

國民中小學團體語文智力測驗編製報告

林幸台

本測驗係以國民中小學學生為適用對象，依記憶、理解、推理、評鑑四種智能運作方式之架構，以語文及算術材料為內涵所編製之團體智力測驗。共有國中、國小高年級、中年級、及低年級四種版本，測驗內容依對象之不同分別編有五至六種分測驗，包括歸類測驗、延宕記憶、數字系列、數的比較、替代比較、算術推理、語文類推、邏輯推理等八種。各版測驗均經兩次預試，編定正式題本，並建立標準分數常模。初步信度及效度考驗包括內部一致性、重測信度、及效標關聯效度、因素分析建構效度等。

關鍵字：智力測驗，團體智力測驗

Keywords: group intelligence test, intelligence test

壹、緒言

在特殊兒童鑑定與診斷過程中，智力測驗所扮演的角色相當重要，而團體智力測驗更是初步篩檢過程中最常使用的工具，然而目前國內適用之團體智力測驗為數甚少，其編製年代或常模大多已有相當時日，無法充分反應社會現況與兒童之智力。林幸台等人（民80）在探討特殊教育機構使用測驗現況的研究中，建議：(1)修改編製過久的測驗之題目；(2)提供詳細的標準化樣本取樣原則與樣本特性資料；(3)建立特殊用途常模；(4)提供完備的信度與效度資料；(5)加強測驗結果解釋及運用之說明；該研究並將中小學學生適用的文字型式之智力測驗列為最優先編製或修訂其常模之重點。周台傑與吳訓生（民82）的研究亦將編製團體文字智力測驗列為優先，由此可見國內對此類測驗需求之殷切程度。

自 A. Binet 編製第一份智力測驗以來，有關智力的理論甚多，除 g 因素論、二因素論、多因素論、階層論、及基爾福特之智能結構模式外，近年有多元智力 (Gardner, 1983)、三元論 (Sternberg, 1985) 以及訊息處理論 (Naglieri & Das, 1990) 等新理論出現，然依此等理論編製評量工具卻非易舉。Gardner 與 Hatch (1989) 依其理論所編之適用於學前、小學與中學學生測驗，為少數可見之成果，但仍未能免除主觀、冗長、昂貴、過於複雜等問題 (Sternberg, 1992)，即使目前認知心理學盛行時期，亦尚未發展出適切的智力測驗；若干學者試圖從「日常」語言中探究智力的本質與新的智力測驗內涵 (Derr, 1989; Sternberg & Wagner, 1986)，均未獲得有效的成

果，因此 Sternberg (1992) 認為目前依據新智力理論發展的測驗工具，仍無法取代傳統測驗，至多僅能補充其之不足；換言之，依據新智力理論編製的智力測驗尚需相當時日始可能推廣供各級學校使用。

近年來，傳統智力測驗受到頗多的質疑（路君約，民 81；Sternberg, 1985）。以新智力理論的觀點論之，傳統智力測驗主要的缺點即在其為靜態的評量、過於簡化智力的範圍、或未考慮受試者其它可能的反應（陳李綢、林清山，民 80）。因此目前在特殊教育中頗受重視的檔案式評量 (portfolio assessment)、實作評量 (performance test) 或課程本位評量 (curriculum-based measurement)，主要即在彌補傳統測驗之不足 (Finch, 1991; Lidz, 1987; Nickerson, 1989; Wainer, 1993; Wiggins, 1989)，然而此等評量方式亦有類似上述之問題，且較傳統測驗更易受個人背景的影響；如做為個別評量的依據，則可顯出其特色，但在團體評量上仍有許多限制，尚難普遍使用。

有關智力的理論甚多，依其特性與內涵，可區分為三大類：(1) 發展理論取向 (developmental approach)；(2) 心理計量取向 (psychometric approach) 取向；(3) 訊息處理取向 (information-processing approach)（蔡崇建，民 80）。各個領域又有不同的模式、編製成多樣的測驗，Sattler (1988) 認為有關智力的界定或其評量方式的概念紛陳，係由於智力本質上是一種歸類而非實體，智力本身是抽象的心理建構而非具體的實體，因此智力的定義仁智互見實不足為奇。綜觀上述所述各取向之智力理論，就現階段而言，心理計量取向的智力理論仍在測驗建構方面居領導地位，而學者努力的重點，在於統合各種取向於一完整的理論系統下，期望能在智力的理論及測量方面，突破舊有各自為政的窠臼，進而提昇測量工具的精確度與預測性（翁儷禎，民 82）。

總之，由有關智力理論與測驗發展背景之分析，可知依據傳統智力理論所編製的智力測驗並非絕無優點，在無其他更適合的智力測驗可資替代的情況下，若能儘量設法改善其缺失之處，如性別歧視、種族或文化歧視、以及因樣本誤差所導致之得分不穩定等問題，則尚有甚多發展空間 (Matarazzo, 1992)。至於文化差異對智力測驗結果的影響素受測驗學者的關注 (Jensen, 1980)，現代社會變遷之急遽，更使測驗內容能否符合國內實際情況的問題受到測驗使用者的質疑，加上近來因著作權問題，引發國內許多測驗無法使用的困境，更迫切需要自行發展適用之測驗。

本研究之目的即以自編題目方式，參酌傳統智力理論之架構，並兼顧古典測驗理論與現代測驗理論的重點，編製國民中小學學生適用之團體語文智力測驗。由於跨越九個年級，在內容、題型、施測方式等方面均有所不同，乃分別編定國中版、國小高年級版、中年級版、及低年級版四種版本，並分別建立總分標準分數 ($M=100$, $SD=15$) 常模。同時為考慮測驗之適用性，各版所擬題目皆以國內學生於學校及

日常生活中實際接觸之事物為範圍，自行編製適用之題目。

貳、研究方法與設計

一、基本設計

依據文獻探討的結果，研究小組嘗試以「智能運作方式」為編製本測驗之基本向度，其內涵係參考 Guilford 理論中的「運作」向度、Thurstone 的群因論及 Cattell & Horn 的理論；其成分包含：記憶、理解、推理及評鑑。研究小組同時以此架構，分析國內現有測驗，嘗試整理現有測驗在各智能運作方式中採用之題型。

根據陳明終等人（民 77）編著之「我國心理與教育測驗彙編」中所列國內現有團體語文智力測驗，研究者針對其構念、理論架構、適用範圍、分測驗題數加以整理、歸納，發現國內現有智力測驗四種智能運作方式之題型大致包括：

1. 記憶：(1)延宕記憶。
2. 認知理解：(1)字詞替換、(2)詞義辨別、(3)語句重組（刪字）、(4)語詞歸類（語文歸類）、(5)比較大小（數學理解）、(6)數系。
3. 推理：(1)算術推理、(2)語文推理。
4. 評鑑：(1)辨別評價、(2)邏輯推理。

研究小組考慮測驗的架構與適用性，乃選出：記憶（延宕記憶）、認知理解（字詞替換、歸類、數系）、推理（算術推理、語文推理）及評鑑（邏輯推理）等較具代表性之題型，作為本測驗之架構。

二、編製原則

本研究目的為編製國民中、小學生適用團體語文智力測驗，乃依一般心理測驗編製之原理，分年採取下述程序進行各項研究工作：

1. 文獻探討：蒐集相關資料，探究智力理論與國內外學者所編製之智力測驗內容與結構。
2. 設計測驗基本架構：依據文獻探討結果，設計本研究所擬編製之測驗架構（包括分測驗之型式與內容），並據以編擬國中版題目。
3. 進行預試前試作：約請數名教師讓學生試作，並就題目性質與內容提供意見，以增刪題目。
4. 進行第一次預試：抽取適當之樣本進行預試，同時蒐集樣本對題目理解情況及作

答之反應。

5. 分析第一次預試資料：根據古典測驗理論之項目分析原則，計算題目之難度、鑑別度等資料，同時亦蒐集預試樣本學業成績及其它測驗資料，計算其相關係數，並依受試者的反應記錄，一併做為修訂題本之參考。
6. 進行第二次預試：抽取第二次預試樣本，施以修訂後預試之題本，同時估算各分測驗使用之時間。
7. 分析第二次預試資料：根據古典測驗理論分析題目之難度、鑑別度指數，並依現代測驗理論，以BILOG套裝軟體估計其難度參數、鑑別度參數、猜測參數，再根據該等指標修刪題目編訂正式題本。
8. 進行正式施測工作：依分層群聚（以班為單位）方法，抽取常模樣本進行正式施測工作，並根據測驗結果進行能力估計。
9. 建立常模：根據常模樣本資料建立標準分數常模。
10. 進行信、效度考驗：依據古典測驗理論進行重測、內部一致性及同時效度與因素分析等考驗。
11. 撰寫研究報告及測驗指導手冊。

本研究自確定架構、編擬題目至建立常模，四個版本共計以32個月時間完成。為確實符合國民中小學之需要，於編題及施測階段邀請數名中小學教師共同研討論議、提供諮詢意見，以便適時修正。研究進行過程中：於測驗架構之擬定、題目之編擬、預試資料分析、正式題本之編定、信度及效度考驗、以及常模建立等階段，均分別邀請二至三位專家審查，以專家會議方式提昇上述各項工作之品質，使整個編製結果臻於完善。

三、研究對象

本研究所擬編製之測驗四個版本適用對象，其母群包括臺灣地區國民中學學生（約120萬人）、國小中高級學生（約154萬人）、及國小一、二年級學生（約76萬人）；預試與常模樣本之選取，係依性別、年級、地區、及城鄉等因素及各因素之比例，以班級為單位進行多階段分層群聚取樣。抽取之樣本數如表一，其中國小低年級因係採個別或小組方式施測，因此取樣人數較少。

表一 各版樣本人數

版 本	第一次預試	第二次預試	常模
國中版	587	738	1932
國小高年級版(甲式)	793	897	1061
國小中年級版(乙式)	785	883	1033
國小低年級版(丙式)	296	359	909

此外，為考慮智能不足及資賦優異兩極端學生之樣本代表性問題，亦抽取若干啓智班及資優班學生為預試樣本，納入項目分析資料中，做為修定題目之參考。

四、研究工具

本研究除編製測驗外，為考驗古典測驗理論所強調之關聯效度，除以學業成績為效標外，另選取國內目前較常使用之國民中學智力測驗（第三種）、國民小學系列學業性向測驗、瑞文式彩色圖形推理測驗等為效標工具進行同時效度考驗。

參、研究結果

一、各版測驗內容

本研究自八十二年九月起至八十五年六月歷時三年，共完成國民中學團體語文智力測驗等四種版本。本系列測驗的目的在評量國小至國中階段學生的一般認知能力，包含：記憶、理解、推理及評鑑等智能運作層次。總括四個版本有八種不同之分測驗，延宕記憶屬記憶部分，歸類測驗、數字系列屬理解部分，算術推理、數的比較、語文類推、邏輯推理屬推理部分。惟測驗結果係以總分呈現。各版之內容如表二所示：

表二 各版分測驗名稱

國民中學	國民小學		
	甲式(高年級)	乙式(中年級)	丙式(低年級)
1. 算術推理	1. 歸類測驗	1. 歸類測驗	1. 歸類測驗
2. 語文類推	2. 算術推理	2. 算術推理	2. 數的比較
3. 數字系列	3. 延宕記憶	3. 延宕記憶	3. 延宕記憶
4. 延宕記憶	4. 語文類推	4. 語文類推	4. 替代比較
5. 邏輯推理	5. 數字系列	5. 數字系列	5. 語文類推
	6. 邏輯推理	6. 邏輯推理	

二、信度考驗

四個版本之信度考驗，包括內部一致性係數及再測信度。大致而言，內部一致性係數在 .70 ~ .80 之間（見表三），與其他一般智力測驗之信度係數相若。表四則為數班再測信度係數之中數值，各分測驗之穩定性大致在 .50 ~ .70 之間，然總分之再測信度則在 .75 以上，由於本測驗之得分係以總分解釋，故整體而言，本測驗之結果尚稱穩定。

表三 各版分測驗之內部一致性係數(α)

國中版(n=1932)		國小甲式(n=1061)		國小乙式(n=1033)		國小丙式(n=909)	
分測驗	信度	分測驗	信度	分測驗	信度	分測驗	信度
算術推理	.730	歸類測驗	.794	歸類測驗	.766	歸類測驗	.817
語文推理	.768	算術推理	.725	算術推理	.740	數的比較	.742
數字系列	.779	延宕記憶	.793	延宕記憶	.760	延宕記憶	.776
延宕記憶	.833	語文類推	.707	語文類推	.623	替代比較	.809
邏輯推理	.770	數字系列	.713	數字系列	.660	語文類推	.742
總分	.854	邏輯推理	.735	邏輯推理	.742	總分	.914
		總分	.931	總分	.919		

表四 各版分測驗之再測信度

國民中學(n=184)		國小甲式(n=78)		國小乙式(n=66)		國小丙式(n=33)	
分測驗	信度	分測驗	信度	分測驗	信度	分測驗	信度
算術推理	.667	歸類測驗	.778	歸類測驗	.610	歸類測驗	.789
語文推理	.793	算術推理	.522	算術推理	.713	數的比較	.724
數字系列	.651	延宕記憶	.735	延宕記憶	.817	延宕記憶	.614
延宕記憶	.788	語文類推	.499	語文類推	.654	替代比較	.778
邏輯推理	.842	數字系列	.465	數字系列	.598	語文類推	.642
總分	.908	邏輯推理	.504	邏輯推理	.751	總分	.913
		總分	.757	總分	.888		

*間隔時間二~四週

三、效度考驗

本測驗效度係數之考驗包括與其他智力測驗及學科成績之效標關聯效度、以及因素分析之建構效度。表五~表六為各分測驗與其他智力測驗及學科成績之相關情形，均分別以其中數為代表值。與其他智力測驗之相關介於.248~.638之間，總分之相關則在.470~.750之間，顯示本測驗與其他團體智力測驗有若干程度之近似性，惟仍有其獨特之處。

效標關聯效度之相關係數值介於.336~.536之間，總分之相關則在.494~.628之間，均達統計上顯著水準。惟使用學科成績為效標時，因僅能以班級為單位，受樣本較少、全距較小之限制，其係數可能較低，惟尚能說明本測驗與學業成績間具有某種程度之關連。

表五 各版分測驗與其他智力測驗之效標關聯效度

國民中學(n=76)*		國小甲式(n=84)**		國小乙式(n=84)**		國小丙式(n=74)***	
分測驗	效度	分測驗	效度	分測驗	效度	分測驗	效度
算術推理	.509	歸類測驗	.638	歸類測驗	.596	歸類測驗	.473
語文推理	.409	算術推理	.581	算術推理	.502	數的比較	.294
數字系列	.248	延宕記憶	.492	延宕記憶	.527	延宕記憶	.257
延宕記憶	.417	語文類推	.526	語文類推	.600	替代比較	.403
邏輯推理	.371	數字系列	.404	數字系列	.503	語文類推	.341
總分	.523	邏輯推理	.496	邏輯推理	.610	總分	.470
		總分	.750	總分	.744		

* 國中版之效標為國民中學智力測驗(第三種)

** 國小甲式和乙式之效標為國民小學系列性向測驗

*** 國小丙式之效標為瑞文式彩色圖形推理測驗

表六 各版分測驗與學校成績之效標關聯效度

國民中學(n=494)*		國小甲式(n=227)**		國小乙式(n=228)**		國小丙式(n=198)***	
分測驗	效度	分測驗	效度	分測驗	效度	分測驗	效度
算術推理	.505	歸類測驗	.463	歸類測驗	.418	歸類測驗	.530
語文推理	.446	算術推理	.415	算術推理	.354	數的比較	.536
數字系列	.340	延宕記憶	.365	延宕記憶	.323	延宕記憶	.465
延宕記憶	.363	語文類推	.393	語文類推	.357	替代比較	.473
邏輯推理	.431	數字系列	.336	數字系列	.398	語文類推	.430
總分	.494	邏輯推理	.453	邏輯推理	.455	總分	.628
		總分	.572	總分	.585		

* 國中版之效標為國文、英文、數學、歷史、地理、生物、公民、健康教育等

** 國小甲式和乙式之效標為國語、算術

*** 國小丙式之效標為國語、算術

因素分析之結果見表七~表十，各版所含分測驗雖有不同，惟大致可歸為兩類因素，即延宕記憶、歸類測驗、語文類推、邏輯推理屬第一因素群--語文部分，而數字系列、算術推理、替代比較或數的比較等為第二因素群--算術部分，與原設計之架構契合，亦即題本內容區分為語文與算術兩大部分。

表七 國中版因素分析結果

分測驗	第一因素	第二因素
延宕記憶	.975	.143
語文類推	.663	-.296
邏輯推理	.618	-.332
數字系列	.066	-.944
算術推理	.158	-.787

表八 國小高年級版因素分析結果

分測驗	第一因素	第二因素
延宕記憶	.863	.155
歸類測驗	.771	.344
語文類推	.747	.373
邏輯推理	.690	.414
數字系列	.216	.911
算術推理	.457	.735

表九 國小中年級版因素分析結果

分測驗	第一因素	第二因素
延宕記憶	.851	.147
歸類測驗	.743	.340
邏輯推理	.708	.426
語文類推	.699	.359
數字系列	.218	.890
算術推理	.404	.740

表十 國小低年級版因素分析結果

分測驗	第一因素	第二因素
語文類推	.853	.174
延宕記憶	.786	.255
歸類測驗	.631	.375
數的比較	.184	.876
替代比較	.347	.742

四、常模

研究者考慮各分測驗之題數及難度有所不同，故決定將各分測驗之原始分數，轉換為標準分數（ $M=13$ ， $SD=4$ ），再以各分測驗標準分數之總合，轉換為離差智商（ $M=100$ ， $SD=15$ ）。此外，經變異數分析結果，發現各版均有年級上顯著差異的情況，而國中版在性別上亦有顯著差異，故各版均建立有年級常模，國中版則特別提供不同性別及年級之常模對照表。

肆、結論與建議

一、本測驗的功能

本測驗具有以下功能：

1. 評量一般學童智能發展情形，測驗結果可配合其他測驗或非正式評量資料（如：觀察資料、作業分析等），作為學習與相關輔導之參考。
2. 四種版本團體語文智力測驗可用於國小至國中階段特殊學生之篩檢、個案學生學

業性向評估，測驗結果可配合其他測驗（如：個別智力測驗、適應行為量表、其他能力或性向測驗）及非正式評量資料，作為特殊學生教育安置、教育輔導或訂定個別化教育方案之參考。

3. 由於本系列測驗均以「記憶、理解、推理及評鑑」等智能運作層次，作為各水準測驗之架構，其中有二個分測驗各版均有（語文類推、延宕記憶），有三個分測驗是前三版共有（算術推理、數字系列、邏輯推理），由於其題型與功能相似，或可作為學生發展性研究的參考工具。
4. 國民小學中高年級兩個版本，六個分測驗均有共同校準試題(anchor items)進行連結(linking)，並以項目反應理論之能力參數(θ)與難度參數(b)為參考依據，分成甲乙(中高)兩版；因此，國民小學中高年級在分測驗的能力量尺上具有相當的連貫性。

二、測驗結果之解釋與運用

本系列測驗為團體語文智力測驗，在必要時需將測驗實施方式加以調整；另本測驗亦與多數團體語文智力測驗可能具有的限制相似，茲將施測的必要調整與限制一併說明如下：

1. 研究初始架構乃欲設計適用於國小低年級學生之團體語文式智力測驗，但於編製過程中卻發現有些許困境，因一年級學生方才稍具簡單閱讀能力，為顧及其發展階段與理解程度，因此測驗呈現型態無法全部純粹以語文表示之，故其中有分測驗以圖形呈現，並輔以文字說明，似乎不宜以語文式智力測驗統稱之。
2. 語文智力測驗可能受文化刺激、語文成就、學校學習經驗等因素影響，而對文化不利、語文成就低落的受試者較為不利，這些受試的智能發展水準可能在本測驗中被低估。對這些學童不可單獨以本語文智力測驗推估其一般能力，宜再選擇非語文智力測驗評估之。
3. 無論施測本系列測驗中何種水準的測驗，對於有兩個以上分測驗原始分數（答對題數）為0分的受試者，測驗使用者宜考慮本測驗所需的反應能力，可能不適合該受試者，如：本測驗的基準對部分兒童可能太高，對於在此基準以下的學童可能無法有效評量，對於這些兒童宜選用其他測驗。
4. 本測驗之施測方式，國小一年級宜採個別施測、二年級可採小團體施測、國小三年級至國中階則採團體施測，施測所需時間約為五十分鐘。對於不適合實施團體測驗的學童，無論在任何年齡階段，主試均應採取個別施測方式，並應將施測方式作必要調整。例如：對於過動或注意力無法持續五十分鐘的學童，應採取個別施測，並考慮分段完成（施測中間給予較長的休息時間，甚至分兩天完成測驗）。

此外，由於本測驗需經由視動管道接收與回答測驗題目，如果受試者有明顯的視覺或視動能力缺陷，可能影響其在本測驗的表現，此時使用者除可考慮選用測驗外，也可以採取較彈性的施測方式，如：採個別施測，對於有特殊困難者由主試唸出題目或代為書寫答案等，惟採取替代方式施測，宜作記錄。

5. 國小低年級版實施對象為一年級學生時，施測者應具備嚴謹之測驗基本概念，因一年級學生之施測方式為施測者朗讀題目，而受試者以手勢表示答題結果，若稍有疏失，可能會有暗示現象發生，影響施測結果之可信度。另一年級學生之施測方式為一對一個別施測，施測情境之營造亦影響其結果。
6. 本測驗四版本常模建立時間為下學期開學後一至二個月（三月至五月間），日後此測驗之運用時機，為求測驗結果之效度，故各版本測驗使用時機，應於下學期施測較為適宜。
7. 由於本測驗係初步完成，實際運用的效果仍待後續的效度研究。本測驗雖可得到五項分測驗分數，但編製者建議本測驗使用者勿由分測驗結果，對受試者的內在能力作過細的差異分析，或只根據一項分測驗解釋受試者的能力，如果受試者顯現在各分測驗明顯的差異，測驗使用者也僅能將這些視為內在能力差異的癥兆或假設，必須經由其他評量方式驗證之。
8. 本測驗國小低年級版本（丙式）之效標關聯效度考驗結果未臻理想，可能與國小低年級學生學業成績之評量方式有關。現今國小低年級學生學業成績之評量方式各校似乎有極大之彈性空間，有些學校採用紙筆測驗；另有學校以以平時教師觀察為主；亦有學校以學生平時活動操作為評量項目；因此並無統一之標準測驗或評量方式，可能偏向教師主觀的評分結果，故影響與本測驗之效度考驗結果。
9. 變異數分析考驗結果，除國中版性別與年級交互作用達顯著水準外，國小部分（高、中、低年級）均未達顯著差異，查有關智力發展之文獻，有男女生於幼兒時期，其語文及數字能力之差異並不明顯，直至中學階段，才因後天環境（學習因素）而逐漸產生差異之現象，本研究之發現與此頗為相符。

三、未來研究發展方向

本研究歷經三年時間完成國中小四個版本團體語文式智力測驗，雖已具相當良好之信度與效度，然於實用上、效度結構上可再加強研究發展之方向為：

1. 進行試探性和驗證性因素分析研究，評估本測驗的內部效度，並進一步探討本測驗的編製與理論依據之間的關係。
2. 將本測驗四版本聯結設計出適用於特殊學生之版本，建立適用於特殊教育學生的常模：本測驗若能建立各類障礙學生及輕度、中度、和重度智能不足等特殊教育

對象常模，可增加特殊教育篩選、診斷的正確性。

3. 進行預測性區別分析，探討不同學科（包括國語、數學、自然、社會）與各分測驗的關係，特別是難界定潛在特質(latent trait) 不明確的科目，如：社會、生活與倫理、自然等。
4. 探討本測驗的線性組合分數與學科成績的線性組合分數間的典型相關分析，藉以進一步了解本測驗的建構效度。
5. 另取適當樣本施測，使用共同校準試題進行垂直等化(vertical equating)之處理，建構從國小一年級至國中三年級間能力參數(θ)之指標，做為電腦適性智力測驗的依據。

參考書目

- 林幸台、張正芬、陳美芳、王寶墉、董媛卿、王純娟、白政民(民80)，**特殊教育機構使用心理與教育測驗之現況與評估研究**。臺北：教育部教育研究委員會。
- 周台傑、吳訓生(民82)，**特殊教育評量工具需求之研究**。臺北：教育部社會教育司。
- 翁麗禎(民82)，**編製推理思考測驗**。教育部輔導工作六年計畫研究報告。
- 陳李綢、林清山(民80)，**多重模式智力測驗的編製報告**。**測驗年刊**，38輯，151-170頁。
- 路君約(民81)，**心理測驗(下)**。臺北：中國行為科學社。
- 蔡崇建(民80)，**智力的評量與分析**。臺北：心理出版社。
- Derr, R.L. (1989). Insights on the nature of intelligence from ordinary discourse. *Intelligence*, 13, 113-118.
- Finch, F.L. (Ed.) (1991). *Educational performance assessment*. Chicago: Riverside.
- Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. NY: Basic Books.
- Gardner, H. & Hatch, T. (1989). Multiple intelligences go to school: Educational implications of the theory of multiple intelligence. *Educational Researcher*, 18(8), 4-10.
- Jensen, A.R. (1980). *Bias in mental testing*. N.Y.: Free Press.
- Lidz, C.S. (Ed.) (1987). *Dynamic assessment*. N.Y.: Guilford.
- Matarazzo, J.D. (1992). Psychological testing and assessment in 21st century. *American Psychologist*, 47, 1007-1018.
- Naglieri, J.A. & Das, J.P. (1990). Planning, attention simultaneous and successive (PASS) cognitive process. *Journal of Psychoeducational Assessment*, 8, 303-337.
- Nickerson, R.S. (1989). New directions in educational assessment. *Educational Researcher*, 18(9), 3-7.
- Sattler, J.M. (1988). Historical survey and Theories of intelligence. In J.M., Sattler: *Assessment of children (3rd ed.)*, pp. 37-60. San Diego: J.M. Sattler, Publisher.
- Sternberg, R.J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. N.Y.: Cambridge University Press.
- Sternberg, R.J. (1992). Ability tests, measurements, and markets. *Journal of Educational Psychology*, 84, 134-140.

- Sternberg, R.J., Wagner, R.K. (1986). *Practical intelligence: Nature and origins of competence in the everyday world*. Cambridge, England: Cambridge University Press.
- Wainer, H. (1993). Measurement problems. *Journal of Educational Measurement*, 30, 1-21.
- Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 70(9), 703-710.

林幸台，美國喬治亞大學教育博士，現任國立台灣師範大學，特殊教育系教授