

美國中小學學業記分政策的探究

楊惠琴

本篇論文深入探究美國中小學各種不同的記分政策，以作為我國實施上的參考。本文依序敘述記分政策的構成要素及其重要性。首先闡明「學業記分」及「成績」的意義。第二部份分別從「記分系統發展的哲學基礎」及「成績決定的哲學基礎」兩方面來闡述學業記分的哲學基礎。第三部份探討學業記分的功能，如通知性、誘因性及行政上的功能。第四部份描述兩種基本的記分模式--比較性的記分、精熟式的記分--之起源、發展、和特質。在這兩種記分模式之下的各種記分符號系統包括：數字成績、字母等第、通過/不及格、非懲罰性的記分、敘述性的報告、契約式的記分、自我評量及學習目標查核表。第五部份探討記分的參照標準：常模參照、標準參照、以及自我參照。第六部份追溯成績的決定要素及來源。第七部份提出有關學業記分的爭論以及新的評量方式。最後建議如何設計及操作一個有效的記分系統。

關鍵字：學業記分、學業記分政策、學業記分系統

Keywords: academic grading、academic grading policies、academic grading systems

壹、導論

學業成就的評量與記分是一項重要的教育措施；它能提供學生、家長、老師、諮商者、以及僱主許多有關學生學業成就的訊息。目前美國已經有許多關於各級學校學業記分政策與實施概況的研究報告。早期的研究顯示出字母等第 (letter grades) 是最普遍的記分措施，而特維傑 (Terwilliger, 1971) 的研究顯示 70% 的學校以五等級系統 (如 A-F) 來登錄各科的學期成績，當前則以數字成績 (numerical grades) 最為普遍。

學業記分政策通常包括以下各種要素：「學業記分」(academic grading) 及「成績」(grades) 的概念性架構；學業記分的哲學基礎；學業記分的目的；基本的記分模式--比較性的記分，精熟式的記分--之起源、發展和特質；以及在這兩種記分模式之下的各種記分符號系統--包括數字成績 (numerical grades)、字母等第 (letter grades)、通過 (pass)/ 不及格 (fail)、非懲罰性 (non-punitive) 的記分、敘述性的報告 (self-assessment)、契約式的記分 (contract grading)、自我評量 (self-assessment)、以及學習目標查核表 (learning-objective checklists)；記分的參照標準--常模參照 (norm-referenced standard)、標準參照 (criterion-referenced standard)、以及自我參

照 (self-referenced standard)；成績的決定要素及來源；成績報告的途徑。最後討論新的評量方式，並提出美國學業記分政策對我國中小學學業成績考查的啓示與建議。

貳、學業記分政策

學業記分政策是由州或地方教育局所制定的一套有關學業記分之法定的、明確的計畫 (Smith & Smittle, 1954)。書面的記分政策有助於對新進人員 (例如行政或教育人員) 的宣導。書面記載的記分政策有很多優點 (Tuttle, 1952)，例如：

- 一、書面的陳述規範責任和權限，因此可以避免誤解和混淆。
- 二、經由書面政策的權威性通告，公眾能較清楚地了解及認識學校。
- 三、行政人員能夠明確地解釋書面的政策。

因此，教育局 (boards of education) 需要制訂一套廣為接納的記分政策以便於評量及報告學業成就。

一、成績和學業記分 (Grades and Academic Grading)

成績是教育過程中某一方面成就的評價性總結 (Thorndike, 1969)；它代表學生的學業能力 (Anderson & Wendel, 1988)。因此，有必要把成績的意義向所有關心者解釋清楚。老師有義務向學生傳達下列事項：(1) 各學科的主要學習目標，(2) 對於這些學習目標的評量標準；(3) 各項記分要素及來源所佔學科成績的比重 (Terwilliger, 1977)。

學業記分是用來報告和傳達評量結果的一種方法 (Robinson & Craver, 1989)。評量包括測量 (measurement) 及評估 (evaluation) (Robinson & Craver, 1989)。測量是指實際觀察某些教育現象、特徵及本質，並且依照既定的步驟或法則將這些觀察轉變成量化的或分類的型式 (Worthen, Borg, & White, 1993)。評估指經由將原始分數參照於某些既定的標準，而後判斷學生表現品質且決定其價值 (Robinson & Craver, 1989)。學業記分包括兩個步驟：(1) 用記號表示 (symbolizing)，和 (2) 報告 (reporting)。所謂「用記號表示」就是將學生的成就水準和所設定的參照標準相比較後，給予一個記號 (例如一個記號字母、一個百分等級、或者一個數字成績) 以代表其成就水準。然後，向父母、學生、或僱主報告這項記號。

二、學業記分的哲學基礎

(一) 記分系統發展的哲學基礎

記分系統的發展是根基於三種主要的哲學論點：(1)選擇性功能，(2)促進社會進步，以及(3)為未來生活作準備。

學校和社會的選擇性功能可以追溯到柏拉圖的「理想社會」。柏拉圖認為，一個好而公平的社會需要我們能辨識每個人間的差異，並且需要提供方法使每個人了解自己的潛力。這種理想可以藉著教育系統的分類及選擇機能而達成 (Plato, 1968)。

歐洲的觀念--教育是困難的--強調了學校的選擇性功能 (Spady, 1992)。教育是很困難的；不僅是因學校的教材愈來愈難，並且因教與學的時間是有限的。這個觀念是基於以下的假設——很多孩子是「不可教的」，因此必須儘早辨識出來以便引導他們去從事非學術性的工作；對於比較可教的小孩，教師需鼓勵他們繼續唸高中及大學 (Tyler, 1986)。基於這個論點，學校有責任選擇及訓練最有能力及天賦的人，使他們成為社會的高階份子。

記分系統發展的另一哲學基礎是根源於這一信念——記分系統確保社會進步。布魯柏橋 (Brubacher, 1952) 認為，成績是一項標準，也就是客觀化的社會理想，它能激發學生努力去達成這個理想以求得到社會的特可。相似地，愛貝爾 (Ebel, 1980) 也說，記分可鼓勵學生設定實際的期望，並且激勵學生努力學習。卡爾門 (Colvin, 1912) 指出，成績是學生努力學習所換得的代價；學生努力學習以期得到好的成績而避免差的成績，就如同工人努力工作以期得到報酬而免於被革職。

記分的過程同時也是為學生的未來工作世界作準備；詳細地說，紙筆測驗及其他課堂作業的成功表現是為未來的教育評量作準備，並使學生具備日後受僱用的能力 (Brubacher, 1952)。在教育及職業輔導上，成績扮演著重要角色。例如，藉著參考學生的所有成績記錄，學生或輔導員便可以洞悉學生的強或弱點，以及興趣，並且為未來的教育或職業計畫作明智的抉擇。

(一) 成績決定的哲學基礎

每一位老師對教育有不同的哲學觀點，因所持的觀點不同而影響到教學、評量、以及記分方式。在特維勒傑 (Terwilliger, 1977) 的書—「指定成績—哲學的議題及實際的建議」 (Assigning Grades-Philosophical Issues and Practical Recommendation) —闡述了各種有關記分的哲學觀點。依特維勒傑的觀點，每一位老師的記分措施是受三種觀點所引導：行為主義 (Behaviorism)、人本主義 (Humanism)、及實用主義 (Pragmatism)。遵奉行為主義者對於個別學生表現的差異並不感興趣；他們所在乎的是學生是否達到既定的學習目標。他們認為成績只不過是人的獎懲制度；它具有傷害性，並且干擾學習。因此，行為主義者不喜歡使用有區別性的成績 (例如 A、B、C、D、F 或者 0-100)。但是若學校政策要求必須記

分時，他們會依既定的表現標準 (Predetermined Performance Criteria) 來記分。

人本主義學家關心的是學生個人的價值、興趣及尊嚴。他們批評機械化及非人性化的傳統教育及傳統記分方法，並呼籲寬鬆的且非正式的教育及評量方法；他們認為成績並沒有助於心理進步、提升學習者動機、或助長真實知識的獲得，因而主張廢除成績。但若依政策要求必要記分，人文主義者會對學生表現作個別化的價值判斷，因而使成績變得人性化。在這種情況下，記分是根據自我參照標準；也就是先將學生的成就水準參照個人的能力或成長，而後再決定成績。實用主義學者強調教育的實用價值及結果。他們認為教育具備為未來生活準備的功能；教育提供學生發展特殊能力及天賦的機會。為了達成這個教育目標，教師必須確認個別學生相對於其他學生的強或弱點，再以常模參照標準加以解釋。同時，實用主義者也相信，在發展過程中，成績可以幫助個人作明智的抉擇。

三、學業記分的目的

第一個記分系統是由耶魯大學 (Yale University) 於 1785 年發展出來的；它是用來評估學生的表現，並且通告學生有關教授對於學生考試表現的看法。當今，重要的議題是成績的各種用途，包括傳達有關學生的訊息、提供學生讀書的誘因、作為選擇的標準、提供行政上的記錄、協助評量和監督教學過程、協助學生、以及在教育及職業上的計畫 (Warren, 1975)。這些記分目的可以歸納為三類：通知性 (informative)、誘因性 (motivational) 及行政上的 (administrative) 功能 (Wrinkle, 1935)。

(一) 通知性的功能 (informative function)

成績的一項主要目的是提供訊息；成績是有關學生學業成就的累積性記錄，它提供給學生、雙親、老師、輔導員、學校行政人員、以及僱主有關學生表現的訊息 (Frisbie & Waltman, 1992; Gage & Berliner, 1988; Spady, 1992)。對學生而言，成績指出有關自己成就表現的優、缺點以及提供改進的參考。對父母而言，成績傳達孩子的成就水準，並且作為父母協助孩子學習的參考。對老師而言，學生成績可以作為改進教學目標及課程計畫的指標。對輔導員而言，由成績可以明瞭學生的特殊能力，藉此幫助學生在教育及職業計畫上作明智的決定。對學校行政人員而言，成績是各種活動的依據，例如：頒發殊榮及獎學金，允許進入特殊的課程，允許參與各種活動。就僱主而言，成績可以預測未來工作的表現，藉此來篩選人才。

(二) 誘因性的功能 (motivational function)

成績被視為學習的誘因 (Colvin, 1912; Ebel, 1980; Spady, 1992)；它能引發教師所需求的教室行為，例如讀書或專心等行為 (McDaniel, 1979)。卡爾門 (Colvin) 是早期提倡將成績視為學習誘因的人。他說，成績激勵學生專心讀書以求獲得好成

績而避免壞成績。相同地，愛貝爾(Ebel, 1980)認為成績是激勵學生用功讀書的有效方法。

(三)行政上的功能(administrative function)

成績是學生成就的一項指標，它常被用為行政上篩選的依據，以確保學生能於某一學程或活動中成功地學習。這些行政上的決定包括升級與留級，允許進入某一門課或科系，獎學金的決定，以及參與各種課外活動(McKeachie, 1986)。

四、學業記分系統

學業記分發生在各級學校，基本上，有兩種基本的記分模式，即比較性的記分與精熟式的記分(Laska & Juarez, 1992a)，其起源、發展、與特質將分別敘述。在這兩種記分模式之下的各種記分符號系統包括：數字成績、字母等第、通過/不及格、非懲罰性的記分、敘述性的報告、契約式的記分、自我評量、及學習目標查核表。在此要探討兩種基本的記分途徑——比較性的記分(comparative grading)及精熟式的記分(mastery grading)。

(一)基本的記分模式

1. 比較性的記分

比較性的記分制度可以追溯到1785年，美國耶魯大學(Yale University)的教授使用記分表來區分學生考試的表現(Smallwood, 1935)。這個量表包含三種描述性形容詞：優等(Optimi)，次等的(Inferiores)，及低劣的(Pejores)——這是比較性記分的源起。

直到十九世紀中期，比較性記分才被廣泛地使用於美國的各級學校(Durm, 1993; Kirschenbaum, Simon, & Napier, 1971)。在西元1850年以前，大部分的小學是集中各種年紀的學生於一間教室內，並沒有意圖將學生按能力或年紀分成不同年級，在這個時期，大部份的學生只上小學而已。西元1850年以後，因政府支持的小學普遍設立，同時，強迫性教育法規的通過，導致就學學生人數大量增加，這種現象需求將學生按年紀或能力分類，因而促進比較性記分的發展。發展比較性記分的另一驅力是幫助大學篩選學生(Hargis, 1990)。因高中入學人數的擴張，大學的入學許可愈來愈競爭。因此，成績被用為篩選學生素質的依據。由上可知，當時的記分系統是用來分類、分組及篩選學生。

上述的兩種社會驅力——為了將學生按能力及年齡分級，並且篩選進入大學的學生——促進量化記分系統的發展。量化的記分系統是因應實驗性教育(experimental education)而發展出來的。實驗性教育源自於科學的訓練(Spady, 1992)。科學知識鼓勵了對於教育測量的研究，其基本假設是教育的結果是可

以測量的(Colvin, 1912)。為了測量教育的結果，必須有一套測量能力及成就的方法。因此，一個有區別性的及比較性的記分系統應運而生。在此記分系統之下，學業成就便成為和同班同學或同年級同學相比較的一項指標。

在二十世紀初，特別是在中等學校，百分制記分變得十分普遍。但有很多小學或中學並不使用百分制，而只是使用符號如S代表滿意(satisfactory)及U代表不滿意(unsatisfactory)。

在西元1920年代，記分的趨勢是採用五等分量表或者字母等第，例如1-5或A、B、C、D、F。這種記分系統的目的在於增進記分的客觀性和可靠性。史達奇(Starch)和伊里特(Elliot)於1913年的研究中批評百分等級記分系統所使用的測量太過於精密，容易受教師價值判斷的影響，而變得主觀且缺乏可靠性(Cited in Kirschenbaum, Simon, & Napier, 1971)。而後也有些研究指出，五等級或十等級記分能較可靠地區分學生的表現。因此，等級記分(如字母等第)或相類似的記分系統，廣泛地使用於美國的各級學校。

比較性的記分是目前美國各級學校最普遍的記分系統。各種不同類型的比較性記分系統一如百分等級、字母等第一有三種特徵：(1)具有兩種以上有區別性的等級，(2)成績的決定是限定於一定的時間內完成，(3)並非所有的學生都會得到最高成績(Laska & Juarez, 1992b)。

關於比較性記分的爭議很多，有些學者認為必須修正或廢除現行的，他們認為比較性的記分是具有破壞性的、誤導的、浮面的、不可靠的、具有區別性的、易引起焦慮的、以及溝通不良等缺點；比較性記分干擾學習，這是由於它忽略對於單元的精通(Morrison, 1926)，製造學生焦慮以及產生對學校的厭惡感(Warren, 1975)，提供欺騙的機會(Blynt, 1992)，逼迫學生花時間及精力在成績上而非學習上(Hargis, 1990)，使得老師和學生相互敵對(Kilpatrick, 1934)，傷害低成就學生的自我形象。反之，另有一派學者支持比較的記分，他們認為這種制度可以確保社會進步及社會穩定；因為好成績是學生努力達成的標竿(Brubacher, 1952)，成績扮演著重新分配有限的教育資源的角色(Ruediger, 1911)，成績作為篩選及安置學生進入高一層級學校及進入特殊學程的依據(Warren, 1975)。同時，支持者也強調成績是必須的；若沒有成績，學生便不願努力讀書也不能明瞭學校活動的意義；況且有效的學習亟需成績作為回饋(Brubacher, 1952; Colvin, 1912)。

2. 精熟式的記分(mastery grading)

直到二十世紀初期這種記分系統才受到明顯地注意。摩勒森(Henry C. Morrison)是精熟式記分理論的一位重要先驅者，他於「中等教育的教學實施」(The Practice of Teaching in the Secondary School, 1926)一書中奠立了精熟理論的基

礎。摩勒森的理论是基於以下假設——教育是個人適應世界的一種過程。依摩勒森的觀點，所謂「適應」是指學習「如何生活」。往後精熟理論的支持者以摩勒森的理论為依歸並加上個人的見解。例如拉斯卡(Laska, 1991)提出，學校的主要目的是幫助每一個學生達成有必需性的學習目標。相同地，傑瑞(Juarez, 1994)說，精熟式的記分目的在於使每一位學生都有成功的可能性。懷汀和雷德(Whiting & Render, 1987)強調，假使每一位學生都明瞭學習目標，被教以最好的學習方式，給予修正錯誤的機會，被允許表現新的學習成果，那麼他們都願意並且能夠學習。

所謂「精熟」是指幫助每一位學生達成必需性的學習目標(essential learning objectives)；也就是說，學習者能100%完全學到所頒佈的精熟課程(mastery curriculum)。依摩勒森的話，「只有當學生完全獲得一項學習可以說是精熟。一半的學習，或者學得相當好，或者正在學都不能算是精熟；換句話說，『精熟』指完完全全做完一件事。」(Morrison, 1926)。摩勒森的「精熟公式」(mastery formula)——「預測、教學、測驗、調整步驟及重新測驗」——提供教學與學習有系統的架構。

凱羅(John B. Carroll)的文章——學校學習的模型(A Model of School Learning, 1963)——對於精熟理論的發展影響很大。傳統觀念認為學生的適性(aptitude)代表學習一特定科目的水準。此一觀點受到凱羅的挑戰；他建議小孩的適性是學習速率的指標。換句話說，每一位學生都有能力精熟學校的課程，但所需的時間會有不同。當適性被視為學習速率的一項指標，孩子應被視為快的或慢的學習者，而非好的或差的學習者。根據這種對於適性(aptitude)的觀點，凱羅又提出一個學校學習的模式。他相信學習者成功的程度完全是看他花了多少時間去做一件工作。假如給予充分的機會和時間去學習，大部分的孩子都能完成既定的學習目標。學習的程度可以用下列公式來表達：

$$\text{學習的程度} = f\left(\frac{\text{所花的時間}}{\text{所需的時間}}\right)$$

也就是說，學習的程度是真正花在學習的時間相對於學習所需要的時間之一種功能。凱羅又進一步確認了影響所花時間及所需時間的五項變數：適性(aptitude)或者學習的速率，對於指導的理解能力(ability to understand instruction)或者學習的速率，對於指導的理解能力(ability to understand instruction)，指導的品質(quality of instruction)，被允許用來學習的時間(time allowed for learning)，以及毅力(perseverance)。他認為所花的時間是由學習者的毅力及被允許用來學習的時間兩者所決定。另一方面，所需的時間是由學習的速率、教學的品質，以及對於教學的理解能力，這三者所決定。凱羅所提出的這一模式對於學習理論貢獻很

大，並且對於學習潛能及學校教育的潛能提出了樂觀的見解。凱羅所提出的五項變數後來被納入布魯門(Benjamin Bloom)的精熟理論裡。

直到布魯門的文章「精熟式的學習(Learning for Mastery)」(Bloom, 1976)以及凱樂(Fred Keller)的模式「個別化的教學系統(Personalized System of Instruction)」(Keller, 1968)兩篇文章的發表，精熟理論才變成教育上的重要特點。這兩種模式(LFM and PSI)有幾項共同點。第一、將教材分割成小的學習單元，並且事先明定表現的標準。第二、在每一學習單元教學完畢時，學生得參與形成性的測驗(formative tests)。第三、根據測驗的結果得知學生對於每一學習單元的理解程度，以及學生為了精熟每一學期單元所必須付出的修正性努力。第四、當學生精熟了一個學習單元就可以繼續下一個單元。

拉斯卡教授歸納出精熟教育的三點特色(Laska, 1988, 1991)。第一、需要有一套明確的精熟課程(mastery curriculum)。精熟課程是由可證明的(verifiable)，可達成的(achievable)及可教的(teachable)學習目標所構成。第二、老師必須提供學生彈性的學習時間以求達成所有必需性的學習目標。第三、老師必須付出時間及精力為學生作有系統的精熟教學；重新教導及重新測驗直到所有學生都學會了精熟課程。

綜合上述，精熟式的記分奠基於精熟教學理論。精熟教學理論是由摩勒森發展出來的，經由凱羅(Carroll)、布魯門(Bloom)、凱樂(Keller)及拉斯卡(Laska)加以闡述發揚。這些精熟理論學家致力於設計精熟式的記分系統，其目的在於鼓勵所有學生完成必需性的學習目標。精熟式的記分系統有三種特色(Laska, 1991)：第一、這種記分系統只有單一的記分符號，它表示學生已經成功地學習；也就是說，學生已完成了必需性的學習目標。第二、成績的決定並不限定於既定的時間內，而是有充分的彈性時間。第三、期望每一位學生最後都能成功。

(二)各種不同的記分符號系統

在兩種基本的記分途徑之下(精熟式記分及比較性記分)，教師使用各種不同的記分符號系統，以便評量學生表現。各種記分符號系統包括：數字成績(numerical grades)、字母等第(letter grades)、通過/不及格(pass/fail)、敘述性的報告(narrative reports)、契約式的記分(contract grading)、自我評量(self-assessment)、以及學習目標查核表(learning-objective checklists)。

1. 數字成績(numerical grades)

在這種記分系統之下，每位學生被指定0-100中的一個數字，這個指定的數字必須相當於所學得教材的百分比(Geisinger, 1982)。因最高分是100，和最高百分等級100相同，因此數字成績系統又叫做「百分等級」系統。這種系統在十九世紀後半期及二十世紀初最普遍(Smith & Dobbin, 1960)。它有二項優

點，第一、數字成績對於學生表現提供了簡便的摘要記錄。第二、數字可以直接平均而不需轉換到不同的量表。例如一百等分的數字成績太過於精密，因而遭到很多批評。佩羅斯特 (Probst, 1988) 說，即使學期成績相差二至三分也不具任何意義。奧因斯坦 (Ornstein, 1989) 說，若將數字轉換成字母則會扭曲成績的意義。因為這些缺點，於 1920 年代，數字成績的使用逐漸下降。

2. 字母等第 (letter grades)

自從 1930 年代以來，字母等第是最普遍的記分系統 (Smith & Dobbin, 1960)。每一字母代表學生表現在某一特定量表上的位置。這個量表的範圍包含從「優秀的」到「失敗的」表現。字母成績及數字成績兩者共同的優點是成績本身具有易於登錄的、易於處理的、以及易於比較的特質。詳細地說，這個記分系統有下列優點 (Worthen, Borg, & White, 1993)，第一、它提供了關於學生表現的簡便摘要性記錄。第二、通常五等級的成績—A、B、C、D 和 F—可以經由附上「+」或「-」號而改變成 A-、B+、B-、C+、D+ 等成績。第三、字母成績很容易轉換成學業平均成績 (grade point average) 以及換算成百分等級成績。

3. 通過/不及格 (pass/fail grading)

在 1960 年代早期，很多高中及學院引入了通過/不及格的記分方法 (Worthen, Borg, & White, 1993)。在小學階段，通過/不及格，常以滿意/不滿意 (Satisfactory/Unsatisfactory) 來代替。通過/不及格記分方法的理論根據如下。第一、以通過/不及格為記分基礎的課程，較能使學生專心於學習而非成績；學生因此敢修比較有挑戰性的科目而不必害怕會影響到學業平均成績 (GPA)。第二、在學校中不應過份強調成績；因為在非學術領域中，成績是不重要的。第三、過度關心好成績妨礙了學生去修習一些對他有益的科目。第四、爭取好成績的壓力干擾了學習並且鼓勵欺騙 (Whorthen, Borg, White, 1993)。然而，這種記分系統也有缺點。第一、它提供太少的訊息。父母親及學校人員可能希望得到更詳細有關成就的訊息，而不是只有通過/不及格這種粗略的二分法記分。另外，學生也需要有關於學業成就更精密的判斷，以便幫助計劃未來的學業或職業生涯。第三、學生可能只考慮通過某一課程而付出最低限度的努力 (Hills, 1976)。

4. 其他非懲罰性 (non-punitive) 的記分方法

當學生在某一學科表現得不夠理想時，非懲罰性的記分方法並不會把學科成績登錄下來，也不會將它保留在學生的永久記錄上 (Geisinger, 1982)。這種記分方法所關心的是，學生不該因表現不夠好而遭到懲罰；也就是說，若學生表現不好就只是給予零學分 (no credit)。當某一學科不及格時，有些學校將它登錄為「零學分」，有些學校甚至不會在學生的永久記錄上註明這門課。其他類

型的非懲罰性記分是允許學生退掉某一門課；當學生覺得表現得不夠好，便可以退掉這個科目 (Oliver, 1971)。很多學生因這種措施而提升了學業平均成績。

5. 敘述性的報告 (narrative reports)

敘述性的報告對於學生在某一學科的學習、技巧、能力、以及表現提供了質的或敘述性的描述。和一般的推薦信一樣，敘述性的報告描述了過程 (process) 及成果 (product) 兩者；換句話說，它包含了有關學生如何學習 (how) 以及學了多少 (how much) 等的訊息。同時，這種記分方法鼓舞老師多思考學生為一個體而非只想到成績簿上的數字。然而，對老師而言，針對每一位學生的表現所做的敘述性報告是一項很大的負擔；書寫敘述性報告是十分花時間的，況且並非每一位老師都懂得如何書寫一份有意義的、有幫助性的、以及個別化的敘述性報告。

6. 契約式的記分 (contract grading)

契約式的記分採用個別化的既定標準 (Hassencahl, 1979)。記分所依據的個別化標準，規範於契約中。這種契約式的記分是經由老師和個別學生共同協商而訂定出學生所應完成的工作以及所應達到的表現水準 (Geisinger, 1982)，它具備下列優點。第一、允許個別化的學習以及鼓勵學習的多樣性。第二、因學生已明瞭預期的標準，所以能去除學習的焦慮。

7. 自我評量 (self-assessment)

自我評量包括兩種類型：自我評估 (self-evaluation) 及自我記分 (self-grading) (Geisinger, 1982)。所謂的「自我評估」指學生在與老師討論之後，對於自己的進步所做的評估。所謂的「自我記分」指學生決定自己的成績。通常自我記分，發生在自我自我評估之後 (Kirschenbaum, Simon & Hapier, 1971)。自我評量的優點是，它提供學生評估自己長短處的機會，並且允許學生在設定教育目標及尋求目標方法上負起責任。然而，這種系統也有缺點。第一、因為成績是由學生自己決定的，成績變得較不精確，並且失去了某種程度的比較性。第二、為了爭取好成績的學生，可能高估自己的表現，進而難以做到誠實地自我記分。因此，自我評估及自我記分尚未被廣泛地使用。

8. 學習目標查核表 (learning-objective checklists)

查核表是學習目標的一份清單檔，某一目標達成時，老師便在上面做個記號。為了提供更詳盡的資料，查核表也包括一系列特質，例如舉止、態度、意願、專心等 (Hills, 1976)，查核表通常被視為其他記分方法的補充而非替代品。查核表作為一項補充的優點是增加訊息。但是，發展一個好查核表並不是一件簡單的事。此外，查核表的結果很難與其他成績相平均而形成總結性記錄。

就德州學區而言，成績可以以數字呈現或者轉換成字母等第(TAC § § 75.191)。對於二至十二年級學生，學業成就是以數字分數 0-100 表示。至於二年級以下，學區可以使用其他記分系統記分。在二至六年級，數字成績使用於語言、藝術(arts)、數學、科學、以及社會研究等科目。至於其他學科包括健康教育、美術(fine arts)、以及體育，可以由學區政策決定使用數字成績、字母等第、或者評論性的文字(如優秀、滿意、不滿意、需要改進)等記分符號系統記分。二年級以下學生，可以使用評論性的文字如：「優秀、滿意、不滿意、需要改進」報告成績(TAC § § 75.191)。

五、記分的參照標準(Standards for Designing Grades)

記分系統除了可以以記分符號區分外，也可以以參照標準區分。三種最常使用的參照標準是標準參照(criterion-referenced standard)、常模參照(norm-referenced standard)、以及自我參照(self-referenced standard)。

(一)標準參照的記分標準(criterion-referenced standard)

標準參照的記分標準又稱為「絕對標準」(absolute standard)，常使用於精熟記分系統中。這種記分方法下，成績是依某一絕對的或固定的標準而定(Docking, 1986)，換句話說，成績是視學生的成就相對於事先設定的標準而定的，而非相對於其他學生的表現。就某方面來說，標準參照的記分方法多多少少有和其學生比較的成份在內，因為既定的標準乃是參照先前學生的表現而訂定的(Geisinger, 1982; Worthen, Borg, & White, 1993)。這種記分方法的一個優點是，只要學生夠努力，而且老師的教學技巧夠好，那麼所有的學生都有可能得到最好的成績。此外，藉著事先設定的標準，老師可以向學生清楚地、有效地傳達對於學業成就的期望。但是，這種記分方法的實施也有困難，客觀標準的設立常使老師感到棘手，設得太高或太低都不得宜。

(二)常模參照標準(norm-referenced standard)

常模參照標準又稱「相對標準(relative standard)」；也就是說，個人學業成就和他人的學業成就相比較後才決定成績。這種記分法的基本假設是，任何人類的特質是以常態分配於群體中。因此，個人的成就只有當與他人的成就相比較時才有意義。相對記分方法是以曲線(curve)為準。早在 85 年以前，美爾(Meyer)就提倡以曲線為記分的基準。直到 1920 和 1930 年代這種方法才變得較為普遍(Meyer, 1988)。構成曲線的常態母群體(normative population)可以是一個班級、某校的一個年級、或者某一特定年級的全國性樣本(Thorndike, 1969)。當常模的樣本愈大時，就愈趨向以常模參照和標準參照為組合的記分(Geisinger, 1982)。

一般人比較容易了解及接受常模參照的記分方法(Cohen, 1982)。然而，這種

記分法也有缺點。摩勒森批評，使用常模參照只是將學生按表現優劣加以排名而非鼓勵有效的學習(Cited in Sparkman, 1992)。學生所得的成績並不是依他對於教材的精熟程度而定，而是依據他相對於其他學生的表現而定。不論學生表現得多好，總有一定比例的人會不及格；同理，不論學生學得如何差，總有一定比例的人會得到 A+。因此有學者建議於課堂內不該採用這種記分方法(Worthen, Borg, & White, 1993)。

(三)自我參照的標準(self-referenced standard)

有兩種常用的自我參照的標準——以成長(growth)為憑藉，或者以學生的能力(student's ability)為憑藉(Terwilliger, 1977)。這種方法的價值在於體認個別差異、對於努力加以獎勵、並且提供能夠增長興趣及動機的環境。當成長被作為記分的憑藉時，會遇到有關「成長」這個概念的技術性問題(Terwilliger, 1977)。首先，為了判斷學生的成長，必須在初期及稍後階段測量。這兩次測量必須在內容及難度上一致，才能將前後測量的差異解釋為成長的指標。第二、起初的表現水準和成長量之間有很大的負相關。由以下公式可以看出：「成長 = (最後的表現水準) - (初期的表現水準)」。由此可知，一個具備較低初期水準的學生能顯示較高的成長量；反之，一個具備較高初期水準的學生卻顯示出較低的成長量。第三、以學生的成長作為記分的憑藉時，很容易導致評量的標準不夠明確，而且評量的過程容易受老師的偏見所左右(Terwilliger, 1971)。

以學生作為記分的憑藉，也牽涉到嚴重的技術性問題。例如老師常常會用主觀的態度來判斷學生的能力，因此對於學生能力的判斷難免受個人偏見所影響。此外，在某些情況下由於測量的誤差、所預測的成就之異質性、對於能力的測量之異質性、或者個人特質與教學型態間的交互作用，導致能力與成就間的相關性並不強(Terwilliger, 1971)。因此特維勒傑(Terwilliger, 1971)並不認為成長或能力是記分的良好憑藉。

六、成績的決定要素及來源(Determinants/Sources of Grades)

(一)成績的決定要素

在每一種記分系統中，某些學生的特質常是記分的參考要素。這些特質包括與成就相關的因素(achievement-related factors)及非成就相關的因素(nonachievement-related factors)。後者如能力、態度、動機、努力、進步、行為、名聲、舉止、課堂上的參與、種族、社會階層、性別等等。隨著學校層級的提高，記分的過程由多重決定要素轉為單一要素。這個單一要素，就是學業成就(Warren, 1975)，例如，在小學階段，非成就相關因素如出席率、行為、愛好乾淨等和其他因素如測驗分數、智商、以及課堂參與討論等對於老師記分有相同程度的影響

(Hesburgh, 1971)。在中學階段，臨時測驗的結果及作業的完成提供了記分的重要基礎 (Hesburgh, 1971)。

學業成就是這些因素中唯一最受推崇的記分要素。就如同桑代克 (Thorndike, 1969) 定義成績為「教育某一領域成就的單一摘要性記號」。根據這個定義，成績應該以學業成就為唯一的根據。相似地，安迪生及王奪 (Anderson & Wendel, 1988) 建議，對於學生能力、人格、行為、出席率等的評估不應納入成績，因為成績只是用來反應學業成就。史地傑等人 (Stiggins, Frisbie, & Griswold, 1989) 認為，以學業成就為基準的記分不僅使成績的意義與解釋較為清楚，而且使得學習目標及老師的期望較易於溝通。特維勒傑 (Terwilliger, 1971) 相信，根據學業成就記分可以使學生適應非學術的世界；因為在校外，學生的學業成就常是外人用來判斷學生能力的指標。

另一方面，有些學者主張除了學業成就外，很多其他因素如學生的成長、進步、態度、能力、及人格特質等，都應納入成績。持這一主張的論點是，我們所評估的學習量 (the amount of learning) 而非所達到的精熟程度 (the level of mastery attained) (Terwilliger, 1971)。丹登 (Denton, 1989) 認為成績是學生進步的一項記錄，也是向所有關心成績者傳達這項進步的一種工具。林氏蘭 (Rinsland, 1937) 倡導成績的純淨 (purity) 以及信度 (validity)；依他的觀點，成績是在任何學習情境下對於每一項學習所進行的完全、可信、及正確的測量。基於這個定義，教師應該正確地測量每一階段的學習及成就。

此外，一些研究指出，其他因素如老師的偏見和期望、老師對於學生行為的感受、學生的舉止等都會影響老師對於學生成就的判斷。陳氏的研究 (Chen, 1993) 顯示出新老師的記分比較容易受對於個別學生的印象所左右。卡特的研究 (Carter, 1952) 也指出男生比同樣程度的女生容易獲得較好的成績。柏納特等的研究也發現行為表現良好的學生比行為差的學生容易被判斷為成績優秀的學生 (Bennett, Gottesman, Rock, & Cerullo, 1993)。

(二) 成績的來源 (Sources of Grades)

很多學者認為成績是由學業成就這一單一因素決定，而學業成就的記錄應該僅限於測驗成績 (Clark & Nelson, 1991)。另有一些學者認為學校成就應該由各種不同活動的分數決定 (Jongsma, 1991)，特維勒傑 (Terwilliger, 1989) 指出教學目標可由小考、測驗、家庭作業、教室表現等評估而得。他更進一步指出，這些成績的來源應該以量化的形式呈現；特別是小考及測驗，應該根據明確的分數基模而以分數、或點數報告出來；至於家庭作業、指定的計畫或者學期報告則應依據確切的量化評量系統來打分數；其他如課堂表現或展示，則應使用數字評量表 (numeric ratings) 或查核表來評估。

七、成績報告的途徑

在完成評估學生成就以及決定成績之後，老師需要將這些成績報告出來。三種成績報告的途徑是成績報告卡 (report card)，家長會 (parent-teacher conferences) 以及給父母的書信。

(一) 成績報告卡

成績報告卡是學期中用來定期報告學業成就以及學期成績的方法。有效的成績報告卡是由什麼構成的呢？第一、成績報告卡必需明示評量的目的、步驟及標準 (Walling, 1975)。第二、成績報告必須是直接而明瞭的。第三、成績報告卡需要預留寫評語的空間 (Robinson & Craver, 1989)。在小學階段，成績報告卡的功能除了報告學科的學業成就外，還傳達學生的社會發展及出席率等訊息 (Perkins & Buchanan, 1983)。

(二) 家長會

家長會被視為成績報告卡的補充 (Hills, 1976)；它能提供父母親更多關於孩子的訊息。家長會的實施，在小學階段通常是針對所有學生，然而在中學階段則只限定於受懲戒的案例 (Ahmann & Glock, 1971)。家長會有很多的優點，例如它允許老師和家長間充分地溝通，允許對於成績意義的釐清，進而獲得父母及老師雙方的支持以便協助孩子更有效率地學習。

(三) 給父母的書信

給父母的信函是成績報告的另一種補充。書信通常以敘述性的方式描述個別學生的學習、技能及能力；也因此它能反映出每一位學生的獨特性 (Hopkins & Antes, 1990)。如同家長會，書信也能提供建設性的意見。然而，它卻未能允許雙向的溝通。另外，很多老師不能夠寫好書信；這大半因為缺乏用來書寫的簡便摘要性記錄。因上述這二項缺點，書信的使用尚未普遍。由上可知，家長會及書信需要花費老師很多的時間和精力；不僅需要時間主持會議或寫信，更需要很多的準備工作。準備工作包括訓練老師有關的實施方法及步驟，以及事先觀察和蒐集資料以便評估學生成就。只有成績報告卡是比較省事而簡便的，因而廣泛使用於大學階段前的各級學校 (Robinson & Craver, 1989)。

八、新的評量方法

由關心傳統記分的意義而引發了新的學業評量法，其目的在於提升高層次的學習以及著重在診斷學習問題，而非只重在記分——僅是為成功的或不成功的學生貼上標籤。亞克伯 (Doug Archbald) 及紐曼 (Fred Newmann) 在他們的書中——「超越標準測驗」 (Beyond Standardized Testing, 1992) 一聲稱學校成績傳達了很少有關學生

特殊成就的品質，而且所測量的學習類型常常是十分瑣碎而毫無意義的。雷勒 (Agnes Lathel) 在「紙筆測驗」(Written Examinations, 1889) 一書中批評，一般的記分措施鼓勵老師以填鴨式的方法向學生灌輸事實。泰勒 (Ralph W. Tyler) 在「評量：對進步教育的挑戰」(Evaluation: A Challenge to Progressive Education, 1935) 指出學校太過於仰賴記分措施；學生成績只反映了有限的教育成長，但卻無助益於其他方面的成長，例如區分閱讀教材與藝術作品的品味。泰勒因此提出「擴大概念的學生評量」(enlarged concept of student evaluation)——評量是根據紙筆測驗外所搜集的充分訊息。相同於此觀念，亞克伯及紐曼指出「真正的學業成就」(authentic academic achievement) 應具備三種特徵：(1) 談話、事物、或表現的產生 (the production of discourse, things or performances)；(2) 時間的彈性運用；(3) 與他人合作 (Archbald & Newmann, 1992)。

有很多名詞已被用來描述新式的學業評量方法。較常用的名詞如「變通性評量」(alternative assessment)、**「真實評量」**(authentic assessment)、**「直接評量」**(direct assessment)、以及**「表現評量」**(performance assessment)，這些評量方法都被視為有別於傳統的選擇題及標準化測驗 (Meyer, 1988)。此外這些新的評量方法有下列幾點共同特徵：

第一、這些新評量方法的理念是，學生成就的直接評量優於非直接評量。直接評量是指於真實生活情境或相似的生活情境下，評估學生在重要及相關工作上的表現 (Perrone, 1991; Wooster, 1993)。被評量的工作包括寫作的樣本、學生的表演及工藝品、學生所提出和解決的問題等等，這些工作被視為有價值的智慧結晶 (Grady, 1992)。

第二、變通性評量及真實評量是經由設計使成為學生經驗的一部分：它和教室內的活動緊密連結著，提供機會得以便修正錯誤及缺點，鼓勵學生朝著建立完美的結果而努力。這種「產生—修正—完美」(create-revise-perfect) 的過程使學生具備更合適的校外生活經驗 (Grady, 1992; Wolf, Bixby, Glenn, & Gardner, 1991)。

第三、變通性評量及直接評量重新肯定老師（而非測驗設計者）為學生學習的主要引導者。老師設計、操作及評估所設計的評量，把診斷學生成就及學習需求當成有組織的評估過程的一部分。因為評量者有協助學生滿足這些學習需求的任務，因此評量者也需要設計合適的教學策略，使成為評量的一部份。

第四、變通性評量及真實評量允許學生直接參與評量工作及評量標準的設立；這不僅讓學生有使用高層次思考的機會，並且能夠積極參與自己的學習，為自己的學習負責任 (Neuman, 1993)。

參、對於我國中小學學業成績評量的啓示與建議

綜觀美國中小學的學業記分政策，大底上都能充分闡述記分的哲學基礎、記分的目的、記分符號系統、記分符號的意義、成績報告的方式、記分的參照標準、記分的決定要素與來源，以及記分要素的比重。因此，在我國教育實施上的建議，首先，呼籲建立一套易瞭解而明確的記分政策。記分政策應該以書面方式呈現，並且將它傳送給相關人員（如老師、學校行政人員、家長、學生、及輔導員）。目前我國國民中小學學生成績考查依德、智、體、群、美五育分別辦理。「國民中小學學生成績考查辦法」中規範了記分的決定要素及來源（如鑑賞、晤談、報告、表演、實作、資料蒐集整理、設計製作、作業、紙筆測驗、實踐等）、記分符號系統（即百分等級、等第——優、甲、乙、丙、丁）、記分的參照標準（即常模參照標準，以班級或年級為參照團體）。除了這三項內涵外，完善且易瞭解的成績考查辦法也應闡述記分的哲學基礎、記分的目的、每一記分符號所代表的意義、成績報告的方式、以及每一記分要素/來源所佔的比重等。

第二、就記分的哲學基礎而言，學業記分不應淪為人為的獎懲制度，使用區別性的成績以達成學校的選擇性功能。理想上，學生的個別化表現應受到尊重；也就是說，學業記分變得人性化，將學生的成就水準參照個人的能力與成長，如此一來，成績可以幫助個人成長，不再是一種獎懲的工具。

第三、成績評量的最重要功能乃是溝通性的功能，而非僅限於行政上的功能。當今我國中小學十分強調以成績作為學校行政決定的參考，如升級與留級、獎學金的決定等，而忽略了成績的主要功能乃是提供有關學生學習的訊息，以作為提供協助與努力的參考。

第四、現在普遍使用的記分乃是一種比較性記分，比較性記分易於干擾學生的學習，製造學生焦慮，產生對學校的厭惡感，逼迫學生花時間及精力在成績上而非在真正的學習上，使得老師和學生相互敵對，傷害低成就學生的自我形象。雖然比較性記分同時也具有正面的價值；例如成績使學生願意努力讀書並且明瞭學校活動的意義，成績的回饋是有效學習的必要條件，成績可以作為選擇與輔導的參考。因此如何妥善運用學業評量，以發揮其正面功能而減少其副作用乃是需要努力的目標。

第五、學業成就——也就是知識或技巧的獲得——應該是智育成績的唯一決定要素。依據學業成就而記分，不僅使得成績的意義與解釋更明確，也使得老師的期望更容易清楚地向學生傳達。其它的學生特質——如學習能力、努力、態度——很難明確地界定，更難有效地評估。當個別老師對這些要素（如能力、態度）賦予不同意涵及不同比重，容易造成成績內涵的混淆。因此這些與成就相關的因素——如

能力、努力、態度——都不應納入智育成績的考量，而是應分開來記錄。

最後，學業成就的評量應該根據紙筆測驗以外的其它評量方法，如另類評量、真實評量、表現評量、或檔案評量。這些新的評量方式把評量活動轉移到更接近真實的工作情境中，在此評量活動中，學生有機會展示能力於真實生活情境中，而評量即發生在整個學習過程中。相反地，傳統的評量（如紙筆測驗）僅發生在非真實的情境中，著重在結果而非學習的過程，而且這種學習是瑣碎且無意義的。

參考書目

- Ahmann, J.S., & Glock, M.D. (1971). *Measuring and evaluation educational achievement*. Boston: Allyn and Bacon Inc.
- Anderson, K.E., & Wendel, F.C. (1988). Pain relief: Make consistency the cornerstone of your policy on grading. *The American school board journal*, 175(10), 36-37.
- Archbald, D.A., & Newmann, F.M. (1992). Beyond standardized testing. In John. A. Laska & T. Juarez (Eds.), *Grading and marking in American schools: Two centuries of debate* (pp. 143-148). Springfield, IL: Charles C Thomas.
- Bennett, R.E., Gottesman, R.L., Rock, D.A., & Cerullo, F. (1993). Influence of behavior perceptions and gender in teachers' judgments of students' academic skills. *Journal of educational psychology*, 85(2), 347-356.
- Bloom, B.S. (1976). *Human characteristics and school learning*. New York: McGraw-Hill.
- Blynt, R.A. (1992). The sticking place: Another look at grades and grading. *English journal*, 81(6), 66-71.
- Brubacher, J.S. (1952). *Eclectic philosophy of education*. New York: Prentice-Hall, Inc.
- Carroll, J.B. (1963). A model of school learning. *Teachers college record*, 64 (8), 723-733.
- Carter, R.S. (1952). How invalid are marks assigned by teachers? *Journal of educational psychology*, 43, 218-228.
- Chen, C.L. (1993). *An investigation of university and college instructors' grading practices in Taiwan*. Unpublished doctoral dissertation, The University of Texas, Austin, TX.
- Clark, H.C., & Nelson, M.N. (1991). Evaluation: Be more than a score-keeper. *Arithmetic teacher*, 38(9), 15-17.
- Cohen, S.B. (1982). Assigning report card grades to the mainstreamed child. *Teaching exceptional children*, 15(1), 86-89.
- Colvin, S.S. (1912). Marks and the marking system as an incentive to study. *Education*, 32, 560-572.
- Denton, J.J. (1989). Selecting an appropriate grading system. *The clearing house*, 63(3), 107-110.
- Docking, R.A. (1986). Criterion-references grading techniques. *Studies in educational evaluation*, 12, 281-294.
- Durm, M.W. (1993). An A is not an A is not an A: A history of grading. *The educational forum*, 57(3), 294-297.
- Ebel, R.L. (1980). Evaluation of students: Implications for effective teaching. *Educational*

- evaluation and policy analysis*, 2(1), 47-51.
- Frisbie, D.A., & Waltman, K.K. (1992). Developing a personal grading plan. *Educational measurement: Issues and practices*, 11(3), 35-42.
- Gage, N.L., & Berliner, D.C. (1988). *Educational psychology* (4th ed.). Boston: Houghton Mifflin.
- Geisinger, K.F. (1982). Marking systems. In H.E. Mitzel (Ed.), *Encyclopedia of educational research* (5th ed., pp.1139-1149). New York: The Free Press.
- Grady, E. (1992). *The portfolio approach to assessment*. Bloomington, IN: Phi Delta Kappa.
- Hargis, C.H. (1990). *Grades and grading practices*. Springfield, IL: Charles C Thomas.
- Hassencahl, F. (1979). Contract grading in the classroom. *Improving college and university teaching*, 27, 30-33.
- Hesburgh, T.M. (1971). Grading systems. In L.C. Deighton (Ed.), *Encyclopedia of education: Vol.1* (pp.181-185). New York: The Macmillan Company & Free Press.
- Hills, J.R. (1976). *Measurement and evaluation in the classroom*. Columbus, OH: Merrill.
- Hopkins, C.D., & Antes, R.L. (1990). *Classroom measurement and evaluation* (3rd ed.). Itasca, IL: F.E. Peacock Publishers, Inc.
- Jongsma, K.S. (1991). Rethinking grading practices. *The reading teacher*, 45(4), 318-320.
- Juarez, T. (1994). Mastery grading to serve student learning in the middle grades. *Middle school journal*, 26(1), 37-41.
- Keller, F.S. (1968). Good-bye teacher ... *Journal of applied behavioral analysis*, 1, 79-89.
- Kilpatrick, W.H. (1934). *Source book in the philosophy of education* (revised edition). New York: The Macmillan Company.
- Kirschenbaum, H., Simon, S.B., & Napier, R.W. (1971). *The grading game in American education*. New York: Hart.
- Laska, J.A. (1988). A new paradigm for teaching and curriculum. *New education*, 10(1), 76-80.
- Laska, J.A. (1991). *Instructional therapy*, course handouts, The University of Texas at Austin.
- Laska, J.A., & Juarez, T. (1992a). Overview. In J.A. Laska & T. Juarez (Eds.), *Grading and marking in American schools: two centuries of debate* (pp.3-6). Springfield, IL: Charles C Thomas.
- Laska, J.A., & Juarez, T. (1992b). Mastery grading. In J.A. Laska & T. Juarez (Eds.), *Grading and marking in American schools: two centuries of debate* (pp.43-44). Springfield, IL: Charles C Thomas.
- Lathel, A.M. (1989). Written examinations — their abuse, and their use. *Education*, 9, 452-456.
- McDaniel, T.R. (1979). The teacher's ten commandments: School law in the classroom. *Phi Delta Kappan*, 60(10), 703-708.
- McKeachie, W.J. (1986). *Teaching tips: A guidebook for the beginning college teacher*. Lexington, MA: Heath.
- Meyer, M. (1988). The grading of students. *Science*, 27, 243-250.
- Morrison, H.C. (1926). *The practice of teaching in the secondary school*. Chicago: University of Chicago Press.

- Neuman, D.(1993). Alternative assessment: Promises and pitfalls. *School library media annual*, 11, 13-21.
- Oliver, F.E.(1971). *The AACRAO survey of grading policies in member institutions*. Washington, DC: American Association of Collegiate Registrars and Admissions Officers. (ERIC Document Reproduction Service No. ED 055 546)
- Ornstein, A.C. (1989). The nature of grading. *Clearing house*, 62(8), 365-369.
- Perkins, J., & Buchanan, A.(1983). *How parents find out about student progress*. Los Alamitos, CA: Southwest Regional Laboratory for Educational Research and Development. (ERIC Document Reproduction Service No. ED 257 787)
- Perrone, V.(1991). Moving toward more powerful assessment. In V. Perrone (Ed.), *Expanding student assessment*(pp.164-166).
- Plato (1968). *The republic*, 415B-C (A. Bloom Trans.). New York: Basic Books.
- Probst, R.E.(1988). *Response and analysis: Teaching literature in junior and senior high school*. Portsmouth, NH: Heinemann.
- Rinsland, H.D. (1937). *Constructing tests and grading in elementary and high school subjects*. New York: Prentice-Hall, Inc.
- Robinson, G.E., & Craver, J.M. (1989). *Assessing and grading student achievement*. Arlington, VA: Education Research Service.
- Ruediger, W.C.(1911). *The principles of education*. Boston, MA: Houghton Mifflin.
- Smallwood, M.L. (1935). *A historical study of examinations and grading systems in early American universities*. Cambridge, MA: Harvard University.
- Smith, A.Z., & Dobbin, J.E.(1960). Marks and marking system. In C.W. Harris(Ed.), *Encyclopedia of educational research* (3rd ed., pp.783-791). New York: Macmillan.
- Smith, M.S., & Smittle, W .R.(1954). *The board of education and educational policy development*. Ann Arbor, MI: Edwards Brothers, Inc.
- Spady, W.(1992). On grades, grading and school reform. In John J.A. Laska & T. Juarez (Eds.), *Grading and marking in American schools: Two centuries of debate* (pp.67-76). Springfield, IL: Charles C Thomas.
- Sparkman, R.S. (1992). *A case study of mastery grading in an elementary school*. Unpublished doctoral dissertation, The University of Texas, Austin, TX.
- Stiggins, R.J., Frisbie, D.A., & Griswold, P.A.(1989). Inside high school grading practices: Building a research agenda. *Educational measurement: Issues and practice*, 8(2), 5-14.
- Terwilliger, J.S. (1971). *Assigning grades to students*. Glenview, IL: Scott, Foresman.
- Terwilliger, J.S. (1977). Assigning grades — Philosophical issues and practical recommendations. *Journal of research and development in education*, 10(3), 21-39.
- Terwiller, J.S(1989). Classroom standard setting and grading practices. *Educational measurement: Issues and practice*, 8(2), 15-19.
- Texas Administrative Code, § § 75.191(1994). Grading and reporting requirements.
- Thorndike, E.L. (1969). Marks and marking systems. In R.L. Ebel (Ed.), *Encyclopedia of educational research*. (4th ed., pp.759-766). New York: American Educational Research Association.
- Tyler, R.W. (1935). Evaluation: A challenge to progressive education. *Educational research bulletin*, 14, 9-16.

- Tyler, R.W.(1986). Changing concepts of educational evaluation. *International journal of educational research*, 10(1), 1-113.
- Tuttle, E.M. (1952). Written polices for boards of education. *American school board journal*, 124(6), 5-6.
- Warren, J.R. (1975). *The continuing controversy over grades*. Princeton, NJ: ERIC Clearing House on Tests, Measurement, and Evaluation. (ERIC Document Reproduction Service No. ED 117 193)
- Whiting, B., & Render, G.F. (1987). Cognitive and affective outcomes of mastery learning: A review of sixteen semesters. *Clearing house*, 60(6), 276-280.
- Wolf, D., Bixby, J., Glenn, J., & Gardner, H.(1991). To use their minds well: investigating new forms of student assessment. *Review of research in education*, 17, 31-74.
- Wooster, J.S. (1993). Authentic assessment: A strategy for preparing teachers to response to curricular mandates in global education. *Theory into practice*, 32(1), 47-51.
- Worthen, B.R., Borg, W.R., & White, K.R. (1993). *Measurement and evaluation in the schools*. New York: Longman.
- Wrinkle, W.L.(1935). School is marks — Why, what and how. *Educational administration and supervision*, 21, 218-225.

楊惠琴，國立成功大學教育研究所副教授