving measurement problems with the Rasch model. Journal of Educational Measurement, 1977,14,97-116.

A BRIEF HISTORY OF PSYCHOLOGICAL AND EDUCATIONAL MEASUREMENT

Professor
Chien, Maw-fa

(Abstract)

The historical origins of psychological and educational measurement are lost in antiquity. There was an elaborate system of civil service examinations in the Chinese empire some three thousand years ago. Modern psychological testing has roots in studies by European and American scientists on the measurement of individual differences in human behavior. A brief overview of the historical antecedents and origins of psychological and educational measurement will provide perspective and should aid in the understanding of present-day tests.

In recent years, numerous changes have taken place in psychological and educational testing. Some of the more important trends concerning current developments in psychological and educational measurement are as follows :

1. There is more use of computers in testing. A complete computer-assisted test construction system should include item banking, item generation, item selection, and item printing. The computer has also been used for tailored (adaptive, sequential) testing.
2. There is increased use of criterion-referenced measurement. The use of tests to improve learning and instruction has created a greater need for criterion-referenced testing. Accountability and minimum-competency testing are designed to determine how well pupils are achieving specific learning tasks; this requires criterion-referenced measures.
3. There is increasing public concern about testing. Legal scrutiny of psychological and educational measurement is both a present and a future reality.