

本文章已註冊DOI數位物件識別碼

► 智慧型線上測驗系統題型之分析與改進

Analysis and Improvement of Test Items for a Network-based Intelligent Testing System

doi:10.6173/CJSE.2002.1004.05

科學教育學刊, 10(4), 2002

Chinese Journal of Science Education, 10(4), 2002

作者/Author：黃國禎(Gwo-Jen Hwang);曾秋蓉(Judy C. R. Tseng);朱蕙君(Carol Chu);蕭經武(Jing-Wu Shiau)

頁數/Page：423-439

出版日期/Publication Date：2002/12

引用本篇文獻時，請提供DOI資訊，並透過DOI永久網址取得最正確的書目資訊。

To cite this Article, please include the DOI name in your reference data.

請使用本篇文獻DOI永久網址進行連結:

To link to this Article:

<http://dx.doi.org/10.6173/CJSE.2002.1004.05>



DOI Enhanced

DOI是數位物件識別碼（Digital Object Identifier, DOI）的簡稱，是這篇文章在網路上的唯一識別碼，用於永久連結及引用該篇文章。

若想得知更多DOI使用資訊，

請參考 <http://doi.airiti.com>

For more information,

Please see: <http://doi.airiti.com>

請往下捲動至下一頁，開始閱讀本篇文獻

PLEASE SCROLL DOWN FOR ARTICLE



智慧型線上測驗系統題型之分析與改進

黃國禎¹ 曾秋蓉² 朱蕙君¹ 蕭經武¹

¹暨南國際大學 資訊管理系

²中華大學 資訊工程學系

(投稿日期：90年8月10日，修訂日期：91年4月9日，接受日期：91年8月22日)

摘要：隨著電腦技術的進步以及網際網路的普及，遠距教學已受到社會大眾的重視。而伴隨教學系統的發展，網路測驗也逐漸扮演重要的角色。為了快速建立測驗所需的題庫，透過網路由多位學者進行合作式出題是一個很有效率的方式。然而，合作式出題在未經協調的情況下，極易產生題目重複、偏重於某些概念或同一試卷中某題目洩漏其他題目答案等情況發生，使得測驗成效大受影響。過去一些相關研究雖然解決了部分的問題，但在準確性方面仍然有待改進。在本論文中，我們嘗試以題意分析架構來改進網路合作式出題的程序，以提高命題品質；同時，為了增加重複性以及一致性檢查的準確度，我們歸納出一些特殊題型，並以高中生物題庫進行測試來驗證新方法的準確性。由實驗的結果發現，使用題意分析架構及特殊題型確實可以提高分析及管理網路命題程序的準確性。

關鍵詞：網路測驗、題庫、重複性分析、一致性分析、遠距學習。

壹、緒論

隨著電腦設備的進步及網際網路的發達，人們獲取知識的管道也產生重大的改變，從傳統的課堂、電視、電影、廣播以及報章雜誌逐漸轉移到網路上。由於遠距教學可節省大量人力，又能突破時間及空間的限制，因此逐漸受到廣泛的重視（莊雅蓁，1999）。遠距教學的型態因播送的方法不同，可區分為「即時群播」與「課程隨選」兩種。「即時群播」類似於視訊會議，採用主播教室傳送到遠端教室的一對

多方式，即教授在主播教室授課，而學生在各個遠端教室聽課。每一間教室都架有錄影器材和麥克風，授課實況經由高速網路傳送，師生們可透過攝影機和麥克風進行授課及問答，我國NII計畫中於1995年7月14日啟用的「遠距教學先導系統」即屬於此類的應用。

「課程隨選」則是先在網路上建立教學平台，將教材儲存於資料庫中。學生不必在固定時間或地點就可透過全球資訊網取得教學資訊、繳交作業或進行測驗。交通大學周倩和簡榮宏（1997）的「遠距合作學習」(Cooperative Remotely Accessible Learning, 簡稱 CORAL)



計畫及美國西北大學的「合作視覺化」(Collaborative Visualization, 簡稱 CoVis)計畫都是著名的「課程隨選」實例。

由於在遠距教學系統中，老師與學生的互動不若一般課堂中頻繁，必須透過教學評量以瞭解學生的學習狀況。Tyler (1950) 指出，過去教學評量一直在各科教學活動中扮演極重要的角色，而測驗為教學評量最主要的工具。何榮桂、蘇建誠和郭再興 (1996) 在研究報告也顯示，如何提供一個豐富完整的題庫及有效率的出題方式，已成為遠距教學中一個重要的課題。

由於數篇研究 (蔡松齡, 1992a, 1992b) 指出，題庫的建立與維護是一項耗費人力的工作；因此，黃國禎、祝鈞毅、許慶昇、杜淑芬和葉士毅 (1997) 認為由多位教育專家或教師透過網路出題已成為未來的趨勢。然而眾專家在未經協調的情形下，出題方向有時無法包含完整的測驗重點，也可能在同一份試卷中出現相同或近似的題目，甚至有某一題的題目包含了另一題的答案等情形，造成命題品質的下降，並使測驗的準確性大打折扣。為因應這些問題，國內學者開始在中文環境下進行研究工作，黃國禎 (2001) 的研究中也試圖提供解決之道，以降低問題發生的機率並能兼顧學習過程中的各項重點。在本論文中，我們透過題意分析的方式來改進試題分析與比對的效果，期望能夠讓題庫在建構時更加符合學習評量的精神，同時也減輕出題者的負擔。除了提出改進的方法外，最後將以實驗的方式來證明我們所提出的方法具有較高的準確度。

貳、相關研究及背景

隨著電腦輔助教學應用的日趨廣泛，藉由電腦來進行測驗已是時代的趨勢 (Dennis & Gruner, 1992)。美國著名的 GRE 測驗已在 1999

年全面廢止現行紙筆測驗方式，並改用電腦進行考試。這些運用電腦進行測驗的系統，近年來快速的成長，並在各種測驗中扮演極重要的角色。

由於傳統電腦輔助教學系統大部份只是作為教材呈現的媒體，與學生的互動及提供個人化學習環境等方面較為欠缺。因此，學者們提出智慧型電腦輔助教學的構想以彌補傳統電腦輔助教學的不足。例如張百慈 (1996) 的研究中所提出的“以卡瑞爾學習模組為基礎建立智慧型電腦輔助教學系統之架構”，試圖以卡瑞爾提出之影響學習成效的五個主因做為基礎，建立一套智慧型電腦輔助教學之基礎架構，並以引導出每位學生的潛力為目標。紐西蘭的研究人員 Fan, Mak 和 Shue (1996) 也提出一套“Knowledge Based Computer Assisted Instruction System”，為避免學生藉由累積考試經驗來背誦固定的答案，因此該測驗系統會在進行測驗時會自動變換題目中數字的部份。由於多數的智慧型電腦輔助教學系統為單機作業，無法對一群學生的測驗作有效的管理與分析；因此，隨著網際網路的發展，智慧型遠距教學系統便成為研究發展的重點。我國政府自 1994 年開始推展全國資訊基礎建設 (NII)，如楊叔卿 (1998) 的研究中認為網路教育的推展由大學而至中小學是既定的政策及方向。在美國大學中多數也設有網路教學系統，其中以加州 UCLA 大學之 Homepage Education Network (<http://www.then.com>) 最受矚目。

在國內，由國科會科學教育處推動的整合型計畫，ITED (Intelligent Tutoring, Evaluation and Diagnosis)，是 CORAL 的延續分支計畫，自民國 88 年至 93 年共分為二個階段各 3 年進行。該計畫成員包括暨南國際大學資管系、教育政策所、比較教育所、中華大學資訊工程系及嶺東技術學院資管系的教授及研究



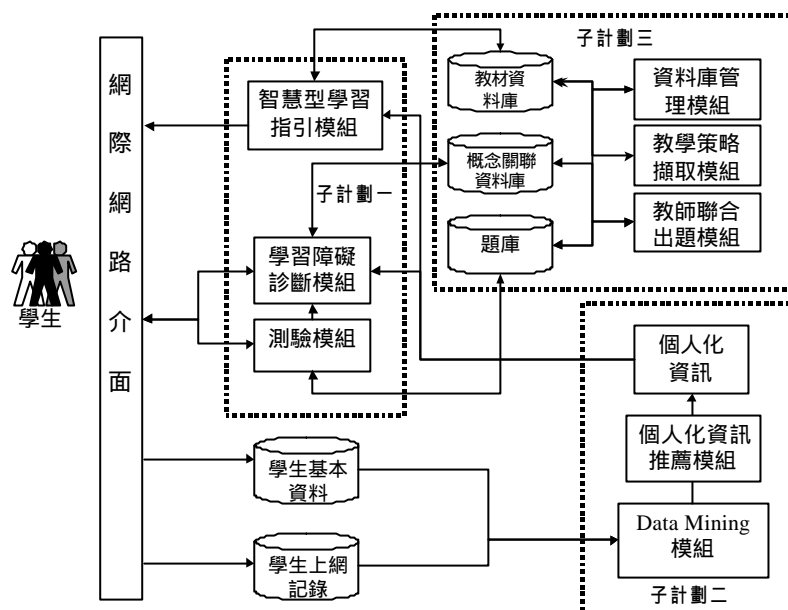


圖 2.1：ITED 系統架構圖

生。計劃的目的在於將專家系統、模糊理論、資料庫系統及個人化技術應用於解決在網路學習環境中進行科學教育相關課程可能產生的學習、測驗與診斷的問題；同時，更提出一套整合的模式，來建立智慧型個人化網路學習、測驗與診斷服務平台，其構想如圖 2.1 所示，共分為三個子計劃：子計劃一「智慧型學習測驗、診斷與導引模組」主要運用學生測驗的表現及課程中概念架構與試題的關係找尋學生的學習障礙，以提高提高學習成效；子計劃二「個人化學習資訊蒐集、分析與推薦模組」嘗試針對不同能力與典型的學生族群來設計教材，而不是只針對一般的學生；子計劃三「多專家教學策略庫分析、協調與管理模組」主要在於運用資訊技術，有效擷取與整合眾多教育專家的教學策略及素材，其中線上試題的擷取與整合為首要的工作。由於科學教育的範疇中，教學單元的主題與目標（重點概念）較為明確，因此，以系統化的方式進行分析與處理較人文、社會

等其他學門更為可行。本論文所提出的為 ITED 子計劃三的研究成果，主要是針對科學教育相關學科在線上共同出題時可能發生的問題進行分析，以確保題庫的品質及測驗的準確性。

在產生試卷的選題過程中，系統會根據使用者要求試題的難易度、鑑別度等參數，動態的配置出一份適當的試題，並且對測驗結果給予評估與建議。而其中直接影響題庫品質的，即為題庫建構的過程，包括對試題進行重複性、一致性及完整性分析，以下將針對這些概念作詳細的說明。

一、重複性分析

「重複性」是指不同教師所出的題目可能有題意相同的現象，因而降低測驗的準確性。為了避免這種情形，黃國禎（2001）提出對題目進行檢查與篩選的演算法，圖 2.2 中為其演算流程的步驟圖示，以下將對每一步驟作說明。

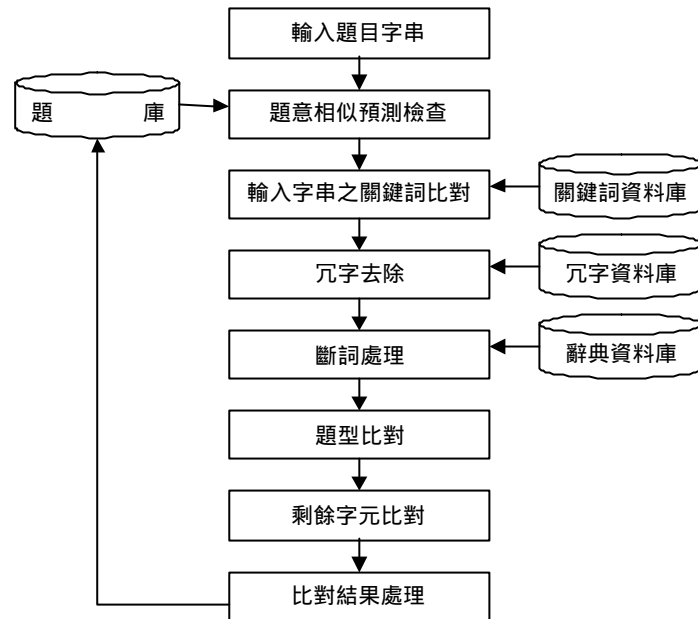


圖 2.2：重複性檢查流程圖

(一)題意相似預測檢查

所謂相似或相同的題目，除了題目的內容類似之外，所包含的各章節測驗重點（或學習關鍵詞）也必定十分相近。為了節省系統在斷字及進行字串搜尋比對時所花的時間，可以先利用題意相似預測檢查來篩選出題意可能相似的題型，再進行比對。

在定義題目的測驗重點時，蕭漢威（1994）提出評估平衡表（Evaluation Balance Table）來表達題目與重點概念的關係值，即 e_{ij} 。 e_{ij} 的值代表以下的意義：

- 5：該題目對於測驗出此重點概念的學習成果非常重要
- 4：該題目對於測驗出此重點概念的學習成果具重要性
- 3：該題目對於測驗出此重點概念的學習成果略微重要
- 2：該題目對於測驗出此重點概念的學習成果有相關性

1：該題目對於測驗出此重點概念的學習成果有少許關聯

0：該題目對於測驗出此重點概念的學習成果無任何關聯

表 2.1 為每個題目與概念之間的關係，此關係可以一個概念向量 Q_i 來表示：

$$Q_i = (P_{i1}, P_{i2}, \dots, P_{im})$$

對於輸入題目之概念向量 Q_{input} 與題庫內題目之概念向量 Q_j ，可以取其概念值間的差之總合作平均：

$$\text{error_distance}(Q_{input}, Q_j) = \sum_{r=1}^m |P_{input\ r} - P_{jr}| \div m$$

其中 m 為重點概念的個數，而誤差距離（error_distance）則表示這兩題所包含重點概念的差距，誤差距離越小表示兩題目間所包含的重點觀念越相近，同時也越有可能產生題目重複的情形；反之，若兩題目所包含重點觀念的誤差距離越大，代表其測驗的內容越不相同，故不需要作進一步的內容比對。這樣的篩選方式可增進系統的效率。



表 2.1：建立題目與概念關係圖（評估平衡表）

	概念 1 (K_1)	概念 2 (K_2)	概念 m
Q_1	P_{11}	P_{12}	P_{1m}
Q_2	P_{21}	P_{22}	P_{2m}
\vdots	\vdots	\vdots		\vdots
Q_i	P_{i1}	P_{i2}	P_{im}

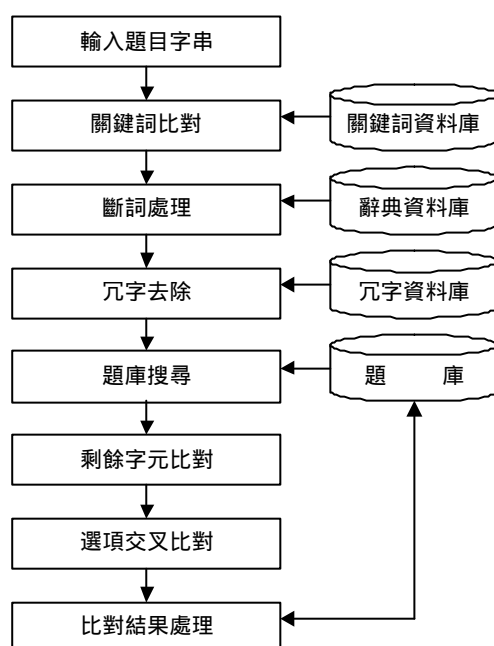


圖 2.3：一致性檢查流程

(二) 關鍵詞比對、冗字去除與斷詞處理

由於課程內容有事先建置的關鍵詞彙，所以可將題目與關鍵詞庫先作比對，以縮小斷詞時所需要的搜尋範圍，並減少斷詞時可能發生錯誤的機率。此外，為了節省斷詞所花費的時間，可將題目中較常出現的一些冗字（不具特殊意義的詞），如：“請”、“問”、“的”、“有”、“個”、“和”、“，”、“？”、“。”等去除，以增進效率。

(三) 題型比對和結果處理

經過以上步驟之後，將會產生三個字串的

集合，分別是關鍵詞字串、冗字字串、和剩餘字串。將這些字串來和原來題庫中已有的題目對應的字串比對，可求出相似百分比為

$$\left(\frac{\text{與題庫題目比對後匹配到的字數}}{\text{原輸入字串總字數}} \right) \times 100\%$$

最後設定一門檻值，決定相似百分比的下限，以作為是否將題目視為「重複」出題的參考依據，只要超過此一門檻，則系統會將這兩個題目列出，供專家參考是否加入題庫或設為互斥。



二、一致性分析

一致性的檢查是為了要避免在一份試題中某題的答案出現在另一題的題目中，而必須在出題時所做的一些控制與篩選。由於這類題目本身的相似性並不一定很高，所以可將相似百分比的門檻值略微調低，最後再針對選出的題目進行選項與題目的交叉比對，檢查是否有答案出現在題目中。若有，則提出警告，並給出題老師作為是否將此兩題設為互斥之參考，其作法及流程如圖 2.3 所示。

三、完整性分析

為了避免題庫的問題都偏重於某些方面，黃國禎（2001）利用評估平衡表作出各重點題數百分比統計圖，以提供教師作為檢查題庫是否重點分佈不均的參考依據。此外，對於題數離平均 3 個標準差之外的重點，也將特別列出給專家參考，亦即

If $(\mu \pm K_j > 1.5 \times S)$

Then { Println(K_j); }

其中 μ 為平均題數、 K_j 為重點題數且

$$S = \sqrt{\sum (x - \bar{x})^2 / N - 1}$$

叁、以語意為基礎的題庫分析策略

雖然上述的方法可以改進題庫在檢驗重複性、一致性及完整性方面可能造成的問題，但仍然有些無法處理的狀況，例如下面兩個題目：

Q1. 請問引起白血病的主要原因是？

Q2. 什麼是造成的血癌重要因素？

由題意看來，我們可以很明顯的看出這是兩個完全相同的題目，但是由先前所提出的方

法來進行重複性比對，卻可以發現：

Q1. 關鍵詞：白血病

一般詞：引起、主要、原因

贅詞：請、問、的、是、？

Q2. 關鍵詞：血癌

一般詞：造成、重要、因素

贅詞：什、麼、是、的、？

代入公式後所得相似度：

$$3 (\text{匹配到的字數}) \div 13 (\text{原輸入字串字數}) \times 100\% = 23.08\%$$

如此低的相似度，顯然不會被檢查出來。其中一個明顯的問題是出在同義詞的部份。其實「血癌」和「白血病」的意義是相同的，「主要」和「重要」兩個詞的相似度很高，「原因」和「因素」可以說是同義詞，但是在之前的研究中並沒有考慮相似詞的關聯性，因此影響了比對的準確度。以下我們將針對可能的改進方式進行探討。

一、同義詞的運用

過去國內已有一些學者（莊雅蓁, 1999; 江玉婷, 1999; 吳忻萍, 1996; 陳佳君, 1995; 黃卓倫, 1996; 國立中央圖書館中文主題用語表研究小組, 1988）在同義詞典對資訊檢索之輔助效益及相關技術進行研究，其中在應用上大多偏重在圖書資訊的處理，包括同義詞典的建構、查詢問句擴展模型的建立、同義詞典及索引典整合應用等。莊雅蓁（1999）指出，經由過去的應用結果可以確定，以同義詞典進行資料比對可獲得良好的效益。

為了解決這個問題，我們在系統中建立關鍵詞與一般詞的概念關聯表格（Conceptual Relation Table），來表達詞與詞之間的關聯性。一般詞之間的關聯性是固定的，在找尋專家並建立表格數據後便不會再更改；而關鍵詞之間的關聯性則可能因課程的不同會有所變



表 3.1：關鍵詞概念關聯表格

	白血病	血友病	血癌
:		:	:	:	:
白血病	:		0	5	:
血友病	0		0	:
血 癌	5	0		:
:	

表 3.2：一般詞概念關聯表格

	引起	因素	造成	主要	重要	原因
:		:	:	:	:	:	:	:
引起		2	5	0	0	2	:
因素	2		2	0	0	5	:
造成	5	2		0	0	2	:
主要	0	0	0		4	0	:
重要	0	0	0	4		0	:
原因	2	5	2	0	0		:
:	

動，所以可由專家來更改其相似度。我們將關聯性分為六個階級，如表 3.1 及 3.2 所示。其中數字 1~5 所代表的意義是：5 代表「完全相同」（同義詞）、4 代表「意義非常接近」、3 代表「意義接近」、2 代表「有一些關聯」、1 代表「只有很小的關聯」、0 代表「完全沒有關聯」，若轉換成百分比數字，則可以由 100%、80%、60%、40%、20%、0% 來表示。

以前面所舉 Q_1 與 Q_2 為例，若引用表 3.2，可得「引起」與「造成」的關聯性為 100%、「主要」與「重要」的關聯性為 80%、「原因」與「因素」的關聯性為 100% 而「血癌」又等於「白血病」；於是 Q_1 與 Q_2 的相似度經計算之後成為：

$$(1 \times 2 + 2 \times 1 + 1 \times 0.8) \div (1 \times 2 + 3 \times 1) \times 100\% = 96\%$$

很顯然引用表 3.2 所獲得的結果比原估算的

23.08% 合理很多；同時，因 96% 高過門檻值（50%），因此 Q_1 與 Q_2 將被列為可能重複的題目。

二、關鍵詞與贅詞之加權處理

在黃國禎（2001）的研究中所提的計算門檻值公式裏，冗字也被列為比較的對象，且其重要性和關鍵詞與一般詞相同，在很多情況下，這並不是很合理的處理方式。所謂「冗字」，顧名思義就是不會對題意造成決定性影響的字與符號，如介係詞、標點符號和一般題目常見的詞語等（例：請問、何者.....）。此外，從另一方面來看，除了關鍵詞以外的一般詞以及冗字即使相似的程度再高，卻並不保證題目的題意會相同，我們可以看看以下的例子：

Q3. 關於腎上腺的敘述何者正確？



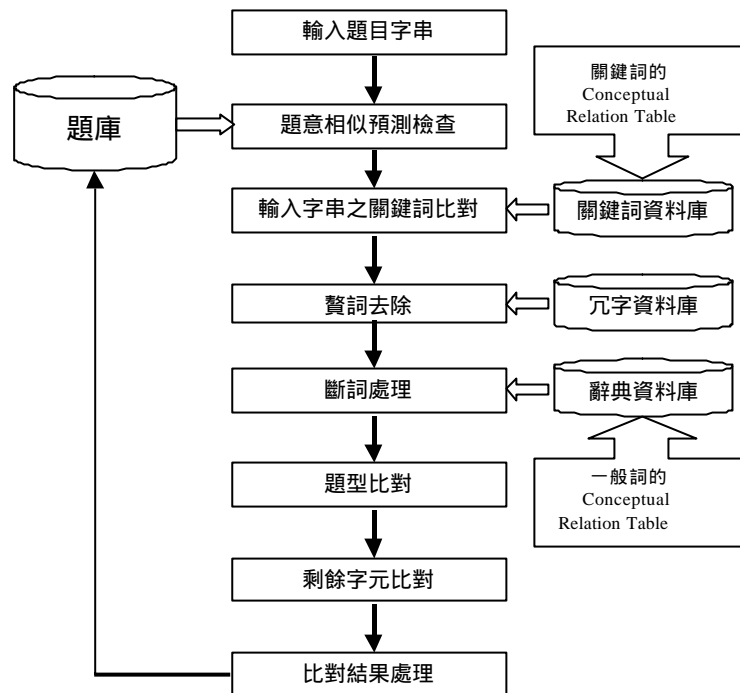


圖 3.1：新的重複性檢查流程圖

Q4. 關於腦垂腺的敘述何者正確？

雖然兩題除了關鍵詞之外的敘述都完全相同，但是問題的重點卻完全不一樣。由此處我們可以清楚的瞭解到，影響一個題目意義最主要的是關鍵詞彙，再來才是一般詞，而冗字的多寡不應造成影響；因此，我們決定採用加權的方式來做處理，可以寫成以下的公式：

$$=C_1:C_2:C_3$$

其中 $C_1 > C_2 > C_3$ 。（例如說我們可以取 $C_1 = 2$ 、 $C_2 = 1$ 、 $C_3 = 0$ ）

注意我們採用「詞」而非「字」為單位，這樣在分析上才具有價值，因為中文的詞具有其結合性，一旦被拆開來，就會失去原來的含意。採用以詞為主的方式，才可表現以「題意」作為重複性比較的精神。然而當一個詞與另一題目中的許多詞都有關聯時，我們將只取關聯性最高的詞與其匹配，可得改進後的流程如圖

3.1 所示，其相似百分比為：

$$\text{Similarity}(Q_i, Q_j) = \left(\sum_r (M_r(Q_i, Q_j) \times W_r) \right) \div \text{Max}(N(Q_i), N(Q_j)) \times 100\%$$

其中 $N(Q_i)$ 與 $N(Q_j)$ 代表 Q_i 和 Q_j 包含的詞數。 $M_r(Q_i, Q_j) \times W_r$ 代表第 r 個 Q_i 和 Q_j 共有的詞， W_r 代表其詞性，若為關鍵詞，則 $W_r = C_1$ ；若為一般詞，則 $W_r = C_2$ ；若為贅詞或冗字，則 $W_r = C_3$ 。

三、特殊題型的應用

在一致性的分析方面，我們必須先了解會造成「某一題的答案出現在另一題的題目中」的原因有哪些。其中一個可能，是答案在另一題的題目中，如：

Q5. 請問人類是屬於哪一類？

(1) 哺乳類 (2)

Q6. 除了人類是哺乳類之外，下列哪一個選項中的動物也是屬於哺乳類？



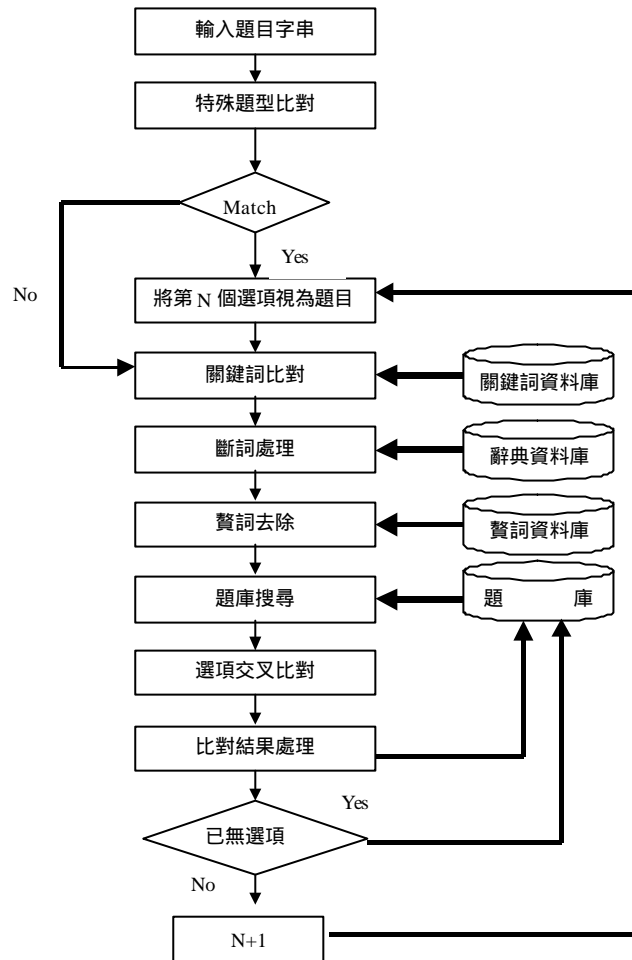


圖 3.2：新的一致性檢查流程圖

這一類型的題目出現的機率較低，但由於題目的相似性不一定很高，只能先降低相似度的門檻值，再對選項作交叉比對，找出是否有在彼此的題目出現，作為專家是否將其設為互斥或重新出題的參考。

而另一種可能，則是答案在另一題的答案中，如：

Q7. 下列敘述有關遺傳的敘述，哪一項是正確的？

- (1) 血友病和色盲均是性聯遺傳
- (2)

Q8. 血友病和色盲是屬於

(1) 基因突變 (2) 性聯遺傳 (3) ...

這種情形出現的機率相當的高；然而，這類型題目之間通常實際計算後獲得的相似度是非常低的。以上述的 Q7 為例，真正的題目重點是在選項當中，若是只比較「純題目部分」，很可能因為相似度太低而沒有發現潛在的一致性問題。

幸好這類的試題大都有固定的題型，如：「試問以下關於 的敘述，哪一項是正確的」或者如「下列敘述何者為非」.....等，我們便可以利用此一特性，將符合此類題型的題目另作處理，把它的每一個選項都視為

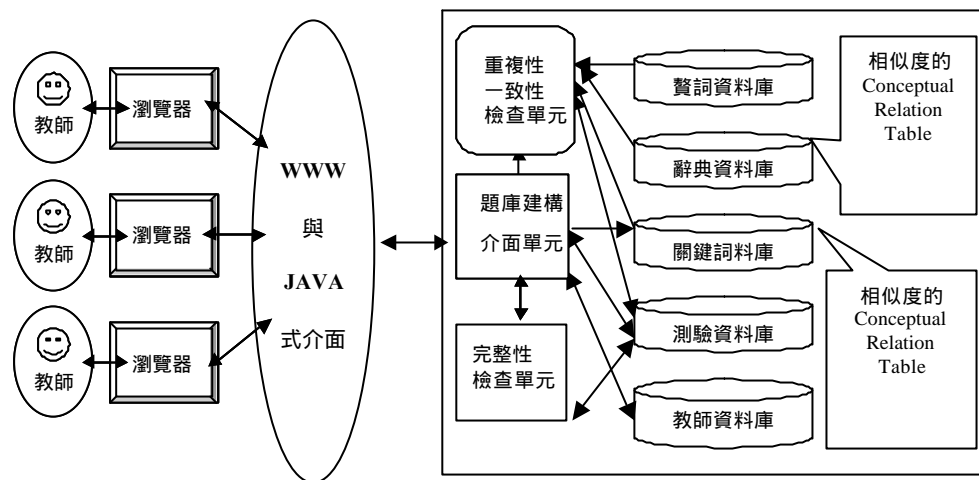


圖 4.1：網路教學智慧型測驗資料庫建構系統架構圖

一個完整的題目來和其他題目作比對，這樣雖然要花費較長的時間，但是卻可以有效判斷題目與題目之間是否存在一致性的問題。修改之後的一致性檢驗流程如圖 3.2 所示。

對特殊題型而言，在的比對過程中，由於關鍵詞對題型沒有影響；因此，關鍵詞的出現與否反而變得不是很重要，也沒有加權計算的問題，這和一般題型的比對有很大的不同。另外，特殊題型的比對還必需包括選項內容的比對。

肆、系統開發與測試

我們提出的系統架構如圖 4.1 所示。出題的學者透過瀏覽器註冊之後，可經由題庫建構之介面進行題目新增、修改及刪除等功能，也可以瀏覽試題、查詢試題的重點分佈狀況或進討論區發表意見。在學者出題時，系統將自動入對該試題作重複性、一致性與完整性的檢查，來確認新增此題目至題庫內是否會造成問題；同時也在有問題時提醒出題者是否有將衝突的題目設為互斥題的必要（即未來在配題時不會出現在同一試卷）。

贅詞資料庫內存放題目中常見之冗字，其作用除了可以減少斷詞的時間，亦可增加比對的正確率；辭典資料庫內存放一般常見詞語，供斷詞時使用；關鍵詞資料庫是專家在建構教材時所建立的，存放測驗範圍內的重點概念（即關鍵詞）；測驗資料庫記錄題目的基本資料，如出題者之帳號、難易度、鑑別度、互斥題號等，供出題時之參考；教師資料庫則存放了教師的基本資料和出過的題數。

我們以 Virtual Basic 作為系統的建構工具，透過 CGI 與 Access 資料庫作溝通，在 Windows NT 的環境中開發系統，並選擇高中生物第 24 章 27 章來作為此次測試所使用的題庫，測試總題數達 362 題。我們使用黃國禎（2001）的論文中所提的方法（以下簡稱為“舊方式”）與本文中所提的方法（以下簡稱為“新方式”）進行比較測試，其結果將在以下說明。

一、重複性的比較與分析

我們以高中生物第 24 章 27 章題庫中的單選題與複選題之重複性進行分析，在出題

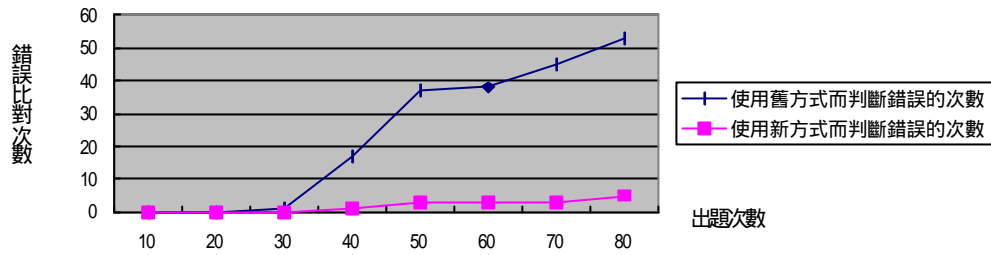


圖 4.2：重複性-單選題-高三生物第 24 章

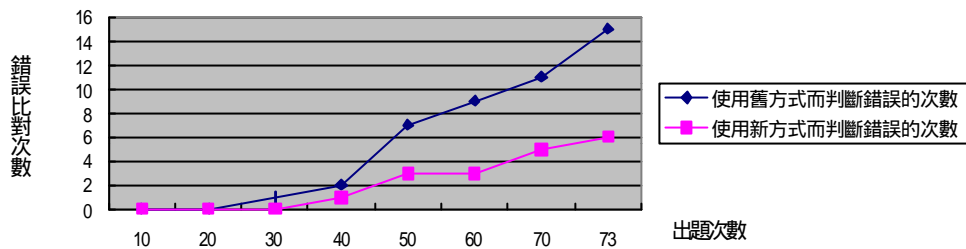


圖 4.3：重複性-單選題-高三生物第 25 章

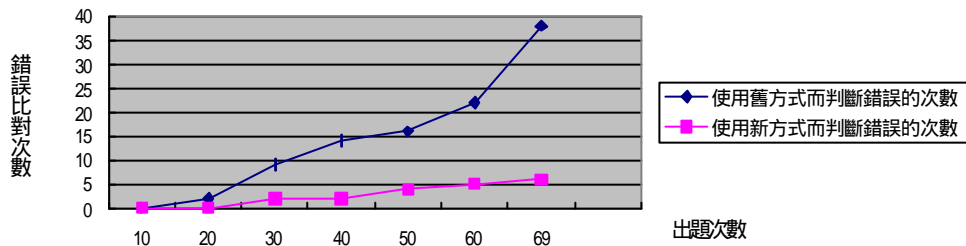


圖 4.4：重複性-單選題-高三生物第 26 章

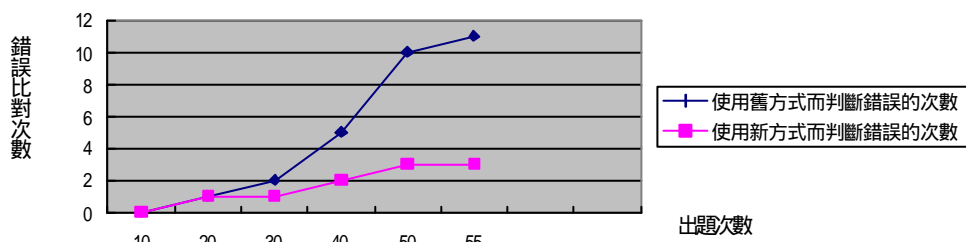


圖 4.5：重複性-單選題-高三生物第 27 章

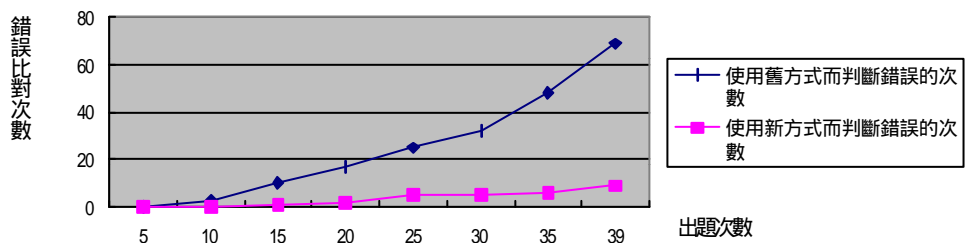


圖 4.6：重複性-複選題-高三生物第 24 章

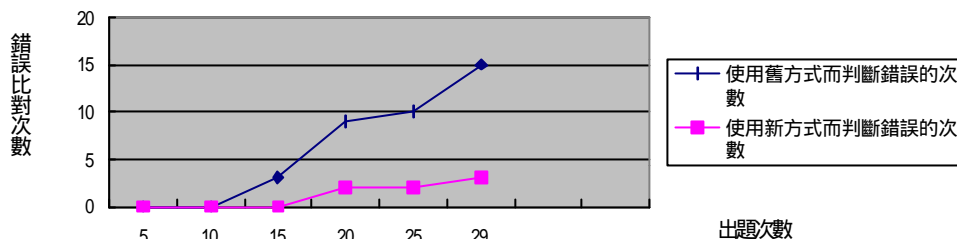


圖 4.7：重複性-複選題-高三生物第 25 章

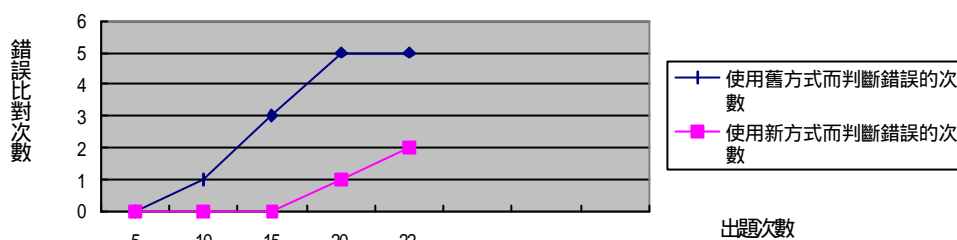


圖 4.8：重複性-複選題-高三生物第 26 章

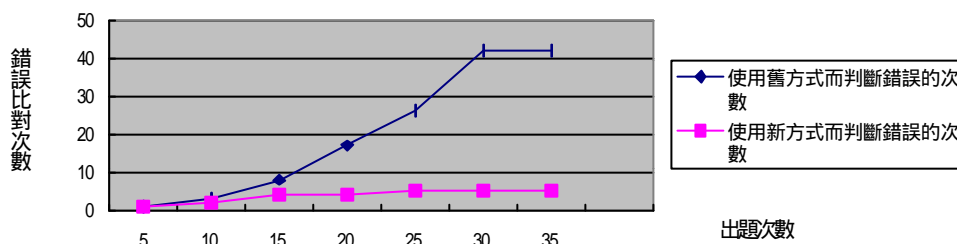


圖 4.9：重複性-複選題-高三生物第 27 章

時若確定題目重複，即不加入題庫中，其實驗結果分別如圖 4.2 4.9 所示。其中橫軸代表在實驗中教師出的題數，而縱軸代表系統在比對過程中錯誤判斷的次數。由於在教師每出一題時，系統將該題與題庫中所有可能相關的題目進行比對，因此每出一題在比對過程中產生的錯誤可能不只一次。從圖 4.2 圖 4.9 的實驗結果可得到以下的結論：

1. 以本文中所提的“新方式”出題時，題目發生重複性錯誤判斷的機率均比“舊方式”少，相對的準確度也較高，可見同義詞典的使用確實有發揮功效。
2. 在高中生物第 24、26、27 章中，由於『下列關於 的敘述，何者(不)

正確？』的題型出現的非常頻繁，所以在特殊題型比對時找出較多重複的題目，由此可以瞭解特殊題型的模式分析確有其必要性。

3. 錯誤的發生頻率與單選或複選並無顯著關係。

二、一致性的比較與分析

以高中生物第 24 章 27 章題庫中的單選題與複選題進行一致性的分析比較，結果分別如圖 4.10 4.17 所示。由實驗結果可以得到以下的結論：

4. 以本文中所提的“新方式”出題時，題目發生一致性錯誤判斷的機率比“舊

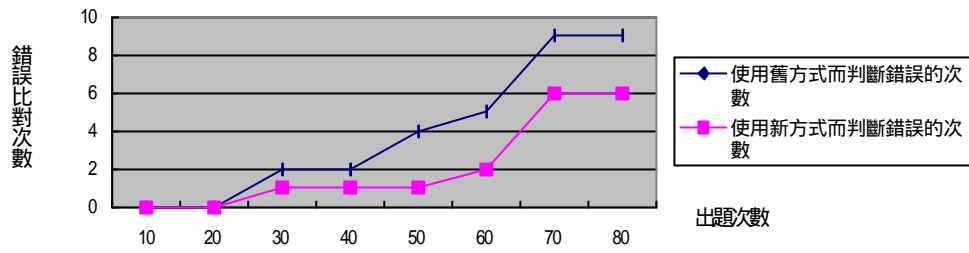


圖 4.10：一致性-單選題-高三生物第 24 章

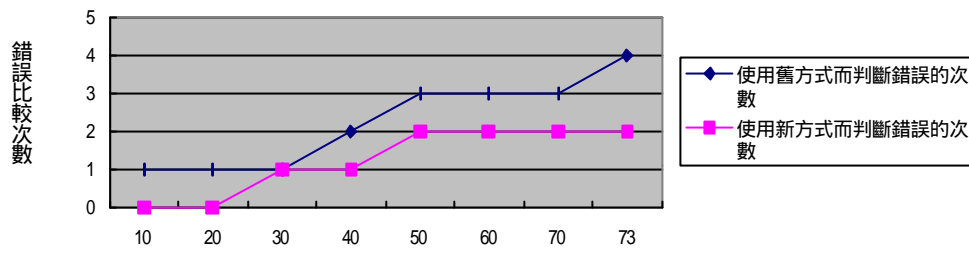


圖 4.11：一致性-單選題-高三生物第 25 章

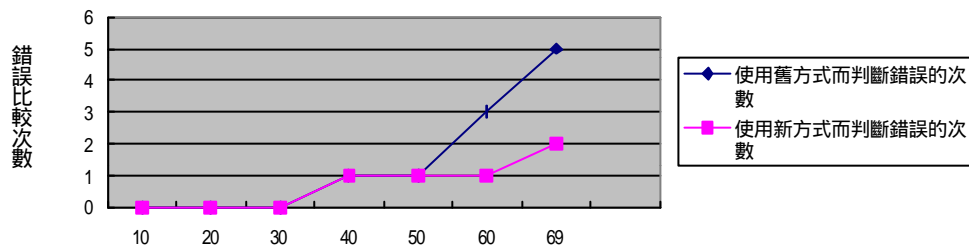


圖 4.12：一致性-單選題-高三生物第 26 章

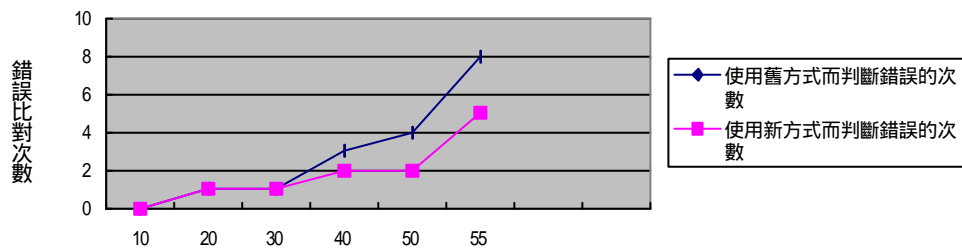


圖 4.13：一致性-單選題-高三生物第 27 章

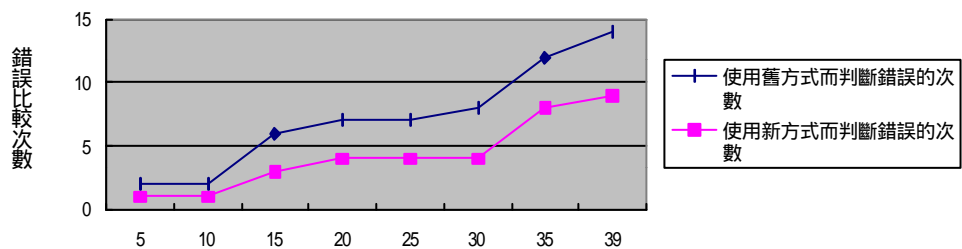


圖 4.14：一致性-複選題-高三生物第 24 章

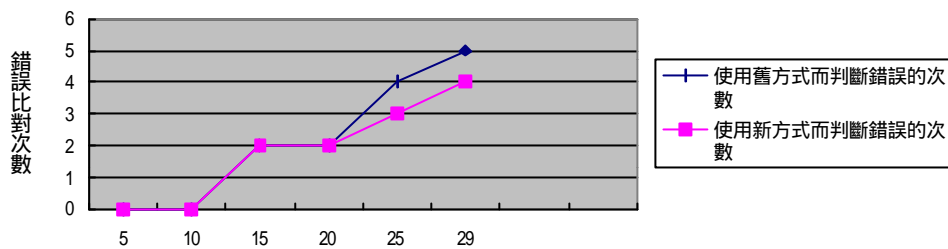


圖 4.15：一致性-複選題-高三生物第 25 章

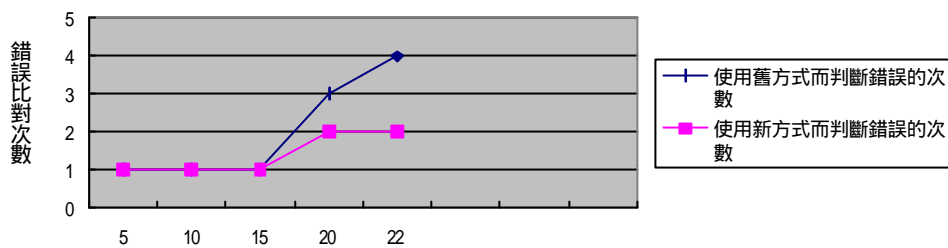


圖 4.16：一致性-複選題-高三生物第 26 章

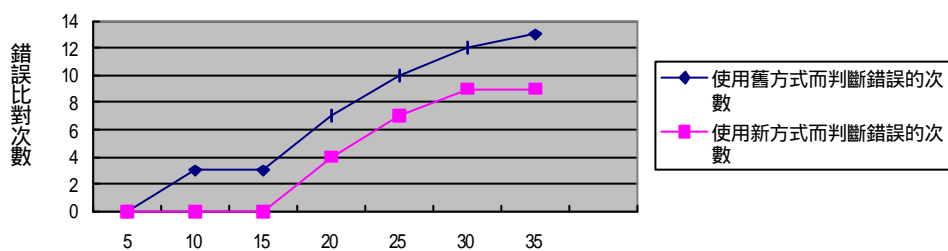


圖 4.17：一致性-複選題-高三生物第 27 章

方式”少，相對的準確度也較高，可見同義詞典的使用確實有發揮功效。

5. 於課程重點範圍越小的章節，檢測出違反一致性的能力越高。
6. 在交叉比對（題目與選項比對）的過程中，由於選項內容通常字數較少，故使用同義詞典的效果並不如作重複性檢查時那麼明顯。

三、完整性的比較與分析

由於完整性的定義，是希望題庫中能包含不同測驗重點的試題，因此，若忽略同義詞的存在，將導致部分重點所包含的題數較多。假設系統的設定是每個測驗重點出 10 題，若未

使用同義詞典，則與「陽光」和「日光」相關的題目將各有 10 題，雖然以完整性的角度而言不是很嚴重，但在出考卷配題時將有嚴重的影響。

伍、結論與討論

學習與測驗評量系統除了教學策略、出題策略之外，題庫的建構與管理是影響測驗測驗品質最主要的因素。為了改善網路多人合作出題時經常發生的現象，包括重複出題、某題目出現另一題的答案，以及在題庫中欠缺某些測驗重點的題目等問題，我們導入了以題意為主的分析比對方式，以及常用題型

結構的概念來強化比對的準確性。由實驗結果顯示，我們所提出的方法，確實在多人合作出題過程中，更精確的找出可能的問題，以確保題庫與測驗的品質。

在系統研發過程中，除了提供多人合作出題的介面並加強比對的準確性之外，有一些系統在使用上的限制尚待解決：

(一)由於在檢驗題目相似性的過程中，關鍵詞的運用相當重要，因此本研究所採用的方法對於科學教育，如自然、數學、物理、化學、生物等學科，較為合適；而對其他如人文、社會或重點不顯著的科目將有適用上的困難。

(二)由於系統需要使用同義詞典以增強字詞比對的準確性，在不同的學科發題庫展過程中，必須建置專屬的同義詞典，因而可能增加教師的負擔。

(三)當有些教師希望以不同角度詮釋同一個問題時，系統必須接受題庫中有相同題目的可能性；因此，題庫中應註明那些是同義題，並且在配題時應有特別的機制來避免相同題目出在同一份考卷中。

未來在合作出題的規模擴大，且題庫資料到達一定量的情況下，可考慮引用國內學者（莊雅蓁, 1999; 江玉婷, 1999; 吳忻萍, 1996; 陳佳君, 1995）研究中指出的自動化索引的技術，以減少教師的負擔。同時，我們也正努力研究適合的配題機制，以配合題庫的使用，提昇測驗品質。

致 謝

感謝審查委員對本研究所提出的改進建議。

本研究由中華民國國科會補助，計劃編號：NSC-90-2520-S-260-002。

參考文獻

1. 江玉婷（1999）：中文資訊檢索測試集設計與製作之研究。台北市：國立台灣大學圖書資訊學系研究所碩士論文（未出版）。
2. 何榮桂、蘇建誠和郭再興（1996）：遠距適性測驗系統架構。資訊與教育學刊, 42, 29-35。
3. 吳忻萍（1996）：以隱藏語意索引為基礎的中文資訊檢索。台北市：國立台灣大學資訊管理學系研究所碩士論文（未出版）。
4. 周倩和簡榮宏（1997, 6月）：網路評量系統之發展與研究。遠距教育學刊, 4, 12-15。
5. 莊雅蓁（1999）。中文查詢問句擴展之研究。台北市：國立台灣大學圖書資訊學系研究所碩士論文（未出版）。
6. 陳佳君（1995）。檢索詞彙來源與檢索詞彙效益之研究。台北市：國立台灣大學圖書館學系研究所碩士論文（未出版）。
7. 黃卓倫（1996）。利用隱藏語意索引進行文件分段檢索之研究。台北市：國立台灣大學資訊管理學系研究所碩士論文（未出版）。
8. 黃國禎、祝鈞毅、許慶昇、杜淑芬和葉士毅（1997, 12月）：遠距學習環境中智慧型學習測驗及評估系統之研製。論文發表於中華民國八十六年全國計算機會議論文集，第二冊，D-129-D-134頁。台中市：私立東海大學。
9. 黃國禎（2001, 8月）：網路智慧型測驗系統題庫重複性、一致性及完整性分析。科學教育研究期刊, 1, 57-85。
10. 國立中央圖書館中文主題用語表研究小組（1988, 6月）：中文主題用語表之研究。中國圖書館學會會報, 42, 21-30。
11. 張百慈（1996）：以卡瑞爾學習模組為基礎建立智慧型電腦輔助教學之架構。資訊與教育學刊, 52。



12. 楊叔卿 (1998, 3月): 電腦網路推展至中小學: 一個超媒體網路師生培訓的研究初探。論文發表於第七屆國際電腦輔助教學研討會論文集, 476-483。高雄市: 國立高雄師範大學。
13. 蔡松齡 (1992a): 電腦在題庫命題作業之運用(上)。師友月刊, 297, 34-38。
14. 蔡松齡 (1992b): 電腦在題庫命題作業之運用(下)。師友月刊, 298, 54-58。
15. 蕭漢威 (1994): 智慧型電腦輔助教學之評估模式。新竹市: 國立交通大學資訊科學研究所碩士論文(未出版)。
16. Alessi, S. M., & Trollip S. R. (1991). *Computer-based instruction: Methods and development* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
17. Dennis, V. E., & Gruner D. (1992). Computer managed instruction at Arthur Anderson & Company: A status report. *Educational Technology*, Vol. 32, No. 8, 7-16.
18. Fan, Joshua Poh-Onn, Mak, Tina Kwai-Lan, & Shue, Li-Yen (1996). *Development of a knowledge based computer assisted instruction system*. Paper presented at the 1996 International Conference Software Engineering: education and Practice, Dnnedin, New Zealand.
19. Tyler, R. W. (1950). *Basic principles of curriculum and instruction*. Chicago: University of Chicago Press.



Analysis and Improvement of Test Items for a Network-based Intelligent Testing System**

Gwo-Jen Hwang^{1*}, Judy C. R. Tseng², Carol Chu¹,
and Jing-Wu Shiau¹

¹Information Management Department, National Chi Nan University,
Pu-Li, Nan-Tou, Taiwan 545, R.O.C.

²Department of Computer Science and Information Engineering,
Chung Hua University, Hsinchu, Taiwan, 300, R.O.C.

Abstract

During the learning process, it is important to evaluate the learning status of the students, especially in a network-based learning environment that provides personalized subject materials. To have an integrated and shared test bank, it is necessary to have several educational experts cooperate through computer networks. In some previous work, researchers have proposed techniques to cope with redundancy, consistency and integrity problems when educators provide test items through computer networks. However, the correctness and accuracy of the redundancy and consistency checking are not desirable, which will significantly affect the quality of the item bank. In this paper, we propose a semantic analysis-based method to cope with these problems. Some experiments on "Biology" course of senior high schools have shown that our approach can achieve more accurate results.

Key words : Network-based Testing, Item Bank, Redundancy Analysis, Consistency Analysis, Distance Learning.

* To whom all correspondences should be sent.

** This study is supported by the National Science Council of the Republic of China under contract number NSC- 90-2520-S-260-002.

