

模糊統計在試題難度上的應用*

謝名娟

國家教育研究院
測量與評量研究中心

吳柏林

國立政治大學
應用數學系

試題難度評量一直是許多人研究的課題。但傳統方法的五點量表問卷只提供固定尺度的選擇，似乎無法完整地表達受測者真實且複雜的思考。因此本研究將以模糊統計的角度進行試題難度的探討。過去許多研究應用模糊平均數、模糊眾數或模糊中位數等概念於試題難度評量。而本文將以此為基礎，定義一種新的距離，再透過一些轉換取得試題的難度指標，進而比較各試題之間難度的差異。本研究的另一個重點，是各個不同難度因子的向度來決定各試題的難度。再以模糊相對權重的概念，對各向度的難度指標作加權，進而比較、分析。在此研究中，模糊難度的數值，和 IRT 的難度 b 值與試題通過率 P 值之相關達到一定水準。代表此方法所得出的結果，已有相當程度的驗證。

關鍵詞：試題難度評量、模糊統計、模糊相對權重

在中國自隋唐開科取士以來，歷經宋、元、明、清諸朝不斷的改進，此一科舉取才的考試制度，雖然源自中國，但中國卻一直沒有針對「考試」這門學問進行比較科學化的量化分析。在綿延數千年後，世界各國爭相採用，以作為文官制度的選拔依據，致使近代的心理計量學（psychometrics）的發展卻起源於外國，西風東漸後，才傳入中國（余民寧，2002）。雖然古今中外對於人才的進用大多數是採用考試的方式產生，而施測單位如欲拔擢適合需求的人才，試題之設計即更顯重要，因為一份良好的試卷，可讓符合需求且真正有能力者出現，一般大小考試試卷的型態不乏選擇題、是非題、填充題、申論題、計算題等等題型，因此試題分析其優劣適當是有其重要性的。

但試題難度的評量尚無一致的標準，每當學測或基測過後，總是在新聞或報紙上看到補習班老師在對今年的試題難度作評論，但這樣的評論的依據為何？靠的是補教老師個人的經驗？或有其他任何科學化的標準？如果說一份題目很難，是難在什麼地方？是難在計算過程很繁瑣，還是難在題目敘述太長太複雜？是想不出解題的策略，或是因為創新的命題方式讓題目變得很難？了解學生覺得困難的部分，是身為一個教師亟欲探究的部分，一旦了解這點，不但對教師的教學有幫助，對學生的診斷性學習更有正確的切入點。

* 本文通訊作者：謝名娟；通訊方式：hm7523@hotmail.com。

大多數的評量中，試題的難度採用答對百分比法來表示難度。答對百分比法，即是計算全體學生中，答對的比例 P 值有多少。 P 值越高，代表答對的人數越多，也就是題目簡單， P 值的範圍在 0~1 之間。例如，若是某題數學題目，全班學生有 70% ($P = 0.7$) 的學生答對，則此題比另一題只有 15% ($P = 0.15$) 學生答對的題目來的簡單。 $P = 0.5$ 左右為較適中的題目，在此難度下比較能夠分別不同程度的學生 (Anastasi & Urbina, 1997)。另外一種計算試題難度的方法為採用現代測驗理論 (Item Response Theory, 簡稱 IRT) 來計算試題參數，在 IRT 的系統下，難度 b 值的範圍為 $+\infty$ 到 $-\infty$ 之間，但大多統計軟體所估計下的 b 值之範圍在 -3 到 +3，一般來說， b 值小於 0 為簡單的試題，而 b 值大於 2 則為困難的題目，簡單的題目代表程度低的學生就可以答對，而難的題目則需程度高的學生才能夠答對， b 值在 0 左右的題目為難易適中的題目 (de Ayala, 2009)。模糊理論為一種新興的研究方法，而近幾年也應用在不同的教育領域中 (胡悅倫、陳皎眉、吳柏林, 2006; 汪慧瑜, 2005; 胡悅倫、吳柏林, 2002)，但多著重在心理量表的應用，並沒有任何相關文獻探討如何將模糊理論應用在計算試題難度上。因此，本研究想要應用模糊統計方法，發展一種新的評量試題難度的方法，提供教師一種新工具，藉由這個方法來了解到底是數學題目是在哪一個面向難度太高導致學生學習困難，或者有其他的原因，以作為教學參考，改進教學方法進而幫助更多學生，期望能提高學生數學學習的動機、興趣及自信。此外，本研究所發展出的難度，亦和傳統常用的難度與現代理論下的難度做比較。

文獻探討

一、難度

常用的難度有兩種，一種是指傳統的通過率，另一種則為現代測驗理論下的難度，以下分別敘述之。

(一) 傳統通過率

難度為用來表示每個試題的難易程度，一般常用答對百分比法來表示。答對百分比法，即是計算全體學生中，答對的比例有多少。難度的值越高，代表答對的人數越多，也就是題目越簡單。難度計算公式如下：

$$P = \frac{R}{N}$$

其中 R 為答對人數， N 為作答學生總人數。

例如，若全班人數有 50 人，答對某數學題的人數有 26 人，則此題的難度則為 $26/50=0.52$ ，這也代表有 52% 的學生答對這一題。

(二) 試題反應理論的難度

教育測驗中，當試題反應只分成對或錯兩種可能性時，即可使用二元化計分的單向度 IRT 模式來進行分析，而在不同的模式中，單參數為一種最簡單，但應用非常廣泛的一種。就單參數的對數模型來說，唯一考量的是試題的難度。此外，這個模組假設鑑別度對每一個題目而言都是相同的，且猜測度都是不存在的。單參數的數學公式可以表達如下：

$$P_i(\theta) = P_i(X_i = 1 | \theta) = \frac{1}{1 + e^{-(\theta - b_i)}}$$

其中 $P_i(\theta)$ 代表某一位受試者，他的能力值為 θ ，而答對第 i 題題目的機率。這個式子也可以用 $P_i(X_i = 1 | \theta)$ 來表示，這代表當受試者的能力值固定為 θ ，他答對第 i 題題目的機率。至於 b_i 則是代表試題的難度。由這個公式可知，在單向度單參數對數模型中，難度值是唯一會影響受試者答對題目機率的試題參數值。

此外，從這個式子可以看出。當受試者的能力值越高，代表這個學生的答對這題的機率越高。而試題難度越高，代表這題題目越難，答對題目的機率就越低。在此有一個重點要注意，在 IRT 的難度，和傳統把通過率當作難度指標的意義是截然不同的。在 IRT 裡，難度越高，代表題目越難，而傳統把通過率當作難度指標，則是數值越高，代表答對人數越多，題目越簡單。傳統的難度值，會受到抽取樣本的影響。若是剛好所選擇的受試者程度都很好，通過率就會高，就顯得題目很容易，但是若對相同的題目，給予一群程度比較差的學生來考，那這題通過率就會低，題目變的很難。傳統的難度值，有這種依賴抽樣的問題。然而，在 IRT 難度值估計上，卻可透過數學模組的轉換，使難度估計不會受到樣本的影響。試題反應理論下的難度值很難直接計算，大多是仰賴統計軟體，如 Bilog、Winsteps 來輔助計算。

二、模糊理論

(一) 模糊數

人類的感覺是模糊的。當有人說今天天氣很熱的時候，究竟他對於『熱』的定義為何？多少度的溫度範圍可以稱為『熱』呢？對於這樣的問題，每個人的回答皆因其主觀性而有不同，即使回答者為同一人，也會因為時地不同而有不同的回應。諸如此類很多的論點和問題，都不是能夠用絕對的二元邏輯所可以界定的。原因則皆來自人類思維的模糊性。但人類卻常常被要求做出絕對的判斷或選擇，以人性的觀點來看，這是十分不合理的。

模糊概念並不只侷限在研究人類的思維與情感而已。在以往嚴謹精確的原則要求下，許多技術層面所衍生出來的灰色地帶，都必須耗費相當大的心力為複雜的系統寫下嚴密的定義與敘述，灰色地帶中的每一個細微末節，都必須完全考慮到，盡全力使得其中的模糊變得明確，但若稍有遺漏，則全盤皆墨，一切又得從頭做起。而模糊理論確提供一種新的思維模式，只需要明瞭各種屬性的狀況，利用軟計算方法建立大略性的處理模式，即可處理系統中灰色地帶的問題。所以我們應該要了解：灰色或是模糊不清的事件是層出不窮的，也是無法完全避免的，也因此才讓我們體認到研究模糊理論的重要性。

隸屬度函數是模糊理論的基礎，它是從傳統集中的特徵函數 (characteristic function) 所衍生出來的，用以表達元素對模糊集合的隸屬度 (membership)，其範圍介於 0 到 1 之間。對於元素和集合關係，古典集合將元素和集合之間的關係以特徵函數來說明，亦即 $I(x) = 1$ ，若 $x \in A$ ； $I(x) = 0$ ，若 $x \notin A$ 。但是 Zadeh (1965) 在模糊集合論中提到，若一個元素屬於某一個集合的程度越大，則其隸屬度值越接近 1，反之則越接近 0 (王文俊，1997；江彥聖，2008；吳柏林，2002)。

隸屬度函數是模糊理論最基本的概念，它不僅可以描述模糊集合的性質，更可以對模糊集合進行量化，並且利用精確的數學方法，來分析和處理模糊性資訊。然而，要建立一個足以表達模糊概念的隸屬度函數，並不是一件容易的事。其原因在於隸屬度函數脫離不了個人的主觀意識，故沒有通用的定理或公式，通常是根據經驗或統計來加以確定，很難像客觀事物一樣有很強的說服力。因此，隸屬度函數的建立經常是具有爭議性的，也沒有一種隸屬度函數是可以被廣泛接受而使用的。

隸屬度函數可分為離散型 (discrete type) 與連續型 (continuous type) 兩種。離散型的隸屬度函數是直接給予有限模糊集合內每個元素的隸屬度，並以向量的形式表現出來；而連續型隸屬度

函數則有幾種常用的函數形式（S-函數、Z-函數、 π 函數、三角形函數、梯形函數、高斯（鐘形）函數）來描述模糊集合。函數定義的表現，可以是無限模糊集合的元素及其隸屬度之間的關係，也可以是有限模糊集合的元素及其隸屬度之間的關係（吳柏林，1997）。

傳統的統計方法透過一般的抽樣調查往往只能得到單一的數值資料、或是固定尺度的選擇，但如此並不足以能夠完整地反應人類個體的想法。若能讓受訪者根據自己的意識，利用隸屬度函數或區間值表達心中對於問項真正屬意的程度，則可更完整地傳達人類真實的思維。再考慮具有模糊特性的問項時，資料本身便具有不確定性與模糊性。

模糊理論是將元素和集合之間的關係，以介於[0,1]之間的隸屬度（membership）描述，若元素屬於集合的程度越大，則其隸屬度值就會越接近 1，反之，元素屬於集合程度越小，則其隸屬度值就越接近 0。有關模糊集合的意義，可簡述如下：令 Ω 表示論域（universal set）， u 為一對應到 [0,1]間的實數函數，即 $u : \Omega \rightarrow [0,1]$ 。U 之模糊子集合 A 的隸屬度函數記為 $u_A(x)$ ，表示元素 x 隸屬於模糊集合 A 的程度。則（1）在離散（discrete）的情形下，模糊集合 A 可表示成：

$$A = \frac{u_A(x_1)}{x_1} + \frac{u_A(x_2)}{x_2} + \dots + \frac{u_A(x_n)}{x_n}$$

其中+代表或的意思， $\frac{u_A(x_i)}{x_i}$ 表示元素 x_i 隸屬於模糊集合 A 的程度。（2）當 U 為連續時，模

糊集合 A 可表示成： $A = \int_{x \in A} \frac{u_A(x)}{x}$ 。

（二）模糊權重

傳統分類方法常把各因子以相同權重看待，但實際上，論域因子的權重應該不同，因為樣本的特徵因素未必同等重要。就像我們選購衣服時，先考慮衣服的價格、舒適度或顏色等等。可知不同的因子有著不同的權重，而對各因子重視的程度，又因人而異，不易評估。若以相同權重看待各因子，或者以主觀的個人意識作判斷去定義因子權重，易產生不適當的分類結果。

模糊統計計算權重時，先將所有因子以模糊問卷的形式依重要程度收集資訊，並採用統計方法轉換得到權重值。

假設論域集合 $S = \{S_1, S_2, \dots, S_k\}$ ，偏好序列 $r = \{r_1, r_2, \dots, r_f\}$ ，且 S_i 在 r_f 的隸屬度為 $\mu_{S_i, f}$ 。則論域因子的模糊權重 $FW = (FW_{S_1}, \dots, FW_{S_k})$ 定義為

$$FW_{S_i} = \sum_{l=1}^f \frac{\mu_{S_i, l}}{r_l} = \frac{\mu_{S_i, 1}}{r_1} + \frac{\mu_{S_i, 2}}{r_2} + \dots + \frac{\mu_{S_i, f}}{r_f}; i = 1, \dots, k$$

模糊權重表示各因子的自我權重分佈，但模糊權重分析的主要目的是求得因子所代表的權重值各是多少，亦即相對的權重（阮亨中、吳柏林，2000）。

爲了求論域因子的模糊相對權重，首先須設計一個模糊評鑑表（fuzzy evaluation table）的問卷形式如下：

表 1 模糊評鑑表 A 在偏好序列 U 的一般隸屬度 μ_A

r S	r_1	r_2	...	r_f
S_1	$\mu_{S_1, 1}$	$\mu_{S_1, 2}$...	$\mu_{S_1, f}$
S_2	$\mu_{S_2, 1}$	$\mu_{S_2, 2}$...	$\mu_{S_2, f}$
\vdots	\vdots	\vdots	\ddots	\vdots
S_k	$\mu_{S_k, 1}$	$\mu_{S_k, 2}$...	$\mu_{S_k, f}$

假設偏好程度為 r_1 最低， r_f 最高（例如非常不同意至非常同意的李克氏量表的偏好程度），並以此問卷訪問專家對論域集合在各偏好序列下的隸屬程度。這些專家評分平均後的各因子之偏好度，即為模糊權重集。則第 k 個因子的模糊權重則為

$$FW_{S_k} = \sum_{l=1}^f \mu_{S_k l} / r_l = \mu_{S_k 1} / r_1 + \mu_{S_k 2} / r_2 + \dots + \mu_{S_k f} / r_f$$

再利用 m 等第評分標準法，分別計算論域中各因子的模糊相對權重 FRW 。所謂 m 等第評分標準法，就是將 m 個偏好序列 r 視為 f 個等第，對此 f 個偏好序列取數量化。亦即，給定 r_1 為一分，給定 r_2 為二分，如此繼續到給定 r_f 為 f 分。根據所得的隸屬度乘上其相對應的分數，分別求它們相對的模糊權重分佈，則為各因子的模糊相對權重（吳柏林，2005）。則第 k 個因子的模糊相對權重 FRW 為

$$FRW_{S_k} = \frac{\sum_{l=1}^f l \cdot \mu_{S_k l}}{\sum_{k=1}^m \sum_{l=1}^f l \cdot \mu_{S_k l}}$$

本研究則將以上的方法，為多變量分類考慮各因子的模糊相對權數法。

例如，考慮以模糊評鑑表訪問專家在評定試題難易度時，衡量試題所採用的評價因素之重要程度。假設訪問數位專家對模糊評鑑表評分，經過統計平均後的資料如下：

表 2 平均後之偏好度

偏好序列 評鑑項目	很不重要	不重要	普通	重要	很重要
題目敘述	0.1	0.37	0.7	0.77	0.67
解題策略	0	0.1	0.4	0.8	0.92
計算過程	0.2	0.47	0.8	0.7	0.3
創新程度	0.3	0.7	0.6	0.2	0.92

我們可算出此四因子的模糊權重分別為

$$FW_{\text{題目敘述}} = 0.1/\text{很不重要} + 0.37/\text{不重要} + 0.7/\text{普通} + 0.77/\text{重要} + 0.67/\text{很重要}$$

$$FW_{\text{解題策略}} = 0/\text{很不重要} + 0.1/\text{不重要} + 0.4/\text{普通} + 0.8/\text{重要} + 0.92/\text{很重要}$$

$$FW_{\text{計算過程}} = 0.2/\text{很不重要} + 0.47/\text{不重要} + 0.8/\text{普通} + 0.7/\text{重要} + 0.3/\text{很重要}$$

$$FW_{\text{創新程度}} = 0.3/\text{很不重要} + 0.7/\text{不重要} + 0.6/\text{普通} + 0.2/\text{重要} + 0.1/\text{很重要}$$

再利用五等第評分標準法，給定「很重要」為 5 分，「重要」為 4 分，「普通」為 3 分，「不重要」為 2 分，而「很不重要」則給定為 1 分。根據表 2 上的隸屬度分別乘以其相對偏好序列給定的數量，各因子的模糊相對權重為：

$$\sum_{l=1}^5 l \cdot \mu_{\text{題目敘述},l} = 1 \times (0.1) + 2 \times (0.37) + 3 \times (0.7) + 4 \times (0.77) + 5 \times (0.67) = 9.37$$

$$\sum_{l=1}^5 l \cdot \mu_{\text{解題策略},l} = 1 \times (0) + 2 \times (0.1) + 3 \times (0.4) + 4 \times (0.8) + 5 \times (0.92) = 9.2$$

$$\sum_{l=1}^5 l \cdot \mu_{\text{計算過程},l} = 1 \times (0.2) + 2 \times (0.47) + 3 \times (0.8) + 4 \times (0.7) + 5 \times (0.3) = 7.84$$

$$\sum_{l=1}^5 l \cdot \mu_{\text{創新程度},l} = 1 \times (0.3) + 2 \times (0.7) + 3 \times (0.6) + 4 \times (0.2) + 5 \times (0.1) = 4.8$$

因為 $\sum_{i=1}^k \sum_{l=1}^f l \cdot \mu_{S,l} = 9.37 + 9.2 + 7.84 + 4.8 = 31.21$ ，故此四個項目的模糊相對權重分別為

$$FRW_{\text{題目敘述}} = \frac{9.37}{31.21} = 0.30$$

$$FRW_{\text{解題策略}} = \frac{9.2}{31.21} = 0.29$$

$$FRW_{\text{計算過程}} = \frac{7.84}{31.21} = 0.25$$

$$FRW_{\text{創新程度}} = \frac{4.8}{31.21} = 0.16$$

故得「題目敘述」因子的權重為 0.30，「解題策略」因子的權重為 0.29，「計算過程」因子的權重為 0.25，「創新程度」因子的權重為 0.16。

(三) 模糊集合的軟運算

軟計算 (soft computing) 與傳統實數計算方法 (硬計算) 不同的是，軟計算乃基於模糊數或模糊樣本，包括區間數，多值數，語言變數等的數學運算。舉例來說：某大學生每週運動 2~4 回，每回 1~2 小時，那一周運動幾小時？又如某人出外旅遊 7~8 天，需攜帶若干胃腸藥，而胃腸藥標示成人每日 3~4 回，每回 5~7 顆，請問他須攜帶多少顆。傳統數學並無法對此作一明確的計算，但是一般人均會概略的計算。這實在應屬於軟計算過程。模糊集合之基本運算有以下幾點原則 (吳柏林, 2005)：

設 A, B 為論域 U 中的兩個模糊集合，其隸屬度表徵為 μ_A, μ_B 。

- (1) 補集： $\mu_{A^c}(x) = 1 - \mu_A(x)$
- (2) 交集： $\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x)) = \mu_A \wedge \mu_B$
- (3) 聯集： $\mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x)) = \mu_A(x) \vee \mu_B(x)$
- (4) 距離： d_1 = 歐基里得 (Euclidian)， d_2 = 漢民 (Hamming)

$$d_1(A, B) = \frac{1}{\sqrt{n}} \sqrt{\sum_{i=1}^n (\mu_A(x_i) - \mu_B(x_i))^2} \quad (n = \text{論域因子集個數})$$

$$d_2(A, B) = \frac{1}{n} \sum_{i=1}^n 1_{\{\mu_A(x_i) - \mu_B(x_i) > 0\}}(x_i) \quad (n = \text{論域因子集個數})$$

舉例來說，假設 A、B 兩國人民對宗教的隸屬度為

$$A = \frac{0.3}{\text{佛教}} + \frac{0.2}{\text{基督教}} + \frac{0.1}{\text{回教}} + \frac{0.3}{\text{其他教}} + \frac{0.1}{\text{未信教}},$$

$$B = \frac{0.2}{\text{佛教}} + \frac{0.3}{\text{基督教}} + \frac{0.2}{\text{回教}} + \frac{0.1}{\text{其他教}} + \frac{0.2}{\text{未信教}}$$

則可得以下之關係

$$A^c = \frac{0.7}{\text{佛教}} + \frac{0.8}{\text{基督教}} + \frac{0.9}{\text{回教}} + \frac{0.7}{\text{其他教}} + \frac{0.9}{\text{未信教}}$$

$$A \cap B = \frac{0.2}{\text{佛教}} + \frac{0.2}{\text{基督教}} + \frac{0.1}{\text{回教}} + \frac{0.1}{\text{其他教}} + \frac{0.1}{\text{未信教}}$$

$$A \cup B = \frac{0.3}{\text{佛教}} + \frac{0.4}{\text{基督教}} + \frac{0.2}{\text{回教}} + \frac{0.3}{\text{其他教}} + \frac{0.2}{\text{未信教}}$$

$$d_1(A, B) = \frac{1}{\sqrt{5}} \sqrt{(0.1^2 + 0.1^2 + 0.1^2 + 0.2^2 + 0.1^2)} = 0.126$$

$$d_2(A, B) = \frac{1}{5} \sum_{i=1}^5 1_{\{\mu_A(x_i) - \mu_B(x_i) > 0\}}(x_i) = 1$$

研究方法

Polya (1945) 提出解題活動的四個主要階段是：了解問題、擬定計畫、執行計畫、回顧過程（驗算、檢討）。了解問題可以讓我們清楚知道，什麼才是我們要尋找的解答，再根據問題中已知數與未知數存在的各種關係，訂定解題計畫，然後就是耐心地進行數學計算，求得答案後，最好能再逐步檢查、驗算，並討論其意義，才能在解題活動中得到最好的結果。

而在基測的考試過程中，七十分鐘的應答時間對大部分的學生來說僅是剛好足夠而已，甚至部分學生連作答全部的題目都有困難。鮮少有學生能夠利用多餘的時間去作回顧過程這個動作。再者，回顧過程這個因素似乎沒有什麼難度的差別。所以，本研究把「回顧過程」這個因子忽略。

從這幾年的基測試題發現，很多學生認為較難的題目，都有個共同的特徵，就是題目很創新。學生對於沒有見過的題目，在作答的表現上往往都不是非常理想。在教學的過程中，我們也常常跟學生灌輸一個概念：基測的題目在新不在難！所以，本研究亦將“創新”這個向度列為決定試題難易的一個因子。

綜合以上，將波利亞的解題的前三個步驟，再加上創新。決定出了四個因子：題目敘述、解題策略、計算過程、創新程度。本研究將以模糊問卷調查的方法，去分析試題分別在這四個因子上的難度。

本研究將以發專家問卷的方式，取得專家對 98 學年度第一次國中基本學力測驗的難度的模糊問卷。而此問卷主要分為兩部分，第一部分為專家們對於題目敘述、解題策略、計算過程、創新程度這四個決定難度的因子的重要性的看法。第二部分為專家對每道個別題目，分別以四個向度去決定其隸屬度。這兩部份所得到的值將用來計算模糊難度。

以下先對本研究所發展的模糊難度進行說明。

一、模糊難度

在說明模糊難度之前，須先了解聚類點、聚類距離、聚類模糊數與聚類比例，以下分別定義之。

(一) 聚類點 (clustering point)

設 U 為一論域，令 $L = \{L_1, L_2, \dots, L_k\}$ 為佈於論域 U 的 k 個語言變數。則定義 $L'_i = \frac{\mu_1}{L_1} + \frac{\mu_2}{L_2} + \dots + \frac{\mu_k}{L_k}$ ，其中 $\mu_j = \begin{cases} 1 & j=i \\ 0 & j \neq i \end{cases}$ 。稱為對於語言變數 L_i 的聚類點 (clustering point)。舉例來說，考慮以模糊評鑑表訪問專家在評定試題難易度時， $L = \{L_1, L_2, L_3, L_4, L_5\} = \{\text{非常容易、容易、中等、困難、非常困難}\}$ 為五個語言變數，則

$$L'_1 = \frac{1}{\text{非常容易}} + \frac{0}{\text{容易}} + \frac{0}{\text{中等}} + \frac{0}{\text{困難}} + \frac{0}{\text{非常困難}},$$

$$L'_2 = \frac{0}{\text{非常容易}} + \frac{1}{\text{容易}} + \frac{0}{\text{中等}} + \frac{0}{\text{困難}} + \frac{0}{\text{非常困難}},$$

$$L'_3 = \frac{0}{\text{非常容易}} + \frac{0}{\text{容易}} + \frac{1}{\text{中等}} + \frac{0}{\text{困難}} + \frac{0}{\text{非常困難}},$$

$$L'_4 = \frac{0}{\text{非常容易}} + \frac{0}{\text{容易}} + \frac{0}{\text{中等}} + \frac{1}{\text{困難}} + \frac{0}{\text{非常困難}},$$

$$L'_5 = \frac{0}{\text{非常容易}} + \frac{0}{\text{容易}} + \frac{0}{\text{中等}} + \frac{0}{\text{困難}} + \frac{1}{\text{非常困難}}.$$

分別為對於五個語言變數的聚類點。

(二) 聚類距離 (distance between sample and clustering point)

設 U 為一論域，令 $L = \{L_1, L_2, \dots, L_k\}$ 為佈於論域 U 的 k 個語言變數。 L'_i 為對於語言變數 L_i 的聚類點。若 A 為一離散模糊數，且 A 對於語言變數 L_i 的隸屬度為 $\mu_i(A)$ ，則定義

$$D_i(A) = \frac{1}{\sqrt{k}} \sqrt{\sum_{j=1}^k (1+(i-j)^2)(\mu_j(A) - \delta_{ij})^2}, \delta_{ij} = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases}$$

為 A 對語言變數 L_i 的聚類距離。本文所定義的聚類距離是根據歐基里德距離

$d(A, B) = \frac{1}{\sqrt{n}} \sqrt{\sum_{i=1}^n (\mu_A(x_i) - \mu_B(x_i))^2}$ 所做的修正，修正的原因是問卷題目的語言變數{非常容易、容易、中等、困難、非常困難}，含有偏好序列的成分。例如，若 $A = \frac{0}{\text{非常容易}} + \frac{0.3}{\text{容易}} + \frac{0.6}{\text{中等}} + \frac{0.1}{\text{困難}} + \frac{0}{\text{非常困難}}$ 為一離散模糊數，若採用歐基里德距離的算法，分別去算 A 與非常容易和非常困難這兩個語言變數聚類點的距離

$$\begin{aligned} d(A, L_1^i) &= \frac{1}{\sqrt{5}} \sqrt{(0-1)^2 + (0.3-1)^2 + (0.6-0)^2 + (0.1-0)^2 + (0-0)^2} \\ &\approx 0.54 \end{aligned}$$

$$\begin{aligned} d(A, L_5^i) &= \frac{1}{\sqrt{5}} \sqrt{(0-0)^2 + (0.3-0)^2 + (0.6-0)^2 + (0.1-0)^2 + (0-1)^2} \\ &\approx 0.54 \end{aligned}$$

可以發現這兩個距離是一模一樣的。但是 A 這個模糊數的隸屬度很明顯的比較偏向中間偏左，但對非常容易和非常困難的距離卻一樣，這就造成不合理的現象。為了修正這個不合理的情況，在原公式中多乘上了一個修正係數： $(1+(i-j)^2)$ ，讓聚類距離的公式變得合理化。這個修正係數是假定樣本與聚類點的距離為 1，而五個語言變數之間的距離也為 1（如圖 1）。

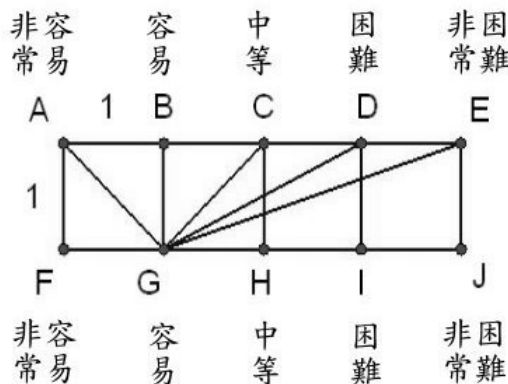


圖 1 語言變數之間的距離

利用畢氏定理，我們可以得到 G 點到 A、B、C、D、E 五個點的距離分別是 $\overline{GA} = \sqrt{2}$ 、 $\overline{GB} = 1$ 、 $\overline{GC} = \sqrt{2}$ 、 $\overline{GD} = \sqrt{5}$ 、 $\overline{GE} = \sqrt{10}$ 。從原本距離公式： $\frac{1}{\sqrt{n}} \sqrt{\sum_{i=1}^n (\mu_A(x_i) - \mu_B(x_i))^2}$ 觀察到

$(\mu_A(x_i) - \mu_B(x_i))^2$ 代表樣本與聚類點在單一語言變項的距離的平方，所以在此公式前，我們嘗試乘上一個係數 $(1+(i-j)^2)$ ，此數代表距離倍數的平方，即 $(\overline{GA})^2 = 2(\overline{GB})^2$ 、 $(\overline{GC})^2 = 2(\overline{GB})^2$ 、 $(\overline{GD})^2 = 5(\overline{GB})^2$ 、 $(\overline{GE})^2 = 10(\overline{GB})^2$ 。如此便可修正原先公式造成帶有偏好序列性質的語言變項中不合理的情況。以下為聚類距離的實例：

$L = \{L_1, L_2, L_3, L_4, L_5\} = \{\text{非常容易、容易、中等、困難、非常困難}\}$ 為五個語言變數，若 $A = \frac{0}{\text{非常容易}} + \frac{0.3}{\text{容易}} + \frac{0.6}{\text{中等}} + \frac{0.1}{\text{困難}} + \frac{0}{\text{非常困難}}$ 為一離散模糊數，則

$$D_1(A) = \frac{1}{\sqrt{5}} \sqrt{(0-1)^2 + 2(0.3-0)^2 + 5(0.6-0)^2 + 10(0.1-0)^2 + 17(0-0)^2} \approx 0.78$$

$$D_2(A) = \frac{1}{\sqrt{5}} \sqrt{2(0-0)^2 + (0.3-1)^2 + 2(0.6-0)^2 + 5(0.1-0)^2 + 10(0-0)^2} \approx 0.50$$

$$D_3(A) = \frac{1}{\sqrt{5}} \sqrt{5(0-0)^2 + 2(0.3-0)^2 + (0.6-1)^2 + 2(0.1-0)^2 + 5(0-0)^2} \approx 0.26$$

$$D_4(A) = \frac{1}{\sqrt{5}} \sqrt{10(0-0)^2 + 5(0.3-0)^2 + 2(0.6-0)^2 + (0.1-1)^2 + 2(0-0)^2} \approx 0.62$$

$$D_5(A) = \frac{1}{\sqrt{5}} \sqrt{17(0-0)^2 + 10(0.3-0)^2 + 5(0.6-0)^2 + 2(0.1-0)^2 + (0-1)^2} \approx 0.86$$

分別為對於五個語言變數的聚類距離。

此時得到的聚類距離，代表一個模糊樣本與五個語言變數的聚類點之間的距離。此距離值越短，表示此樣本越接近該語言變數。而為了表示此樣本對於各個語言變數的隸屬度，我們利用倒數轉換，將此距離轉化成另外一個數值，即以下定義的聚類比例。

(三) 聚類比例 (clustering ratio) 和聚類模糊數 (clustering fuzzy number)

設 U 為一論域，令 $L = \{L_1, L_2, \dots, L_k\}$ 為佈於論域 U 的 k 個語言變數。 L'_i 為對於語言變數 L_i 的聚類點。若 A 為一離散模糊數，且 $D_i(A)$ 為對於語言變數 L_i 的聚類距離，則定義

$$R_i(A) = \frac{\frac{1}{D_i(A)}}{\sum_j \left(\frac{1}{D_j(A)} \right)}$$

為 A 對語言變數 L_i 的聚類比例。

又定義

$$R(A) = \sum_i \frac{R_i(A)}{L_i}$$

為 A 對於對語言變數 L_i 的聚類模糊數

舉例來說， $L = \{L_1, L_2, L_3, L_4, L_5\} = \{\text{非常容易、容易、中等、困難、非常困難}\}$ 為五個語言變數，若 $A = \frac{0}{\text{非常容易}} + \frac{0.3}{\text{容易}} + \frac{0.6}{\text{中等}} + \frac{0.1}{\text{困難}} + \frac{0}{\text{非常困難}}$ 為一離散模糊數，可算出 $D_1(A) \approx 0.78$ 、 $D_2(A) \approx 0.50$ 、 $D_3(A) \approx 0.26$ 、 $D_4(A) \approx 0.62$ 、 $D_5(A) \approx 0.86$ ，則

$$R_1(A) = \frac{\frac{1}{0.78}}{\frac{1}{0.78} + \frac{1}{0.50} + \frac{1}{0.26} + \frac{1}{0.62} + \frac{1}{0.86}} \approx 0.13$$

$$R_2(A) = \frac{\frac{1}{0.50}}{\frac{1}{0.78} + \frac{1}{0.50} + \frac{1}{0.26} + \frac{1}{0.62} + \frac{1}{0.86}} \approx 0.21$$

$$R_3(A) = \frac{\frac{1}{0.26}}{\frac{1}{0.78} + \frac{1}{0.50} + \frac{1}{0.26} + \frac{1}{0.62} + \frac{1}{0.86}} \approx 0.38$$

$$R_4(A) = \frac{\frac{1}{0.62}}{\frac{1}{0.78} + \frac{1}{0.50} + \frac{1}{0.26} + \frac{1}{0.62} + \frac{1}{0.86}} \approx 0.16$$

$$R_5(A) = \frac{\frac{1}{0.86}}{\frac{1}{0.78} + \frac{1}{0.50} + \frac{1}{0.26} + \frac{1}{0.62} + \frac{1}{0.86}} \approx 0.12$$

分別為對於五個語言變數的聚類比例。

$$\text{而 } R(A) = \frac{0.13}{\text{非常容易}} + \frac{0.21}{\text{容易}} + \frac{0.38}{\text{中等}} + \frac{0.16}{\text{困難}} + \frac{0.12}{\text{非常困難}}$$

為 A 對於對語言變數 L 的聚類模糊數。

之前定義的聚類模糊數本身已經是具有相當程度參考價值的資料，也具有相當程度的解釋能力。但大眾較習慣難度為一個單一的數值，而非一組模糊數，因此，在此進一步定義難度指標如下。

(四) 難易度指標 (difficulty index)

設 A 為一代表試題難度之離散模糊數， $R(A)$ 為 A 之聚類模糊數， $L = \{L_1, L_2, \dots, L_k\}$ 為佈於論域 U 的 k 個語言變數。 $R_i(A)$ 為對於語言變數 L_i 的聚類比例， $\sum_{i=1}^k R_i(A) = 1$ 。 m_i 為被賦予 L_i 的難度權重值。則稱 $DI(A) = \sum_{i=1}^k m_i R_i(A)$ 為模糊數 A 的難易度指標。

舉例來說， $L = \{L_1, L_2, L_3, L_4, L_5\} = \{\text{非常容易、容易、中等、困難、非常困難}\}$ 為五個語言變數，若 $A = \frac{0}{\text{非常容易}} + \frac{0.3}{\text{容易}} + \frac{0.6}{\text{中等}} + \frac{0.1}{\text{困難}} + \frac{0}{\text{非常困難}}$ 為一離散模糊數，而聚類比例

$$R(A) = \frac{0.13}{\text{非常容易}} + \frac{0.21}{\text{容易}} + \frac{0.38}{\text{中等}} + \frac{0.16}{\text{困難}} + \frac{0.12}{\text{非常困難}}。若分別賦予五個語言變數的權重為：$$

非常容易=0，容易=0.25，普通=0.5，困難=0.75，非常困難=1

則 A 的難易度指標 $DI(A)$ 為

$$0.13 \times 0 + 0.21 \times 0.25 + 0.38 \times 0.5 + 0.16 \times 0.75 + 0.12 \times 1 = 0.48$$

總結來說，本文將以歐基里德距離為基礎，定義出另一種新的距離，稱之為聚類距離。此距離由問卷中所得到的模糊樣本值加以平均，而其基本概念是，距離越短，代表越偏向該語言變數；反之，距離越長，則代表越偏離該語言變數。接著再將聚類距離轉換成聚類比例，以此比例當成原模糊數對於五個語言變數的隸屬度，而得到一個新的模糊數—聚類模糊數。

其實這時所得到的聚類模糊數，其本身即具備一定程度的解釋能力。但若要比較各題之間難易度，則可利用反模糊化的概念，最後將每一個聚類距離模糊數的反模糊化值求出，即可以此數代表原樣本所代表的題目在某個向度的難度指標。依照此難易度指標，我們除了可以分向度來比較各題之間的難度順序之外，甚至可以將各項度的難度指數分別乘上模糊相對權重值，即得到各題目的綜合加權難易度指標。

二、研究工具

(一) 抽樣調查與問卷設計

本研究利用軟統計方法來進行問卷調查之解析，期以更接近人類思維模式來反應專家們對於 98 學年度第一次國中基本學力測驗數學科試題難度的看法。本問卷的抽取對象是台北縣市國高中的數學科教師，有效問卷共 20 份。填寫問卷時間在 98 學年度下學期第 17 周。問卷問題共 34 題，共分為兩部分。第一部分請填寫者填寫決定難度的因子的重要性，即請教師在權重部分以 0 到 1 之間的數字，給予以隸屬度，且每題隸屬度總和為 1；第一部份的指導語為：請針對您個人教學經驗，對於決定試題難度的各項因子之重要性來進行填答。各能力重要性之權重，請依各項填入 0~1，合計權重為 1。例如：您覺得題目敘述對於試題難度並非很重要，感覺介於很不重要到不重要之間。則可將權重設為很不重要 0.8，不重要 0.2。您在各項的可能填法為：

語言變數難度 專家	很不重要	不重要	普通	重要	非常重要
題目敘述	0.5	0.4	0.1		
解題策略			0.4	0.6	
計算過程		0.6	0.4		
創新程度		0.4	0.6		

第二部分則是填寫各向度在每個題目下的難度。每題我們請填寫者以 0 到 1 之間的數字，給予以隸屬度，且每題隸屬度總和為 1。例如：第 1 題題目敘述，若您覺得「普通偏容易」，則可在「容易」項目下填入 0.6，「普通」項目下填入 0.4；若覺此題的解題策略非常容易，則在該題「容易」項目下填入 1 分。若此題的計算過程覺得較偏向困難，則可在「普通」項目下填入 0.2，「困難」項目下填入 0.8；若覺得此題很創新，但學生會感到達對此題非常困難，則在該題「困難」項目下 x 填入 1 分。分數可以依您的感受作調整，但每項目的總分為 1 分。您的可能填法為：

		非常容易	容易	普通	困難	非常困難
第 1 題	題目敘述	0.6		0.4		
	解題策略	1				
	計算過程			0.2	0.8	
	創新程度					1

(二) 試題內容

本研究所採用的試題為 98 學年度第一次國中基本學力測驗數學科試題的試題，題目總共有 34 題，全部均為四選一的單選題，測驗時間為 70 分鐘。國中基測為教育部委託臺灣師大心測中心設計編製的標準化測驗，主要目的為評量國中生在校的學習成就表現，是屬於總結性評量的測驗工具。基測題庫中的試題都以試題反應理論的估計程序，將試題難度連結在同一把尺上，以國中基測而言，由於測量的是學生的基本學力，題目設計的難度為中間偏易，大部分題目的答對率均在 50%~75% 左右（國民中學學生基本學力測驗推動工作委員會，2009）。題目詳細內容可至國民中學學生基本學力測驗推動工作委員會網站下載（<http://www.bctest.ntnu.edu.tw>）。

結果

一、各因子的模糊權重

根據問卷的第一部分的結果將其平均，求得四個決定試題難度因子的模糊權重以及模糊相對權重，整理如下表 3：

表 3 難度四因子的模糊相對權重

	題目敘述	解題策略	計算過程	創新程度
很不重要	0	0	0	0.02
不重要	0.03	0	0.05	0.19
普通	0.40	0.11	0.27	0.47
重要	0.55	0.72	0.63	0.31
非常重要	0.02	0.17	0.05	0.01
模糊權重	3.56	4.06	3.68	3.10
模糊相對權重	24.72%	28.19%	25.55%	21.52%

由此表可知，本問卷的研究對象認為決定難度最重要的因子是解題策略，其次是計算過程，再來是題目敘述，最後才是創新程度。而其重要性的權重之百分比分別為 28.19%、25.55%、24.72%、21.52%。

二、各難度向度的結果分析

將問卷數據的模糊均數求出，再依照研究方法分四個向度，算出各題對於五個語言變數的聚類距離與聚類比例，以表格方式整理如下表 4~7。

表 4 題目敘述向度各題的聚類比例

題號	題目敘述向度的聚類比例				
	非常容易	容易	普通	困難	非常困難
1	48.64	20.34	13.23	9.95	7.85
2	14.38	23.37	34.75	15.58	11.91
3	61.65	14.87	10.04	7.52	5.92
4	10.96	14.91	44.79	17.69	11.65
5	50.15	19.57	12.93	9.69	7.65
6	42.93	22.54	14.74	11.05	8.73
7	43.67	22.15	14.64	10.91	8.64
8	15.91	35.58	23.60	14.14	10.77
9	11.64	17.89	45.21	14.50	10.77
10	37.95	25.43	15.58	11.76	9.28
11	24.63	37.05	16.39	12.34	9.58
12	53.21	18.00	12.35	9.18	7.26
13	12.19	18.63	42.82	15.16	11.21
14	19.07	43.60	16.26	11.92	9.15
15	10.25	15.16	51.23	13.52	9.85
16	48.72	20.31	13.18	9.95	7.84
17	9.97	14.68	52.51	13.23	9.62
18	11.59	16.22	43.37	17.04	11.78
19	8.20	11.60	60.23	11.75	8.23
20	10.87	14.06	21.97	35.50	17.60
21	11.68	15.26	28.02	29.31	15.75
22	11.34	15.26	41.51	19.38	12.51
23	9.63	13.17	52.16	15.00	10.04
24	10.80	14.20	23.92	35.26	15.82
25	9.90	12.94	19.37	41.00	16.79
26	11.56	15.34	39.86	20.16	13.07
27	11.94	15.87	34.28	23.70	14.22
28	8.77	11.44	16.10	46.47	17.22
29	12.04	15.99	33.19	24.36	14.41
30	12.17	16.25	28.67	27.61	15.29
31	5.20	7.29	74.85	7.43	5.23
32	10.10	14.69	52.24	13.24	9.72
33	8.97	11.70	16.82	45.68	16.82
34	9.59	12.46	16.75	40.19	21.00

表 5 解題策略向度各題的聚類比例

題號	解題策略向度的聚類比例				
	非常容易	容易	普通	困難	非常困難
1	12.20	59.80	12.62	8.77	6.61
2	23.52	38.50	16.27	12.23	9.47
3	11.79	15.35	25.37	30.35	17.14
4	13.40	19.85	34.92	18.80	13.03
5	48.94	19.87	13.33	9.98	7.88
6	7.88	11.40	62.09	10.85	7.78
7	9.69	13.99	53.81	13.09	9.43
8	9.12	12.39	54.91	14.11	9.48
9	9.62	13.23	52.51	14.68	9.97
10	15.74	37.37	22.44	13.94	10.52
11	14.31	24.17	33.60	15.96	11.96
12	19.29	44.37	15.72	11.68	8.94
13	6.65	9.46	68.04	9.25	6.60
14	18.82	46.32	15.09	11.20	8.57
15	16.31	32.80	25.03	14.62	11.24
16	15.25	45.14	18.29	12.12	9.20
17	15.86	35.84	23.35	14.20	10.76
18	13.65	22.64	36.32	15.65	11.74
19	7.87	10.86	61.69	11.60	7.98
20	9.30	12.80	54.05	14.23	9.62
21	12.42	19.09	41.86	15.30	11.33
22	13.85	22.57	36.14	15.63	11.81
23	9.91	13.55	50.75	15.42	10.37
24	6.02	8.67	71.01	8.33	5.97
25	11.48	15.10	24.04	32.55	16.84
26	9.94	14.68	52.51	13.23	9.64
27	10.76	13.86	19.02	34.01	22.36
28	8.92	11.61	16.32	45.51	17.65
29	6.58	9.09	67.99	9.66	6.68
30	8.28	11.88	60.05	11.57	8.22
31	9.93	13.04	19.99	40.85	16.19
32	10.37	13.39	17.97	35.41	22.85
33	10.68	13.96	22.23	36.59	16.54
34	11.44	15.55	41.58	18.98	12.45

表 6 計算過程向度各題的聚類比例

題號	計算過程向度的聚類比例				
	非常容易	容易	普通	困難	非常困難
1	45.50	21.40	14.17	10.57	8.36
2	51.66	19.05	12.49	9.39	7.41
3	14.43	24.20	32.54	16.60	12.23
4	23.71	37.38	16.75	12.46	9.69
5	13.78	22.97	36.02	15.51	11.72
6	14.28	24.81	33.60	15.55	11.76
7	15.23	28.56	29.15	15.38	11.67
8	18.70	46.14	15.27	11.26	8.63
9	23.52	38.50	16.27	12.23	9.47
10	14.25	24.13	33.58	16.05	11.98
11	8.22	11.93	60.48	11.27	8.10
12	16.77	40.57	19.41	13.22	10.03
13	17.19	33.53	23.19	14.79	11.30
14	39.45	25.22	14.98	11.38	8.96
15	36.91	26.73	15.41	11.72	9.22
16	11.35	16.98	46.19	14.74	10.74
17	12.44	18.86	41.14	15.96	11.60
18	15.96	37.49	22.34	13.74	10.47
19	15.52	27.75	29.35	15.53	11.86
20	59.49	15.81	10.55	7.92	6.23
21	10.35	15.43	50.65	13.64	9.93
22	12.60	20.06	40.90	15.14	11.31
23	60.84	15.16	10.27	7.68	6.05
24	10.70	14.96	47.26	16.07	11.01
25	8.56	11.81	58.16	12.69	8.78
26	8.49	12.44	59.29	11.47	8.31
27	9.89	13.06	20.13	41.11	15.81
28	10.18	15.10	51.99	13.05	9.67
29	11.55	16.59	45.80	15.10	10.96
30	14.49	24.17	32.99	16.24	12.12
31	6.92	9.18	13.07	57.95	12.88
32	9.79	13.50	51.66	14.91	10.14
33	7.48	10.35	63.49	11.08	7.61
34	10.07	14.95	51.91	13.36	9.71

表 7 創新程度向度各題的聚類比例

題號	創新程度向度的聚類比例				
	非常容易	容易	普通	困難	非常困難
1	60.90	15.23	10.20	7.65	6.02
2	12.59	61.45	11.39	8.32	6.25
3	16.47	36.78	22.33	13.82	10.61
4	9.36	12.74	53.56	14.60	9.75
5	59.49	15.81	10.55	7.92	6.23
6	56.77	16.96	11.21	8.43	6.64
7	63.86	14.02	9.46	7.09	5.57
8	9.86	13.09	20.65	41.03	15.36
9	11.48	17.82	45.82	14.26	10.62
10	18.21	45.75	15.80	11.45	8.79
11	62.36	14.63	9.84	7.38	5.80
12	56.77	16.96	11.21	8.43	6.64
13	9.74	13.68	52.36	14.34	9.89
14	62.36	14.63	9.84	7.38	5.80
15	13.88	50.70	16.35	10.89	8.18
16	63.86	14.02	9.46	7.09	5.57
17	15.84	50.64	14.81	10.63	8.09
18	11.93	15.94	33.33	24.48	14.32
19	11.25	15.29	42.37	18.89	12.20
20	12.21	16.60	33.85	23.22	14.13
21	15.71	34.70	24.17	14.49	10.93
22	9.53	13.62	53.83	13.52	9.51
23	8.92	11.60	16.06	45.01	18.41
24	11.15	16.52	46.61	14.96	10.75
25	9.83	12.59	16.88	35.31	25.38
26	15.67	33.26	25.15	14.77	11.14
27	10.81	13.77	20.03	34.15	21.23
28	11.44	15.55	41.58	18.98	12.45
29	10.75	15.12	47.02	16.10	11.01
30	11.84	17.82	43.56	15.55	11.23
31	10.34	13.68	22.33	38.07	15.58
32	9.70	12.25	16.49	26.30	35.26
33	9.46	11.97	15.95	26.42	36.19
34	8.99	11.38	15.14	24.26	40.23

三、模糊難度指標及傳統試題難度

將問卷數據的模糊均數求出，再依照研究方法分四個向度，算出各題對於五個語言變數的各向度的難度指標，以表格方式整理如下表 8。

表 8 各向度的各題難度指標及排名

題號	題目敘述		解題策略		計算過程		創新程度	
	難度	排名	難度	排名	難度	排名	難度	排名
1	0.27	30	0.34	33	0.29	31	0.21	30
2	0.47	23	0.36	31	0.25	32	0.34	26
3	0.20	34	0.56	7	0.47	17	0.41	22
4	0.51	13	0.50	20	0.37	26	0.51	13
5	0.26	32	0.27	34	0.47	16	0.21	29
6	0.30	28	0.50	18	0.46	20	0.23	27
7	0.30	29	0.50	19	0.45	22	0.19	33
8	0.42	24	0.51	10	0.36	28	0.60	7
9	0.49	21	0.51	11	0.36	27	0.49	19
10	0.32	27	0.42	28	0.47	18	0.37	24
11	0.36	26	0.47	25	0.50	7	0.20	31
12	0.25	33	0.37	30	0.40	25	0.23	27
13	0.49	22	0.50	15	0.42	23	0.50	15
14	0.37	25	0.36	32	0.31	30	0.20	31
15	0.49	20	0.43	26	0.32	29	0.37	23
16	0.27	31	0.39	29	0.49	13	0.19	33
17	0.49	18	0.42	27	0.49	14	0.36	25
18	0.50	15	0.47	23	0.41	24	0.53	9
19	0.50	16	0.50	13	0.45	21	0.51	11
20	0.59	5	0.51	12	0.21	33	0.53	10
21	0.56	7	0.49	22	0.49	10	0.43	21
22	0.52	12	0.47	24	0.48	15	0.50	16
23	0.51	14	0.51	9	0.21	34	0.63	5
24	0.58	6	0.50	17	0.50	4	0.49	17
25	0.60	4	0.57	6	0.50	5	0.63	4
26	0.52	11	0.49	21	0.50	8	0.43	20
27	0.53	10	0.61	3	0.60	2	0.60	6
28	0.63	1	0.63	1	0.49	12	0.51	12
29	0.53	9	0.50	14	0.49	11	0.50	14
30	0.54	8	0.50	16	0.47	19	0.49	18
31	0.50	17	0.60	4	0.65	1	0.59	8
32	0.49	19	0.62	2	0.51	3	0.66	3
33	0.62	3	0.59	5	0.50	6	0.67	2
34	0.63	2	0.51	8	0.49	9	0.69	1

從表 8 可看出，各題在各向度的難度指標及難度排名。例如在題目敘述和解題策略向度難度最高的都是第 28 題，在計算過程向度難度最高的是第 31 題；而在創新程度向度的難度最高的則是第 34 題。

將各向度難度的聚類比例，按照第一部分得到的模糊相對權重加權，即得到各題的綜合向度的聚類比例，其綜合向度的難度指標及排序如 9 所示。

由表 9 可看出來，在經過各向度的模糊相對權重加權之後，各題的綜合難度隸屬。進一步可以看出，整體難度最高的是第 33 題，第二名是第 31 題，第三名是第 27 題。比較分向度難度值和綜合難度值，可以發現一些有趣的現象。例如第 28 題在題目敘述和解題策略兩個向度中都是難度最高的，但經過加權之後的難度僅排名第六。而像第 33 題，在各向度的排名分別是第 3、第 5、第 6 和第 2。沒有一個向度是最難，但整體的難度卻是排名第一。

此外，表 9 亦比較兩種常用的試題難度。如文獻探討所述，常用的試題難度有兩種，第一種為試題反應理論下的難度 b 值，另一種則為通過率 P 值。在試題反應理論下的 b 值越大，代表試題越難，然而， P 所代表的是試題通過率， P 值越大，反倒代表試題越簡單，此研究所發展出的模糊加權難度，則類似 b 值的向度，值越大，代表試題越難。表 9 將每一題的模糊加權難度並與學生的作答結果作一比較。學生作答結果是採 98 第一次國中基本學力測驗的實徵數據，隨機取樣 5000 位考生的資料進行分析，並根據這 5000 名考生的資料，計算答對率 P 值與試題反應理論單參數模式所估計的 b 值。難度相對比較見表 9。

表 9 難度指標之綜合比較

題號	模糊指標		P 值	b 值
	加權難度	排名		
1	0.28	34	0.74	-1.02
2	0.36	29	0.76	-1.10
3	0.42	23	0.80	-1.33
4	0.47	17	0.59	-0.38
5	0.31	33	0.78	-1.24
6	0.38	27	0.79	-1.25
7	0.37	28	0.75	-1.09
8	0.47	18	0.68	-0.73
9	0.46	19	0.80	-1.32
10	0.40	25	0.71	-0.87
11	0.39	26	0.76	-1.12
12	0.32	32	0.67	-0.69
13	0.48	16	0.70	-0.83
14	0.32	31	0.67	-0.72
15	0.41	24	0.63	-0.52
16	0.34	30	0.62	-0.48
17	0.44	22	0.65	-0.61
18	0.48	15	0.60	-0.41
19	0.49	13	0.65	-0.64

表 9 (續)

20	0.46	21	0.58	-0.34
21	0.49	11	0.65	-0.63
22	0.49	12	0.64	-0.57
23	0.46	20	0.56	-0.27
24	0.52	8	0.61	-0.46
25	0.58	4	0.57	-0.28
26	0.49	14	0.50	-0.02
27	0.59	3	0.42	0.32
28	0.57	6	0.40	0.37
29	0.51	9	0.37	0.52
30	0.5	10	0.47	0.11
31	0.59	2	0.49	0.02
32	0.57	7	0.42	0.32
33	0.59	1	0.27	0.95
34	0.57	5	0.39	0.42

經由計算，試題答對率與模糊難度指標的相關係數為 $(u)-0.76$ ，達到相當程度的負相關，而與單參數 IRT 所算出的 b 值相關係數為 0.75 ，也達到相當程度的正相關。可見此方法得到的難度指標與傳統常用的試題難度是相當幅合的。

結論與建議

相較於傳統的試題難度評量方法，模糊問卷所取得的數據資料，更能真實反映試題難度的模糊性。將決定難度的因子分成數個不同的向度，可從不同的角度去檢視試題的難度。而將模糊相對權重的概念應用於試題難度評量，能將試題依不同向度的權重作加權，以此得到試題的綜合難度指數，便於比較與分析。與傳統只從單一向度去考慮難度的方法來比較，分多向度去決定難度似乎更具有說服力。

模糊統計為一新的統計技術，在計算上的確較傳統的統計方法來的複雜。但此方法有其優點，尤其能夠改善二元邏輯，非黑即白的計算方式，並加以考量人類的認知的灰色地帶。研究者目前在發展模糊統計軟體，希冀未來的研究者能夠藉由統計軟體來方便進行模糊統計的運算。

用模糊統計進行試題難度運算，還有許多技術層面值得後續的研究，但不失為一種計算試題難度的另類思考方式。而令人感到鼓舞的為在此研究中，雖然受測的教師人數並不多，且主要侷限為台北縣市的中學教師，但依其問卷所估計出的模糊難度的數值，和使用全國基測實徵數據下，所估算的 IRT 難度 b 值與通過率 P 值之相關達到一定水準。代表此方法所得出的難度結果，已達到相當程度的驗證。此外，此研究所提出計算難度的方法，可以提供高風險性考試組題的參考。例如在某些檢定考試或入學考試時，為擔心會有洩題的危險而題目無法進行預試，則可在入闖時，請參與命題的老師，對各向度的難度指標進行評估，來建構試題。在此方法下，只要所參與的老師是具有代表性的，應該所估算的難度也能接近實徵數據的難度指標。

本文提出聚類距離與聚類比例的新定義，取代直接使用模糊平均數或模糊眾數等概念，對所取得的模糊問卷作解釋。研究者發現，這樣的方法具有相當好的解釋能力。而利用反模糊化概念取得的難度指標，則是便於對試題難度作排序及分類。

最後，研究者對未來的研究方向及應用提出幾點建議：

一、本研究所採用的四個決定難度的因子，並非固定不變，本文僅提出分向度評量試題難度的概念，後續研究者可視不同情況需要，採取不同的因子來決定試題的難度，或者可對本文採取的因子，蒐集更多專家學者的意見，取得一致的共識，得到更精準的模糊相對權重值。

二、在本研究中，發現某些題目，使用模糊理論所估計的難度值，和傳統難度 P 與 b 值差異較大，然造成教師認知與學生實際作答難度差異的成因，未來研究值得進行深入的探討。

三、本研究所定義的聚類距離，是針對具有偏好序列性質的語言變數所定義的。而此種特殊的距離在模糊理論上所具有的性質，也值得後續深入研究。

四、本研究所提出的分向度決定難度的概念，以及將試題依照難度指標範圍分類的作法，可應用於命題卡的製作或是題庫系統的建立。

參考文獻

- 王文俊 (1997): **認識 Fuzzy**。台北：全華。
- 江彥聖 (2008): **模糊相關係數及其應用**。國立政治大學應用數學系數學教學碩士在職專班碩士論文。
- 余民寧 (2002): **教育測驗與評量—成就測驗與教學評量**。台北：心理。
- 汪慧瑜 (2005): 模糊統計分析在網路成癮行為的調查應用。**測驗學刊**, 52 (1), 83-104。
- 阮亨中、吳柏林 (2000): **模糊數學與統計應用**。台北：俊傑。
- 吳柏林 (2005): **模糊統計導論：方法與應用**。台北：五南。
- 吳柏林 (2002): **現代統計學**。台北：前程。
- 吳柏林 (1997): 社會科學研究中的模糊邏輯與模糊統計分析。**國立政治大學研究通訊**, 7, 17-38。
- 胡悅倫、陳皎眉、吳柏林 (2006): 應用模糊統計於 A 型量表分類之研究。**教育與心理研究期刊**, 29 (1), 151-181。
- 胡悅倫、吳柏林 (2002): 模糊統計在分析樂觀量表之應用。**教育與心理研究**, 25 (3), 457-484。
- 國民中學學生基本學力測驗推動工作委員會 (2009): **98 年國民中學學生基本學力測驗專輯**。取自國民中學學生基本學力測驗推動工作委員會網站：<http://www.bctest.ntnu.edu.tw/98annuals/index2.html>, 2011 年 5 月 1 日。
- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice-Hall.
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford.
- Polya, G. (1945). *How to solve it: A new aspect of mathematical method*. Princeton, NJ: Princeton University.
- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8, 338-353.

收稿日期：2011 年 05 月 30 日

一稿修訂日期：2011 年 10 月 21 日

二稿修訂日期：2011 年 11 月 02 日

接受刊登日期：2011 年 11 月 02 日

Bulletin of Educational Psychology, 2012, 44(2), 207-228

National Taiwan Normal University, Taipei, Taiwan, R.O.C.

Applying of Fuzzy Statistics in Item Difficulty

Ming-chuan Hsieh

Research Center for Testing and Assessment
National Academy for Educational Research

Ber-lin Wu

Department of Mathematics Science
National Chengchi University

The traditional method of using the Likert scale in questionnaires provides researchers with information of fixed scale choices, which does not allow participants fully express their candid and complex thinking. The current paper applies fuzzy statistics to examine item difficulty of questionnaires. Concepts such as fuzzy mean, fuzzy mode or fuzzy median are applied. Based on the conceptions of fuzzy statistics, we try to define a new distance, which was transformed to obtain an item difficulty index that allows the comparisons between test items. Another focus of this paper is to determine the difficulty level of test items by examining various dimensions of difficulty factors. The weights of the dimensions of the item difficulty were compared and analyzed with the concept of fuzzy relative weight. The fuzzy difficulty index derived from this study shows promising features that is highly correlated with traditional difficulty indices.

KEY WORDS: assessment for test difficulty, fuzzy relative weight, fuzzy statistics