科學教育學刊
2002, 第十卷第四期, 407-421

Chinese Journal of Science Education
2002, 10(4), 407-421

發展多元評量模糊複合分數之初探

羅豪章
中山醫學大學 教育學程中心

(投稿日期：民國 90 年 2 月 27 日，修訂日期：90 年 12 月 11 日，接受日期：91 年 9 月 17 日)

摘要：隨著評量測驗的時代演進，傳統的單一標準化測驗已逐漸為多元評量所取代。然而，在多元評量強調評量工具、評量方式以及評量標準之多元性的前提下，教師如何依據多元化的評量結果進行複合分數的計算，並對學生的學習情形之良窳提出適當的建議，將是教學的首要工作，也是教育研究者所當深思的課題。本研究旨在利用模糊理論符合人類語意描述與思考邏輯之特性，捨棄線性且絕對性的傳統分數處理方式，改以非線性的模糊邏輯推理方式，發展一個整合多元評量分數的運算機制，經由該機制所運算所得的結果研究者稱其為“模糊複合分數（fuzzy composite score, FCS）”，而該機制則稱其為“FCS 評分架構”。研究者並依據教學情境之可能性，設計一組研究模擬樣本分數，藉由模擬樣本分數為輸入，進行模糊複合分數之整合評分工作。經結果分析與討論可得知：利用本研究所發展之評分架構推論而得到的模糊複合分數，能夠與本研究之假設與目的相互吻合，並可看出本研究所提出之 FCS 評分架構對於不同量尺的評量分數之整合，具有其便利性與適切性。同時，研究者也對於專家資料庫的建置以及專業評分軟體之設計，提出相關後續研究之建議。相信以本研究為基礎，持續相關之後續研究，應該可以使模糊複合分數在實際教學評量上的實用性與普適性更向前邁進一步。

關鍵字：多元評量、模糊、模糊邏輯、模糊複合分數。

壹、緒論

一、研究背景與動機

教學上，評量最主要目的在於瞭解學生學習狀況，並據以檢視教師教學之良窳，進而對整體的教學活動進行反思與修正。隨著評量測驗的時代演進，傳統的單一標準化測驗為另類、多元的評量所取代；試題反應理論（item response theory, IRT）克服了古典測驗理論（classical test theory, CCT）高度依賴測驗群體的明顯限制（Hambleton & Rogers, 1991）；在僅注重學習整體表現的靜態評量逐漸被人質疑下，強調學習歷程軌跡的動態評量，開始在教室中萌芽；個人化、適性化的評量，逐漸被大家所重視、接受；評量試題中，已經由以往假設性（spurious）的類型，逐漸轉變成著重考量真實情境的真實評量（authentic assessments）。這些教育評量上的變革，我們不難在過去幾十年的研究文獻脈絡中，覓得有力的實證。

然而，在多元評量強調評量工具、方式以及標準之多元性的前提下，教師對學生所進行的評量工作，每次得到的結果可能因教學的目標、評量的目的而異，如何在一個學習階段結束，統整這些在不同學習情境下，經由不同的評量工具和方法（筆試、口試、學習中觀察、師生溝通對話等）所獲得之多項評量結果，進而對學生的學習成就整合出一個複合分數（composite score），並據以對學生的學習做出相關之價值評判，這是學校教育的必然工作，也是教育研究者所當深思的課題。

本研究有鑑於傳統使用統計方式描述學生學習成就時，對於分數的絕對性所產生之疑慮（如：評量結果 91 分的學生學習成就優於 90 分者？是否 60 分及格的學生代表學習成功，而 59 分者則代表表示學習失敗？），故研究者捨棄線性且絕對性的分數處理方式，改以較符合人類語意、思考形式的模糊理論，進行多元評量分數之統整。

二、研究目的

本研究基於多元評量教育時代的來臨，以及模糊理論符合人類語意描述語思考邏輯之特性，發展出一個整合多元評量分數的運算機制。此一機制以非線性運算取代傳統線性之處理方式，對多種量尺下所獲得的評量結果進行整合之工作。研究者稱其為“FCS 評分架構”，而經由該機制所整合出的結果稱為“模糊複合分數（fuzzy composite score, FCS）”。研究者希望在多元評量教育時代來臨的同時，透過模糊理論的加入，能為教師在整合與評估學生成就時，提供教師一個新的做法和思考空間。

三、研究假設

本研究基於學習者接受評量之表現除了本身真實之學習狀況之外，仍受到許多無法完全掌握之因素影響（例如：諸多之心理和生理之狀態以及評量環境等），致使評量表現蘊含著相當程度的不確定性。再者，教學者在依據學習者之評量表現進行評比時，其思考、推論、以及下決定的運作過程，是以相當之模糊程度進行的。所以，研究者認為：多元評量的評分工作應以模糊理論為邏輯基礎，配合各種量尺之多元評量特質，進而發展出 FCS 之運算模式。同時，由於 FCS 是經由非線性的模糊推論而得，故其推論得到的結果勢必與傳統線性運算所得之結果不盡相同，但不論其結果為何，仍須符合我們一般的常識性判斷（例如：每次均獲得較高評量分數者其 FCS 亦較大），並且對於具某種程度差距之輸入評量分數，能夠具有區辨之能力。

四、研究限制

利用模糊理論所獲得的模糊複合分數，在真實的教學情境中，對於整合多元評量計分上，雖具有較合理的量化詮釋空間，然而對於多元評量在學生學習歷程的質性詮釋上，仍需佐以多元管道評量過程所蒐集的資料（如教學觀察、對談紀錄、歷程檔案等），才能對學生在學習情境中所發生的學習活動，有較為全面性的瞭解與剖析，兼顧學習評量的信度與效度。

貳、文獻探討

一、多元評量

多元評量（multiple assessment），基本上是指應用各種不同評量方式之最有效組合，藉以瞭解學生的學習情形與成就。一般而言，評量者需基於教師的教學目標與學生的學習成果來量身定做評量設計，以往傳統上是以美國教育學家 Bloom 所提出的認知、技能、情意三大領域為依歸，但由於其各領域在評鑑細目過

於繁瑣，因此國外學者 Stiggins 將其化簡為知識、推理與問題解決、技能、產品、情意五大類（Stiggins, 1994）。

在評量方式上，國內學者則認為：挑選反應評量 (selected response assessments)、建構反應評量 (constructed response assessments)、實作評量 (performance assessments)、與人際溝通評量 (interpersonal communication assessments) 是最常見的四種評量方式（王文中等人, 1999）。

這些多元性的評量可以讓評鑑者透過不同的評量內容、方式與時機，取得完整之學習者學習狀況的資訊，並據此對學習進行價值的判斷。

以科學教室中的科學學習評量為例，除了利用傳統之標準化測驗，針對學習單元內容知識，獲得學生答對百分比的紙筆成績外，科學教師亦可利用口頭問答（口試）的方式，去獲得學生有關其學習過程之動機、情意、參與程度、和認知程度等資訊，再依據教師的教學目標（可能是單一或是多元的），給予適當的評分。近年來，國外學者大力倡導在教學課程進行中，評分者可依據相關科學活動的評分規準（scoring rubrics）進行學習歷程評鑑，教師經由觀察學生在教學活動中的表現與預設之評分規準對照比較之下，依其表現之情形，給予相對應之評鑑分數，這些評分規準往往是由教師對學習者的成就期望（performance expections）所建立起的（Champagne & Kouba, 2000）。

此外，實作評量（performance assessments），提供一個真實情境的工作（task），讓學生能以實際操作（hands-on）表現出學習的成效，以建立真實的評量。檔案評量（portfolio assessments），則是透過學習者的學習歷程檔案呈現，讓評分者能夠瞭解學生在相關學習過程中的努力與自我反思和成長。

誠如上述，諸如此類的教學評量方式都是多元評量的有效可能途徑，也是活化學生學習的強而有利之催化劑。

二、模糊理論

模糊（fuzzy）的概念是由美國加州柏克萊大學教授 Lofti Zadeh 於 1965 年所提出的，其主要目的在於試圖以模糊集合的數學方法作為語意上不精確性（vagueness）的表徵方式，以便解決蒸氣渦輪引擎的控制問題（Zadeh, 1965）。事實上，十九世紀以來，在物理與數學兩大領域中逐漸興起所謂“不確定性（uncertainty）”的觀點，這使得精確的科學量化工具，有重新思考的必要性，而模糊理論則是電腦科技在處理此等模糊性與不確定性的最佳利器（Klir & Yuan, 1995）。

換個角度來看，人類在真實生活中的語意描述、思考運作、決策等行為，基本上皆是以模糊（vague）、不精準（imprecise）不確定（uncertain）的方式進行訊息處理的工作，而非絕對明確(crisp) 的。例如，在教室情境中，師生會以“作業好多”、“題目好難”、“表現很好”、“秩序很差”等，諸如此類的語詞進行對話，這些語義上的描述往往並無明確的界定，而師生卻是以這些充滿模糊與不確定性的語義傳達，去推論、詮釋、建構真實的世界。因此，人類以數學量化方式詮釋真實世界時，適度地加入模糊性之考量，才能切合真實的情境。這就是模糊理論之所以在近年來，廣泛地被運用至各領域中的原因。

模糊邏輯（fuzzy logic, FL）是以模糊集合（fuzzy subset）為基礎，取代傳統雙值（two-valued）函數（classical sets; crisp sets），所進行之邏輯推論。其最大優點在於其多值（multi-valued）函數的特色，能改善僅以「真」（true）、「偽」（false）兩個絕對極端值，進行相關邏輯推論時所產生與真實情境不盡相符的窘境。尤其是涉及到人類的思考、互動、判斷，或是

複雜、多變的環境系統，都是以模糊邏輯解決問題之最佳時機（"*What is Fazzy Logic*?", 1997; Dounias, *et al.*, 1998; Zimmermann, Zadah, & Gaines, 1984; Zopounidis & Documpos, 1998）。

基本上，使用模糊邏輯來處理有關人類思考推理與決策的問題時，就是以電腦模擬人類的推理歷程。在作法上，是以函數的形式來定義相關變數與其所詮釋之語意世界的相符程度（degree of membership function）。這個多值函數稱為歸屬函數（membership function）；而這個變數則稱為語意變數（linguistic variable）。在邏輯推論的執行上，則是由資料庫（data base）、知識庫（knowledge base）、以及推論引擎（inference engine）所共同完成。資料庫是用以儲存相關工作資料的集合，相當於人腦中之短期記憶體（short-term memory）；知識庫是描述經驗領域的資訊所形成的集合，相當於人腦中之長期記憶體（long-term memory），包含了推論法則（通常是一些"IF-THEN"的條件敘述句）和實例資料庫；推論引擎則是執行奠基於提示法則（production rules）系統的近似推理（approximate reasoning）運作。由上述組成所建立起的電腦運作機制，即所謂的專家系統（expert system）。

如同傳統演繹推論一般，模糊近似推理是根據一個普遍性程度較高的命題（proposition），推導出一個普遍性程度較低的命題。在自然語意中，全稱性質命題和假設命題為主要的兩種命題形式，前者是用以斷定一類事物全有或全無具備某種屬性的推論過程，常被稱為"三段論"；後者乃是根據事物間的條件關係，所進行之演繹過程，又稱為"假設推理"。不論其性質為何，其基本過程均是由一組前提進而遞推出其結果之過程。常見的模糊近似推理模組有四：一、明確推理（categorical reasoning），這種類型的推理在於前提部分不具備模糊機率和模糊數量詞（Zadeh, 1992）；二、質性推理（qualitative reasoning），此一類型的推理其前提和結論均涉及語意或模糊變數，並且輸入與輸出的關係是建立於一系列的模糊規則組合（Zadeh, 1975）；三、三段論式推理（syllogistic reasoning），相對於明確推理，這一類型推理式中的前提具備了模糊數量詞（Zadeh, 1992; Dubois & Prade, 1988）；四、意向推理（dispositional reasoning），這一類型的推理，其前提意涵著某種程度的「經常性」（usually），此一類型的推理常常被視為建立電腦進行常識推理架構的理論之一（Zadeh, 1992）。

以本研究所進行之性質推理為例，其推論過程可以用圖示表述如下：

(一)單一法則推理過程：若 x 為 A_1 且 y 為 B_1 時，則結果 z 是 C_1，命題型式如下：

"IF x is A_1 AND y is B_1, THEN z is C_1"

其中 x、y 為輸入資料，A_1、B_1、與 C_1 分別為語意變數可由歸屬函數定義之，而 z 則為輸出結果。

一般而言，在模糊近似推理最常使用的推論法則為 Mamdani 最小模糊關聯規則(Mamdani's minimum fuzzy implication rule)，其推論過程可圖示如圖 1。

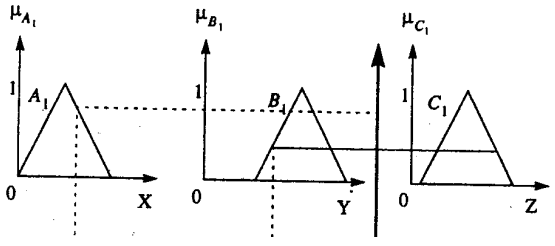

圖 1：Mamdani 最小模糊關聯規則之圖示

(二)兩個以上（含兩個）法則推理過程：以兩個法則推理為例，推論規則命題及結論如下：

IF x is A_1 AND y is B_1, THEN z is C_1
IF x is A_2 AND y is B_2, THEN z is C_2

結論：　z 為 c'

其中，x 與 y 為語意變數，其推論過程可用圖 2 表示之：

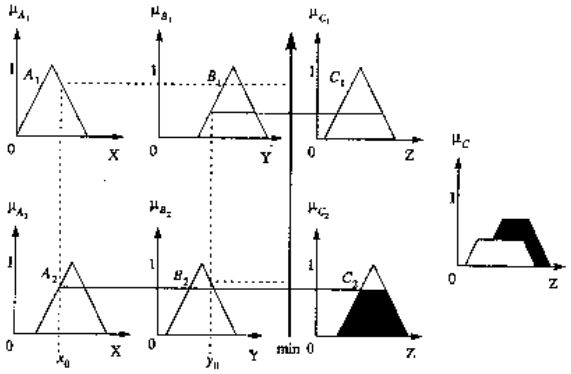

圖 2：Mamdani 近似推理之圖示

Note. From Neural fuzzy system: A Neuro-fuzzy synergism to intelligent systems (p.155), by Lin & Lee, 1996, New York: Prentice-Hall International, Inc.

三、模糊複合分數

鑑於現行傳統計分處理方式未能合理反映教育評量先天上所具備的模糊與不確定性特質，再加上未來教育實施多元評量的制度之下，對於不同量尺所得到的評量分數，在整合上勢必無法均以傳統百分數之平均值算法來計算。於是乎，一套能兼顧評量之特質與多樣性，可供多元分數整合的運作機制，更顯現其重要性和迫切性。

由於模糊邏輯推論具備語意上某種程度的不確定性，而教學評量過程含有相當成分的模糊性，故將其運用至教學評量的實際狀況中，將比傳統以絕對性的分數量度學生學習的成就，來得更適切且更具彈性（孫光天, 1998）。

於是，研究者基於未來教育評量上的需求，嘗試以模糊理論為基礎，發展出整合不同量尺所得到評量分數的機制，並稱經由這種機制所推論出之評量分數為「模糊複合分數」（fuzzy composite score，以下簡稱 FCS）。

建構 FCS 評分架構之概念是源自於模糊邏輯控制器（fuzzy logic controller，FLC）的架構。研究者根據研究之需求，將 FLC 改編而成 FCS 的評分架構如（圖 3）所示。其功能是要將數個多元評量分數，經由模糊推論的過程，整合出一個學習成效的參考值，以供教育評估之用。

FCS 的運算流程主要可以分為三大部分如下：

(一)**模糊化（fuzzification in）：**

將明確的輸入數值 x（crisp input）依據資料庫中相對應語意變數之歸屬函數，轉換為模糊語意值 $\mu_{(x)}$（linguistic values）。

(二)**推論（inference）：**

依使用者所自行設定之 IF-THEN 模糊推論法則（fuzzy inference rule，FIR）、Mamdani 的最小模糊關聯規則（Mamdani' s minimum fuzzy implication ruler）、以及相關之模糊理論運算關係，進行推論工作（過程如前述圖 1、2 所示）。

(三)**解模糊化（defuzzification）：**

為了將推論引擎所推論出來的模糊演繹結果（fuzzy sets），轉換成明確的輸出值（crisp sets），第三個部份的工作是以中心法（center of area method, COA）的方法，對推論引擎所得的結果的進行下列運算：

$$z^*_{COA} = \frac{\int_z \mu_C(z)\, z\, dz}{\int_z \mu_C(z)\, dz}$$

最後，經由上式運算所獲致的中心點（centroid point），此即為本研究的輸出：FCS。

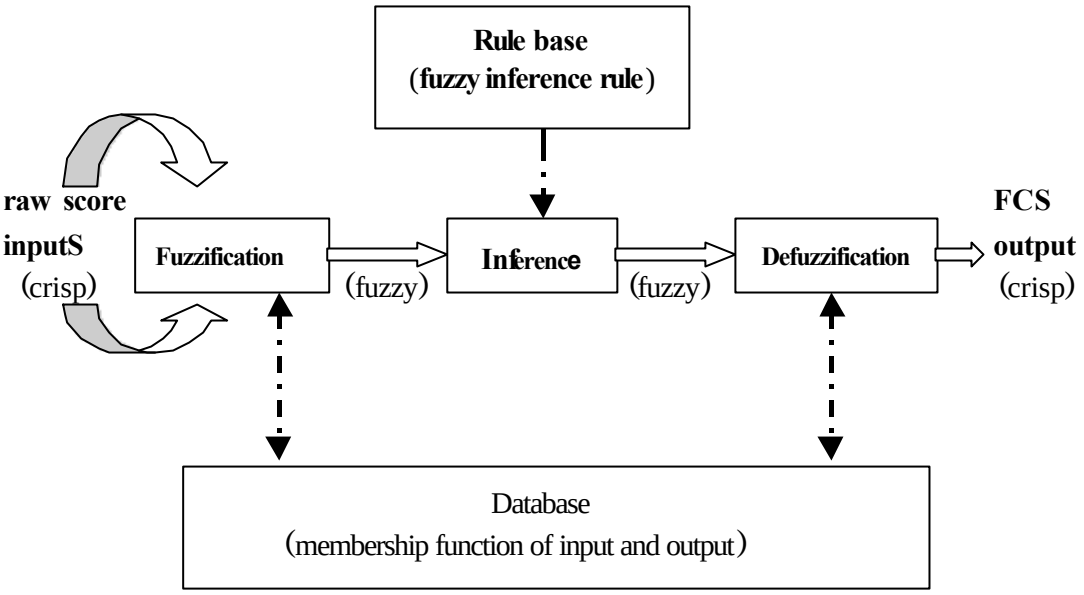

圖 3：FCS 評分架構

叁、研究設計

一、研究樣本

本研究的樣本以模擬方式，建構一個為數三十人的班級成績，全班學生在各項評量之得分情形如表 1 所示。其中包含了五項多元評量成績：

(一)紙筆測驗：教師依據教學需要，所實施之隨堂紙筆測驗，採百分數量尺，依據學生答對百分比計分。

(二)口頭發問：課堂中，教師依據學生對口頭發問回答情形給予 1　10 的級分。

(三)實作評量：實驗室中，學生針對實驗目的所進行之實際操作，教師則以預先所制定的實驗操作評分歸準，給予學生 1　20 的級分。

(四)檔案評量：教師將學生所製作之學習檔案，分成五個等第分別給予 1~5 的級分。

(五)期中標準化測驗：全校性之定期考察，採百分數量尺，依據學生答對百分比計分。

在設計樣本模擬時，研究者希望能盡可能符合真實的教學評量情境，並且涵蓋較廣泛的成績可能分布情形，以便研究結果在未來延伸至一般的教學情境可以具備較高之信效度。因此，在模擬樣本設計上，研究者之考量有二：

(一)需具備各項評量表現一致的樣本：例如 5 號學生表現皆優秀，而 16 號學生表現皆不佳。

(二)需具備在各項評量的表現各有優劣之樣本：在多元評量制度中，學生們在不同情境以及不同的評量方式之下，極有可能在各項評量中表現會互有優劣。這個部分之模擬得分，在設計是採隨機方式產生。

二、FCS 相關參數設計

(一)歸屬函數：將多元評量明確之學習評

表 1：本研究模擬之學生得分表

座號	評量一 (1-100)	評量二 (1-10)	評量三 (1-20)	評量四 (1-5)	評量五 (1-100)
1	75	8	12	3	80
2	82	2	10	4	72
3	47	5	8	1	60
4	68	3	14	3	52
5	92	9	18	5	95
6	74	2	16	5	70
7	78	4	8	2	68
8	84	9	4	4	62
9	53	8	12	3	74
10	80	3	14	5	60
11	45	4	5	2	42
12	65	2	10	2	50
13	75	3	7	4	70
14	90	6	15	3	40
15	83	5	11	3	80
16	40	2	3	1	40
17	58	6	8	4	75
18	74	2	15	3	90
19	62	3	5	4	65
20	70	4	12	2	54
21	80	2	4	5	82
22	54	4	8	4	73
23	69	3	7	2	65
24	80	1	18	3	70
25	72	6	7	3	55
26	66	4	9	4	60
27	70	6	10	3	85
28	50	2	12	4	60
29	45	5	15	2	73
30	88	8	7	5	82

量分數，轉換成合乎我們一般口語化的模糊語意值時，可由歸屬函數的定義而得到其對應關係。在實際轉換時，用來定義歸屬函數

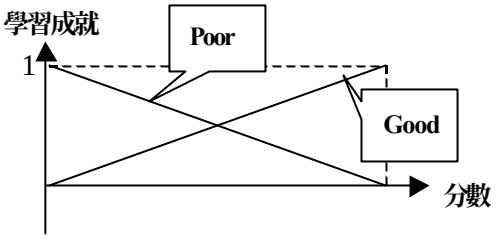

圖 4：學習成就之歸屬函數

的函數形式相當多，如：三角形函數（triangular functions）、鐘形函數（bell-shaped functions）、梯形函數（trapezoidal functions）等。在一般的應用上，以三角形模糊函數來定義歸屬函數，具有其簡單性及快速性（黃士滔和劉天賜, 1997）。同時，由於教師在評斷學生學習之優劣時，基本上是和學習者在評量中表現優劣成正比的。基於上述之考量，故本研究採用三角形模糊函數為研究所需之歸屬函數型式，藉以定義輸入及輸出。

以歸屬函數進行模糊化的過程中，本研究採最簡單的兩區段模糊分割（fuzzy partition with two terms）之模糊分割。也就是說，我們將學生的學習成就分成好、壞兩種語義變數來對其所得分數加以模糊推論，如圖 4 所示。

圖 4 中，歸屬函數“Good”表示“我們在語意上說一個學生學習情況良好的程度，是與其學習成就之得分成正比”；相反的，歸屬函數“Poor”則表示“我們在語意上說一個學生學習情況不良的程度，是隨著其學習成就得分越低而越高”。

利用上述語義變數的歸屬函數“Good”和“Poor”，配合研究模擬樣本的五項評量分數之量尺，即可建構出本研究 FCS 評分架構中 Database 所對應的五個評量輸入歸屬函數及其圖形，分別如圖 5、圖 6、圖 7、圖 8、和圖 9 所示。

$$\begin{cases} \mu_{good}(x) = 0.01X \\ \mu_{poor}(x) = 1-0.01X \end{cases} \quad \text{for } 0 \leqq X \leqq 100$$

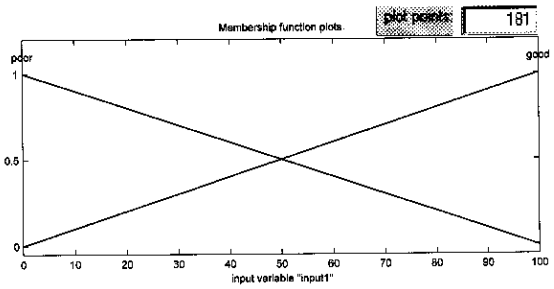

圖 5：評量一得分之歸屬函數

$$\begin{cases} \mu_{good}(x) = 0.1X \\ \mu_{poor}(x) = 1-0.1X \end{cases} \quad \text{for } 0 \leqq X \leqq 10$$

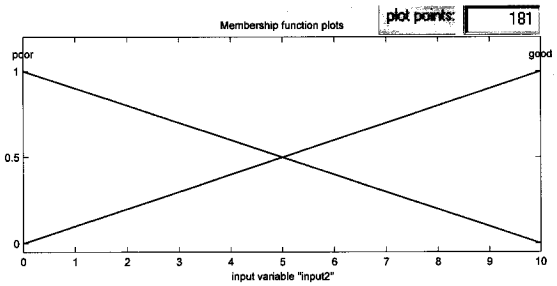

圖 6：評量二得分之歸屬函數

$$\begin{cases} \mu_{good}(x) = 0.05X \\ \mu_{poor}(x) = 1-0.05X \end{cases} \quad \text{for } 0 \leqq X \leqq 20$$

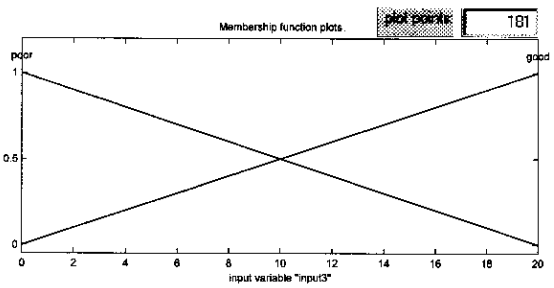

圖 7：評量三得分之歸屬函數

$$\begin{cases} \mu_{good}(x) = 0.2X \\ \mu_{poor}(x) = 1-0.2X \end{cases} \quad \text{for } 0 \leqq X \leqq 5$$

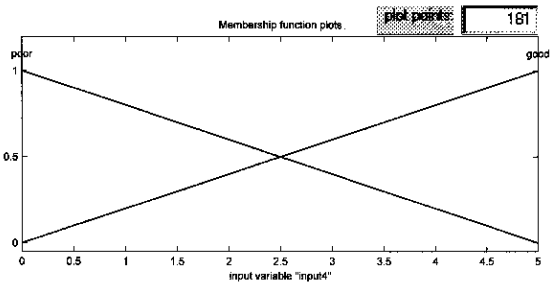

圖 8：評量四得分之歸屬函數

$$\begin{cases} \mu_{good}(x) = 0.01X \\ \mu_{poor}(x) = 1-0.01X \end{cases} \quad \text{for } 0 \leqq X \leqq 100$$

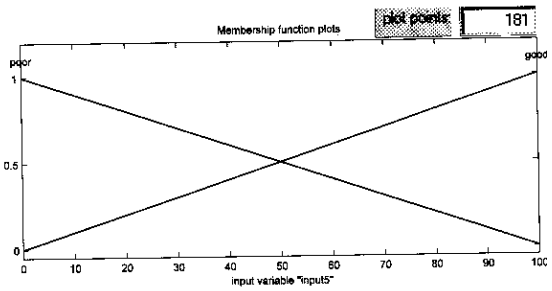

圖 9：評量五得分之歸屬函數

由於本研究希望將五項多元評量成績整合成一個學習成就參考值，也就是說：處理過程模糊化，結果明確化。故研究參數之設定仍需考量到 Database 中，解模糊化的輸出歸屬函數。研究者擬將 FCS 評分架構的輸出值設計為介於零到五之間的數值，而我們即可將其轉換成五等第級數，進而對其學習進行評比，並決定學生是否通過學習之評量。

為了達到上述輸出之設計理念與目的，研究者將其輸出值範圍設定於 [-5 10]，如圖 10 所示。為何範圍設定不是 [0 5]？蓋本研究 FCS 採用 COA 的方法進行解模糊化運算，所以若

$$\begin{cases} \mu_{good}(x) = (1/15)(X+5) \\ \qquad\qquad \text{for } -5 \leqq X \leqq 10 \\ \mu_{poor}(x) = 1-(1/15)(X+5) \end{cases}$$

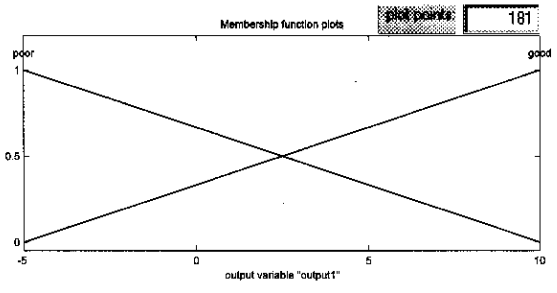

圖 10：輸出值之歸屬函數

是將輸出值範圍設定在 [0 5]，則輸出值將只會落在 [1.65 3.35]的區域中。若將其輸出範圍設定於 [-5 10]，則其輸出值將會大致落於 [0 5]（實際之準確值為 [-0.05 5.05]）。

(二)模糊推論法則：由於本研究具有兩個語意變數以及五個輸入，故所有可能的模糊推論法則會有 $2^5 = 32$ 條，如下表 2 所述。

以 Rule 1 為例，其推理邏輯前提條件之語意為：

“若學習者在評量一的表現良好、在評量二的表現良好、在評量三的表現良好、在評量四的表現良好、在評量五的表現良好時”。

至於輸出結果的訂定，則視其前提條件中“Good”與“Poor”的語意變數出現多寡而決定其輸出為“Good”或“Poor”。例如：表 2 中的 Rule 7 前提條件為“若學習者在評量一的表現良好、在評量二的表現良好、在評量三的表現不良、在評量四的表現不良、在評量五的表現良好時”，語意變數“Good”與“Poor”出現次數分別為三次與兩次，故其輸出選定為“Good”；又如 Rule 7 前提條件為“若學習者在評量一的表現良好、在評量二的表現良好、在評量三的表現不良、在評量四的表現不良、在評量五的表現不良時”，語意變數

表 2：兩個語意變數以及五個輸入之條件下，所有模糊推論法則

	Input 1	Input 2	Input 3	Input 4	Input 5	Output
Rule 1	Good	Good	Good	Good	Good	Good
Rule 2	Good	Good	Good	Good	Poor	Good
Rule 3	Good	Good	Good	Poor	Good	Good
Rule 4	Good	Good	Good	Poor	Poor	Good
Rule 5	Good	Good	Poor	Good	Good	Good
Rule 6	Good	Good	Poor	Good	Poor	Good
Rule 7	Good	Good	Poor	Poor	Good	Good
Rule 8	Good	Good	Poor	Poor	Poor	Poor
Rule 9	Good	Poor	Good	Good	Good	Good
Rule 10	Good	Poor	Good	Good	Poor	Good
Rule 11	Good	Poor	Good	Poor	Good	Good
Rule 12	Good	Poor	Good	Poor	Poor	Poor
Rule 13	Good	Poor	Poor	Good	Good	Good
Rule 14	Good	Poor	Poor	Good	Poor	Poor
Rule 15	Good	Poor	Poor	Poor	Good	Poor
Rule 16	Good	Poor	Poor	Poor	Poor	Poor
Rule 17	Poor	Good	Good	Good	Good	Good
Rule 18	Poor	Good	Good	Good	Poor	Good
Rule 19	Poor	Good	Good	Poor	Good	Good
Rule 20	Poor	Good	Good	Poor	Poor	Poor
Rule 21	Poor	Good	Poor	Good	Good	Good
Rule 22	Poor	Good	Poor	Good	Poor	Poor
Rule 23	Poor	Good	Poor	Poor	Good	Poor
Rule 24	Poor	Good	Poor	Poor	Poor	Poor
Rule 25	Poor	Poor	Good	Good	Good	Good
Rule 26	Poor	Poor	Good	Good	Poor	Poor
Rule 27	Poor	Poor	Good	Poor	Good	Poor
Rule 28	Poor	Poor	Good	Poor	Poor	Poor
Rule 29	Poor	Poor	Poor	Good	Poor	Poor
Rule 30	Poor	Poor	Poor	Good	Poor	Poor
Rule 31	Poor	Poor	Poor	Poor	Good	Poor
Rule 32	Poor	Poor	Poor	Poor	Poor	Poor

“Good”與“Poor”出現次數分別為兩次與三次，故其輸出選定為“Poor”。

依據上述推論法則制定之理念，表 2 中 Rule 1 之邏輯推理完整語意為：

當評量一的表現良好、評量二的表現良好、評量三的表現良好、評量四的表現良好、評量五的表現良好時、我們認為其學習表現整體而言是良好的。

三、FCS 運算程式撰寫與執行

本研究所提出之 FCS 評分架構是使用美國 Math Works 公司所推出的 MATLAB 軟體，完成其程式撰寫及執行的工作。MATLAB 基本上是以數學矩陣（matrix）的運算方式為主，同時為了因應不同的應用並方便使用者操作，業已發展出四十餘種相關之應用程式，即為 MATLAB 工具箱 (MATLAB Toolboxes)，為使用者提供了方便、快速的操作模組。

本研究便是以其中的模糊邏輯工具箱（Fuzzy Logic Toolboxes）為主要的運算工具，藉由研究者所撰寫的主程式，讀取研究模擬樣本數據，形成樣本得分矩陣，再以呼叫子程式的方式，將樣本數據輸入已完成相關參數設定之模糊推論系統（fuzzy inference system, FIS）程式，即可獲致本研究所要的 FCS。

以下是本研究於 FCS 評分程式設計時，主要的工作項目：

(一)依據研究目的，設定模糊推論系統（FIS）中，符合語義變數之歸屬函數，以及所需之模糊推論法則。

(二)建立原始分數資料檔案（將模擬樣本分數轉換為文字檔（data.txt））。

(三)撰寫 FCS 評分主程式，其功能在於讀取 data.txt 之輸入資料，並呼叫執行副程式 (FCS.FIS)，執行模糊推論。

(四)除錯（Debug）。

(五)解釋輸出的結果。

肆、結果與分析

依據前述之相關設定與步驟，研究者以模擬樣本分數輸入，可得到結果如表 3 所示。表 3 中的學習評比和建議是依據前述研究者之研究設計及目的，將五項評量成績整合為一個學習總成績，再依據所得到的 FCS 和表 4 之關係，轉換成五等第制，並給予一個簡短之描述詞以及該名學生是否通過學習評量的建議。

表 3：模擬評鑑結果（五等第）

座號	評量一	評量二	評量三	評量四	評量五	FCS	學習評比	建議
1	75	8	12	3	80	3.88	佳	及格
2	82	2	10	4	72	3.44	佳	及格
3	47	5	8	1	60	2.38	差	不及格
4	68	3	14	3	52	2.98	差	不及格
5	92	9	18	5	95	4.93	優	及格
6	74	2	16	5	70	4.09	優	及格
7	78	4	8	2	68	1.75	差	不及格
8	84	9	4	4	62	4.12	優	及格
9	53	8	12	3	74	3.02	佳	及格
10	80	3	14	5	60	3.67	佳	及格
11	45	4	5	2	42	1.91	差	不及格
12	65	2	10	2	50	2.50	可	及格
13	75	3	7	4	70	3.81	佳	及格
14	90	6	15	3	40	3.25	佳	及格
15	83	5	11	3	80	2.91	可	及格
16	40	2	3	1	40	0.93	劣	不及格
17	58	6	8	4	75	3.19	佳	及格
18	74	2	15	3	90	3.84	佳	及格
19	62	3	5	4	65	3.39	佳	及格
20	70	4	12	2	54	2.80	可	及格
21	80	2	4	5	82	4.51	優	及格
22	54	4	8	4	73	2.80	可	及格
23	69	3	7	2	65	1.75	差	不及格
24	80	1	18	3	70	3.67	佳	及格
25	72	6	7	3	55	3.09	佳	及格
26	66	4	9	4	60	3.09	佳	及格
27	70	6	10	3	85	2.91	可	及格
28	50	2	12	4	60	2.91	可	及格
29	45	5	15	2	73	2.50	可	及格
30	88	8	7	5	82	4.26	優	及格

註：若有補考制度時，則可建議“差”等級的學生參加補考。

表 4：模擬評鑑之標準

FCS	評比	建議
4~5	優	及格
3~4	佳	及格
2~3	可	及格
1~2	差	補考或重修
0~1	劣	重修

從表 3 中，研究者發現下列幾點值得注意的地方：

(一)依照研究模擬樣本之設計理念，5 號學生的各項評量成績得分均為最高，結果以模糊推論整合後之 FCS（4.93）亦為全班之冠；16 號學生的各項評量成績得分均為最低，結果以模糊推論整合後之 FCS（0.93）則在全部學生中敬陪末座。這個結果完全符合我們之常理推判。

(二)在結果之中，有相同 FCS 出現，例如 15 號、27 號、28 號三位之 FCS 均為 2.91；7 號 23 號之 FCS 均為 1.75；12 號、29 號之 FCS 均為 2.50；20 號、22 號之均為 2.80；以及 25 號、26 號 FCS 均為 3.09。仔細分別比較各組 FCS 同分的學習者，在各項評量中得分情形，即可發現 FCS 相同的學生在不同的學習評量得分上各有優劣，所以在經過最後的整合推論，認定其學習成就是約略相當的。

以 15 號、27 號、28 號之 FCS 均為 2.91 的情形為例來看，三位的各項原始評量得分表列如下表 5 所示，我們可以看到：在紙筆測驗、口頭發問以及期中標準化測驗部分，28 號表現雖不如 15 和 27 號，然而在實作評量和檔案評量上則優於其他兩者，以致於可以彌補他在紙筆測驗方面之不足；而 15 及 27 號兩位學生除了在檔案評量上得分相同之外，其餘各項得分上則呈現得分相近且優劣參半的局勢，而造成最後的 FCS 相同。

(三)本研究之模擬樣本設計，除了考量設計在各評量表現一致的樣本之外，亦以隨機方式產生在各評量上表現不一致的樣本，結果分析顯示，本研究所發展出的 FCS 評分架構對於學生在不同評量象度上的表現，具有統整分析的功能。

表 5：對相同 FCS 之原始評量比較

座號	紙筆測驗	口頭發問	實作評量	檔案評量	期中標準化測驗	FCS
15	83	5	11	3	80	2.91
27	70	6	10	3	85	2.91
28	50	2	12	4	60	2.91

表 6：模擬評鑑結果（百分數）

座號	五分制 FCS	百分數 FCS	座號	五分制 FCS	百分數 FCS
1	3.88	77	16	0.93	19.3
2	3.44	68.5	17	3.19	63.4
3	2.38	47.7	18	3.84	76.2
4	2.98	59.5	19	3.39	67.5
5	4.93	97.7	20	2.80	55.9
6	4.09	81.3	21	4.51	89.3
7	1.75	35.3	22	2.80	55.9
8	4.12	81.7	23	1.75	35.3
9	3.02	60.2	24	3.67	73.0
10	3.67	73	25	3.09	61.5
11	1.91	38.5	26	3.09	61.5
12	2.50	50	27	2.91	58.1
13	3.81	75.5	28	2.91	58.1
14	3.25	64.7	29	2.50	50.0
15	2.91	58.1	30	4.26	84.5

(四)若是希望輸出之 FCS 為百分數，理論上應該僅需將輸出之歸屬函數的範圍改定為 [-97 197]，則 FCS 之輸出將落於 [0 100]（實際為 [0.02 100]）的區間內。研究者亦實際進

行上述歸屬函數範圍之更改，重新執行程式一次，所得結果與表 3 作一比較如表 6。

從表 6 的結果顯示：五分制得分高低與百分數得分高低結果是完全一致的，也就是說，五分制得分高者其百分數也高，五分制得分相同者百分數得分亦相同。這個結果可以說明兩個重要的觀點，第一點是說明本研究所提出之 FCS 評分架構在輸出上，可以依據使用者實際需求，僅改變輸出之歸屬函數的範圍，即可迅速地獲致所需之 FCS；第二點則說明利用模糊推論進行多元評量之評分工作，是過程模糊化，而非結果“模糊化”，所以只要 FCS 評分系統中的相關參數一但設定後，對每一筆輸入資料處理時的模糊化程度就會採取一樣的標準，故不會產生因為歸屬函數範圍的改變，而造成 FCS 是評比時的不一致情形。

伍、結論與建議

一、結論

(一)以模糊推論過程來進行多元評量分數之整合是基於多元評量教育時代的來臨，而多元評量之所以異於過去傳統教學評量之最大不同點，便是教學評量的工作可以依照教學的目標和實際內容，選擇在學生學習過程中的任何一個時間點（甚至是全程的），以任何一種可行、有效之評量方式，在不同的象度上以不同之量尺去檢視學生的學習成效，以期得到學生最完整和詳盡之相關學習資訊（包含了動機、歷程、價值觀、結果乃至於學習之動態成長）。而多元評量分數之整合勢必成為未來的一個重要議題。更重要的是：如何整合才能符合真實教育情境中，教師對於學生學習行為之判斷與推論過程。人工智慧中的模糊理論，便提供我們一個模擬教師對學生學習成就推論的較佳工具。

(二)透過研究所設計之 FCS 評分架構以模擬樣本為輸入，可以獲致下列之結論：

1. 本研究發展之 FCS 評分架構所進行之非線性的模糊推論而得到的 FCS 符合我們一般性的常識判斷，並且對於輸入之評量分數具有顯著差異者，能夠區辨其間之優劣，同時並顯示出其異於傳統以線性、精準、明確方式處理之結果，這與研究者的研究假設相互吻合。
2. 對於多元評量所獲得之不同量尺得分，使用者只需設定各量尺的範圍（如：1~100、1~5…等），FCS 評分機制即可以對輸入值直接進行整合工作。
3. 在 FCS 評分過程中，可以找到在以往的研究文獻中，所提及有關模糊理論在應用時的優、缺點，例如：推論快速、合乎人類實際語意及思考的方式…為其優點；需要專家設定相關之語意變數、歸屬函數以及推論法則，繁瑣且不易之特性…則為其缺點。
4. 對於相同之原始數據，若考慮到不同 FCS 輸出型式，使用者僅需依實際需求修改輸出歸屬函數的範圍即可。

(三)對於重視學習歷程與相關技能學習之學科，尤其是自然科學，多元評量提供一個教育者重新思考及調整教學的機會，而本研究則提供一個多元評量分數整合與評比的思考方向。

(四)透過本試探性研究，可以初步獲致以模糊理論整合多元評量分數之適切性，並可供未來相關發展多元評量分數整合之研究基礎。

二、建議

經由本研究之相關模擬與討論，可瞭解到如何以模糊理論整合多元評量分數及其相關之理論背景，未來後續的研究方向，研究者有以下幾點建議：

(一)教師專家資料庫之建立：未來台灣全

面實施九年一貫教育制度後，多元評量將扮演一個更重要的角色，除了教學教材、評量之獨特性之外，仍需考量到各年級、各學校在橫向與縱向的成績認定問題，若是能在發展 FCS 評分架構時便一併考慮到建置一個較能普遍符合相關教師的專家資料庫，則相信此 FCS 評分更能為一般大眾所接受。

(二)專業 FCS 評分軟體之設計：由本研究可得知：一般教師以模糊推論進行多元評量分數整合之最大困難在於相關之電腦程式設計與執行以及其參數設定。所以，若要 FCS 評分方式能落實至實際教學評量中，一個具有簡化的操作介面之專業FCS 評分軟體勢必被提出。研究者建議後續程式研究開發人員可以直接利用 MATLAB 軟體中，握把式圖形（handle graphics）之概念，設計一個和善（Friendly）、直覺（intuitive）、簡單（simple）的圖形使用者介面（graphic user interface, GUI），讓使用教師可以簡單地以滑鼠便能操作 FCS 評分系統。

(三)建置具有自我學習（self-learning）功能之 FCS 評分架構：一個具有公認性的專家資料庫固然重要，然而在一些實際應用上，個別教學之設計和評分者之信念因素仍需加以考量。所以，若是能夠在 FCS 評分機制中加入自我學習功能（例如結合模糊邏輯（fuzzy logic）與類神經網路（artificial neural network）的模糊類神經整合系統（fuzzy neural integrated systems)），讓系統能夠依據使用者所輸入之訓練值（training data），而自行演繹出運算所需的定義歸屬函數和模糊推論法則，如此可以使 FCS 在實際教學評量上的實用性與普遍性更向前邁進一步。

參考文獻

1. 王文中、呂今燮、吳毓瑩、張郁雯和張淑慧（1999）。教育測驗與評量。台北：五南出版社。
2. 郭生玉（1985）。心理與教育測驗。台北：精華書局。
3. 孫光天（1998）。模糊理論應運於學習評量之研究。高雄：第七屆國際電腦輔助教學研討會大會論文, 358-363。
4. 黃士滔和劉天賜（1997）。模糊理論在 ISO14001 環境考量面重要程度判定上之應用。中華民國品質發展學會八十六年會, 42-47。
5. 張智星（2000）。MATLAB 程式設計與應用。新竹：清蔚科技出版事業部。
6. 蒙以正（1998）。專業設計技巧MATLA5。台北：碁峰資訊。
7. Champagne, A. B., & Kouba, V. L. (2000). Writing to inquire: Written products as performance measures. In J. Mintzes & J. Wandersee & J. Novak (Eds), *Assessing Science Understanding : A human constructivist view* (pp. 224-250). San Diego : Academic Press.
8. Dounias, G., Havellas, S., &Tselentis, G. (1998). A fuzzy rule based approach for decision making in commercial food chains. *EUPIT' 98* Aachen, Germany, September 7-10.
9. Dubois, D., & Prade, H. (1988). On fuzzy syllogisms. *Computational Intelligence, 4*, 171-179.
10. Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
11. Klir, G J., & Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Application.* New Jersey : Prentice-Hall Inc.
12. Lin, C. T., & Lee, C. S. (1996). *Neural fuzzy system - A Neuro - fuzzy synergism to intelligent systems.* New York : Prentice-Hall International, Inc.

13. Battelle Memorial Institute (1997). *Fuzzy logic: What is fuzzy logic*? [Online]. Available: http:// emsl.npl. gov. 2080 / proj / neuron / fuzzy / what. html (March 10, 1999).
14. Mintzes, J. J., & Wandersee, J. H., & Novak, J. D. (2000). *Assessing Science Understanding : A human constructrvist view.* San Diego : Academic Press.
15. "*What is Fuzzy Logic*?" (1997) [www page] URL: http://emsl.npl.gov.2080/proj/neuron/ fuzzy/what.html (visited 1999/3/10).
16. Stiggins, R. J. (1994). *Student-centered classroom assessment.* New York : Macmillan College Publishing Company.
17. Zadeh, L. A. (1965). Fuzzy set *Information and Control, 8,* 338-353.
18. Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning. *Information Science, 8*, 199-249.
19. Zadeh, L. A. (1992). Knowledge repre- sentation in fuzzy logic. In R. R. Yager & L. A. Zadeh (Eds), *An Introduction to Fuzzy Logic Applications in Intelligent Systems.* Boston : Kluwer Academic.
20. Zimmermann, H. J., Zadeh, L. A., & Gaines, B. R. (1984). *Fuzzy Sets and Decision Analysis.* Netherland: Eiaevier Science Publishers.
21. Zopounidis, C., & Documpos, M. (1998). Developing a mulyicriteria decision support system for financial classification problems: The finclas system. *Optimization Methods and Software, 8*, 277-304.

A Preliminary Study of Development of Fuzzy Composite Score for Multiple Assessments

Hao-Chang, Lo
Chung Sun Medical University
Center for Teacher Education

Abstract

Multiple assessments have taken the place of single standardized test recently. Tools, forms, and standards are multiple in the process of assessments. How to integrate scores of multiple assessments into a composite score will be an important subject in educational research. The aim of this research was to propose an operating mechanism based on fuzzy theory to integrate scores of multiple assessments. The character of fuzzy theory was to process information vaguely, imprecisely, and uncertainly. The composite score was called "fuzzy composite score, FCS" in this research. A simulated sample was used to test this fuzzy scoring frame. As the result shown, the operating mechanism agreed to the hypotheses and purpose of this research. It was convenient to integrate scores in different scales, too. Finally, some suggestions for future researches were proposed including how to design the professional software and build knowledge base of expert systems. It's believed that works based on this research will make FCS more close to the educational practice in the future.

Key word: Multiple Assessments, Fuzzy, Fuzzy Logic, Fuzzy Composite Score.