

信心加權對測驗分數信度與效度之影響 林邦傑

(作者現為本校文理學院教育學系專任副教授)

壹、前言

近年來，臺灣工商進步，人民生活富裕，各級學校畢業生欲繼續升學者甚衆。許多學生在升學競爭激烈之情況下，往往由於一分之差，而無法進入理想的學府，有的甚至還因之名落孫山。因此聯考試題之是否合理及計分方式是否客觀公平，乃成為大眾最關注之事。聯考之論文式試題雖具有某些優點，但由於閱卷費時，評分主觀，因此漸為客觀式試題 (objective tests) 所取代，尤其自從六十二學年度大學聯考開始採用電腦閱卷以來，客觀式試題更成為大專聯招及各縣市中學聯考試題的主幹。在一般中小學的課室教學中，也大行其道。

在客觀式試題中，最常見的是選擇式的試題 (Multiple-choice test)。它通常是由一個題幹 (stem) 以及幾個可能的選目 (option) 所組成。在這些選目中，只有一個是正確答案 (answer)，其他幾個都是錯誤的答案。有時候，也有所謂複選題的出現，每一題幹後面所列的選目中，可能有二個或二個以上是正確的答案。假如受試者所挑選的答案完全正確，即得一分；空題不答，得零分；答錯了，則得零分或負分。學生在測驗上答對的總題數，即為其在該測驗的總分。這類試題具有下列優點：1.評分客觀，不因不同評分者而有所影響，2.題目之取樣較廣，效度較高，3.評分快速簡便，可用機器代替人工評分。因此在大規模的考試中可縮短評閱試卷時間。它的重大缺點，則為兒童的測驗成績容易受學生猜題 (guessing) 行為的影響。假如參加考試的每個學生能力相等，猜題的次數也一樣，那麼猜題所造成的影響並不嚴重。不過心理學家發現兒童的能力不但有很大的個別差異，而且每個人的猜測行為也非常不同 (Granich, 1931; Jackson, 1955; Slakter, 1967; Swineford, 1938, 1941; 吳靜吉, 1974)。有些受試者，即使主試人員告訴他一定要猜，他遇到沒把握的題目寧可空著也不肯猜。有些受試者您不准他猜，猜錯了要受懲罰，他仍然每題都做。因此試題評分時，如不針對兒童的猜題行為加以校正，那麼喜歡猜題的學生就比不猜的學生佔很多的便宜。Davis (1951) 的研究可顯示出這問題的嚴重性。他以納廸閱讀測驗乙式 (Form B of The Nelson-Denny Reading Test) 測量 393 名中學生。根據測驗結果從中挑取成績最差的 47 名，告訴他們下次測驗時，不管是否知道正確答案如每題都回答，比較可能得高分。隔幾天他以同一測驗之複本 (即該測驗之甲式)，再度測量這些學生，結果發現其前後二次之平均數，相差將近一倍 (由 25.53 分增加到 46.32 分)。由此可見兒童猜題行為所造成第二次測驗成績之巨幅增加與其能力沒什麼關係，但與其人格特質却有密切關係 (Swineford and Miller, 1953; Hamilton, 1950; Sherriffs and Boomer, 1954; Votaw, 1936; Ziller, 1957)。猜題行為造成了測量結果的誤差，如何加以校正，一直為教育及心理學者所重視。目前國內外考試在這方面最常用的校正公式如下：

$$S = R - W / (k - 1)$$

S 代表某一受試者校正後的分數

R 為其答對的總題數

W 為其答錯的題數

K 為每一試題可選的答案數

此一公式所隱含的假設是，凡學生答錯的題目都是猜錯的，而學生答對的題目則有二種不同的情況，一是受試者真正知道正確答案，一是受試者完全不知道，但由於運氣好，碰巧猜對了。因此計分時須根據機遇率每答錯一題要倒扣 $1/(k-1)$ 分 (Davis, 1959; Lord and Novick, 1968; Thorndike, 1971; Diamond and Evans, 1973; Lord, 1975)。

事實上，大部分兒童對同一試題所列的各種選目，並非都具有同樣程度的了解和信心，他們很可能對其中一二個選目比較有信心，對其他的選目則一無所知。因此他在答題時往往先將有把握的選目剔除，再對其他選目加以猜測。像這類滲雜部份知識 ((partial knowledge) 的猜題與盲目的猜題 (blind guessing) 有很大的差別，因此任何根據盲目猜題的假設，所建立起的校正公式，都很難令人滿意 (Lord, 1975)。國內大專及高中聯招試題都是在考前幾天由聯招委員會臨時聘請學科專家與教師命題的，由於時間短促，及心理測驗知識的不足，每一題目上所列的選目，並非都具有同樣的吸引力 (attractive)。有些選目往往一看就知道不可能是正確答案。在這情況下大專聯招目前所採用的校正公式，更不適宜了。像這類部份知識對兒童考試分數影響之大，由Little and Creaser(1966) 的研究可明顯的看出。他以三選一的選擇式試題測驗一百名學生，要其在每個題目上都表示是否知道正確答案，是否純是猜測。為了使學生坦白回答，沒有顧忌起見，考前告訴學生這次考試只供研究之用，答錯了並不扣分，結果發現學生很有把握的題目上所做的反應，有百分之九十三是正確的，在沒有什麼把握的題目上，有百分之六十七答對了，在純粹猜測的題目上有百分之五十五答對。從機遇率的觀點言，假如受試者對題目內容完全不知道，則其猜中的機率應是百分之三十三，而非百分之六十七和百分之五十五。Jackson (1955) 以五選一的選擇題測驗學生，也發現學生在純粹猜測的題目中，答對的比率是三分之一，而非機遇率所謂的五分之一。因此即使使用傳統的機率校正公式，喜歡猜題的人仍比不敢猜題者佔很大便宜。

爲了考慮部分知識對兒童測試成績之影響，心理學家想出一些新的計分方式，試圖取代以答對分數爲總分，或以機率公式校正的傳統計分法。這些新計分法中最爲晚近心理學家及教育學者感興趣的是信心測驗法 (Confidence Testng) (Echternacht, 1972)。信心測驗法與一般傳統方法最大的不同是，計分時不但考慮兒童之答案是否正確，還要考慮其對自己所做答案的信心程度，信心愈高，得分也愈高。信心測驗法種類很多，大略可分爲兩大類：一爲信心加權法(Confidence Weighting procedures)，一爲個人機率計分法 (the personal probability approach)。前者是要受試者在一組選目中選出一個正確答案，並對所選出的答案的正確性表示自己的信心程度。二個學生在某一題目上，雖有同樣答案，但如二者對自己答案的正確性，信心程度不同，則信心高的人得分較多，信心低的人得分較少，有時甚至不計分。個人機率計分法也和信心加權法一樣是根據信心程度來調整受試者得分之高低，不過它要求受試者標明每個選目成爲正確答案之或然率多高。假如，一個題目有五個選

目，則受試之得分即由其在這五個選目上所做的各種反應來決定。因此它比同一題目用信心加權法填答及評分，要花更多的時間，計分的方式也更為複雜，通常皆需大型電腦襄助。因此這方法在臺灣目前的情況下尚無實用價值。而信心加權法至少在理論上言是刪減猜題行為對測試分數之影響的一種比較精密合理的方法，其計分方式却較個人機率計分法簡單，其應用價值頗值得國人做進一步的探討。

信心加權法（以後簡稱 CW 法），在客觀性試題上的運用是自 1930 年代開始。早期的研究有 Hevner (1932), Soderquist (1939)，和 Dressel and Schmid (1953) 等，這方法後來就不再引起心理學家的重視，直到心理測驗專家 Ebel 在 1961 和 1965 年的研究，發現 CW 法可提高測驗的信度，這方法才再度激起學者和專家的強烈興趣。有的心理學家甚至籌組公司，設計信心測驗計分工具 (Valid Confidence Testing Demonstration Kit, 簡稱 VCTDK)，對外出售 (Shuford and Massengill, 1967)。雖然 VCTDK 是否可提高測驗之信度與效度，學者看法並不一致，不過很多人主張 CW 法應在課堂教學及各種考試上大量試用 (Stokes, 1966)。有關 cw 法對測驗信度及效度之影響的研究，目前正方興未艾，茲將過去有關研究摘述於後：

表一 過去學者有關信心加權法 (cw) 信度之研究結果摘要表

受試者	人數	測驗學科	信 度			研 究 者
			C W	傳統	計算方法	
大學生	100	音樂欣賞	.76	.67	折半法	Hevner (1932)
大學生	100	音樂欣賞	.56	.56	折半法	Hevner (1932)
大學生	62	美術判斷	.85	.72	折半法	Soderquist (1936)
大學生	90	物理	.73	.70	KR20	Dressel and Schmid (1953)
中學生	834	字彙	.73	.72	KR20	Coombs, et al (1956)
中學生	30	教育測量	.71	.57	KR20	Ebel (1961)
中學生	30	教育測量	.83	.77	KR20	Ebel (1961)
中學生	30	教育測量	.82	.73	KR20	Ebel (1961)
小學生	57	社會學科	.67	.78	副本法	Archer (1963)
小學生	49	社會學科	.69	.83	副本法	Archer (1963)
大學生	79	教育測量	.85	.82	KR20	Ebel (1965)
大學生	78	教育測量	.79	.79	KR20	Ebel (1965)
大學生	74	教育測量	.86	.85	KR20	Ebel (1965)
中學生	432	社會學科	.84	.76	折半法	Michael (1968)

中學生	22	寫 作	.70	.53	Hoyt	Rippey (1969)
中學生	65	化 學	.89	.81	再測法	Armstrong and Mooney (1969)
中學生	65	化 學	.86	.82	折半法	Armstrong and Mooney (1969)
中學教師	73	教育測量	.66	.71	折半法	Hambleton, et al (1970)
研究生	63	統 計 學	.92	.88	折半法	Hopkins, et al (1973)
中學生	94	智力測驗	.91	.88	折半法	Wen (1975)

由表一可知對於同一試卷，以信心加權法計分所得的信度似乎有比傳統計分法高的傾向。只有Hambleton, et al (1970) 的研究結果例外。

至於信心加權法對效度的影響，有二種不同的看法，在1970年以前，大部份有關學者雖未明言，但似乎都認為由於信心加權法可增加測驗的信度，故亦可間接增進測驗之效度。不過很少學者實際做研究加以證實。Hambleton, Roberts, and Traub (1970) 發現兒童在學科成就測驗上信心加權法所得的分數與效標（即期末考試的成績）之相關 ($r = .72$) 高於傳統計分法與同一效標之相關($r = .62$)。一九七〇年以後，有些學者對此採不同的看法 (Wang and Stanley, 1970; Stanley and Hopkins, 1972)。譬如，Hopkins, Hakstian and Hopkins(1973)即認為信心加權法雖可使測驗信度增加，但這增加是由於受試者某些反應模式(Response style)所造成的。這些反應模式與兒童的能力無關，故其雖可增進信度，却對測驗之效度有害。他以63名研究生在統計學測驗上的成績加以研究，結果證實信心加權法的效度 ($r = .667$) 確低於傳統計分法 ($r = .701$)。綜合過去的研究吾人可發現，一九六八年以前，有關信心加權對測驗分數的影響，似乎都偏重於信度的研究上，一九六八年以後，則開始著重於效度的研究，有的學者認為信心加權可提高測驗的效度，有的則認為反而有害。不過這二派看法尚乏有力的證據支持，因為在其研究的結果中，傳統方法和信心加權法的效度係數表面上看起來雖有差異，但這些差異都未達顯著水準 ($p > .05$)。因此信心加權法的效度，尚待學者進一步的探討。由於信心加權分數是由兒童答案的正確性和其信心分數所組成。因此信心分數之信度，及其和其他因素之關係亦為信心加權法成敗的重要關鍵。

過去學者的研究發現，兒童的猜題行為和其性別，測驗指導語，及人格特質有關。一般而言，男孩比女孩喜歡猜題與冒險 (Swineford, 1938; 吳靜吉, 1974)。個性內向、自卑、順從及小心的兒童對沒有把握的題目常空著不答。如果考試時不准學生猜，則這些學生之成績通常比大膽，敢於冒險的學生來得低 (Votaw, 1936; Sherriffs and Boomer, 1954; Slakter, 1968, 1969)。此外學者亦發現成就動機高的學生測驗分數較高 (Anastasi, 1958; Yamamoto and Dizney, 1965; 林邦傑, 1971)，兒童測試焦慮通常和智力及成就測驗之成績呈負相關（約在-.1到-.3之間）(Ruebush, 1963; French, 1962; Allison, 1970; Chambers, Hopkins, and Hopkins, 1972; Silverstein, 1964; 李克明，黃國彥，1976)

Witkin (1954, 1962) 認為人類的認知模式 (Cognitive style) 有二；一是場地獨立

型 (Field-independent) , 一是場地依賴型 (Field-dependent) 。場地獨立型的人，比較能够區分視覺界中的物體與背景，場地依賴型的人，在這方面能力較差。這種區分的能力通常是一種連續的現象，其一極端為場地獨立型，亦稱分析的場地傾向 (Analytical field approach) ，這種類型的人智力較高，個性較主動，獨立和自我信賴。另一極端則為場地依賴型，又稱籠統的場地傾向 (Global field approach) ，這種人較被動，順從，缺乏自信，並有顯著的焦慮傾向 (Vernon, 1962) 。

前述的性別、測驗指導語、人格特質及認知模式既與兒童的猜測行為及成就測驗成績有關，它們與信心分數間可能也會具有類似的關係存在。

本文研究之目的在探討 1. 信心加權對測驗信度與效度的影響，2. 信心分數與人格特質及認知模式之關係，3. 性別和指導語對信心加權分數的影響。

貳、研究方法

一、受試者

本研究受試對象係由民國六十五年二月臺北縣永和國中二年級四個班 220 名學生所組成，其中男女各半。有關本研究受試者之選取及分派，將於測驗的程序及計分方法中詳談。

二、測量工具

1. 中國兒童測試焦量表 (Test Anxiety Scale for Chinese Children, 簡稱 TASCC)：本量表係由林碧峯、楊國樞、繆瑜、和楊有維等 (1973) 根據 Sarason 等人 (1960) 所編的「兒童測試焦量表」 (Test Anxiety Scale for Children) 加以修訂而成，用以測量兒童在某些類似測驗或考試的情境下的焦慮情形。本量表之再測信度和折半信度在 .70 至 .98 之間；效度係數也尚令人滿意。全部題目三十七題，其中九題是測謊量表的題目，用來測量學生說謊的傾向。

2. 基氏人格測驗自卑感量表：本量表係賴保禎先生 (1976) ，根據 Guilford 等所編的「人格測驗 I 量表」 (Personality Inventory, I Scale) 修訂而成，全部量表共十題，用以測量兒童的自卑感 (Inferiority Feeling) 。分數愈低，表示愈缺乏自信，對自己評估過低，易受別人影響。本量表之折半信度為 .82 ；再測信度為 .72 ，其對學生自卑感之鑑別力尚佳。

3. 藏圖測驗：本測驗係由吳靜吉博士 (1974) 根據 Messick (1962) 的「藏圖測驗」 (Hidden Figure Test) 翻譯而成。整個測驗包含 32 題，用以測量兒童的場地獨立性 (field-independent) ，測驗時間 20 分鐘。藏圖測驗是 Messick 仿照 Witkin 等的「圖形隱藏測驗」 (The Embedded Figures Test) 編製而成，不過施測手續較前者簡易。本測驗之信度、效度尚佳。

4. 成就動機測驗：本測驗係美國米蘇里大學羅素 (Russell) 教授所編製，羅氏先邀請二十四位經驗豐富的中學教師，根據學校的生活情境，編撰五十個題目，用以測量學生的學業成就動機和努力的意願。然後將這些問題測驗美國肯塔基州某市中學學生 100 人，根據項目分析結果將區別力低的題目淘汰，保留了二十六題，後來羅氏又添加了三十二個題目，以之

測量該州某鄉村中學學生 100 名，再根據項目分析結果，將區別力較高的題目保留，編成本測驗，此一測驗由林邦傑 (1971) 首先在國內試用，本測驗之折半信度係數，在 .94左右，其與加州成就測驗 (The California Achievement tests) 之相關為 .705。羅氏又將受試者分為兩組，一組是志願參加學業競賽的，一組是非志願參加的。結果發現志願組在學業成就動機上的分數顯著的高於非志願組。可見本測驗能够把學生的競爭心和努力意願鑑別出來。

5. 永和國中二年級數學成就測驗：本測驗係由臺北縣永和國中和國立師範大學教育系合作編製的。原測驗包括一百道四選一的選擇題，測驗時間為 70 分鐘。由於本研究目的之一在了解傳統計分方式和信心加權計分方式對測驗信度的影響。而本文所用的信度係數係由折半法 (Split-half method) 估計而得。折半法不能用於速度測驗，為免測驗題目太長，影響了信度的估計，本研究只取其中八十道題目，作為研究之用。

為研究指導語不同，對兒童信心加權分數的影響，本研究特設計甲乙二種不同的答案紙，其中甲式只告訴受試者如何填寫正確答案，和表示對自己所做答案的信心程度。乙式除告訴受試者如何作答外，還告訴他們信心程度之高低，會影響其測驗成績。茲將各式的作法說明和指導語列於後：

1. 本測驗甲式作法說明

「本測驗每一題目在答案紙上均包括兩部分，左邊部分為答案欄，請將您認為正確的答案填答於此。本測驗涉及的範圍較廣，有些題目，您可能比較熟悉，回答起來比較有信心與把握，有些題目，可能比較不熟悉，或難於回答，作答起來，比較沒有把握與信心。請您在選擇正確答案時，順便在題目右邊表示您對自己的答案，有多少信心與把握。

例一：一小時有 60 分鐘，一分鐘有 60 秒，請問五小時共有 (1)12000 秒 (2)15000 秒 (3)18000 秒 (4)21000 秒

答 案				很 有 把 握 (1)	有 把 握 (2)	稍 有 把 握 (3)	沒 把 握 (4)
(1)	(2)	(3)	(4)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

在本題中，如您認為第二個答案是正確的，就在左邊答案欄第二個方格內打個「×」號，如對您所選擇的這個答案是否正確，稍有把握與信心，就在右邊第三個方格內打個「×」號。

每題在作答時務請表示，您對自己所選的答案的信心程度，不要遺漏。」

2. 本測驗乙式作法說明

本測驗乙式則除包含甲式之說明外，並添加下列詞句：

「每題在作答時，務請誠實地表示您對自己所選答案的信心程度，因為這會影響到您分數

的高低。例如張三和李四第一題均答對了，但張三在回答時表示很有把握，李四却覺得一點把握也沒有，則二人雖都答對了，但張三的得分將高於李四。相反的，如果二人第二題都答錯了，而張三對自己的答案很有把握，李四一點把握也沒有，則二人雖都答錯了，但張三扣分比較多，李四扣分比較少。

每題作答時，一定要表示您對自己選答的把握程度，不要遺漏。」

本測驗為本研究探討信心加權對測驗分數影響的主要工具。本測驗題數縮短及形式改變後之信度在.87到.92之間，其與師大教研所國中數學成就測驗之相關在.73到.85之間，與國中二年級上學期數學成績之相關在.47到.78之間。

6.師大教育研究所國中數學成就測驗：本測驗係由國立師範大學教育研究所（1972）編製。原測驗共有五十道四選一的選擇題，測驗時限40分鐘。本測驗之折半相關為.89，其與國中學生數學科學業成績之相關在.61到.84之間。本研究之重點是以永和國中數學成就測驗為工具，比較傳統計分法與信心加權計分法對信度及效度之影響，所用效標之一即為兒童在本測驗上之成績。但前述二種計分法從理論上言最大的不同是傳統計分法容易受猜測之影響，且忽略了兒童對測驗之部份知識。為了客觀起見，本研究所用的效標自應不可受兒童猜題的影響。因此筆者特邀請了臺北市二位資深的數學教師，把所有的題目由選擇式改為填充式，俾盡可能的減低猜題因素的影響。此外原測驗題目係於三年多以前根據當時國中一、二、三年級數學教科書編撰而成的，由於本研究的受試者只讀了一年級，及二年級上學期的數學，而且這三年多來，數學教科書已重新編訂，很多教材已更換，有些雖沒更換，但所用的名詞及符號，與前略有不同。因此筆者亦請這些國中教師根據新版本略加修訂，從中挑選出，目前國中數學教科書前三冊範圍內的題目共25題，測驗時間也相對的縮短為二十分鐘。本測驗改編後，與永和國中數學成就測驗之相關在.73到.85之間，與國中二年級上學期數學成績之相關在.53到.63之間。可見本測驗改編後，雖然試題數量削減了一半，其效度尚佳。

7.國中二年級上學期數學成績：永和國中二年級上學期數學成績是由幾次月考成績及平時成績計算而得。雖然該校月考採統一命題方式，不過題目多數為選擇題，且各班平時成績非由同一教師評閱，因此本成績可能受到兒童猜題，及教師評分標準不一的影響。不過由於本成績是許多次考試成績組合而成，而且一般教師、家長仍以之為評量兒童學習成就之重要根據，故筆者也將之列為重要之效標。

三、實驗設計

本研究共有二個獨立變項 (independent variables)，一為指導語，一為性別。指導語有二種，一種刊載於永和國中數學成就測驗答案紙甲式。拿到此種試卷的受試者，只知道如何填答試卷，而不知道信心分數對其測驗成績的影響，這組受試者稱為控制組 (control group)。另外一種指導語刊載於該測驗答案紙乙式。此種指導語，告訴受試者對自己所做的答案要誠實地表示其信心程度。信心分數的高低會影響其測驗之得分。這組受試者，視之為實驗組 (experimental group)，因此本研究之受試者共分為實驗組男生，實驗組女生，控制組男生，控制組女生等四組。本研究之因變項(dependent variables)有下列幾種：1.傳統分數，包括各信心選目下之傳統分數，及傳統總分，2.信心分數，3.信心加權分數。為探

討人格特質，認知模式與上述分數之相關，我們並求它們在各組組內之相關係數。

四、實驗程序：

臺北縣永和國中二年級共有二十五班，其編班方式有二種，一為男女合班，一為男女分班。兩種班數約略相同。由於行政上的困難，測驗實施時，常須以班級為單位進行。班級與班級間，由於能力背景等之不同，可能會影響了因變項，混淆了實驗結果。因此筆者在男女合班中隨機選取四個班級學生為受試者，將永和國中數學成就測驗答案紙甲、乙兩式，以隨機的方式混雜排列，分發各班填答。如此每班男女生都有同樣的機會被分派到本研究的各組中。筆者曾以同一方式在男女分開的班級中，選取四班。以這些學生組成實驗組和控制組。由於來自各班學生的比率一致（樣本之選取，各班一致），班級差異對這兩組成績的影響，可能不大。但由於是男女分開編班，則如同一組男女結果不同，究竟這是由於性別差異造成，或由實驗組男生來自男生班，女生來自女生班，班級本身差異造成，很難劃分。為求精確起見，本文的分析完全以永和國中二年級男女合班的四個班級 220 名學生為對象，男女各半。

五、計分方法

測驗完畢後，即進行閱卷及計分，永和國中數學成就測驗每份答案紙均有二種不同的分數。傳統分數總分是根據下列步驟計分而得的：1.答對一題得一分，2.空題不答者得零分，3.答錯了，以零分計算，不倒扣，4.兒童答對題數之總和，即為受試者的傳統總分。信心加權分數是根據下列步驟計分而得：1.答對一題得一分，答錯一題得分，空題不答得零分，2.凡兒童對自己所做的答案「很有把握」者得信心分數 4 分；凡覺得「有把握」者得 3 分；凡稍有把握者得 2 分；沒有把握得 1 分，3.答案正確（得 1 分）、錯誤（得 -1 分）、或空題不答（得 0 分）所得的分數乘以該題之信心分數，即為該題的信心加權分數。假如張三對自己所挑的答案很有信心（4 分），而且答對了（1 分），則其信心加權分數為 $4 \times 1 = 4$ 分，如答錯了（-1 分），則其信心加權分數為 $4 \times (-1) = -4$ 分。餘類推。4.兒童各題信心加權分數之總和，即為信心加權總分。5.如不論兒童答案之正誤，只求各題信心分數的總和，即為信心總分。至於其他測驗之計分方法，完全按照原測驗手冊之規定，不擬贅述。

三、結果

一、各組學生信心分數及各種數學分數的比較：

表二所列者為本研究實驗組與控制組男女學生在永和國中數學成就測驗上之傳統分數，信心加權分數，二年級上學期數學成績，及師大教研所國中數學成就測驗成績之平均數和標準差。至於各組學生上述成績之比較則列於表三至表十二。根據表三和表四的統計結果可知：各組學生在二年級上學期之數學成績和師大教研所國中數學測驗成績上都無顯著之差異。根據表五、表六、表七、表八、表九、表十、表十一、表十二分析的結果發現：各組在永和國中數學成就測驗上之傳統總分，沒有作答的題數和信心加權分數上彼此無顯著差異。惟女生在「沒有信心」選項下之傳統分數得分顯著地高於男生 ($P < 0.1$)，男生之信心分數

得分顯著地高於女生 ($P < 0.5$)。由此可見男女兩組之數學能力（指測驗總分）雖相等；但男生對自己的答案之信心較女生為高。

表二 各組男女學生在永和國中數學成就測驗上傳統分數、信心加權分數、二年級上學期數學成績、及師大教研所國中數學成就測驗成績之平均數及標準差

受試組別	受試性別	平均數及標準差	永和國中數學成就測驗								二年級上學期數學成績	師學大成教研究所國中成績		
			不同信心選項中之傳統分數				沒有作答的題數	傳統總分	信心分數	信心加權分數				
			很有信心	有信心	稍有信心	沒有信心								
實驗組	男	\bar{X}	31.19	11.51	5.88	1.46	0.42	50.04	256.96	89.16	81.98	8.67		
		SD	19.81	10.34	6.49	2.06	0.85	11.81	44.00	77.29	10.43	4.38		
	女	\bar{X}	29.25	11.14	6.78	2.94	0.67	80.96	244.18	88.11	80.96	7.45		
		SD	14.54	6.06	5.40	3.35	1.88	9.55	35.38	70.81	9.55	4.46		
控制組	男	\bar{X}	30.18	12.43	6.32	1.94	0.47	50.87	251.56	94.09	81.36	8.40		
		SD	18.21	10.80	7.51	2.40	0.97	11.46	38.02	78.14	9.86	4.63		
	女	\bar{X}	25.71	11.48	2.80	2.62	0.98	48.60	237.45	80.96	81.73	7.64		
		SD	17.11	9.53	7.30	3.94	1.72	10.19	56.30	66.56	12.81	3.89		

表三 各組男女生二年級上學期數學成績變異數分析摘要表

變異來源	離均差平方和	自由度	均方	F
A (性別)	7.62	1	7.62	0.07
B (指導語)	2.10	1	2.10	0.02
AB (性別×指導語)	24.38	1	24.38	0.21
S/AB (誤差)	25367.42	216	117.442	

表四 各組男女生在師大教研所國中數學成就測驗上成績之變異數分析摘要表

變異來源	離均差平方和	自由度	均方	F
A (性別)	54.01	1	54.01	2.81
B (指導語)	.12	1	.12	.01
AB (性別×指導語)	2.83	1	2.83	.15
S/AB (誤差)	4157.68	216	19.25	

表五 各組男女生在永和國中數學成就測驗傳統分數變異數分析摘要表

變異來源	離均差平方和	自由度	均方	F
A (性別)	66.55	1	66.55	.55
B (指導語)	6.23	1	6.23	.05
AB (性別×指導語)	75.64	1	75.64	.62
S/AB (誤差)	26190.59	216	121.25	

表六 各組男女生在永和國中數學成就測驗「很有信心」選項下傳統分數之比較

變異來源	離均差平方和	自由度	均方	F
A (性別)	564.96	1	564.96	1.81
B (指導語)	284.67	1	284.67	.91
AB (性別×指導語)	88.01	1	88.01	.28
S/AB (誤差)	67545.25	216	312.71	

表七 各組男女生在永和國中數學成就測驗「有信心」選項下傳統分數之比較

變異來源	離均差平方和	自由度	均方	F
A (性別)	23.96	1	23.96	.27
B (指導語)	21.83	1	21.83	.24
AB (性別×指導語)	4.63	1	4.63	.05
S/AB (誤差)	19316.66	216	89.43	

表八 各組男女生在永和國中數學成就測驗「稍有信心」選項下傳統分數之比較

變異來源	離均差平方和	自由度	均方	F
A (性別)	157.10	1	157.10	3.41
B (指導語)	83.22	1	83.22	1.81
AB (性別×指導語)	34.31	1	34.31	.74
S/AB (誤差)	9949.83	216	46.06	

表九 各組男女生在永和國中數學成就測驗「沒有信心」選項下傳統分數之比較

變異來源	離均差平方和	自由度	均方	F
A (性別)	64.15	1	64.15	6.86*
B (指導語)	.35	1	.35	.04
AB (性別×指導語)	8.80	1	8.80	.94
S/AB (誤差)	2020.44	216	9.35	

*P<.001

表十 各組男女生在永和國中數學成就測驗上「沒有作答」題數之比較

變異來源	離均差平方和	自由度	均方	F
A (性別)	4.37	1	4.37	2.10
B (指導語)	0.37	1	0.37	0.18
AB (性別×指導語)	0.04	1	0.04	0.02
S/AB (誤差)	448.58	216	2.08	

表十一 各組男女生在永和國中數學成就測驗上信心加權分數變異數分析摘要表

變異來源	離均差平方和	自由度	均方	F
A (性別)	2765.51	1	2765.51	.51
B (指導語)	67.72	1	67.72	.01
AB (性別×指導語)	2003.93	1	2003.93	.37
S/AB (誤差)	1183818.30	216	5480.64	

表十二 各組男女生在永和國中數學成就測驗上信心分數變異數分析摘要表

變異來源	離均差平方和	自由度	均方	F
A (性別)	9943.01	1	9943.01	5.00*
B (指導語)	2023.05	1	2023.05	1.02
AB (性別×指導語)	24.12	1	24.12	.01
S/AB (誤差)	429155.12	216	1986.83	

* $P < 0.05$

二、傳統計分法與信心加權法對測驗信度與效度的影響

表十三所列者係各組學生在永和國中數學成就測驗上傳統分數與信心加權分數之折半相關係數。這二種係數均由同一樣本求得，故其差異之顯著性須由下列公式(Kenny, 1975, Peters & VAN Voorhis, 1940) 加以考驗。

設 r_{14} = 傳統分數之奇偶相關

r_{23} = 信心加權分數之奇偶相關

r_{12} = 傳統分數之奇數總分和信心加權

分數之奇數總分之相關

r_{34} = 傳統分數之偶數總分和信心加權

分數之偶數總分之相關

r_{13} = 傳統分數之奇數總分和信心加權

分數之偶數總分之相關

r_{24} = 傳統分數之偶數總分和信心加權

分數之奇數總分之相關

則 $K = (r_{12} - r_{24} r_{14}) (r_{34} - r_{24} r_{23})$

$+ (r_{13} - r_{12} r_{23}) (r_{24} - r_{12} r_{14})$

$) + (r_{12} - r_{13} r_{23}) (r_{34} - r_{13}$

$r_{14}) + (r_{13} - r_{14} r_{34}) (r_{24} -$

$r_{34} r_{23})$

$$Z = \frac{\sqrt{N} (r_{14} - r_{23})}{\sqrt{(1 - r_{12}^2)^2 + (1 - r_{23}^2)^2 - K}}$$

由表十三可知，1.除控制組女生外，其他各組信心加權分數之信係度數都略高於傳統分

數，不過其差異由上述公式考驗後均未達顯著水準($P>0.05$)。2.各組信心分數之信度極高，其折半相關校正後，約在 0.897 到 0.997 之間。

表十三 各組學生在永和國中數學成就測驗上傳統分數與信心加權分數之信度係數

組 別	性別	傳 統 分 數		加 權 分 數		信 心 分 數	
		折半相關	校正後* 之信度	折半相關	校正後* 之信度	折半相關	校正後* 之信度
實 驗 組	男	.840	.913	.848	.918	.898	.946
	女	.765	.867	.803	.891	.813	.897
控 制 組	男	.801	.890	.835	.910	.943	.971
	女	.843	.915	.817	.899	.995	.997

*表內係數係以 Spearman-Brown Prophecy formula 加以校正。

各組男女生在永和國中數學測驗上之信心分數與數學能力之相關係數列於表十四。傳統總分與信心加權分數之相關係數列於表十五。根據上述二表分析可發現：1.傳統總分與信心加權分數具有很高的相關($r=0.958\sim0.976$)。這表示學生在這兩種分數上的排列等第(Rank order) 非常一致。2.信心分數與信心加權分數之相關在 0.401 到 0.640 之間，與傳統總分之相關在 0.282 到 0.645 之間，與二年級上學期數學成績之相關在 -0.277 到 0.594，與師大教研所國中數學成就測驗之相關在 0.298 到 0.591 之間。可見受試者信心分數與其數學能力雖具有顯著的相關，但並非完全正相關。受試者對自己答案正確性的信心程度，除其數學知識與能力外，可能還受到性別等其他因素的影響，譬如男生對自己所做的數學答案正確性之信心程度與其數學知識（指學期成績及測驗總分）之相關 ($r=0.448\sim0.645$) 就普遍地高於女生 ($r=-0.277\sim0.401$)。3.受試者之信心總分以及各信心選目下之答對題數的相關，在「很有信心」選目下為 0.778~0.887，在「有信心」選目下為 -0.210~-0.418，在「稍有信心」選目下為 -0.674~-0.698，在「沒有信心」選目下為 -0.400 到 -0.612。可見信心分數與能力之相關視測驗情境而定 (Situational)，在很有信心選項下有相當高的正相關，在其他信心選項下則有顯著的負相關。這種趨勢在各組中都很一致。4.信心總分數與「空題不答」的題數之相關近於零。

表十四 信心分數與數學能力之相關係數

受試組別	性別	永和國中數學成就測驗						二年級上成績	師國就大中測數數驗研學分所成數	
		不同信心選項下之傳統分數				沒的題作答數	傳統總分	信心加權分數		
		很有信心	有信心	稍有信心	沒有信心					
實驗組	男	.887	-.418	-.698	-.612	.116	.645	.640	.582	
	女	.826	-.336	-.699	-.660	-.056	.391	.430	.330	
控制組	男	.887	-.397	-.688	-.400	.039	.502	.548	.594	
	女	.778	-.210	-.674	-.600	.036	.282	.401	-.277	

表十五 傳統分數與信心加權分數之相關係數

受試組別	性別	傳統分數與信心加權分數之相關	
		男	女
實驗組	男	.958	
	女		.976
控制組	男	.972	
	女		.968

筆者為研究傳統計分法與信心加權計分法對測驗分數效度的影響，特以國中二年級數學成績與師大教研所數學成就測驗之成績為效標，茲將求出的效度係數列於表十六。這兩種效度係數係由同一樣本計算所得，須根據下列公式 (GLass & Stanley, 1970) 對兩個效度係數差量的顯著性加以考驗。

設 r_{12} = 傳統分數與效標之相關

r_{13} = 信心加權分數與效標之相關

r_{23} = 傳統分數與信心加權分數之相關

n = 樣本人數

則

$$Z = \frac{\sqrt{n} (r_{12} - r_{13})}{\sqrt{(1 - r_{12}^2)^2 + (1 - r_{13}^2)^2 - 2r_{23}^2 - (2r_{23} - r_{12}r_{13})(1 - r_{12}^2 - r_{13}^2 - r_{23}^2)}}$$

根據表十六分析的結果發現：1.如以師大教研所國中數學成就測驗為效標，則各組信心

加權計分法所得的效度係數($r = .757 \sim .848$)均高於傳統計分法所得的效度($r = .724 \sim .831$)，惟二者之差異只有控制組男生達到顯著水準 ($P < .05$)。2.如以國中二年級上學期數學成績為效標，則實驗組女生和控制組男生信心加權計分法所得的效度 ($r = .686$ 和 $.726$) 高於傳統計分法所得的係數 ($r = .650$ 和 $.683$) 二者之差量在這幾組中均未達顯著水準。實驗組男生和控制組女生信心加權分數之效度 ($r = 0.731$ 和 0.472) 則低於傳統分數之效度係數 ($r = .788$ 和 $.553$)，二者差異顯著 ($P < .05$)，這兩種計分方法在不同的效標上所顯示的趨勢並不一致。這可能由於這二個效標性質彼此不同所致。師大數學成就測驗是填空式試題，受兒童猜題行為之影響較少，而且每個學生試題及評分標準完全一致。學期成績則由月考及平時成績計算而成，月考採選擇式試題，容易受猜題行為的影響，平時成績又因各班師教不同，評分標準可能不會完全一致。因此在這兩種效標中，自以前者較佳。3.由各組在永和國中數學成就測驗上各信心選目之答對題數與師大教研所國中數學成就測驗成績之相關可看出，「很有信心」選項下之傳統分數（即答對題數）效度甚高 ($r = .515 \sim .734$)，而其他信心選項之傳統效度係數絕大部份則為顯著的負相關或近於零相關。如改以二年級上學期數學成績為效標，除控制組女生外，亦得類似結果。由此可見兒童在測驗上的分數，除了考慮其答案之正確性外，亦應就其信心程度給予不同的比重 (differential weighting)。4.各組空題不答的題數與效標之相關受性別的影響。女生較男生空題不答的題數多，個別差異大，但其與效標之相關都近於零。男生空題不答的題數與師大教研所國中數學測驗成績有顯著的相關，但與國中二年級上學期數學成績之相關則不顯著。

三、人格特質與認知模式對信心加權的影響

信心加權法之成效，視信心分數是否可靠而定。由過去文獻的探討發現，信心分數可能受到某些人格特質及認知模式的影響。茲將兒童在各種人格測驗及藏圖測驗成績之平均數和標準差列於表十七。各組男女生傳統分數、信心加權分數、信心分數與人格特質之相關係數列於表十八，表十九，表二十和表二十一。

表十六 信心加權分數與傳統分數之效度係數

計分方法	受試組別	二年級上學期數學成績				師大教研所數學成就測驗分數			
		實驗組		控制組		實驗組		控制組	
		男	女	男	女	男	女	男	女
信 心 加 權 分 數		0.731	0.686	0.726	0.472	0.757	0.848	0.806	0.787
傳 統 分 數		0.788	0.650	0.683	0.553	0.734	0.831	0.766	0.781
不 同 之 信 心 傳 統 選 項 數	很 有 信 心	0.590	0.525	0.713	0.167	0.734	0.652	0.627	0.515
	有 信 心	-0.058	0.069	-0.180	0.263	-0.333	0.072	-0.003	0.263
	稍 有 信 心	-0.136	-0.168	-0.422	-0.117	-0.313	-0.165	-0.353	-0.356
	沒 有 信 心	-0.455	-0.147	-0.025	-0.070	-0.305	-0.124	0.015	-0.242
	沒 有 作 答 的 題 數	0.186	0.119	0.218	0.140	0.351	0.080	0.375	0.083

由上列各表可知：1.各組藏圖測驗成績與傳統總分、信心加權分數及信心分數均呈正相關，且大多達顯著水準。2.測試焦慮與控制組女生之傳統總分呈顯著負相關 ($r = -0.267$)，與其信心分數則呈顯著正相關 ($r = .306$)，其餘之相關係數均未達顯著水準。3.說謊量表的分數與女生信心加權分數呈顯著負相關 ($r = -.290 \sim -.313$)，與控制組男生之信心分數呈顯著正相關 ($r = .307$)。4.自卑量表分數與控制組女生之信心分數呈正相關 ($r = .40$)，與控制組男生則呈顯著負相關 ($r = -.322$)。5.學業成就動機與控制組女生之信心分數有顯著正相關 ($r = .593$)，與其他各組之相關則未達顯著水準。

表十七 各組受試者人格測驗及藏圖測驗成績之平均數與標準差

人格測驗	受試組別	實驗組				控制組			
		男*		女*		男*		女*	
		\bar{X}	SD	\bar{X}	SD	\bar{X}	SD	\bar{X}	SD
兒 童 測 試 焦 慮 量 表		12.27	5.85	13.65	5.22	13.91	5.60	14.09	8.20
說 謊 量 表		4.80	2.07	4.80	1.91	4.75	2.21	4.49	1.64
自 卑 量 表		6.98	4.13	7.29	4.67	7.36	4.29	8.89	8.76
藏 圖 測 驗		10.11	5.39	9.27	4.77	10.96	5.77	9.07	5.89
學 業 成 就 動 機 測 驗		21.27	2.56	21.64	2.19	21.47	2.98	21.69	5.23

* 各組之人數均為55人

表十八 實驗組男生之人格特質、認知模式與永和國中數學成就測驗上各種分數之相關

人 格 測 驗	傳 統 總 分	信 心 加 權 分 數	信 心 分 數
兒童測試焦慮量表	-0.169	0.088	-0.239
說謊量表	-0.085	-0.095	-0.142
自卑量表	-0.139	-0.096	-0.105
藏圖測驗	0.422*	0.430*	0.276*
學業成就動機測驗	0.173	0.169	0.081

* p<.05

表十九 控制組男生之人格特質、認知模式與永和國中數學成就測驗上各種分數之相關

人 格 測 驗	傳 統 總 分	信 心 加 權 分 數	信 心 分 數
兒童測試焦慮量表	0.106	0.096	-0.155
說謊量表	-0.115	-0.081	0.307*
自卑量表	-0.100	-0.111	-0.322*
藏圖測驗	0.064	0.078	0.304*
學業成就動機測驗	0.039	0.006	0.234

* p<.05

表二十 實驗組女生之人格特質與永和國中數學成就測驗上各種分數之相關

人 格 測 驗	傳 統 總 分	信 心 加 權 分 數	信 心 分 數
兒童測試焦慮量表	-0.079	-0.020	0.043
說 謠 量 表	-0.259	-0.313*	-0.225
自 卑 量 表	-0.162	-0.135	0.139
藏 圖 測 驗	0.432*	0.424*	0.307*
學業成就動機測驗	0.113	0.076	0.102

* $P < .05$

表二十一 控制組女生之人格特質與永和國中數學成就測驗上各種分數之相關

人 格 測 驗	傳 統 總 分	信 心 加 權 分 數	信 心 分 數
兒童測試焦慮量表	-0.267*	-0.202	0.306*
說 謠 量 表	-0.227	-0.290*	-0.235
自 卑 量 表	-0.260	-0.170	0.410*
藏 圖 測 驗	0.328*	0.304*	0.118
學業成就動機測驗	-0.029	0.065	0.593*

* $P < .05$

肆、討 論

本研究發現信心加權計分法所得的信度和效度雖比傳統計分法略高，但效果不很顯著。從傳統分數和信心加權分數間之高度相關亦可看出此現象。信心加權計分法所得之效度和信度之所以不如理想那麼高，可能是由於其信心分數除了兒童學科知識外，還受其他因素的影響。

響。本研究即曾發現信心分數和學科知識之間，只有中度之相關。它與性別、指導語、人格特質、及認知模式有顯著的關連。不過這些缺點，正如心理與教育學者所言，不但不應成為忽視信心加權法的藉口，反而是大家要重視它的理由（Ebel, 1965; Wen, 1975; Adams & Adams, 1957, 1961）。Ebel (1965) 和 Wen (1975) 認為信心加權法效果之所以不顯著，是由於兒童無法根據自己答案的正確程度，賦予適當的信心。他們主張學校今後應加強訓練兒童，使之能更客觀的區別不同的信心水準（level of confidence）。Adams & Adams (1957, 1961) 更以實驗證實信心的判斷（confidence judgement）是可經訓練加以增進的。

筆者認為信心加權法效果不顯著，可能和信心之分等及加權有關。在信心分等方面，有的學者主張分成二個等第，有的主張分為三個或四個等第（本研究所用的是四個等第）。在信心加權方面，有的學者主張對信心相同的答案，不管其正確與否，一律給予相同的加權（Weighting），有的人則主張獎勵重於懲罰，正確答案的加權數應大於錯誤的答案（本研究所採用的是前者）。由於信心的分等、及正確和錯誤答案的加權，每個研究不一致，使得不同的研究結果，難以相互比較。尤其值得注意的是，每個研究的信心加權，都很主觀。今後信心加權似應以學業成績或標準成就測驗為效標（criteria），以各信心選項下正確答案及錯誤答案之題數為預測變項（predictor），利用多重迴歸（Multiple Regression）及交叉驗證（Cross Validation）等科學方法來決定各種答案應得的比重（或加權）。過去的研究發現，兒童的猜題行為與學生之能力及試題之難度有關。受試者之能力及試題之難度對信心加權分數究竟有什麼影響，亦值進一步的探討。

總之，臺灣升學競爭如此激烈，許多學子往往由於一分之差，而無法進入理想的學府，有的甚至還因之名落孫山。如何使測驗結果更為客觀公平，殊屬重要，而目前臺灣各級學校招生入學考試及課堂教學所大量採用的選擇式試題，雖可使評分更為客觀快速，題目取樣更為廣泛，但其測驗結果却容易受兒童猜題行為的影響。有些考試（如大專聯考）為減少猜題行為的影響，採用機率校正公式。每題如有 k 個選目，每答錯一題，即倒扣 $1/(k-1)$ 分。但由於學生對試題所列的各種選目，常擁有部份知識。他答題時往往先將有把握的選目剔除，再對其他一無所知的選目加以猜測。像這類滲雜部份知識的猜題，與盲目的猜題有很大的差別。況且目前各級學校聯招試題，為了保密，撰題時間都很短促，加以沒有測驗專家從旁協助，所編列的選目很多只是為了湊數，和題幹的關連不大，一看就知道不是正確答案。在這種情況下，傳統計分方法就更不適用了。本研究曾發現兒童很有信心的答案，不但正確率高，且其正確答案與標準成就測驗之分數，具有很高的正相關，信心程度較低的題目，則正確率較低，且其正確答案與效標呈零相關和顯著的負相關。由此可見兒童的測驗分數，確應根據信心程度予以不同的加權。既然傳統計分法本身具有重大的缺點，而兒童的測驗分數又有根據其信心程度予以不同加權的必要，則如何發展出一套更精密的信心加權法，以適用於不同性別及人格特質的受試者，將是今後學術界的重大課題。

伍、參考書目

- 吳靜吉 猜題行為 未發表
- 李克明、黃國彥 國小學生測試行為與課堂測驗之關係
測驗年刊二十三輯 民國六十五年
- 林邦傑 國民中學學生學業成就與人格特質關係之研究
國立政治大學學報第二十三期 民國六十年
- 林碧峯、楊國樞、繆瑜、楊有維 中國學童焦慮量表之修訂 見楊國樞、張春興編著 中國兒童行為的發展 臺北：環宇出版社 民國六十三年
- 國立臺灣師範大學教育研究所 數學成就測驗測驗手冊 民國六十一年
- 賴保禎 基氏人格測驗手冊 中國行為科學社出版 民國六十五年
- Adams, J. K. and Adams, P. A. Realism of Confidence-Judgements. *Psychological Reviews*, 1961, 68, 33~45
- Adams, P. A. and Adams, J. K. Training in Confidence-Judgements. *American Journal of Psychology*, 1957, 70, 432~436
- Allison, D. E. Test Anxiety, Stress, and Intelligence-test Performance. *Canadian Journal of Behavioral Science*, 1970, 2, 26~37
- Anastasi, A. Heredity, Environment, and the Question "How?" *Psychological Review*, 1958, 65, 197~208
- Archer, N. S. Effects of confidence Weighting by fifth and sixth grade students on objective test scores. Abstract of a paper presented at the annual meeting of the American Educational Research Association, Chicago, 1963 (Multilith)
- Armstrong, R. J. and Mooney, R. F. Confidence testing: Is it reliable? Paper read at the annual meeting of the National council on Measurement in Education, Los Angeles, February, 1969.
- Coombs, D. H. Milholland, J. E. and Womer, F. B. The assessment of Partial Knowledge. *Educational and Psychological Measurement*, 1956, 16, 13~37.
- Chambers, A. C. Hopkins, K. D. and Hopkins, B. R. Anxiety, psychologically and psychologically measured: Its effects on mental test performance, *Psychology in the schools* 1972.
- Davis, F. B. Item selection techniques. In E. F. Lindquist, ed, *Educational measurement*, Washington, D. C: American Council on Education, 1951.
- Davis, F. B. Use of correction for chance success in test scoring. *Journal of Educational Research*, 1959, 52, 279~280.
- Diamond, J. J. and Evans, W. B. The Correction for guessing. *Review of Educational Research*, 1973, 43, 181~191.
- Dressel, P. J. and Schmid, P. Some modifications of the multiple choice item. *Educational and Psychological Measurement*, 1953, 13, 574~594.

- Echternacht, G. J. The use of Confidence testing in objective test. Review of Educational Research, 1972, 42, 217~236.
- Ebel, R. L. Confidence Weighting and test reliability. Journal of Educational Measurement, 1965, 2, 49~57.
- Ebel, R. L. Some effects of credit for confidence and Penalty for error on true-false test scores. Abstract of a paper presented at the annual meeting of the American Educational Research Association, Chicago, 1961 (Multilith)
- French, J. W. Effects of Anxiety on Verbal and Mathematical examination Scores. Educational and Psychological Measurement, 1962, 22, 553~564.
- Glass, G. V. and Stanley, J. C. Statistical Methods in Education and Psychology. New Jersey: Prentice-Hall 1970.
- Granich, L. A. A technique for experimenting on guessing in objective tests. Journal of Educational Psychology, 1931, 22, 145~156
- Hansen, R. The influence of variables other than knowledge on probabilistic test. Journal of Educational Measurement, 1971, 8, 9~14.
- Hambleton, R. K., Roberts, D. M. and Traub, R. E. A Comparison of the reliability and validity of two methods for assessing partial knowledge on a multiple-choice test. Journal of Educational Measurement, 1970, 7, 75~82.
- Hamilton, C. H. Bias and Error in Multiple-choice test. Psychometrika, 1950, 15, 151~168.
- Hevner, K. A. A method of correcting for guessing in true-false tests and empirical evidence in support of it, Journal of Social Psychology, 1932, 3, 359~362.
- Hopkins, K. D., Hakstian, A. R. and Hopkins, B. R. Validity and reliability consequences of confidence Weighting. Educational and Psychological Measurement, 1973, 33, 135~141.
- Jackson, R. A. Guessing and Test Performance. Educational and Psychological Measurement, 1955, 15, 74~79.
- Jacobs, S. S. Correlates of unwarranted confidence in response to objective test items. Journal of Educational Measurement, 1971, 8, 15~19.
- Johnson, D. M. A systematic introduction to the psychology of thinking. New York: Harper and Row, 1972.
- Kenny, D. A. Cross-Lagged panel correlation: A test for spuriorisness. Psychological Bulletin, 1975, Vol. 82, 6, 887~903.
- Little, E. and creaser uncertain responses on Multiple-choice examinations. Psychological Reports, 1966, 18, 801~802.
- Lord, F. M. Formula Scoring and Number-right Scoring. Journal of Educational Measurement, 1975, 12, 7~11
- Lord, F. M. and Novick, M. R. Statistical theories of mental test scores, Reading Mass: Addison-Wesley, 1968.

- Messick, S. Hidden Figure Test, Princeton, New Jersey: Educational Testing Service, 1962.
- Michael, J. J. The reliability of a multiple-choice examination under various test-taking instructions Journal of Educational Measurement, 1968, 5, 307~314.
- Peters, C. C. and Van Voorhis, W. R. Statistical procedures and the mathematical bases. New York: McGraw-Hill, 1940.
- Rippey, R. Probabilistic testing, Journal of Educational Measurement, 1968, 5, 211~215.
- Ruebush, B. K. Anxiety. In the Yearbook committee and Associated Contributors, H. W. Stevenson, J. Kagan and C. Spiker, eds. Child Psychology: The Sixty-Second Yearbook of the National Society for the Study of Education. Chicago: Univrsity of Chicago Press, 1963.
- Sarason, S. B., Davidson, K. S., Lighthall, F.F. Waite, R. R. and Ruebush, B. K. Anxiety in elementary School Children: A Report of Research. New York: Wiley 1960.
- Schariffs, A. C. and Boomer .D. S. Who is penalized by the penalty for guessing?, Journal of Educational Psychology, 1954, 45, 81~90.
- Shuford, E. H. and Massengill, H. E. Valid Confidence Testing Demonstration Kit. The Shuford-Massengill Corp. 1967.
- Silverstein, A. B., Moham, P. J., Franken, R. E. and Rhone, D. E. Test anxiety and intellectual performnace in mentally retarded school children. Child Development, 1964, 35, 1137~1146.
- Slakter, M. J. Risk-taking on objective examinations. American Educational Research Journal, 1967, 4, 31~43.
- Slakter, M. J. The penalty for not guessing. Journal of Educational Measurement, 1968, 5, 141~144.
- Slakter, M. J. Generality of risk taking on objective examinations. Educational and Psychological Measurement, 1969, 29, 115~128,
- Soderquist, H. O. A New method of Weighting Scores in a true-false test. Journal of Educational Research, 1936, 30, 290~292.
- Stokes, R. R. The Split-response technique. Phi Delta Kappan, 1966, 47, 271~272.
- Stanley, J. C. and Hopkins, K. D. Educational and Psychological Measurement and Evaluation. Englewood cliffs, N. J.: Prentice-Hall, 1972
- Stanley, J. and Wang, M. D. Weighting test items and test-item options, An overview of the analytical and Empirical literature. Educational and Psychological Measurement, 1970, 30, 21~35.
- Swineford, F. Measurement of a personality trait. Journal of Educational Psychology, 1938, 29, 295~300
- Swineford, F. Analysis of a personality trait. Journal of Educational Psychology, 1941, 32, 438~444.

- Swineford, F. and Miller, P. M. Effects of directions regarding guessing on item statistics of a Multiple-choice Vocabulary test. *Journal of Educational Psychology*, 1953, 44, 129~139.
- Thorndike, R. L. The problem of guessing. In R. L. Thorndike (ed.) *Educational Measurement* (2n ed.) Washington, D. C.: American Council on Education 1971, 59~61.
- Vernon, M. D. *The psychology of perception*, England: Penguin, 1962.
- Votaw, D. F. The effect of "Do not guess" directions on the validity of true-false and Multiple-choice tests. *Journal of Educational Psychology*, 1936, 27, 699~704.
- Wen, S. The relationship between Verbal-Meaning test scores and degree of confidence in item responses. *Journal of Educational Measurement*, 1975, 12, 197~200.
- Witkin, H. A., Lewis, H. B. and Hertzman, H. *Personality through perception*, New York: Harper, 1954.
- Witkin, H. A., Dyk, R. B., Faterson, H. F., Goodenough D. R. and Karp, S. A. *Psychological differentiation*. New York: Wiley, 1962.
- Yamamoto, K. and Dizney, H. F. Effects of three sets of test instructions on Scores on an intelligence Scale. *Educational and Psychological Measurement*, 1965, 25, 87~94.
- Ziller, R. C. A measure of the gambling response Set in objective tests *Psychometrika*, 1957, 22, 289~292.

本文承國科會助補，特此誌謝！